APPLICATION OF RIGOROUS HIGH-ORDER METHODS AND NORMAL FORMS TO NONLINEAR SYSTEMS

By

Adrian Weisskopf

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Physics – Doctor of Philosophy

2021

ABSTRACT

APPLICATION OF RIGOROUS HIGH-ORDER METHODS AND NORMAL FORMS TO NONLINEAR SYSTEMS

By

Adrian Weisskopf

The nonlinearities of dynamical systems often display the most interesting and fascinating behavior. At the same time, those nonlinearities complicate finding closed form analytic solutions, especially for complex systems, to the point where it is often impossible. Differential algebra (DA) based methods allow us to analyze those systems with all their nonlinearities up to arbitrary order in an automated, computer based framework that operates with floating point accuracy.

This thesis will investigate repetitive dynamical systems from seemingly unrelated fields of study using DA methods such as DA based transfer and Poincaré maps, the DA normal form algorithm, normal form defect studies, and verified methods based on Taylor Models. The common mathematical underpinnings of those dynamical systems allow us to analyze them with different techniques that have the same methods at their core.

Specifically, we will analyze resonances, associated fixed point structures, and oscillation periods of particles in the accelerator storage ring of the Muon g-2 Experiment at Fermilab to gain a detailed understanding of the stability of the system and the potential loss mechanism of particles. If successful, the Muon g-2 Experiment raises existential questions about the completeness of the Standard Model of particle physics, which makes our efforts to understand the stability of the system highly relevant.

The same methods used for the analysis of the accelerator storage ring will also be used to generate far reaching sets of satellite orbits for formation flying missions under the Earth's gravitational zonal perturbations. Our approach is particularly elegant and precise, and its theoretical limits are beyond the range of practical applications.

One central method in both of those analyses is the DA normal form algorithm. Using the mechanical device of the centrifugal governor as an illustrative example problem, the special

properties of the resulting normal form, the sensitivities and limitations of the algorithm, and its resulting quantities are explained in detail.

We also will provide first results and an outlook for future work of the presented methods in the realm of verified methods, and illustrate the current possibilities as well as future opportunities and challenges. In particular, Taylor Model based verified global optimization is introduced and used to calculate rigorous stability estimates for different configurations of the Muon g-2 Storage Ring.

To my parents and grandparents.

ACKNOWLEDGEMENTS

First of all, I would like to thank my academic advisor Professor Martin Berz for his continuous support, patience, and helpful guidance not only in my research, but also during my Ph.D. time in general. I really enjoyed diving into complex problems with him and discovering the key mechanisms at play. His approach of analyzing the fundamental components of a problem strongly influenced my way of structuring problems and attempting their solution.

Furthermore, I particularly appreciated the collaborative work with Roberto Armellin, David Tarazona, Kyoko Makino, and Eremey Valetov and would like to thank all of them for the insightful discussions and welcoming atmosphere on and off work. I am proud of our joint contributions to the scientific community and really enjoyed the process of getting there. I would like to additionally thank Kyoko for her remarkable support and attention to detail, which contributed to this work.

I also shall not forget to thank Scott Pratt, Mark Dykman, Vladimir Zelevinsky, together with Martin and Kyoko for being so kind and agreeing to serve on my thesis committee.

Last but not least, I cannot thank enough my girlfriend Franzi, my family and friends for their continuous support during my Ph.D. time. This thesis and Ph.D. was only possible due to the generous scholarship by the Studienstiftung des deutschen Volkes. Additionally, it was supported by the US Department of Energy (DOE) and the Mathematical Sciences Research Institute (MSRI) in Berkeley. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xiv

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 METHODS 5
- 2.1 The Differential Algebra (DA) Framework 5
- 2.2 DA Transfer Maps and Poincaré Maps 6
- 2.3 The DA Normal Form Algorithm 8
  - 2.3.1 Tunes, Tune Shifts, and Normal Form Radii 16
  - 2.3.2 Resonances 17
- 2.4 The Normal Form Defect 19
- 2.5 Verified Computations Using Taylor Models (TM) 22
  - 2.5.1 Interval Arithmetic 22
  - 2.5.2 Taylor Models 24
- 2.6 Taylor Model based Verified Global Optimizers 27

CHAPTER 3 AN EXAMPLE-DRIVEN WALK-THROUGH OF THE DA NORMAL FORM ALGORITHM 30
- 3.1 The Centrifugal Governor 30
  - 3.1.1 Units 31
  - 3.1.2 The Equilibrium Point 31
  - 3.1.3 The Equations of Motion 33
  - 3.1.4 Illustration of System Dynamics 35
- 3.2 Map Calculation via Integration 37
- 3.3 The DA Normal Form Algorithm 38
  - 3.3.1 The Parameter Dependent Fixed Point 40
  - 3.3.2 The Linear Transformation 41
  - 3.3.3 The Nonlinear Transformations 44
    - 3.3.3.1 General *m*th Order Nonlinear Transformation 44
    - 3.3.3.2 Explicit Second Order Nonlinear Transformation 46
    - 3.3.3.3 Explicit Third Order Nonlinear Transformation 49
    - 3.3.3.4 The Effect of the Second Order Transformation on Third Order Terms 50
  - 3.3.4 Transformation back to Real Space Normal Form 52
  - 3.3.5 Invariant Normal Form Radius 55
  - 3.3.6 Angle Advancement, Tune and Tune Shifts 56
- 3.4 Visualization of the Different Order Normal Forms and Conclusion 58

CHAPTER 4 BOUNDED MOTION PROBLEM 61
- 4.1 Introduction to Bounded Motion 61

4.2 Understanding Orbital Motion Under Gravitational Perturbation 64
4.2.1 The Perturbed Gravitational Potential 64
4.2.2 The Equations of Motion 65
4.2.3 The Kepler Orbit 66
4.2.4 Orbits Under Gravitational Perturbation 67
4.2.5 The Bounded Motion Conditions by Xu *et al.* 69
4.2.6 The Fixed Point Orbit 70
4.3 Method of Bounded Motion Design Under Zonal Perturbation 71
4.3.1 The Poincaré Surface Space 72
4.3.2 The Fixed Point Orbit 72
4.3.3 The Calculation of Poincaré Return Map 73
4.3.4 The Normal Form Averaging 74
4.4 Bounded Motion Results 76
4.4.1 Bounded Motion in Low Earth Orbit 77
4.4.2 Bounded Motion in Medium Earth Orbit 82
4.4.3 Testing the Limitations of the DANF Method 86
4.5 Conclusion 89

CHAPTER 5 STABILITY ANALYSIS OF MUON g-2 STORAGE RING 92
5.1 Introduction 92
5.2 Storage Ring Simulation Using Poincaré Maps 95
5.3 The Closed Orbit 97
5.3.1 The Closed Orbit Under Perturbation 97
5.3.2 The Momentum Dependence of the Closed Orbit 100
5.3.3 The Relevance of Closed Orbits 101
5.4 Tune Analysis 103
5.4.1 Tunes of the Momentum Dependent Closed Orbit 103
5.4.2 The Amplitude Dependent Tune Shifts 105
5.4.3 The Tune Footprint 111
5.5 Stability and Loss Mechanisms 113
5.5.1 The Normal Form Defect of Tracked Particles 115
5.5.2 Lost Muon Studies 117
5.5.3 Period-3 Fixed Point Structures 134
5.5.4 Muon Loss Rates from Simulation 138
5.6 Conclusion 141

CHAPTER 6 VERIFIED CALCULATIONS USING TAYLOR MODELS 143
6.1 The Rosenbrock Optimization Problem 144
6.1.1 The Rosenbrock Function 144
6.1.2 Global Optimization Using COSY-GO 146
6.1.2.1 Illustration of the Cluster Effect and Dependency Problem using the 2D Rosenbrock Function 147
6.1.2.2 Performance of COSY-GO for High Dimensional Rosenbrock Function 150
6.2 The Lennard-Jones Potential Problem 152

6.2.1 Introduction 152
6.2.1.1 The Lennard-Jones Potential U_{LJ} 153
6.2.1.2 Configurations of Particles $\mathcal{S}_k$ 154
6.2.1.3 The Lennard-Jones Optimization Problem and its Challenges . . 155
6.2.2 Minimum Energy Lennard-Jones Configurations in 1D 157
6.2.2.1 Coordinate System, Numbering Scheme, and Variable Definition in 1D 157
6.2.2.2 Upper Bounds $r_{k,\mathrm{UB}}$ and $v_{x,\mathrm{UB}}$ on Inter-Particle Distances of Minimum Energy Configurations in 1D 158
6.2.2.3 The Upper Bound $U_{k,\mathrm{UB}}$ on the Minimum Energy 159
6.2.2.4 The Lower Bound $r_{k,\mathrm{LB}}$ on r_{ij} 160
6.2.2.5 The Infinite 1D Equidistant Configuration 161
6.2.2.6 The Verified Global Optimization Results for Configurations of k Particles in 1D 163
6.2.2.7 The Verified Global Optimization Results for Symmetric Configurations of k Particles in 1D 169
6.2.2.8 Redundancies and Penalty Functions 175
6.2.3 Minimum Energy Lennard-Jones Configurations in 2D and 3D 176
6.2.3.1 Coordinate System, Numbering Scheme, and Variable Definition 176
6.2.3.2 Upper Bounds $r_{k,\mathrm{UB}}$ and $v_{x,\mathrm{UB}}$ on Inter-Particle Distances of Minimum Energy Configurations in 2D and 3D 179
6.2.3.3 The Upper Bound $U_{k,\mathrm{UB}}$ on the Minimum Energy 180
6.2.3.4 Setup of Initial Search Domain of the Optimization Variables in 2D and 3D 181
6.2.3.5 The Evaluation of the Objective Function 183
6.2.3.6 Taylor Model Evaluation of Piecewise Defined Functions 184
6.2.3.7 The Verified Global Optimization Results for Configurations of k Particles in 2D 187
6.2.3.8 The Verified Global Optimization Results for Configurations of k Particles in 3D 197
6.2.4 Summary 201
6.3 Verified Stability Analysis of Dynamical Systems 202
6.3.1 The Potential Implications for the Bounded Motion Problem 202
6.3.2 The Implications for the Stability Analysis of the Muon g-2 Storage Ring . 204
6.3.3 The Normal Form Defect as the Objective Function for the Optimization . . 206
6.3.4 The Search Domain in the Form of Onion Layers 206
6.3.5 The Complexity and Nonlinearity of the Normal Form Defect Function . . 208
6.3.6 The Results of the Verified Global Optimization of the Normal Form Defect 212
6.3.7 Comparison of Verified Nekhoroshev-type Stability Estimates to Actual Rates of Divergence in Example Island Structure 218
6.3.8 Relevance of the ESQ Voltage on the Stability 219
6.3.9 Comparison of Nonverified and Verified Normal Form Defect Analysis . . 224
6.3.10 The Analysis of the Effect of Normal Form Transformations of Different Order on the Normal Form Defect 229

CHAPTER 7 CONCLUSION 241

APPENDIX 244

BIBLIOGRAPHY 246

LIST OF TABLES

Table 3.1: List of stable equilibrium angles ϕ_0 of the centrifugal governor arms for some specific rotation frequencies ω. 33

Table 3.2: Integration result for map around equilibrium state $(\phi_0(\omega = \sqrt{2}) = \pi/3, 0)$ integrated until $t = 1$ using an order 20 Picard-iteration based integrator with stepsize $h = 10^{-3}$ over 1000 iterations within COSY INFINITY. The component $\mathcal{M}_0^+ = Q(q_0, p_0)$ is on the left, $\mathcal{M}_0^- = P(q_0, p_0)$ on the right. 39

Table 3.3: Coefficients of $\mathcal{M}_1$ up to order three. Note the complex conjugate property $\mathcal{S}^{\pm}_{m(k_+,k_-)} = \bar{\mathcal{S}}^{\mp}_{m(k_-,k_+)}$. 43

Table 3.4: The values of the $\mathcal{T}^{\pm}_{2(k_+,k_-)}$ and $\mathcal{O}^{\pm}_{3(k_+,k_-)}$. Note that $\mathcal{T}_2$ and $\mathcal{O}_3$ and therefore $\mathcal{A}_2$ and its inverse are real with $\mathcal{A}^+_{m(k_+,k_-)} = \mathcal{A}^-_{m(k_-,k_+)}$. 47

Table 3.5: New coefficients of third order of $\mathcal{M}_2$ after the second order transformation. Note that the first order terms remain unchanged and that the second order terms are all eliminated by the second order transformation. Interestingly, the second order transformation caused some terms of the third order to disappear in this specific case, which is not a general property of the second order transformation. The emphasized terms are surviving the third order transformation as explained in the following subsection. 48

Table 3.6: The values of the $\mathcal{T}^{\pm}_{3(k_+,k_-)}$. The values for $\mathcal{T}^+_{3(2,1)}$ and $\mathcal{T}^-_{3(1,2)}$ cannot be calculated because the denominator in Eq. (3.36) is zero. Note that $\mathcal{T}^+_{3(k_+,k_-)} = \mathcal{T}^-_{3(k_-,k_+)}$. 49

Table 3.7: The normal form map $\mathcal{M}_{\text{NF}}$ up to order three. The component $\mathcal{M}^+_{\text{NF}}$ is on the left, $\mathcal{M}^-_{\text{NF}}$ on the right. 53

Table 3.8: The normal form transformation $\mathcal{A}$ up to order three. The component $\mathcal{A}^+$ is on the left, $\mathcal{A}^-$ on the right. 54

Table 3.9: Tune and coefficients of amplitude and parameter $\delta\omega$ dependent tune shifts for centrifugal governor with $\omega_0 = \sqrt{2}$. 58

Table 4.1: The expansion of $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ for relative bounded motion orbits with an average nodal period $\overline{T_d} = 7.64916169$ (≈103 min) and an average ascending node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad. The expansion is relative to the pseudo-circular LEO from [42]. 78

Table 4.2: The LEOs below are all initiated at $v_{r,0} = -1.05621369\text{E-}3$ and $r_0 = 1.14016749 + \delta r$, and have an average nodal period of $\overline{T_d} = 7.64916169$ ($\approx$103 min) and an average ascending node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad. The pseudo-circular LEO from [42] is denoted by $\mathcal{O}_0$. 79

Table 4.3: Expansion of $\omega_p(\delta r, \delta v_r = 0)$ of relative bounded motion LEOs with an average nodal period $\overline{T_d} = 7.64916169$ ($\approx$103 min) and an average node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad. The expansion is relative to the pseudo-circular LEO from [42]. 81

Table 4.4: The expansion of $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ for relative bounded motion MEOs with an average nodal period of $\overline{T_d} = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The expansion is relative to the pseudo-circular MEO from [10]. 83

Table 4.5: The MEOs below are all initiated at $v_{r,0} = -1.14150072\text{E-}4$ and $r_0 = 4.17198963 + \delta r$, and have an average nodal period of $\overline{T}_d = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The orbit $\mathcal{O}_0$ is the pseudo-circular MEO from [10]. 83

Table 4.6: Expansion of $\omega_p(\delta r, \delta v_r = 0)$ of relative bounded motion orbits with an average nodal period of $\overline{T}_d = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The expansion is relative to the pseudo-circular MEO from [10]. 84

Table 4.7: The following orbit parameters are obtained by evaluating $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ from Tab. 4.1 and Tab. 4.4 for various δr keeping $\delta v_r = 0$. 86

Table 5.1: Percentages of different characterization groups. Read as follows: x % of *Base* particles have the property *Property*. All particles that hit a collimator during the 4500 turns of tracking are considered lost. 140

Table 6.1: Verified global optimization results for $U^\star_{k,1\text{D}}$. The k particles form n_pairs pairwise interactions. The upper bound $U_{k,1\text{D},\text{UB}}$ on the minimum energy (see Sec. 6.2.2.3) was used to calculate $r_{k1\text{D},\text{UB}}$, which sets the lower bound of the initial search domain (see Sec. 6.2.2.4 and Eq. (6.45)). The optimizer COSY-GO with QFB/LDB enabled was operated with Taylor Models of third order. The number of remaining boxes with all side-lengths $s < s_\text{min}$ is denoted by n_fin. 165

Table 6.2: Verified global optimization results for configurations of k particles in 1D. The variable $v^\star_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, with $1 \leq i < k$, and the mirror symmetry can be observed. 166

Table 6.3: Performance of verified global optimization using COSY-GO with QFB/LDB enabled on minimum energy search of a 1D configuration of k particles. The Taylor Model orders are denoted by 'O'. Since QFB requires a minimum order of two, order one calculations are not listed. 168

Table 6.4: Verified global optimization results on the minimum energy $U^{\star}_{k,\mathrm{1D}}$ of symmetric configurations. The upper bound $U_{k,\mathrm{1D,UB}}$ on the minimum energy (see Sec. 6.2.2.3) was used to calculate $r_{k\mathrm{1D,UB}}$, which sets the lower bound of the initial search domain (see Sec. 6.2.2.4 and Eq. (6.45)). The optimizer COSY-GO with QFB/LDB enabled was operated with Taylor Models of third order. The number of remaining boxes with all side-lengths $s < s_{\mathrm{min}}$ is denoted by n_{fin}. 170

Table 6.5: Verified global optimization results for symmetric configurations of k particles in 1D for $k = 3$ to $k = 20$. $v^{\star}_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, and p_{k-i-1} and p_{k-i}. The results for $k = 21$ to $k = 26$ are listed in Tab. 6.6. 171

Table 6.6: Verified global optimization results for symmetric configurations of k particles in 1D for $k = 21$ to $k = 26$. $v^{\star}_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, and p_{k-i-1} and p_{k-i}. The results for $k = 3$ to $k = 20$ are listed in Tab. 6.5. 172

Table 6.7: Performance of verified global optimization using COSY-GO with QFB/LDB enabled for the minimum energy search of a 1D symmetric configuration of k particles. The Taylor Model orders are denoted by 'O'. Note that we use n_{var} as a shorthand notation for $n_{\mathrm{1D,sym,var}}$ in the table. 174

Table 6.8: Verified global optimization results for the minimum energy configurations of four particles in 2D, $\mathcal{S}^{\star}_{4,\mathrm{2D}}$. The $r^{\star}_{ij}$ yield the optimal distances between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1}, and $v^{\star}_{y,i}$ is the optimal y position of particle p_i. 190

Table 6.9: Verified global optimization results for the minimum energy configurations of five particles in 2D, $\mathcal{S}^{\star}_{5,\mathrm{2D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i. 192

Table 6.10: Performance of verified global optimization using COSY-GO with QFB/LDB enabled on minimum energy search of a 2D configuration of k particles. The Taylor Model orders are denoted by 'O'. 192

Table 6.11: Performance of verified global optimization using parallel COSY-GO with QFB/LDB enabled for minimum energy search of a 2D configuration of six particles, $\mathcal{S}^{\star}_{6,\mathrm{2D}}$. The parallel computations are run on Cori at NERSC using different communication timing t_{com}. 193

Table 6.12: Verified global optimization results for the minimum energy configurations of six particles in 2D, $\mathcal{S}^{\star}_{6,\mathrm{2D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i. 195

Table 6.13: Verified global optimization results for the minimum energy configurations of seven particles in 2D, $\mathcal{S}^{\star}_{7,\mathrm{2D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i. The values below are for the configuration $(\kappa, \nu) = (5, 6)$ from Fig. 6.23. 197

Table 6.14: Verified global optimization results for the minimum energy configurations of five particles in 3D, $\mathcal{S}^{\star}_{5,\mathrm{3D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1}. $v^{\star}_{y,i}$ and $v^{\star}_{z,i}$ are the optimal y and z positions of particle p_i. 200

Table 6.15: Summary of verified global optimization results on the minimum energy of configurations in 2D and 3D. 201

Table 6.16: Analysis of the normal form radius range of the five island particles in Fig. 5.30 and the number of turns it requires to get from the lower end of the range to the upper end. The islands are numbered from smallest (1) to largest (5). 218

Table 6.17: Number of turns required by the islands from Fig. 5.30 to cross the given onion layers in $r_{\mathrm{NF},2}$ direction. The islands are numbered from smallest (1) to largest (5). Additionally, the minimum number of turns required to cross the onion layer determined by the verified analysis is shown. 219

Table A.1: The left columns list the variables of [59], denoted by $v'_{\cdot,i}$, and their respective initial search domains. The middle columns list the variables of our optimization and their respective initial search domains. The right columns show the optimized variables of our optimization. 245

LIST OF FIGURES

Figure 2.1: Schematic illustration of the various normal form quantities involved in the calculation of the minimum iteration number within allowed region $\mathbb{D}$. 21

Figure 2.2: Verified representation of $f(x) = \sin(\pi x/2) - \exp(x)$ over the domain $\mathbb{D} = I_1 = [-1, 1]$ with interval methods using $f(\mathbb{D})$ and with Taylor Models $(\mathcal{P}_{m,f}, \epsilon_{m,\mathbb{D},f})$ of various orders m. The original function $f(x)$ is indicated by the black line, while its DA polynomial representation is shown in green. The bounds at a distance $\epsilon_{m,\mathbb{D},f}$ from the DA polynomial are red. The two straight blue lines indicate the bounds of the interval evaluation. Note that the scale of the y axis is changing to better illustrate the tightness of the Taylor Model representation with higher orders. Accordingly, the interval bounds are only shown for order $m = 1$ and order $m = 2$. 26

Figure 3.1: Schematic illustration of centrifugal governor. 30

Figure 3.2: Illustration of the stable equilibrium angle ϕ_0 of the arms of the centrifugal governor as a function of the rotation frequency ω. For $\omega > \omega_{\min} = 1$, $\phi_0 = 0$ is an unstable equilibrium angle. 32

Figure 3.3: Potential well of U_{eff} for multiple oscillation frequencies ω. The equilibrium angle ϕ_0 corresponds to the minimum of the potential well. 34

Figure 3.4: Dynamics of the centrifugal governor for a rotation frequency of $\omega = \sqrt{2}$. The centrifugal governor arms were initiated with $\dot{\phi} = p_\phi = 0$ and at the following angles: 60°, 65.5°, 69.5°, 73.5°, 77.5°, 81.5°, 85.5°, and 89.5°. The left plot shows the oscillatory behavior around the equilibrium angle at $\phi_0 = 60°$ over time. The right plot shows the stroboscopic phase space behavior from repetitive map evaluation. To relate the phase space behavior to the position behavior in time, the ϕ axis of both plots are aligned. 35

Figure 3.5: Phase space behavior of the centrifugal governor arms around their equilibrium angle of $\phi_0(\omega = \sqrt{2}) = 60°$ provided by a tenth order Poincaré map of the system. a) shows the original phase space behavior. b) shows the associated circular behavior in normal form. 39

Figure 3.6: Comparison between the calculated period with normal form methods $T_{\text{NF}} = 1/\nu(q_0, p_0)$ for calculation order ten (O10) and calculation order three (O3) to the actual period of oscillation given by the oscillatory behavior of the centrifugal governor arms for $\omega = \sqrt{2}$ from Fig. 3.4. 57

Figure 3.7: Phase space tracking of incomplete normal form maps of order ten of the centrifugal governor arms with a fixed rotation frequency of $\omega = \sqrt{2}$. The original map (a), only linear normal form transformation (b), and only normal form transformations up to order two (c) and three (d), respectively. The normal form up to the full tenth order was illustrated in Fig. 3.5. 60

Figure 4.1: The behavior of the Keplerian elements of a low Earth orbit under zonal gravitational perturbations up to J_{15} (purple) and as a regular Kepler orbit in the unperturbed gravitational field (green) over time. Left and right plots show different time scales of the behavior. Note that the vertical scales of Ω and ω are adjusted in the right plot to show the long term behavior. 68

Figure 4.2: Keplerian elements of a quasi-circular low Earth orbit under Earth's zonal gravitational perturbation. 70

Figure 4.3: a) Distorted phase space behavior in the original phase space (q, p) and b) circular behavior in the corresponding normal form phase space $(q_{\mathrm{NF}}, p_{\mathrm{NF}})$. In a), the phase space angle advancement Λ_k and the phase space radius r_i are not constant by continuously change along each of the phase space curves. In b), the phase space behavior is rotationally invariant ('normalized') with a constant radius r_{NF} and a constant but amplitude dependent angle advancement $\Lambda(r_{\mathrm{NF}})$. 75

Figure 4.4: Oscillatory behavior of the bounded motion quantities T_d and $\Delta\Omega$ of the bounded LEOs $\mathcal{O}_1$ and $\mathcal{O}_2$ initiated at $\delta r = 0.06$ and $\delta r = 0.13$, respectively. Additionally, the constant nodal period $T_d^\star = 7.64916169$ and constant ascending node drift of $\Delta\Omega^\star = 0.0704°$ of the fixed point orbit $\mathcal{O}_0$ are shown. The periods of oscillation are 1763 orbital revolutions (126 days) for $\mathcal{O}_2$, 1810 orbital revolutions (129 days) for $\mathcal{O}_1$, and 1823 orbital revolutions (130 days) for $\delta r \to 0$ of $\mathcal{O}_0$. The shown results are generated by numerical integration. . . 79

Figure 4.5: Relative bounded motion of LEOs with an average nodal period of $\overline{T_d} = 7.64916169$ ($\approx$103 min) and an average node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad for 14 years. The total relative distance between the orbits is shown in the left plot and the right plot shows the relative radial and along-track distance between orbit pairs from the perspective of one of the orbits in the pair. The oscillation in the relative distance between $\mathcal{O}_2$ and $\mathcal{O}_1$ is caused by the rotating orbital orientation of the orbits at different frequencies. 80

Figure 4.6: Oscillatory behavior of the bounded motion quantities T_d and $\Delta\Omega$ of the bounded MEOs $\mathcal{O}_1$ and $\mathcal{O}_2$ initiated at $\delta r = 0.24$ and $\delta r = 0.52$, respectively. Additionally, the constant nodal period $T_d^\star = 53.5395648$ and constant ascending node drift of $\Delta\Omega^\star = -0.0192176316$ deg of the fixed point orbit $\mathcal{O}_0$ are shown. The periods of oscillation are 38682 orbital revolutions (52.9 years) for $\mathcal{O}_2$, 34621 orbital revolutions (47.4 years) for $\mathcal{O}_1$, and 33671 orbital revolutions (46.1 years) for $\delta r \rightarrow 0$ of $\mathcal{O}_0$. The shown results are generated by numerical integration. 84

Figure 4.7: Relative bounded motion of MEOs from Tab. 4.5 with an average nodal period of $\overline{T_d} = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad over 70 years. The total relative distance between the orbits is shown in the left plots and the right plot shows the relative radial and along-track distance between orbit pairs from the perspective of one of the orbits in the pair. The ‘breathing’ of the relative total distance between $\mathcal{O}_2$ and $\mathcal{O}_0$ originates from the rotating orbital orientation of pseudo-elliptical $\mathcal{O}_2$ relative to the pseudo-circular $\mathcal{O}_0$. Due to the very long rotation periods, only the first 70 years of the relative distance oscillation and radial/along-track behavior between $\mathcal{O}_2$ and $\mathcal{O}_1$ could be shown. 85

Figure 4.8: The behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the test orbits from Tab. 4.7 of the calculated LEO bounded motion set generated by numerical integration. For large δr, the influences of higher order oscillations are apparent. The frequency and amplitude of oscillation increase with increasing δr. The amplitude of $\Delta\Omega$ is particularly sensitive to δr. 87

Figure 4.9: Distance between the orbits in the calculated bounded motion set and $\mathcal{O}_0$ is determined in regular time intervals with numerical integration over more than ten years. The left plot only shows the upper bound to avoid overlaps. Thin horizontal lines at the initial upper bound emphasize small changes. The dotted light blue curve (right) originates from an unintended near-resonance between the chosen time interval for distance evaluations and the orbital behavior. A measurable increase in relative distances (left) over 10 years for $\delta r \geq 0.3$ is supported by thickening curves in the radial/along-track behavior (right). 88

Figure 4.10: Behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the test orbits from Tab. 4.7 of the calculated MEO bounded motion set generated by numerical integration. In contrast to the investigated LEOs, the frequency and amplitude of oscillation decrease with increasing δr such that $\mathcal{O}_{1.4}$ appears almost steady. For $\delta r \geq 0.8$ the center of oscillation of $\Delta\Omega$ start to drift to more negative values and away from $\Delta\Omega^\star$. To capture both, the oscillatory behavior around $\Delta\Omega^\star$ and the drift of the center of oscillation for very large δr, two plots with a different scale and range are shown for $\Delta\Omega$. 89

Figure 4.11: Distance between the orbits in the calculated bounded motion set and $\mathcal{O}_0$ is determined in regular time intervals by numerical integration over more than 70 years. The left plot only shows the upper bound to avoid overlaps. Thin horizontal lines at the initial upper bound emphasize small changes. The 'breathing' of the total relative distance from the orbital rotation is clearly visible. Its period increases with increasing δr until being unrecognizable due to the strong divergence for $\delta r \geq 1.4$, which is supported by thinker curves in the right plot. The weaker divergence over the 70-year timespan is already noticeable for $\delta r \geq 0.9$. The divergence is caused by the offset in respective bounded motion quantities (see Fig. 4.10). 90

Figure 5.1: The fixed points of Poincaré return maps from various azimuthal locations around the ring indicate the behavior of the closed orbit (for $\delta p = 0$). The projections of the four dimensional fixed points into subspaces illustrate the influence of the magnetic field perturbations on the closed orbit around the ring. The results from the five collimator locations (C1-C5) are highlighted with red color. The x coordinate corresponds to displacements in the radially outward direction, while the y coordinate indicates the displacement in the vertically upward direction. 99

Figure 5.2: Changes of the closed orbits due to relative changes δp in the total initial momentum. The plots illustrate absolute coordinates with respect to the ideal orbit at the center of the ring for the five collimator locations (C1-C5). 100

Figure 5.3: Phase space behavior of four particles in different phase space regions with various amplitudes and momentum offsets. Particle 4 (yellow) hits the collimator (circle in the xy plot) and is lost. The momentum dependent radial position x of the particles is particularly prominent. The individual particles are characterized by the parameter set $\left(x_{\text{amp}}, y_{\text{amp}}, \delta p\right)$. For particle 1 (P1) the parameter set is $(6\,\text{mm}, 12\,\text{mm}, -0.39\%)$. For particle 2 (P2) the parameter set is $(12\,\text{mm}, 6\,\text{mm}, -0.39\%)$. For particle 3 (P3) the parameter set is $(27\,\text{mm}, 16\,\text{mm}, +0.13\%)$. For particle 4 (P4) the parameter set is $(6\,\text{mm}, 25\,\text{mm}, +0.39\%)$. 102

Figure 5.4: Schematic illustration of viable xy region around a momentum dependent fixed point. The region contains all rectangles centered at the fixed point, which do not overlap with the collimator circle. 103

Figure 5.5: Vertical and horizontal tune dependence in the model of the Muon g-2 Storage Ring of E989 on relative offsets δp from the reference momentum p_0. 104

Figure 5.6: Amplitude dependent tune shifts in the model of the Muon g-2 Storage Ring of E989. The black line indicates the amplitude dependent tune shifts for $\delta p = 0$, while the other lines have a momentum offset specified by their color. For the left plots regarding the radial amplitude dependence, the vertical amplitude relative to the momentum dependent fixed point is set to zero and vice versa for the plots regarding the vertical amplitude dependence on the right. The lines end when the total xy amplitude of the particle relative to the ideal orbit reaches the collimator at $r_0 = 45$ mm. 106

Figure 5.7: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV. 108

Figure 5.8: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV. 109

Figure 5.9: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV. 110

Figure 5.10: Projections of the distribution of the variables $(x, a, y, b, \delta p)$ in the realistic beam simulation at the azimuthal ring location of the central kicker. 112

Figure 5.11: The tune footprint of a realistic beam distribution at the azimuthal ring location of the central kicker. The tune footprint from the 10th order calculation is colored according to the momentum offset of the individual particles. The black lines correspond to resonance conditions. In a) the 8th order calculation (green) is overlaid to illustrate the drastic influence of the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential. In b) the particles with a momentum offset $-0.3\% < \delta p < 0.1\%$ are overlaid in green. In c) the particles with a momentum offset $0.1\% < \delta p < 0.28\%$ are overlaid in green. In d) the particles with a momentum offset $0.28\% < \delta p < 0.5\%$ are overlaid in green. 113

Figure 5.12: The left plot shows in violet the ratio of particles that have a larger $d_{\text{NF,max}}$ then the corresponding x value. The green boxes indicate the ratio of particles lost in each $d_{\text{NF,max}}$ group, e.g., particles that encounter a maximum normal form defect larger than $2^{-7} = 7.8 \cdot 10^{-3}$ are all lost (loss ratio of 1). The right plot shows the ratio between the normal form radius range over 4500 turns ($d_{\text{NF,lt}}$) and the maximum encountered normal form defect $d_{\text{NF,max}}$ for each particle. The particles are grouped into surviving and lost particles. 115

Figure 5.13: The plots show the long term normal form defect dependent on the calculated tune range of each particle. The dots are the minimum calculated tune of each particle while tracking. Red dots indicate that the respective particle is lost over the 4500 tracking turns. The gray lines show the calculated tune range of each particle. The left plot illustrates the radial long term normal form defect with respect to the radial tune and the 17/18 resonance (green line). The right plot shows the vertical long normal form defect with respect to the vertical tune and the 1/3 resonance (green line). 116

Figure 5.14: The tune range of the particles forming the spike in Fig. 5.13 are shown on the left. The right plot shows the normal form defect of the particles depends on their closeness to the $6\nu_x + 4\nu_y = 7$ resonance (green line). 117

Figure 5.15: The radial and vertical phase space behavior indicates that this particle ($\delta p = 0.015\%$) oscillates at constant amplitudes around its momentum dependent reference orbit. The overall normal form radius is constant and confirms this. Accordingly, the tune footprint of the particle is a single dot. This is a trivial large amplitude loss. 119

Figure 5.16: The vertical phase space behavior of this particle ($\delta p = 0.196\%$) has a slight triangular deformation. The overall normal form radius indicates a modulated amplitude and the spread out tune footprint starts right after the vertical 1/3-resonance line. Despite slight influence of the resonance, the rather elliptical phase space behavior makes this a trivial large amplitude loss. 120

Figure 5.17: This particle ($\delta p = -0.088\%$) is caught around a period-3 fixed point structure in the vertical phase space, which is related to the vertical 1/3-resonance. We refer to these structures as islands and the loss mechanisms is called island related loss. 121

Figure 5.18: This particle ($\delta p = -0.015\%$) forms large islands around a period-3 fixed point structure in the vertical phase space, which is associated with a major modulation of the oscillation amplitude. 122

Figure 5.19: This particle ($\delta p = -0.127\%$) jumps between the islands. The large radial amplitude and/or the closeness to the $(17/18, 1/3)$ resonance point might have triggered the jump. This is an example of moderate unstable behavior around a period-3 fixed point structure. 123

Figure 5.20: This particle ($\delta p = 0.024\%$) shows a different kind of moderate unstable behavior around a period-3 fixed point structure, where the island size varies. The particle has both, a large radial amplitude and the closeness to the $(17/18, 1/3)$ resonance point. 124

Figure 5.21: This particle ($\delta p = 0.140\%$) forms a shuriken like shape in the vertical phase space. In this pattern there are two period-3 fixed point structures involved indicated by the double crossing of the vertical $1/3$ resonance line. 125

Figure 5.22: This particle ($\delta p = 0.196\%$) illustrates moderate unstable behavior in a shuriken pattern. The radial amplitude is not particularly large, but the resonance point $(17/18, 1/3)$ is very close, which might be the trigger of the unsuitability. The unstable behavior is also visible in the continuously increasing normal form radius. 126

Figure 5.23: This particle ($\delta p = 0.242\%$) illustrates a shuriken pattern, where the two period-3 fixed point structures are more obvious. The muon experiences a major modulation in the vertical oscillation amplitude and performs a double crossing of the vertical $1/3$ resonance line. 127

Figure 5.24: This particle ($\delta p = -0.096\%$) shows a shuriken pattern with unstable tendencies. The large radial amplitude and/or the closeness to the radial $17/18$ resonance line might be the trigger for the instability. 128

Figure 5.25: This particle ($\delta p = -0.159\%$) shows a shuriken pattern with a moderate instability. The two period-3 fixed point structures are so close together that the particle gets temporarily caught around the inner one of them in an island pattern. 129

Figure 5.26: This particle ($\delta p = 0.181\%$) shows the pattern of a very blunt shuriken. The vertical amplitude oscillation is only moderate and illustrates there can be almost regular behavior between two period-3 fixed point structures. 130

Figure 5.27: This particle ($\delta p = 0.106\%$) is characterized by a very large vertical amplitude, which is additionally modulated by the shuriken pattern. Its one of the very few particles for which the orbit considerably overlaps with the collimator boundary. 131

Figure 5.28: This particle ($\delta p = 0.118\%$) shows strong instabilities caused by a combination of a very large vertical amplitude in combination with a period-3 fixed point structure, which occasionally captures the orbit in an island pattern. 132

Figure 5.29: This particle ($\delta p = 0.010\%$) diverges due to its unstable orbit. The approach of the unstable fixed point with such a large vertical amplitude are likely the trigger of the divergence. 133

Figure 5.30: The left plot shows stroboscopic tracking in the vertical phase space illustrating orbit behavior with a single period-3 fixed point structure present. The orbits only differ in their vertical phase space behavior – they all have the same momentum offset of $\delta p = 0.126$ % and are at the momentum dependent equilibrium point in radial phase space ($x = 10.64$ mm, $a = 0.045$ mrad) having no radial oscillation amplitude. The blue orbits indicate the island patterns around the stable fixed points in the middle of the islands. The red orbits are right at the edge before being caught around the fixed points. The three unstable fixed points are in the space between the two red orbits, where the islands almost touch. In the right plot, the attractive (green) and repulsive (violet) eigenvectors of the linear dynamics around the unstable fixed points are schematically shown. 135

Figure 5.31: Stroboscopic tracking in the vertical phase space illustrating orbit behavior with two period-3 fixed point structures present. The orbits in each plot only differ in their vertical phase space behavior. All orbits have the same momentum offset of $\delta p = 0.339$ %. The four plots differ by their radial amplitude around the momentum dependent equilibrium point in radial phase space at ($x = 27.7$ mm, $a = 0.144$ mrad). The radial amplitudes are: a) $x_{\text{amp}} = 6$ mm, b) $x_{\text{amp}} = 4.8$ mm, c) $x_{\text{amp}} = 4$ mm, d) $x_{\text{amp}} = 1$ mm. The blue orbits indicate the island patterns around the stable fixed points. The red orbits are right at the edge before being caught around the period-3 fixed points. The green orbits are caught around both period-3 fixed point structures. The gray orbits in d) emphasize that half of the fixed points from c) have indeed been annihilated. 137

Figure 5.32: a) Shows how the muon loss ratio is composed of regular particles (purple) and particles involved with period-3 fixed point structures (green). Of the period-3 particles (green), the fraction caught in islands structures is indicated by the blue stripe pattern. In b) the loss ratio over time is shown for each subgroup of lost particles to better understand which losses drive to overall loss from plot a). The tracking starts after the initial 30 μs of final beam preparation when data taking is initiated. 141

Figure 6.1: A contour plot of the Rosenbrock function with $(a, b) = (1, 100)$. 144

Figure 6.2: Projections of the multidimensional generalizations of the Rosenbrock function (Eq. (6.2)) into 2D-subspaces around minimum at $\vec{x} = (1, 1, ..., 1)$, i.e., all variables are equal one except for the ones shown in the respective plot. 146

Figure 6.3: Verified global optimization of the 2D Rosenbrock function using COSY-GO in different operation modes with fourth order Taylor Models for all modes except interval evaluations (IN). 148

Figure 6.4: No cluster effect for the COSY-GO operating mode QFB/LDB, but a significant cluster effect for the IN evaluation. 149

Figure 6.5: Splitting comparison between fourth order Taylor Model approach with QFB/LDB enabled and interval evaluation using the example of the modified 2D Rosenbrock function. 150

Figure 6.6: Time consumption and number of steps in the optimization of the regular n dimensional Rosenbrock function from Eq. (6.2) at various orders with COSY-GO and QFB/LDB enabled. Additionally, the interval evaluation performance is also shown for comparison. 151

Figure 6.7: Time consumption and number of steps in the optimization of the n dimensional Rosenbrock function with an additional artificial dependency problem $f = f_{\mathrm{2D}} - f_{\mathrm{2D}} + f_{\mathrm{2D}}$ for various Taylor Model orders with COSY-GO and QFB/LDB enabled. 152

Figure 6.8: The Lennard-Jones potential for a pairwise interaction between two particles as defined in Eq. (6.8). The potential well around the minimum is shown on the left, while the right plot emphasizes its shallowness compared to the steep potential wall for $r < 1$. The potential is offset for the convenience of calculation so that the single minimum $U^{\star}_{\mathrm{LJ}}$ has an energy of zero. 154

Figure 6.9: The particles p_i are numbered according to their x position. The variable $v_{x,l}$ denotes the distance between particle p_l and p_{l+1}. A configuration with any $v_{x,l} > 1$ (left picture) is never optimal, because the U_k can always be lowered by setting $v_{x,l} = 1$ (right picture). 159

Figure 6.10: The relation between U_{LJ} and the corresponding inter-particle distance(s). Note that $r_{\max}(U_{\mathrm{LJ}})$ is only defined for $U_{\mathrm{LJ}} \leq 1$. $r_{\min}(U_{\mathrm{LJ}})$ is only decreasing very slowly with increasing U_{LJ} as the logarithmic plot on the right shows. . . 160

Figure 6.11: The plots show the values of the optimization variables $v^{\star}_{x,i}$ of the minimum energy configuration of k particles in 1D that resulted from the verified global optimization using COSY-GO. The minimum energy configuration is mirror symmetric with the middlemost distances between adjacent particles asymptotically approaching $r_{\star} \approx 0.998724135$, the solution of the infinite equidistant configuration from Eq. (6.41). The right plot shows the logarithm of the difference between the calculated distances from the verified optimization and $r_{\star}$. The ranges reflect the side-length of the remaining box. 167

Figure 6.12: The plots show the values for the optimized variables $v^{\star}_{x,i}$ of the symmetric minimum energy configuration of k particles that resulted from the verified global optimization. Again, the middlemost distances asymptotically approach $r_{\star} \approx 0.998724135$, the solution of the infinite equidistant configuration from Eq. (6.41). The right plot shows the logarithm of the difference between the calculated distances from the verified optimization and $r_{\star}$. The ranges reflect the side-length of the remaining box. 173

Figure 6.13: Performance of the minimum energy search for configurations of k particles in 1D using COSY-GO with different Taylor Model orders with QFB/LDB enabled. The order of the Taylor Models is denoted by ‘O’. The results from both Sec. 6.2.2.6 and Sec. 6.2.2.7 (’sym’) are shown. 175

Figure 6.14: The particles p_i are numbered according to their x position. The variable $v_{x,l}$ denotes the x distance between particle p_l and p_{l+1}. A configuration with any $v_{x,l} > 1$ (left picture) is never optimal, because the overall potential can always be lowered by setting $v_{x,l} = 1$ (right picture). 180

Figure 6.15: The optimal 2D configuration of five particles is denoted by five dots. Enclose all the five particles by a 2D rectangle using the minimum and maximum coordinates in x and y, shown by a solid line rectangle. Surround the resulting rectangle with a band of width 1, and we have a initial search domain for the sixth particle (shaded area). 181

Figure 6.16: Schematic illustration of the upper bound on distance perpendicular to the x axis (major axis) due to the requirement of having the longest distance between p_1 and p_k. p_1 at $x_1 = 0$ and p_k at x_k, where the major axis length $r_{1k} = x_k$. . . . 182

Figure 6.17: Initial search domain for a configuration of k particles in 2D. Note that the initial domain width in x direction is always 1 (see Eq. (6.72)) and that the x position of particle p_i determines the starting position in x of the domain of particle p_{i+1}. Particle p_1 is fixed to the origin, particle p_2 is bound by $y_2 \leq 0$, and particle p_k has a fixed y value of zero. 183

Figure 6.18: Piecewise defined modified Lennard-Jones potential $\tilde{U}_{\text{LJ,sqr}}$ shown by the black curve. The red curves shows the Lennard-Jones potential and the green line shows the tangent of this Lennard-Jones potential of r_{sqr}. The plot shown here is an example case with $r^2_{\text{LB}} = 0.9^2$. 185

Figure 6.19: Taylor Model description of piecewise defined function. Each Taylor Model is represented by three lines as previously done in Fig. 2.2. The central curve denotes the polynomial part of the Taylor Model, while the curves above and below it indicate the bounds. 186

Figure 6.20: Minimum energy configurations of four particles in 2D, $\mathcal{S}^{\star}_{4,\mathrm{2D}}$. Interestingly, the minimum energy configuration is not a square, an obvious symmetric object with fourfold symmetry, but a rhombus with two almost equilateral triangles very slightly squished in the horizontal direction. 190

Figure 6.21: Minimum energy configuration of five particles in 2D, $\mathcal{S}^{\star}_{5,\mathrm{2D}}$. 191

Figure 6.22: Minimum energy configuration of six particles in 2D, $\mathcal{S}^{\star}_{6,\mathrm{2D}}$. 194

Figure 6.23: Minimum energy configuration of seven particles in 2D, $\mathcal{S}^{\star}_{7,\mathrm{2D}}$. The configuration is represented by two equivalent numbering schemes. 196

Figure 6.24: Minimum energy configuration of five particles in 3D, $\mathcal{S}^{\star}_{5,\mathrm{3D}}$. The configuration is shown in 2D projections, the xy plane projection (left), the xz plane projection (middle), and the yz plane projection (right). The solution consists of a central equilateral triangle spanned by the particles p_2, p_3, and p_4 in the yz plane, and one particle each centered above and below that triangle. In other words, it is similar to a double tetrahedron, which is slightly squished along the major axis (the x axis) increasing the side-length of the equilateral triangle in the middle to values slightly larger than one. The inter-particle distances are shown in Tab. 6.14. 200

Figure 6.25: The left and the middle plot show the representation of an onion layer (black region) in regular phase space coordinates. The thickness of the onion layer is determined by the range in $r_{\mathrm{NF},1}$ and $r_{\mathrm{NF},2}$ as well as the range in δp. For this particular example, we set δp to a fixed value of $\delta p = 0\%$ instead of a range. The range in the normal form radii is given by $r_{\mathrm{NF},1} \in [0.15, 0.25]$ and $r_{\mathrm{NF},2} \in [0.7, 0.75]$. Note that the thickness in $r_{\mathrm{NF},1}$ is twice the thickness in $r_{\mathrm{NF},2}$. Accordingly, the projection of the onion layer into the radial phase space (x, a) appears roughly twice as thick as the projection into the vertical phase space (y, b). 207

Figure 6.26: Normal form defect landscape of the radial phase space in $\phi_{\mathrm{NF},1}$ and $\phi_{\mathrm{NF},2}$ for fixed normal form amplitudes of $r_{\mathrm{NF},1} = 0.4$ and $r_{\mathrm{NF},2} = 0.4$, and with $\delta p = 0\%$. The underlying map considers an ESQ voltage of 18.3 kV. 209

Figure 6.27: The normal form defect landscape of the radial (left side) and vertical (right side) phase space for multiple onion layers of zero thickness, which are characterized by $(r_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \delta p)$. The top row corresponds to $(0.1, 0.2, 0.24\%)$, the middle row corresponds to $(0.2, 0.05, 0.24\%)$, and the bottom row corresponds to $(0.56, 0.72, 0.04\%)$. The underlying map considers an ESQ voltage of **18.3 kV**. 210

Figure 6.28: The normal form defect landscape of the radial (left side) and vertical (right side) phase space for multiple onion layers of zero thickness, which are characterized by $(r_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \delta p)$. The top row corresponds to $(0.1, 0.2, 0.24\%)$, the middle row corresponds to $(0.2, 0.05, 0.24\%)$, and the bottom row corresponds to $(0.56, 0.72, 0.04\%)$. The underlying map considers an ESQ voltage of **17.5 kV**. 211

Figure 6.29: **Nonverified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 214

Figure 6.30: **Verified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 215

Figure 6.31: **Verified** normal form defect for the phase space storage regions of the Muon g-2 Storage ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 216

Figure 6.32: **Nonverified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. 217

Figure 6.33: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV. 221

Figure 6.34: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV. 222

Figure 6.35: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV. 223

Figure 6.36: Difference between verified normal form defect analysis and nonverified normal form defect analysis for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference of the evaluated normal form defects of the specific onion layer. The white boxes for lower normal form radii indicate differences below 10^{-5}. The yellow boxes denote differences up to 10^{-4}, the orange boxes correspond to differences up to 10^{-3}, the red boxes denote differences up to $10^{-2.5}$ and the black boxes indicate differences larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. . . . 225

Figure 6.37: Difference between verified normal form defect analysis and nonverified normal form defect analysis for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the difference of the evaluated normal form defects of the specific onion layer. The white boxes for lower normal form radii indicate a difference below 10^{-5}. The yellow boxes denote differences up to 10^{-4}. The orange boxes correspond to differences up to 10^{-3}. The red boxes denote differences up to $10^{-2.5}$, and the black boxes indicate differences larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. 226

Figure 6.38: Difference between the rigorously guaranteed upper bound and the lower bound of the maximum normal form defect using Taylor Model based verified global optimization. The analysis is for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference between the upper bound and the lower bound of the maximum normal form defect of the specific onion layer. All boxes are white because the difference is below 10^{-5}. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04‰ in δp. 227

Figure 6.39: Difference between the rigorously guaranteed upper bound and the lower bound of the maximum normal form defect using Taylor Model based verified global optimization. The analysis is for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference between the upper bound and the lower bound of the maximum normal form defect of the specific onion layer. All boxes are white because the difference is below 10^{-5}. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04‰ in δp. 228

Figure 6.40: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 1** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04‰ in δp. 230

Figure 6.41: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 2** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 231

Figure 6.42: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 3** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 232

Figure 6.43: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 4** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 233

Figure 6.44: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 5** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 234

Figure 6.45: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 6** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 235

Figure 6.46: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 7** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 236

Figure 6.47: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 8** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. 237

Figure 6.48: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 9** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. 238

Figure 6.49: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to the **full tenth order**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04 × 0.04 box in normal form space with a thickness of 0.04% in δp. 239

Figure 6.50: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using an **eleventh order map** and its normal form transformation up to **tenth order**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp. 240

CHAPTER 1

INTRODUCTION

Henri Poincaré was a pioneer – his three volumes on 'New Methods of Celestial Mechanics' [74] were one of the greatest methodological contributions not only to the field of celestial mechanics, but also for the mathematical theory of dynamical systems in general. Numerous methods to describe and analyze dynamical systems in various research areas have been established and developed based on his work.

Poincaré's ideas and concepts were groundbreaking, but strongly limited in their application. Performing his perturbation theory approaches by hand requires a certain simplicity or algebraic structure of the considered system. Many complex systems do not exhibit this simplicity and are impossible to solve in a purely analytic closed form. Consequently, those systems are often reduced in their complexity to ideal cases or simplified versions to solve them analytically.

Computer based numerical methods have been developed to solve complex systems for very specific initial conditions with floating point accuracy. However, to develop sophisticated solutions of complex systems, which are more general than just for a specific set of initial conditions, it is critical to capture as much of the algebraic structure of the problem as possible. The differential algebra (DA) framework developed by Berz *et al.* [19, 15, 18, 14] (Sec. 2.1) constitutes a hybrid structure that manages both of these aspects. It captures the algebraic structure of a system up to arbitrary order to carry out the perturbation part going back to Poincaré's theory, while its implementation in COSY INFINITY [27, 25, 61] allows for an automated calculation of algebraic solutions in a computer environment based on floating point arithmetic.

This thesis will use this powerful hybrid and its associated methods to dive into the fascinating world of nonlinear dynamical systems. The common mathematical underpinnings of many of those systems make it possible to apply the highly developed DA methods to seemingly unrelated fields of study using suitable transformations and projections. To emphasize this versatility of the methods, we analyze one problem from the field of astrodynamics in Chapter 4, and one problem from the field of accelerator physics in Chapter 5. Additionally, we introduce a key technique – the DA normal form algorithm [19] – in Chapter 3, where we analyze the well known system of the centrifugal governor not in its usual linearized version, but with its high order nonlinearities.

The analysis in the field of accelerator physics in Chapter 5 is concerned with the stability and the oscillation frequencies of particles in the storage ring of the Muon g-2 Experiment at Fermilab (E989). We investigate the dependence of these frequencies on offsets in the momentum of the particles and on the amplitudes of oscillation. Nonlinear effects of the various electric field and magnetic field components of the storage rings that are used to confine the particles and bend their trajectory cause these shifts in the frequencies, which potentially influences the beam's susceptibility to resonances. In fact, for the specific ring configurations considered in this thesis, the resonance behavior and their associated fixed point structures make this analysis particularly interesting from a dynamical systems point of view.

In contrast, the analysis in the field of astrodynamics in Chapter 4 is concerned with the trajectories of satellites in low and medium Earth orbits under zonal gravitational perturbation. The perturbation significantly distorts the orbits from their Keplerian form, causing them to rotate within their orbital plane and precess around the Earth at different frequencies. We present a method that elegantly solves one of the elementary challenges in astrodynamics, namely the bounded motion problem, for orbits in the Earth's zonally perturbed gravitational field. Our method generates large continuous sets of orbits, for which any two orbits remain in bounded motion for time periods of decades despite the perturbation.

An essential tool in all of those applications are DA transfer maps and Poincaré maps [19, 40] (see Sec. 2.2). Instead of continuously working with the equations of motions in the form of ordinary differential equations (ODE) as Poincaré did, we work with maps generated from those ODEs. They yield an arbitrarily high order description of a system's behavior between two discrete instances of time or location. Maps are particularly useful for the analysis of repetitive systems in the form of Poincaré return maps, where the maps represent the system's behavior in a chosen cross section of the motion for each turn. A repetitive application of the map to a state in that cross section corresponds to the propagation of the state in the system. Accordingly, the repetitive application allows for a stroboscopic study of the repetitive motion with all the implications regarding its stability.

Origin preserving Poincaré return maps, which are expanded around a linearly stable fixed point, are the starting point of the DA normal form algorithm [19, 17, 16] (see Sec. 2.3). The linearly stable fixed point corresponds to a stable equilibrium state in the Poincaré projection of the system. With the DA normal form algorithm, the phase space behavior around the fixed point of the map is transformed to normalized coordinates, which are closely related to action-angle coordinates. In those normal form coordinates, the phase space behavior is rotationally invariant with only amplitude dependent angle advancements up to the order of calculation. Accordingly, the angle advancements and the amplitude describe the key aspects of the dynamics straightforwardly (see Sec. 2.3.1).

This generalized nonlinear normalization method up to arbitrary order is very powerful and has many applications making it the main component of many techniques used in this thesis. Chapter 3 focuses on a detailed walk through of the DA normal form algorithm using the centrifugal governor as an example. While the principal structure of the process is rather straightforward, the implications of individual steps are not always obvious. This chapter allows discussing those intricacies in full detail.

One critical aspect of the normal form transformation is its sensitivity to resonances (see Sec. 2.3.2). Resonances can affect the normalization process such that the rotationally invariant structure of the resulting normal form is perturbed depending on the strength of the resonances. Hence, those resonances constitute one of the driving factors of the normal form defect (see Sec. 2.4), which is a measure of the variance of the (pseudo-)invariants produced by the normal form. This variance yields a local rate of divergence and can therefore be used as a stability estimate. Phase space regions with large normal form defects can trigger diverging phase space behavior and indicate less stable motion.

As an outlook for future developments, Chapter 6 discusses the first steps of enhancing the methods for these specific applications by making them completely verified. We will see that fully transferring these methods to a verified version is everything but trivial and still to be further investigated. As a starting point for the verified analysis, we introduce verified global optimization [12, 69, 29, 63, 57, 43] and its application for a verified stability estimate of the Muon g-2 Storage Ring.

The basis of this discussion and the global optimization method (see Sec. 2.6) are Taylor Models [53, 58, 54, 55, 21, 75] (see Sec. 2.5), which yield a structure for verified computations by enhancing the DA framework with rigorous remainder bounds.

CHAPTER 2

METHODS

The methods used for this thesis are hybrids of numerical and analytical techniques based on a differential algebra (DA) framework, which was first developed to its current extent by Berz *et al.* [19, 14, 15]. The following summary and introduction to the DA framework (Sec. 2.1), DA maps (Sec. 2.2), and the DA normal form algorithm (Sec. 2.3) are based on [19] and have been given in similar form in my previous publications [95, 96, 93, 94].

In Sec. 2.3.1, the resulting quantities of the normal form, namely the tune, tune shifts, and normal form radii, are discussed in more detail. The influence of resonances on the normal form is described in Sec. 2.3.2. Sec. 2.4 yields an introduction to the normal form defect, a measure for the non-invariance of the normal form radii, based on [29].

The introduction to Taylor Models (Sec. 2.5) for verified computations and their applications including verified global optimization (Sec. 2.6) are based on the work of Makino and Berz *et al.* [53, 58, 54, 55, 29, 62].

2.1 The Differential Algebra (DA) Framework

The fundamental purpose of the DA framework [19] is to provide a mathematical backbone for computer based storage and manipulation of analytic functions. In principle, this is done by representing an analytic function f in terms of its Taylor polynomial expansion $\mathcal{P}_f$ up to order m, similar to how real numbers are represented by an approximation up to a certain arbitrary number of significant digits. In order to discuss the mathematical construction of the differential algebra framework in more detail, we require the notation '$=_m$' instead of just '$\approx$' to clarify that both sides of such an equation are equivalent up to order m.

A Taylor polynomial expansion $\mathcal{P}_f$ up to order m represents multiple analytic functions which are equivalent up to order m. This gives rise to the definition of equivalence classes following [19, p. 91]. The equivalence class $[f]_m$ represents all elements f of the vector space of m times differentiable functions $\mathcal{C}^m(\mathbb{R}^n)$ with n real variables that have identical derivatives at the origin up to order m. The origin is chosen out of convenience and without loss of generality – any other point may be selected. In the DA framework, the equivalence class $[f]_m$ is represented by a DA vector, which stores all the coefficients of the Taylor expansion of f and the corresponding order of the terms in an orderly fashion. Operations are defined on the vector space ${}_mD_n$ of all the equivalence classes $[]_m$.

There are three operations: addition, vector multiplication, and scalar multiplication, which yield results equivalent to the result up to order m of adding two polynomials, multiplying two polynomials, and multiplying a polynomial with a scalar. The first two operations on the equivalence classes (DA vectors) form a ring. The scalar multiplication makes the three operations on the real (or complex) DA vectors an algebra, where not every element has a multiplicative inverse. An example of such elements without a multiplicative inverse is functions without a constant part like $f(x) = x$, since $1/f(x) = 1/x$ is not defined at the origin and can therefore not be expanded around it.

To make the algebra a differential algebra, the derivation D satisfying the Leibniz rule

$$D(fg) = fD(g) + gD(f)) \tag{2.1}$$

is introduced, which is almost trivial in the picture of differentiating polynomial expansions. The derivation opens the door to the algebraic treatment of ordinary and partial differential equations as it is common in the study of differential algebras [77, 76, 45].

Implemented in COSY INFINITY [27, 25, 61], the DA framework allows preserving the algebraic structure up to arbitrary order while manipulating the coefficients of the DA vectors with floating point accuracy. Detailed examples of the operations on ${}_1D_1$ and ${}_2D_1$ are given in [19] and [93], respectively. An example of a DA vector in the application of DA transfer maps and Poincaré maps is given in Sec. 2.2.

2.2 DA Transfer Maps and Poincaré Maps

The dynamics of a system are often described by a set of ordinary differential equations (ODE) $\dot{\vec{z}} = f(\vec{z}, t)$, which describe the incremental change of a state $\vec{z}$ over an independent variable t like time. For practical purposes, it is often advantageous to generally describe the long term propagation of a state $\vec{z}$.

In the terminology of dynamical system theory, a so-called flow operator $\mathcal{M}_T$ is used to describe the action of the system on a state $\vec{z}$ after a fixed time T. Since it is often impossible to determine the flow in a closed form, numerical integration of the ODE is required. The DA framework allows for a hybrid integration that conserves the algebraic structure up to arbitrary order during the integration. Integrating a local expansion $\delta\vec{z}_{\text{ini}}$ around an initial state $\vec{z}_0$ yields the final state $\vec{z}_{\text{fin}}$ in form of an m order flow map $\mathcal{M}_T$, which depends on the expansion in $(\delta\vec{z}_{\text{ini}}, \delta\vec{\eta})$, where $\delta\vec{\eta}$ is the expansion around a reference set of parameters $\vec{\eta}_0$.

More generally speaking, a transfer map $\mathcal{M}$ algebraically expresses how a final state $\vec{z}_{\text{fin}}$ is dependent on an initial state $\vec{z}_{\text{ini}}$ and system parameters $\vec{\eta}$, as

$$\vec{z}_{\text{fin}} = \mathcal{M}\left(\vec{z}_{\text{ini}}, \vec{\eta}\right). \tag{2.2}$$

Transfer maps are also called propagators or simply maps. The expansion point of the map belongs to a chosen reference orbit/state of the system, e.g. a (pseudo-)closed orbit for a fixed point map and/or the ideal orbit of the unperturbed system.

There are special transfer maps called Poincaré maps [74] that constrain the initial and final state to Poincaré surfaces $\mathbb{S}_{\text{ini}}$ and $\mathbb{S}_{\text{fin}}$, respectively. For the simulation of storage rings and their particle optical elements, this concept is used to represent how the state after a storage ring element depends on system parameters and the state before the element. A setup of multiple consecutive storage ring elements is described by the composition of their Poincaré maps.

Poincaré return maps represent the case where $\mathbb{S}_{\text{ini}}$ is equal to $\mathbb{S}_{\text{fin}}$. They are particularly useful for the representation of dynamics in repetitive systems like the ones considered in this thesis. Multiple applications of a Poincaré return map correspond to the propagation of the system. The Poincaré return maps are particularly advantageous when they are origin preserving, i.e., the expansion point is a fixed point of the map, because system dynamics represented by origin preserving Poincaré return maps can be further analyzed by normal form methods and for the asymptotic stability of the system.

Constraining the map to the Poincaré surface $\mathbb{S}$ is often done by calculating the flow of an ODE and projecting it onto the surface $\mathbb{S}$. This reduces the dimension of the original map and generates the Poincaré map. An implementation of a timewise projection onto a surface $\mathbb{S}$ defined by $\sigma(\vec{z}, \vec{\eta}) = 0$ is outlined in [40].

The projection uses DA inversion methods that compute the inverse $\mathcal{A}^{-1}$ to the auxiliary map $\mathcal{A}$, which contains the constraining conditions of the Poincaré surface $\mathbb{S}$. Given that $\mathcal{A}$ has no constant part, the auxiliary map and its inverse satisfy

$$\mathcal{A}^{-1} \circ \mathcal{A} =_m \mathcal{A} \circ \mathcal{A}^{-1} =_m \mathcal{I}. \tag{2.3}$$

The basic idea of the projection of a transfer map $\mathcal{M}$ onto a surface defined by $\sigma(\vec{z}, \vec{\eta}) = 0$ is to replace one of the variables or parameters of $\mathcal{M}$ by an expression in terms of all the other variables and parameters such that the constraint $\sigma(\mathcal{M}) = 0$ is satisfied. This eliminates the corresponding component of the map and thereby reduces its dimensionality. In [40], the timewise projection is prepared by calculating an expansion of the map $\mathcal{M}$ in time t. The DA inversion methods are then used to find the intersection time $t^\star(\vec{z}, \vec{\eta})$ dependent on the state variables $\vec{z}$ and system parameters $\vec{\eta}$ such that

$$\sigma(\mathcal{M}(\vec{z}, \vec{\eta}, t^\star(\vec{z}, \vec{\eta}))) = 0. \tag{2.4}$$

2.3 The DA Normal Form Algorithm

The DA normal form algorithm [19] is an advancement from the DA-Lie based version, the first arbitrary order algorithm by Forest, Berz, and Irwin [38]. Given an origin preserving map $\mathcal{M}$ of a repetitive Hamiltonian system, where the components of the map are in phase space coordinates, the DA normal form algorithm provides a nonlinear change of the phase space variables by an order-by-order transformation to rotationally invariant normal form coordinates.

Implemented in COSY INFINITY [27, 25, 61], this is a fully automated process, which can be performed up to arbitrary order. It is only limited by floating point accuracy and the capability of the computer system to handle DA vectors of the chosen computation order m_{calc} and dimension.

In the standard configuration, order ten calculations of a three dimensional system (with six phase space variables) are easily manageable.

In Chapter 3, the normal form algorithm is explained in great detail for the one dimensional system (with two phase space variables) of a centrifugal governor. Here we want to explain the more general form for a $2n$ dimensional symplectic system with an optional parameter dependence on n_η parameters summarized in $\vec{\eta}$. The explanations are largely based on [19].

For parameter dependent maps, the algorithm starts by expanding the origin preserving map $\mathcal{M}(\vec{z}, \vec{\eta})$ around its parameter dependent fixed point $\vec{z}_{\text{PDFP}}(\vec{\eta})$, which satisfies

$$\mathcal{M}\left(\vec{z}_{\text{PDFP}}(\vec{\eta}), \vec{\eta}\right) = \vec{z}_{\text{PDFP}}(\vec{\eta}). \tag{2.5}$$

Defining the extended map $\mathcal{N} = (\mathcal{M} - \mathcal{I}_{\vec{z}}, \vec{\eta})$, the parameter dependent fixed point $\vec{z}_{\text{PDFP}}$ is determined by evaluating the inverse of $\mathcal{N}$ at the expansion point $\vec{z} = \vec{0}$:

$$\left(\vec{z}_{\text{PDFP}}\left(\vec{\eta}\right), \vec{\eta}\right) = \mathcal{N}^{-1}\left(\vec{0}, \vec{\eta}\right). \tag{2.6}$$

The map $\mathcal{M}$ is then expanded around its parameter dependent fixed point $\vec{z}_{\text{PDFP}}$.

The resulting map $\mathcal{M}_0 = \mathcal{L} + \sum_m \mathcal{U}_m$ consists of a linear part $\mathcal{L}$ and the nonlinear parts $\mathcal{U}_m$ of order m. Due to the transformation to the parameter dependent fixed point, the map has no terms that only depend on parameters. Accordingly, the entire linear part is independent of parameters.

The variables of the map are the canonical phase space coordinates $\vec{z} = (\vec{q}_0, \vec{p}_0)$ and, if applicable, parameters $\vec{\eta}$. The normal form algorithm transforms this map order by order up to the full calculation order m_{calc} of the map. For each order m, the transformation step has the following form

$$\mathcal{M}_m = \mathcal{A}_m \circ \mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1}, \tag{2.7}$$

where $\mathcal{A}_m$ is the transformation map (also just called transformation) and $\mathcal{A}_m^{-1}$ is its inverse. The result of the mth order transformation step is the map $\mathcal{M}_m$. Hence, $\mathcal{M}_{m-1}$ is the result from the previous transformation step or $\mathcal{M}_0$ from above. The last transformation step is for order $m = m_{\text{calc}}$.

The first step of the algorithm, for $m = 1$, is to linearly decouple the map into n two dimensional subspaces. The linear transformation diagonalizes the system, transforming the (parameter dependent) fixed point map into the complex conjugate eigenvector space of its linear part. We assume linearly stable behavior around the (parameter dependent) fixed point of the map with distinct complex conjugate eigenvalue pairs of magnitude one since this property is shared among all systems considered in this thesis (see [19] for other cases). If any of the eigenvalues $\lambda_\star$ had an absolute value larger than 1, the motion would be unstable since the state on the corresponding eigenvector $\vec{v}_\star$ would grow in magnitude by a factor of $|\lambda_\star| > 1$ with each iteration. Additionally, eigenvalues of symplectic maps come in reciprocal pairs such that eigenvalues with a magnitude smaller than 1 have a reciprocal partner eigenvalue $|\lambda_\star| > 1$, which are again linearly unstable.

The complex conjugate eigenvalue pairs $e^{\pm i\mu_j}$ of the diagonalized linear part are grouped together such that the matrix $\hat{R}$ of the diagonalized linear part $\mathcal{R}$ has the following decoupled form

$$\hat{R} = \begin{pmatrix} \hat{R}_1 & & & & \\ & \ddots & & & \\ & & \hat{R}_l & & \\ & & & \ddots & \\ & & & & \hat{R}_n \end{pmatrix} \qquad \text{where} \qquad \hat{R}_j = \begin{pmatrix} e^{+i\mu_j} & 0 \\ 0 & e^{-i\mu_j} \end{pmatrix}. \tag{2.8}$$

The resulting map of the first transformation step – linear transformation – is $\mathcal{M}_1 = \mathcal{R} + \sum_m \mathcal{S}_m$, where the new nonlinear terms of order m that resulted from the linear transformation are denoted by $\mathcal{S}_m$. The complex phase $\pm\mu_j$ of the eigenvalue pairs will be of critical importance in the nonlinear transformations of the algorithm.

In summary, the first transformation step performed the following operation

$$\mathcal{M}_1 = \mathcal{A}_1 \circ \mathcal{M}_0 \circ \mathcal{A}_1^{-1} = \mathcal{A}_1 \circ \mathcal{L} \circ \mathcal{A}_1^{-1} + \sum_m \mathcal{A}_1 \circ \mathcal{U}_m \circ \mathcal{A}_1^{-1} = \mathcal{R} + \sum_m \mathcal{S}_m, \tag{2.9}$$

where $\mathcal{A}_1$ is the linear transformation from the original coordinate space $(\vec{q}_0, \vec{p}_0)$ to the complex conjugate coordinate space $(\vec{q}_1, \vec{p}_1)$ and $\mathcal{A}_1^{-1}$ is its inverse for the transformation in the opposite direction.

With the linearly decoupled map, the following steps of the normal form algorithm can be performed for each of these linearly decoupled subspaces separately. The jth subspace of the linearly decoupled map $\mathcal{M}_1$ can be explicitly written as

$$
\begin{aligned}
\mathcal{M}_{1,j}\left(\vec{q}_1, \vec{p}_1, \vec{\eta}\right) = \mathcal{R}_j + \sum_m \mathcal{S}_{m,j} = &\begin{pmatrix} e^{+i\mu_j} & 0 \\ 0 & e^{-i\mu_j} \end{pmatrix} \begin{pmatrix} q_{1,j} \\ p_{1,j} \end{pmatrix} \\
&+ \sum_{m=||\vec{k}^{+}+\vec{k}^{-}||_1+||\vec{k}^{\eta}||_1} \begin{pmatrix} \mathcal{S}^{+}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} \\ \mathcal{S}^{-}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} \end{pmatrix} \prod_{l=1}^{n} \left(q_{1,l}\right)^{k_l^{+}} \left(p_{1,l}\right)^{k_l^{-}} \prod_{u=1}^{n\eta} \left(\eta_u\right)^{k_u^{\eta}}, \quad (2.10)
\end{aligned}
$$

where k_l^+ represents the positive integer exponent of $q_{1,l}$, k_l^- represents the positive integer exponent of $p_{1,l}$, and k_u^η represents the positive integer exponent of η_u. The positive integer exponents are summarized in the vectors $\vec{k}^+$, $\vec{k}^-$, and $\vec{k}^\eta$, respectively. The L^1-Norm $||\cdot||_1$ of the sum of these vectors is used to ensure that only polynomial terms of order m are considered.

To better understand the expression in Eq. (2.10), we present some terms of the $\mathcal{M}^-_{1,j}$ component

$$
\begin{aligned}
\mathcal{M}^{-}_{1,j}\left(\vec{q}_1, \vec{p}_1, \vec{\eta}\right) = e^{-i\mu_j} \cdot p_{1,j} &+ \mathcal{S}^{-}_{2\left((2,0,\ldots,0)^T,(0,\ldots,0)^T,(0,\ldots,0)^T\right),j} \cdot q_{1,1}^2 + \ldots \quad (2.11) \\
&+ \mathcal{S}^{-}_{2\left((0,\ldots,0,k_j^{+}=1,0,\ldots,0)^T,(0,\ldots,0,k_l^{-}=1,0,\ldots,0)^T,(0,\ldots,0)^T\right),j} \cdot q_{1,j} p_{1,l} + \ldots \\
&+ \mathcal{S}^{-}_{2\left((0,\ldots,0)^T,(0,\ldots,0,1)^T,(1,0,\ldots,0)^T\right),j} \cdot p_{1,n} \eta_1 + \ldots
\end{aligned}
$$

Due to the linear transformation into the complex conjugate eigenvector space of the purely real linear part, the two components of each subspace form a complex conjugate pair. The '+' and '-' notation is used, where the sign corresponds to the sign of the complex eigenvalue phase of the map component of that subspace. Specifically, this means that

$$
\mathcal{M}^{+}_{1,j} = \overline{\mathcal{M}^{-}_{1,j}} \quad \text{with} \quad q_{1,j} = \overline{p}_{1,j} \quad \text{and} \quad \mathcal{S}^{+}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} = \overline{\mathcal{S}^{-}}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}. \quad (2.12)
$$

This property is maintained throughout all the following nonlinear transformation steps, which are performed order by order starting with order two. The general form of the nonlinear transformation is $\mathcal{A}_m =_m \mathcal{I} + \mathcal{T}_m$, where $\mathcal{T}_m$ is a polynomial containing only terms of order m. Hence, the transformation $\mathcal{A}_m$ is a near-identity transformation and a full identity up to order $m-1$. The transformation $\mathcal{A}_m$ is determined by finding $\mathcal{T}_m$ such that the mth order of the resulting map $\mathcal{M}_m$

is simplified or even eliminated when the transformation $\mathcal{A}_m$ and its inverse $\mathcal{A}_m^{-1} =_m \mathcal{I} - \mathcal{T}_m$ are applied to $\mathcal{M}_{m-1}$ in the mth order nonlinear transformation step (see Eq. (2.7)).

The higher order terms of the transformation $\mathcal{A}_m$ do not influence the mth order terms of the map. Hence, they are irrelevant for the mth order transformation step and can be chosen freely, e.g. to make the transformation symplectic with $\mathcal{A}_m = \exp(L_{\mathcal{T}_m})$ which we will do (see [19]). However, the higher orders of the resulting map $\mathcal{M}_m$ are strongly dependent on $\mathcal{A}_m$, its higher order terms, and its corresponding inverse. In Chapter 3, the influences of the second order transformation on the third order terms of the resulting map are analyzed in great detail. While these influences are not to be dismissed, the key element of this mth order transformation step is the elimination of as many mth order terms of the resulting map $\mathcal{M}_m$ as possible by a smart choice of $\mathcal{T}_m$.

Given the map $\mathcal{M}_{m-1}$, representing $\mathcal{M}$ simplified up to order $m-1$ and applying $\mathcal{A}_m$ and its inverse to it, yields [19, Eq. (7.60)]:

$$\begin{aligned}\mathcal{A}_m \circ \mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1} &=_m (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} + \mathcal{S}_m) \circ (\mathcal{I} - \mathcal{T}_m) \\ &=_m (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} - \mathcal{R} \circ \mathcal{T}_m + \mathcal{S}_m) \\ &=_m \mathcal{R} + \mathcal{S}_m + [\mathcal{T}_m, \mathcal{R}]\,, \end{aligned} \tag{2.13}$$

where $\mathcal{R}$ is the diagonalized linear part and $\mathcal{S}_m$ represents only the mth order terms of the map $\mathcal{M}_{m-1}$ (the leading order of terms that have not been simplified yet).

The equations above only consider terms up to order m, since terms of order $m+1$ and larger are irrelevant for determining $\mathcal{T}_m$. The maximum simplification would be achieved by finding $\mathcal{T}_m$ such that the commutator

$$\mathcal{C}_m = \mathcal{T}_m \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_m = [\mathcal{T}_m, \mathcal{R}] = -\mathcal{S}_m, \tag{2.14}$$

which would eliminate all nonlinear terms $\mathcal{S}_m$ of order m.

Since the commutator only involves $\mathcal{T}_m$ and $\mathcal{R}$ we can investigate this transformation separately in the n individual subspaces. The components of the jth subspace of the commutator $\mathcal{C}_m = [\mathcal{T}_m, \mathcal{R}]$

are

$$\mathcal{C}_{m,j} = \sum_{m=||\vec{k}^{+}+\vec{k}^{-}||_1+||\vec{k}^{\eta}||_1} \begin{pmatrix} \mathcal{C}^{+}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} \\ \mathcal{C}^{-}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} \end{pmatrix} \prod_{l=1}^{n} (q_l)^{k_l^{+}} (p_l)^{k_l^{-}} \prod_{u=1}^{n_\eta} (\eta_u)^{k_u^{\eta}}, \tag{2.15}$$

where

$$\mathcal{C}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} = \mathcal{T}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} \left(e^{i\vec{\mu}\left(\vec{k}^{+}-\vec{k}^{-}\right)} - e^{\pm i\mu_j} \right). \tag{2.16}$$

Accordingly, the commutator terms $\mathcal{C}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}$ can eliminate their corresponding nonlinear terms of the map $\mathcal{S}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}$ by choosing

$$\mathcal{T}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j} = \frac{-\mathcal{S}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}}{e^{i\vec{\mu}\left(\vec{k}^{+}-\vec{k}^{-}\right)} - e^{\pm i\mu_j}}, \tag{2.17}$$

if

$$e^{i\vec{\mu}\left(\vec{k}^{+}-\vec{k}^{-}\right)} - e^{\pm i\mu_j} \neq 0. \tag{2.18}$$

In other words, only the $\mathcal{S}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}$ terms corresponding to $\mathcal{C}^{\pm}_{m(\vec{k}^{+},\vec{k}^{-},\vec{k}^{\eta}),j}$ for which the condition (see [19, Eq. (7.65)])

$$\operatorname{mod}_{2\pi}\left(\mu_j (k_j^{+} - k_j^{-} \mp 1) + \sum_{l \neq j} \mu_l \left(k_l^{+} - k_l^{-} \right) \right) = 0, \tag{2.19}$$

is satisfied, survive.

A straightforward solution of the condition in Eq. (2.19) is

$$k_j^{+} - k_j^{-} = \pm 1 \quad \wedge \quad k_l^{+} = k_l^{-} \, \forall l \neq j, \tag{2.20}$$

where the first condition concerns the jth subspace and the second condition is regarding all the other subspaces l with $l \neq j$.

The surviving terms of the mth order transformation step in the jth subspace can be generally written as

$$\mathcal{S}^{+}_{m(\vec{k}+\vec{e}_j,\vec{k},\vec{k}^{\eta}),j} \quad \text{and} \quad \mathcal{S}^{-}_{m(\vec{k},\vec{k}+\vec{e}_j,\vec{k}^{\eta}),j} \quad \text{with} \quad 2||\vec{k}||_1 + 1 + ||\vec{k}^{\eta}||_1 = m, \tag{2.21}$$

where the unit vector $\vec{e}_j$ consists only of zeros except for a 1 at the jth entry.

From Eq. (2.21) it becomes clear that only certain terms of uneven order in the phase space coordinates $(\vec{q}, \vec{p})$ survive. These terms have the special property that each complex conjugate phase space variable pair is raised to the same exponent except for the phase space variable pair of the respective subspace. So, all even order terms in phase space coordinates can be eliminated by the nonlinear normal form transformations.

The remaining terms of $\mathcal{S}_m$ (from Eq. (2.21)) describe the entire dynamics of the systems in a nutshell and are the key elements of the normal form and therefore essential for further analysis of the dynamics.

Resonances between the complex phases $\vec{\mu}$ of the different subspaces in the denominator of Eq. (2.17) can break this special structure and therefore the rotational invariance of the normal form as will be discussed in Sec. 2.3.2. For now, we will continue only with the terms that are supposed to survive, namely the terms specified in Eq. (2.21).

Once the nonlinear transformation steps transformed the map up to its full order $m = m_{\text{calc}}$, the map has been significantly simplified to

$$\begin{pmatrix} \mathcal{M}^+_{m,j} \\ \mathcal{M}^-_{m,j} \end{pmatrix} = \begin{pmatrix} q_{m,j}\, f^+_j\left(q_{m,1}p_{m,1}, q_{m,2}p_{m,2}, \ldots, q_{m,n}p_{m,n}, \vec{\eta}\right) \\ p_{m,j}\, f^-_j\left(q_{m,1}p_{m,1}, q_{m,2}p_{m,2}, \ldots, q_{m,n}p_{m,n}, \vec{\eta}\right) \end{pmatrix}, \tag{2.22}$$

where

$$f^+_j = e^{+i\mu} + \sum_{m=2\|\vec{k}\|_1+1+\|\vec{k}^\eta\|_1} \mathcal{S}^+_{m(\vec{k}+\vec{e}_j,\vec{k},\vec{k}^\eta),j} \prod_{l=1}^{n} \left(q_{m,l}p_{m,l}\right)^{k_l} \prod_{u=1}^{n_\eta} (\eta_u)^{k^\eta_u}. \tag{2.23}$$

Since the original map is real, the last step of the algorithm is transforming the resulting map to the real normal form basis $(\vec{q}_{\text{NF}}, \vec{p}_{\text{NF}})$, which is composed of the real and imaginary parts of the current complex conjugate basis $(\vec{q}_m, \vec{p}_m)$. The relation between the bases is

$$q_{\text{NF},j} = \frac{q_{m,j} + p_{m,j}}{2}, \quad p_{\text{NF},j} = \frac{q_{m,j} - p_{m,j}}{2i}, \quad \text{and} \tag{2.24}$$

$$q_{m,j} = q_{\text{NF},j} + i p_{\text{NF},j}, \quad p_{m,j} = q_{\text{NF},j} - i p_{\text{NF},j}. \tag{2.25}$$

The squared normal form radius $r^2_{\text{NF},j}$ is given by the product of $q_{m,j}p_{m,j}$, with

$$q_{m,j}p_{m,j} = q^2_{\text{NF},j} + p^2_{\text{NF},j} = r^2_{\text{NF},j}. \tag{2.26}$$

Applying the basis transformation to the map components of $\mathcal{M}_m$ in each subspace yields

$$\begin{aligned}
\mathcal{M}_{\mathrm{NF},j} &= \mathcal{A}_{\mathrm{real},j} \circ \mathcal{M}_{m,j} \circ \mathcal{A}^{-1}_{\mathrm{real},j} \\
&= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \cdot \begin{pmatrix} f_j^+\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\left(q_{\mathrm{NF},j} + i\, p_{\mathrm{NF},j}\right) \\ f_j^-\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\left(q_{\mathrm{NF},j} - i\, p_{\mathrm{NF},j}\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}\left(f_j^+ + \bar{f}_j^+\right) q_{\mathrm{NF},j} + \frac{i}{2}\left(f_j^+ - \bar{f}_j^+\right) p_{\mathrm{NF},j} \\ \frac{-i}{2}\left(f_j^+ - \bar{f}_j^+\right) q_{\mathrm{NF},j} + \frac{1}{2}\left(f_j^+ + \bar{f}_j^+\right) p_{\mathrm{NF},j} \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Re}\left(f_j^+\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) & -\mathrm{Im}\left(f_j^+\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) \\ \mathrm{Im}\left(f_j^+\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) & \mathrm{Re}\left(f_j^+\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) \end{pmatrix} \cdot \begin{pmatrix} q_{\mathrm{NF},j} \\ p_{\mathrm{NF},j} \end{pmatrix}.
\end{aligned} \tag{2.27}$$

Writing f_j^+ and its complex conjugate counterpart f_j^- in terms of complex phases with

$$f_j^{\pm}\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right) = e^{\pm i \Lambda_j\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)} \tag{2.28}$$

yields the following normal form

$$\mathcal{M}_{\mathrm{NF},j} = \begin{pmatrix} \cos\left(\Lambda_j\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) & -\sin\left(\Lambda_j\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) \\ \sin\left(\Lambda_j\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) & \cos\left(\Lambda_j\left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}, \vec{\eta}\right)\right) \end{pmatrix} \cdot \begin{pmatrix} q_{\mathrm{NF},j} \\ p_{\mathrm{NF},j} \end{pmatrix}, \tag{2.29}$$

which clearly shows the circular phase space behavior in normal form subspaces with only amplitude

$$\vec{r}_{\mathrm{NF,sqr}} = \left(r^2_{\mathrm{NF},1}, ..., r^2_{\mathrm{NF},n}\right) \tag{2.30}$$

and parameter $\vec{\eta}$ depended angle advancements $\vec{\Lambda}$.

The radii of the circular motion – the normal form radii – are constants of motion up to the calculation order. The entire dynamics in the normal form are given by the constant angle advancements $\vec{\Lambda}$ along the circular phase space curves. The rotational invariance implies an interpretation of the normal form as an averaged representation of the original Poincaré return map $\mathcal{M}$, in the limit where the map application is repeated infinitely many times.

Normalizing the angle advancements $\vec{\Lambda}$ to $[0, 1]$ yields the tunes $\vec{\nu}$ and amplitude and parameter dependent tune shifts $\delta\vec{\nu}(\vec{r}_{\mathrm{NF,sqr}}, \vec{\eta})$. Accordingly,

$$\frac{\vec{\Lambda}(\vec{r}_{\mathrm{NF,sqr}}, \vec{\eta})}{2\pi} = \vec{\nu} + \delta\vec{\nu}(\vec{r}_{\mathrm{NF,sqr}}, \vec{\eta}). \tag{2.31}$$

The normal form transformation $\mathcal{A}$ and its inverse $\mathcal{A}^{-1}$ are given by the composition of all the individual transformations of each transformation step with

$$\mathcal{M}_{\mathrm{NF}} = \underbrace{\mathcal{A}_{\mathrm{real}} \circ \mathcal{A}_m \circ \mathcal{A}_{m-1} \circ ... \circ \mathcal{A}_1}_{\mathcal{A}} \circ \mathcal{M} \circ \underbrace{\mathcal{A}_1^{-1} \circ ... \circ \mathcal{A}_{m-1}^{-1} \circ \mathcal{A}_m^{-1} \circ \mathcal{A}_{\mathrm{real}}^{-1}}_{\mathcal{A}^{-1}} . \tag{2.32}$$

The normal form transformation $\mathcal{A}$ yields how the normal form variables $(q_{\mathrm{NF},j}, p_{\mathrm{NF},j})$ depend on the original phase space variables $(\vec{q}_0, \vec{p}_0)$ and, if considered, system parameters $\vec{\eta}$, which suggests the following notation for $\mathcal{A}$ and its inverse

$$\mathcal{A} = \left(\vec{q}_{\mathrm{NF}}\left(\vec{q}_0, \vec{p}_0, \vec{\eta}\right), \vec{p}_{\mathrm{NF}}\left(\vec{q}_0, \vec{p}_0, \vec{\eta}\right)\right) \tag{2.33}$$

$$\mathcal{A}^{-1} = \left(\vec{q}_0\left(\vec{q}_{\mathrm{NF}}, \vec{p}_{\mathrm{NF}}, \vec{\eta}\right), \vec{q}_0\left(\vec{q}_{\mathrm{NF}}, \vec{p}_{\mathrm{NF}}, \vec{\eta}\right)\right) . \tag{2.34}$$

2.3.1 Tunes, Tune Shifts, and Normal Form Radii

DA normal form methods are used to transform the origin preserving phase space Poincaré return map to the rotationally invariant normal form up to calculation order. From the normal form, the angle advancements $\vec{\Lambda}(\vec{r}_{\mathrm{NF,sqr}}, \vec{\eta})$ as a functions of amplitude $\vec{r}_{\mathrm{NF,sqr}}$ and parameters $\vec{\eta}$ are particularly straightforward to extract. Scaling the angle advancements in each of the normal form phase spaces to $[0, 1]$ instead of $[0, 2\pi]$ provides the average number of phase space revolutions per system revolution represented by the Poincaré return map. In beam physics terminology, the frequencies of normal form phase space revolutions is known as the tunes $\vec{\nu}$ and their amplitude and parameter dependent tune shifts $\delta\vec{\nu}(\vec{r}_{\mathrm{NF,sqr}}, \vec{\eta})$ [19].

The tune ν_j corresponds to the scaled complex phase μ_j of the complex conjugate eigenvalues $\lambda_j^{\pm}$ of the linear transformation. Hence, the tune is related to the linear motion around the expansion point, i.e., the motion ‘infinitely close’ to the expansion point. Interpreting the tune and its tune shifts as the phase space rotation frequency suggests that the tune – the phase space rotation frequency of the expansion point – is a rotation with no amplitude, where the frequency is determined by the linear motion around the expansion point. In particular, this means that different maps with the same expansion point can have different tunes depending on the linear motion around the expansion point. Since the tunes are calculated from the linear coefficients directly without any nonlinear transformations, performing the tune calculation with parameter dependent linear coefficients directly yields the parameter dependent tune shifts.

The tune shifts indicate the change of the phase space rotation frequency dependent on the phase space amplitudes $r_{\mathrm{NF},j}$ and variations in the system parameters $\vec{\eta}$. Since the normal form transformation is symplectic, it preserves the phase space volume, which is critical to understanding the connection between the original phase space coordinates and their normal form radii. If the system is only weakly coupled between the different phase spaces, the normal form radius $r_{\mathrm{NF},j}$ is a measure for the invariant phase space area of the jth subspace denoted by A_j. Hence, the original phase space coordinates of an invariant phase space orbit in the jth subspace enclose the area A_j, which roughly corresponds to the normal form radius of $r_{\mathrm{NF},j} = \sqrt{A_j/\pi}$.

The normal form radii are the link between the tune dependencies and the original coordinates. The dependency of the tune shifts on the normal form radii is a result of the surviving terms $\mathcal{S}_m$ of the nonlinear normal form transformations. However, the crucial terms are the $\mathcal{T}_m$ terms from Eq. (2.17) that are used to cancel all the other nonlinear terms $\mathcal{S}_m$. On the one hand, the $\mathcal{T}_m$ terms determine how the original coordinates $\vec{z} = (\vec{q}, \vec{p})$ and the system parameters $\vec{\eta}$ relate to the normal form radii $r_{\mathrm{NF},j}$, since the $\mathcal{T}_m$ are the essential part of the normal form transformation. On the other hand, they influence the higher order nonlinear terms $\mathcal{S}_l$ with $l > m$, which either survive and determine the dependency of the tune shifts on the normal form radii, or they determine the higher order terms $\mathcal{T}_l$.

2.3.2 Resonances

The denominator of $\mathcal{T}_m$ in Eq. (2.17) has a potentially large effect on the size of $\mathcal{T}_m$ the closer it is to satisfying the resonance condition in Eq. (2.19). If the condition is satisfied, the corresponding nonlinear terms in $\mathcal{S}_m$ cannot be eliminated. Accordingly, terms survive which do not fit the normal form structure. They break the normal form by the size of their respective coefficient.

If the condition is almost satisfied close to a resonance, then the denominator of $\mathcal{T}_m$ becomes very small, making $\mathcal{T}_m$ very large. In this situation, there are two options. One option is to continue the procedure with the very large $\mathcal{T}_m$ coefficient, which conserves the normal form structure but yields diverging coefficients in all higher order terms. The other option is to let the corresponding term in $\mathcal{S}_m$ survive, which breaks the normal form structure but avoids a divergence of the coefficients. In practice, one chooses a cutoff value for the size of the denominator, which restricts the size of potentially diverging coefficients. If the denominator is smaller than the cutoff value, the $\mathcal{T}_m$ coefficient is set to zero, letting the corresponding $\mathcal{S}_m$ term survive.

Rewriting the resonance condition in terms of tunes yields

$$\vec{w} \cdot \vec{\nu} = g, \tag{2.35}$$

where $\vec{w}$ consists only of integer values and g is a natural number $\mathbb{N}_0$. The values in $\vec{w}$ and g are chosen such that the greatest common divisor of all values is 1. With this definition, the order of the resonance is given by $m_{\text{res}} = ||\vec{w}||_1$.

In the normal form algorithm a tune resonance defined by $(\vec{w}, g)$ appears in all terms $\mathcal{S}^{\pm}_{m(\vec{k}^+,\vec{k}^-,\vec{k}^\eta),j}$ for which

$$w_j = k_j^+ - k_j^- \mp 1 \quad \wedge \quad w_l = k_l^+ - k_l^- \quad \forall l \neq j, \tag{2.36}$$

$$\text{and} \quad -w_j = k_j^+ - k_j^- \mp 1 \quad \wedge \quad -w_l = k_l^+ - k_l^- \quad \forall l \neq j, \tag{2.37}$$

according to Eq. (2.20). Resonances of order m_{res} appear for the first time in the normal form transformation step of order $m_{\text{NF}} = m_{\text{res}} - 1$.

Consider a four dimensional phase space system ($n = 2$) without parameter dependence, where the eigenvalue phases μ_i satisfy the following order seven resonance $2\mu_1 - 5\mu_2 = -4\pi$. This corresponds to the tune resonance condition of $-2\nu_1 + 5\nu_2 = 2$ denoted by $(\vec{w}, g) = \left((-2, 5)^T, 2\right)$. Given $\vec{w}$, the corresponding terms $\mathcal{S}^{\pm}_{m(\vec{k}^+,\vec{k}^-,\vec{k}^\eta),j}$ in the normal form that encounter this resonance are determined by all vectors $\vec{k}^+$ and $\vec{k}^-$ that satisfy the conditions in Eq. (2.36) and Eq. (2.37). Hence, the first terms of the normal form to encounter this resonance are the sixth order complex conjugate terms

$$\mathcal{S}^{+}_{6((0,5)^T,(1,0)^T),1} \quad \text{and} \quad \mathcal{S}^{-}_{6((1,0)^T,(0,5)^T),1} \tag{2.38}$$

$$\text{as well as} \quad \mathcal{S}^{-}_{6((0,4)^T,(2,0)^T),2} \quad \text{and} \quad \mathcal{S}^{+}_{6((2,0)^T,(0,4)^T),2}. \tag{2.39}$$

Hence, for each subspace, one complex conjugate pair survives due to the resonance between μ_1 and μ_2, which break the rotational symmetry structure of the resulting normal form.

As we will see later on and as discussed in [98], resonances in the tune space correspond to fixed point structures in the phase space, which often yields fascinating behavior especially for low order resonances.

2.4 The Normal Form Defect

The volume conserving property of Hamiltonian systems expressed by Liouville's theorem is maintained by the normal form transformation. Given the rotational invariants of the normal form, the size of the phase space volume is determined by the normal form radii. Accordingly, the normal form phase space radii constitute invariants of motion up to the order of the normal form transformation if no resonance conditions were encountered. However, they are usually not invariants of the full (order) motion.

While the expansion of the transfer map improves in accuracy with every additional order considered, the same is not guaranteed for the normal form transformation. It is unknown how well or even if the normal form converges with higher orders. This is due to its sensitivity to resonances, which may initiate asymptotic behavior once the order of a close-by resonance is reached. The higher the order of the computation, the more resonances are potentially relevant. Depending on the complexity of the original transfer map, it is usually unpredictable which resonances may affect the normal form and in what way.

However, if the normal form transformation converges, its high order limit will yield the exact invariants. In the case of exact invariants, the system is integrable and can be transformed into a trivial system by introducing the invariants as variables. Those variables are known as action-angle coordinates, where the action is constant and unique for each phase space curve and each point on the phase space curve is associated with the action-angle. For complex systems such as the ones discussed in this thesis, there are no exact invariants that can be expressed in terms of finite order terms. Thus, tools to assess the error of the calculated pseudo-invariants in the form of normal form radii are useful.

The normal form defect represents the inaccuracy of the normal form radii as invariants and is locally defined for each phase space state. Given an origin preserving fixed point map $\mathcal{M}(q, p) = (Q, P)$ of a repetitive system and the corresponding normal form transformation $\mathcal{A}(q, p) = (q_{\mathrm{NF}}, p_{\mathrm{NF}})$, the normal form defect $d_{\mathrm{NF}}(\vec{z}_0)$ of the phase space state $\vec{z}_0 = (q, p)$ is given by the difference between the normal form radius $r(\vec{z}_1 = \mathcal{M}(\vec{z}_0))$ of the mapped phase space state $\vec{z}_1 = \mathcal{M}(\vec{z}_0)$ and the normal form radius $r(\vec{z}_0)$ of the original phase space state $\vec{z}_0$. Generally, the normal form radius r of a phase space state $\vec{z}$ is the magnitude of the vector formed by the normal form phase space state $(q_{\mathrm{NF}}, p_{\mathrm{NF}}) = \mathcal{A}(\vec{z}_0)$, specifically

$$r(\vec{z}) = \sqrt{(q_{\mathrm{NF}}(\vec{z}))^2 + (p_{\mathrm{NF}}(\vec{z}))^2}. \tag{2.40}$$

Accordingly, the normal form defect is given by

$$\begin{aligned} d_{\mathrm{NF}}(\vec{z}_0) &= r_1 - r_0 = r(\vec{z}_1) - r(\vec{z}_0) = r(\mathcal{M}(\vec{z}_0)) - r(\vec{z}_0) \\ &= \sqrt{(q_{\mathrm{NF}}(\mathcal{M}(\vec{z}_0)))^2 + (p_{\mathrm{NF}}(\mathcal{M}(\vec{z}_0)))^2} - \sqrt{(q_{\mathrm{NF}}(\vec{z}_0))^2 + (p_{\mathrm{NF}}(\vec{z}_0))^2}. \end{aligned} \tag{2.41}$$

The application of the one turn map represents the evolution of the system by describing how each phase space state changes after one revolution of the system. The normal form defect indicates how much the normal form radii, i.e., a (pseudo-)invariants of the motion, change between two states of the motion connected by the map $\mathcal{M}$. An increasing normal form radius with time indicates diverging phase space behavior with larger amplitudes, i.e. the normal form defect measures the local rate of divergence per map application.

Analyzing the normal form defect for a whole set of states within a certain phase space domain $\mathbb{D}$ allows for stability estimations by placing an upper bound on the rate of divergence. The upper bound can be determined in various ways, including rigorous global optimization methods on the normal form defect over the given domain. The upper bound can serve as a Nekhoroshev-type stability estimate [73] that allows for the calculation of the minimum amount of revolutions of the system N, for which the motion will be guaranteed to stay within the allowed region $\mathbb{D}$:

$$N = \frac{r_{\text{max}} - r\left(\vec{z}_{\text{ini}}\right)}{\max\left(d_{\text{NF}}\left(\vec{z}\right)\right)} \quad \text{with} \quad \vec{z} \in \mathbb{D}, \tag{2.42}$$

where $r\left(\vec{z}_{\text{ini}}\right)$ is the upper bound of the normal form radius of the initial state of the system and r_{max} is the lower bound of the maximum normal form radius corresponding to motion still within the allowed region $\mathbb{D}$ (see Fig. 2.1).

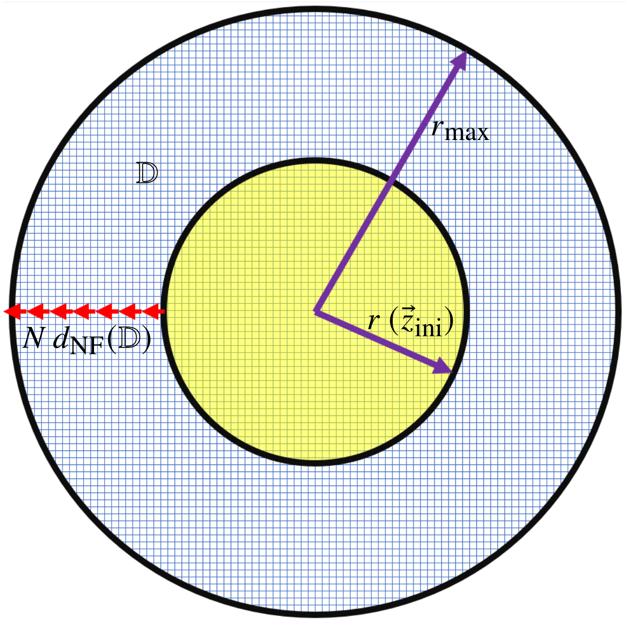
r_{max}
$\mathbb{D}$
$r\left(\vec{z}_{\text{ini}}\right)$
$N\,d_{\text{NF}}(\mathbb{D})$

Figure 2.1: Schematic illustration of the various normal form quantities involved in the calculation of the minimum iteration number within allowed region $\mathbb{D}$.

The concept of a Nekhoroshev-type stability estimate based on the normal form defect is comparable to an augmented Lyapunov function [51]. A regular Lyapunov function L is not increasing along any phase space curve, with $L(\mathcal{M}(\vec{z})) \leq L(\vec{z})$. This works very well for systems with damping. For damped motion in a convex potential, the total energy function can serve as a Lyapunov function. For systems without damping, this is a lot less straightforward. Under the assumption that the normal form algorithm produces a normal form radius which is a true invariant of the motion, the normal form transformation to calculate the normal form radius is a regular Lyapunov function proving eternal stability. However, the errors to the limited floating point accuracy already break this hypothetical scenario. An augmented or pseudo-Lyapunov function $L_\star = L + \max\left(d_{\mathrm{NF}}\left(\mathbb{D}\right)\right)$ is increasing in a very slow and well estimated way with a verified upper bound on the rate of increase per iteration

$$L(\mathcal{M}(\vec{z})) \leq L_\star = L(\vec{z}) + \max\left(d_{\mathrm{NF}}\left(\mathbb{D}\right)\right). \tag{2.43}$$

Thus, it cannot prove eternal stability, but it can rigorously estimate the long term stability. See [23] and [48] for a detailed discussion.

In [29], this method was successfully used to analyze the long term stability of the Tevatron storage ring at the Fermi National Accelerator Laboratory. However, it can be generally used in dynamical systems applications to assess stability. Particularly, in complex systems where the stability in different phase space regions is not evident, the Nekhoroshev-type stability estimate based on the normal form defect is a great tool to capture the maximum rate of divergence.

2.5 Verified Computations Using Taylor Models (TM)

Based on DA vectors (Sec. 2.1), Taylor Models (TM) were developed by Makino and Berz [53, 58, 54, 55, 21, 75] as a structure for rigorously verified computations, which deals much better with issues known from interval arithmetic like the dependency problem [55], the wrapping effect [62, 60, 24], and linear scaling of the overestimation with domain size. Accordingly, the following introduction to TM and their application is largely based on their work [53, 58, 54, 55, 21, 75, 62].

To better understand the advantages of TM, we will first take a quick look at the alternative of using interval arithmetic for verified computations.

2.5.1 Interval Arithmetic

Intervals are a basic concept to represent a range of numbers and are often used to capture uncertainty. The interval $I = [a, b] = \{x \,|\, a \leq x \leq b\}$ represents all numbers between a and b, and the values a and b themselves.

The basic interval arithmetic [67, 68, 47] for the addition, subtraction, multiplication, and division of two intervals $I_1 = [a_1, b_1]$ and $I_2 = [a_2, b_2]$ are given by the following operations. The addition yields

$$I_1 + I_2 = [a_1 + a_2, b_1 + b_2] . \tag{2.44}$$

The subtraction operation $I_1 - I_2$ works equivalently by performing the addition of I_1 with $-I_2 = [-b_2, -a_2]$.

The multiplication yields

$$I_1 \cdot I_2 = [\min(a_1a_2, a_1b_2, b_1a_2, b_1b_2), \max(a_1a_2, a_1b_2, b_1a_2, b_1b_2)] . \tag{2.45}$$

The division is only possible if the divisor interval does not contain zero. If the divisor does not contain zero, the division I_1/I_2 is equivalently defined by multiplying I_1 with

$$\frac{1}{I_2} = \left[\frac{1}{b_2}, \frac{1}{a_2}\right] \quad \text{for} \quad 0 \notin I_2. \tag{2.46}$$

This arithmetic provides the mathematically tightest bounds for certain operations like the square of an interval, and when the quantities represented by I_1 and I_2 are independent. However, in many cases, the calculated bounds are an overestimation due to the dependency problem, which is easily illustrated by considering the difference between an interval and itself. The result of the expression $x - x$ should be zero, but from the arithmetic above the difference between two identical intervals is

$$I - I = [a, b] - [a, b] = [-(b - a), (b - a)] , \tag{2.47}$$

which has a width of $2(b - a)$ instead of zero width.

Compared to DA vectors (see Sec. 2.1), which form a ring structure, intervals do not even form a group structure, because neither for addition nor multiplication there is an inverse for intervals of nonzero width.

For the interval evaluation of functions, further rules can be established. Monotonically increasing functions $f_{\text{mon}\nearrow}$ like $\exp(x)$ can be evaluated by

$$f_{\text{mon}\nearrow}([a, b]) = \left[f_{\text{mon}\nearrow}(a), f_{\text{mon}\nearrow}(b)\right] . \tag{2.48}$$

Monotonically decreasing functions $f_{\text{mon}\searrow}$ can be equivalently evaluated by

$$f_{\text{mon}\searrow}([a,b]) = \left[f_{\text{mon}\searrow}(b), f_{\text{mon}\searrow}(a)\right]. \tag{2.49}$$

Trigonometric functions are compositions of monotonically increasing and monotonically decreasing sections, which are well known. Accordingly, the interval evaluation of a trigonometric function can be implemented based on many subcases depending on the size and position of the interval.

Considering the function $f(x) = \sin(\pi x/2) - \exp(x)$ and evaluating it over the domain interval $I_1 = [-1, 1]$ yields

$$f\left(I_1\right) = \sin\left(\frac{\pi I_1}{2}\right) - \exp\left(I_1\right) = I_1 - \left[\exp\left(-1\right), \exp\left(1\right)\right] \tag{2.50}$$

$$= \left[-1 - e, 1 - e^{-1}\right] \subset [-3.718282, 0.632121]. \tag{2.51}$$

We will compare this interval evaluation to the performance of different order Taylor Models in the following section.

2.5.2 Taylor Models

Taylor Models [53, 58, 54, 55, 21, 75] are remainder-enhanced DA vectors. The DA part of the TM of the function f is an mth order Taylor polynomial in form of a regular DA vector. The remainder part complements this by rigorously verified bounds on the error of using the truncated Taylor expansion of f up to order m in form of a DA vector compared to f itself. In contrast to regular DA vectors, the function f must be $(m + 1)$ times continuously partially differentiable to evaluate the reminder using the Taylor Remainder Theorem. Additionally, TM need to be defined over a domain $\mathbb{D}$ to be able to rigorously bound the remainder.

The Taylor Remainder Theorem says: Given a function $f : \vec{\mathbb{D}} = \left[\vec{a}, \vec{b}\right] \subset \mathbb{R}^n \to \mathbb{G} \subset \mathbb{R}$ being $(m + 1)$ times continuously partially differentiable on the domain $\vec{\mathbb{D}}$ with $\vec{x}_0 \in \vec{\mathbb{D}}$. Then for each

$\vec{x} \in \vec{\mathbb{D}}$ there is an $\eta \in (0, 1)$ such that

$$f(\vec{x}) = \underbrace{\sum_{k=0}^{m} \left. \frac{\left((\vec{x} - \vec{x}_0) \cdot \vec{\nabla}_{\vec{y}}\right)^k f(\vec{y})}{k!} \right|_{\vec{y}=\vec{x}_0}}_{\mathcal{P}_{m,f}} + \underbrace{\left. \frac{\left((\vec{x} - \vec{x}_0) \cdot \vec{\nabla}_{\vec{y}}\right)^{m+1} f(\vec{y})}{(m+1)!} \right|_{\vec{y}=\vec{x}_0+(\vec{x}-\vec{x}_0)\eta}}_{\mathcal{E}_{m,\mathbb{D},f}}, \tag{2.52}$$

where $\mathcal{P}_{m,f}$ is the polynomial part and $\mathcal{E}$ is an expression for the remainder.

A Taylor Model is characterized by its order m, the function f that it is representing, and the domain $\mathbb{D}$ over which the representation of f is within the verified bounds of the Taylor Model. We denote a Taylor Model with

$$\mathcal{T}_{m,\mathbb{D},f} = \left(\mathcal{P}_{m,f}, \epsilon_{m,\mathbb{D},f}\right), \tag{2.53}$$

where $\mathcal{P}_{m,f}$ is the Taylor polynomial term of order m and $\epsilon_{m,\mathbb{D},f}$ is a rigorous verified estimation $\epsilon_{m,\mathbb{D},f}$ of the remainder size over the domain $\mathbb{D}$ such that for function f

$$\left|f(\vec{x}) - \mathcal{P}_{m,f}(\vec{x})\right| \leq \epsilon_{m,\vec{\mathbb{D}},f} \quad \forall \vec{x} \in \vec{\mathbb{D}}. \tag{2.54}$$

A Taylor Model can be visualized as a tube that wraps around the mth order DA representation with a distance ϵ such that the original expression is guaranteed to lie within the tube over the given domain $\mathbb{D}$ (see Fig. 2.2).

Except for order $m = 1$, the Taylor Model bounding of f significantly outperforms the interval bounding. The tightness of the bounding also improves drastically with higher order Taylor Models. With every additional order, the polynomial part clings closer to f, and the reminder gets smaller and smaller.

This tighter and tighter bounding with higher orders shows how the DA part of the Taylor Models avoids more and more of the dependency problem. Dependent expressions like $1 + x - x$, which may arise as the first order part of expressions like $\exp(x) - \sin(x)$ are reduced to just $1 + 0$ in the DA part of the Taylor Model description. As we saw in Sec. 2.5.1, interval arithmetic is not able to avoid this dependency problem.

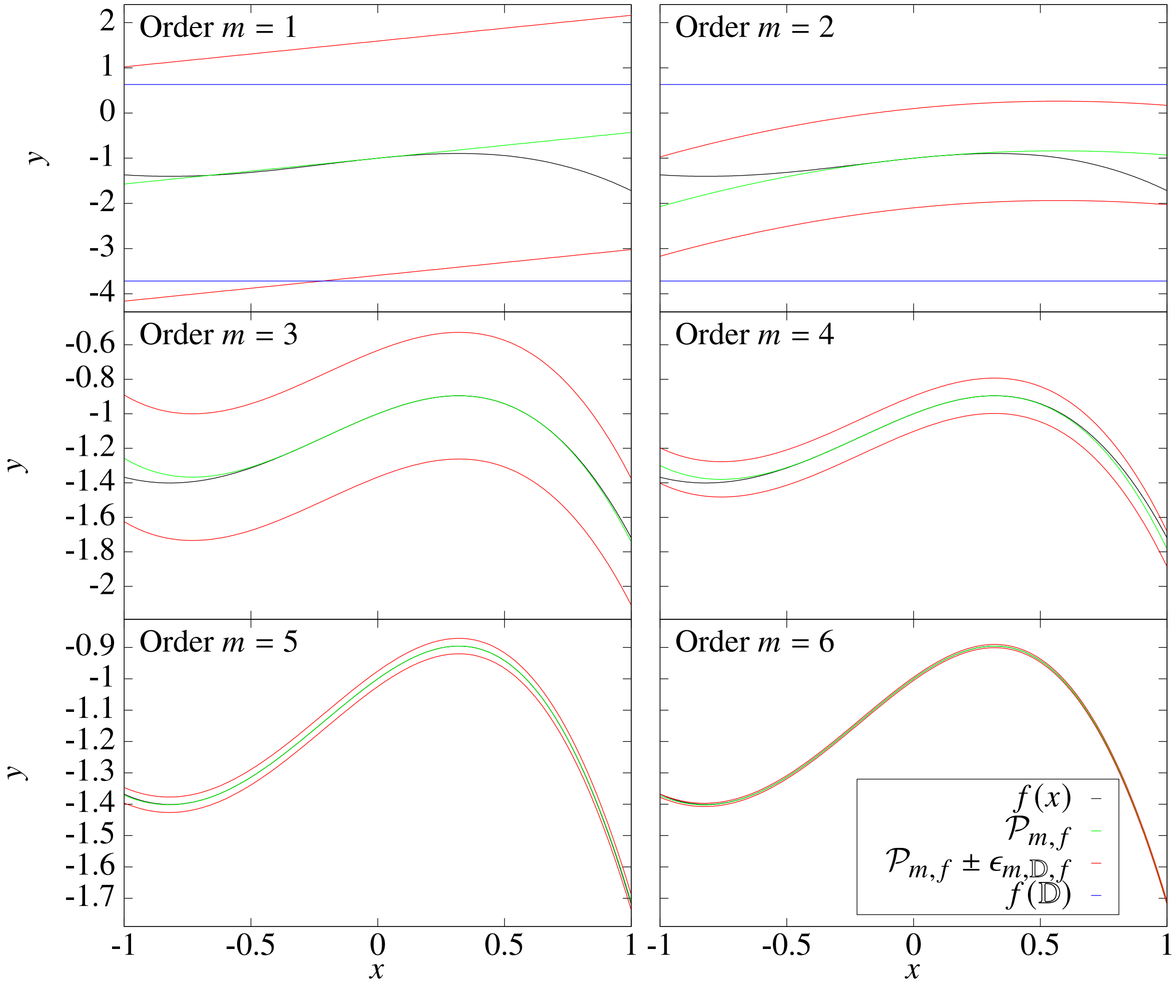

Figure 2.2: Verified representation of $f(x) = \sin(\pi x/2) - \exp(x)$ over the domain $\mathbb{D} = I_1 = [-1, 1]$ with interval methods using $f(\mathbb{D})$ and with Taylor Models $(\mathcal{P}_{m,f}, \epsilon_{m,\mathbb{D},f})$ of various orders m. The original function $f(x)$ is indicated by the black line, while its DA polynomial representation is shown in green. The bounds at a distance $\epsilon_{m,\mathbb{D},f}$ from the DA polynomial are red. The two straight blue lines indicate the bounds of the interval evaluation. Note that the scale of the y axis is changing to better illustrate the tightness of the Taylor Model representation with higher orders. Accordingly, the interval bounds are only shown for order $m = 1$ and order $m = 2$.

The fourth order Taylor Model representation of the function $f(x) = \sin(\pi x/2) - \exp(x)$ over the domain $\mathbb{D} = I_1 = [-1, 1]$ would be

$$\mathcal{T}_{4,I_1,f}(x) = \left(-1 + \frac{(\pi - 2)x}{2} - \frac{x^2}{2} - \frac{(\pi^3 - 8)x^3}{48} - \frac{x^4}{24}, 0.102345\right). \tag{2.55}$$

2.6 Taylor Model based Verified Global Optimizers

The goal of a verified global optimizer [12, 69, 29, 63, 57, 43] is finding the optimum of a given scalar objective function $f(\vec{x})$ of n_{var} variables x_i over a predefined n_{var} dimensional global search domain box $\vec{\mathbb{B}}$. Without loss of generality, it is assumed that the optimum is a minimum. If the optimum is a maximum, consider the optimization of $-f(\vec{x})$.

Ideally, the result of global optimization yields the minimum $f^{\star}$ of the objective function $f(\vec{x})$ and all locations $\vec{x}^{\star}$, where the minimum is assumed within the global search domain box $\vec{\mathbb{B}}$. However, straightforward and exact analytic solutions of the optimization problem only exist for elementary objective functions. As soon as higher order terms and multiple variables are involved, iterative algorithms to track down the optimum are inevitable. Consequently, results are often only approximations of the actual minimum and all their locations where it is assumed. Verified global optimizers compensate for the shortcoming of being unable to pinpoint the exact minimum by yielding rigorously verified bounds on the minimum and its locations.

The fundamental idea of a verified global optimization algorithm is the efficient elimination of subdomains/subboxes of the initial search box $\vec{\mathbb{B}}$ by proving that those eliminated subboxes do not contain the minimum. The basic steps of the algorithm are the following:

1. Split domain box $\vec{\mathbb{B}}$ into subdomains $\vec{\mathbb{B}}_i$.
2. Determine a lower bound $f_{i,\text{LB}}$ of f over $\vec{x} \in \vec{\mathbb{B}}_i$.
3. Calculate/Update the cutoff value $\mathcal{C}$ – the currently lowest known upper bound of the minimum.
4. Eliminate all boxes $\mathbb{B}_i$ with a lower bound $f_{i,\text{LB}}$ larger than the cutoff value $\mathcal{C}$.
5. Restart the algorithm at step 1 for each of the non-eliminated domain boxes $\vec{\mathbb{B}}_i^{\#}$.

The more subdomain boxes are eliminated in step 4 in each iteration, the more effective the algorithm. Accordingly, it is essential to use methods for very tight bounding in step 2 (making $f_{i,\text{LB}}$ as large as possible), and to use heuristics to significantly improve the cutoff value $\mathcal{C}$ in step 3, making it as small as possible.

For the determination of the cutoff value $\mathcal{C}$ in step 3, any method or combination of methods that produce a tight verified upper bound on the global minimum of the search domain are useful. A typical technique is the verified evaluation of individual points within the domain box. The testing points are chosen either randomly in a Monte-Carlo based approach or by heuristics, e.g., the results of non-verified optimization over the domain. Depending on the computational effort of those methods, the improvement of the cutoff value and its benefits for the algorithm must be weighed against the computation time of the cutoff method.

For step 2, Taylor Models (see Sec. 2.5) are particularly useful, especially high order Taylor Models, since they allow for very tight bounding compared to interval methods. This property can mainly be ascribed to the avoidance of the dependency problem due to the DA vector part of the TM. For very complex objective functions like the normal form defect (see Sec. 2.4), the evaluation with very high order Taylor Models (e.g. order ten) can take considerably more time compared to evaluations with lower order Taylor Models (e.g. order three). Again, the benefits of the more precise bounding with high order Taylor Model evaluation have to be weighed against the associated computation time. A rule of thumb is that the larger the evaluation domain and the more complex the objective function, the larger the benefit of higher order Taylor Models.

For the rigorous bounding of Taylor Models, there are multiple approaches. The standard method uses order bounds, where the terms belonging to each order are bound and summed up together with the remainder bound. More sophisticated methods are discussed in great detail in [64]. They can be briefly summarized as follows. The linear dominated bounder (LDB) is very efficient for linear dominated domains. The quadratic dominated bounder (QDB) is good at determining the minimum of a multidimensional quadratic dominated function but losses its efficiency with very high dimensional problems. The quadratic fast bounder (QFB) is not as exact as the QDB but very efficient in providing a good lower bound near a local minimum, where the Hessian matrix of the objective function over the domain is positive definite.

To avoid an infinite continuation of the splitting, stop conditions are implemented, which are checked before a domain box is split. A typical stop condition sets a lower bound on the size of the domain, either by setting a lower bound on the volume of the domain box or its side length. Non-eliminated domain boxes below such a threshold values are not split.

Another possible stop condition is a lower bound on the tightness of the bounding of the minimum of the objective function rather than the domain size. With such a stop condition in place, the algorithm would not split a non-eliminated domain box over which the bounds of the minimum are tighter than a certain given value. This is particularly useful if the exact minimum is not relevant but rather the order of magnitude of the minimum.

CHAPTER 3

AN EXAMPLE-DRIVEN WALK-THROUGH OF THE DA NORMAL FORM ALGORITHM

This chapter is based on my arXiv preprint and MSU Report MSUHEP-190617 *Introduction to the Differential Algebra Normal Form Algorithm using the Centrifugal Governor as an Example* [94].

We provide a very detailed description of the steps involved in the DA normal form algorithm (Sec. 2.3) and their implications for the normal form using the example of the centrifugal governor. We pick this example because it is one dimensional and the derivation of the equations of motion and the linearization of the motion are well known. This understanding yields the groundwork for the non-trivial analysis of the nonlinear phenomena using the steps of the DA normal form algorithm.

3.1 The Centrifugal Governor

The centrifugal governor (see Fig. 3.1) is a device involving gravitational and centrifugal forces with the rotation axis parallel to the direction of the gravitational force. We consider a mathematically idealized governor, which consists of two massless rods of equal length R suspended in a common

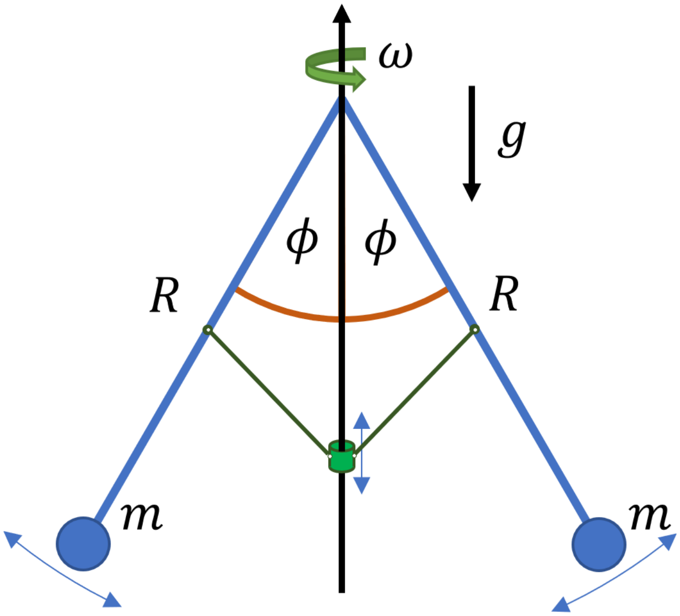

Figure 3.1: Schematic illustration of centrifugal governor.

plane with the rotation axis. A point mass m is attached at the end (opposite to where the rod is mounted) of each of the rods. The angle between the rotation axis and the rod is denoted by the angle ϕ. A mechanism links the two rods and the rotation axis, which guarantees identical angles and therefore identical behavior on both sides. An external torque applied via the rotation axis ensures that the rotation frequency ω of the centrifugal governor arms is kept constant.

In the usual application of a centrifugal governor, the rotation frequency is not fixed but negatively coupled to the angle ϕ through an additional mechanism external to the governor itself. This additional mechanism makes the system self regulating by decreasing ω for an increase in ϕ. Accordingly, in those applications, e.g. the steam engine, the rotation frequency ω changes during the regulating process. However, as already mentioned above, for the introduction to the DA normal form algorithm, we consider the motion of the system for a fixed rotation frequency ω, i.e. no self-regulating coupling mechanism between ϕ and ω.

3.1.1 Units

To limit the number of parameters in the following calculations to just the rotation frequency ω, we scale time, distance, and mass in such a way that the mass m, the gravitational constant g, and the length of the rods R are all equal to one in their respective scaled units and therefore disappear from the equations. Specifically, mass is considered in units of the point mass m, distances are considered in units of the rod length R, and time is considered in units of

$$T_0[\mathrm{s}] = \sqrt{\frac{R[\mathrm{m}]}{g\left[\frac{\mathrm{m}}{\mathrm{s}^2}\right]}}, \tag{3.1}$$

such that the gravitational constant g equal one in units of distance R and time T_0.

3.1.2 The Equilibrium Point

For any given fixed rotation frequency ω, there is an angle ϕ_0 so that $\phi(t) = \phi_0$ is a solution of the motion of the centrifugal governor arms. This equilibrium angle is characterized by the alignment of the rods with the vector sum of the vertical gravitational force F_{grav} and the radial centrifugal force F_{cent} such that there is no torque acting on the rods in the common plane of the rods and the rotation axis.

For any frequency ω, $\phi_0 = 0$ satisfies this requirement, since the centrifugal force is zero and there is only the gravitational force acting vertically downwards. However, if the rotation frequency ω is sufficiently high enough (see Eq. (3.3)), a bifurcation of the equilibrium angle occurs – the angle $\phi_0 = 0$ becomes an unstable equilibrium state, while stable equilibrium angle $\phi_0(\omega) > 0$ arises, which satisfies the alignment condition with

$$\tan\phi_0 = \frac{F_{\text{cent}}}{F_{\text{grav}}} = \frac{m\omega^2 R \sin\phi_0}{mg} = \omega^2 \sin\phi_0. \tag{3.2}$$

For $\phi_0 > 0$, this corresponds to

$$\cos\phi_0 = \frac{1}{\omega^2} \quad \Rightarrow \quad \phi_0 = \arccos\left(\frac{1}{\omega^2}\right) \quad \text{for} \quad \omega > 1 = \omega_{\text{min}}. \tag{3.3}$$

Fig. 3.2 visualizes the stable equilibrium angle as a function of the rotation frequency ω.

Since the vertical contribution of the gravitational force to the vector sum is nonzero and independent of the rotation frequency, an equilibrium angle of $\phi_0 = 90°$ is only approached asymptotically for the rotation frequency ω approaching infinity. The bifurcation of the equilibrium state at $\omega_{\text{min}} = 1$ is also clearly visible.

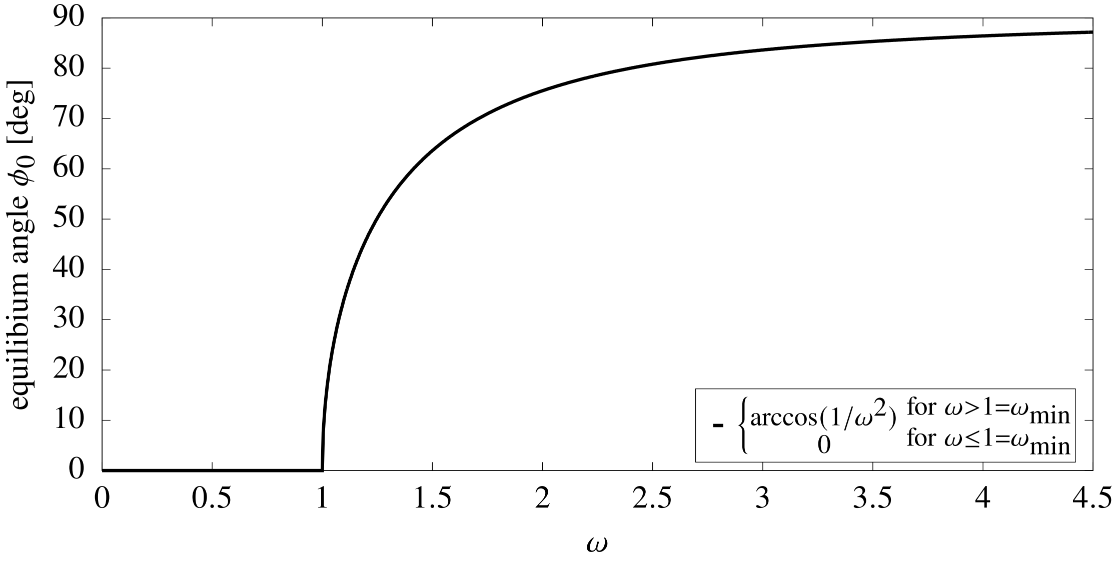

Figure 3.2: Illustration of the stable equilibrium angle ϕ_0 of the arms of the centrifugal governor as a function of the rotation frequency ω. For $\omega > \omega_{\text{min}} = 1$, $\phi_0 = 0$ is an unstable equilibrium angle.

Tab. 3.1 lists stable equilibrium angles for some specific rotation frequencies, especially for the fast-changing region between $\omega = 1$ and $\omega = 2$.

Table 3.1: List of stable equilibrium angles ϕ_0 of the centrifugal governor arms for some specific rotation frequencies ω.

ω	ϕ_0 [deg]	ϕ_0 [rad]
1	0°	0
$\sqrt{2}/\sqrt[4]{3}$	30°	$\pi/6$
$\sqrt[4]{2}$	45°	$\pi/4$
$\sqrt{2}$	60°	$\pi/3$
2	$\approx 75.52°$	≈ 1.318
20	$\approx 89.86°$	≈ 1.568
$\lim_{\omega\to\infty}$	90°	$\pi/2$

3.1.3 The Equations of Motion

To understand the dynamics of the centrifugal governor arms around an equilibrium state, we derive the equations of motion for one of the two masses starting with the Lagrangian formulation of the problem. It yields

$$L = \frac{m}{2}\left(\dot{\phi}^2 R^2 + \omega^2 R^2 \sin^2\phi\right) - mgR\left(1 - \cos\phi\right) = \frac{\dot{\phi}^2}{2} - \underbrace{\left(\frac{-\omega^2 \sin^2\phi}{2} + (1 - \cos\phi)\right)}_{U_{\text{eff}}}, \tag{3.4}$$

where U_{eff} is the effective or centrifugal-gravitational potential. In Fig. 3.3, we illustrate the centrifugal-gravitational potential U_{eff} for multiple rotation frequencies ω.

The minimum of the effective potential well corresponds to the stable equilibrium angle discussed in Sec. 3.1.2. The axis notations indicate that the width and the depth of the potential increase with increasing rotation frequency ω. The higher the rotation frequency ω, the less relevant are the gravitational influences and the deeper and the more symmetric the potential well. The asymmetry of the effective potential is also apparent in the dynamics of the system, which we discuss in Sec. 3.1.4.

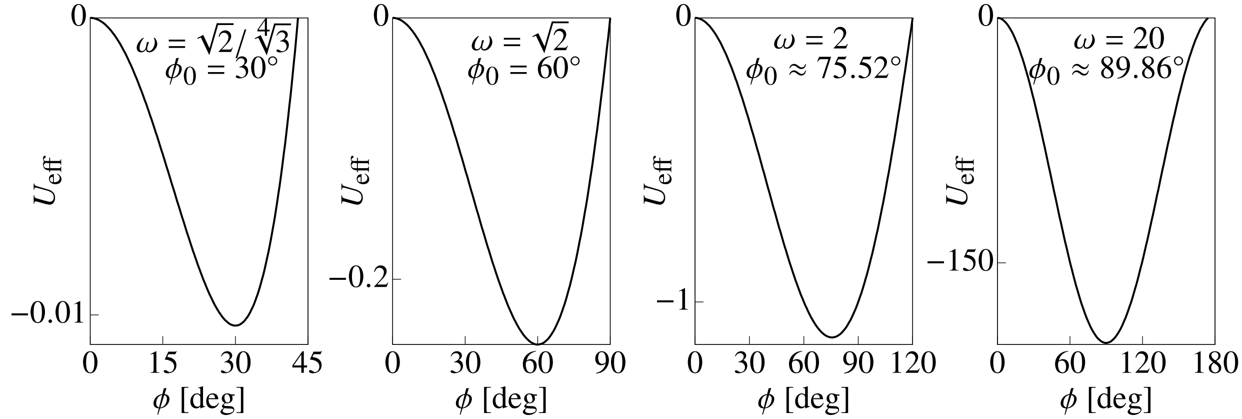

Figure 3.3: Potential well of U_{eff} for multiple oscillation frequencies ω. The equilibrium angle ϕ_0 corresponds to the minimum of the potential well.

For the rest of the chapter, we will focus on the case $\omega = \sqrt{2}$, which yields a clear 2:1 asymmetry left and right of its equilibrium angle.

To continue the derivation of the equations of motion, we derive the generalized canonical momentum p_ϕ to the position variable ϕ from the Lagrangian, where

$$p_\phi = \frac{dL}{d\dot{\phi}} = mR^2\dot{\phi} = \dot{\phi}. \tag{3.5}$$

Using the Legendre transformation, the Hamiltonian

$$H = \frac{p_\phi^2}{2mR^2} - \frac{m\omega^2 R^2 \sin^2\phi}{2} + mgR\,(1 - \cos\phi) = \frac{p_\phi^2}{2} + U_{\text{eff}} = E \tag{3.6}$$

is obtained, which is not explicitly time dependent and therefore a constant of motion. The Hamiltonian also happens to correspond to the energy E of this system.

The equations of motions are derived from the Hamiltonian via Hamilton's equations where

$$\dot{\phi} = \frac{dH}{dp_\phi} = \frac{p_\phi}{mR^2} = p_\phi \tag{3.7}$$

$$\text{and} \quad \dot{p}_\phi = -\frac{dH}{d\phi} = -mgR\sin\phi + m\omega^2R^2\sin\phi\cos\phi = \sin\phi\left(\omega^2\cos\phi - 1\right). \tag{3.8}$$

In coordinates $(\delta\phi, \delta p_\phi)$ relative to the equilibrium state $(\phi_0, 0)$, the equations of motions are

$$\frac{\mathrm{d}\delta\phi}{\mathrm{d}t} = \delta p_\phi \qquad \text{and} \qquad \frac{\mathrm{d}\delta p_\phi}{\mathrm{d}t} = \sin\left(\phi_0 + \delta\phi\right)\left(\omega^2\cos\left(\phi_0 + \delta\phi\right) - 1\right). \tag{3.9}$$

3.1.4 Illustration of System Dynamics

With the equations of motion relative to the equilibrium state (Eq. (3.9)) and the understanding of how the shape of the effective potential well changes with the rotation frequency ω, we can now interpret the dynamics of the centrifugal governor when the angle of the rods is perturbed from the equilibrium angle $\phi_0(\omega)$.

In Fig. 3.4, the dynamics of the rods are shown for a rotation frequency of $\omega = \sqrt{2}$, which corresponds to an equilibrium angle of $\phi_0 = 60°$. While the oscillation is periodic, it is asymmetric around the equilibrium point, as we would expect from the asymmetric effective potential for $\omega = \sqrt{2}$ in Fig. 3.3. The asymmetry of the oscillation is larger, the larger the angle during initiation. The maximum downward angle displacement (often more generally referred to as amplitude) and the maximum upward angle displacement of the governor's arms relative to their equilibrium angle are related through the effective potential, which corresponds to the energy of the vertical motion for

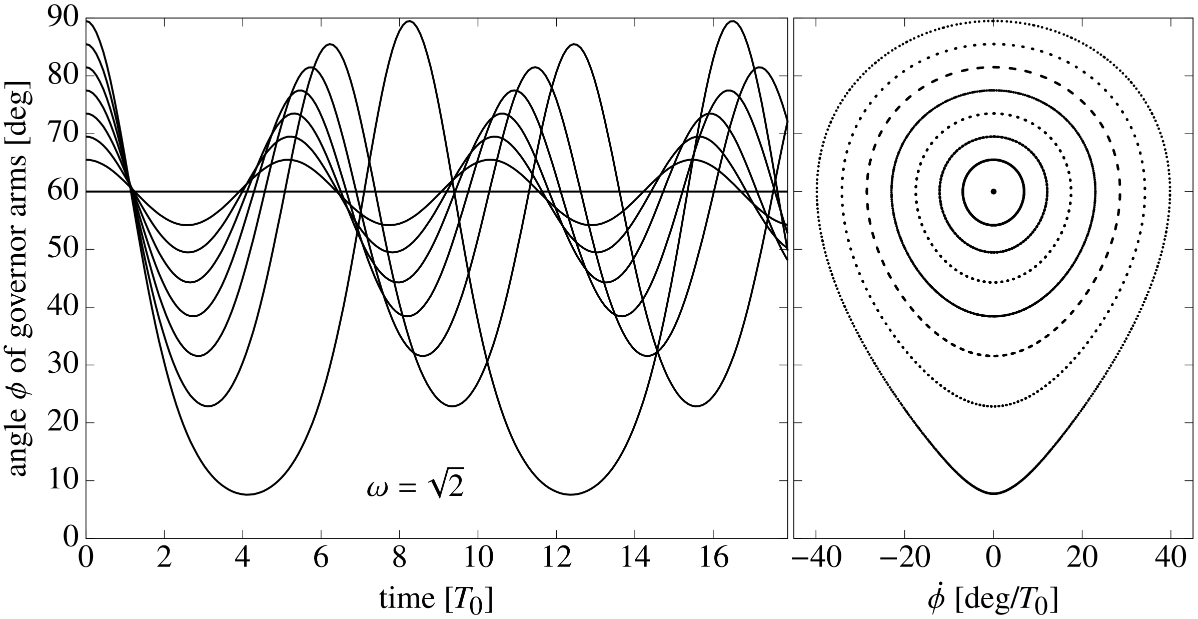

Figure 3.4: Dynamics of the centrifugal governor for a rotation frequency of $\omega = \sqrt{2}$. The centrifugal governor arms were initiated with $\dot{\phi} = p_\phi = 0$ and at the following angles: 60°, 65.5°, 69.5°, 73.5°, 77.5°, 81.5°, 85.5°, and 89.5°. The left plot shows the oscillatory behavior around the equilibrium angle at $\phi_0 = 60°$ over time. The right plot shows the stroboscopic phase space behavior from repetitive map evaluation. To relate the phase space behavior to the position behavior in time, the ϕ axis of both plots are aligned.

$\delta p_\phi = 0$. For both those angle displacements, the effective potential has the same maximum value or 'invariant amplitude' corresponding to the energy. The maximum amplitudes in the momentum space in the right plot of Fig. 3.4 are related the same way. In other words, the phase space motion in Fig. 3.4 corresponds to contour lines of the energy.

For future reference, it is useful to associate the term 'amplitude' not only with a physical displacement or a maximum/minimum momentum but also with an abstract quantity that relates all the different versions of phase space amplitudes like the energy in this case.

Apart from the asymmetric upward and downward position amplitudes, the left plot in Fig. 3.4 clearly shows a change in the period of oscillation depending on the angle during initiation, or more generally speaking, depending on the invariant amplitude of the motion, e.g., the total energy of the system. The larger the amplitude, the longer is the period of oscillation. This is particularly prominent for the oscillation with the largest amplitude. It is also obvious, especially for the larger amplitudes that the relation between the amplitude and the period is nonlinear.

However, there is no trivial way of extracting this nonlinear relation between the amplitude and the period of oscillation from the equations of motion and/or the energy. Additionally, if we were unaware of the function for the effective potential and energy, or were considering a more complex system, it would also be very difficult to relate the different phase space amplitudes to each other. The DA normal form algorithm generates both relations in an automated process up to calculation order. In the order-by-order process, it determines an invariant amplitude up to calculation order as a function of the original phase space variables and also determines the period of oscillation as a function of that invariant amplitude.

All the normal form algorithm requires is an origin preserving transfer map (see Sec. 2.2), which represents the flow of the ODEs (see Eq. (3.9)) relative to the linearly stable fixed point of the considered phase space motion. For the centrifugal governor example, the equilibrium phase space state $(\phi_0, 0)$ constitutes such a phase space fixed point, as the right plot in Fig. 3.4 already indicated. In other words, we require a functional description of how the relative phase space state $z_{\text{fin}} = (\delta\phi_{\text{fin}}, \delta p_{\phi,\text{fin}})$ after a fixed time t_0 depends on the initial relative phase space state $z_{\text{ini}} = (\delta\phi_{\text{ini}}, \delta p_{\phi,\text{ini}})$. DA based maps (Sec. 2.2) can provide this functional description up to arbitrary order. We will use them to represent the dynamics around the equilibrium state corresponding to a rotation frequency of $\omega = \sqrt{2}$, for the later analysis with the DA normal form algorithm.

3.2 Map Calculation via Integration

As mentioned above, the following analysis of the centrifugal governor considers the system at a fixed rotation frequency of $\omega = \sqrt{2}$. We are interested in the dynamics around the corresponding equilibrium state of the centrifugal governor arms at $(60°, 0)$. The goal of this section is to generate a DA map describing the phase space dynamics relative to that equilibrium state.

For consistency with the notation introduced in Sec. 2.3, we denote the phase space coordinates relative to the equilibrium point with (q_0, p_0) instead of the previously used $\left(\delta\phi, \delta p_\phi\right)$. We will also conduct the calculations in radians rather than degrees due to their slightly easier implementation.

The map is calculated by integrating the ODEs (see Eq. (3.9)) from the initial phase space state

$$(q_{\text{ini}}, p_{\text{ini}}) = \left(\phi_0\left(\omega = \sqrt{2}\right) + \delta\phi, \delta p_\phi\right) = \left(\frac{\pi}{3} + q_0, p_0\right) \tag{3.10}$$

from $t = 0$ until $t = t_0 = 1$. Since the flow of the ODEs in Eq. (3.9) remains expanded around the equilibrium state for any t_0, the time of the integration can be chosen freely.

The resulting map of the integration $\mathcal{M}_0 = (Q(q_0, p_0), P(q_0, p_0))^T$ has the following form: $\mathcal{M}_0 = \mathcal{C} + \mathcal{L} + \sum_m \mathcal{U}_m$, where the constant part is denoted by $\mathcal{C}$, the linear part with $\mathcal{L}$ and each of the nonlinear parts of order m with $\mathcal{U}_m$. Since the system is expanded around the equilibrium point, the constant part of the map corresponds to the equilibrium state $(\pi/3, 0)$. The following explicit formulation of $\mathcal{M}_0$ up to order three introduces the notation of various coefficients of the map:

$$\begin{aligned}
\mathcal{M}_0\left(q_0, p_0\right) = \begin{pmatrix} \mathcal{M}_0^+\left(q_0, p_0\right) \\ \mathcal{M}_0^-\left(q_0, p_0\right) \end{pmatrix} = \begin{pmatrix} Q\left(q_0, p_0\right) \\ P\left(q_0, p_0\right) \end{pmatrix} &= \underbrace{\begin{pmatrix} q_{\text{const}} \\ p_{\text{const}} \end{pmatrix}}_{\mathcal{C}} + \underbrace{\begin{pmatrix} (Q|q_0) & (Q|p_0) \\ (P|q_0) & (P|p_0) \end{pmatrix}\begin{pmatrix} q_0 \\ p_0 \end{pmatrix}}_{\mathcal{L}} \\
&\underbrace{+\begin{pmatrix} \mathcal{U}^+_{2(2,0)} \\ \mathcal{U}^-_{2(2,0)} \end{pmatrix} q_0^2 + \begin{pmatrix} \mathcal{U}^+_{2(1,1)} \\ \mathcal{U}^-_{2(1,1)} \end{pmatrix} q_0 p_0 + \begin{pmatrix} \mathcal{U}^+_{2(0,2)} \\ \mathcal{U}^-_{2(0,2)} \end{pmatrix} p_0^2}_{\mathcal{U}_2} \\
&\underbrace{+\begin{pmatrix} \mathcal{U}^+_{3(3,0)} \\ \mathcal{U}^-_{3(3,0)} \end{pmatrix} q_0^3 + \begin{pmatrix} \mathcal{U}^+_{3(2,1)} \\ \mathcal{U}^-_{3(2,1)} \end{pmatrix} q_0^2 p_0 + \begin{pmatrix} \mathcal{U}^+_{3(1,2)} \\ \mathcal{U}^-_{3(1,2)} \end{pmatrix} q_0 p_0^2 + \begin{pmatrix} \mathcal{U}^+_{3(0,3)} \\ \mathcal{U}^-_{3(0,3)} \end{pmatrix} p_0^3}_{\mathcal{U}_3} + \ldots \qquad (3.11)
\end{aligned}$$

The position Q and momentum P components of the map $\mathcal{M}_0$ correspond to the upper and lower component and are denoted by '+' and '-', respectively. The coefficients in the upper and lower component for the nonlinear $m(= a + b)$th order terms $q^a p^b$ are denoted by $\mathcal{U}^{\pm}_{m(a,b)}$. The coefficients in the linear matrix $(a|b)$ indicate the factor with which a is linearly dependent on b.

The following Tab. 3.2 lists the values of the coefficients in Eq. (3.11) above. The integration was performed with an order 20 Picard-iteration based integrator with stepsize $h = 10^{-3}$ over 1000 iterations within COSY INFINITY. Details on the implementation of the integrator are given in [93].

3.3 The DA Normal Form Algorithm

In Sec. 2.3, the general DA normal form algorithm [19] was introduced for a linearly stable $2n$ dimensional system with optional parameter dependence. This section provides a detailed example-driven walk-through of the differential algebra based normal form algorithm for the symplectic one dimensional (1D) system of the centrifugal governor without a parameter dependence.

The normal form resulting from the DA normal form algorithm constitutes circular motion with a quasi-invariant as radius and only normal form phase space amplitude (and parameter) dependent angle advancements. Fig. 3.5 illustrates the oscillatory phase space behavior of the governor's arms around the equilibrium point (left plot already seen in different orientation in Fig. 3.4) and compares

Table 3.2: Integration result for map around equilibrium state ($\phi_0(\omega = \sqrt{2}) = \pi/3, 0$) integrated until $t = 1$ using an order 20 Picard-iteration based integrator with stepsize $h = 10^{-3}$ over 1000 iterations within COSY INFINITY. The component $\mathcal{M}_0^+ = Q(q_0, p_0)$ is on the left, $\mathcal{M}_0^- = P(q_0, p_0)$ on the right.

Order	Coeff.	Value	Coeff.	Value
0	q_{const}	1.04719755	p_{const}	0
1	$(Q\|q_0)$	0.33918599	$(P\|q_0)$	-1.15214118
1	$(Q\|p_0)$	0.76809412	$(P\|p_0)$	0.33918599
2	$\mathcal{U}^+_{2(2,0)}$	-0.44622446	$\mathcal{U}^-_{2(2,0)}$	-0.55821731
2	$\mathcal{U}^+_{2(1,1)}$	-0.29304415	$\mathcal{U}^-_{2(1,1)}$	-0.64033440
2	$\mathcal{U}^+_{2(0,2)}$	-0.08403817	$\mathcal{U}^-_{2(0,2)}$	-0.29304415
3	$\mathcal{U}^+_{3(3,0)}$	0.31844278	$\mathcal{U}^-_{3(3,0)}$	0.50817317
3	$\mathcal{U}^+_{3(2,1)}$	0.29904862	$\mathcal{U}^-_{3(2,1)}$	0.76091921
3	$\mathcal{U}^+_{3(1,2)}$	0.13758223	$\mathcal{U}^-_{3(1,2)}$	0.46230241
3	$\mathcal{U}^+_{3(0,3)}$	0.03017663	$\mathcal{U}^-_{3(0,3)}$	0.13758223

it to its associated rotationally invariant phase space behavior in the normal form representation. The orientation of the phase space in Fig. 3.5 is according to the usual convention, where the position q is on the horizontal axis and the momentum p on the vertical axis. In Fig. 3.4, this convention

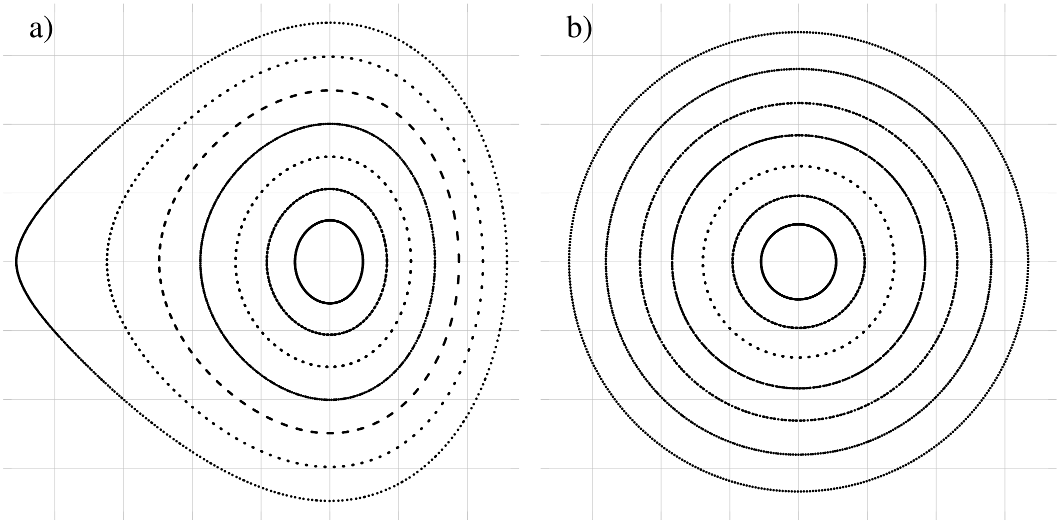

Figure 3.5: Phase space behavior of the centrifugal governor arms around their equilibrium angle of $\phi_0(\omega = \sqrt{2}) = 60°$ provided by a tenth order Poincaré map of the system. a) shows the original phase space behavior. b) shows the associated circular behavior in normal form.

was ignored for the sake of a better understanding when comparing the phase space behavior to the position behavior over time. Accordingly, the asymmetry with larger downwards amplitudes is shown in the horizontal (ϕ) direction in Fig. 3.5a.

The transformation steps of the normal form algorithm are done order by order. With each transformation step, the index of the map and the variables is going to increase by 1, i.e. as a result of the first (order) transformation we get $\mathcal{M}_1$ dependent on the variables (q_1, p_1). For each order m there is a transformation $\mathcal{A}_m$ and its inverse $\mathcal{A}_m^{-1}$, which are applied to resulting map of the previous transformation $\mathcal{M}_{m-1}$ to yield the resulting map of the mth order transformation

$$\mathcal{M}_m(q_m, p_m) = (\mathcal{A}_m \circ \mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1})(q_m, p_m). \tag{3.12}$$

The transformation $\mathcal{A}_m^{-1}$ transforms (q_m, p_m) to (q_{m-1}, p_{m-1}), which are the variables of the map of the previous order $\mathcal{M}_{m-1}$. The transformation $\mathcal{A}_m$ transforms the intermediate result of $\mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1}$, which is in the (q_{m-1}, p_{m-1}) phase space, back to the new phase space in (q_m, p_m).

The nonlinear normal form transformation steps below are calculated up to third order. It will become obvious during the process that transformations of higher even and odd orders follow the same pattern as the second and third order transformation, respectively.

3.3.1 The Parameter Dependent Fixed Point

The DA normal form algorithm starts with an origin preserving map. Accordingly, the result from the integration is shifted to the equilibrium/fixed point $\mathcal{M}_{\mathrm{FP}} = \mathcal{M}_0 - \mathcal{C}$, hence $\mathcal{M}_{\mathrm{FP}} = \mathcal{L} + \sum_m \mathcal{U}_m$ is an origin preserving fixed point map with $\mathcal{M}_{\mathrm{FP}}(\vec{0}) = \vec{0}$.

If the map were dependent on changes $\delta\eta$ of a system parameter η, e.g., changes in the driving frequency $\omega = \omega_0 + \delta\omega$, the normal form algorithm would require the calculation of the parameter dependent fixed point $\vec{z}(\delta\eta) = (q_{\mathrm{FP}}(\delta\eta), p_{\mathrm{FP}}(\delta\eta))$ such that $\mathcal{M}_{\mathrm{FP}}(\vec{0}, \delta\eta) = \vec{0}$. In Eq. (3.3), the relation of the equilibrium point (fixed point) and the driving frequency was already calculated yielding the parameter dependent fixed point

$$\vec{z}\,(\delta\omega) = \left(\arccos\left(\frac{1}{(\omega_0 + \delta\omega)^2}\right), 0\right) \quad \text{for} \quad (\omega_0 + \delta\omega)^2 \geq 1.$$

For less straightforward systems, one uses the following inversion method on the extended map $(\mathcal{M}_{\mathrm{FP}} - \mathcal{I}_{\vec{z}}, \mathcal{I}_{\delta\vec{\eta}})$ to find the parameter dependent fixed point $\vec{z}(\delta\vec{\eta})$ [19, Eq. (7.47)]:

$$\left(\vec{z}\left(\delta\vec{\eta}\right), \mathcal{I}_{\delta\vec{\eta}}\right) = \left(\mathcal{M}_{\mathrm{FP}} - \mathcal{I}_{\vec{z}}, \mathcal{I}_{\delta\vec{\eta}}\right)^{-1} \left(\vec{0}, \delta\vec{\eta}\right), \tag{3.13}$$

where $\mathcal{I}_{\vec{z}}$ and $\mathcal{I}_{\delta\vec{\eta}}$ are the identity map of $\vec{z}$ and $\delta\vec{\eta}$, respectively.

Given the parameter dependent fixed point, the map is expanded around it:

$$\mathcal{M}_{\mathrm{PDFP}} = \mathcal{M}_{\mathrm{FP}}\left(\vec{z}\left(\delta\vec{\eta}\right) + \vec{z}, \delta\vec{\eta}\right) - \mathcal{M}_{\mathrm{FP}}\left(\vec{z}\left(\delta\vec{\eta}\right), \delta\vec{\eta}\right). \tag{3.14}$$

To limit the complexity of the walk-through of the DA normal form algorithm, we will not consider parameter dependence in the further calculations of this chapter and therefore proceed with $\mathcal{M}_{\mathrm{FP}}$.

3.3.2 The Linear Transformation

The first order transformation is the diagonalization, transforming the system into the eigenvector space of the linear part $\mathcal{L}$. In order to determine the transformation $\mathcal{A}_1$ and its inverse $\mathcal{A}_1^{-1}$ for the diagonalization, we determine the eigenvalues $\lambda_\pm$ and eigenvectors $\vec{v}_\pm$ of the linear matrix $\hat{L}$ in the linear part $\mathcal{L}$. For this, we require that all eigenvalues of $\mathcal{M}_{\mathrm{FP}}$ are distinct. Furthermore, we only consider cases where $\mathcal{M}_{\mathrm{FP}}$ is linearly stable, which means that all eigenvalues have an absolute value $|\lambda| \leq 1$. This also means that $\det(\hat{L}) \leq 1$, otherwise at least one of the eigenvalues is larger than 1, making the system linearly unstable. Particularly interesting is the case $\det(\hat{L}) = 1$, which indicates that the system is symplectic and only stable in the case of complex conjugate eigenvalues $\lambda_\pm = e^{\pm i\mu}$. While there are procedures for the cases of real and degenerate eigenvalues with a magnitude smaller than one (see [19]), this chapter only illustrates the procedures for the most relevant and common symplectic case of only complex conjugate eigenvalues and eigenvectors.

Solving the characteristic polynomial yields the eigenvalues

$$\lambda_\pm = \frac{\mathrm{tr}\left(\hat{L}\right)}{2} \pm \sqrt{\frac{\mathrm{tr}\left(\hat{L}\right)^2}{4} - \det\left(\hat{L}\right)} = re^{\pm i\mu}$$

$$\text{with} \quad r = \sqrt{\det\left(\hat{L}\right)} \quad \text{and} \quad \mu = \mathrm{sign}\left(Q|p_0\right) \arccos\left(\frac{\mathrm{tr}\left(\hat{L}\right)}{2r}\right).$$

To generalize the procedure of diagonalization, the Twiss parameters [32] are used with

$$\alpha = \frac{(Q|q_0) - (P|p_0)}{2r\sin\mu}, \quad \beta = \frac{(Q|p_0)}{r\sin\mu}, \quad \text{and} \quad \gamma = \frac{-(P|q_0)}{r\sin\mu}.$$

With this notation the linear matrix $\hat{L}$ can be generally written as

$$\hat{L} = r \cdot \begin{pmatrix} \cos\mu + \alpha\sin\mu & \beta\sin\mu \\ -\gamma\sin\mu & \cos\mu - \alpha\sin\mu \end{pmatrix}.$$

The complex conjugate eigenvectors $\vec{v}_\pm$ associated with the complex conjugate eigenvalues $\lambda_\pm$ of $\hat{L}$ are then obtained by solving $\left(\hat{L} - \lambda_\pm \mathcal{I}\right)\vec{v}_\pm = \vec{0}$.

As a result, the following eigenvectors are calculated

$$\vec{v}_\pm = \begin{pmatrix} \beta \\ -\alpha \pm i \end{pmatrix} \quad \text{or} \quad \vec{v}_\pm = \begin{pmatrix} \alpha \pm i \\ -\gamma \end{pmatrix},$$

for the case that either $\beta = 0$ or $\gamma = 0$. The transformation $\mathcal{A}_1^{-1}$ consist of the two complex conjugate eigenvectors $\vec{v}_\pm$, guaranteeing that $\mathcal{A}_1^{-1}\,(q_1, p_1)$ is real just like the original variables (q_0, p_0) and the fixed point map $\mathcal{M}_{\mathrm{FP}}$. The transformation $\mathcal{A}_1$ is calculated accordingly such that the resulting map $\mathcal{M}_1 = \mathcal{A}_1 \circ \mathcal{M}_{\mathrm{FP}} \circ \mathcal{A}_1^{-1}$ is in the complex conjugate eigenvector space and has complex conjugate components $\bar{\mathcal{M}}_1^+ = \mathcal{M}_1^-$. For $\beta \neq 0$, the transformations are

$$\mathcal{A}_1^{-1} = \begin{pmatrix} (q_0|q_1) & (q_0|p_1) \\ (p_0|q_1) & (p_0|p_1) \end{pmatrix} = \frac{1}{2\sqrt{\beta}} \begin{pmatrix} \beta & \beta \\ i - \alpha & -i - \alpha \end{pmatrix} \quad \text{and} \tag{3.15}$$

$$\mathcal{A}_1 = \begin{pmatrix} (q_1|q_0) & (q_1|p_0) \\ (p_1|q_0) & (p_1|p_0) \end{pmatrix} = \frac{i}{\sqrt{\beta}} \begin{pmatrix} -i - \alpha & -\beta \\ -i + \alpha & \beta \end{pmatrix}. \tag{3.16}$$

For the centrifugal governor example with $\omega = \sqrt{2}$, the eigenvalues are $\lambda_\pm = re^{\pm i\mu}$ with

$$r = 1 \quad \text{and} \quad \mu = 1.22474487. \tag{3.17}$$

The Twiss parameters are

$$\alpha = 0, \quad \beta = 0.816496581 \approx \sqrt{\frac{2}{3}}, \quad \text{and} \quad \gamma = 1.22474487 \approx \sqrt{\frac{3}{2}}.$$

The resulting diagonalized map is of the form $\mathcal{M}_1 = \mathcal{R} + \sum_m S_m$, where $\mathcal{S}_m$ are the transformed nonlinear parts of order m in the eigenvector space of $\hat{L}$ and $\mathcal{R}$ is the diagonalized linear part, where the linear matrix $\hat{R}$ of $\mathcal{R}$ only consist of the eigenvalues $e^{\pm i\mu}$ on its main diagonal:

$$\mathcal{M}_1\left(q_1, p_1\right) = \underbrace{\begin{pmatrix} e^{i\mu} & 0 \\ 0 & e^{-i\mu} \end{pmatrix}\begin{pmatrix} q_1 \\ p_1 \end{pmatrix}}_{\mathcal{R}} + \underbrace{\begin{pmatrix} \mathcal{S}^+_{2(2,0)} \\ \mathcal{S}^-_{2(2,0)} \end{pmatrix} q_1^2 + \begin{pmatrix} \mathcal{S}^+_{2(1,1)} \\ \mathcal{S}^-_{2(1,1)} \end{pmatrix} q_1 p_1 + \begin{pmatrix} \mathcal{S}^+_{2(0,2)} \\ \mathcal{S}^-_{2(0,2)} \end{pmatrix} p_1^2}_{\mathcal{S}_2}$$

$$+ \underbrace{\begin{pmatrix} \mathcal{S}^+_{3(3,0)} \\ \mathcal{S}^-_{3(3,0)} \end{pmatrix} q_1^3 + \begin{pmatrix} \mathcal{S}^+_{3(2,1)} \\ \mathcal{S}^-_{3(2,1)} \end{pmatrix} q_1^2 p_1 + \begin{pmatrix} \mathcal{S}^+_{3(1,2)} \\ \mathcal{S}^-_{3(1,2)} \end{pmatrix} q_1 p_1^2 + \begin{pmatrix} \mathcal{S}^+_{3(0,3)} \\ \mathcal{S}^-_{3(0,3)} \end{pmatrix} p_0^3}_{\mathcal{S}_3} + \ldots \qquad (3.18)$$

Tab. 3.3 lists the values to the coefficients above for the centrifugal governor example for a rotation frequency corresponding to an equilibrium angle of $\phi_0(\omega = \sqrt{2}) = \pi/3 = 60°$.

Table 3.3: Coefficients of $\mathcal{M}_1$ up to order three. Note the complex conjugate property $\mathcal{S}^{\pm}_{m(k_+,k_-)} = \bar{\mathcal{S}}^{\mp}_{m(k_-,k_+)}$.

Order	Coeff.	Real Part	Imaginary Part
1	$e^{i\mu}$	0.339185989	0.940719334
1	$e^{-i\mu}$	0.339185989	-0.940719334
2	$\mathcal{S}^+_{2(2,0)}$	-0.216977793	-0.059191831
2	$\mathcal{S}^-_{2(2,0)}$	0.072325931	-0.102961500
2	$\mathcal{S}^+_{2(1,1)}$	-0.258557455	0.368076331
2	$\mathcal{S}^-_{2(1,1)}$	-0.258557455	-0.368076331
2	$\mathcal{S}^+_{2(0,2)}$	0.072325931	0.102961500
2	$\mathcal{S}^-_{2(0,2)}$	-0.216977793	0.059191831
3	$\mathcal{S}^+_{3(3,0)}$	0.068036138	0.047162997
3	$\mathcal{S}^-_{3(3,0)}$	-0.045160062	-0.016282923
3	$\mathcal{S}^+_{3(2,1)}$	0.259415349	-0.130475661
3	$\mathcal{S}^-_{3(2,1)}$	-0.022283986	0.239186527
3	$\mathcal{S}^+_{3(1,2)}$	-0.022283986	-0.239186527
3	$\mathcal{S}^-_{3(1,2)}$	0.259415349	0.130475661
3	$\mathcal{S}^+_{3(0,3)}$	-0.045160062	-0.016282923
3	$\mathcal{S}^-_{3(0,3)}$	0.068036138	-0.047162997

3.3.3 The Nonlinear Transformations

The nonlinear transformations are the key steps of the normal form algorithm. In this first part of this subsection, we are going to look at an mth order transformation in general, before going through the nonlinear transformation for orders two and three in detail.

3.3.3.1 General mth Order Nonlinear Transformation

All the following nonlinear transformation steps are done order by order and are all of the same form: $\mathcal{M}_m = \mathcal{A}_m \circ \mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1}$, where the mth transformation does **not** change any of the lower order terms of $\mathcal{M}_{m-1}$ that have already been transformed in the previous transformations. Hence, $\mathcal{M}_m$ differs from $\mathcal{M}_{m-1}$ only in the orders m and larger. The mth order transformation $\mathcal{A}_m = \mathcal{I} + \mathcal{T}_m + \mathcal{O}_{\geq m+1}$, specifically the polynomial $\mathcal{T}_m$ of only mth order terms, is chosen such that the mth order terms $\mathcal{S}_m$ of the map $\mathcal{M}_{m-1}$ are simplified or even eliminated.

Effects on the higher orders of $\mathcal{M}_m$ due to the mth order transformation can only be considered by adjusting the terms of order higher than m of $\mathcal{A}_m$, namely $\mathcal{O}_{\geq m+1}$. In other words, finding $\mathcal{T}_m$ is essential to the DA normal form algorithm, while the terms $\mathcal{O}_{\geq m+1}$ can be chosen freely, e.g., to make the transformation symplectic by choosing $\mathcal{A}_m = \exp(L_{\mathcal{T}_m})$ or to avoid higher order resonances. Usually, the symplectic transformation is chosen since the calculation of the transformation $\mathcal{A}_m$ and its inverse are straightforward.

The flow operator $L_{\mathcal{T}_m} = (\mathcal{T}_m^+ \partial_q + \mathcal{T}_m^- \partial_p)$ in the exponential behaves in the following way:

$$\begin{aligned}
\exp\left(L_{\mathcal{T}_m}\right)\mathcal{I} &= \left(L^0_{\mathcal{T}_m} + L^1_{\mathcal{T}_m} + \frac{1}{2}L^2_{\mathcal{T}_m} + \mathcal{O}_{>(m+1)}\right)\mathcal{I} \\
&= \left(1 + (\mathcal{T}_m^+ \partial_q + \mathcal{T}_m^- \partial_p) + \frac{1}{2}L_{\mathcal{T}_m}(\mathcal{T}_m^+ \partial_q + \mathcal{T}_m^- \partial_p) + \mathcal{O}_{>(m+1)}\right)(q,p)^T \\
&= \mathcal{I} + \mathcal{T}_m + \frac{1}{2}L_{\mathcal{T}_m}\mathcal{T}_m + \mathcal{O}_{>(m+1)}. \qquad (3.19)
\end{aligned}$$

So, the inverse is given by

$$\mathcal{A}_m^{-1} = \exp\left(-L_{\mathcal{T}_m}\right) = \mathcal{I} - \mathcal{T}_m + \frac{1}{2}L_{\mathcal{T}_m}\mathcal{T}_m - \mathcal{O}_{>(m+1)}. \qquad (3.20)$$

In the example case of the centrifugal governor, we investigate the DA normal form algorithm up to order three, which means for $m = 3$:

$$\mathcal{A}_3 = \exp\left(L_{\mathcal{T}_3}\right)\mathcal{I} =_3 \mathcal{I} + \mathcal{T}_3 \tag{3.21}$$

$$\mathcal{A}_3^{-1} = \exp\left(-L_{\mathcal{T}_3}\right)\mathcal{I} =_3 \ \mathcal{I} - \mathcal{T}_3. \tag{3.22}$$

For the second order transformation it is necessary to consider the third order terms $\mathcal{O}_3$, since they influence the third order terms of $\mathcal{M}_2$:

$$\mathcal{A}_2 = \exp\left(L_{\mathcal{T}_2}\right)\mathcal{I} =_3 \mathcal{I} + \mathcal{T}_2 + \mathcal{O}_3 \tag{3.23}$$

$$\mathcal{A}_2^{-1} = \exp\left(-L_{\mathcal{T}_2}\right)\mathcal{I} =_3 \ \mathcal{I} - \mathcal{T}_2 + \mathcal{O}_3, \tag{3.24}$$

with

$$\mathcal{O}_3 = \frac{1}{2}L_{\mathcal{T}_m}\mathcal{T}_m = \frac{1}{2}(\mathcal{T}_2^+\partial_q + \mathcal{T}_2^-\partial_p)\mathcal{T}_2. \tag{3.25}$$

As introduced in Sec. 2.1, the notation '$=_m$' indicates that the quantities on both sides are equal up to expansion order m.

In order to determine $\mathcal{T}_m$, we analyze the mth order transformation and only look at terms up to order m [19, Eq. (7.62)]:

$$\begin{aligned}\mathcal{A}_m \circ \mathcal{M}_{m-1} \circ \mathcal{A}_m^{-1} &=_m (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} + \mathcal{S}_m) \circ (\mathcal{I} - \mathcal{T}_m)\\ &=_m (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} - \mathcal{R} \circ \mathcal{T}_m + \mathcal{S}_m)\\ &=_m \mathcal{R} + \mathcal{S}_m + [\mathcal{T}_m, \mathcal{R}]\,. \end{aligned}\tag{3.26}$$

Various terms with orders higher than m are ignored in the equations above. The goal is to choose $\mathcal{T}_m$ such that the commutator $[\mathcal{T}_m, \mathcal{R}] = \mathcal{T}_m \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_m = -\mathcal{S}_m$ to simplify $\mathcal{M}_m$, i.e. the result of Eq. (3.26). The polynomials in the upper and lower component of $\mathcal{T}_m$ can be express as

$$\mathcal{T}_m^\pm(q, p) = \sum_{\substack{m=k_++k_-\\ k_\pm\in\mathbb{N}_0}} \mathcal{T}_{m(k_+,k_-)}^\pm q^{k_+}p^{k_-}. \tag{3.27}$$

Hence, the commutator $\mathcal{C}_m = [\mathcal{T}_m, \mathcal{R}]$ yields

$$\mathcal{C}_m^\pm(q, p) = \sum_{\substack{m=k_++k_-\\ k_\pm\in\mathbb{N}_0}} \mathcal{T}_{m(k_+,k_-)}^\pm \left(e^{i\mu(k_+-k_-)} - e^{\pm i\mu}\right) q^{k_+}p^{k_-}. \tag{3.28}$$

A term in $\mathcal{S}_m$ can only be removed if and only if the corresponding term in the commutator $\mathcal{C}_m$ is not zero. Terms of the commutator are zero, whenever the condition

$$e^{i\mu(k_+-k_-)} - e^{\pm i\mu} = 0 \tag{3.29}$$

is satisfied, which is the case for $k_+ - k_- = \pm 1$. This (Eq. (3.29)) is the key condition of the DA normal form algorithm, since it determines the surviving nonlinear terms $\mathcal{S}_m$. All other terms that do not satisfy the condition are eliminated by choosing the coefficients of $\mathcal{T}_m$ as follows

$$\mathcal{T}^{\pm}_{m(k_+,k_-)} = \frac{-\mathcal{S}^{\pm}_{m(k_+,k_-)}}{e^{i\mu(k_+-k_-)} - e^{\pm i\mu}}. \tag{3.30}$$

Specifically, this means that the terms $\mathcal{S}^{+}_{m(k,k-1)}$ and $\mathcal{S}^{-}_{m(k-1,k)}$ always survive for all uneven orders m with $m = k + k - 1 = 2k - 1$.

3.3.3.2 Explicit Second Order Nonlinear Transformation

The polynomial $\mathcal{T}_m$ from Eq. (3.27) for $m = 2$ yields

$$\begin{aligned}\mathcal{T}_2(q,p) &= \left(\mathcal{T}^{\pm}_2|2,0\right) q^2 + \left(\mathcal{T}^{\pm}_2|1,1\right) qp + \left(\mathcal{T}^{\pm}_2|0,2\right) p^2 \\ &= \begin{pmatrix}\mathcal{T}^{+}_{2(2,0)} \\ \mathcal{T}^{-}_{2(2,0)}\end{pmatrix} q^2 + \begin{pmatrix}\mathcal{T}^{+}_{2(1,1)} \\ \mathcal{T}^{-}_{2(1,1)}\end{pmatrix} qp + \begin{pmatrix}\mathcal{T}^{+}_{2(0,2)} \\ \mathcal{T}^{-}_{2(0,2)}\end{pmatrix} p^2.\end{aligned} \tag{3.31}$$

The commutator $\mathcal{C}_2 = [\mathcal{T}_2, \mathcal{R}] = \mathcal{T}_2 \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_2$ of the second order nonlinear transformation has only nonzero terms with

$$\begin{aligned}\mathcal{C}_2(q,p) &= [\mathcal{T}_2, \mathcal{R}](q,p) = (\mathcal{T}_2 \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_2)(q,p) \\ &= \left(\mathcal{T}^{\pm}_2|2,0\right) e^{2i\mu} q^2 + \left(\mathcal{T}^{\pm}_2|1,1\right) qp + \left(\mathcal{T}^{\pm}_2|0,2\right) e^{-2i\mu} p^2 - e^{\pm i\mu} \mathcal{T}^{\pm}_2(q,p) \\ &= \begin{pmatrix}\mathcal{T}^{+}_{2(2,0)}\left(e^{2i\mu} - e^{i\mu}\right) \\ \mathcal{T}^{-}_{2(2,0)}\left(e^{2i\mu} - e^{-i\mu}\right)\end{pmatrix} q^2 + \begin{pmatrix}\mathcal{T}^{+}_{2(1,1)}\left(1 - e^{i\mu}\right) \\ \mathcal{T}^{-}_{2(1,1)}\left(1 - e^{-i\mu}\right)\end{pmatrix} qp + \begin{pmatrix}\mathcal{T}^{+}_{2(0,2)}\left(e^{-2i\mu} - e^{i\mu}\right) \\ \mathcal{T}^{-}_{2(0,2)}\left(e^{-2i\mu} - e^{-i\mu}\right)\end{pmatrix} p^2\end{aligned} \tag{3.32}$$

eliminating all $\mathcal{S}_2$ terms by choosing

$$\mathcal{T}^{\pm}_{2(k_+,k_-)} = \frac{-S^{\pm}_{2(k_+,k_-)}}{\left(e^{i\mu(k_+-k_-)} - e^{\pm i\mu}\right)}, \tag{3.33}$$

since the condition from Eq. (3.29) is not satisfied:

$$e^{i\mu(k_+-k_-)} - e^{\pm i\mu} \neq 0 \quad \forall k_+, k_- \in \mathbb{N}_0 \quad \text{with} \quad k_+ + k_- = 2.$$

The values of the $\mathcal{T}^{\pm}_{2(k_+,k_-)}$ for the centrifugal governor example are given in Tab. 3.4. The terms of $\mathcal{O}_3$ are calculated via Eq. (3.25) from $\mathcal{T}_2$ and are also given in Tab. 3.4 yielding all terms of the transformation $\mathcal{A}_2$ and its inverse $\mathcal{A}_2^{-1}$ from Eq. (3.23) and Eq. (3.24).

Table 3.4: The values of the $\mathcal{T}^{\pm}_{2(k_+,k_-)}$ and $\mathcal{O}^{\pm}_{3(k_+,k_-)}$. Note that $\mathcal{T}_2$ and $\mathcal{O}_3$ and therefore $\mathcal{A}_2$ and its inverse are real with $\mathcal{A}^{+}_{m(k_+,k_-)} = \mathcal{A}^{-}_{m(k_-,k_+)}$.

Order	Coeff.	Value	Coeff.	Value
2	$\mathcal{T}^{+}_{2(2,0)}$	-0.195635573	$\mathcal{T}^{-}_{2(2,0)}$	0.065211858
2	$\mathcal{T}^{+}_{2(1,1)}$	0.391271145	$\mathcal{T}^{-}_{2(1,1)}$	0.391271145
2	$\mathcal{T}^{+}_{2(0,2)}$	0.065211858	$\mathcal{T}^{-}_{2(0,2)}$	-0.195635573
3	$\mathcal{O}^{+}_{3(3,0)}$	0.051031036	$\mathcal{O}^{-}_{3(3,0)}$	0
3	$\mathcal{O}^{+}_{3(2,1)}$	-0.034020691	$\mathcal{O}^{-}_{3(2,1)}$	0.051031036
3	$\mathcal{O}^{+}_{3(1,2)}$	0.051031036	$\mathcal{O}^{-}_{3(1,2)}$	-0.034020691
3	$\mathcal{O}^{+}_{3(0,3)}$	0	$\mathcal{O}^{-}_{3(0,3)}$	0.051031036

To study how the second order transformation affects the third order terms $\mathcal{S}_3$ of the map $\mathcal{M}_2$, the transformation is considered up to third order:

$$\begin{aligned}
\mathcal{M}_2 &=_3 \mathcal{A}_2 \circ \mathcal{M}_1 \circ \mathcal{A}_2^{-1} \\
&=_3 (\mathcal{I} + \mathcal{T}_2 + \mathcal{O}_3) \circ (\mathcal{R} + \mathcal{S}_2 + \mathcal{S}_3) \circ (\mathcal{I} - \mathcal{T}_2 + \mathcal{O}_3) \\
&=_3 (\mathcal{I} + \mathcal{T}_2 + \mathcal{O}_3) \circ \Big(\mathcal{R} \circ (\mathcal{I} - \mathcal{T}_2 + \mathcal{O}_3) + \underline{\mathcal{S}_2 \circ (\mathcal{I} - \mathcal{T}_2 + \mathcal{O}_3)} + \underline{\mathcal{S}_3 \circ (\mathcal{I} - \mathcal{T}_2 + \mathcal{O}_3)}\Big) \\
&=_3 (\mathcal{I} + \mathcal{T}_2 + \mathcal{O}_3) \circ \left(\mathcal{R} - \mathcal{R} \circ \mathcal{T}_2 + \mathcal{R} \circ \mathcal{O}_3 + \overbrace{\mathcal{S}_2 + \mathcal{S}_{2\to3} + \cancel{\mathcal{O}_{\geq4}}} + \overbrace{\mathcal{S}_3 + \cancel{\mathcal{O}_{\geq4}}}\right) \\
&=_3 \mathcal{R} - \mathcal{R} \circ \mathcal{T}_2 + \mathcal{R} \circ \mathcal{O}_3 + \mathcal{S}_2 + \mathcal{S}_{2\to3} + \mathcal{S}_3 \\
&\qquad + \underline{\mathcal{T}_2 \circ (\mathcal{R} - \mathcal{R} \circ \mathcal{T}_2 + \mathcal{R} \circ \mathcal{O}_3 + \mathcal{S}_2 + \mathcal{S}_{2\to3} + \mathcal{S}_3)} + \cancel{\mathcal{O}_{\geq4}} \\
&=_3 \overbrace{\mathcal{T}_2 \circ \mathcal{R} + \mathcal{K}_{2\to3} + \cancel{\mathcal{O}_{\geq4}}} + \mathcal{R} - \mathcal{R} \circ \mathcal{T}_2 + \mathcal{R} \circ \mathcal{O}_3 + \mathcal{S}_2 + \mathcal{S}_{2\to3} + \mathcal{S}_3 \\
&=_3 \mathcal{R} + \underbrace{\mathcal{S}_2 + [\mathcal{T}_2 \circ \mathcal{R}]}_{=0} + \underbrace{\mathcal{S}_3 + \mathcal{S}_{2\to3} + \mathcal{K}_{2\to3} + \mathcal{R} \circ \mathcal{O}_3}_{\mathcal{S}_{3,\text{new}}}\,.
\end{aligned} \tag{3.34}$$

All the crossed-out terms $\cancel{\mathcal{O}_{\geq 4}}$ represent terms that do not contribute to the result up to order three, since they are at least of order four. As a result of the second order transformation, the third order terms have changed and are summarized by $\mathcal{S}_{3,\text{new}}$. They are composed of the third order terms from after the linear transformation $\mathcal{S}_3$ and three new terms: $\mathcal{S}_{2\to 3} =_3 \mathcal{S}_2 \circ (\mathcal{I} - \mathcal{T}_2) - \mathcal{S}_2$, $\mathcal{K}_{2\to 3} =_3 \mathcal{T}_2 \circ (\mathcal{R} - \mathcal{R} \circ \mathcal{T}_2 + \mathcal{S}_2) - \mathcal{T}_2 \circ \mathcal{R}$ and $\mathcal{R} \circ \mathcal{O}_3$. While the last one is self-explanatory, the first two are not intuitively understood. In Sec. 3.3.3.4 these terms are calculated more explicitly, however, we recommend this section only for the very intrigued reader and encourage everyone else to skip it to follow the steps in the normal form algorithm.

The result of the second order transformation $\mathcal{M}_2 = \mathcal{R} + \mathcal{S}_{3,\text{new}}$ for the example case of the centrifugal governor is given in Tab. 3.5.

Table 3.5: New coefficients of third order of $\mathcal{M}_2$ after the second order transformation. Note that the first order terms remain unchanged and that the second order terms are all eliminated by the second order transformation. Interestingly, the second order transformation caused some terms of the third order to disappear in this specific case, which is not a general property of the second order transformation. The emphasized terms are surviving the third order transformation as explained in the following subsection.

Order	Coeff.	Real Part	Imaginary Part
3	$\mathcal{S}^+_{3,\text{new}(3,0)}$	0.061270641	0.073920008
3	$\mathcal{S}^-_{3,\text{new}(3,0)}$	0	0
3	$\mathcal{S}^+_{3,\text{new}(2,1)}$	**0.470359667**	**-0.169592994**
3	$\mathcal{S}^-_{3,\text{new}(2,1)}$	0	0.288035295
3	$\mathcal{S}^+_{3,\text{new}(1,2)}$	0	-0.288035295
3	$\mathcal{S}^-_{3,\text{new}(1,2)}$	**0.470359667**	**0.169592994**
3	$\mathcal{S}^+_{3,\text{new}(0,3)}$	0	0
3	$\mathcal{S}^-_{3,\text{new}(0,3)}$	0.061270641	-0.073920008

3.3.3.3 Explicit Third Order Nonlinear Transformation

The third order transformation follows the same scheme as above (see Eq. (3.26)) only that the commutator $\mathcal{C}_3 = [\mathcal{T}_3 \circ \mathcal{R}]$ has terms that are zero

$$\mathcal{C}_3 = \begin{pmatrix} \mathcal{T}^{+}_{3(3,0)}\left(e^{3i\mu} - e^{i\mu}\right) \\ \mathcal{T}^{-}_{3(3,0)}\left(e^{3i\mu} - e^{-i\mu}\right) \end{pmatrix} q^3 + \begin{pmatrix} 0 \\ \mathcal{T}^{-}_{3(2,1)}\left(e^{i\mu} - e^{-i\mu}\right) \end{pmatrix} q^2 p$$
$$+ \begin{pmatrix} \mathcal{T}^{+}_{3(1,2)}\left(e^{-i\mu} - e^{i\mu}\right) \\ 0 \end{pmatrix} q p^2 + \begin{pmatrix} \mathcal{T}^{+}_{3(0,3)}\left(e^{-3i\mu} - e^{i\mu}\right) \\ \mathcal{T}^{-}_{3(0,3)}\left(e^{-3i\mu} - e^{-i\mu}\right) \end{pmatrix} p^3, \tag{3.35}$$

with $\mathcal{C}^{+}_{3(2,1)} = \mathcal{C}^{-}_{3(1,2)} = 0$. This means that the terms $\mathcal{S}^{+}_{3,\text{new}(2,1)}$ and $\mathcal{S}^{-}_{3,\text{new}(1,2)}$ cannot be eliminated. All the other terms are eliminated by choosing

$$\mathcal{T}^{\pm}_{3(k_+,k_-)} = \frac{-\mathcal{S}^{\pm}_{3,\text{new}(k_+,k_-)}}{\left(e^{i\mu(k_+-k_-)} - e^{\pm i\mu}\right)} \quad \text{for} \quad k_+ - k_- \neq \pm 1. \tag{3.36}$$

The values of $\mathcal{T}^{\pm}_{3(k_+,k_-)}$ for the centrifugal governor example are given in Tab. 3.6.

After the third order transformation the resulting map is of the following form

$$\mathcal{M}_3 = \underbrace{\begin{pmatrix} e^{i\mu} & 0 \\ 0 & e^{-i\mu} \end{pmatrix}\begin{pmatrix} q_3 \\ p_3 \end{pmatrix}}_{\mathcal{R}} + \underbrace{\begin{pmatrix} \mathcal{S}^{+}_{3,\text{new}(2,1)} \\ 0 \end{pmatrix} q_3^2 p_3 + \begin{pmatrix} 0 \\ \mathcal{S}^{-}_{3,\text{new}(1,2)} \end{pmatrix} q_3 p_3^2}_{\mathcal{S}_{3,\text{transformed}}}$$
$$= \begin{pmatrix} \left(e^{i\mu} + \mathcal{S}^{+}_{3,\text{new}(2,1)} q_3 p_3\right) q_3 \\ \left(e^{-i\mu} + \mathcal{S}^{-}_{3,\text{new}(1,2)} q_3 p_3\right) p_3 \end{pmatrix} = \begin{pmatrix} f^{+}\left(q_3 p_3\right) q_3 \\ f^{-}\left(q_3 p_3\right) p_3 \end{pmatrix}. \tag{3.37}$$

Table 3.6: The values of the $\mathcal{T}^{\pm}_{3(k_+,k_-)}$. The values for $\mathcal{T}^{+}_{3(2,1)}$ and $\mathcal{T}^{-}_{3(1,2)}$ cannot be calculated because the denominator in Eq. (3.36) is zero. Note that $\mathcal{T}^{+}_{3(k_+,k_-)} = \mathcal{T}^{-}_{3(k_-,k_+)}$.

Order	Coeff.	Value	Coeff.	Value
3	$\mathcal{T}^{+}_{3(3,0)}$	0.051031036	$\mathcal{T}^{-}_{3(3,0)}$	0
3	$\mathcal{T}^{+}_{3(2,1)}$	–	$\mathcal{T}^{-}_{3(2,1)}$	-0.153093109
3	$\mathcal{T}^{+}_{3(1,2)}$	-0.153093109	$\mathcal{T}^{-}_{3(1,2)}$	–
3	$\mathcal{T}^{+}_{3(0,3)}$	0	$\mathcal{T}^{-}_{3(0,3)}$	0.051031036

The corresponding values for the coefficients can be found in Tab. 3.3 for the linear terms and in Tab. 3.5 for the third order terms. The complex conjugate property of the map $\mathcal{M}_3^+ = \overline{\mathcal{M}_3^-}$ is maintained.

While all nonlinear transformations follow the same structure, there is a fundamental difference between even and odd order transformation steps. For even order transformations there are no regularly surviving terms as shown for the second order transformation. For uneven order transformations, there are some terms of a special structure that do survive as shown for third order transformation. Higher even and odd order transformations will behave in the same way, and we will stop the process of the detailed walk-through here, after the third order transformation. In principle, the calculation of the transformations can be continued up to arbitrary order. With each transformation, the higher order terms are changed and in the end only the terms $\mathcal{S}^+_{m(k,k-1)}$ and $\mathcal{S}^-_{m(k-1,k)}$ of uneven orders survive. Hence, the components $\mathcal{M}_m^\pm$ can also be factorized into the $f^\pm\left(q_m p_m\right)$ notation (see Eq. (3.37)) for higher orders.

3.3.3.4 The Effect of the Second Order Transformation on Third Order Terms

The following calculation investigates the term $\mathcal{S}_{2\to3}$ as was previously done in [93] and was added here for sake of completeness.

$$
\begin{aligned}
\mathcal{S}_{2\to3} &=_3 \mathcal{S}_2 \circ \left(\mathcal{I} - \mathcal{T}_2\right) - \mathcal{S}_2 \\
&=_3 \mathcal{S}_{2(2,0)}\left(q - \mathcal{T}_2^+\right)^2 + \mathcal{S}_{2(0,2)}\left(p - \mathcal{T}_2^-\right)^2 + \mathcal{S}_{2(1,1)}\left(q - \mathcal{T}_2^+\right)\left(p - \mathcal{T}_2^-\right) - \mathcal{S}_2 \\
&=_3 \underbrace{\cancel{\mathcal{S}_{2(2,0)} q^2 + \mathcal{S}_{2(1,1)} qp + \mathcal{S}_{2(0,2)} p^2 - \mathcal{S}_2}}_{=0} \\
&\quad + \underbrace{\cancel{\mathcal{S}_{2(2,0)}\left(\mathcal{T}_2^+\right)^2 + \mathcal{S}_{2(1,1)}\mathcal{T}_2^+\mathcal{T}_2^- + \mathcal{S}_{2(0,2)}\left(\mathcal{T}_2^-\right)^2}}_{\geq\mathcal{O}_4} \\
&\quad - 2\mathcal{S}_{2(2,0)}\mathcal{T}_2^+ q - \mathcal{S}_{2(1,1)}\left(\mathcal{T}_2^+ p + \mathcal{T}_2^- q\right) - 2\mathcal{S}_{2(0,2)}\mathcal{T}_2^- p. \qquad (3.38)
\end{aligned}
$$

As derived in the beginning of Sec. 3.3.3.3, the surviving parts of $\mathcal{S}_{2\to3}$ after the third order transformation are $\mathcal{S}^+_{2\to3(2,1)}$ and its complex conjugate counterpart $\mathcal{S}^-_{2\to3(1,2)}$:

$$\begin{aligned}
\mathcal{S}^+_{2\to3(2,1)} &= -2\mathcal{S}^+_{2(2,0)}\mathcal{T}^+_{2(1,1)} - \mathcal{S}^+_{2(1,1)}\left(\mathcal{T}^+_{2(2,0)} + \mathcal{T}^-_{2(1,1)}\right) - 2\mathcal{S}^+_{2(0,2)}\mathcal{T}^-_{2(2,0)} \\
&= \frac{2\mathcal{S}^+_{2(2,0)}\mathcal{S}^+_{2(1,1)}}{1-e^{i\mu}} + \frac{\mathcal{S}^+_{2(1,1)}\mathcal{S}^+_{2(2,0)}}{e^{2i\mu}-e^{i\mu}} + \frac{\mathcal{S}^+_{2(1,1)}\mathcal{S}^-_{2(1,1)}}{1-e^{-i\mu}} + \frac{\mathcal{S}^+_{2(0,2)}\mathcal{S}^-_{2(2,0)}}{e^{2i\mu}-e^{-i\mu}}.
\end{aligned} \tag{3.39}$$

This illustrates the complexity of these terms since every single term from $\mathcal{S}_2$ is relevant for them. Each term of $\mathcal{S}_2$ is again dependent on the terms of $\mathcal{U}_2$. The relation is given by the linear transformation $\mathcal{S}_2 = \mathcal{A}_1 \circ \mathcal{U}_2 \circ \mathcal{A}_1^{-1}$. In principle, one can extend the calculation above to express $\mathcal{S}^+_{2\to3(2,1)}$ in terms of $\mathcal{U}_2$ and the Twiss parameters as done in [93]. The main insight however is that due to the significant influence of lower order transformation on higher order terms it is almost impossible to determine a priory which terms are the relevant ones for characteristics of the normal form.

In the following calculation we are investigating the term $\mathcal{K}_{2\to3}$, which was not previously investigated in [93].

$$\begin{aligned}
\mathcal{K}_{2\to3} &= \mathcal{T}_2 \circ (\mathcal{R} - \mathcal{R}\circ\mathcal{T}_2 + \mathcal{S}_2) - \mathcal{T}_2\circ\mathcal{R} \\
&= \mathcal{T}_2 \circ (\mathcal{R} - \mathcal{K}_2) - \mathcal{T}_2\circ\mathcal{R} \\
&=_3 \mathcal{T}_{2(2,0)}\left(e^{i\mu}q - \mathcal{K}_2^+\right)^2 + \mathcal{T}_{2(0,2)}\left(e^{-i\mu}p - \mathcal{K}_2^-\right)^2 \\
&\quad + \mathcal{T}_{2(1,1)}\left(e^{i\mu}q - \mathcal{K}_2^+\right)\left(e^{-i\mu}p - \mathcal{K}_2^-\right) - \mathcal{T}_2\circ\mathcal{R} \\
&=_3 \underbrace{\cancel{\mathcal{T}_{2(2,0)}e^{2i\mu}q^2 + \mathcal{T}_{2(1,1)}qp + \mathcal{T}_{2(0,2)}e^{-2i\mu}p^2 - \mathcal{T}_2\circ\mathcal{R}}}_{=0} \\
&\quad \underbrace{\cancel{+\,\mathcal{T}_{2(2,0)}\left(\mathcal{K}_2^+\right)^2 + \mathcal{T}_{2(1,1)}\mathcal{K}_2^+\mathcal{K}_2^- + \mathcal{T}_{2(0,2)}\left(\mathcal{K}_2^-\right)^2}}_{\geq\mathcal{O}_4} \\
&\quad - 2\mathcal{T}_{2(2,0)}\mathcal{K}_2^+e^{i\mu}q - \mathcal{T}_{2(1,1)}\left(\mathcal{K}_2^+e^{-i\mu}p + \mathcal{K}_2^-e^{i\mu}q\right) - 2\mathcal{T}_{2(0,2)}\mathcal{K}_2^-e^{-i\mu}p,
\end{aligned} \tag{3.40}$$

where

$$\mathcal{K}_2 = \mathcal{R}\circ\mathcal{T}_2 - \mathcal{S}_2 \qquad \to \qquad \mathcal{K}_2^\pm = e^{\pm i\mu}\mathcal{T}_2^\pm - \mathcal{S}_2^\pm. \tag{3.41}$$

So,

$$\begin{aligned}\mathcal{K}_{2\to3} =_3 \; & 2\mathcal{T}_{2(2,0)}\mathcal{S}_2^+ e^{i\mu} q + \mathcal{T}_{2(1,1)}\left(\mathcal{S}_2^+ e^{-i\mu} p + \mathcal{S}_2^- e^{i\mu} q\right) + 2\mathcal{T}_{2(0,2)}\mathcal{S}_2^- e^{-i\mu} p \\ & - 2\mathcal{T}_{2(2,0)}\mathcal{T}_2^+ e^{2i\mu} q - \mathcal{T}_{2(1,1)}\left(\mathcal{T}_2^+ p + \mathcal{T}_2^- q\right) - 2\mathcal{T}_{2(0,2)}\mathcal{T}_2^- e^{-2i\mu} p. \end{aligned} \tag{3.42}$$

The surviving terms of $\mathcal{K}_{2\to3}$ after the third order transformation are $\mathcal{K}^+_{2\to3(2,1)}$ and its complex conjugate counterpart $\mathcal{K}^-_{2\to3(1,2)}$ are

$$\begin{aligned}\mathcal{K}^+_{2\to3(2,1)} &= 2\mathcal{T}^+_{2(2,0)}\mathcal{S}^+_{2(1,1)}e^{i\mu} + \mathcal{T}^+_{2(1,1)}\left(\mathcal{S}^+_{2(2,0)}e^{-i\mu} + \mathcal{S}^-_{2(1,1)}e^{i\mu}\right) \\ &\quad + 2\mathcal{T}^+_{2(0,2)}\mathcal{S}^-_{2(2,0)}e^{-i\mu} - 2\mathcal{T}^+_{2(2,0)}\mathcal{T}^+_{2(1,1)}e^{2i\mu} \\ &\quad - \mathcal{T}^+_{2(1,1)}\left(\mathcal{T}^+_{2(2,0)} + \mathcal{T}^-_{2(1,1)}\right) - 2\mathcal{T}^+_{2(0,2)}\mathcal{T}^-_{2(2,0)}e^{-2i\mu} \\ &= \frac{-2\mathcal{S}^+_{2(2,0)}\mathcal{S}^+_{2(1,1)}}{e^{i\mu}-1} - \frac{\mathcal{S}^+_{2(1,1)}}{1-e^{i\mu}}\left(\mathcal{S}^+_{2(2,0)}e^{-i\mu} + \mathcal{S}^-_{2(1,1)}e^{i\mu}\right) + \frac{2\mathcal{S}^+_{2(0,2)}\mathcal{S}^-_{2(2,0)}}{e^{2i\mu}-e^{-i\mu}} \\ &\quad + \frac{\mathcal{S}^+_{2(1,1)}\left(2\mathcal{S}^+_{2(2,0)} + \mathcal{S}^-_{2(1,1)}\right)}{2\left(\cos\mu - 1\right)} - \frac{\mathcal{S}^+_{2(1,1)}\mathcal{S}^+_{2(2,0)}}{2e^{2i\mu}-e^{i\mu}-e^{3i\mu}} - \frac{2\mathcal{S}^+_{2(0,2)}\mathcal{S}^-_{2(2,0)}}{2e^{2i\mu}-e^{-i\mu}-e^{5i\mu}}. \end{aligned} \tag{3.43}$$

Also for $\mathcal{K}^+_{2\to3(2,1)}$ and $\mathcal{K}^-_{2\to3(1,2)}$ the intertwine dependency on all terms of $\mathcal{S}_2$ becomes apparent highlighting the complex relation between lower order and higher order terms.

3.3.4 Transformation back to Real Space Normal Form

Since the original map $\mathcal{M}_0$ only operates in real space, the normal form map $\mathcal{M}_{\text{NF}}$ should also only operate in real space. This is why the current map $\mathcal{M}_m$, where m is the order of last transformation, is transformed to a real normal form basis $(q_{\text{NF}}, p_{\text{NF}})$ composed of the real and imaginary parts of the current complex conjugate basis (q_m, p_m). Based on [19, Eq. (7.58) and (7.59) and (7.67)] the bases are related as follows

$$q_{\text{NF}} = \frac{q_m + p_m}{2} \quad \text{and} \quad p_{\text{NF}} = \frac{q_m - p_m}{2i}, \tag{3.44}$$

and

$$q_m = q_{\text{NF}} + i\,p_{\text{NF}} \quad \text{and} \quad p_m = q_{\text{NF}} - i\,p_{\text{NF}} \quad \text{with} \tag{3.45}$$

$$q_m p_m = q_{\text{NF}}^2 + p_{\text{NF}}^2 = r_{\text{NF}}^2. \tag{3.46}$$

The associated transfer matrix

$$\mathcal{A}_{\text{real}} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} = \begin{pmatrix} (q_{\text{NF}}|q_m) & (q_{\text{NF}}|p_m) \\ (p_{\text{NF}}|q_m) & (p_{\text{NF}}|p_m) \end{pmatrix} \tag{3.47}$$

to the real normal form basis is obtained from the equations above. The inverse relation is given by

$$\mathcal{A}_{\text{real}}^{-1} = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} = \begin{pmatrix} (q_m|q_{\text{NF}}) & (q_m|p_{\text{NF}}) \\ (p_m|q_{\text{NF}}) & (p_m|p_{\text{NF}}) \end{pmatrix}. \tag{3.48}$$

The transformation back to the real space (into normal form space) yields

$$\begin{aligned} \mathcal{M}_{\text{NF}} = \mathcal{A}_{\text{real}} \circ \mathcal{M}_m \circ \mathcal{A}_{\text{real}}^{-1} &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \cdot \begin{pmatrix} f^+\left(r_{\text{NF}}^2\right)(q_{\text{NF}} + i\,p_{\text{NF}}) \\ f^-\left(r_{\text{NF}}^2\right)(q_{\text{NF}} - i\,p_{\text{NF}}) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}\left(f^+ + \bar{f}^+\right) q_{\text{NF}} + \frac{i}{2}\left(f^+ - \bar{f}^+\right) p_{\text{NF}} \\ \frac{-i}{2}\left(f^+ - \bar{f}^+\right) q_{\text{NF}} + \frac{1}{2}\left(f^+ + \bar{f}^+\right) p_{\text{NF}} \end{pmatrix} \\ &= \begin{pmatrix} \operatorname{Re}\left(f^+\left(r_{\text{NF}}^2\right)\right) & -\operatorname{Im}\left(f^+\left(r_{\text{NF}}^2\right)\right) \\ \operatorname{Im}\left(f^+\left(r_{\text{NF}}^2\right)\right) & \operatorname{Re}\left(f^+\left(r_{\text{NF}}^2\right)\right) \end{pmatrix} \cdot \begin{pmatrix} q_{\text{NF}} \\ p_{\text{NF}} \end{pmatrix}. \end{aligned} \tag{3.49}$$

For the example of the centrifugal governor up to order three the normal form is

$$\mathcal{M}_{\text{NF}} = \begin{pmatrix} \cos\mu + \frac{1}{2}\operatorname{Re}\left(\mathcal{S}_{3,\text{new}(2,1)}^+\right) r_{\text{NF}}^2 & -\sin\mu - \frac{1}{2}\operatorname{Im}\left(\mathcal{S}_{3,\text{new}(2,1)}^+\right) r_{\text{NF}}^2 \\ \sin\mu + \frac{1}{2}\operatorname{Im}\left(\mathcal{S}_{3,\text{new}(2,1)}^+\right) r_{\text{NF}}^2 & \cos\mu + \frac{1}{2}\operatorname{Re}\left(\mathcal{S}_{3,\text{new}(2,1)}^+\right) r_{\text{NF}}^2 \end{pmatrix} \cdot \begin{pmatrix} q_{\text{NF}} \\ p_{\text{NF}} \end{pmatrix}. \tag{3.50}$$

The Tab. 3.7 below yields the values for the normal form map of our example case.

Table 3.7: The normal form map $\mathcal{M}_{\text{NF}}$ up to order three. The component $\mathcal{M}_{\text{NF}}^+$ is on the left, $\mathcal{M}_{\text{NF}}^-$ on the right.

Order	Coeff.	Value	Coeff.	Value
1	$\mathcal{M}_{\text{NF}(1,0)}^+$	0.339185989	$\mathcal{M}_{\text{NF}(1,0)}^-$	0.940719334
1	$\mathcal{M}_{\text{NF}(0,1)}^+$	-0.940719334	$\mathcal{M}_{\text{NF}(0,1)}^-$	0.339185989
3	$\mathcal{M}_{\text{NF}(3,0)}^+$	0.470359667	$\mathcal{M}_{\text{NF}(3,0)}^-$	-0.169592994
3	$\mathcal{M}_{\text{NF}(2,1)}^+$	0.169592994	$\mathcal{M}_{\text{NF}(2,1)}^-$	0.470359667
3	$\mathcal{M}_{\text{NF}(1,2)}^+$	0.470359667	$\mathcal{M}_{\text{NF}(1,2)}^-$	-0.169592994
3	$\mathcal{M}_{\text{NF}(0,3)}^+$	0.169592994	$\mathcal{M}_{\text{NF}(0,3)}^-$	0.470359667

The normal form transformation from $\mathcal{M}_0$ to $\mathcal{M}_{\text{NF}}$ can be obtained by the combination of all the single transformations yielding

$$\begin{aligned}\mathcal{M}_{\text{NF}} = & \underbrace{\mathcal{A}_{\text{real}} \circ \mathcal{A}_m \circ \mathcal{A}_{m-1} \circ ... \circ \mathcal{A}_1 \circ \mathcal{A}_{\text{FP}}}_{\mathcal{A}} \circ \mathcal{M}_0 \\ & \circ \underbrace{\mathcal{A}_{\text{FP}}^{-1} \circ \mathcal{A}_1^{-1} \circ ... \circ \mathcal{A}_{m-1}^{-1} \circ \mathcal{A}_m^{-1} \circ \mathcal{A}_{\text{real}}^{-1}}_{\mathcal{A}^{-1}} . \end{aligned} \tag{3.51}$$

The values of the coefficients of the full normal form transformation $\mathcal{A}$ are given in Tab. 3.8.

Table 3.8: The normal form transformation $\mathcal{A}$ up to order three. The component $\mathcal{A}^+$ is on the left, $\mathcal{A}^-$ on the right.

Order	Coeff.	Value	Coeff.	Value
1	$\mathcal{A}^+_{1(1,0)}$	1.106681920	$\mathcal{A}^-_{1(1,0)}$	0
1	$\mathcal{A}^+_{1(0,1)}$	0	$\mathcal{A}^-_{1(0,1)}$	-0.903602004
2	$\mathcal{A}^+_{2(2,0)}$	0.319471552	$\mathcal{A}^-_{2(2,0)}$	0
2	$\mathcal{A}^+_{2(1,1)}$	0	$\mathcal{A}^-_{2(1,1)}$	0.521694860
2	$\mathcal{A}^+_{2(0,2)}$	0.425962069	$\mathcal{A}^-_{2(0,2)}$	0
3	$\mathcal{A}^+_{3(3,0)}$	-0.046111747	$\mathcal{A}^-_{3(3,0)}$	0
3	$\mathcal{A}^+_{3(2,1)}$	0	$\mathcal{A}^-_{3(2,1)}$	-0.414150918
3	$\mathcal{A}^+_{3(1,2)}$	-0.399635138	$\mathcal{A}^-_{3(1,2)}$	0
3	$\mathcal{A}^+_{3(0,3)}$	0	$\mathcal{A}^-_{3(0,3)}$	0.025100056

Writing the complex conjugate functions $f^{\pm}$ from the equations above (particularly Eq. (3.49)) in a complex notation as

$$f^{\pm}\left(r_{\text{NF}}^2\right) = e^{\pm i \Lambda\left(r_{\text{NF}}^2\right)} \tag{3.52}$$

illustrates circular behavior of the normal form:

$$\mathcal{M}_{\text{NF}} = \begin{pmatrix} \cos\left(\Lambda\left(r_{\text{NF}}^2\right)\right) & -\sin\left(\Lambda\left(r_{\text{NF}}^2\right)\right) \\ \sin\left(\Lambda\left(r_{\text{NF}}^2\right)\right) & \cos\left(\Lambda\left(r_{\text{NF}}^2\right)\right) \end{pmatrix} \cdot \begin{pmatrix} q_{\text{NF}} \\ p_{\text{NF}} \end{pmatrix} . \tag{3.53}$$

It shows that the normal form $\mathcal{M}_{\text{NF}}$ consists of circular curves in phase space with only amplitude dependent angle advancements $\Lambda\left(r_{\text{NF}}^2\right)$.

3.3.5 Invariant Normal Form Radius

The squared normal form radius r_{NF}^2 is related to the original coordinates (q_0, p_0) by the normal form transformation $\mathcal{A}$, where

$$\begin{aligned} r_{\mathrm{NF}}^2(q_0, p_0) &= \left(q_{\mathrm{NF}}^2(q_0, p_0) + p_{\mathrm{NF}}^2(q_0, p_0)\right) \\ &= \left(\mathcal{A}_+^2 + \mathcal{A}_-^2\right)(q_0, p_0) \,. \end{aligned} \tag{3.54}$$

Explicitly calculating the squared normal form radius with the normal form transformation $\mathcal{A}$ up to order three from Tab. 3.8 yields

$$r_{\mathrm{NF}}^2 =_3 1.224745 q_0^2 + 0.816497 p_0^2 + 0.707107 q_0^3 \tag{3.55}$$

$$\approx \sqrt{\frac{3}{2}} q_0^2 + \sqrt{\frac{2}{3}} p_0^2 + \frac{1}{\sqrt{2}} q_0^3 \tag{3.56}$$

$$=_3 \frac{2\sqrt{2}}{\sqrt{3}} \left(E\left(\frac{\pi}{3} + q_0, p_0\right) - E\left(\frac{\pi}{3}, 0\right)\right), \tag{3.57}$$

where

$$E(q, p) = \frac{p}{2} + \frac{-\omega^2 \sin^2 \phi}{2} + (1 - \cos\phi) \tag{3.58}$$

can be straightforwardly derived from Eq. (3.4) and Eq. (3.6). This direct relationship between the energy E, as an invariant or constant of motion, and the squared normal form radius up to order three confirms that the normal form radius constitutes a constant of motion up to calculation order.

The invariant of motion is a family of functions that remain constant for all phase space states (q, p) along their phase space motion. In particular, if $I(q, p)$ is an invariant of motion, then so is $I^2(q, p)$ or any other function $f(I)$, which is defined by the resulting values of I. Furthermore, $I(Q(q, p), P(q, p))$ is also an invariant if (Q, P) belong to the same phase space curve as (q, p). Transfer maps can yield such relations $(Q(q, p), P(q, p))$, since they can represent how a phase space final state (Q, P) depends on the phase space initial state (q, p).

Accordingly, the energy E and the normal form radius r_{NF}^2 are both functions of the same family and related by the transformations explained in the paragraph above. Up to order three, this relation includes a shift by a constant and scaling, but the relation might reveal itself to be more complex than this with higher orders.

3.3.6 Angle Advancement, Tune and Tune Shifts

In the beam physics terminology, the angle advancements $\Lambda\left(r_{\mathrm{NF}}^2\right)$ are scaled to the interval $[0,1]$ instead of $[0,2\pi]$ and referred to as the tune and amplitude dependent tune shifts [19]. The angle advancement can be calculated from the normal form map via

$$\Lambda\left(r_{\mathrm{NF}}^2 = q_{\mathrm{NF}}^2, p_{\mathrm{NF}} = 0\right) = \arccos\left(\frac{\left.\mathcal{M}_{\mathrm{NF}}^+\right|_{p_{\mathrm{NF}}=0}}{q_{\mathrm{NF}}}\right) = \arccos\left(\mathrm{Re}\left(f^+\left(r_{\mathrm{NF}}^2\right)\right)\right). \tag{3.59}$$

For the centrifugal governor example up to order three, the angle advancement is given by

$$\begin{aligned}\Lambda\left(r_{\mathrm{NF}}^2 = q_{\mathrm{NF}}^2\right) &= \arccos\left(\cos\mu + \frac{1}{2}\mathrm{Re}\left(\mathcal{S}_{3,\mathrm{new}(2,1)}^+\right) r_{\mathrm{NF}}^2\right)\\ &= \mu - \frac{\mathrm{Re}\left(\mathcal{S}_{3,\mathrm{new}(2,1)}^+\right)}{2\sin\mu} r_{\mathrm{NF}}^2.\end{aligned} \tag{3.60}$$

Note that μ is the eigenvalue phase of the original linear part (see Eq. (3.17)). The tune and tune shifts are calculated from Eq. (2.31), with

$$\frac{\Lambda\left(r_{\mathrm{NF}}^2\right)}{2\pi} = \nu\left(r_{\mathrm{NF}}^2\right) = 0.1949242 - 0.07957747 r_{\mathrm{NF}}^2, \tag{3.61}$$

where the constant part is the tune already known from the linear transformation with

$$\nu = \frac{\mu}{2\pi} = \frac{1.22474487}{2\pi} = 0.1949242. \tag{3.62}$$

With the expression of r_{NF}^2 in terms of the original coordinates (q_0, p_0) from Eq. (3.55) the tune and tune shifts are evaluated to

$$\nu\left(q_0, p_0\right) = 0.1949242 - 0.0974621 q_0^2 - 0.0649747 p_0^2 - 0.05626977 q_0^3. \tag{3.63}$$

This yields a key insight into the centrifugal governor behavior for $\omega = \sqrt{2}$. We already know that the centrifugal governor is rotating at $\sqrt{2}/(2\pi) \approx 0.225$ revolutions per T_0 for $\omega = \sqrt{2}$. The tune of about 0.195 tells us that the centrifugal governor arms oscillate at a frequency of about $0.195 + c$ oscillations per T_0 around their equilibrium position. The negative tune shifts additionally show that this frequency is decreasing for increasing amplitude of oscillation.

Since the map can only compare initial and final state of the oscillation after the integration time of 1 T_0 we only know how much the oscillation cycle has advanced over this period, but not how many additional full oscillations c have been completed in the meantime. By doing the same process as above for the centrifugal governor with $\omega = \sqrt{2}$ for a Poincaré map after time $t = 2\pi/\sqrt{2}$, i.e. one full centrifugal governor revolution, yields

$$\nu\left(q_0, p_0\right) = 0.8660254 - 0.4330127 q_0^2 - 0.2886751 p_0^2 - 0.25000000 q_0^3, \tag{3.64}$$

which is exactly a factor of $2\pi/\sqrt{2}$ larger than the tunes from Eq. (3.63). This means that c must be zero and we did not miss any full oscillations during the integration up to $t = 1$.

From Eq. (3.63) we can directly calculate the period of oscillation from normal form T_{NF}, which is just $1/\nu\left(q_0, p_0\right)$.

To compare the calculated normal form period T_{NF} to the actual period of oscillation, we flip the horizontal and vertical axis from Fig. 3.4 and overlay the oscillatory plot with the calculated periods (see Fig. 3.6). The centrifugal governor arms are initiated with multiple angle offsets with $p_\phi = 0$ relative to their equilibrium angle at $\phi_0 = 60°$. If the normal form calculation of the period

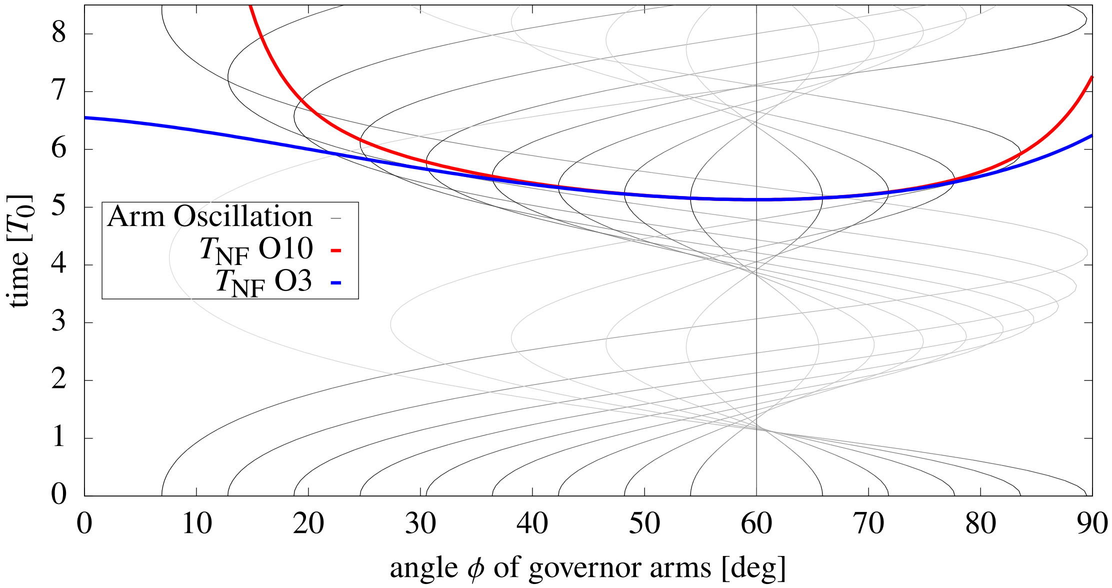

Figure 3.6: Comparison between the calculated period with normal form methods $T_{\mathrm{NF}} = 1/\nu(q_0, p_0)$ for calculation order ten (O10) and calculation order three (O3) to the actual period of oscillation given by the oscillatory behavior of the centrifugal governor arms for $\omega = \sqrt{2}$ from Fig. 3.4.

is correct, the calculated period will agree with the time when the equilibrium governor arms reach their initial position amplitude after one actual period of oscillation.

The higher the amplitude of oscillation, the more relevant are higher order effects. Accordingly, the accuracy drops with larger amplitudes. The order three calculation performs well between 35°($\delta\phi = -25°$) and 75°($\delta\phi = +15°$), while the order ten calculation can extend an accurate description over the range from 25°($\delta\phi = -35°$) to 85°($\delta\phi = +25°$).

The normal form algorithm can also be performed with parameters, e.g., depending on changes to ω. In Tab. 3.9 result for the amplitude and parameter $\delta\omega$ dependent tunes shifts are listed. It shows that the $\delta\omega$ dependent tune shifts are positive, which means that an increase in ω increases the oscillation frequency of centrifugal governor arms. This is related to the deeper potential well.

This knowledge about the dependency of the tunes on parameter shifts can help by the selection of a suitable ω, e.g., to avoid resonances between the governors revolution frequency and the oscillation frequency of the arms. While such a resonance is irrelevant in this simplified example it might be critical when the governor is part of a more complex system.

Table 3.9: Tune and coefficients of amplitude and parameter $\delta\omega$ dependent tune shifts for centrifugal governor with $\omega_0 = \sqrt{2}$.

	Exponents				Exponents		
Coefficient	q_0	p_0	$\delta\omega$	Coefficient	q_0	p_0	$\delta\omega$
0.1949242003	0	0	0	-0.0562697698	3	0	0
0.3355884937	0	0	1	-0.0307638305	2	0	1
-0.0974621002	2	0	0	0.1741334861	1	1	1
-0.0649747334	0	2	0	0.1123973696	0	2	1
0.1591549431	1	0	1	-0.0435458248	1	0	2
-0.5753522001	0	0	2	-0.0142179396	0	1	2
				0.0866936204	0	0	3

3.4 Visualization of the Different Order Normal Forms and Conclusion

In this chapter, we considered the system of a centrifugal governor with a fixed rotation frequency of $\omega = \sqrt{2}$ and analyzed it using the DA normal form algorithm.

To visualize the effect of the different steps in the DA normal form algorithm, Fig. 3.7 shows phase space tracking pictures for incomplete normal form maps. Given the tenth order Poincaré map which describes the behavior of the centrifugal governor for $\omega = \sqrt{2}$, these incomplete normal form maps stopped the normal form transformations at an order $n < 10$ such that the resulting incomplete normal form map is only normalized up to order n. There is no practical use for these incomplete normal form maps other than showing the progress of the normal form algorithm, since to make use of the normal form properties completion of the normal form transformation to the full order of the map is required. The phase space behavior in the full order normal form with its rotationally invariant property was previously shown in Fig. 3.5.

The difference between a) and b) in Fig. 3.7 shows the effect of the linear transformation, which scales the variables to create circles close to the expansion point. The nonlinear distortions for larger amplitudes are still present. With the second and third order transformation, these distortions are removed in the normal form, however still not forming perfect circles for larger amplitudes.

As a result of the DA normal form algorithm, we were able to produce invariants of motion up to calculation order. Specifically, we could show how the squared normal form radius is directly related to the energy E up to calculation order (see Eq. (3.57)), which is a constant of motion for this system.

The normal form algorithm also provided transformations from the original coordinates to the normal form coordinates, which were used to relate the phase space amplitudes to the normal form invariant.

Finally, the normal form produced the period of oscillation of the centrifugal governor arms around their equilibrium angle depending on the amplitude of oscillation. The preformed calculation of order ten did not capture all the relevant high order effects at vary large amplitudes. However, yet higher order calculation would describe the period of oscillation for these amplitudes more accurately.

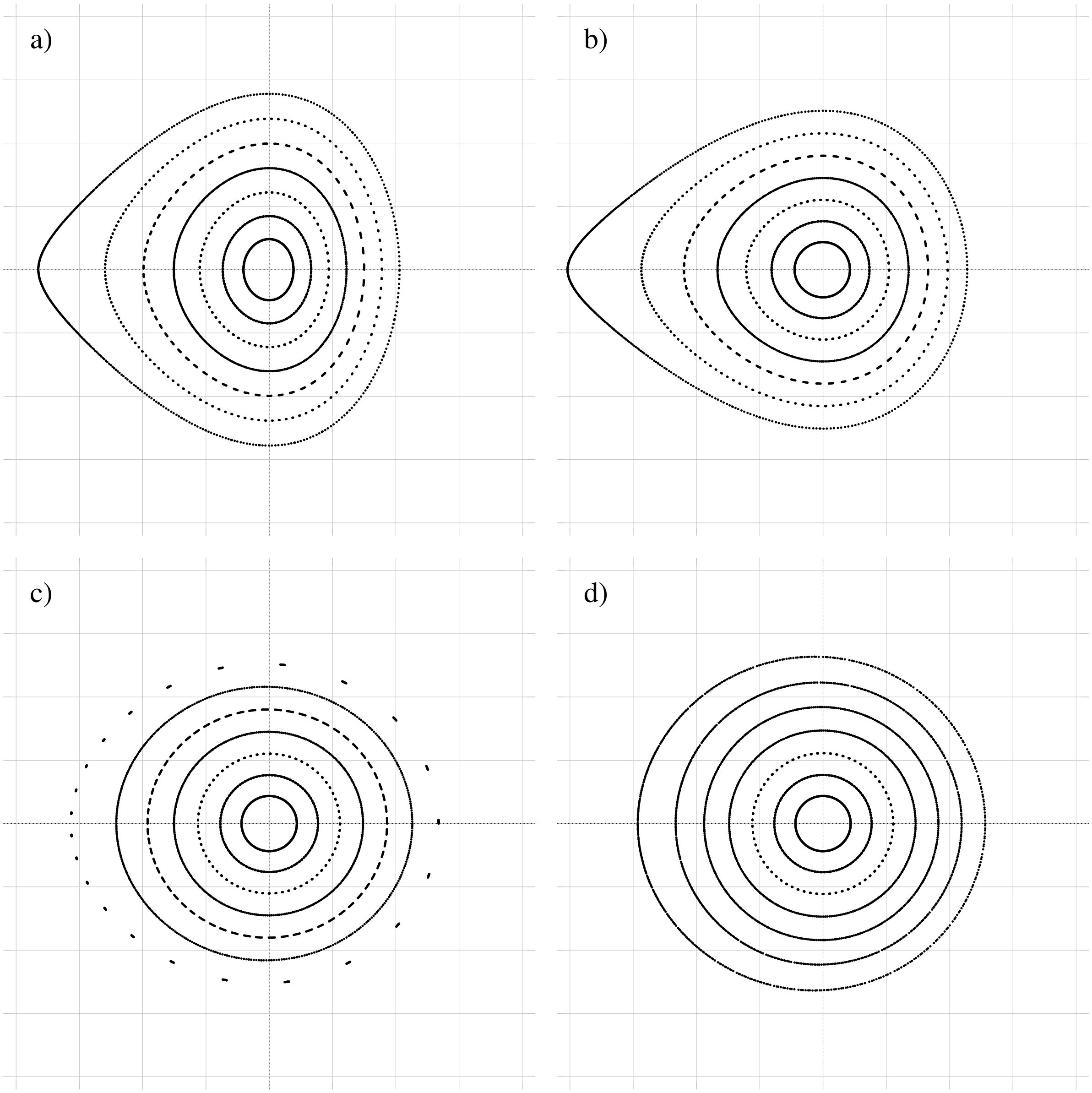

Figure 3.7: Phase space tracking of incomplete normal form maps of order ten of the centrifugal governor arms with a fixed rotation frequency of $\omega = \sqrt{2}$. The original map (a), only linear normal form transformation (b), and only normal form transformations up to order two (c) and three (d), respectively. The normal form up to the full tenth order was illustrated in Fig. 3.5.

CHAPTER 4

BOUNDED MOTION PROBLEM

This chapter contains large parts of my paper *Bounded motion design in the Earth zonal problem using differential algebra based normal form methods* published in *Celestial Mechanics and Dynamical Astronomy, Vol. 132, 14 (2020)* [95]. The paper was authored by Roberto Armellin, Martin Berz, and me.

Given the detailed understanding of the differential algebra (DA) normal form algorithm from Sec. 2.3 and Chapter 3, we present its application in a new technique for the calculation of entire continuous sets of orbits, which remain in long term relative bounded motion under zonal gravitational perturbation. We will see that the application of the DA normal form algorithm in this particular case is only possible due to a well-chosen Poincaré surface for the Poincaré return map (Sec. 2.2), which captures the critical phase space behavior at the right space-time instance, which requires a combination of dimension-reducing phase space projections.

4.1 Introduction to Bounded Motion

The term ‘bounded motion’ is used in the field of astrodynamics to describe a special orbital flight pattern of two objects (usually man-made satellites), where the two objects remain in close proximity to each other over an extended period of time. Both objects are on orbits around a common central gravitational body like a planet, moon, asteroid, or star, and their relative distance is bounded.

In practice, bounded motion finds application in cluster flight [31] and formation flying [5] missions, which can offer many advantages compared to single spacecraft missions. From the scientific standpoint, they enable measurements of unprecedented spatial and temporal correlation, but they also have economic advantages such as allowing for redundancies within the spacecraft group, a distribution of the payload, and the adaptability of the mission by exchanging modules of the group. Missions such as PRISMA [33], GRACE [66], and TerraSAR-X and TanDEM-X [34] demonstrated the practicability of formation flying and stimulated further research in the field.

Moving from an ideal unperturbed system with elliptical Kepler orbits to the realistic mission case by considering perturbations to the dynamics makes it not trivial to find bounded motion orbits. The dominating perturbation is often due to the oblateness of the central body and the associated zonal perturbation from the second zonal harmonic coefficient J_2 of the gravitational potential. This zonal perturbation introduces a drift in the right ascension of the ascending node (RAAN) $\Delta\Omega$, the argument of periapsis, and the mean anomaly. The drift in each of the quantities is oscillating at different frequencies, which drastically increases the complexity of the bounded motion problem. Additional non-zonal gravitational perturbations break the rotational symmetry of the system and the regular oscillations in each of the quantities mentioned above, which complicates the problem even more.

To minimize the extent of formation-keeping maneuvers with control strategies during a mission, it is of great interest to the astrodynamical community to find 'naturally' bounded motion orbits for models considering as many perturbations as possible, which leave only the unmodeled perturbations to be corrected by control maneuvers. In this chapter and in [95], we present a method that allows for the design of long term relative bounded motion considering a zonal gravitational model using normal form methods. Since [95] contains an extensive literature review of previous approaches, only contributions directly linked to our technique for the zonal problem will be mentioned below.

The pioneering work by Broucke [30] on families of two dimensional quasi-periodic invariant tori around stable periodic orbits of the Ruth-reduced axially symmetric system was used by Koon *et al.* [46] in combination with Poincaré section techniques to study the J_2 problem. While this method improved first order approaches, long term bounded motion was still not achieved by placing orbits on the center manifold. Xu *et al.* [100] pointed out that long term bounded motion in the zonally perturbed system could only be achieved when the RAAN drift $\Delta\Omega$ and nodal period T_d are on average the same for each of the bounded modules (see Sec. 4.2.5). These constraints are weaker than the constraints originally derived by Martinusi and Gurfil [65].

In [9], a fully numerical technique based on stroboscopic maps was used to obtain entire families of quasi-periodic orbits producing bounded relative motion about a periodic one. This method was then used to study both: bounded motion about asteroids [8] and in low Earth, medium Earth, and geostationary orbits [10]. Numerical approaches yield bounded relative orbits with arbitrary size over very long periods of time (or infinite time in theory). However, they require complex and time-consuming algorithms.

In [42], a compromise between the analytic and numerical technique was presented based on the use of DA. DA techniques were used to expand to high order the mapping between two consecutive equatorial crossings (i.e., Poincaré maps). This enabled the study of the motion of a spacecraft for many revolutions by the fast evaluation of Taylor polynomials. The problem of designing bounded motion orbits was then reduced to the solution of two nonlinear polynomial equations, namely constraining the mean nodal period T_d and drift of the right ascension of the ascending node $\Delta\Omega$. The derived method showed an accuracy comparable with that of fully numerical methods but with a reduced complexity due to the introduced polynomial approximations. The main drawback of this technique consisted of the calculation of the mean T_d and $\Delta\Omega$ using numerical averaging over thousands of nodal crossings. This process resulted in the computationally intensive part of the algorithm and was also responsible for accuracy degradation in the case of very large separations.

The advantage of our approach is that it overcomes this limitation when calculating bounded motion orbits under zonal perturbation by the introduction of DA based normal form (DANF) methods. In particular, the high-order DANF algorithm is used to transform the Poincaré map into normal form space, in which the phase space behavior is circular and can be easily parameterized by action-angle coordinates (see Fig. 4.3). The action-angle representation of the normal form coordinates is then used to parameterize the original phase space coordinates of the Poincaré return map. The original map is averaged over a full phase space revolution by a path integral along the angle parameterization, yielding the Taylor expansion of the averaged bounded motion quantities T_d and $\Delta\Omega$, for which the bounded motion conditions are straightforwardly imposed. Sets of highly accurate bounded orbits are obtained in the full zonal problem, extending over several thousand kilometers and valid for decades. This method avoids the numerical averaging introduced in [42]. The superiority in terms of elegance, computational time, and accuracy of the new algorithm will be demonstrated using similar test cases to those presented in [42] and [10].

Before introducing our approach from [95], we start with some basics on the orbital motion under gravitational perturbation. Later we will show our results for the full zonal problem [95].

4.2 Understanding Orbital Motion Under Gravitational Perturbation

We consider the orbital motion around a single central body of mass, where the motion is only determined by the gravitational potential of the central body. Perturbations due to atmospheric drag, solar radiation pressure, or the gravitational field of other space bodies are ignored. We also ignore parabolic and hyperbolic orbits, which escape the gravitational potential due to their large enough kinetic energy.

4.2.1 The Perturbed Gravitational Potential

Any gravitational potential U can be expressed in terms of spherical harmonics $Y_{l,m}$ and the corresponding coefficients $k_{l,m}$:

$$U\left(r,\theta,\phi\right) = -\frac{\mu}{r}\left(1 + \sum_{l=1}^{\infty}\sum_{m=-l}^{l} k_{l,m}\left(\frac{R_0}{r}\right)^{l} Y_{l,m}\left(\theta,\phi\right)\right), \tag{4.1}$$

where (r,θ,ϕ) are spherical coordinates with the origin at the center of mass and where μ is the product of the gravitational constant and the mass of the central body. The coefficients of the $Y_{l,m}$ are often split into $k_{l,m} \cdot R_0^l$ to make them independent of the size R_0 of the central body.

The orientation of the coordinate system is usually chosen such that $\hat{z}$ $(\theta = 0)$ aligns with the dominating symmetry axis of the central body. The plane perpendicular to $\hat{z}$, i.e. the xy plane or $\theta = \pi/2$ plane, is referred to as the equatorial plane.

The spherical harmonics can be grouped into three categories. Zonal terms $(m = 0)$ are independent of the longitude ϕ creating zones in the vertical/latitudinal direction. Sectional terms $(m = l)$ on the other hand are independent of the latitude θ creating sections longitudinally. Tesseral terms $(0 < m < l)$ are dependent on both ϕ and θ creating a chessboard pattern on the sphere. Each of these terms is considered a gravitational perturbation to the spherically symmetric potential

$$U_0 = -\frac{\mu}{r}, \tag{4.2}$$

which only depends on the distance r.

The gravitational potentials of many rotating central bodies are dominated by their low order zonal terms, in particular, $Y_{2,0}$, since centrifugal effects of the rotation often cause a zonally dependent mass distribution with more mass at the equator and less mass at the poles compared to the sphere. Considering only the effects of zonal perturbations is also referred to as the zonal problem and is going to be the basis of our analysis. The axial symmetry conserves the angular momentum component along the symmetry axis and simplifies the potential significantly as the spherical harmonics $Y_{l,m}$ reduce to the ordinary Legendre polynomials P_l, with

$$U(r,\theta) = -\frac{\mu}{r}\left(1 + \sum_{l=1}^{\infty} J_l \left(\frac{R_0}{r}\right)^l P_l(\cos\theta)\right). \tag{4.3}$$

4.2.2 The Equations of Motion

To calculate the behavior of an object in the perturbed gravitational field, we derive the equations of motion, which describe the dynamics as a set of mathematical functions. To be consistent with previous approaches and [95], we will use cylindrical coordinates. The starting point of the derivation is the Lagrangian

$$L = \frac{1}{2}\left(\dot{\rho}^2 + \dot{z}^2 + \rho^2\dot{\phi}^2\right) - U(\rho, z, \phi) \tag{4.4}$$

of the system in cylindrical coordinates (ρ, z, ϕ), where ρ is the distance in the equatorial plane such that $r = \sqrt{\rho^2 + z^2}$ yields the total distance between the orbiting object and the center of mass. The potential takes the following form in cylindrical coordinates

$$U(\rho, z, \phi) = -\frac{\mu}{r}\left[1 + \sum_{l=1}^{\infty}\sum_{m=0}^{l}\left(\frac{R_0}{r}\right)^l P_{l,m}\left(\frac{z}{r}\right)\left(C_{l,m}\cos(m\phi) + S_{l,m}\sin(m\phi)\right)\right], \tag{4.5}$$

where $P_{l,m}$ are the associated Legendre polynomials.

For the zonal problem ($m = 0$), the $P_{l,m}$ reduce to the ordinary Legendre polynomials P_l. The coefficients $C_{l,0}$ of the zonal problem are often denoted by J_l.

The canonical momenta $\left(v_\rho, v_z, v_\phi\right)$ to the position variables (ρ, z, ϕ) are given by

$$v_\rho = \frac{\partial L}{\partial \dot{\rho}} = \dot{\rho} \quad v_z = \frac{\partial L}{\partial \dot{z}} = \dot{z} \quad v_\phi = \frac{\partial L}{\partial \dot{\phi}} = \rho^2 \dot{\phi} \doteq \mathcal{H}_z, \tag{4.6}$$

where $\mathcal{H}_z$ is the angular momentum component along the symmetry axis $\hat{z}$ and the canonical momentum to the angle ϕ. From the Lagrange-Euler equations it follows that

$$\dot{\mathcal{H}}_z = -\frac{\partial U}{\partial \phi}, \tag{4.7}$$

which is zero for the zonal problem due to the axial symmetry making $\mathcal{H}_z$ a constant of motion.

Using the Legendre transformation, the Hamiltonian

$$H = \frac{1}{2}\left(v_\rho^2 + v_z^2 + \frac{\mathcal{H}_z^2}{\rho^2}\right) + U\left(\rho, z, \phi\right) \tag{4.8}$$

is obtained. Due to the time independence of the system ($\mathrm{d}_t H = 0$), the Hamiltonian is equivalent to the energy E, which is a constant of motion.

The equations of motion are derived from the Hamiltonian via the Hamilton equations

$$\dot{\rho} = v_\rho \qquad \dot{z} = v_z \qquad \dot{\phi} = \frac{\mathcal{H}_z}{\rho^2} \tag{4.9}$$

$$\dot{v}_\rho = \frac{\mathcal{H}_z^2}{\rho^3} - \frac{dU}{d\rho} \qquad \dot{v}_z = -\frac{dU}{dz} \qquad \dot{\mathcal{H}}_z = -\frac{dU}{d\phi}. \tag{4.10}$$

The time evolution $\mathcal{X}\left(t\right)$ of the state $\mathcal{X} = \left(r, v_r, z, v_z, \phi, \mathcal{H}_z\right)^T$ of a spacecraft is determined by integrating the system of ODEs $\dot{\mathcal{X}} = f\left(\mathcal{X}\right)$ from above. The orbit $\mathcal{O}$ of the spacecraft is described by the set of all states $\mathcal{X}\left(t\right)$.

4.2.3 The Kepler Orbit

Before we investigate the orbital behavior under perturbation, it is advisable to understand the unperturbed system with the spherically symmetric gravitational potential U_0. The orbiting motion of an object in the unperturbed potential takes the Keplerian form of a closed ellipse, which makes the motion two dimensional. The plane in which the ellipse lies is called the orbital plane.

The traditional orbital elements $(a, e, i, \Omega, \omega, \nu(t))$, also called Keplerian elements, characterize the position and orbit of the object using the elliptical shape as well as the equatorial plane of the central body as a reference (see [5] for a detailed description). The variables a and e define the size (semi-major axis) and shape (eccentricity) of the ellipse, respectively. To describe the orientation of the orbital plane with respect to the central body, the reference direction $\hat{x}$ within the equatorial plane is defined. Except for orbits within the equatorial plane, the elliptical orbit intersects with the equatorial plane in two places. The intersection in the $\hat{z}$ direction (from south to north) is called the ascending node ☊. The angle between the equatorial plane and the orbital plane is called the inclination i. The angle between the reference direction $\hat{x}$ and the ascending node within the equatorial plane is the longitude or right ascension of the ascending node (RAAN) Ω. The argument of periapsis ω describes the orientation of the ellipse within the orbital plane as the angle between the ascending node and the periapsis (closest point of the ellipse to the origin). The true anomaly $\nu(t)$ yields the position of the object along the ellipse as the angle between the periapsis and the object. The time between two consecutive ascending nodes is called the nodal period T_d, with $T_d = t\,(☊_{n+1}) - t\,(☊_n)$.

4.2.4 Orbits Under Gravitational Perturbation

The elliptical orbits deform under gravitational perturbations such that the orbits no longer close after a revolution around the central body.

The description of perturbed orbits using Keplerian elements has to be carefully considered, since the four elements (a, e, ω, ν) are based on the assumption of an elliptical orbit in an unperturbed system. The elements i and Ω, on the other hand, only describe the orientation of the orbital plane, determined by position and velocity vectors of the orbiting object, but make no assumptions about the shape of the orbit. In practice, the Keplerian elements are calculated at each point in time assuming the orbit is an ellipse in an unperturbed system while propagating the object in the perturbed system.

This representation is particularly helpful when the gravitational potential is only slightly perturbed. It shows how the unperturbed elliptical orbit is influenced by the perturbations at each point in time. In Fig. 4.1, the orbital elements of a low Earth orbit ($\mathcal{O}_2$ from Sec. 4.4.1) under zonal perturbation are shown. As a reference, the orbit is also initiated with the same starting conditions but propagated considering only the spherically symmetrical part of the Earth's gravitational field.

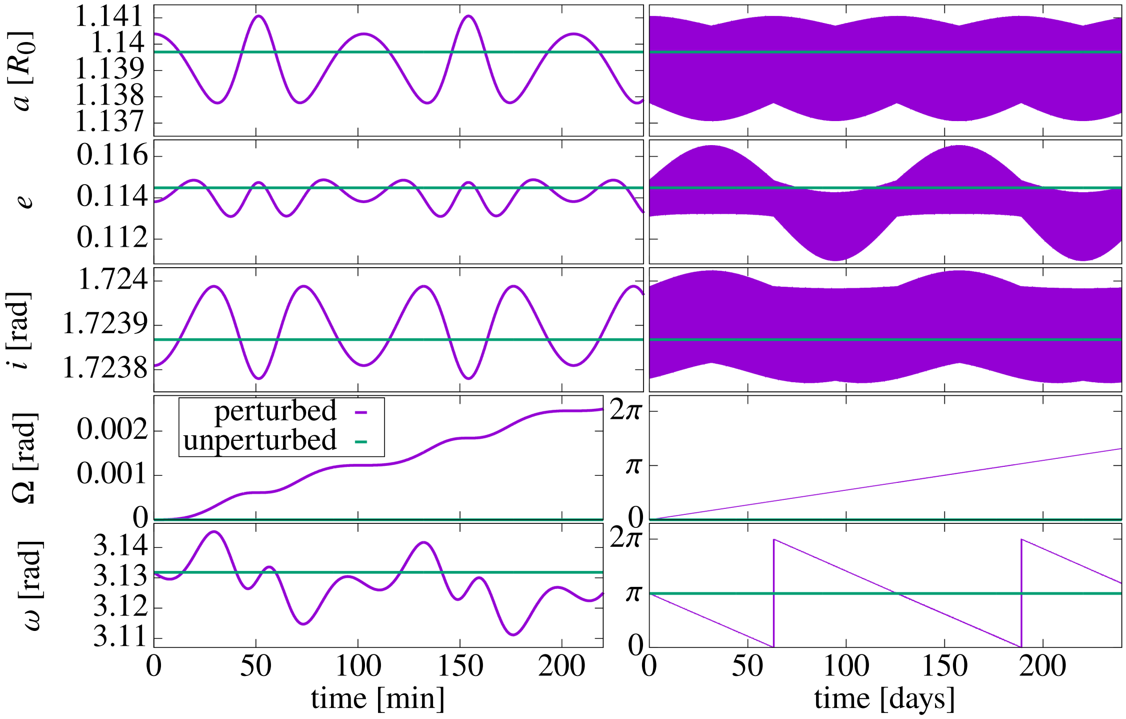

Figure 4.1: The behavior of the Keplerian elements of a low Earth orbit under zonal gravitational perturbations up to J_{15} (purple) and as a regular Kepler orbit in the unperturbed gravitational field (green) over time. Left and right plots show different time scales of the behavior. Note that the vertical scales of Ω and ω are adjusted in the right plot to show the long term behavior.

Compared to the unperturbed motion, the behavior of the Keplerian elements under zonal perturbation is quite complex. There are multiple oscillations happening at different frequencies. On the short time scale (see left plots in Fig. 4.1), there is the semi-periodic behavior associated with one orbital revolution with a nodal period of roughly 103 min. As already mentioned in the introduction, the zonal perturbation introduces a drift of the orbital plane, which is indicated by the increasing Ω in Fig. 4.1. The corresponding long term behavior suggests that the orbital plane is rotating around the symmetry axis in about 365 days. However, as we will discover in Sec. 4.4.1 and in particular in Fig. 4.4 neither the nodal period T_d nor the drift in the ascending node are constant, but they are also oscillating. The nodal period T_d, the RAAN-drift $\Delta\Omega$, and the long term behavior of a, e, and i are oscillating at the frequency of the rotation of the argument of periapsis ω, which has a period of roughly 129 days.

4.2.5 The Bounded Motion Conditions by Xu *et al.*

Considering that each orbit is individually influenced by the gravitational perturbations determining its shape and orbital period, bounded motion conditions link two orbits in space-time.

Xu *et al.* [100] showed that the conditions for bounded motion between two orbits $\mathcal{O}_1$ and $\mathcal{O}_2$ require the following conditions to be met:

$$\overline{T}_d\left(\mathcal{O}_1\right) = \overline{T}_d\left(\mathcal{O}_2\right) \tag{4.11}$$

$$\overline{\Delta\Omega}\left(\mathcal{O}_1\right) = \overline{\Delta\Omega}\left(\mathcal{O}_2\right). \tag{4.12}$$

In other words, any two orbits are in sync, if both, their average nodal period $\overline{T}_d$ and their average drift of the ascending node $\overline{\Delta\Omega}$, are the same.

The time related condition is linked to the space related condition by the space-time event at the ascending node, where the object passes through the equatorial plane from south to north. The time difference between two consecutive ascending nodes is the nodal period T_d. The angular difference between two consecutive ascending nodes is denoted by $\Delta\Omega$, also referred to as the RAAN-drift. It is defined by

$$\Delta\Omega = \phi\left(\Omega_{n+1}\right) - \phi\left(\Omega_n\right) - 2\pi\operatorname{sgn}\left(\mathcal{H}_z\right), \tag{4.13}$$

where $-2\pi\operatorname{sgn}\left(\mathcal{H}_z\right)$ ensures that $\Delta\Omega$ is the shortest angular distance between the two consecutive ascending nodes.

Under zonal perturbation, the nodal period T_d and the RAAN-drift $\Delta\Omega$ show regular oscillatory behavior (see Fig. 4.4), making their average values constants of motion. The basic goal of our approach is finding a way of cleverly calculating those average values and relating them to the constants of motion $\mathcal{H}_z$ and E. Given the relation, $\mathcal{H}_z$ and E can be chosen such that the bounded motion conditions are satisfied and the associated orbits are bound.

4.2.6 The Fixed Point Orbit

Under zonal perturbation, there are special orbits for which the nodal period T_d and the RAAN-drift $\Delta\Omega$ are constant. The associated reduced state $\mathcal{Z} = (\rho, v_\rho, z = 0, v_z)$ at the ascending nodes remains unchanged, which is why these orbits are called fixed point orbits. The orbits are also known as quasi-circular orbits, which originates from the idea of having the elliptical reference shape of the orbit rotate within the orbital plane under zonal perturbation. The fact that $r = \rho$ is constant at the ascending node for those orbits suggests that the rotating reference shape of the orbit in the orbital plane is a circle. The Keplerian elements of such a quasi-circular orbit (see Fig. 4.2) show however that e oscillates around a value slightly greater than zero, which is the reason for the word ‘quasi’. More insightful is the idea that the perturbations influence the orbit just right to yield periodic behavior after just one orbital revolution around the central body.

Compared to the Keplerian elements of non-quasi-circular orbits like the one shown in Fig. 4.1, the orbital behavior of the quasi-circular orbit is a lot more regular. Its nodal period T_d and ascending node drift $\Delta\Omega$ are constant and not oscillating as Fig. 4.4 reveals. Since the long term oscillation has

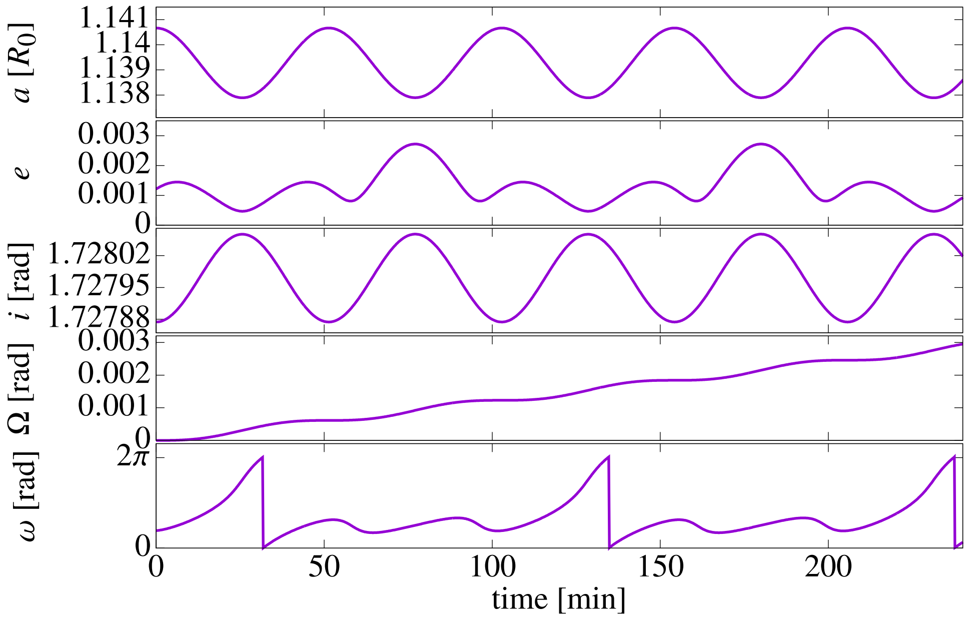

Figure 4.2: Keplerian elements of a quasi-circular low Earth orbit under Earth’s zonal gravitational perturbation.

no amplitude, the entire dynamics of a quasi-circular orbit are already captured by the time scale of minutes shown in Fig. 4.2.

For our approach, these fixed point orbits serve as a reference for entire families of orbits which all share the same average nodal period T_d and the same average RAAN-drift $\Delta\Omega$. Our method calculates a manifold in $(\rho, v_\rho, z, v_z, \mathcal{H}_z, E)$ around the fixed point, where the manifold is defined such that any two points on the manifold satisfy the bounded motion condition.

In the fully gravitationally perturbed system the axial symmetry vanishes, which introduces a ϕ dependence and results in $\mathcal{H}_z$ no longer being a constant of motion. Accordingly, fixed point orbits in the fully gravitationally perturbed systems must have a fixed point property in the full state $\mathcal{X} = (\rho, v_\rho, z = 0, v_z, \phi, \mathcal{H}_z)$. We will discuss fixed point orbits in the fully perturbed system and the possibilities of creating bounded motion manifolds around them in more detail later in this chapter, but first, we will present the method and results from [95], where manifolds of bounded motion orbits for the zonal problem are calculated.

4.3 Method of Bounded Motion Design Under Zonal Perturbation

This section is from [95]. The goal is to generate a Poincaré return map $\mathcal{P}$ that describes the dynamics of the system by characterizing how a state $(\mathcal{X}_{\text{ini}}, t = 0) \in \mathcal{O}$ within a Poincaré surface $\mathbb{S}$ returns to $\mathbb{S}$. Defining a suitable Poincaré surface is the first step in generating the map. Secondly, a reference orbit with fixed point properties has to be identified to ensure that the expansion point of the map returns to itself. The Poincaré return map is then calculated as an expansion around the reference orbit before being averaged using DA normal form methods. This yields the average nodal period $\overline{T}_d$ and average ascending node drift $\overline{\Delta\Omega}$ as a function of the system parameters and expansion variables around the reference orbit. Using DA inversion methods, the system parameters can be determined such that the bounded motion conditions are met.

4.3.1 The Poincaré Surface Space

The bounded motion conditions are defined regarding the ascending node of two orbits. To be able to enforce the bounded motion condition on our map, we choose the set of ascending nodes ($z = 0, v_z \geq 0$) as the Poincaré surface. The Poincaré surface $\mathbb{S}_{\Omega}$ can be divided into subsurfaces $\mathbb{S}_{\Omega,\mathcal{H}_z,E}$ for specific angular momentum components $\mathcal{H}_z$ and energies E. These surfaces contain all states with the parameters $(\mathcal{H}_z, E)$ that lie in the equatorial plane ($z = 0$) and satisfy $v_z > 0$. The restriction of v_z to positive values makes the relation between E and v_z (see Eq. (4.8)) bijective and therefore locally invertible in $\mathbb{S}_{\Omega,\mathcal{H}_z,E}$, so

$$\mathbb{S}_{\Omega,\mathcal{H}_z,E} = \left\{ \mathcal{X} \mid z = 0,\; v_z = \sqrt{2\left(E - U\left(r\right)\right) - v_r^2 - \left(\frac{\mathcal{H}_z}{r}\right)^2} \right\}. \tag{4.14}$$

This means that any state $\mathcal{X} \in \mathbb{S}_{\Omega,\mathcal{H}_z,E}$ is uniquely determined by (r, v_r, ϕ), since $z = 0$ and $v_z\left(r, v_r, \mathcal{H}_z, E\right)$.

4.3.2 The Fixed Point Orbit

The orbit associated with the fixed point state is called reference orbit. The reference orbit has the special property that it returns to the same reduced state $\mathcal{Z} = (r, v_r, z, v_z)^T$ after each revolution with a constant nodal period $T_d^\star$ and a constant angle advancement in ϕ, which is also referred to as the fixed point drift in the ascending node $\Delta\Omega^\star$.

For a certain set of parameters $(\mathcal{H}_z, E)$, we use DA inversion techniques iteratively to find the fixed point orbit (see Sec. 2.2 and Sec. 2.3). The iteration is initialized with an educated guess of the fixed point corresponding to the ascending node state

$$\mathcal{Z}_0 = \left(r = -\frac{1}{2E}, v_r = 0, z = 0, v_z\left(r, \mathcal{H}_z, E\right)\right)^T, \tag{4.15}$$

where r is set to the apsides of an elliptical orbit in an unperturbed gravitational field with a specific orbital energy E (see [97]).

For each iteration step n, the state $\mathcal{Z}_{n-1}$ is expanded in the variables (r, v_r). After a full orbit integration until the next ascending node intersection, the map $\mathcal{M}$ is timewise projected onto the

Poincaré surface $\mathbb{S}_{\Omega,\mathcal{H}_z,E}$ (see Sec. 2.2). The resulting Poincaré map $\mathcal{P}$ represents the one turn map in dependence on variations $(\delta r, \delta v_r)$ in the variables (r, v_r). The difference between the constant part of the map $\mathcal{P}$ and the initial state $\mathcal{Z}_{n-1}$ in the components r and v_r is denoted by Δr and Δv_r, respectively. The Poincaré map without its constant part is indicated by $\mathcal{P}'$. The next initial state $\mathcal{Z}_n$ for the iterative process will be given by the evaluation of

$$\begin{pmatrix} \mathcal{Z}_{r,n} \\ \mathcal{Z}_{v_r,n} \end{pmatrix} = \begin{pmatrix} \mathcal{P}'_r\left(\delta r, \delta v_r\right) - \delta r \\ \mathcal{P}'_{v_r}\left(\delta r, \delta v_r\right) - \delta v_r \end{pmatrix}^{-1} \left(\delta r = -\Delta r, \delta v_r = -\Delta v_r\right). \tag{4.16}$$

The process is repeated until the offset $(\Delta r, \Delta v_r)$ is smaller than a threshold value e.g. 1E-14. The resulting $\mathcal{Z}_n$ is then the ascending node state of the fixed point orbit.

4.3.3 The Calculation of Poincaré Return Map

Given a fixed point state $\mathcal{Z}^\star$ from Sec. 4.3.2 for the parameter set $(\mathcal{H}_z, E)$, the Poincaré return map $\mathcal{P} : (\mathbb{S}_\Omega, t) \to (\mathbb{S}_\Omega, t)$ is calculated as a DA expansion around that reference orbit. In the first step, the flow $\mathcal{M}$ of the fixed point and its neighborhood in $\mathbb{S}_\Omega$ (expansion in $(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$) is obtained by integrating the system of ODEs from the initial state until the reference/fixed point orbit is an element of $\mathbb{S}_{\Omega,\mathcal{H}_z,E}$ again after $T_d^\star$. In other words, the state is integrated until the orbit of $\mathcal{X}^0$ intersects with the equatorial plane from south to north again.

While the reference orbit itself is in $\mathbb{S}_{\Omega,\mathcal{H}_z,E} \subset \mathbb{S}_\Omega$ after $T_d^\star$, the expansion around the reference orbit is not in $\mathbb{S}_{\Omega,\mathcal{H}_z+\delta\mathcal{H}_z,E+\delta E} \subset \mathbb{S}_\Omega$ due to changing nodal periods of the orbits within the expansion. In order to project the flow $\mathcal{M}$ after $T_d^\star$ onto the Poincaré surface $\mathbb{S}_{\Omega,E+\delta\mathcal{H}_z,E+\delta\mathcal{H}_z}$, a timewise projection is calculated following Sec. 2.2 and [40]. The flow $\mathcal{M}$ is expanded in time to find the intersection time $t_{\text{intersec}}(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$ such that

$$\mathcal{P}_z = \mathcal{M}_z\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E, t_{\text{intersec}}\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right)\right) = 0 \tag{4.17}$$

and $\mathcal{P} = (\mathcal{M}(t_{\text{intersec}}), T_d^\star + t_{\text{intersec}}) \in (\mathbb{S}_{\Omega,\mathcal{H}_z+\delta\mathcal{H}_z,E+\delta E}, t) \subset (\mathbb{S}_\Omega, t)$.

The time component $\mathcal{P}_{T_d}$ of the Poincaré return map yields the dependence of the nodal period T_d on the system parameters and expansion variables.

4.3.4 The Normal Form Averaging

Given the fixed point Poincaré return map $\mathcal{P}$ with

$$\mathcal{P}\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) = \begin{pmatrix} \mathcal{P}_r\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) \\ \mathcal{P}_{v_r}\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) \\ \mathcal{P}_z = 0 \\ \mathcal{P}_{v_z}\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) \\ \mathcal{P}_\phi\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) \\ \mathcal{P}_{T_d}\left(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E\right) \end{pmatrix} \tag{4.18}$$

we are using only the first two components (in r and v_r) of the Poincare map for the calculation of phase space transformation provided by the DA normal form algorithm, since the motion is determined by only the (r, v_r) phase space and the parameters $(\mathcal{H}_z, E)$. The reduced map is denoted by $\mathcal{K} = (\mathcal{P}_r, \mathcal{P}_{v_r})^T$.

The normal form transformation $\mathcal{A}(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$ (see Eq. (2.33)) and its inverse are used to transform the map $\mathcal{K}$ such that

$$\mathcal{A} \circ \mathcal{K} \circ \mathcal{A}^{-1}\ \left(q_{\mathrm{NF}}, p_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E\right) = \mathcal{K}_{\mathrm{NF}}\left(q_{\mathrm{NF}}, p_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E\right) \tag{4.19}$$

is rotational invariant in the normal form phase space coordinates $(q_{\mathrm{NF}}, p_{\mathrm{NF}})$ up to the order of calculation. In other words, the distorted phase space curves in original phase space coordinates $(\mathcal{P}_r(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E), \mathcal{P}_{v_r}(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E))$ are transformed to circles in the normal form coordinates $(Q_{\mathrm{NF}}(q_{\mathrm{NF}}, p_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E), P_{\mathrm{NF}}(q_{\mathrm{NF}}, p_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E))$ as Fig. 4.3 illustrates.

By rewriting the normal form coordinates $(q_{\mathrm{NF}}, p_{\mathrm{NF}})$ in an action-angle representation $(r_{\mathrm{NF}}, \Lambda)$ with

$$\begin{pmatrix} q_{\mathrm{NF}} \\ p_{\mathrm{NF}} \end{pmatrix} = r_{\mathrm{NF}} \begin{pmatrix} \cos\Lambda \\ \sin\Lambda \end{pmatrix}, \tag{4.20}$$

each normal form phase space curve is characterized by the normal form radius (action) r_{NF} and the path along each curve is parameterized by the angle Λ. Using the inverse normal form transformation $\mathcal{A}^{-1}$ (see Eq. (2.34)), the original phase space variables $(\delta r, \delta v_r)$ of $\mathcal{P}$ (and $\mathcal{K}$) are expressed in

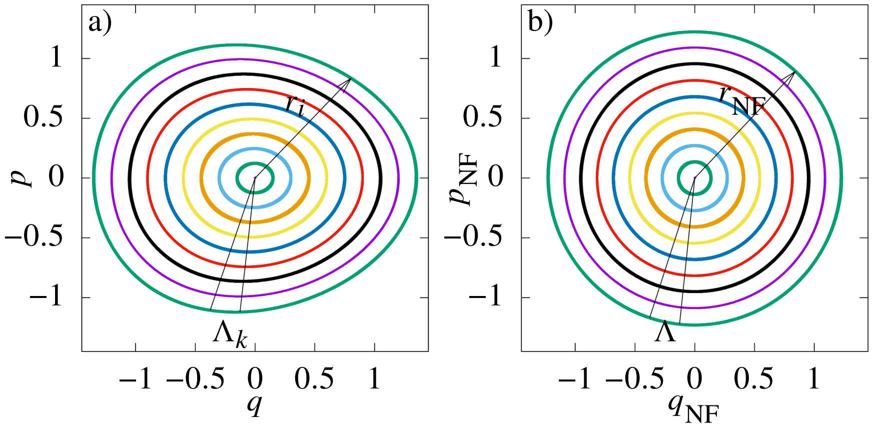

Figure 4.3: a) Distorted phase space behavior in the original phase space (q, p) and b) circular behavior in the corresponding normal form phase space $(q_{\mathrm{NF}}, p_{\mathrm{NF}})$. In a), the phase space angle advancement Λ_k and the phase space radius r_i are not constant by continuously change along each of the phase space curves. In b), the phase space behavior is rotationally invariant ('normalized') with a constant radius r_{NF} and a constant but amplitude dependent angle advancement $\Lambda(r_{\mathrm{NF}})$.

terms of the action-angle representation and variations in the system parameters $(\delta\mathcal{H}_z, \delta E)$:

$$(\delta r, \delta v_r) = \mathcal{A}^{-1}\left(q_{\mathrm{NF}}\left(r_{\mathrm{NF}}, \Lambda\right), p_{\mathrm{NF}}\left(r_{\mathrm{NF}}, \Lambda\right), \delta\mathcal{H}_z, \delta E\right). \tag{4.21}$$

The Poincaré map $\mathcal{P}(r_{\mathrm{NF}}, \Lambda, \delta\mathcal{H}_z, \delta E)$ is then averaged over a full phase space revolution, by integrating along the angle Λ:

$$\overline{\mathcal{P}}\left(r_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E\right) = \frac{1}{2\pi}\oint \mathcal{P}\left(r_{\mathrm{NF}}, \Lambda, \delta\mathcal{H}_z, \delta E\right) d\Lambda. \tag{4.22}$$

The numerical averaging presented in [42] is done in the time domain, which cannot incorporate the slightly different oscillation frequencies of the relevant quantities T_d and $\Delta\Omega$ for the different orbits. The key advantage of the normal form representation is that the different oscillation frequencies are captured by the amplitude dependent angle advancement in the normal form. The generalized parameterization of all normal form phase space curves makes the averaging independent of those differences in the frequency.

Splitting the integration into subsections minimizes the error of the numerical integration and considerably improves the quality and accuracy of the averaging. For n separate parameterization

$$\begin{pmatrix} q_{\mathrm{NF}} \\ q_{\mathrm{NF}} \end{pmatrix} = r_{\mathrm{NF}} \begin{pmatrix} \cos\left(\frac{2\pi(k-1)}{n}\right) & -\sin\left(\frac{2\pi(k-1)}{n}\right) \\ \sin\left(\frac{2\pi(k-1)}{n}\right) & \cos\left(\frac{2\pi(k-1)}{n}\right) \end{pmatrix} \begin{pmatrix} \cos\Lambda \\ \sin\Lambda \end{pmatrix} \quad k \in \{1, 2, ..., n\}, \tag{4.23}$$

each section is integrated over the symmetric interval of $\Lambda \in [-\pi/n, \pi/n]$.

The result of the averaging yields every component of $\mathcal{P}$ averaged over a full phase space curve. In particular, it yields the averaged drift in the ascending node $\overline{\Delta\Omega}\,(r_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E)$ and average nodal period $\overline{T}_d\,(r_{\mathrm{NF}}, \delta\mathcal{H}_z, \delta E)$.

For mission design purposes the abstract quantity r_{NF} is expressed by the original coordinates $(\delta r, \delta v_r)$ and the parameters $(\delta\mathcal{H}_z, \delta E)$ with

$$r_{\mathrm{NF}}^2(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E) = \left(q_{\mathrm{NF}}^2 + p_{\mathrm{NF}}^2\right)(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E) \tag{4.24}$$

using the normal form transformation $\mathcal{A}$, which yields how $(q_{\mathrm{NF}}, p_{\mathrm{NF}})$ depend on the original coordinates $(\delta r, \delta v_r)$ and the parameters $(\delta\mathcal{H}_z, \delta E)$.

The average drift in the ascending node $\overline{\Delta\Omega}\,(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$ and the average nodal period $\overline{T}_d\,(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$ are then projected such that the bounded motion conditions are satisfied, with

$$\Delta\Omega^\star = \overline{\Delta\Omega}\,(\delta r, \delta v_r, \delta\mathcal{H}_z\,(\delta r, \delta v_r)\,, \delta E\,(\delta r, \delta v_r)) \tag{4.25}$$

$$T_d^\star = \overline{T}_d\,(\delta r, \delta v_r, \delta\mathcal{H}_z\,(\delta r, \delta v_r)\,, \delta E\,(\delta r, \delta v_r))\,. \tag{4.26}$$

In this process, DA inversion methods are used to find $\delta\mathcal{H}_z(\delta r, \delta v_r)$ and $\delta E(\delta r, \delta v_r)$. The dependence of $\mathcal{H}_z$ and E on orbital parameters for bounded motion orbits were previously discussed in [90, 79].

Theoretically, one could have proceeded with the abstract invariant of motion r_{NF} to satisfy the bounded motion condition with $\delta\mathcal{H}_z(r_{\mathrm{NF}})$ and $\delta E(r_{\mathrm{NF}})$. For specific bounded orbits one would then have chosen a value for r_{NF} to calculate $(\delta\mathcal{H}_z, \delta E)$ and afterwards the initial values for (r, v_r) by using Eq. (4.21), where Λ can be chosen freely.

4.4 Bounded Motion Results

The following results were already discussed in [95]. In this section, we will apply the normal form methods for bounded motion of low Earth and medium Earth orbits. For this, we use fixed point orbits of the zonal problem that have previously been investigated by He *et al.* [42] for the low Earth orbit (LEO) and Baresi and Scheeres [10] for the medium Earth orbit (MEO).

As explained above, the fixed point Poincaré maps $\mathcal{P}$ are calculated as an expansion in the variables $(\delta r, \delta v_r, \delta\mathcal{H}_z, \delta E)$ around the respective fixed point orbit. In the calculation we consider zonal perturbations up to the J_{15}-term, since investigations in [42] indicated no considerable influence of J_k terms for $k > 15$. We are using DA maps of 8th order, which provide the best balance of accuracy and computation time. Additionally, the following dimensionless units are used: distances are considered in units of the average Earth radius $R_0 = 6378.137$ km and time is considered in units of $T_0 = 806.811$ s such that the gravitational constant assumes the value $\mu = 1$. Thus, velocities are in units of $R_0/T_0 = 7.905$ km/s.

It will be shown that the DANF method provides entire sets of bounded motions that extend far beyond the realistic/practical scope. Since the approach is based on polynomial expansions, it is obvious it will have to fail at some point. After presenting the bounded motion results for the LEO and MEO case, we take a look at the limitations of the DANF method and the resulting sets for very large distances between orbits.

4.4.1 Bounded Motion in Low Earth Orbit

In a first comparison, we are investigating bounded motion around a pseudo-circular LEO that was also considered in [42]. The pseudo-circular orbit corresponds to the reduced fixed point state

$$(r^\star, v_r^\star) = (1.14016749, -1.05621369\text{E-}3) \tag{4.27}$$

for the parameters $(\mathcal{H}_z, E) = (-0.16707295, -0.43870527)$. The orbit has a fixed nodal period of $T_d^\star = 7.64916169$ ($\approx$ 103 min) and a constant ascending node drift of $\Delta\Omega^\star = 1.22871195\text{E-}3$ rad (0.0704°). The vertical position z of the Poincaré fixed point orbit are defined by the Poincaré section ($z = 0$) and Eq. (4.14) with $v_z^\star\left(r^\star, v_r^\star, \mathcal{H}_z, E\right) = 0.92518953$.

The computation of the Poincaré map took 165 seconds on a Lenovo E470 with an Intel®Core$^{\text{TM}}$ i5-7200U CPU 2.5GHz. The map confirms the fixed point property of the orbit, since the offset of the constant part of the map from the initial coordinates is well within the numerical error of the integration with $(\Delta r, \Delta v_r, \Delta z, \Delta v_z) = (4\text{E-}15, 5\text{E-}13, -1\text{E-}15, -4\text{E-}15)$. The normal form transformation of the reduced fixed point Poincaré map $\mathcal{K} = (\mathcal{P}_r, \mathcal{P}_{v_r})^T$ is calculated via the DA normal form algorithm (in 90 milliseconds). The circular phase space behavior in normal form space is parameterized using the action-angle notation $(r_{\mathrm{NF}}, \Lambda)$. The phase space parameterization is transformed back to the original coordinates of the Poincaré map. The Poincaré map is averaged (in 52 milliseconds) over a full phase space rotation using 8 subsections following the procedure outlined in Sec. 4.3.4. Afterwards, the variable r_{NF} is expressed in terms of δr, δv_r, $\delta \mathcal{H}_z$ and δE before the variations in the constants of motion $(\delta \mathcal{H}_z, \delta E)$ are matched dependent on $(\delta r, \delta v_r)$ such that the averaged expressions for T_d and $\Delta\Omega$ satisfy the bounded motion conditions (Eq. (4.25) and Eq. (4.26)). Note that above we are not listing the computation time for the computation steps that are performed very quickly.

Considering bounded orbits initiated with the same v_r as the pseudo-circular orbit ($\delta v_r = 0$), the dependence of $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ are provided in Tab. 4.1.

Table 4.1: The expansion of $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ for relative bounded motion orbits with an average nodal period $\overline{T_d} = 7.64916169$ (≈103 min) and an average ascending node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad. The expansion is relative to the pseudo-circular LEO from [42].

$\mathcal{H}_z(\delta r, \delta v_r = 0) =$		$E(\delta r, \delta v_r = 0) =$	
− 0.16707295		− 0.43870527	
+ 0.32072807	δr^2	− 0.31602983E-3	δr^2
+ 0.25767948E-3	δr^3	− 0.25390482E-6	δr^3
− 0.19132824	δr^4	− 0.31003174E-3	δr^4
+ 0.53296708E-4	δr^5	− 0.85361819E-6	δr^5
+ 0.12006391E-1	δr^6	− 0.32152252E-3	δr^6
+ 0.60713391E-3	δr^7	− 0.24661573E-5	δr^7
− 0.19751494	δr^8	− 0.21784073E-3	δr^8

To show that the expansion of $\delta \mathcal{H}_z$ and δE provide relative bounded motion orbits, we illustrate the long term behavior of three LEOs relative to one another. The first orbit is the fixed point/pseudo-circular orbit and is denoted by $\mathcal{O}_0$. The second orbit ($\mathcal{O}_1$) is initiated at $\delta r = 0.06$ with $\delta v_r = 0$. The third orbit ($\mathcal{O}_2$) is initiated at $\delta r = 0.13$ with $\delta v_r = 0$. The last two both have an initial longitudinal offset of $\phi = 0.5°$ relative to $\mathcal{O}_0$. The specific values of the orbits are given in Tab. 4.2.

In the $(\delta r, \delta v_r)$ phase space, $\mathcal{O}_1$ and $\mathcal{O}_2$ oscillate around the fixed point $(r_0, v_{r,0})$ of $\mathcal{O}_0$.

Table 4.2: The LEOs below are all initiated at $v_{r,0} = -1.05621369\text{E-}3$ and $r_0 = 1.14016749 + \delta r$, and have an average nodal period of $\overline{T_d} = 7.64916169$ (≈103 min) and an average ascending node drift of $\overline{\Delta\Omega} = 1.22871195\text{E-}3$ rad. The pseudo-circular LEO from [42] is denoted by $\mathcal{O}_0$.

	δr	δv_r	ϕ	$\mathcal{H}_z$	E
$\mathcal{O}_0$	0.00	0	0.0°	-0.16707295	-0.43870527
$\mathcal{O}_1$	0.06 (383 km)	0	0.5°	-0.16592075	-0.43870642
$\mathcal{O}_2$	0.13 (829 km)	0	0.5°	-0.16170668	-0.43871071

Accordingly, the altitude of those orbits is also changing and roughly captured by $r_0 \pm \delta r$, which means that $\mathcal{O}_2$ already reaches very low altitudes around $r = 1.01$.

In Fig. 4.4 we show that the bounded motion conditions are met: the oscillatory behavior of the nodal period T_d and the ascending node drift $\Delta\Omega$ of the two orbits $\mathcal{O}_1$ and $\mathcal{O}_2$ average out to the same value, respectively, which corresponds to the constant nodal period $T_d^\star$ and constant ascending node drift $\Delta\Omega^\star$ of the fixed point orbit $\mathcal{O}_0$.

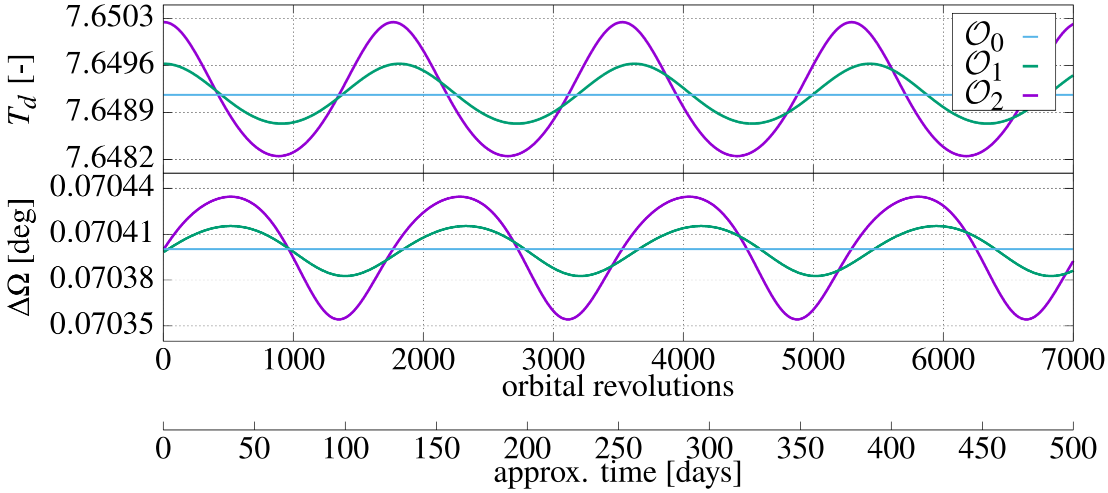

Figure 4.4: Oscillatory behavior of the bounded motion quantities T_d and $\Delta\Omega$ of the bounded LEOs $\mathcal{O}_1$ and $\mathcal{O}_2$ initiated at $\delta r = 0.06$ and $\delta r = 0.13$, respectively. Additionally, the constant nodal period $T_d^\star = 7.64916169$ and constant ascending node drift of $\Delta\Omega^\star = 0.0704°$ of the fixed point orbit $\mathcal{O}_0$ are shown. The periods of oscillation are 1763 orbital revolutions (126 days) for $\mathcal{O}_2$, 1810 orbital revolutions (129 days) for $\mathcal{O}_1$, and 1823 orbital revolutions (130 days) for $\delta r \longrightarrow 0$ of $\mathcal{O}_0$. The shown results are generated by numerical integration.

The bounded motion is further confirmed by Fig. 4.5, which shows the total distance between the three LEOs respectively for 14 years. Furthermore, Fig. 4.5 illustrates the relative radial and along-track distance between the orbit pairs from the perspective of one of the orbits in the pair.

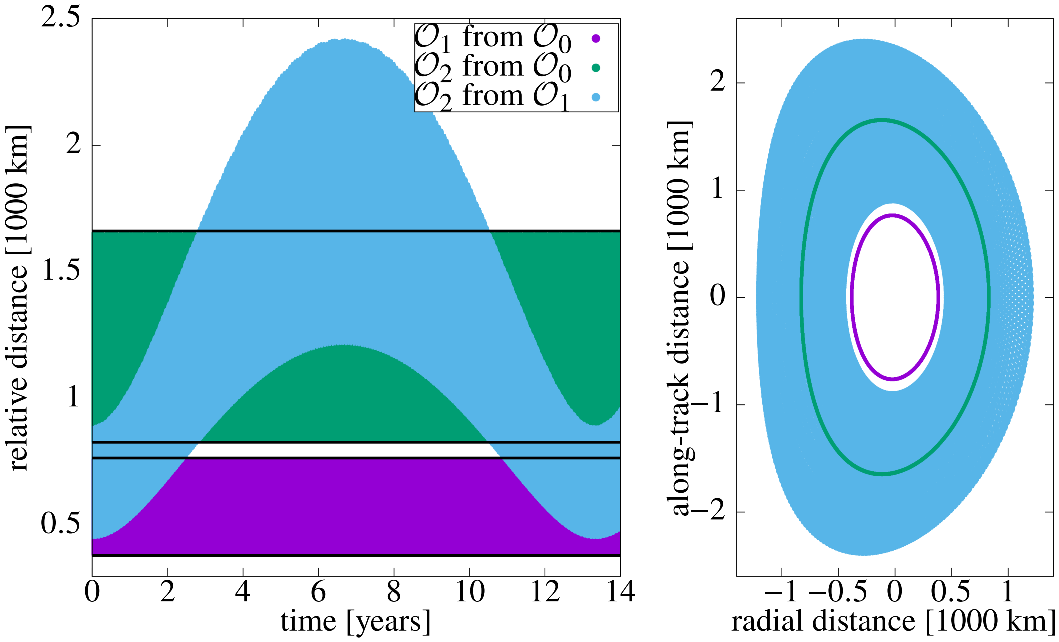

Figure 4.5: Relative bounded motion of LEOs with an average nodal period of $\overline{T_d}$ = 7.64916169 (≈103 min) and an average node drift of $\overline{\Delta\Omega}$ = 1.22871195E-3 rad for 14 years. The total relative distance between the orbits is shown in the left plot and the right plot shows the relative radial and along-track distance between orbit pairs from the perspective of one of the orbits in the pair. The oscillation in the relative distance between $\mathcal{O}_2$ and $\mathcal{O}_1$ is caused by the rotating orbital orientation of the orbits at different frequencies.

Apart from yielding long term bounded motion, the normal form methods also provide the average angle advancement Λ in the (r, v_r) phase space. This angle advancement is directly linked to the rotation frequency ω_p of the orbit (and its apsides) within its orbital plane, which causes the oscillation of T_d and $\Delta\Omega$ shown in Fig. 4.4 with ω_p. One (r, v_r) phase space rotation corresponds to one revolution of the orbit (and its apsides) within its orbital plane. Accordingly, the frequency $\omega_p = \Lambda/2\pi$ is equivalent to the definition of the tune and the tune shifts $\nu + \delta\nu$, which are just the normalized angle advancement separated into its constant part (the tune ν) and its amplitude dependent part (the tune shifts $\delta\nu$). The normal form yields the average angle advancement Λ dependent on $(r_{\text{NF}}, \delta\mathcal{H}_z, \delta E)$. After normalizing Λ, by division by 2π, and replacing r_{NF} by an expression of $(\delta r, \delta v_r)$ and $(\delta\mathcal{H}_z, \delta E)$ according to Eq. (4.24), and using the expressions for $(\delta\mathcal{H}_z(\delta r, \delta v_r), \delta E(\delta r, \delta v_r))$ from earlier, the frequency $\omega_p(\delta r, \delta v_r)$ is obtained for the bounded motion orbits around the fixed point LEO. The coefficients of ω_p for $\delta v_r = 0$ are given in Tab. 4.3.

Table 4.3: Expansion of $\omega_p(\delta r, \delta v_r = 0)$ of relative bounded motion LEOs with an average nodal period $\overline{T_d} = 7.64916169$ ($\approx$103 min) and an average node drift of $\overline{\Delta\Omega} = 1.22871195$E-3 rad. The expansion is relative to the pseudo-circular LEO from [42].

$\omega_p(\delta r, \delta v_r = 0) =$	
$+\,0.54868728$E-3	
$+\,0.10803872$E-2	δr^2
$+\,0.86800515$E-6	δr^3
$+\,0.10552068$E-2	δr^4
$+\,0.29106874$E-5	δr^5
$-\,0.76284414$E-3	δr^6
$+\,0.39324207$E-5	δr^7
$-\,0.35077526$E-1	δr^8

Accordingly, the periods of the oscillations of the nodal periods T_d and the ascending node drifts $\Delta\Omega$ in Fig. 4.4 (in units of orbital revolutions) are just the inverse of the frequencies $\omega_p(\delta r = 0.06) = 5.52590498$E-4 and $\omega_p(\delta r = 0.13) = 5.67242676$E-4. These frequencies also help explain the oscillation of the total relative distance range between $\mathcal{O}_1 \rightleftarrows \mathcal{O}_2$ over 13.3 years in Fig. 4.5.

While $\mathcal{O}_1$ shows repetitive behavior after 1809.7 orbital revolutions (129.3 days), the behavior of $\mathcal{O}_2$ is repetitive after 1762.9 orbital revolutions (125.9 days). Accordingly, the two orbits will be in and out of sync regarding their orbital orientation, while maintaining bounded due to the matching average nodal period and ascending node drift. Specifically, the two orbits will be back in sync after about 68170 orbital revolutions (4869 days $\equiv$ 13.3 years) as Fig. 4.5 illustrates, since $\mathcal{O}_1$ will have turned 37.7 times while $\mathcal{O}_2$ will have turn exactly once less, namely, 36.7 times, bringing them both back into the same orbital orientation to one other before moving apart again.

In conclusion, our first comparison showed the superiority of the normal form methods, particularly, compared to the iterative map evaluation method in [42], where numerical adjustments to the method were required to provide long term relative bounded motion for $\delta r = 0.11$.

In Sec. 4.4.3 we will show that the DANF method even provides hypothetical long term bounded motion up to $\delta r = 0.3$, which covers all realistic cases until $\delta r \approx 0.14$ and further hypothetical (non-practical) cases with altitudes below the Earth's surface.

In the next comparison, we are going to investigate bounded motion much farther from the Earth's surface. Accordingly, we expect a larger theoretical and practical bounded motion range from the DANF method, due to a weaker influence of the zonal perturbations.

4.4.2 Bounded Motion in Medium Earth Orbit

In this comparison, we are considering a medium Earth orbit (MEO) from [10, p. 11] initiated at $r = 26562.58$ km, $v_r = -9.05\text{E-}4$ km/s and $v_z = 3.18$ km/s. In the units of $R_0 = 6378.137$ km and $T_0 = 806.811$ s, the zonal problem with J_2 to J_{15} yields a fixed point orbit at $(r^\star, v_r^\star) = (4.17198963, -1.14150072\text{E-}4)$ and $v_z^\star = 0.40154964$ for the parameters $(\mathcal{H}_z, E) = (1.16863390, -0.11984818)$. The fixed point orbit has a fixed nodal period $T_d^\star = 53.5395648$ ($\approx$12 hours) and constant drift in the ascending node of $\Delta\Omega^\star = -3.35410945\text{E-}4$ rad (-0.0192°). The angular momentum component $\mathcal{H}_z$ is positive for this orbit in contrast to the LEO from Sec. 4.4.1, which means that $\dot{\phi}$ is positive and the orbit is moving eastwards.

The same computer system as in Sec. 4.4.1, took 131 seconds for the computation of the map. The offset of the integration with $(\Delta r, \Delta v_r, \Delta z, \Delta v_z) = (-4\text{E-}15, -2\text{E-}13, -4\text{E-}15, 2\text{E-}16)$ is well within the range of the numerical error of the integration. After the normal form transformation (in 100 milliseconds) and the averaging (in 62 milliseconds) following the same procedure as in Sec. 4.4.1, the dependencies of the constants of motion $(\mathcal{H}_z, E)$ on $(\delta r, \delta v_r)$ were calculated. Below, Tab. 4.4 yields $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$.

To illustrate that the DANF methods also provide bounded motion for this set of parameters, we consider the long term behavior of three MEOs relative to one another. The fixed point/pseudo-circular orbit denoted by $\mathcal{O}_0$. Since $r^\star$ of the fixed point MEO is about four times the $r^\star$ of the low Earth fixed point orbit from the previous section, the bounded orbits are initiated at four times the distance compared to the LEO investigation in Sec. 4.4.1. The orbit $\mathcal{O}_1$ is initiated at $\delta r = 0.24$ (1531 km) with $\delta v_r = 0$ and $\mathcal{O}_2$ is initiated at $\delta r = 0.52$ (3317 km) with $\delta v_r = 0$. These relative distances are already larger than distances that are used in practice. Again, both orbits have an initial longitudinal offset of $\phi = 0.5°$ relative to $\mathcal{O}_0$. The specific values of the orbits are given in Tab. 4.5.

Table 4.4: The expansion of $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ for relative bounded motion MEOs with an average nodal period of $\overline{T_d} = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The expansion is relative to the pseudo-circular MEO from [10].

$\mathcal{H}_z(\delta r, \delta v_r = 0) =$		$E(\delta r, \delta v_r = 0) =$	
$+\,1.16863390$		$-\,0.11984818$	
$-\,0.16787983$	δr^2	$-\,0.11295792\text{E-}05$	δr^2
$-\,0.57819536\text{E-}5$	δr^3	$-\,0.38903865\text{E-}10$	δr^3
$+\,0.72342680\text{E-}2$	δr^4	$-\,0.16786161\text{E-}07$	δr^4
$+\,0.16208617\text{E-}6$	δr^5	$-\,0.34176382\text{E-}11$	δr^5
$-\,0.69493130\text{E-}4$	δr^6	$-\,0.28279909\text{E-}08$	δr^6
$+\,0.11561378\text{E-}6$	δr^7	$+\,0.27190622\text{E-}12$	δr^7
$+\,0.54888817\text{E-}4$	δr^8	$-\,0.51224108\text{E-}10$	δr^8

Table 4.5: The MEOs below are all initiated at $v_{r,0} = -1.14150072\text{E-}4$ and $r_0 = 4.17198963 + \delta r$, and have an average nodal period of $\overline{T}_d = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The orbit $\mathcal{O}_0$ is the pseudo-circular MEO from [10].

	δr	δv_r	ϕ	$\mathcal{H}_z$	E
$\mathcal{O}_0$	0.0	0	0.0°	1.16863390	-0.119848175
$\mathcal{O}_1$	0.24 (1531 km)	0	0.5°	1.15898794	-0.119848240
$\mathcal{O}_2$	0.52 (3317 km)	0	0.5°	1.123766254	-0.119848482

Equivalent to Fig. 4.4 we show that the bounded motion conditions are met for the chosen MEOs in Fig. 4.6. The oscillatory behavior of the nodal period T_d and the ascending node drift $\Delta\Omega$ of the two orbits $\mathcal{O}_1$ and $\mathcal{O}_2$ average out to the same value, respectively, which correspond to the constant nodal period $T_d^\star$ and constant ascending node drift $\Delta\Omega^\star$ of the fixed point orbit $\mathcal{O}_0$. In contrast to the investigated LEOs, the oscillation period of the bounded motion quantities of the MEOs increases with increasing δr. The period of oscillation in the MEO cases is also about two orders of magnitude longer with periods of 47 and 53 years for $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively, compared to the LEOs.

Using the normal form methods, the rotation frequency ω_p of the orbital orientation within its orbital plane is calculated as described in Sec. 4.4.1. The results from the expansion of ω_p confirm these periods of oscillation with $\omega_p(0.24) = 2.88842404\text{E-}5$ and $\omega_p(0.52) = 2.58516089\text{E-}5$. The expansion of ω_p dependent on δr is given in Tab. 4.6.

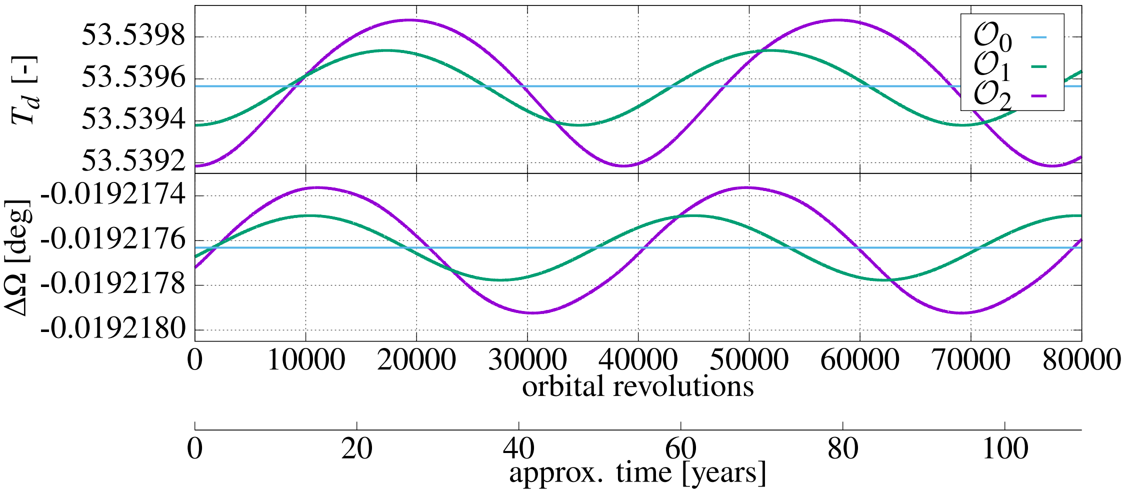

Figure 4.6: Oscillatory behavior of the bounded motion quantities T_d and $\Delta\Omega$ of the bounded MEOs $\mathcal{O}_1$ and $\mathcal{O}_2$ initiated at $\delta r = 0.24$ and $\delta r = 0.52$, respectively. Additionally, the constant nodal period $T_d^\star = 53.5395648$ and constant ascending node drift of $\Delta\Omega^\star = -0.0192176316$ deg of the fixed point orbit $\mathcal{O}_0$ are shown. The periods of oscillation are 38682 orbital revolutions (52.9 years) for $\mathcal{O}_2$, 34621 orbital revolutions (47.4 years) for $\mathcal{O}_1$, and 33671 orbital revolutions (46.1 years) for $\delta r \to 0$ of $\mathcal{O}_0$. The shown results are generated by numerical integration.

Table 4.6: Expansion of $\omega_p(\delta r, \delta v_r = 0)$ of relative bounded motion orbits with an average nodal period of $\overline{T}_d = 53.5395648$ ($\approx$ 12 h) and an average ascending node drift of $\overline{\Delta\Omega} = -3.35410945\text{E-}4$ rad. The expansion is relative to the pseudo-circular MEO from [10].

$\omega_p(\delta r, \delta v_r = 0) =$	
+ 0.29699500E-04	
− 0.14137545E-04	δr^2
− 0.48691156E-09	δr^3
− 0.22644327E-06	δr^4
− 0.43912160E-10	δr^5
− 0.10717280E-05	δr^6
− 0.10374073E-09	δr^7
+ 0.23789772E-05	δr^8

Fig. 4.7 shows the long term bounded motion behavior by illustrating the relative total distance between the orbits and their relative radial and along-track distances. Due to the long oscillation periods in the bounded motion quantities of 47 and 53 years for $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively, the oscillation in the total distance between $\mathcal{O}_1$ and $\mathcal{O}_2$ is about 456 years and can therefore only be partially shown. After 456 years the orbital orientation of $\mathcal{O}_1$ will have turned 9.6 times and align again with the orbital orientation of $\mathcal{O}_2$, which will have turned 8.6 times.

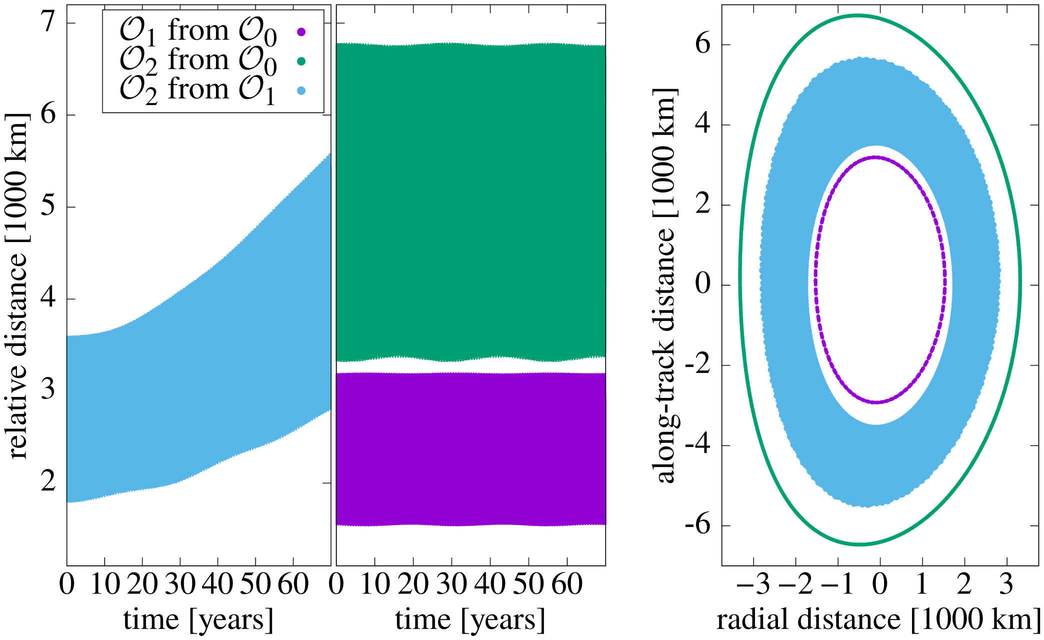

Figure 4.7: Relative bounded motion of MEOs from Tab. 4.5 with an average nodal period of $\overline{T_d}$ = 53.5395648 (≈ 12 h) and an average ascending node drift of $\overline{\Delta\Omega}$ = −3.35410945E-4 rad over 70 years. The total relative distance between the orbits is shown in the left plots and the right plot shows the relative radial and along-track distance between orbit pairs from the perspective of one of the orbits in the pair. The 'breathing' of the relative total distance between $\mathcal{O}_2$ and $\mathcal{O}_0$ originates from the rotating orbital orientation of pseudo-elliptical $\mathcal{O}_2$ relative to the pseudo-circular $\mathcal{O}_0$. Due to the very long rotation periods, only the first 70 years of the relative distance oscillation and radial/along-track behavior between $\mathcal{O}_2$ and $\mathcal{O}_1$ could be shown.

The 'breathing' of the relative distance between the orbits is particularly noticeable for the orbit pair of $\mathcal{O}_2$ and $\mathcal{O}_0$. The frequency of the 'breathing' is $2\omega_p$ which is a result of the rotation of the orbital orientation of the pseudo-elliptical $\mathcal{O}_2$ relative to the pseudo-circular $\mathcal{O}_0$. Since the orbital shape of the pseudo-elliptical $\mathcal{O}_2$ is approximately symmetric along its semi-major axis, one full rotation of the orbital orientation corresponds to two breathing cycles.

In conclusion, our methods also provided an entire set of long term relative bounded motion around the considered fixed point MEO from [10], which was validated far beyond practical relative distances. In the following section, the limitations of our method are investigated. The investigations will show that the validity of the sets presented in Sec. 4.4.1 and Sec. 4.4.2 extends over about twice the already presented distance from their respective fixed point orbits.

4.4.3 Testing the Limitations of the DANF Method

The previous two sections illustrated the validity of the DANF method for all practical relative distances for bounded motion and beyond. In this section, we move even further away from any practical relevance of the calculated sets of bounded motion to the limitations of our method. Since it is based on polynomial expansions, it is obvious it will fail at some point and we want to show when and how this failing process takes place.

First we pick a number of test orbits from the calculated bounded motion sets (see Tab. 4.7). In contrast to previous examples, no initial longitudinal offset relative to the respective fixed point orbits are set.

Table 4.7: The following orbit parameters are obtained by evaluating $\mathcal{H}_z(\delta r, \delta v_r = 0)$ and $E(\delta r, \delta v_r = 0)$ from Tab. 4.1 and Tab. 4.4 for various δr keeping $\delta v_r = 0$.

Test LEOs				Test MEOs			
	δr	$\mathcal{H}_z$	E		δr	$\mathcal{H}_z$	E
$\mathcal{O}_0$	0.00	-0.16707295	-0.43870527	$\mathcal{O}_0$	0.0	1.1686339	-0.11984817
$\mathcal{O}_{0.15}$	0.15	-0.15995246	-0.43871254	$\mathcal{O}_{0.6}$	0.6	1.1091311	-0.11984854
$\mathcal{O}_{0.20}$	0.20	-0.15454760	-0.43871843	$\mathcal{O}_{0.7}$	0.7	1.0881027	-0.11984873
$\mathcal{O}_{0.25}$	0.25	-0.14777078	-0.43872632	$\mathcal{O}_{0.8}$	0.8	1.0641420	-0.11984890
$\mathcal{O}_{0.30}$	0.30	-0.13975416	-0.43873648	$\mathcal{O}_{0.9}$	0.9	1.0373802	-0.11984910
$\mathcal{O}_{0.35}$	0.35	-0.13066556	-0.43874929	$\mathcal{O}_{1.0}$	1.0	1.0079682	-0.11984932
$\mathcal{O}_{0.40}$	0.40	-0.12071669	-0.43876526	$\mathcal{O}_{1.1}$	1.1	0.97607833	-0.11984957
				$\mathcal{O}_{1.2}$	1.2	0.94190725	-0.11984984
				$\mathcal{O}_{1.3}$	1.3	0.90567972	-0.11985014
				$\mathcal{O}_{1.4}$	1.4	0.86765361	-0.11985047

Fig. 4.8 illustrates the behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the chosen orbits of the LEO bounded motion set. Both quantities show oscillatory behavior centered at or close to $T_d^\star$ and $\Delta\Omega^\star$, respectively. With increasing distance δr, the influence of higher order oscillations becomes apparent. The frequency and amplitude of oscillation of the bounded motion quantities also increase with increasing distance δr.

If the bounded motion conditions are not met or only met approximately, the orbits will start drifting apart. This effect is illustrated in Fig. 4.9, which shows very slowly diverging behavior of approximately 2.6 km/year for $\delta r = 0.3$ (1913 km) and a stronger divergence of approximately 10.6

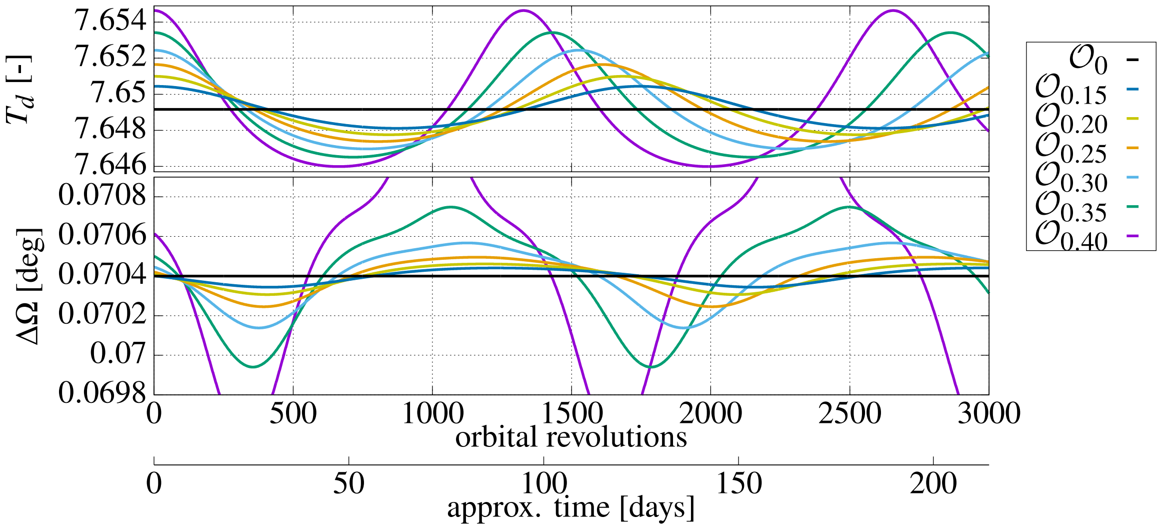

Figure 4.8: The behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the test orbits from Tab. 4.7 of the calculated LEO bounded motion set generated by numerical integration. For large δr, the influences of higher order oscillations are apparent. The frequency and amplitude of oscillation increase with increasing δr. The amplitude of $\Delta\Omega$ is particularly sensitive to δr.

km/year for $\delta r = 0.4$ (2551 km) in the left plot. The thickening curves in the radial/along-track representation of the relative orbit motion are a further indication of divergence. The strength of divergence in Fig. 4.9 can be directly linked to the size of the offsets in the bounded motion quantities from $T_d^\star$ and $\Delta\Omega^\star$, shown in Fig. 4.8.

From Fig. 4.8 and Fig. 4.9 we conclude that our method and the resulting expansions in $\mathcal{H}_z$ and E for long term bounded motion of at least 10 years around the fixed point LEO from [42] start to lose their significant accuracy for $\delta r \geq 0.3$ to satisfy the bounded motion conditions with the required precision. Note that $\delta r = 0.3$ (1913 km) is already a purely theoretical orbit with altitudes of more than 1000 km below the Earth's surface, which means that our expansions in $\mathcal{H}_z$ and E provided reliable bounded motion beyond realistic distances ($\delta r \leq 0.14$) between orbits.

The behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the chosen orbits of the MEO bounded motion set (from Tab. 4.7) are shown in Fig. 4.10. In contrast to the test LEOs, the amplitude and period of oscillation of the bounded motion quantities are decreasing with increasing distance δr, which causes the almost steady behavior of $\delta r = 1.4$ over the shown timespan and generally suppresses higher order oscillations that were seen for the LEOs. While the oscillations of

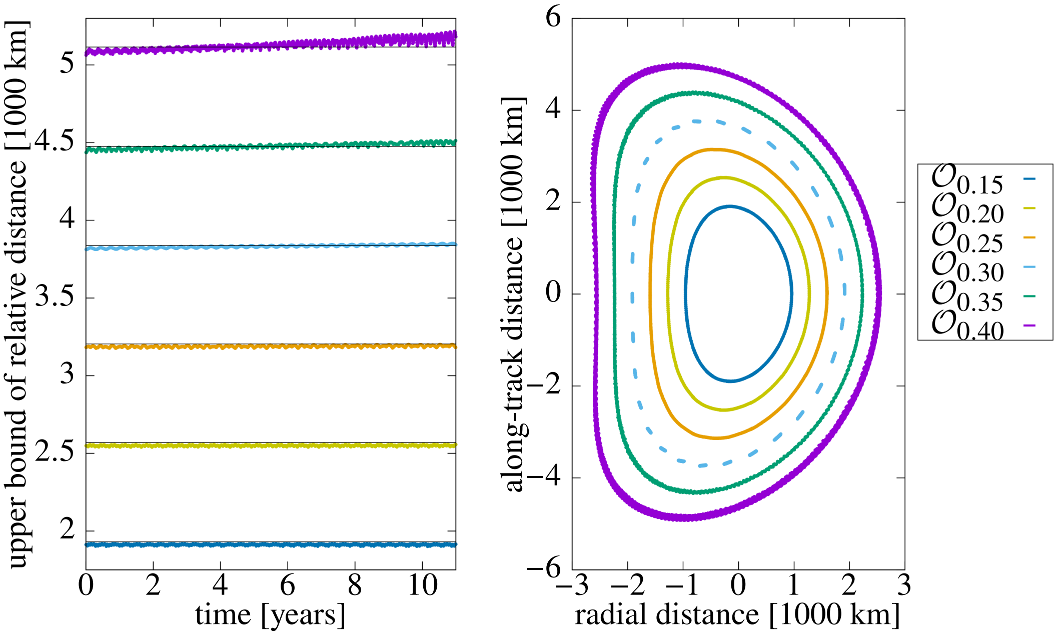

Figure 4.9: Distance between the orbits in the calculated bounded motion set and $\mathcal{O}_0$ is determined in regular time intervals with numerical integration over more than ten years. The left plot only shows the upper bound to avoid overlaps. Thin horizontal lines at the initial upper bound emphasize small changes. The dotted light blue curve (right) originates from an unintended near-resonance between the chosen time interval for distance evaluations and the orbital behavior. A measurable increase in relative distances (left) over 10 years for $\delta r \geq 0.3$ is supported by thickening curves in the radial/along-track behavior (right).

T_d are approximately centered around $T_d^\star$ (except for $\mathcal{O}_{1.4}$), the center of oscillation is increasingly diverging from $\Delta\Omega^\star$ to lower $\Delta\Omega$ for $\delta r \geq 0.8$. In other words, the expansions in $\delta\mathcal{H}_z$ and δE start failing in producing related orbits that satisfy the bounded motion condition.

The consequence of this offset in the bounded motion condition is diverging behavior between the orbits, which can be seen in Fig. 4.11. The upper bound of the total distance between the orbits starts diverging for those very large distances and the thickening curves in the radial/along-track representation of the distance of the orbits from the perspective of $\mathcal{O}_0$ further indicate this divergence. Additionally, Fig. 4.11 shows the ‘breathing’ in the total relative distance between the orbits with $2\omega_p$, which is due to the rotating orbital orientation of the orbits relative to the pseudo-circular fixed point orbit as already mentioned in the section above.

From Fig. 4.10 and Fig. 4.11, we conclude that our method and the resulting expansions in $\mathcal{H}_z$ and E for long term bounded motion of at least 70 years around the fixed point MEO from [10]

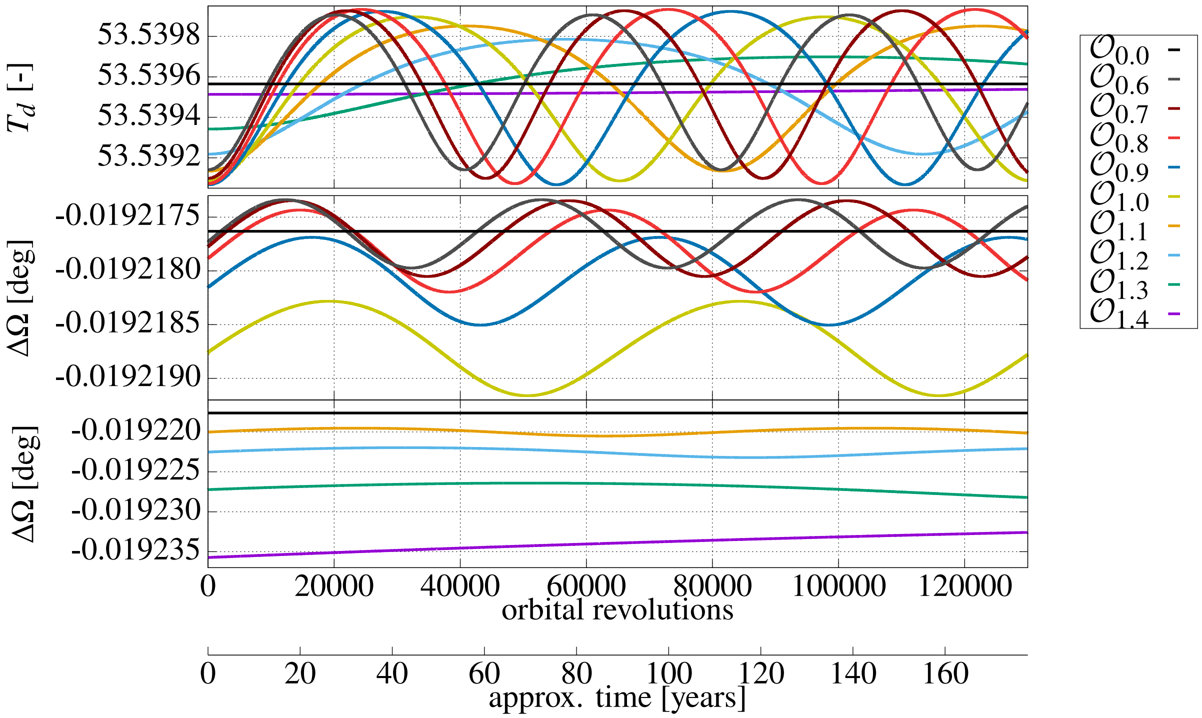

Figure 4.10: Behavior of the bounded motion quantities T_d and $\Delta\Omega$ for the test orbits from Tab. 4.7 of the calculated MEO bounded motion set generated by numerical integration. In contrast to the investigated LEOs, the frequency and amplitude of oscillation decrease with increasing δr such that $\mathcal{O}_{1.4}$ appears almost steady. For $\delta r \geq 0.8$ the center of oscillation of $\Delta\Omega$ start to drift to more negative values and away from $\Delta\Omega^\star$. To capture both, the oscillatory behavior around $\Delta\Omega^\star$ and the drift of the center of oscillation for very large δr, two plots with a different scale and range are shown for $\Delta\Omega$.

start to lose their significant accuracy for $\delta r \geq 0.9$ to satisfy the bounded motion conditions with the required precision. Interestingly, the very long 'breathing' periods for very large distances like $\delta r = 1.3$ suggested (temporary) bounded motion for the first 70 years when looking at Fig. 4.11, while Fig. 4.10 reveals the underlying diverging behavior due to the mismatched bounded motion conditions.

4.5 Conclusion

The normal form methods presented in this chapter yield parameterized sets of the constants of motion $(\mathcal{H}_z(\delta r), E(\delta r))$ for bounded orbits with an average nodal period and average ascending node drift corresponding to the fixed nodal period and ascending node drift of the reference (fixed

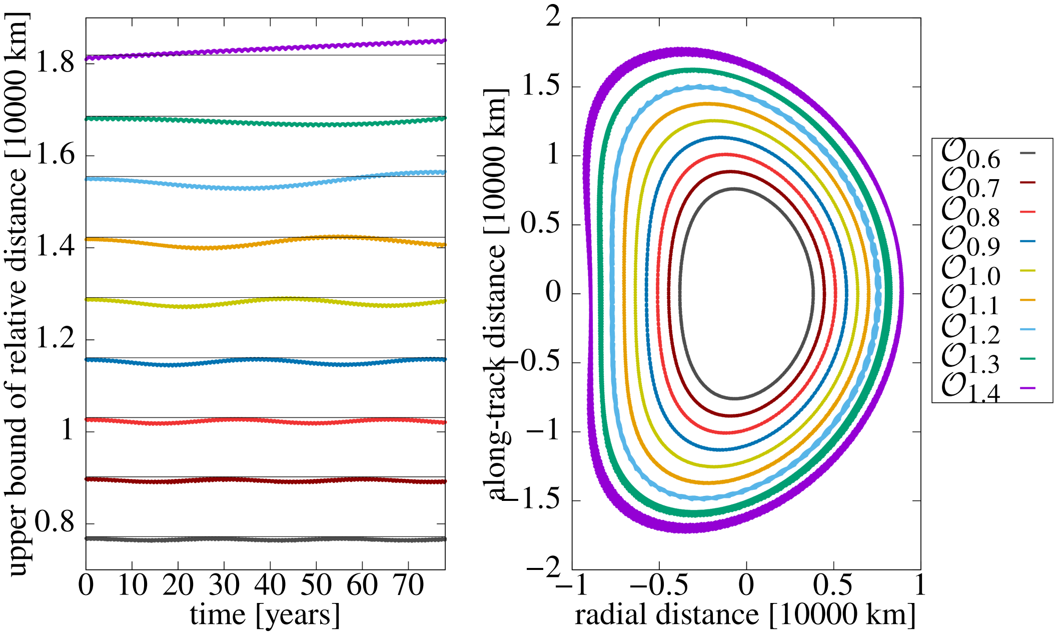

Figure 4.11: Distance between the orbits in the calculated bounded motion set and $\mathcal{O}_0$ is determined in regular time intervals by numerical integration over more than 70 years. The left plot only shows the upper bound to avoid overlaps. Thin horizontal lines at the initial upper bound emphasize small changes. The 'breathing' of the total relative distance from the orbital rotation is clearly visible. Its period increases with increasing δr until being unrecognizable due to the strong divergence for $\delta r \geq 1.4$, which is supported by thinker curves in the right plot. The weaker divergence over the 70-year timespan is already noticeable for $\delta r \geq 0.9$. The divergence is caused by the offset in respective bounded motion quantities (see Fig. 4.10).

point) orbit. The range of δr for which bounded motion orbits can be obtained is dependent on the closeness to the Earth. The closer to the Earth, the stronger the influence of the zonal perturbation on the orbits. Hence, the dynamics of bounded orbits initiated with δr differ much more when they are in a LEO than when they are in a MEO.

In comparison to the approach in [42], our method avoided the time-consuming and inaccurate numerical averaging, by using a normal form based parameterization for the averaging. As a result, the range of the bounded motion provided by our methods is more than twice as large as the range of the results in [42]. Additionally, our method does not require a separate calculation for each δr, but rather provides an expansion in $(\delta r, \delta v_r)$, which covers all orbits up to a certain maximum range that varies with the altitude of the reference trajectory.

While the method in [10] has the advantage of allowing for the calculation of bounded orbits up to arbitrary distances δr, it lacks the ability to provide parameterized sets of bounded motion just like [42].

The normal form methods are also able to provide parameterized sets of the rotation frequency of the orbits within their orbital plane. This rotation is due to the zonal perturbations in the gravitational field of the Earth since there is no rotation of the orbit for the spherically symmetric case. With increasing distance from the Earth's center ρ, the zonal perturbations J_l fall off with ρ^{-l-1}. Accordingly, it is not surprising that the rotation frequency of the MEOs is so much lower than the rotation frequency of the LEOs. Similarly, the δr dependence of the bounded motion is a lot less sensitive for the MEOs compared to the LEOs.

CHAPTER 5

STABILITY ANALYSIS OF MUON g-2 STORAGE RING

This chapter contains parts from my paper *Computation and consequences of high order amplitude- and parameter-dependent tune shifts in storage rings for high precision measurements* published in the *International Journal of Modern Physics A, Vol. 34, No. 36, 1942011 (2019)* [96]. The paper was authored by David Tarazona, Martin Berz, and me. The analysis and results from [96] are presented here and they are complemented by additional investigations into period-3 fixed point structures and their relevance in muon loss mechanisms, which was partly discussed in [88].

The differential algebra (DA) map methods (Sec. 2.2) and DA normal form methods (Sec. 2.3) are used to analyze the dynamics of particles in the storage ring of the Muon g-2 Experiment at Fermilab. In contrast to the scenarios in Chapter 3 and Chapter 4, this case of study considers two phase space dimensions. We chose a configuration of the storage ring which was utilized during one of the first data-collecting stages. This configuration is particularly interesting because of the closeness to a low-order resonance and its influence on the stability and loss rates of particles.

5.1 Introduction

Nonlinear effects of electric and magnetic field components of storage rings to confine the particles and bend their trajectory can cause substantial amplitude dependent tune shifts within the beam. Additionally, tune shifts are often sensitive to variations of system parameters, e.g., offsets of the total particle momentum δp relative to the reference momentum p_0 of the storage ring. Such amplitude and parameter dependent tune shifts lead to particles within the beam that oscillate at different frequencies, which potentially affects the beam's susceptibility to resonances and therefore its dynamics and stability. Thus, it is critical for high precision measurements like the Muon g-2 Experiment to analyze and understand these influences.

In this chapter, we investigate the dynamics within the Muon g-2 Storage Ring, which is the fundamental component of the Muon g-2 Experiment, using Poincaré return maps and DA normal form methods. A one-turn Poincaré return map yields the state of particles at a certain azimuthal location within the ring dependent on their state in the previous turn and on system parameters. The application of DA normal form methods to such maps allows for the calculations of the tune shifts and quasi-invariants of the motion around a (stable) fixed point of the map. Additionally, these maps can be used to track the phase space behavior stroboscopically. Before explaining the methods and the results, the following paragraphs will yield a short introduction to the Muon g-2 Experiment and its relevance.

The goal of the Muon g-2 Experiment at Fermilab (E989) [1] is the measurement of the anomalous magnetic dipole moment of the muon

$$a_\mu \equiv \frac{g_\mu - 2}{2}, \tag{5.1}$$

where the g-factor relates the spin and magnetic moment of a particle. Dirac theory predicts the factor to be two for charged leptons like the muon [35, 36], but hyperfine structure experiments showed that $g \neq 2$ [70, 71]. The largest radiative correction was introduced by Schwinger to explain the difference [80, 81]. Over the years more corrections were explored to gain an understanding of the deviation (g-2, where the name of the experiment comes from) [6].

Today, the most successful theory in particle physics is the standard model (SM). The most accurate calculation of the magnetic dipole moment anomaly of the muon using the standard model, a_μ^{SM}, reaches a precision of 0.39 ppm [6]. The Muon g-2 Experiment E821 conducted at the Brookhaven National Laboratory (BNL) yielded a result with a precision of 0.54 ppm [11], which differed from the SM calculation by 3.6 standard deviations. The E989 at Fermilab is the latest experiment in a series of measurements aimed at pushing the precision of the measured result even higher to reach a precision of 0.14 ppm such that the discrepancy between measurement and calculation reaches more than five standard deviations [92]. In this case, the result would be a very strong indication that the standard model is unable to describe this anomaly and would call for adjustments to the model or new theories. The results from the first set of measurements at E989 [1, 4, 2, 3] were in agreement with the measurement from BNL and had a precision of 0.46 ppm. Combined with the result from BNL, this yields an experimental precision of 0.35 ppm and a discrepancy of 4.2 standard deviations from theory predictions [1]. Expectations are that the five standard deviations are reached with the next sets of results from E989.

The experimental technique of the Muon g-2 Experiment can be briefly summarized as follows [39]: a highly spin-polarized beam of muons is created as a decay product of high energy protons, which decay into pions, which then decay into (positively charged) muons. The muons are delivered through the Muon Campus, which is part of the accelerator complex at Fermilab [87, 83], and injected into the Muon g-2 Storage Ring. During the first revolutions within the storage ring, the muon beam is prepared to adjust the emittance and limit fluctuations of the total momentum of the muons to the acceptance range of about $\pm 0.5\%$ relative to the reference momentum p_0. The prepared beam then orbits in the storage ring with only the vertical magnetic field and the four electrostatic quadrupole systems (ESQ) acting on it. The constant magnetic field forces the beam to circle around within the ring and causes the spin of the muons to precess. The four ESQ confine and focus the muons vertically [82]. The muons decay while they orbit and their spins precess. Their decay products, positrons, are measured by the calorimeter system [39] around the beamline in order to determine the spin precession frequency of the muons, which is then used together with a high-precision measurement of the magnetic field [3] to calculate the muon anomalous magnetic moment [11].

Understanding the behavior of the muon beam in the storage ring is particularly important to identify and address problems. One issue is muon loss, which introduces a systematic bias for the average spin precession frequency of the remaining particles, which affects the overall result of the measurement.

To better understand the dynamics of the muons and our methods of their analysis, we start with the introduction from [96] into how the Poincaré maps for the storage ring are generated. Then, we discuss the concept of closed orbits and the relevance of the momentum dependent closed orbit. Afterwards, the tune shift analysis from [96] is presented, which serves as the basis of our subsequent investigations into muon loss and the relevance of resonances and their associated fixed point structures.

5.2 Storage Ring Simulation Using Poincaré Maps

A storage ring is composed of various particle optical elements, each of which can be simulated in COSY INFINITY [61, 26]. For each particle optical element, there is a hypothetical ideal orbit, usually along the center of the element [19]. The ideal orbit is often characterized by a predetermined set of system parameters $\vec{\eta}_0$, for example, a specific total reference momentum of the particles. If the element is simulated as ideal, namely without perturbations, the actual trajectory of a particle initiated on the ideal orbit when entering the element (at $\vec{z}_0$) follows the ideal orbit throughout the element. However, with perturbations like imperfections in the associated fields of the element, a particle initiated at $\vec{z}_0$ might follow a trajectory different from the ideal orbit. Hence, the ideal orbit describes the actual trajectory of a particle initiated at $\vec{z}_0$ in the unperturbed case.

To analyze how an element influences the transverse phase space behavior around the ideal orbit, Poincaré maps (see Sec. 2.2) are used. The Poincaré surfaces correspond to the transverse storage ring cross section perpendicular to the optical axis at azimuthal locations before ($\mathbb{S}_i$) and after the element ($\mathbb{S}_f$). The Poincaré map $\mathcal{P}$ is expanded around the ideal orbit and expresses how the relative phase space state $\vec{z}_f \in \mathbb{S}_f$ after the particle optical element depends on variations in the system parameters $\vec{\eta}$ and on the relative phase space state $\vec{z}_i \in \mathbb{S}_i$ before the element, with $\vec{z}_f = \mathcal{P}(\vec{z}_i, \vec{\eta})$. The phase space states relative to the ideal orbit $\vec{z}$ consist of the horizontal $(q_1, p_1) = (x, a)$ and vertical $(q_2, p_2) = (y, b)$ phase space components within the Poincaré surface $\mathbb{S}$. For unperturbed elements, the Poincaré map $\mathcal{P}$ is origin preserving, with $\mathcal{P}(\vec{0}, \vec{0}) = \vec{0}$, since the trajectory follows the ideal orbit.

The transverse phase space behavior after a full revolution in the storage ring is given by the Poincaré return map $\mathcal{M}$, which is generated by composing the individual Poincaré maps $\mathcal{P}_i$ of the individual storage ring elements according to the storage ring setup ($\mathcal{M} = \mathcal{P}_k \circ \mathcal{P}_{k-1} \circ \ldots \circ \mathcal{P}_2 \circ \mathcal{P}_1$) such that the ideal orbits connect.

For the simulation of the Muon g-2 Storage Ring, a detailed nonlinear model [85, 86] of the storage ring particle optical elements has been set up using COSY INFINITY. The simulation considers the magnetic field that guides the beam around the storage ring and the four-fold symmetric electrostatic quadrupole system (ESQ) [82], which focuses the beam vertically. The ESQ is not ideal, which makes the simulation of the higher multipole components a critical aspect of the model. Additionally, perturbations due to the ESQ fringe fields and imperfections in the vertical magnetic field can be taken into account based on experimental field measurement data [3].

The model represents the magnetic field inhomogeneities by fitting 2D magnetic multipoles up to fifth order to measurement data of the magnetic field within the Muon g-2 Storage Ring (see [3, 85, 86] for details). The ESQ [82] is considered by the corresponding electrostatic potential as a 2D multipole expansion up to tenth order to accurately model the nonlinearities of the system up to the significant contribution of the 20th-pole. The fringe fields of the ESQ – the fall-off of the electric field at the edges of the ESQ components – are simulated based on numerical calculations performed with the code COULOMB [91].

The generated Poincaré return maps are expanded in the transverse phase space plane relative to the ideal orbit, where the radial phase space is denoted by (x, a) and the vertical phase space is denoted by (y, b). The coordinates x and y indicate the position in the radially outward and vertically upward direction relative to the ideal orbit. The components $a = p_x/p_0$ and $b = p_y/p_0$ are the momenta perpendicular to the ideal orbit, namely p_x and p_y, scaled by the reference momentum of the particles p_0. Additionally, the maps are expanded in the relative offset $\delta p = \Delta p/p_0$ with respect to the reference momentum p_0 to represent particles with a relative momentum offset. The relative change δp corresponds to the change of the system parameter $\vec{\eta}$.

To distinguish the effect of various elements of the storage ring and their perturbations on the dynamics of the particles, we simulated different configurations of the components in [96]. Specifically, the influence of perturbations due to ESQ fringe fields and influence from imperfections in the vertical magnetic field were studied separately. We also considered the system for two ESQ voltages, namely 18.3 kV and 20.4 kV. In this chapter of the thesis, however, we will only consider an ESQ voltage of 18.3 kV, since it offers the most interesting nonlinear dynamics and was a set-point used during the first data collection of the Muon g-2 Experiment. We are also only considering the map with the magnetic field imperfections since investigations in [96] indicated that it is the dominating perturbation and therefore yields the most realistic results. The main insights from [96] regarding the other cases will still be mentioned at the appropriate places in the text below.

5.3 The Closed Orbit

Closed orbits return to themselves after each storage ring revolution, which makes them fixed points of the Poincaré return maps. There are also low period closed orbits that return to themselves after a few turns n. These orbits correspond to low period fixed point structures in the n-turn Poincaré return map. While there are also unstable fixed points, which are discussed later, we will first focus on the properties of the stables ones.

The closed orbit is a reference for the associated particles since they oscillate around it with the closed orbit representing an equilibrium state. Accordingly, the closed orbit is sometimes also referred to as the reference orbit. In the stroboscopic view of the Poincaré return maps, the fixed point mimics an equilibrium point of the oscillatory phase space behavior around it. Using the DA normal form algorithm (see Sec. 2.3) on an origin preserving Poincaré return map, the transverse oscillation frequencies around the fixed point can be calculated. In the rest of this section, we will focus on how these closed orbits and their associated fixed points in the Poincaré return maps are determined.

5.3.1 The Closed Orbit Under Perturbation

If all components are simulated to be unperturbed, then the Poincaré return map is a composition of origin preserving Poincaré maps and hence also origin preserving. However, if the simulation considers perturbations, the actual trajectory of the expansion point may be distorted from the ideal orbit and hence not a closed orbit. Accordingly, the expansion point of the associated Poincaré return map may not be a fixed point and the map may not be origin preserving.

However, if the perturbation is sufficiently small, a fixed point $\vec{z}_{\mathrm{FP}}$ will continue to exist. Parameterizing the strength of the perturbation with $\vec{\eta}$, the origin preserving fixed point map of the unperturbed system is given by $\mathcal{M}\left(\vec{z}, \vec{\eta} = 0\right)$. To analyze the preservation of the parameter dependent fixed point, an extended map $\mathcal{N}(\vec{z},\vec{\eta}) = (\mathcal{M}(\vec{z},\vec{\eta}) - \vec{z},\vec{\eta})$ is defined [19]. If $\det(\mathrm{Jac}(\mathcal{N}(\vec{z},\vec{\eta})))|_{(\vec{z},\vec{\eta})=(\vec{0},\vec{0})} \neq 0$ then, according to the inverse function theorem, an inverse of the map $\mathcal{N}$ exists for a neighborhood $\mathbb{D}$ around the evaluation point $(\vec{0},\vec{0})$ of the Jacobian. The parameter dependent fixed point $\vec{z}_{\mathrm{FP}}(\vec{\eta})$ of $\mathcal{M}$ and hence the closed orbit of the system exists as long as $(0,\vec{\eta})$ is within the neighborhood for which invertibility has been asserted. If this is the case and the inverse $\mathcal{N}^{-1}$ around $(\vec{0},\vec{0})$ is given, then the parameter dependent fixed point can be calculated via

$$(\vec{z}_{\mathrm{FP}}(\vec{\eta}),\vec{\eta}) = \mathcal{N}^{-1}\left(\vec{0},\vec{\eta}\right). \tag{5.2}$$

Expanding the map around the parameter dependent fixed point yields the origin preserving Poincaré return map under perturbations in the system parameters.

The perturbation due to imperfections in the magnetic field distorts particles from the ideal orbit of the E989 storage ring. Accordingly, the Poincaré return map from the composition of the individual particle optical elements is not origin preserving. Using the method above, the fixed point of the map – the phase space coordinates of the closed orbit at the azimuthal location of the map – is calculated and the map is expanded around it. The result is an origin preserving fixed point map.

Calculating the fixed point for Poincaré return maps at multiple azimuthal locations of the ring indicates the form of the closed orbit (see Fig. 5.1).

The collimator locations are highlighted because they are of particular relevance for muon losses. They constitute the narrowest part around the storage region restricting the muons to amplitudes of $r = \sqrt{x^2 + y^2} < 45\text{ mm} = r_0$ relative to the center of the ring, i.e. the ideal orbit. Muons hitting a collimator during data taking for the measurement are known as *lost muons*.

While the radial/horizontal motion of the closed orbit along the storage ring is close to sinusoidal, the vertical phase space motion is disturbed into more complex behavior. In the xy projection, distorted elliptical motion around the ideal orbit along the center of the ring is indicated. All these deviations from the ideal orbit are triggered by the weak coupling of radial and vertical motion due to ppm-level imperfections of the skew quadrupole magnetic field. The form of the closed orbit is determined by the distribution of such magnetic field imperfections as well as the fields of the ESQ

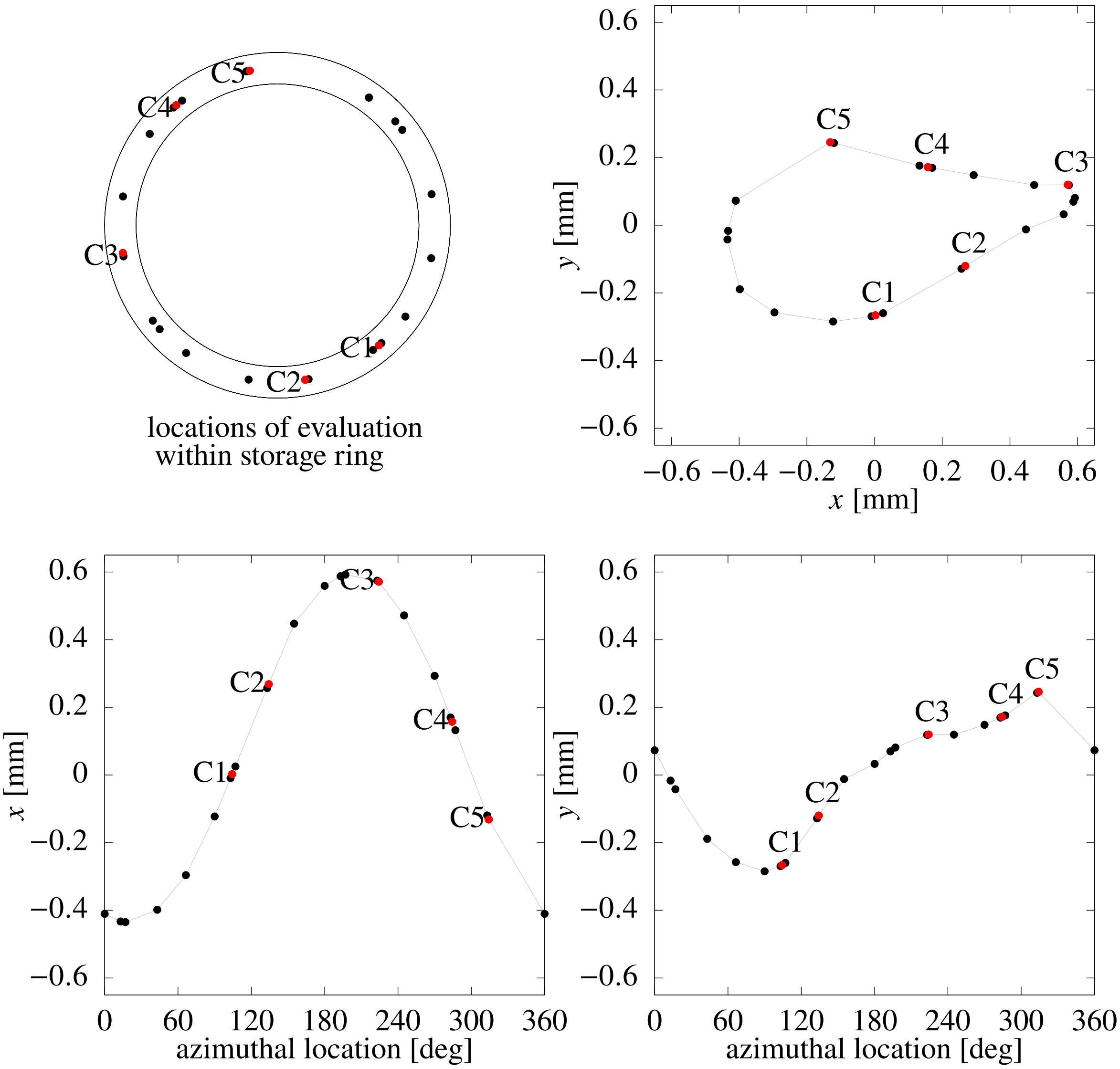

Figure 5.1: The fixed points of Poincaré return maps from various azimuthal locations around the ring indicate the behavior of the closed orbit (for $\delta p = 0$). The projections of the four dimensional fixed points into subspaces illustrate the influence of the magnetic field perturbations on the closed orbit around the ring. The results from the five collimator locations (C1-C5) are highlighted with red color. The x coordinate corresponds to displacements in the radially outward direction, while the y coordinate indicates the displacement in the vertically upward direction.

specified by the voltage.

The closed orbit we found here and showed in Fig. 5.1 is considering a particle with no momentum offset ($\delta p = 0$). Following the argumentation above the closed orbit continues to exist with perturbations in δp as will be investigated in the next section.

5.3.2 The Momentum Dependence of the Closed Orbit

The closed orbit additionally depends on system parameters like the momentum offset of the particles. Just like for the magnetic field perturbation, Eq. (5.2) is used to calculate the parameter dependent fixed point (PDFP) of the origin preserving Poincaré return map, where the parameter is the momentum offset δp. The phase space coordinates of the momentum dependent fixed point at the collimator locations in the ring are shown in Fig. 5.2.

The primary effect of the momentum offset comes from the interaction of the charged particles

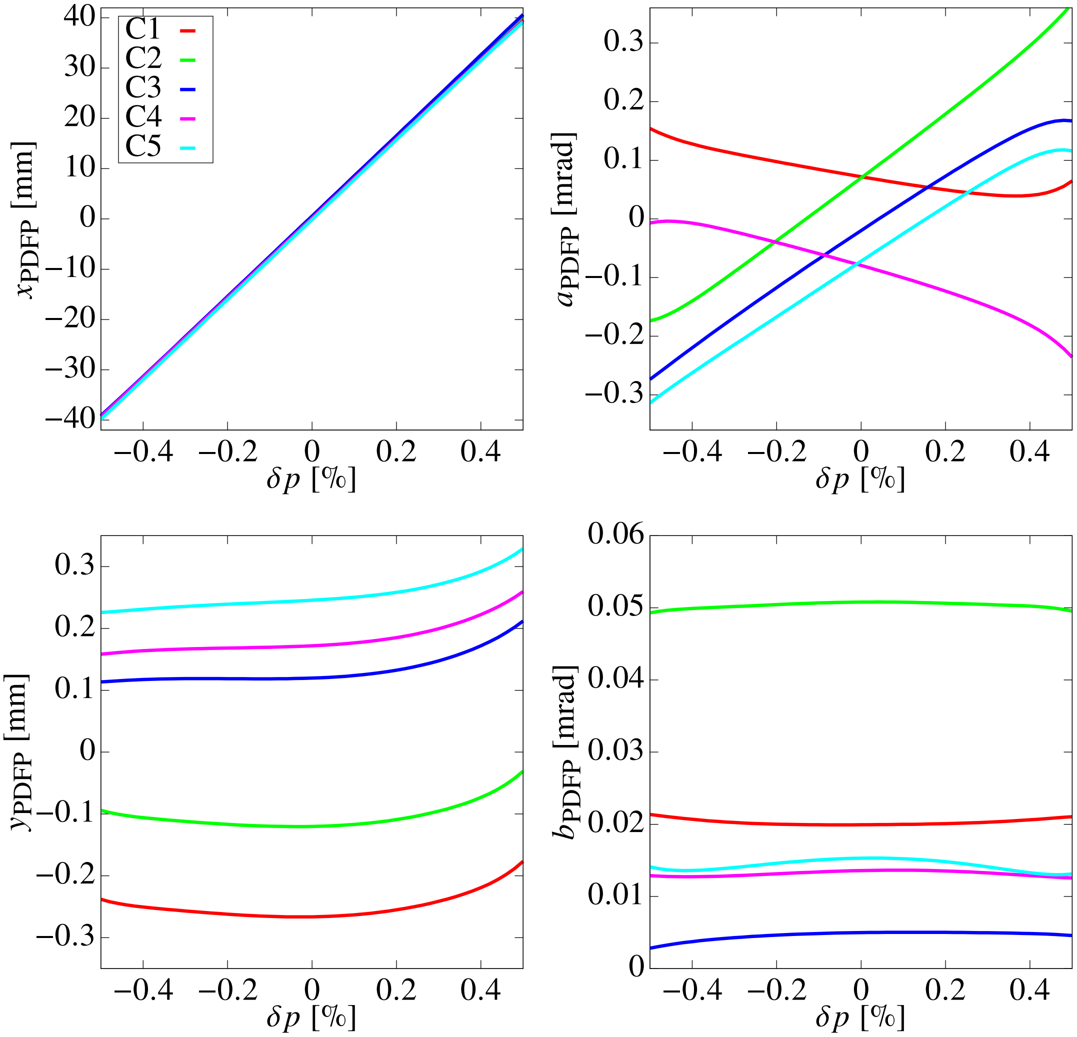

Figure 5.2: Changes of the closed orbits due to relative changes δp in the total initial momentum. The plots illustrate absolute coordinates with respect to the ideal orbit at the center of the ring for the five collimator locations (C1-C5).

with the unperturbed part of the vertical magnetic field. The Lorentz force, which determines the orbit radius, is directly proportional to the velocity of the particle, which is relativistically related to the momentum of the particle. This behavior is clearly visible in the horizontal components of Fig. 5.2. The radial position of the parameter dependent fixed point x_{PDFP} changes dominantly linearly at about 79 mm/% with the momentum offset at all collimator locations. The associated dependence of the horizontal momentum a_{PDFP} incorporates the changing radial orientation of the momentum dependent closed orbit with respect to the Poincaré surface and the different orientations at the various collimator locations.

The vertical components y_{PDFP} and b_{PDFP} of the closed orbit are mostly dependent on the azimuthal location of the map and change only slightly with a momentum offset.

5.3.3 The Relevance of Closed Orbits

The momentum dependent closed orbits correspond to fixed points in the Poincaré return maps. Particles that are not on a closed orbit oscillate around the momentum dependent closed orbit corresponding to their specific momentum offset. In the stroboscopic view of the Poincaré return maps, this corresponds to stroboscopic oscillatory behavior around the fixed point in both phase spaces in form of distorted ellipses as Fig. 5.3 indicates.

Given a particle with distorted elliptical phase space behavior and its corresponding momentum dependent fixed point $\vec{z}_{\text{PDFP}}(\delta p)$, we define the oscillation amplitudes x_{amp} and y_{amp} independently from each other. In the radial phase space $x_{\text{amp}} = |x_0 - x_{\text{PDFP}}(\delta p)|$ for $a_0 = a_{\text{PDFP}}(\delta p)$ and $y_{\text{amp}} = |y_0 - y_{\text{PDFP}}(\delta p)|$ for $b_0 = b_{\text{PDFP}}(\delta p)$ in the vertical phase space.

The oscillation amplitudes of these transverse oscillations are determined by the phase space position of the particle and the momentum dependent fixed point. Particles with the same oscillation amplitudes but different momentum offsets will follow roughly the same motion, but at different locations in phase space. On the other hand, particles at the same phase space location may follow entirely different orbital motion depending on their corresponding momentum dependent fixed point. In summary, the phase space motion of a particle is characterized by its momentum dependent fixed

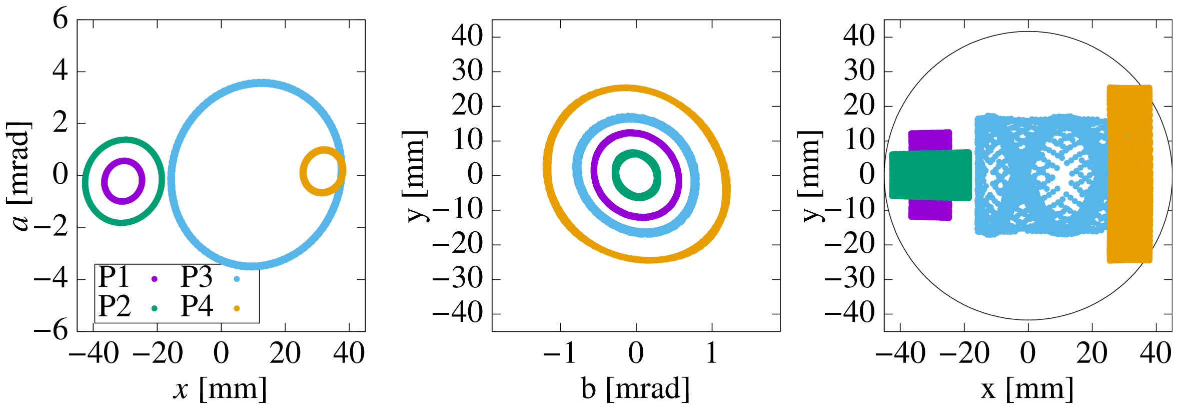

Figure 5.3: Phase space behavior of four particles in different phase space regions with various amplitudes and momentum offsets. Particle 4 (yellow) hits the collimator (circle in the xy plot) and is lost. The momentum dependent radial position x of the particles is particularly prominent. The individual particles are characterized by the parameter set $\left(x_{\text{amp}}, y_{\text{amp}}, \delta p\right)$. For particle 1 (P1) the parameter set is $(6\,\text{mm}, 12\,\text{mm}, -0.39\text{‰})$. For particle 2 (P2) the parameter set is $(12\,\text{mm}, 6\,\text{mm}, -0.39\text{‰})$. For particle 3 (P3) the parameter set is $(27\,\text{mm}, 16\,\text{mm}, +0.13\text{‰})$. For particle 4 (P4) the parameter set is $(6\,\text{mm}, 25\,\text{mm}, +0.39\text{‰})$.

point, its amplitudes of oscillation, and its oscillation frequencies, which are addressed in detail in Sec. 5.4.

The collimators restrict the maximum amplitudes of oscillation around the associated momentum dependent fixed points. Fig. 5.4 illustrates the shape of the viable phase space region around a momentum dependent fixed point. The closeness of the reference closed orbit to the collimators increases the risk of muon loss. While particles with low momentum offset are only at risk of getting lost when they have relatively large oscillation amplitudes, particles with a large momentum offset may already be lost with seemingly small amplitudes of oscillation.

Since the semi-major and semi-minor axis of the distorted elliptical phase space behavior are not necessarily aligned with the position and momentum axis and vary for each particle, there is no straightforward definition of the amplitude of oscillation. The DA normal form algorithm takes care of this by transforming the distorted ellipses in phase space to circles such that the amplitudes of oscillation are just the radii of the circles – the normal form radii. We will investigate the relationship between the original phase space coordinates and the normal form radii more closely later on and also use its advantages, but for now, we want to focus on practically relatable quantities in the original phase space, rather than abstract quantities like the normal form radius.

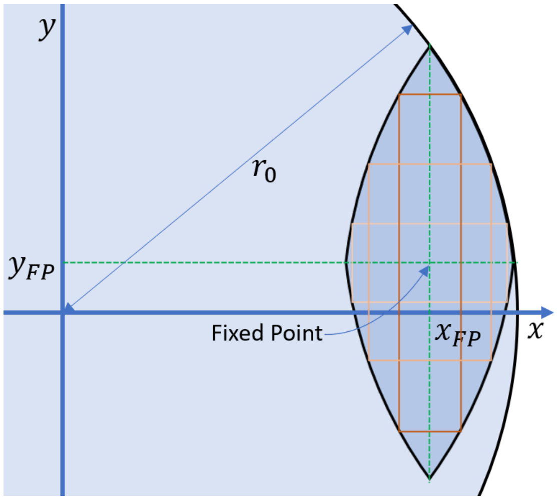

Figure 5.4: Schematic illustration of viable xy region around a momentum dependent fixed point. The region contains all rectangles centered at the fixed point, which do not overlap with the collimator circle.

5.4 Tune Analysis

The following tune analysis investigates the oscillation frequency around the reference closed orbits depending on the momentum offset and the amplitude of oscillation. The tunes shall shed light on average loss times and the involvement of resonances.

5.4.1 Tunes of the Momentum Dependent Closed Orbit

Given the parameter dependent fixed point map representing the phase space behavior around the momentum dependent closed orbit of the Muon g-2 Storage Ring model, the diagonalization in the DA normal form algorithm is used to determine the tunes of the momentum dependent closed orbit.

The calculated tunes of the closed orbit (for $\delta p = 0$) differ only very slightly depending on the azimuthal location of the Poincaré return map yielding

$$\nu_x = 0.944462633(8 \pm 3) \quad \text{and} \quad \nu_y = 0.330814444(7 \pm 6), \tag{5.3}$$

which is expected since they all describe the linear motion around same closed orbit. The proximity of the vertical tune ν_y to the low $1/3$-resonance will be investigated more closely later. The radial tune ν_x is even closer to a higher order resonance namely the $17/18$-resonance. Without loss of generality, we will use the Poincaré return map at collimator C3 for our further map investigations.

The Fig. 5.5 illustrates the momentum dependence of the tunes over the momentum offset range of $\delta p \in [-0.5\%, 0.5\%]$ and indicates the linear dependence (chromaticities) ξ_i as a reference.

For $|\delta p| < 0.25\%$ the momentum dependence of both tunes is predominantly linear with

$$\xi_x = -0.131999346 \quad \text{and} \quad \xi_y = 0.389753993. \tag{5.4}$$

For $|\delta p| > 0.33\%$ however, the tunes are dominated by an order eight dependence on relative momentum offsets δp. This eighth order dependence results from the strong ninth order terms in the original map, which are linear in the phase space components and of order eight in the momentum dependence, representing the earlier mentioned significant influence of the 20th-pole of the ESQ potential (see Sec. 5.2).

Interestingly, the linear coefficient and the eighth order coefficient of the vertical momentum dependent tune shifts are both larger by a factor of three and opposite in sign compared to their radial counterparts. Additionally, the momentum dependent vertical tune shifts away from the $1/3$-resonance.

The investigation in [96] indicated a strong influence of the ESQ voltage on the linear motion

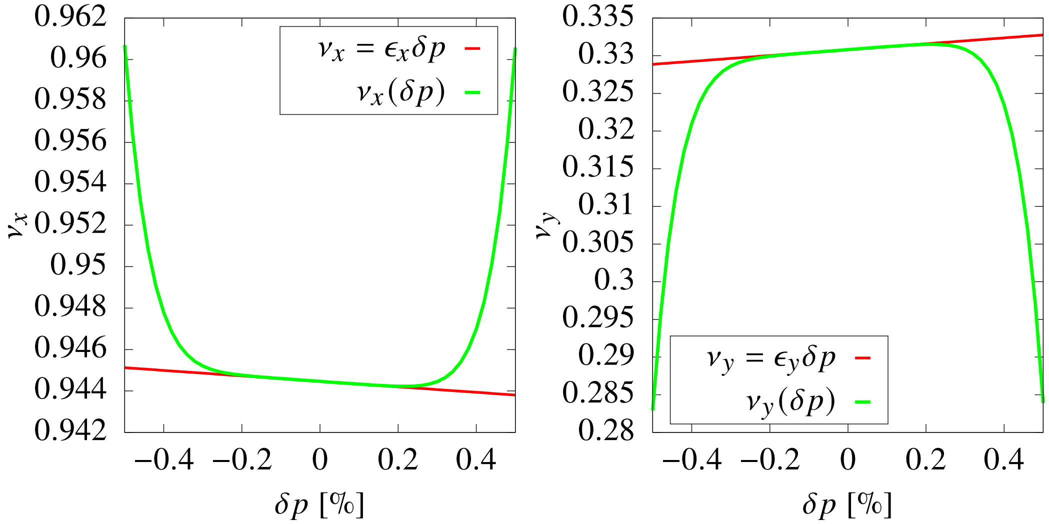

Figure 5.5: Vertical and horizontal tune dependence in the model of the Muon g-2 Storage Ring of E989 on relative offsets δp from the reference momentum p_0.

around the respective expansion points and therefore the tunes. The momentum dependence of the tunes – the momentum dependent tune shifts – however is only slightly changed by the ESQ voltage (see [96] for more details).

5.4.2 The Amplitude Dependent Tune Shifts

The DA normal form algorithm provides the transformation $\mathcal{A}_{\mathrm{NF}}$ from the original phase space coordinates (x, a) and (y, b) to rotationally invariant normal form coordinates $(q_{\mathrm{NF},1}, p_{\mathrm{NF},1})$ and $(q_{\mathrm{NF},2}, p_{\mathrm{NF},2})$. The amplitude and parameter dependent tune shifts $\nu_i(r_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \delta p)$ can be extracted from the normal form map, where the amplitudes are given by the normal form radii $r_{\mathrm{NF},i} = \sqrt{q^2_{\mathrm{NF},i} + p^2_{\mathrm{NF},i}}$.

This full description of the tunes and their dependence on phase space amplitudes and momentum offsets is extremely powerful. However, the abstract normal form radii are not as practically useful as the previously defined oscillation amplitudes x_{amp} and y_{amp} in original phase space coordinates. To address this, Fig. 5.6 illustrates the dependence of the tunes on the radial phase space amplitude x_{amp} and the dependence on the vertical phase space amplitude y_{amp}, separately. This is done by calculating the corresponding normal form coordinates and normal form radii and using those for the tune evaluation.

The amplitude dependence is never linear but always appears as even orders. Investigations in [96] indicated that amplitude dependent tune shifts, just like momentum dependent tune shifts, are only weakly influenced by the ESQ voltages and the field perturbations. Similar to the purely momentum dependent tune shifts, the sign of the momentum offset seems to only play a minor role compared to the magnitude of the offset.

The radial amplitude dependence of the tunes is relatively well behaved. Again, there is the dominating eighth order dependence related to the strong ninth order nonlinear terms resulting from the 20th-pole of the ESQ potential, which shifts the tunes of the radial phase space up and tunes of the vertical phase space down with increasing radial amplitude and magnitude of the momentum offset.

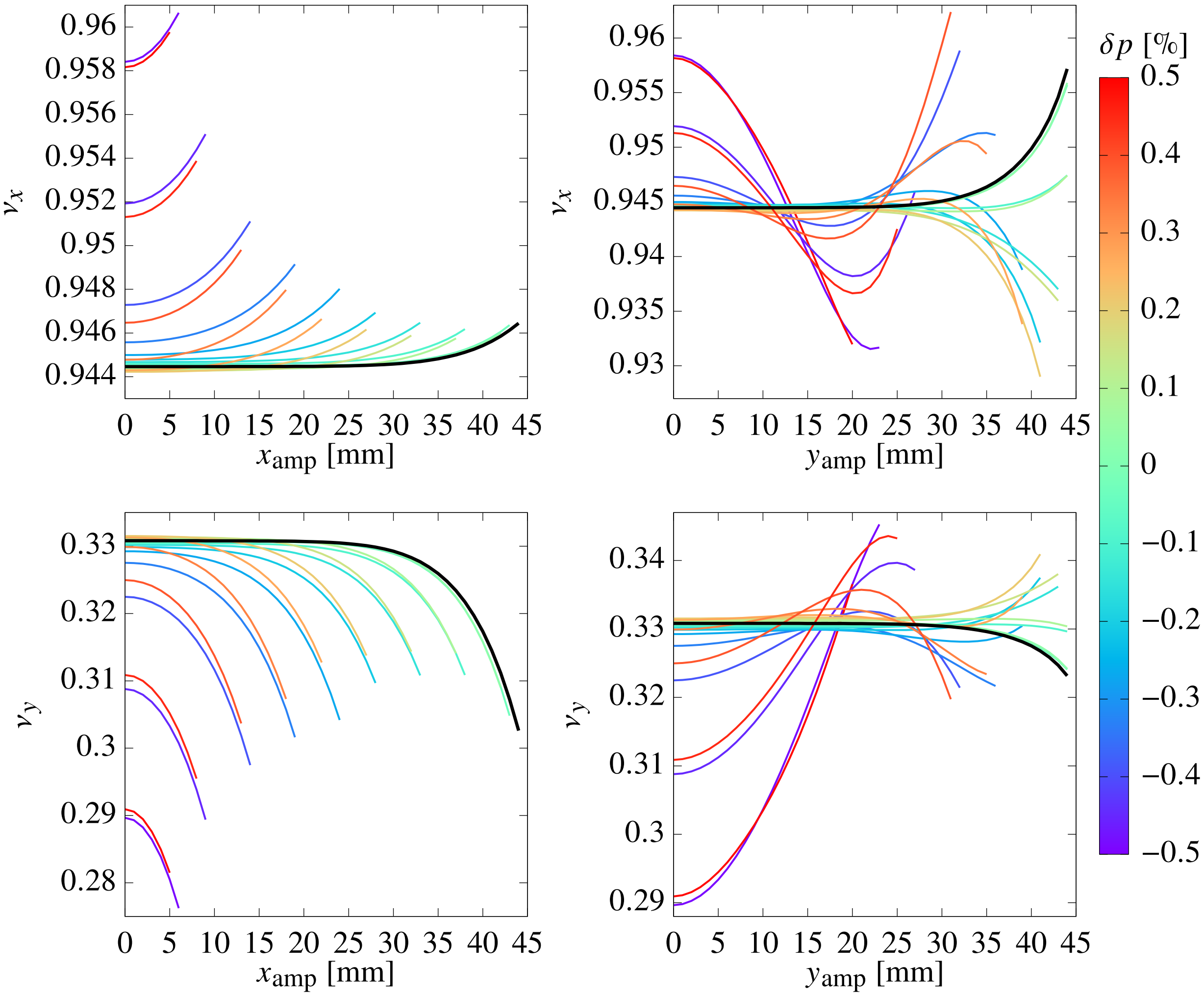

Figure 5.6: Amplitude dependent tune shifts in the model of the Muon g-2 Storage Ring of E989. The black line indicates the amplitude dependent tune shifts for $\delta p = 0$, while the other lines have a momentum offset specified by their color. For the left plots regarding the radial amplitude dependence, the vertical amplitude relative to the momentum dependent fixed point is set to zero and vice versa for the plots regarding the vertical amplitude dependence on the right. The lines end when the total xy amplitude of the particle relative to the ideal orbit reaches the collimator at $r_0 = 45$ mm.

The vertical amplitude dependence however is more complex as it varies strongly with the magnitude of the momentum offset. Regarding the vertical tune, this is particularly critical due to the crossing of the 1/3-resonance tune for some vertical amplitude and momentum offset combinations. Such low resonances can have a major influence on the dynamics of particles which is why we will closely investigate these cases later.

Even though the purely momentum dependent tune shifts ($x_{\text{amp}} = 0$, $y_{\text{amp}} = 0$) and the tune shifts purely dependent on the vertical amplitude ($x_{\mathrm{amp}} = 0, \delta p = 0$) shift in the same direction – up for radial tunes and down for vertical tunes – there are opposing cross-terms, which depend both on the vertical amplitude and the momentum offset that trigger this nontrivial tune shift behavior.

In Fig. 5.7 to Fig. 5.9 the combined effects of simultaneous radial and vertical amplitudes on the tune shifts are illustrated for selected momentum offsets. The behavior for the intermediate momentum offsets may be interpolated from the given plots. Again, the sign of the momentum offset has only a minor influence on the form of the tune shifts compared to its magnitude.

Note that Fig. 5.7 to Fig. 5.9 only illustrates tunes for phase space states within the viable phase space around the corresponding momentum dependent fixed point. Accordingly, not all lines extend over the full 45 mm range of y_{amp} and some lines for large x_{amp} are not shown, since their total xy amplitude of the particle relative to the ideal orbit reaches the collimator at $r_0 = 45$ mm.

The combined effects in Fig. 5.7 to Fig. 5.9 emphasize the strong nonlinear character of the tune dependencies, which was already indicated in Fig. 5.6. The wave-like structure illustrates how different order terms dominate at different vertical amplitudes y_{amp} depending on both, the radial amplitude x_{amp} and the momentum offset δp. Additionally, for almost every momentum offset there are combinations of oscillation amplitudes for which the vertical $1/3$-resonance tune is crossed. Investigations in [96] did not show this strong nonlinear behavior of the combined effects on the tune shifts in such clarity.

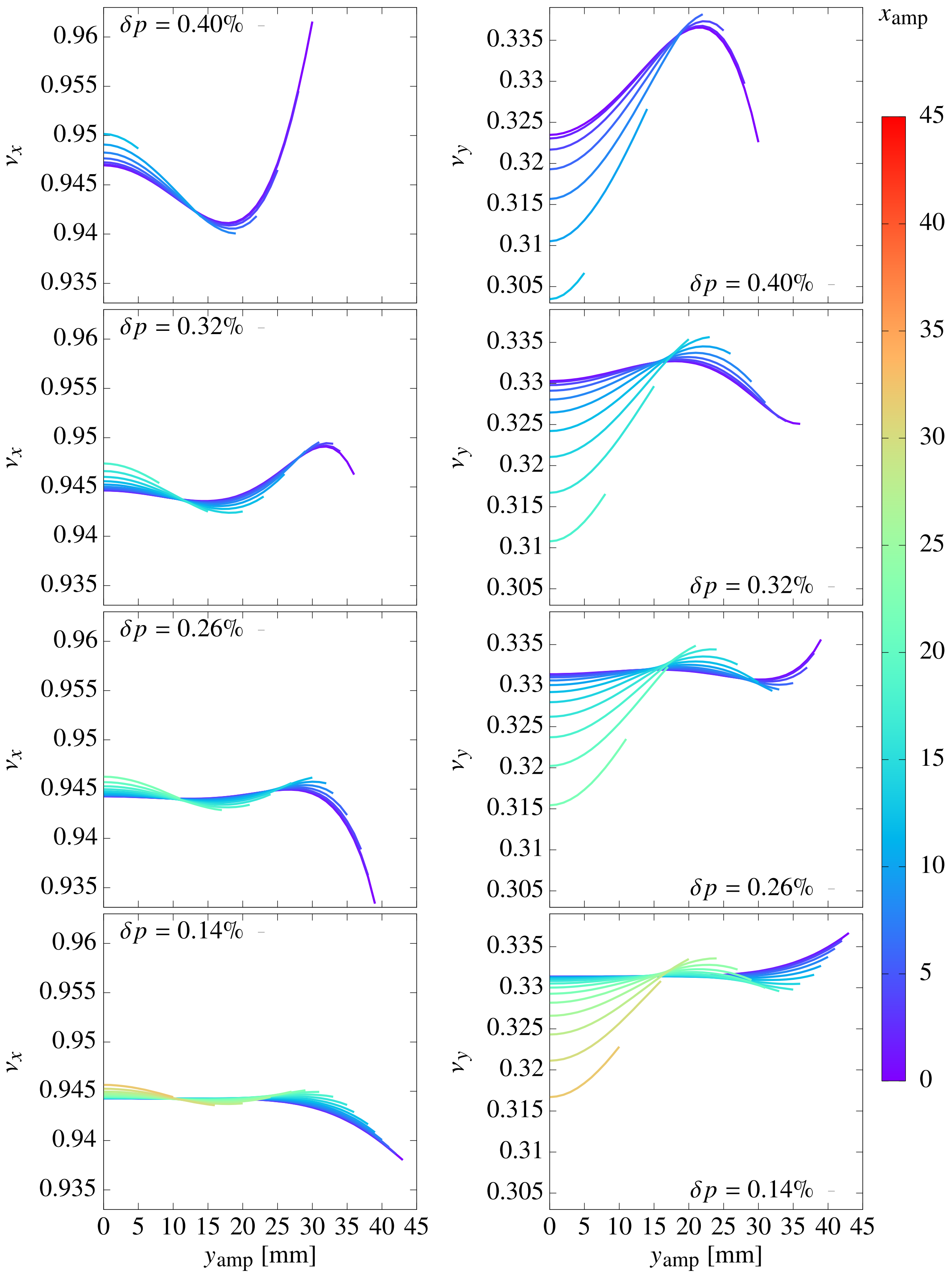

Figure 5.7: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV.

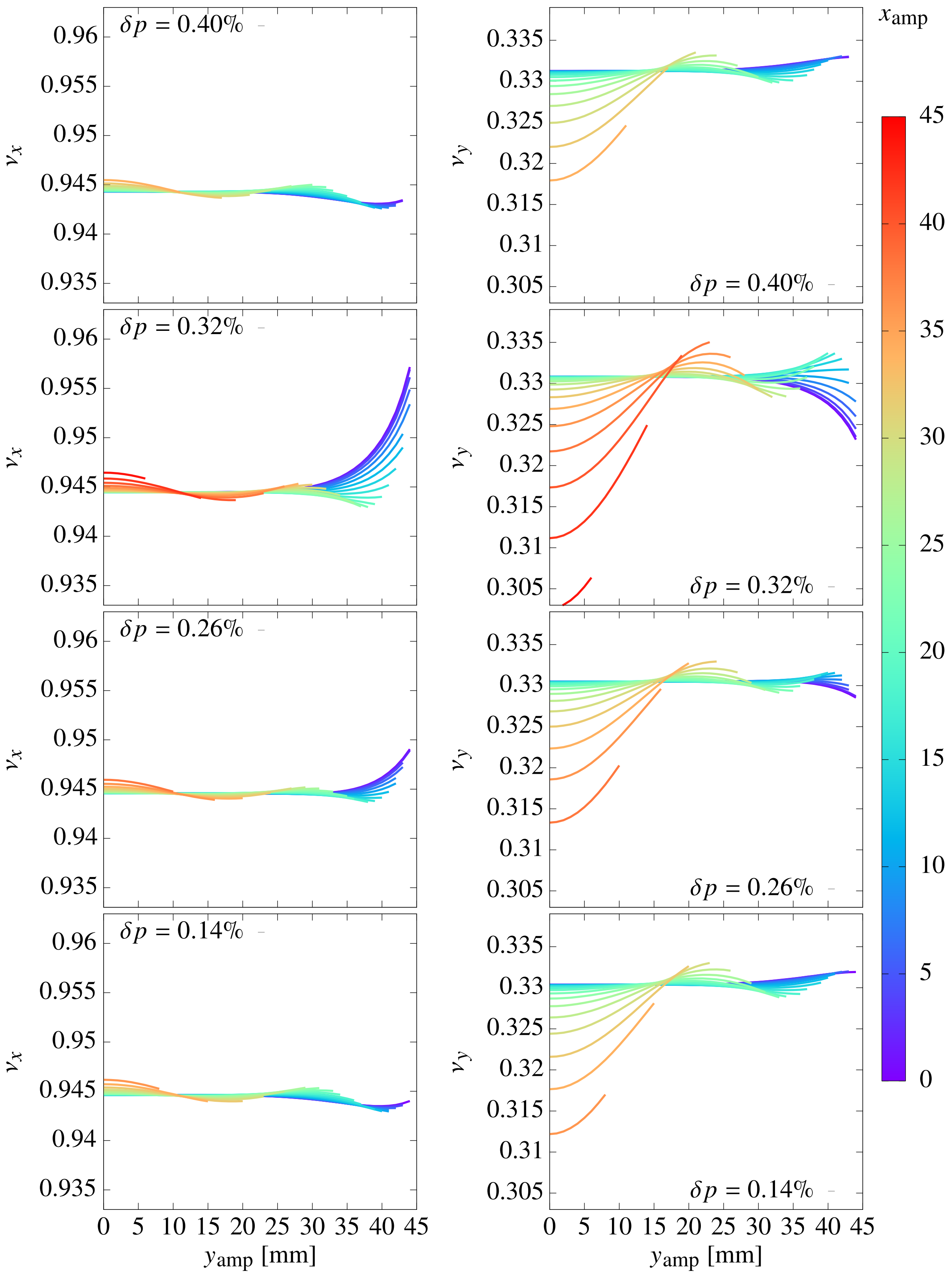

Figure 5.8: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV.

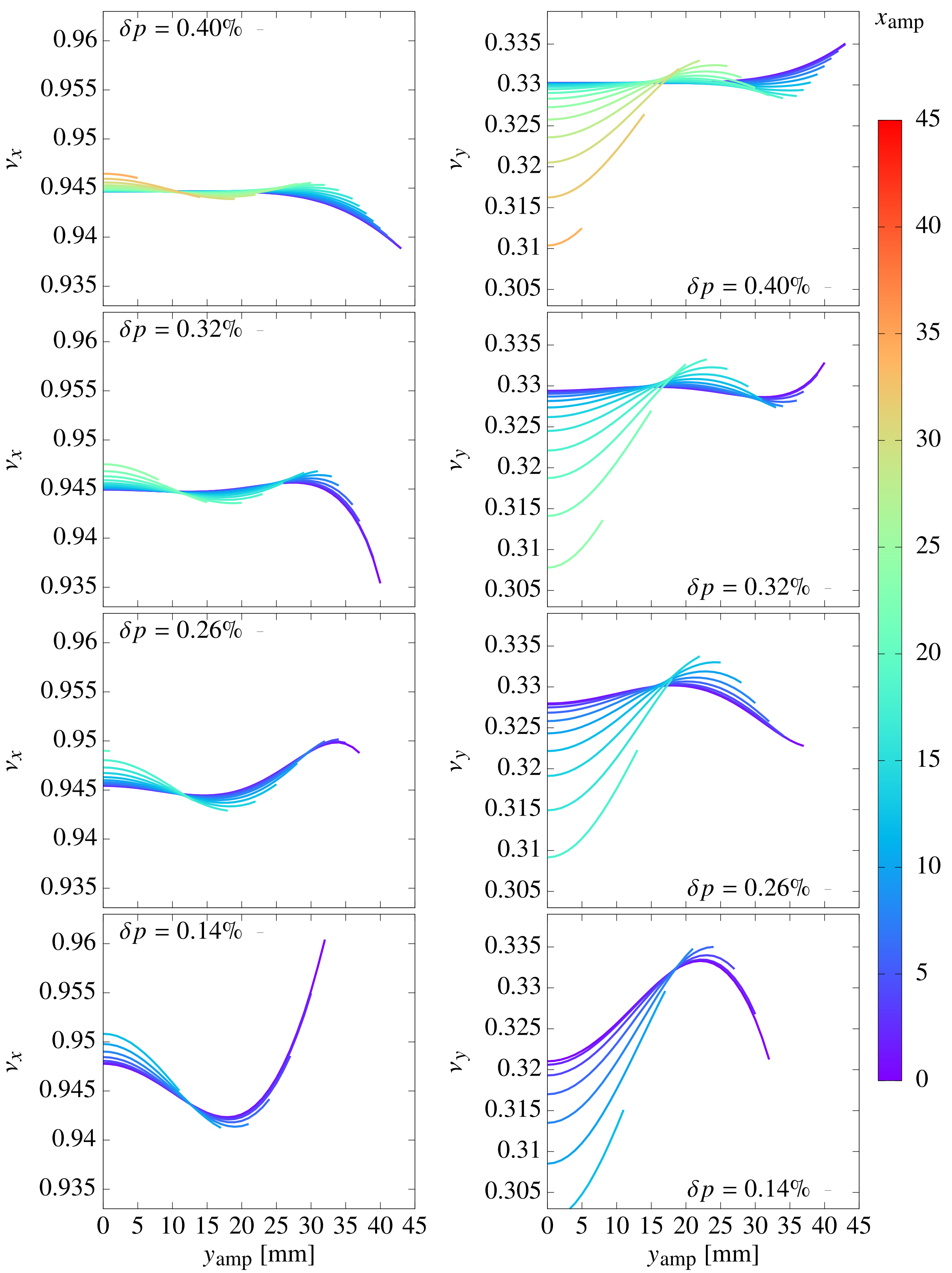

Figure 5.9: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 18.3 kV.

5.4.3 The Tune Footprint

The tune footprint visualizes the projection of a beam distribution into tune space. The COSY INFINITY based model [85, 86] of the Muon g-2 Storage Ring is used to generate the realistic beam distribution of 37738 particles from orbit tracking of the muon beam until it is circulating in the storage ring, prepared for data analysis. In particular, the model considers the imperfect injection process, which attempts to align the injected beam with the ideal orbit of the storage ring as well as possible. The model also considers the mispowered ESQ components to imitate the preparation mechanism during the first turns of the beam in the storage ring at E989. Further details of the tracking model and on how a realistic distribution of particles is obtained are elaborated in [85, 86].

The variables $(x, a, y, b, \delta p)$ relative to the ideal orbit are illustrated in Fig. 5.10 as projections into the (x, a), (y, b), and (x, y) subspaces.

The beam distribution tends towards higher total momenta in the range of $\delta p \in [-0.2\%, 0.4\%]$ while overall staying well within the momentum acceptance range of ±0.5%. The spread of the vertical momentum component b is about a factor two to three smaller than its horizontal counterpart a. The position space (xy) is filled up to the limitations due to the collimators.

The distributions of the horizontal and vertical tunes are illustrated by the tune footprint in Fig. 5.11, where the vertical tunes of the particle distribution are plotted against their horizontal tunes as previously done in [56]. The tune footprint of the tenth order calculation is overlaid by the result of an eighth order calculation to emphasize the influence of the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential. The tune footprint of the tenth order calculation is five to six times larger in each dimension than its eighth order counterpart.

Additionally, particles in different momentum offset ranges are highlighted to illustrate the behavior of this specific group. The tune footprint can be segmented into three groups characterized by their momentum offset which generates a tune footprint in the shape of a ‘T’.

The tune footprint for the other ESQ voltages in [96] has a similar distribution for the order eight and order ten calculations, respectively. While the reference tunes are mainly determined by the ESQ voltage, the relative tune shifts behave very similarly. If the ESQ voltage were to place the

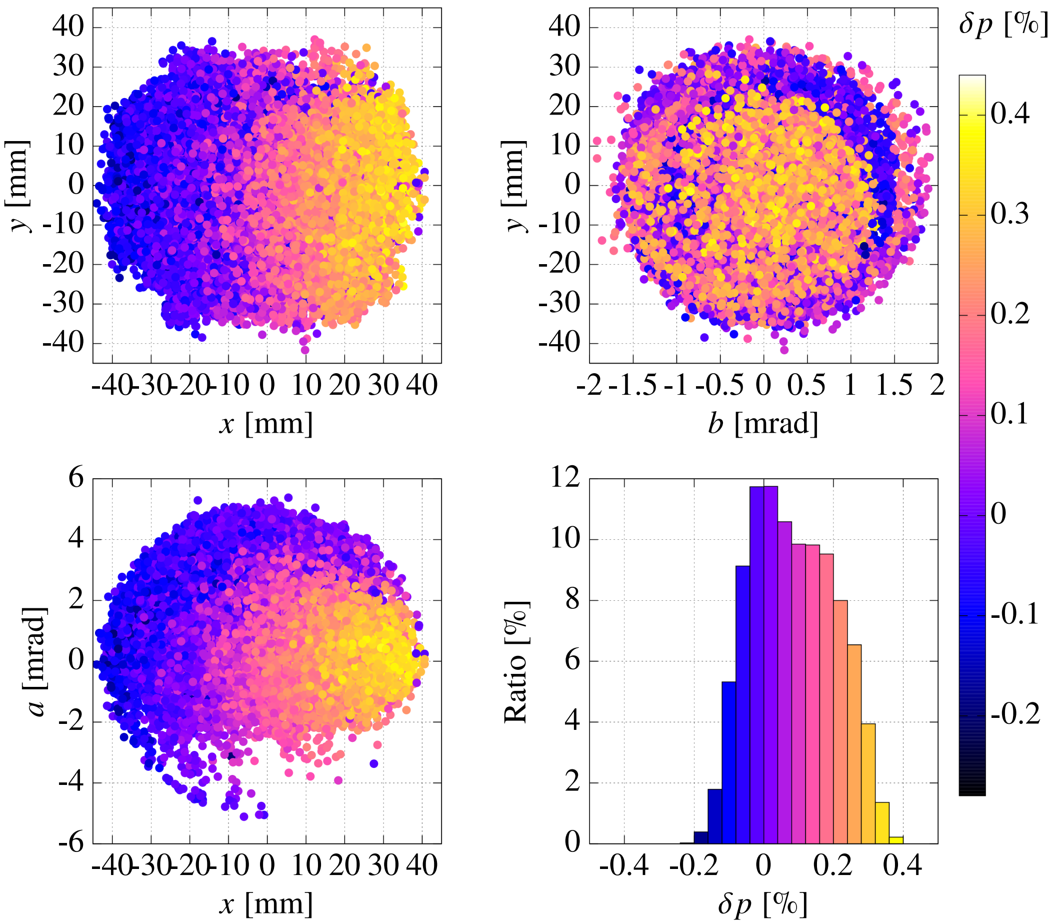

Figure 5.10: Projections of the distribution of the variables $(x, a, y, b, \delta p)$ in the realistic beam simulation at the azimuthal ring location of the central kicker.

reference tunes very close to a resonance line, we expect the tune distribution and tune shifts to behave differently.

Fig. 5.11 shows that the vertical 1/3-resonance tune cannot only be reached hypothetically for the apparent case of a nominal set-point away from resonances. A substantial part of particles is close to or on this low order resonance. The overlaid eighth order calculation shows that this is triggered by the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential. The segmentation with regard to the momentum offset of the particles into subgroups additionally shows that the vertical 1/3-resonance tune is crossed in each of those groups. The resonance point (17/18, 1/3) is also covered and surrounded by many particles and might have a

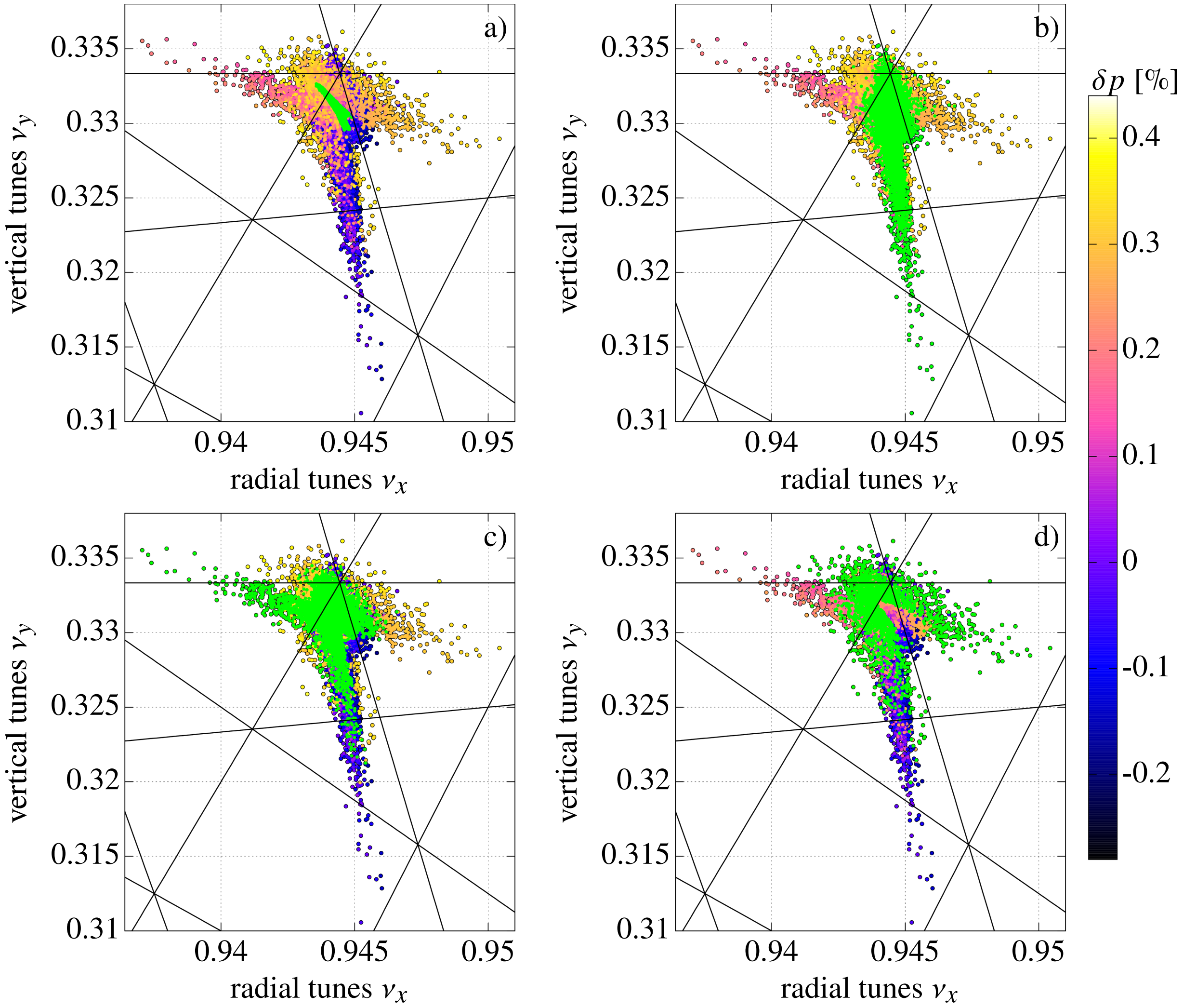

Figure 5.11: The tune footprint of a realistic beam distribution at the azimuthal ring location of the central kicker. The tune footprint from the 10th order calculation is colored according to the momentum offset of the individual particles. The black lines correspond to resonance conditions. In a) the 8th order calculation (green) is overlaid to illustrate the drastic influence of the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential. In b) the particles with a momentum offset $-0.3\% < \delta p < 0.1\%$ are overlaid in green. In c) the particles with a momentum offset $0.1\% < \delta p < 0.28\%$ are overlaid in green. In d) the particles with a momentum offset $0.28\% < \delta p < 0.5\%$ are overlaid in green.

particularly strong impact.

5.5 Stability and Loss Mechanisms

Muons are lost when they hit structural parts of the storage ring and lose the energy necessary to remain within the storage region during data taking. Collimators, which are inserted at various azimuthal locations in the ring (see Fig. 5.1), constitute the narrowest part around the storage region. They restrict the muons to amplitudes of $r = \sqrt{x^2 + y^2} < 45$ mm $= r_0$ relative to the ideal orbit.

Our previous analysis is very helpful for gaining a general understanding of certain properties of the system, e.g. the momentum dependence of the reference orbit, and the momentum and amplitude dependent shifts in the oscillation frequency of orbits around their repetitive reference orbit. This analysis showed that the vertical $1/3$-resonance tune is relevant for various combinations of amplitudes and momentum offsets. However, only tracking analysis can yield the actual phase space behavior of lost particles and particles involved with the vertical $1/3$-resonance tune. Additionally, we saw that the radial tune is very close to the high order $17/18$-resonance, which we will also look at more closely.

For the tracking analysis, we use both one-turn maps as well as sectional maps. The one-turn Poincaré return map yields the state of a muon at the azimuthal location of the central kicker depending on its state in the previous turn. Sectional maps transfer the state of a muon to the azimuthal location of the collimators. Accordingly, the muons are not tracked continuously, but stroboscopically at specific azimuthal locations e.g. at the respective collimator locations.

There are two common approaches for tracking analysis. For a general understanding of the phase space dynamics of the storage ring, one could track a particle distribution, which is evenly distributed in all phase space dimensions and over momentum offset range. However, the implication from such an analysis for the actual muon beam might be limited, since the actual muon beam is not evenly distributed. Accordingly, we track the realistic particle distribution of 37738 particles from Sec. 5.4.3.

We take the distribution of particles at the central kicker after 200 turns (corresponding to about 30 μs) from the injection, which is when data taking begins. During this initial 30 μs after injection, the muon beam and the system are conditioned for data taking [2]. The beam is tracked for additional 4500 turns (670 μs), while determining and recording various orbit parameters that shall be analyzed in detail below.

5.5.1 The Normal Form Defect of Tracked Particles

As explained in Sec. 2.4, the normal form defect yields the inaccuracies in the normal form, i.e., how much the pseudo-invariants (the normal form radii) vary per turn. Using tracking simulations, one can evaluate a related quantity that we will call the long term normal form defect

$$d_{\mathrm{NF,lt}} = r_{\mathrm{NF,max}} - r_{\mathrm{NF,min}}. \tag{5.5}$$

It yields the difference between the maximum and the minimum normal form radius of a single particle orbit over the many turns of the long term tracking. Thus, provides the normal form radius range of the orbital pattern. The maximum per turn normal form defect $d_{\mathrm{NF,max}}$ of the particle is the maximum rate of change of the normal form radius during the orbital pattern.

In Fig. 5.12 the particles are grouped by the maximum per turn normal form defect they encountered during the 4500 turns of tracking. The rate of particles getting lost is strongly correlated with the size of the maximum normal form defect they encountered. More than 91% of particles

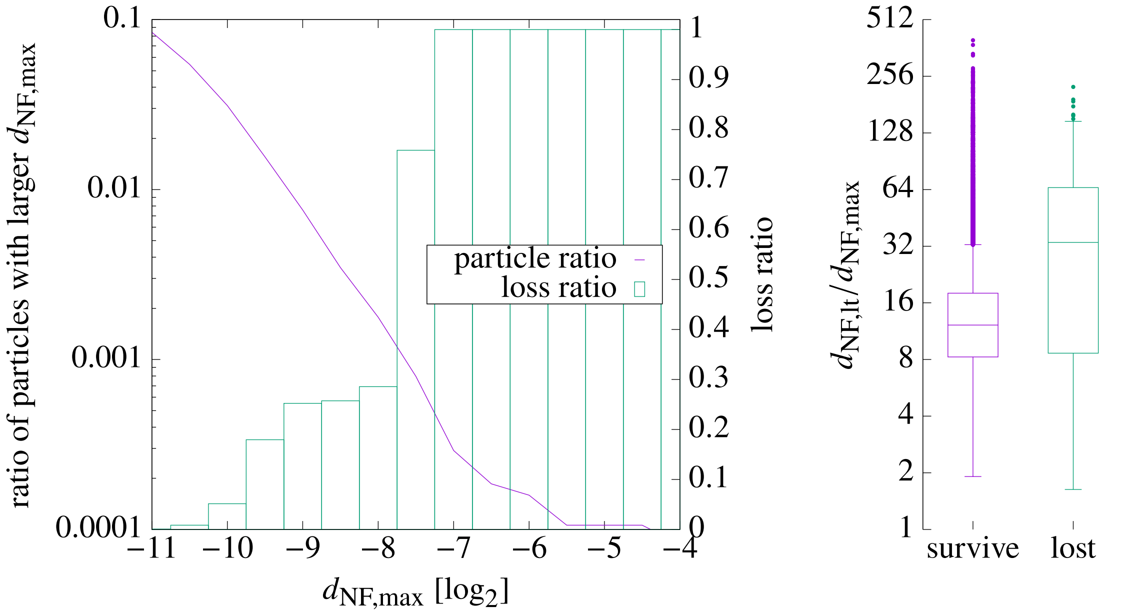

Figure 5.12: The left plot shows in violet the ratio of particles that have a larger $d_{\mathrm{NF,max}}$ then the corresponding x value. The green boxes indicate the ratio of particles lost in each $d_{\mathrm{NF,max}}$ group, e.g., particles that encounter a maximum normal form defect larger than $2^{-7} = 7.8 \cdot 10^{-3}$ are all lost (loss ratio of 1). The right plot shows the ratio between the normal form radius range over 4500 turns ($d_{\mathrm{NF,lt}}$) and the maximum encountered normal form defect $d_{\mathrm{NF,max}}$ for each particle. The particles are grouped into surviving and lost particles.

have a maximum normal form defect smaller than $2^{-11} = 4.9 \cdot 10^{-4}$ and none of those particles is lost. With larger maximum normal form defects, the loss rate increases significantly. Particles that encounter a maximum normal form defect larger than $2^{-7} = 7.8 \cdot 10^{-3}$ are all lost. This confirms that the size of the per turn normal form defect is a good indication for losses, in the sense that the larger the per turn normal form defect the more likely the particle gets lost.

The right side of Fig. 5.12 illustrates the ratio of the long term normal form defect $d_{\mathrm{NF,lt}}$ to the maximum per turn normal form defect $d_{\mathrm{NF,max}}$ of a particle. Considering that $d_{\mathrm{NF,lt}}$ over 4500 turns is only a factor of eight to 16 larger than $d_{\mathrm{NF,max}}$ for surviving particles illustrates the overestimation of Nekhoroshev-type stability estimates (see Sec. 2.4) based on the per turn normal form defect. Additionally, the ratio is much more shifted to higher factors for lost particles, indicating less overestimation for lost particles with Nekhoroshev-type stability estimates.

In Fig. 5.13 the relevance of the resonances – especially low order resonances like the vertical 1/3-resonance tune – on the long term normal form defect becomes obvious. Since the tunes are dependent on the normal form radii, a larger long term normal form defect automatically corresponds to a larger tune range of a particle.

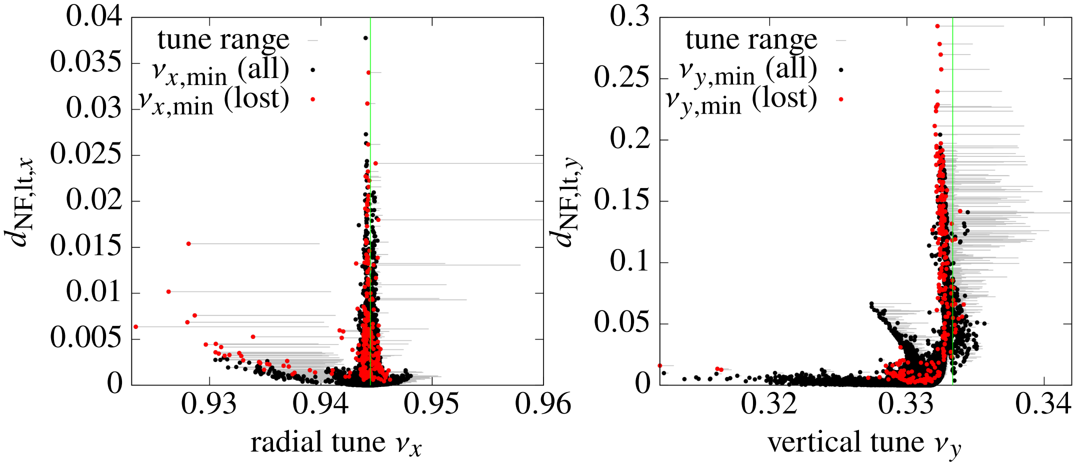

Figure 5.13: The plots show the long term normal form defect dependent on the calculated tune range of each particle. The dots are the minimum calculated tune of each particle while tracking. Red dots indicate that the respective particle is lost over the 4500 tracking turns. The gray lines show the calculated tune range of each particle. The left plot illustrates the radial long term normal form defect with respect to the radial tune and the 17/18 resonance (green line). The right plot shows the vertical long normal form defect with respect to the vertical tune and the 1/3 resonance (green line).

In the plot of the vertical tune against the vertical long term normal form defect, there is a 'spike' facing roughly 45° away from the resonance line. In Fig. 5.14, the tune range of these 'spike' particles is analyzed to determine a resonance as a potential trigger of the increasing normal form defect. The analysis indicates that the 10th order $6\nu_x + 4\nu_y = 7$ resonance might be the cause of this spike, but it remains unknown why the normal form defect increases along this resonance with increasing distance from the $1/3$-resonance.

The normal form radii are the oscillation amplitudes in the high order normalized, linearly decoupled phase space. They are closely related to the oscillation amplitudes in the respective phase spaces relative to the momentum dependent closed orbit. The strong variation in the normal form radii (the large long term normal form defect) of some orbits indicates that the corresponding oscillation amplitude of those orbits around their respective reference orbits is also not constant. To investigate this more closely, the following section investigates the orbits of all lost particles.

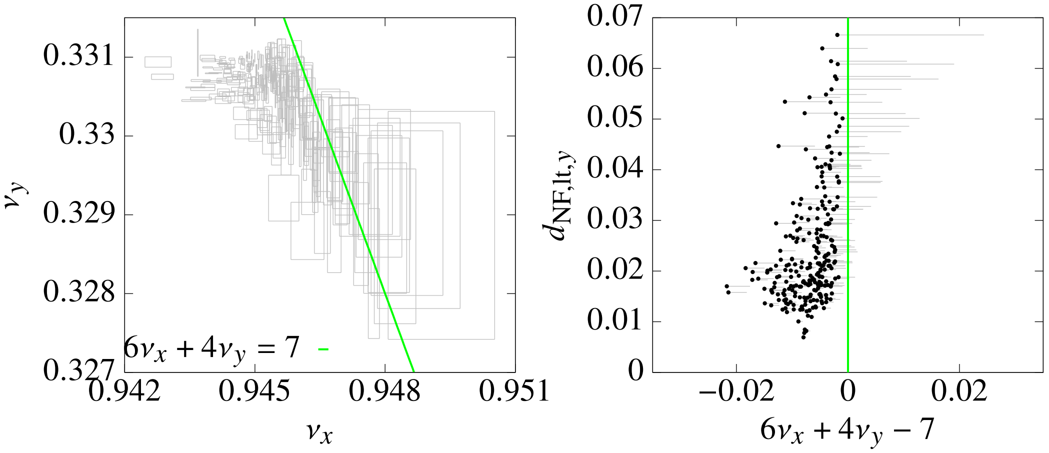

Figure 5.14: The tune range of the particles forming the spike in Fig. 5.13 are shown on the left. The right plot shows the normal form defect of the particles depends on their closeness to the $6\nu_x + 4\nu_y = 7$ resonance (green line).

5.5.2 Lost Muon Studies

In this section, we track and investigate all 259 muons of the distribution from Sec. 5.4.3 that are lost at collimator C3 and/or C4 over the 4500 turns. In Fig. 5.15 to Fig. 5.29, 15 of those lost particles are picked to illustrate the different phase space behaviors observed for lost particles. Each figure illustrates the behavior of a different particle and is made up of six plots. The scaling of the plots is the same for all figures. In each figure, the left two plots show the radial and vertical phase space behavior. The xy behavior is shown in the top center plot above the normal form phase space behavior $(q_{\mathrm{NF},2}, p_{\mathrm{NF},2})$. The top right plot shows the normal form radius $r_{\mathrm{NF}} = \sqrt{r^2_{\mathrm{NF},1} + r^2_{\mathrm{NF},2}}$ over the number of turns, and the bottom right plot shows the tune footprint of the particle. In the caption, the momentum offset of the particle is mentioned.

One striking property that many of the lost muons share is the appearance of threefold-symmetry patterns in the vertical phase space projections. The calculated tunes of these lost particles are all crossing or proceed very close to the vertical $1/3$-resonance. These threefold-symmetry patterns often include significant modulations in the vertical oscillation amplitude, which is additionally shown by the changing overall normal form radius r_{NF} and the variations in the calculated tunes shown in the tune footprint. While there are many patterns, there are two that stick out, namely, the island pattern (see for example Fig. 5.17) and the shuriken pattern (see for example Fig. 5.23). In Sec. 5.5.3, we will understand how all these patterns are related to period-3 fixed point structures.

The patterns come in stable, semi-stable, and unstable forms. This tendency to unstable behavior is often associated with a large radial amplitude and/or a closeness to the $(\nu_x, \nu_y) = (17/18, 1/3)$ resonance point. The 'fuzzyness' of the vertical phase space pattern in (y, b) compared to the pattern in the corresponding normal form phase space $(q_{\mathrm{NF},2}, p_{\mathrm{NF},2})$ is related to the radial phase space motion. Due to the weak coupling between the radial and vertical phase space from the imperfections in the magnetic field, large amplitudes in (x, a) notably affect the motion in (y, b), which does not happen in the decoupled normal form phase space.

This 'fuzzyness' might also trigger the jumping between different patterns for orbits, which are close to the border between two patterns (see Fig. 5.31). While studying the figures below, pay attention to the modulation frequency of the vertical amplitude / the normal form radius. Shuriken and unstable patterns yield the slowest modulations followed by large and small island patterns. The fastest modulations occur in almost regularly looking patterns in form of distorted ellipses. There also seems to be a correlation between the size of the modulation and its frequency, i.e., the larger the modulation the slower its frequency.

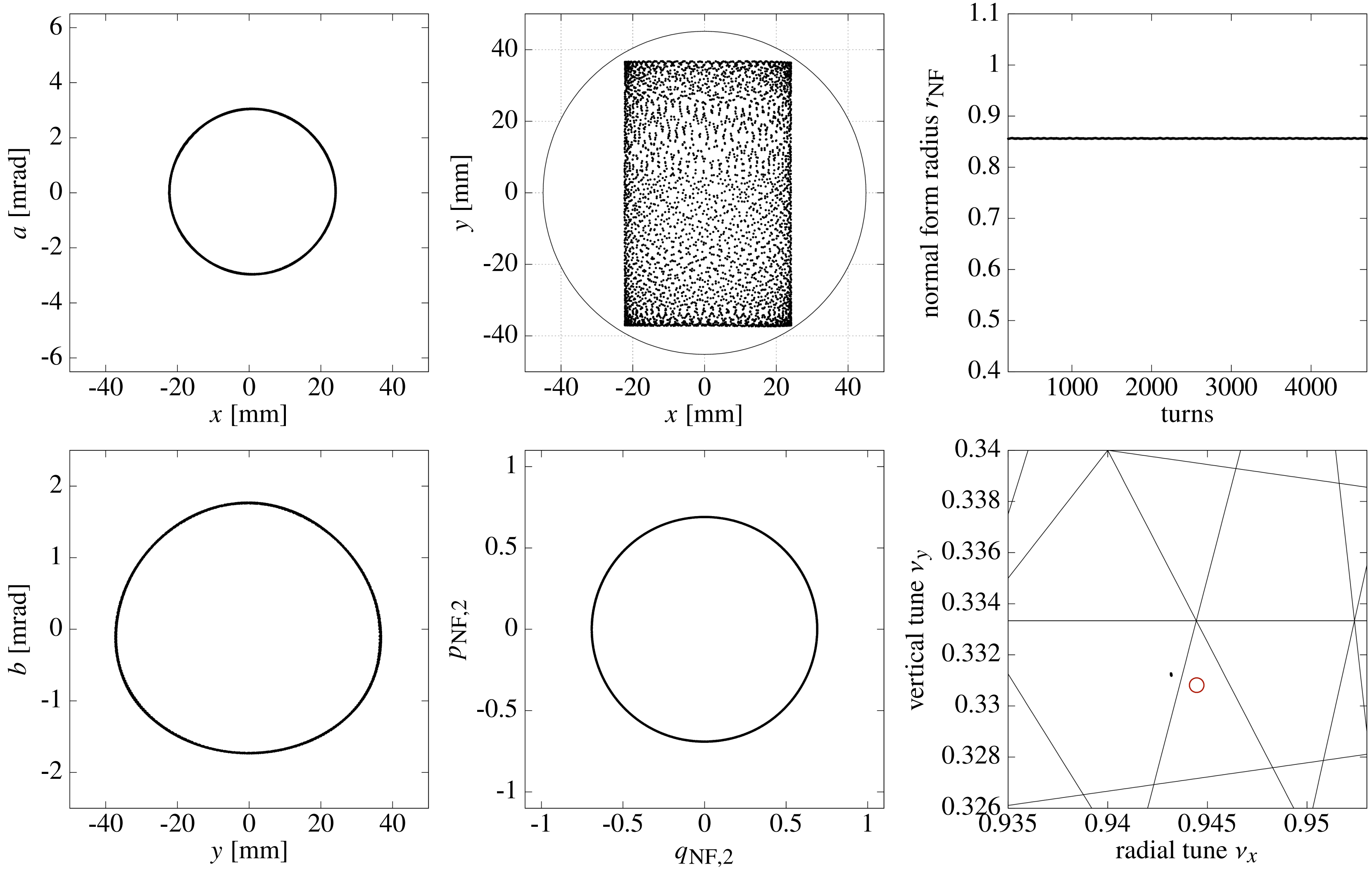

Figure 5.15: The radial and vertical phase space behavior indicates that this particle (δp = 0.015%) oscillates at constant amplitudes around its momentum dependent reference orbit. The overall normal form radius is constant and confirms this. Accordingly, the tune footprint of the particle is a single dot. This is a trivial large amplitude loss.

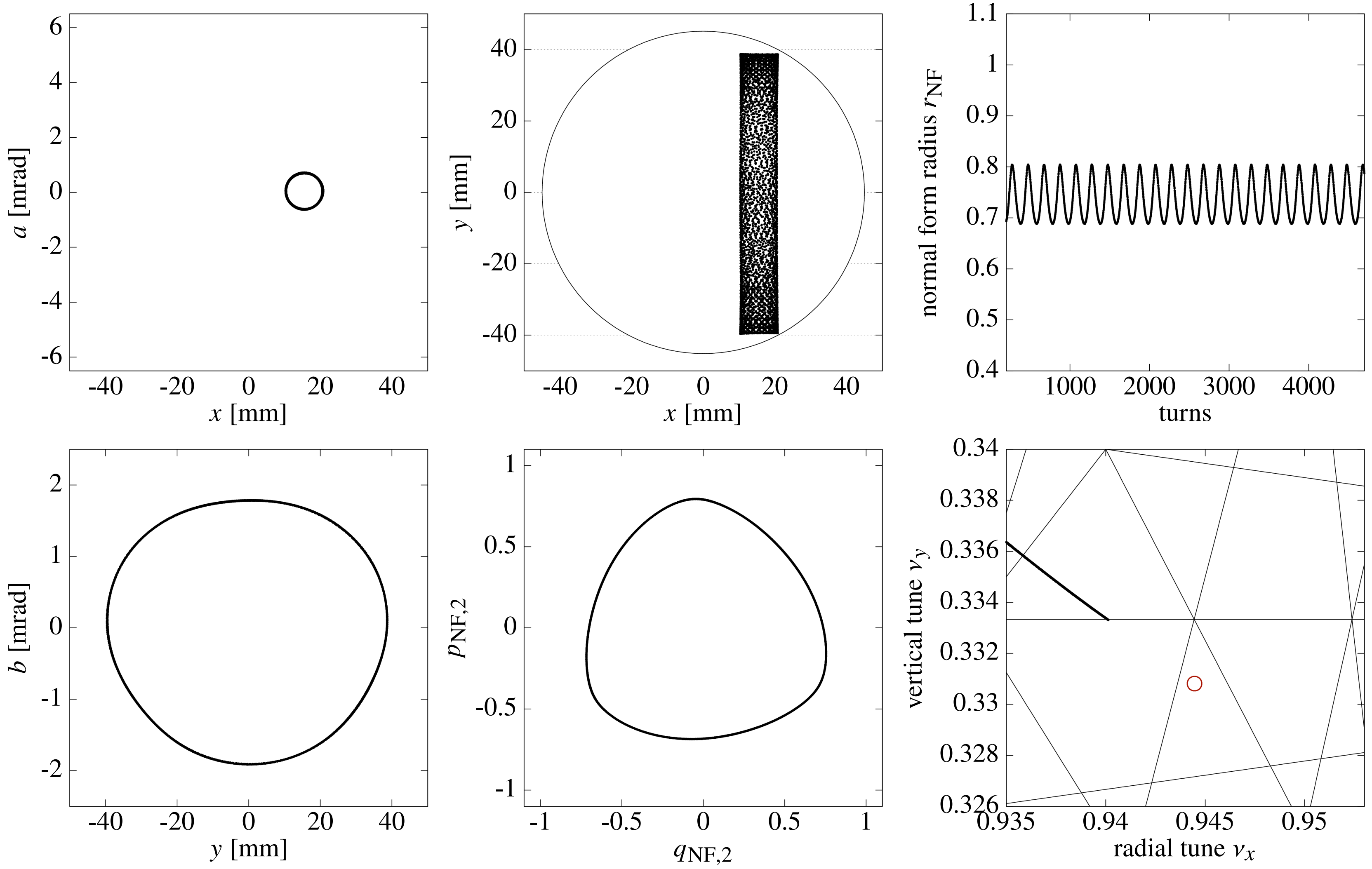

Figure 5.16: The vertical phase space behavior of this particle ($\delta p = 0.196\%$) has a slight triangular deformation. The overall normal form radius indicates a modulated amplitude and the spread out tune footprint starts right after the vertical 1/3-resonance line. Despite slight influence of the resonance, the rather elliptical phase space behavior makes this a trivial large amplitude loss.

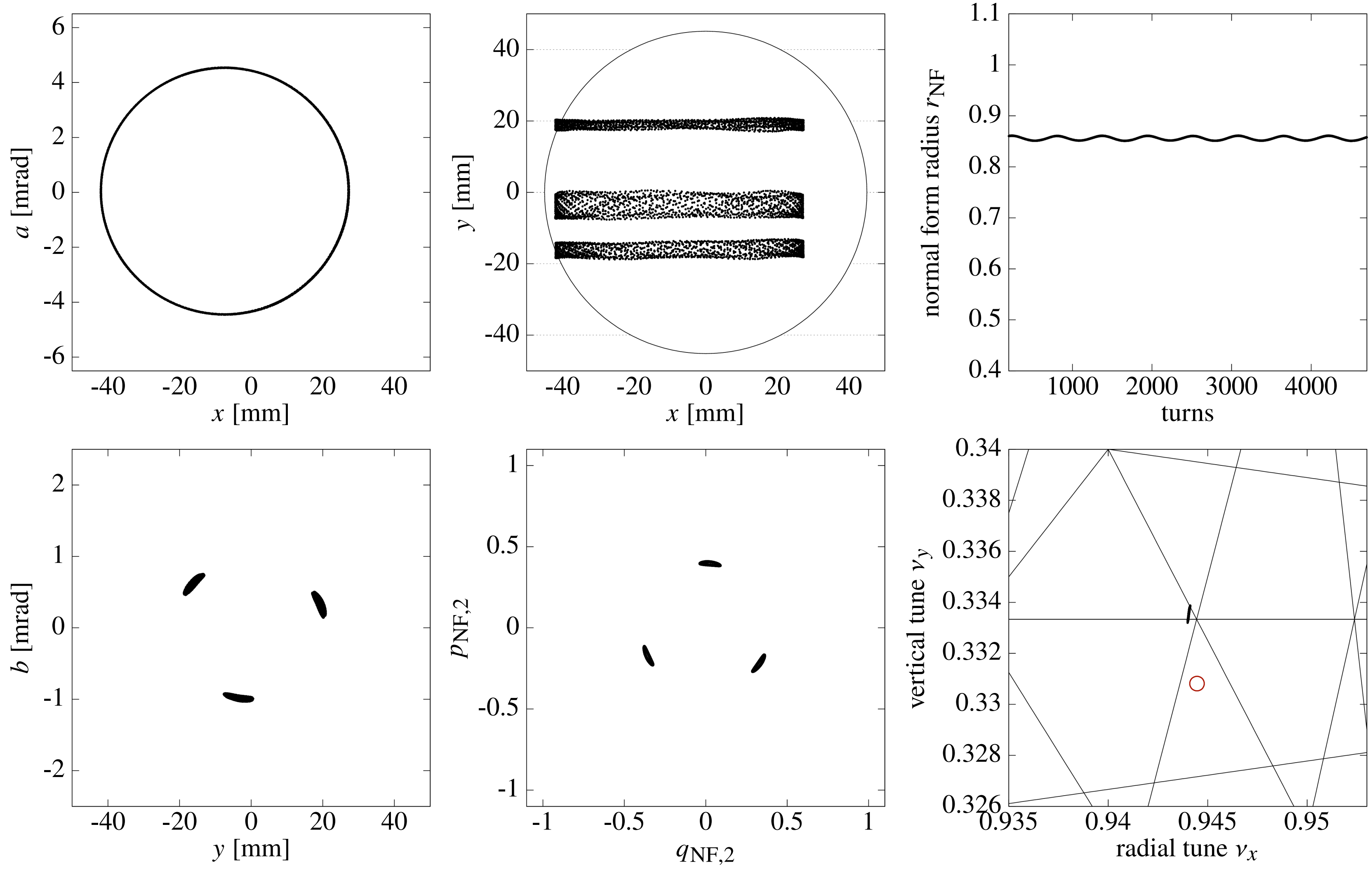

Figure 5.17: This particle ($\delta p = -0.088\%$) is caught around a period-3 fixed point structure in the vertical phase space, which is related to the vertical 1/3-resonance. We refer to these structures as islands and the loss mechanisms is called island related loss.

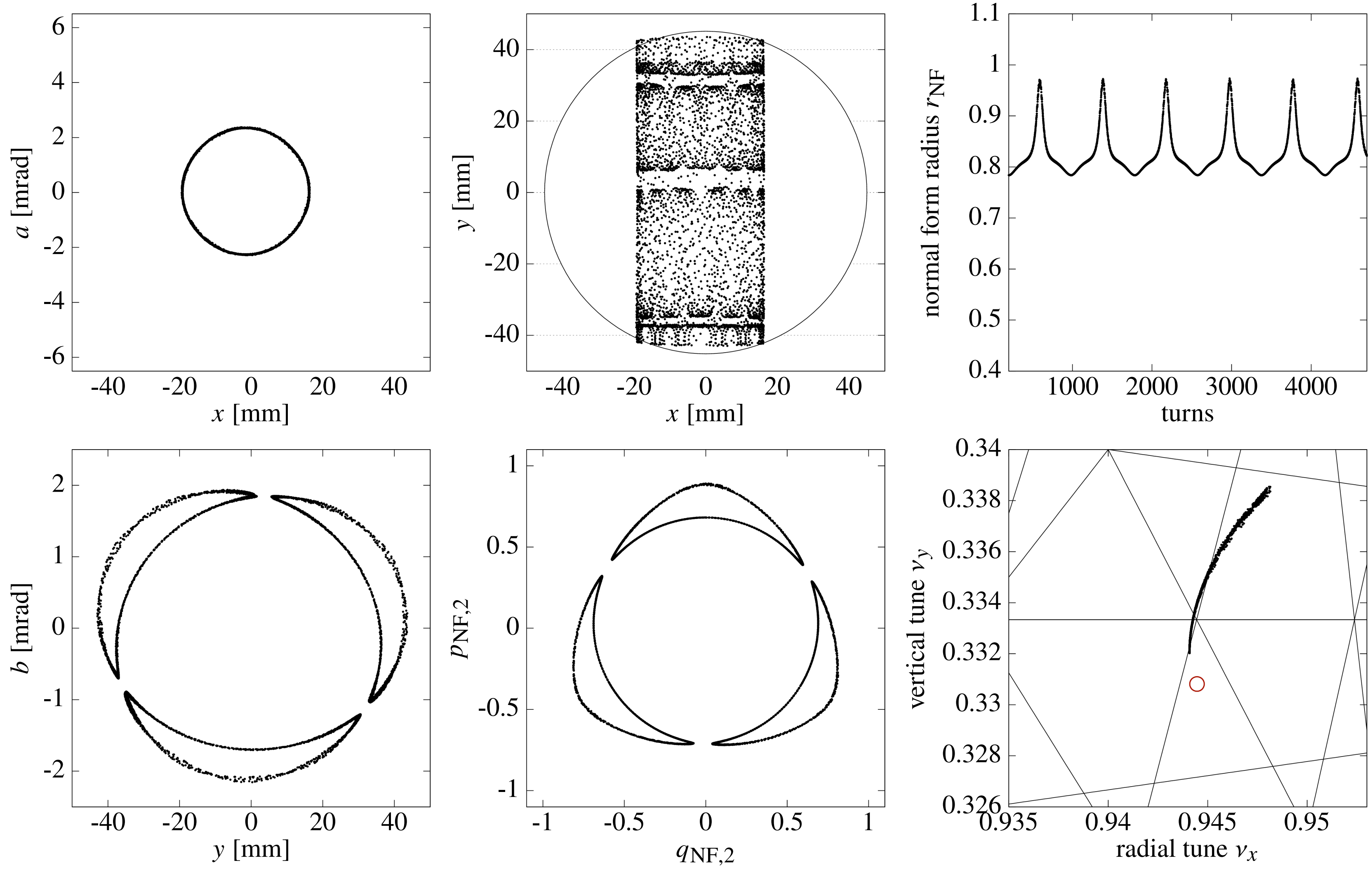

Figure 5.18: This particle ($\delta p = -0.015\%$) forms large islands around a period-3 fixed point structure in the vertical phase space, which is associated with a major modulation of the oscillation amplitude.

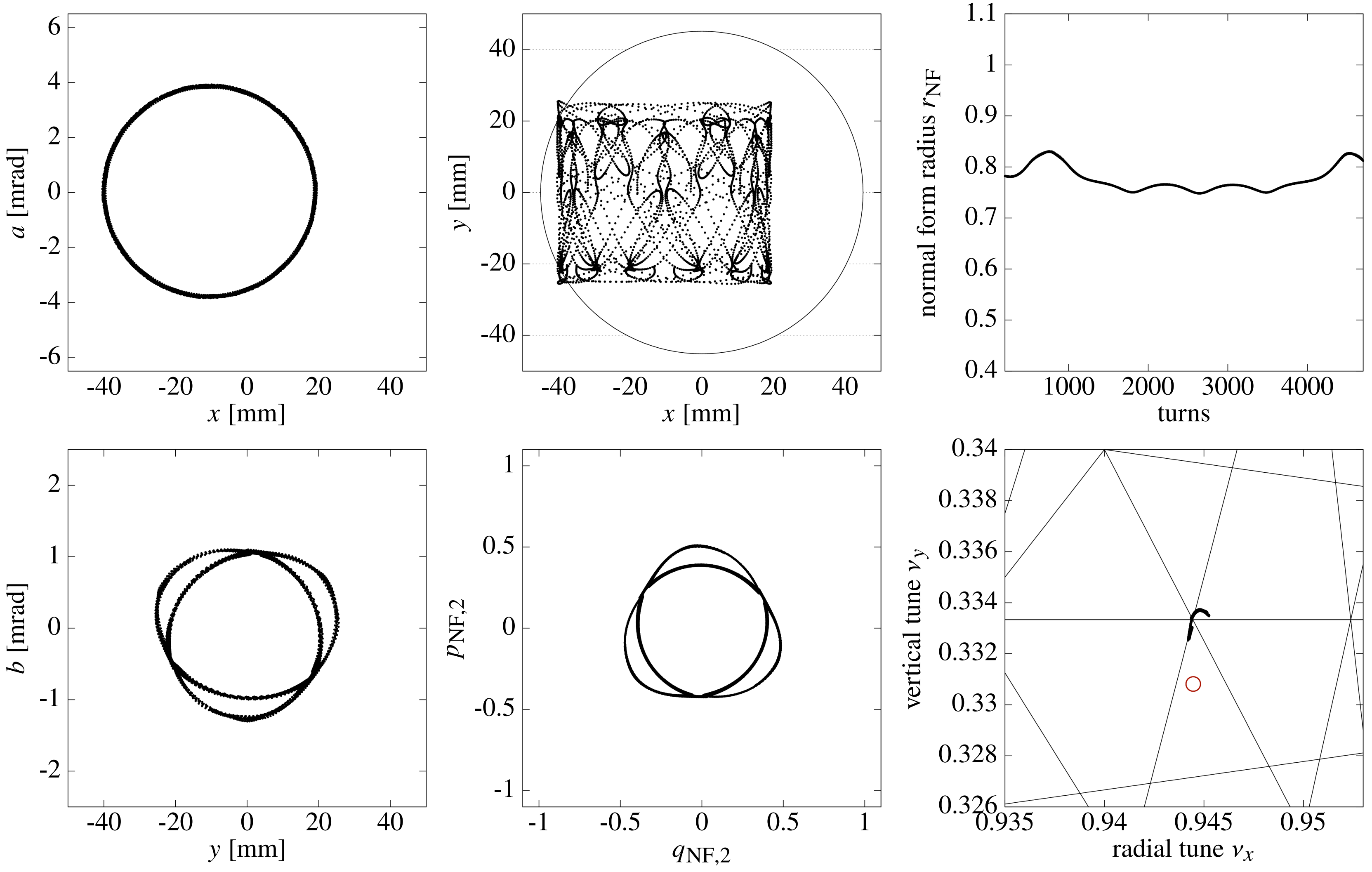

Figure 5.19: This particle ($\delta p = -0.127\%$) jumps between the islands. The large radial amplitude and/or the closeness to the $(17/18, 1/3)$ resonance point might have triggered the jump. This is an example of moderate unstable behavior around a period-3 fixed point structure.

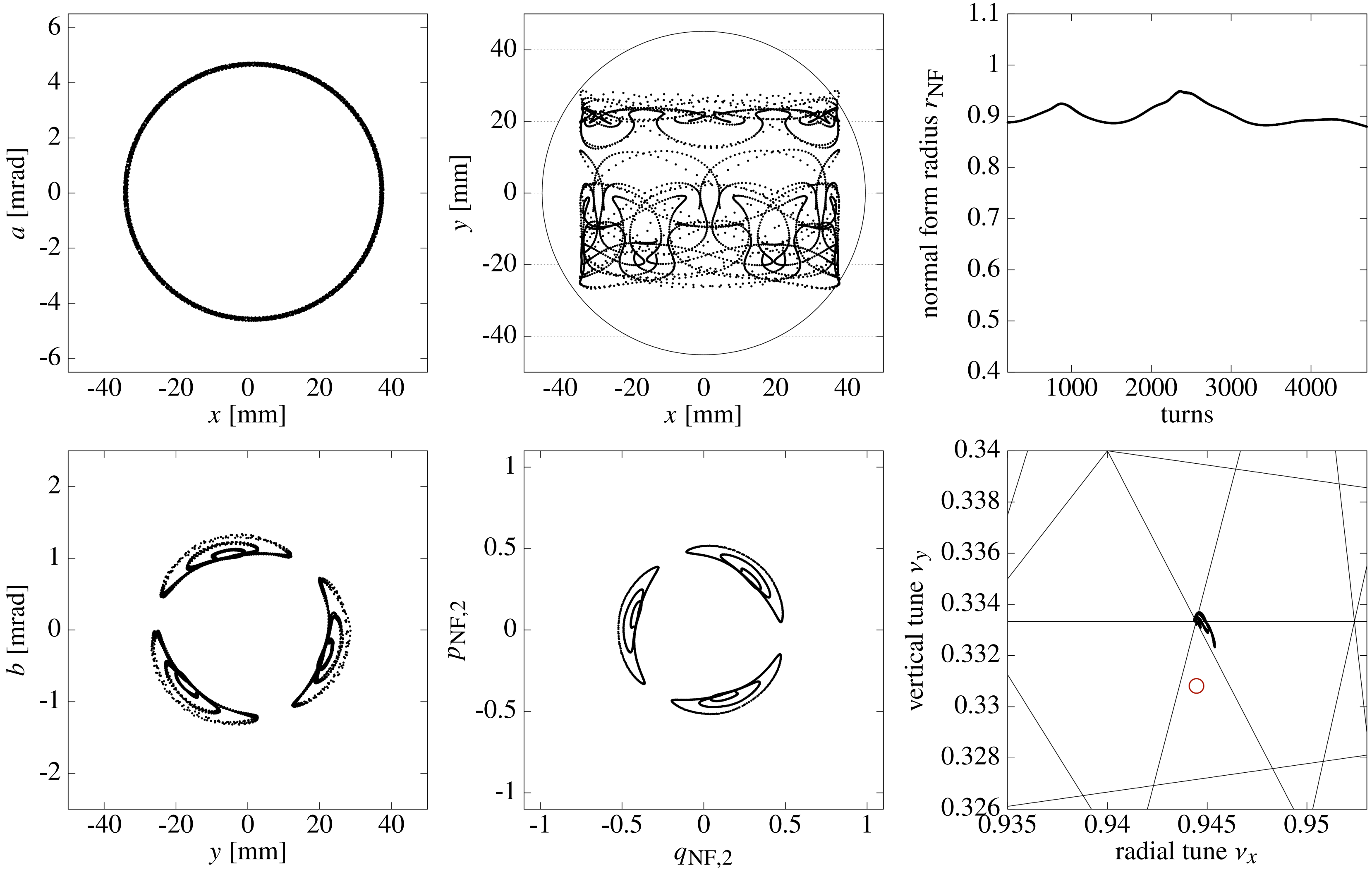

Figure 5.20: This particle ($\delta p = 0.024\%$) shows a different kind of moderate unstable behavior around a period-3 fixed point structure, where the island size varies. The particle has both, a large radial amplitude and the closeness to the $(17/18, 1/3)$ resonance point.

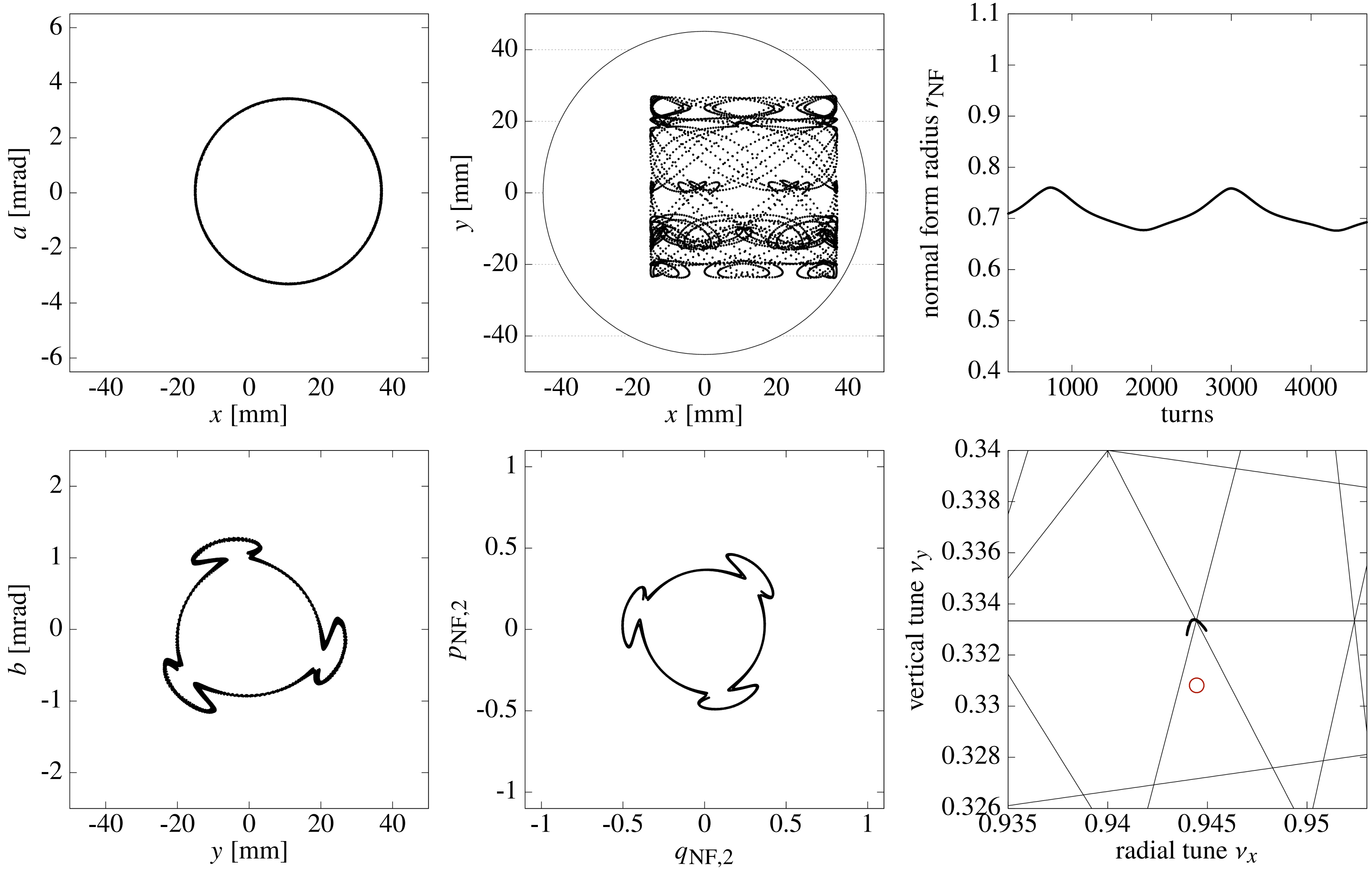

Figure 5.21: This particle ($\delta p = 0.140\%$) forms a shuriken like shape in the vertical phase space. In this pattern there are two period-3 fixed point structures involved indicated by the double crossing of the vertical $1/3$ resonance line.

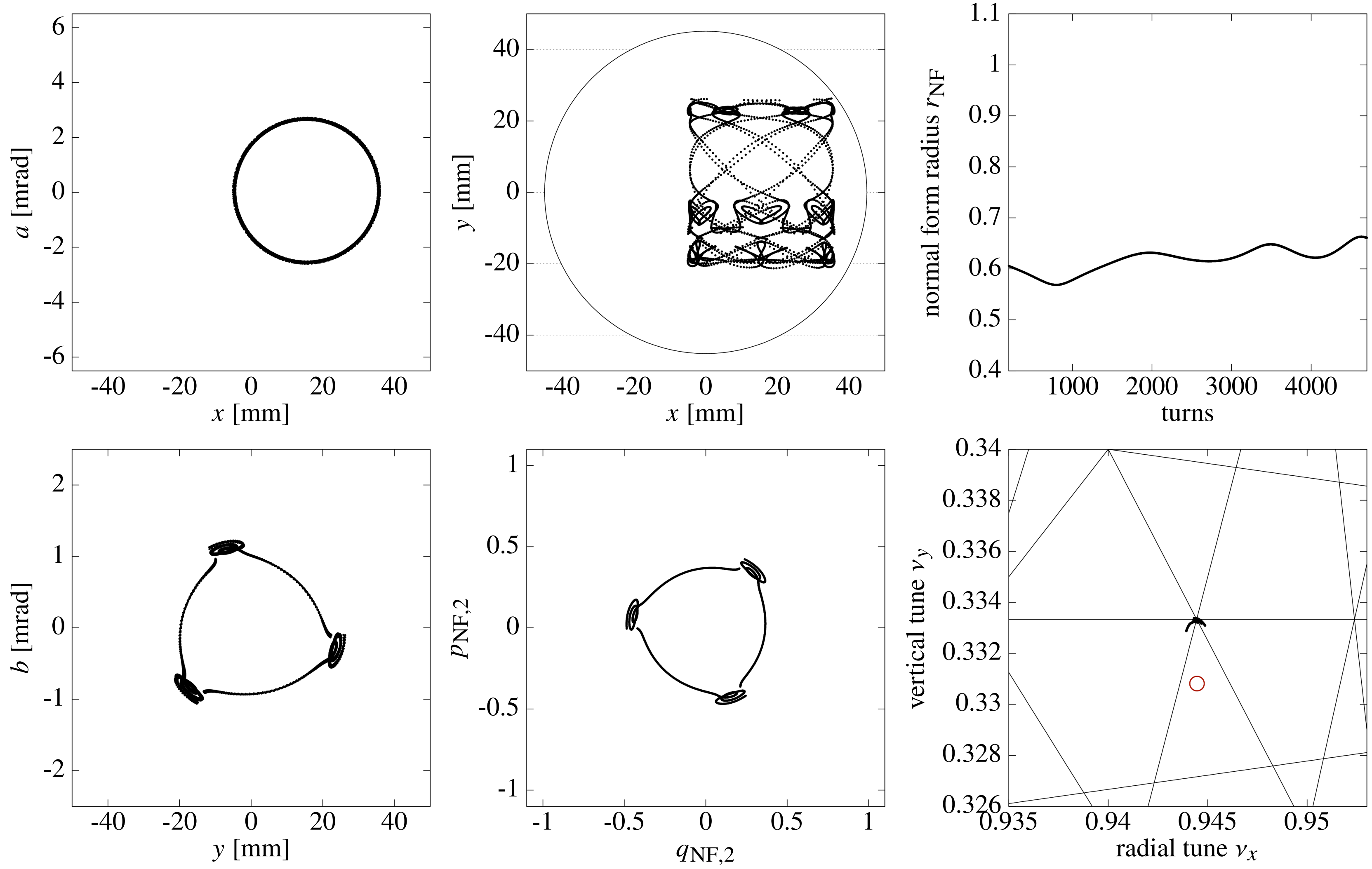

Figure 5.22: This particle ($\delta p = 0.196\%$) illustrates moderate unstable behavior in a shuriken pattern. The radial amplitude is not particularly large, but the resonance point $(17/18, 1/3)$ is very close, which might be the trigger of the unsuitability. The unstable behavior is also visible in the continuously increasing normal form radius.

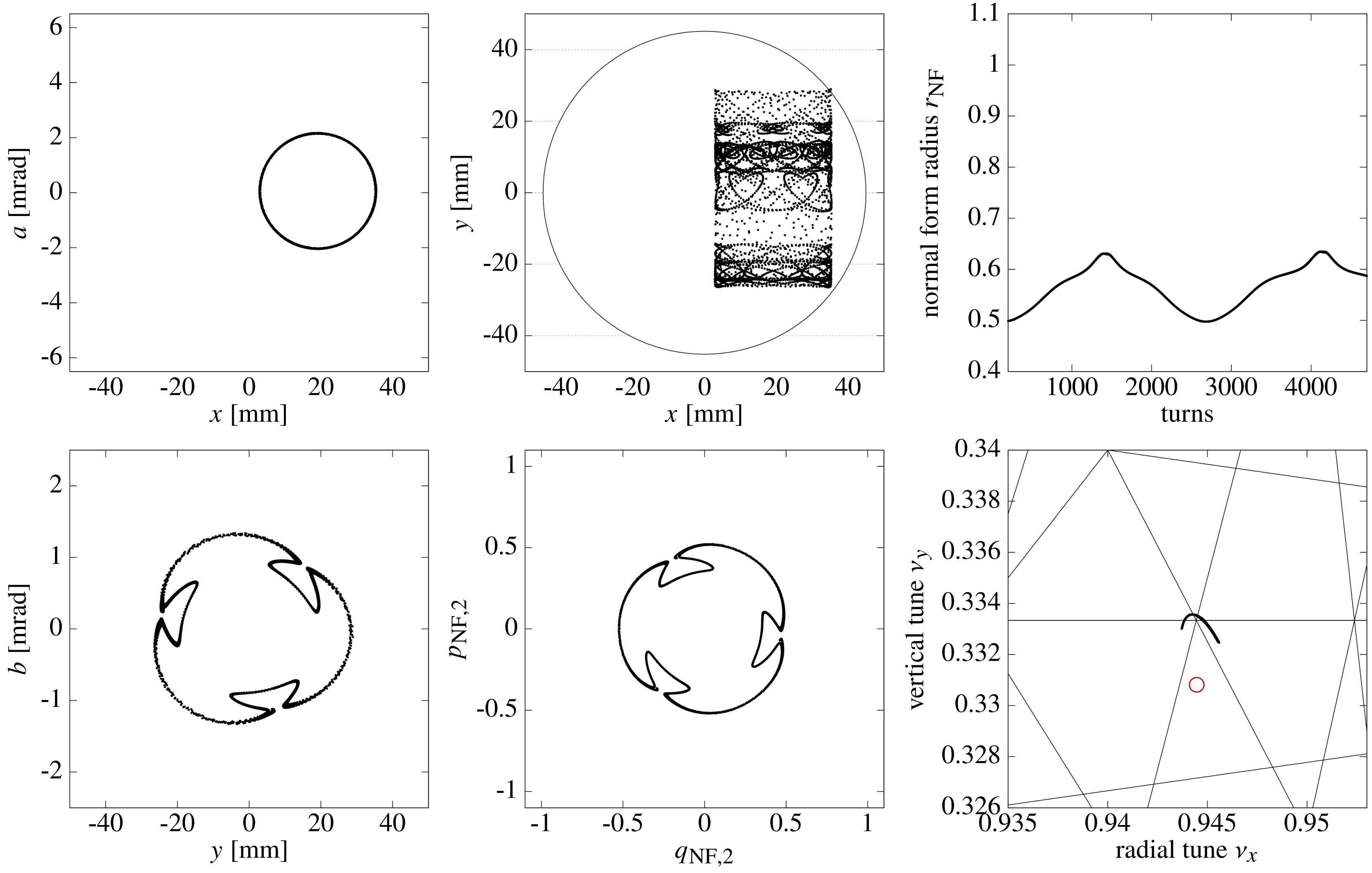

Figure 5.23: This particle ($\delta p = 0.242\%$) illustrates a shuriken pattern, where the two period-3 fixed point structures are more obvious. The muon experiences a major modulation in the vertical oscillation amplitude and performs a double crossing of the vertical $1/3$ resonance line.

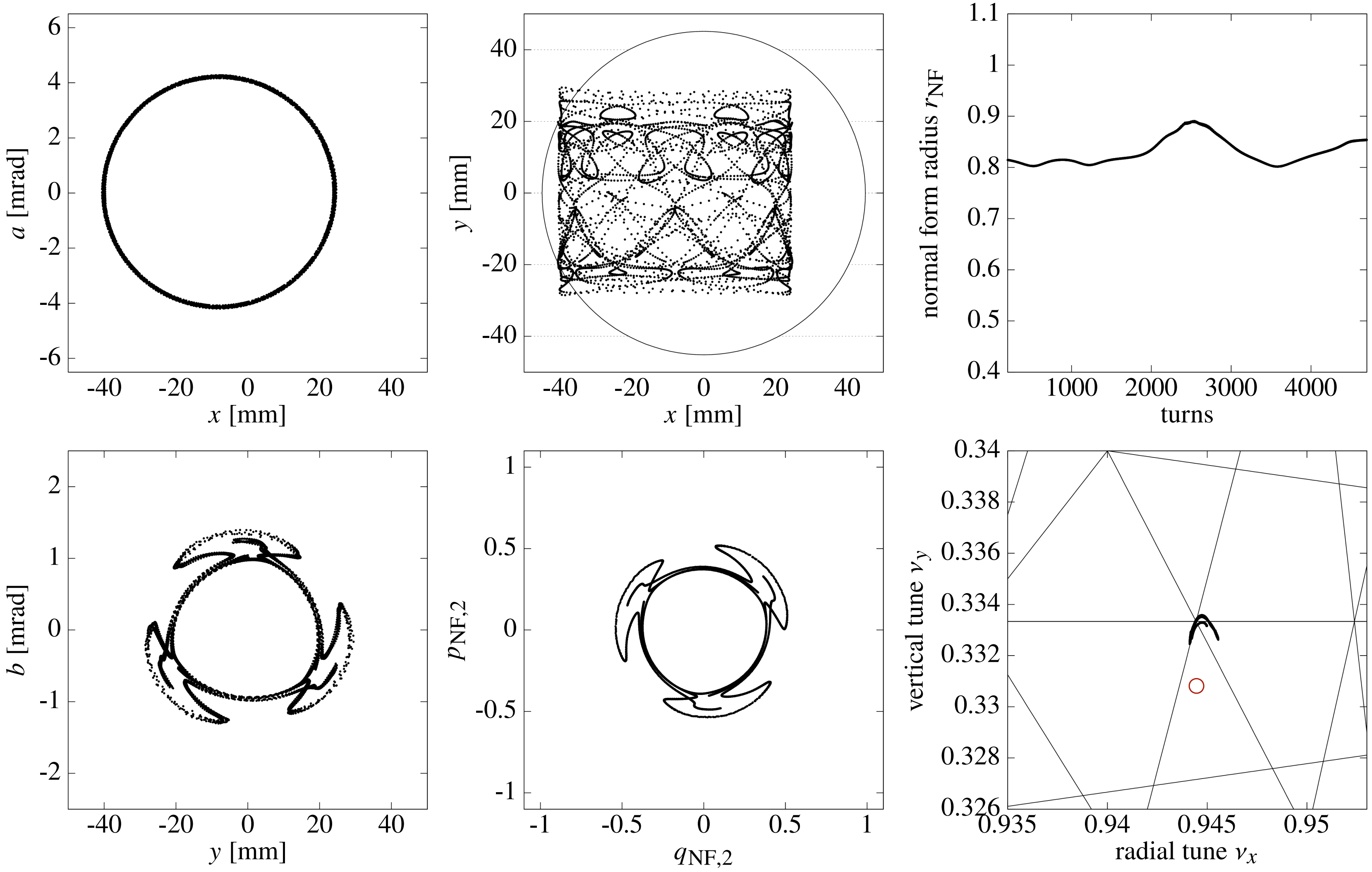

Figure 5.24: This particle ($\delta p = -0.096\%$) shows a shuriken pattern with unstable tendencies. The large radial amplitude and/or the closeness to the radial 17/18 resonance line might be the trigger for the instability.

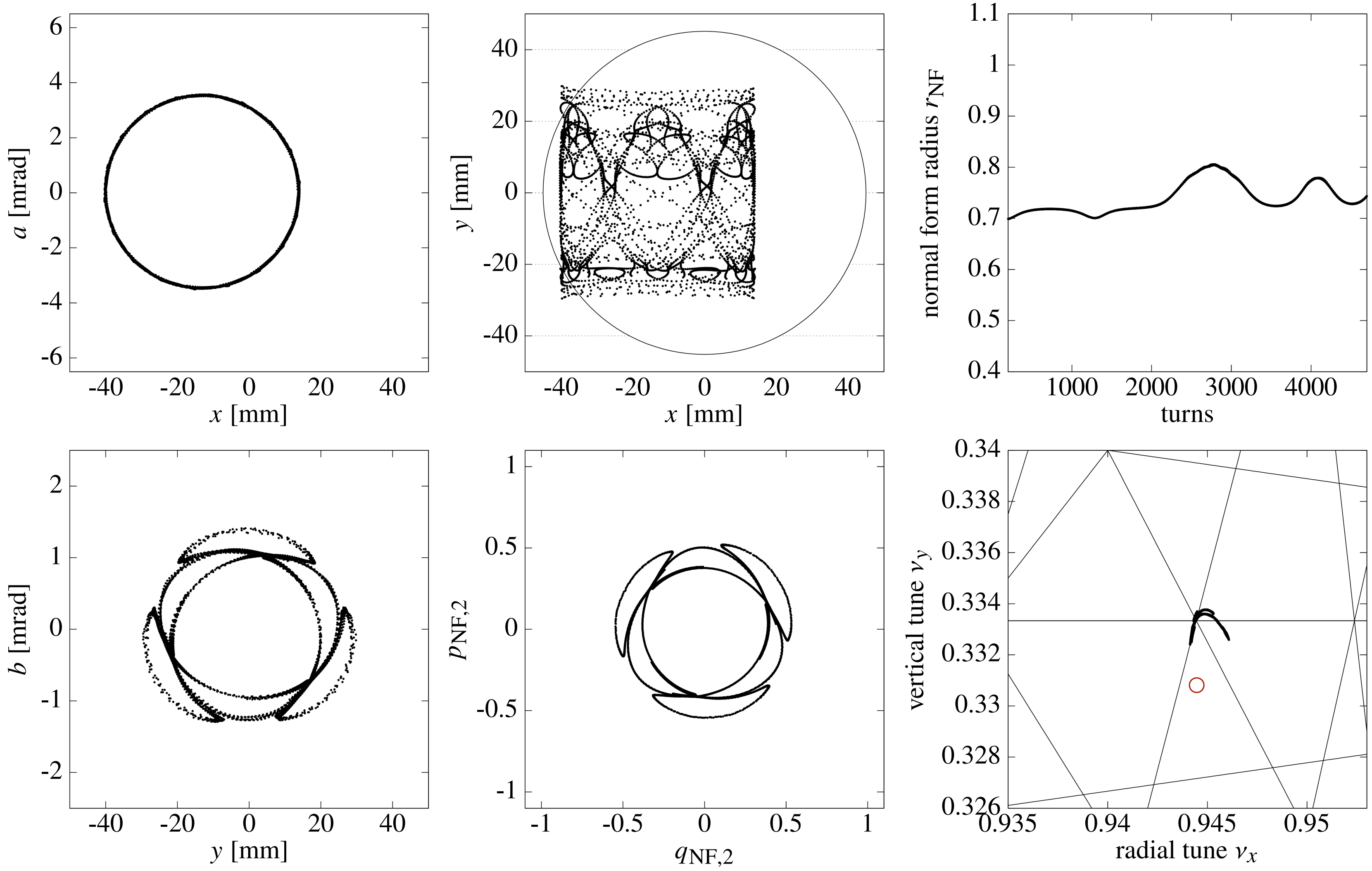

Figure 5.25: This particle ($\delta p = -0.159\%$) shows a shuriken pattern with a moderate instability. The two period-3 fixed point structures are so close together that the particle gets temporarily caught around the inner one of them in an island pattern.

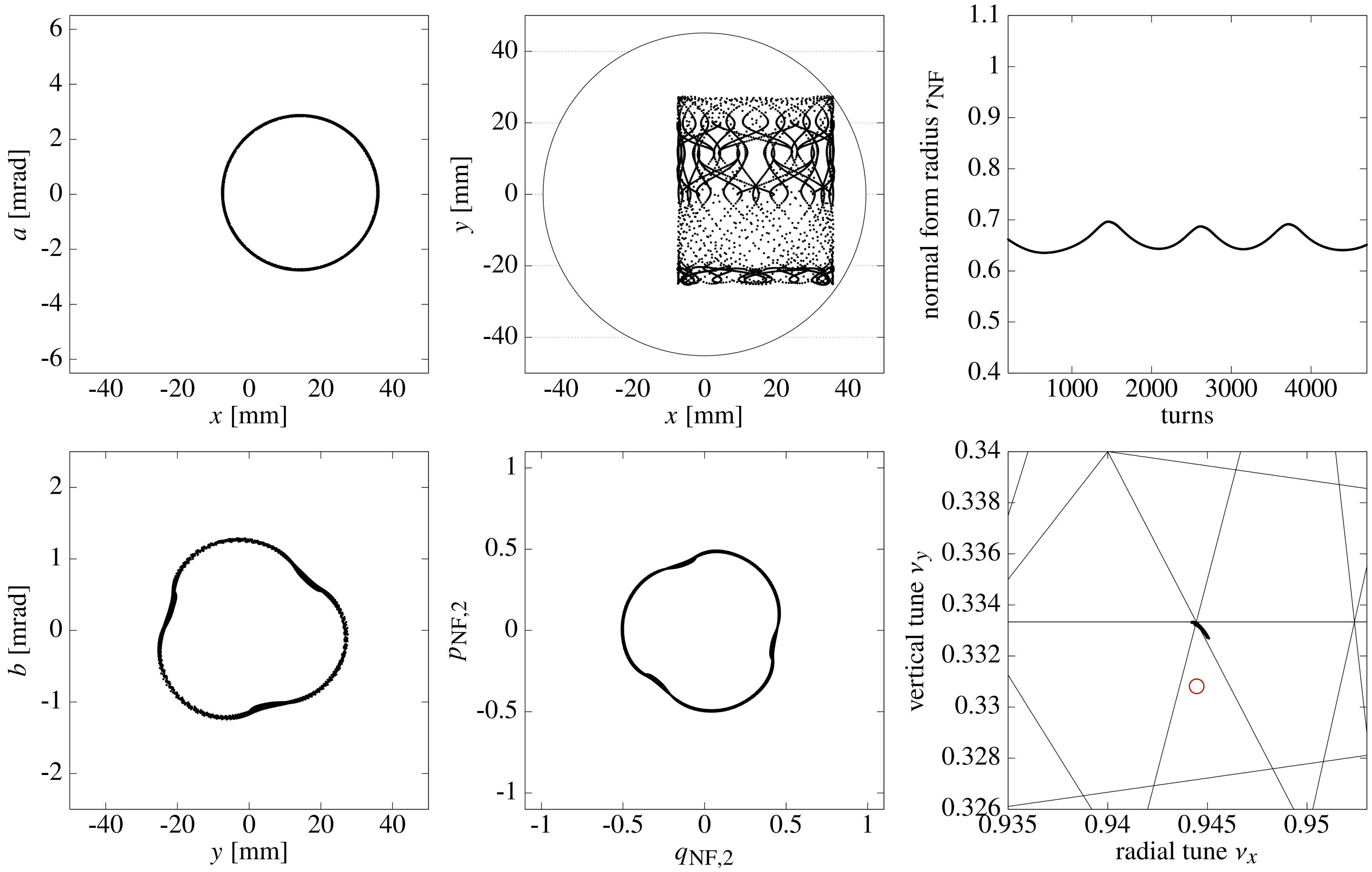

Figure 5.26: This particle ($\delta p = 0.181\%$) shows the pattern of a very blunt shuriken. The vertical amplitude oscillation is only moderate and illustrates there can be almost regular behavior between two period-3 fixed point structures.

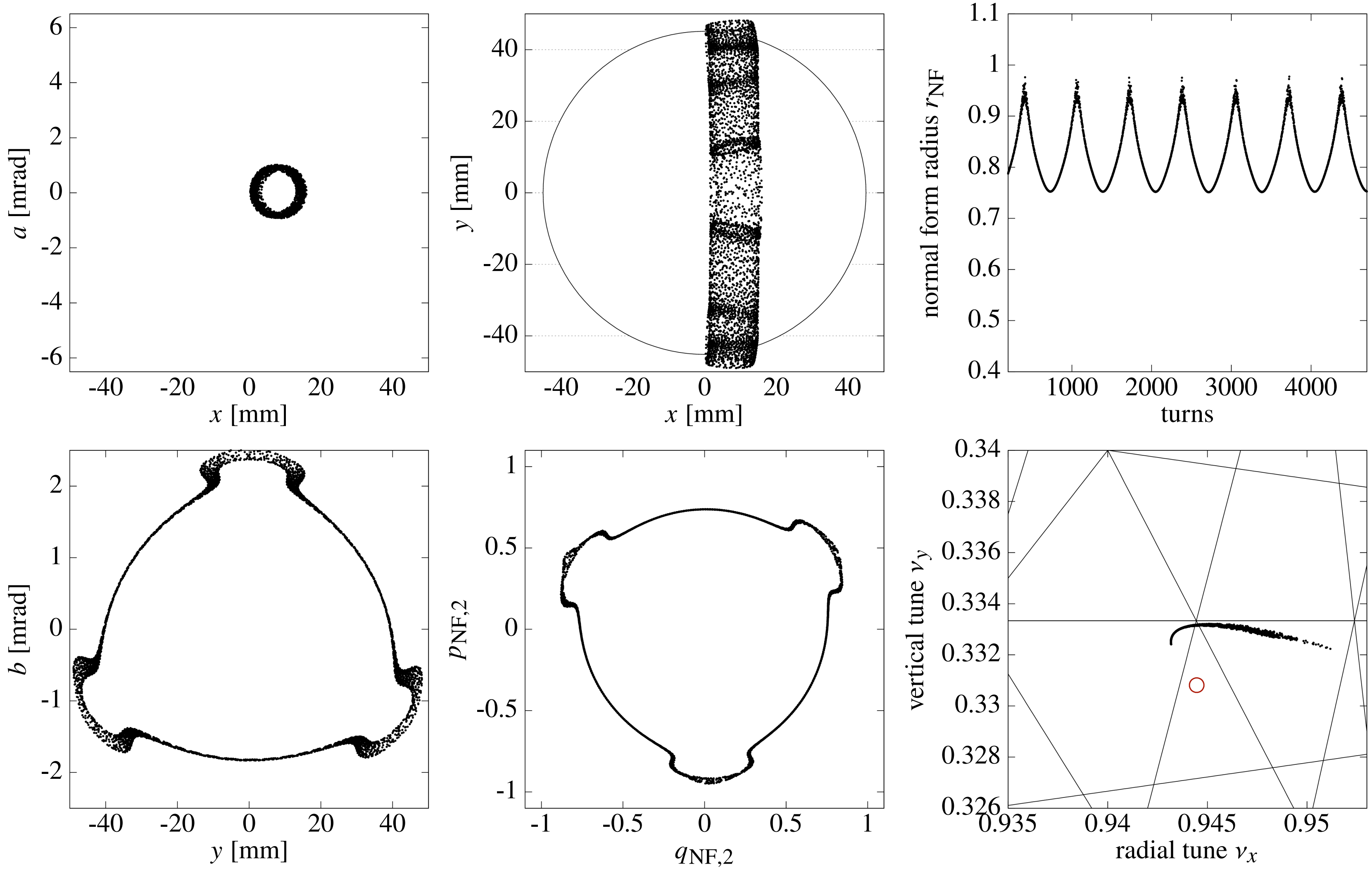

Figure 5.27: This particle ($\delta p = 0.106\%$) is characterized by a very large vertical amplitude, which is additionally modulated by the shuriken pattern. Its one of the very few particles for which the orbit considerably overlaps with the collimator boundary.

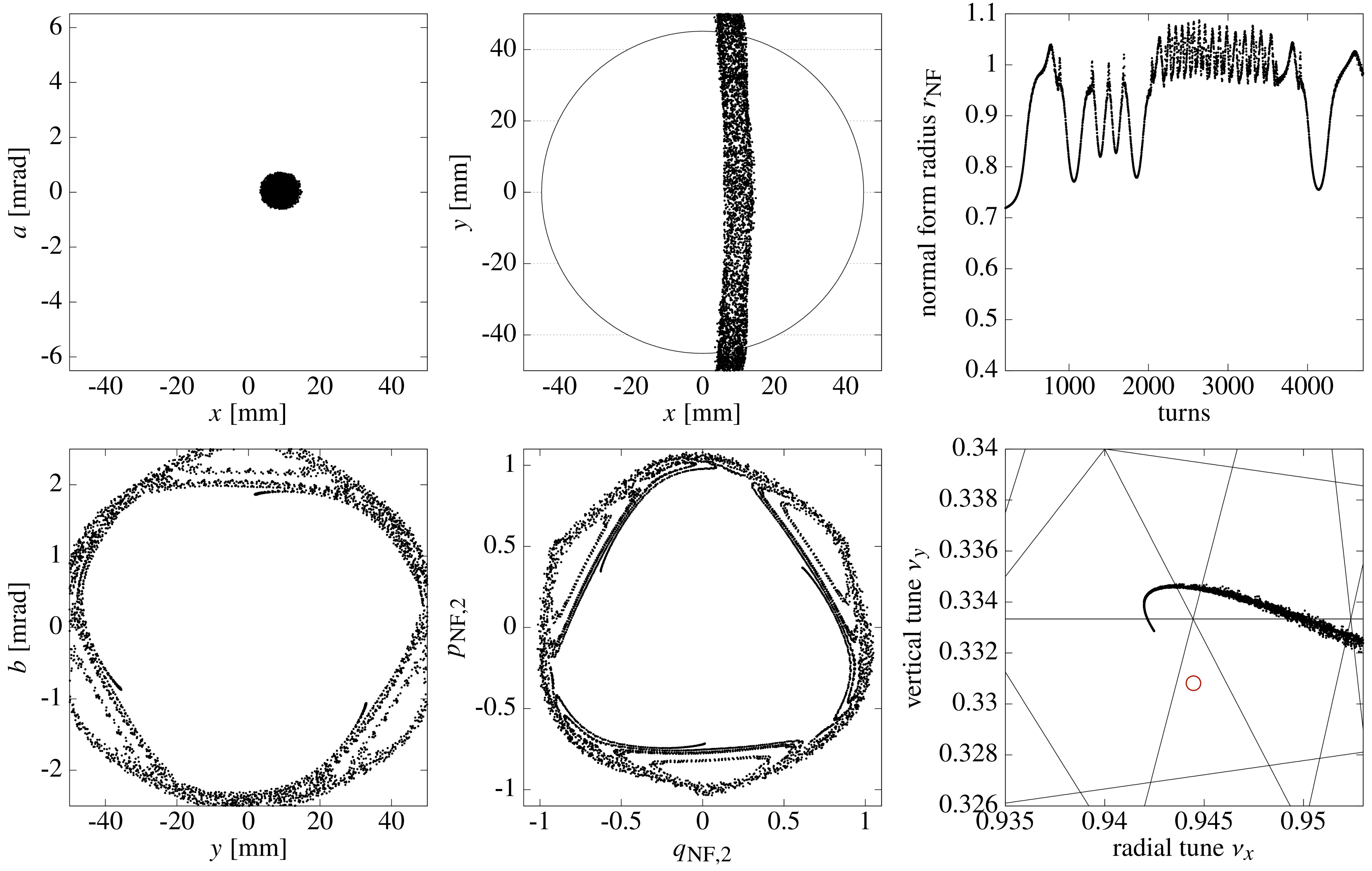

Figure 5.28: This particle (δp = 0.118%) shows strong instabilities caused by a combination of a very large vertical amplitude in combination with a period-3 fixed point structure, which occasionally captures the orbit in an island pattern.

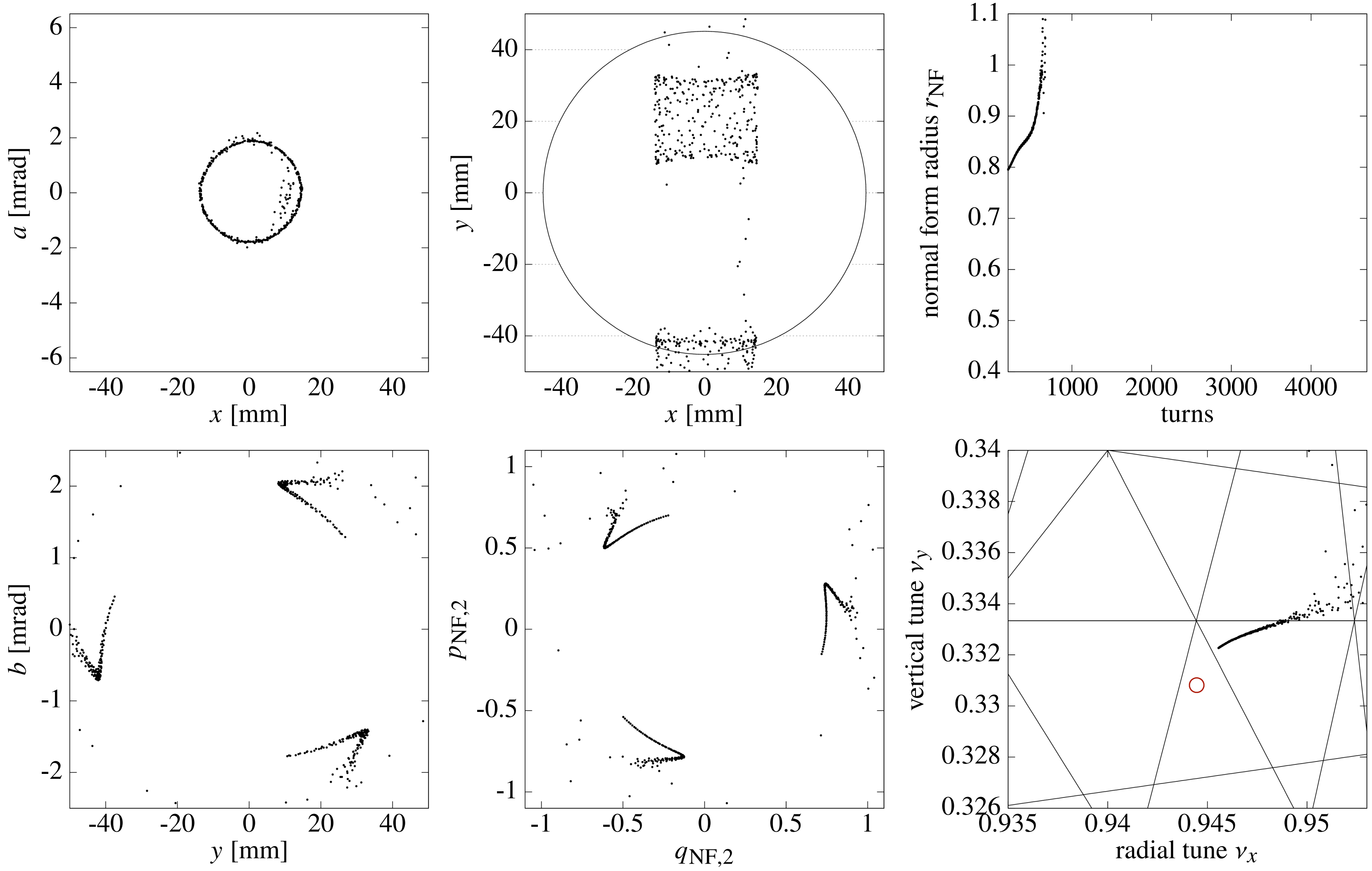

Figure 5.29: This particle ($\delta p = 0.010\%$) diverges due to its unstable orbit. The approach of the unstable fixed point with such a large vertical amplitude are likely the trigger of the divergence.

Another property that many lost particles share is a significant momentum offset, which radially shifts their respective reference orbit closer to the boundaries of the collimator. The dependence of the radial position of the reference orbit on the momentum offset decreases the maximum survivable size of those rectangular shapes in xy space significantly as previously discussed in Sec. 5.3.3 and illustrated in Fig. 5.4.

Last but not least, there are also particles like the one shown in Fig. 5.15, which get lost simply because of their constant but large oscillation amplitudes in the radial and/or vertical direction. However, it is not always obvious to distinguish them from particles that are under the influence of a period-3 fixed point structure like the particle in Fig. 5.16.

Fig. 5.15 to Fig. 5.29 also indicate that the xy pattern of lost particles often only barely touches the collimator boundary. For these cases, it may take many revolutions for both oscillations, in the radial and vertical direction, to reach their maximum simultaneously [84].

5.5.3 Period-3 Fixed Point Structures

There are period-3 fixed point structures in the vertical phase space as seen, for example, in Fig. 5.17. The period-3 fixed points are a property of the vertical projection of the stroboscopic muon tracking. They are associated with the vertical $1/3$-resonance, which is particularly relevant due to the strong eight order nonlinear tune shifts from the strong ninth order nonlinear field contributions of the 20th order multipole of the potential from the ESQ [82].

The period-3 fixed point structure corresponds to an orbit, which vertically oscillates around its momentum dependent reference orbit with a period of exactly three turns, i.e. a vertical betatron tune of $1/3$. However, such an orbit is not necessarily a closed orbit, which closes after three turns, because while the vertical behavior might be exactly resonant after three turns, the radial behavior is not.

There are stable fixed points and unstable fixed points within the period-3 fixed point structures. Accordingly, the term ‘period-3 fixed points’ describes a set of 6 fixed points at the same amplitude in yb, where every other fixed point is stable. The positions of the period-3 fixed points in the vertical phase space depend on the momentum offset δp and the radial phase space (due to coupling). The combination of stable and unstable fixed points creates island patterns around the stable fixed points as shown in Fig. 5.30.

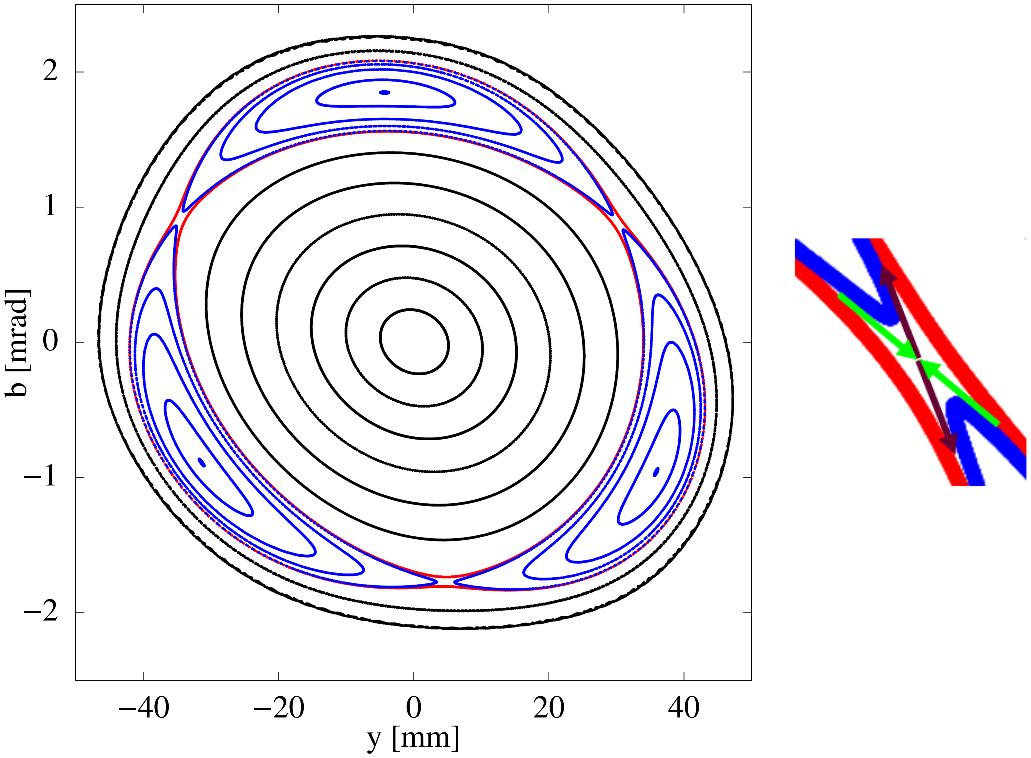

Figure 5.30: The left plot shows stroboscopic tracking in the vertical phase space illustrating orbit behavior with a single period-3 fixed point structure present. The orbits only differ in their vertical phase space behavior – they all have the same momentum offset of $\delta p = 0.126$ % and are at the momentum dependent equilibrium point in radial phase space ($x = 10.64$ mm, $a = 0.045$ mrad) having no radial oscillation amplitude. The blue orbits indicate the island patterns around the stable fixed points in the middle of the islands. The red orbits are right at the edge before being caught around the fixed points. The three unstable fixed points are in the space between the two red orbits, where the islands almost touch. In the right plot, the attractive (green) and repulsive (violet) eigenvectors of the linear dynamics around the unstable fixed points are schematically shown.

The unstable fixed points are located in the blank space between the islands. The linear dynamics around them are characterized by an attractive eigenvector with a corresponding eigenvalue smaller than one and a repelling eigenvector with an eigenvalue larger than one. Those unstable fixed points give rise to chaotic behavior because two phase space orbits that are initially near yield widely diverging dynamical behavior from each other once they come close to the unstable fixed point. The inner red orbit and the adjacent blue island orbit in Fig. 5.30 illustrate this chaotic behavior. They both approach the unstable fixed points along the attractive eigenvector (green), but the unstable eigenvector (violet) ejects them in opposite directions.

While the inner red orbit appears almost regular like the black orbits of lower amplitudes, the muon on the blue orbit with a slightly larger amplitude gets ejected outwards by the unstable fixed point, which drastically increases its vertical amplitude. In the case shown in Fig. 5.30, the stable fixed point is able to keep the particle in an island orbit. In Fig. 5.29, on the other hand, the particle cannot remain on the island orbit and diverges. In [98, 99], a similar analysis of the accelerator transfer map representing the Tevatron is performed and rigorous methods to determine the position of those fixed point structures are presented.

It is also not uncommon for two period-3 fixed point structures to be present simultaneously in the vertical phase space. Often the structures are oriented such that a stable fixed point of structure with the larger amplitude is 'above' an unstable fixed point of the structure with the lower amplitude. In Fig. 5.31 a phase space region with two period-3 fixed point structures for orbits with $\delta p = 0.339$ ‰ is shown.

The different plots illustrate how the relative position and interaction of the two period-3 fixed point structures drastically changes the dynamics for particles that are initiated on initially very near orbits. This further emphasizes the potential to chaotic behavior caused by the unstable fixed points within those period-3 fixed point structures.

The two period-3 fixed point structures can be well separated, as shown in a), yielding the known island patterns with 'regular' orbits in between. However, the structures can also move into each other such that some orbits are caught between the two period-3 fixed point structures and follow the shape of a threefold shuriken around the two island patterns as shown in green in b) and c). When the two period-3 fixed point structures come even closer, the opposite fixed points of the two period-3 fixed point structures can annihilate each other, resulting in triangular patterns with rounded corners (see gray patterns in d)).

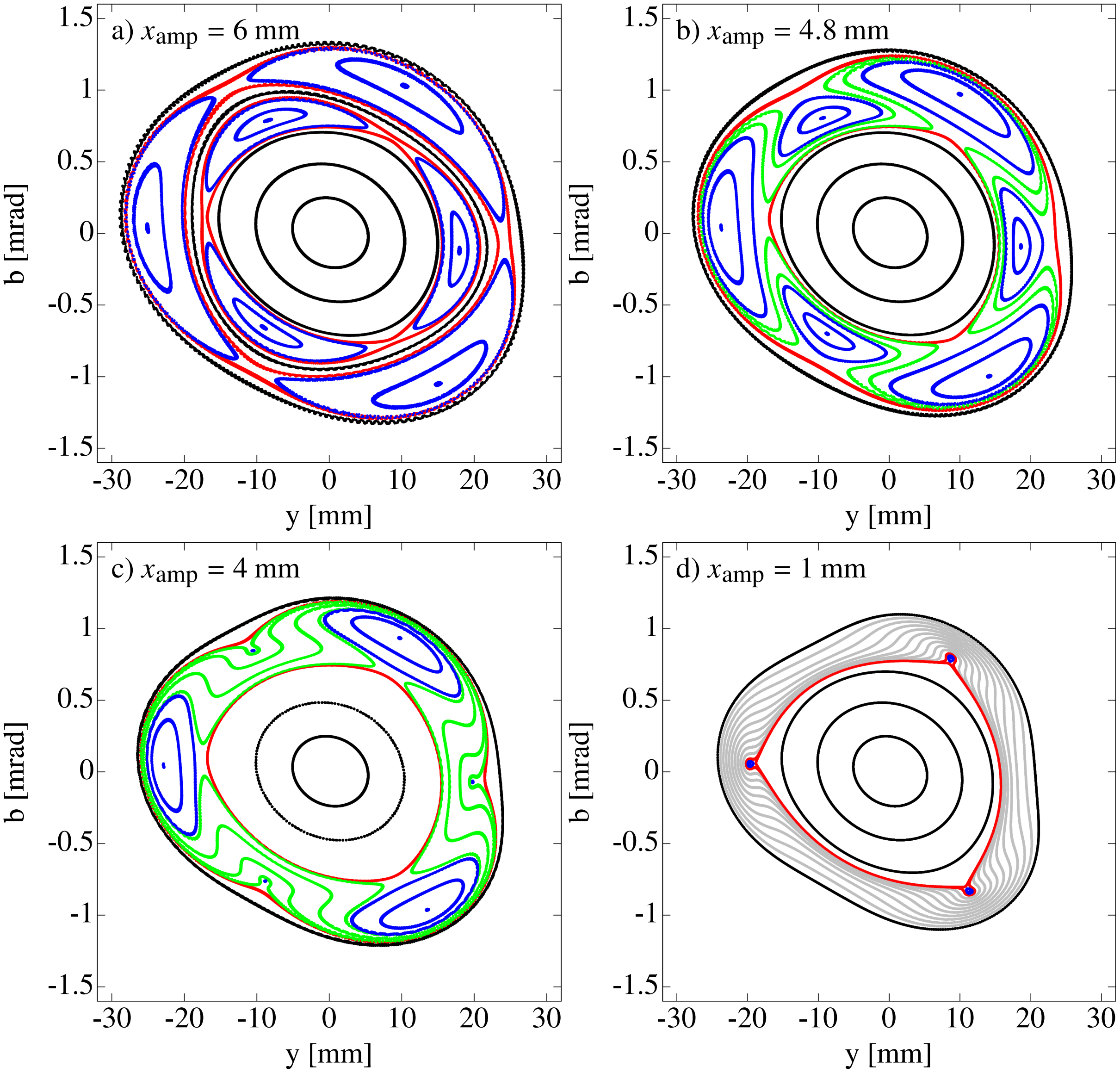

Figure 5.31: Stroboscopic tracking in the vertical phase space illustrating orbit behavior with two period-3 fixed point structures present. The orbits in each plot only differ in their vertical phase space behavior. All orbits have the same momentum offset of $\delta p = 0.339$ %. The four plots differ by their radial amplitude around the momentum dependent equilibrium point in radial phase space at ($x = 27.7$ mm, $a = 0.144$ mrad). The radial amplitudes are: a) $x_{\text{amp}} = 6$ mm, b) $x_{\text{amp}} = 4.8$ mm, c) $x_{\text{amp}} = 4$ mm, d) $x_{\text{amp}} = 1$ mm. The blue orbits indicate the island patterns around the stable fixed points. The red orbits are right at the edge before being caught around the period-3 fixed points. The green orbits are caught around both period-3 fixed point structures. The gray orbits in d) emphasize that half of the fixed points from c) have indeed been annihilated.

While the period-3 fixed point structures often lead to a significant vertical amplitude modulation, many of them are well within the boundary of the collimators like the examples shown in Fig. 5.30 and Fig. 5.31. So, the involvement of a particle in a period-3 fixed point structure or two does not necessarily mean that it is lost, but the additional modulation of the vertical amplitude drastically increases the risk of getting lost for those particles.

All orbit patterns shown in Fig. 5.15 to Fig. 5.29 can be found in a similar form either in Fig. 5.30 or Fig. 5.31. In other words, we fully understand what is causing the different types of patterns. The major difference for some particles is the stability of their pattern. The phase space regions chosen in Fig. 5.30 and Fig. 5.31 are stable and do not share the characteristics of unstable orbits which are large radial amplitudes and/or closeness to the 17/18 resonance point.

5.5.4 Muon Loss Rates from Simulation

We have seen what different phase space tracking patterns can arise due to period-3 fixed point structures. We also saw that these structures can be responsible for losses due to the modulation of the oscillation amplitude in the vertical phase space. To get a more general understanding of how prominent these patterns are among the entire distribution and how common they are among lost particles, we need a mechanism to characterize these patterns in a way that can be automatically detected.

The various degrees of instabilities, especially among particles involved with period-3 fixed point structures make a generalized categorization difficult. There is no obvious distinction between certain unstable islands and certain shuriken patterns, and also no clear distinction between very blunt shuriken patterns and very triangularly deformed regular elliptical patterns. Accordingly, we only make two distinctions. First, we try to distinguish between particles involved with the vertical 1/3-resonance and particles that are not. Among the particles that are involved with the vertical 1/3-resonance, we make a further distinction between pure island patterns and everything else. A pure island pattern is a (non-across-jumping) island pattern. Fig. 5.19 shows an across-jumping island structure, where the orbit jumps from one fixed point island to another. In comparison, Fig. 5.20 also shows an unstable island pattern, but one that remains on the island around the fixed points.

For reference, we will call particles that we detect as being involved with the vertical 1/3-resonance 'period-3 particles' and all the others 'regular particles'. Of the period-3 particles, we will only give a special name to the 'island particles', because the period-3 non-island particles are a very diverse group, which is not easily described by a single word without mischaracterizing at least some of its elements.

Since the transition between patterns is continuous as the gray orbits in Fig. 5.31d illustrates, the category of period-3 particles and the category of regular particles might have elements that are almost identical.

The following paragraph clarifies how we define the different categories with our detection mechanisms. We start by explaining how we identify period-3 particles. We consider the vertical phase space in polar coordinates and look at the phase space behavior in steps of three. The first three vertical phase space angles during tracking are denoted by $\phi_{1,1}$, $\phi_{2,1}$ and $\phi_{3,1}$, and the next three angles are denoted by $\phi_{1,2}$, $\phi_{2,2}$ and $\phi_{3,}$. Starting from the initial angle $\phi_{1,1}$, we have angles after each turn as $\phi_{2,1}$, $\phi_{3,1}$, $\phi_{1,2}$, $\phi_{2,2}$, $\phi_{3,2}$, ..., $\phi_{1,n}$, $\phi_{2,n}$, $\phi_{3,n}$ until the total turn number is reached. For our tracking, its 4500 turns in total, which corresponds to $n = 1500$. Additionally, we define the angle advances $\Delta\phi_{i,n} = \phi_{i,n} - \phi_{i,n-1}$. To avoid ambiguity in the value for the angles, we require that value for $\phi_{i,n}$ is chosen such that $\phi_{i,n} \in \left[\phi_{i,n-1} - \pi, \phi_{i,n-1} + \pi\right]$. If there is a sign change from $\Delta\phi_{i,n-1}$ to $\Delta\phi_{i,n}$ the sign-change-count κ_i is increased by one. If all κ_i are nonzero after the 4500 turns, then we categorize the particle as period-3 particle.

To identify island particles, we use the definitions from above and additionally introduce the range $\mathbb{D}_{i,1} = \left[\phi_{i,1}, \phi_{i,1}\right]$ of the angles for each of the three potential island locations. With every iteration step the ranges of the angles are updated to

$$\mathbb{D}_{i,n} = \left[\mathbb{D}_{i,n,\mathrm{LB}}, \mathbb{D}_{i,n,\mathrm{UB}}\right] = \left[\min\left(\phi_{i,n}, \mathbb{D}_{i,n-1,\mathrm{LB}}\right), \max\left(\phi_{i,n}, \mathbb{D}_{i,n-1,\mathrm{UB}}\right)\right] . \tag{5.6}$$

The abbreviations 'LB' and 'UB' denote the lower and upper bound of the range respectively. Note that the rule to avoid ambiguity in the value for the angles from above also applies here. All particles for which the sum of the three ranges is less than a full revolution (2π) after the 4500 tracking turns are considered island particles. In other words, island particles satisfy

$$\sum_{i=1}^{3} |\mathbb{D}_{i,1500}| < 2\pi. \tag{5.7}$$

With these recognition mechanisms implemented, we were able to characterize all particles and determine their proportion as presented in Tab. 5.1. Period-3 particles are dominating among lost particles. Accordingly, period-3 particles and island particles, in particular, are more prone to be lost. But by far not every period-3 particle or island particle is lost. More than 77% of island particles and more than 92% of period-3 particles survive the 4500 turns. As Fig. 5.30 illustrates, sometimes the amplitude of these period-3 structures is so low that the additional modulation of the amplitude is not enough to be critical.

Table 5.1: Percentages of different characterization groups. Read as follows: *x* % of *Base* particles have the property *Property*. All particles that hit a collimator during the 4500 turns of tracking are considered lost.

Base Property	All	Lost	Period-3	Island
Lost	0.686%	100%	7.44%	22.2%
Period-3	7.06%	76.4%	100%	100%
Island	1.00%	32.4%	14.2%	100%

While island particles make up only 1/7 of period-3 particles, they are responsible for almost half the losses associated with period-3 particles. This is particularly surprising because the island particle category excludes most unstable patterns by definition (exceptions are moderate instabilities that do not contravene the recognition criteria like the particle shown in Fig. 5.20). On the other hand, period-3 particles cover a wide range of patterns some of which barely show a modulation of the vertical amplitude as the example of the gray orbits in Fig. 5.31d shows.

To understand how the losses occur over time, we plot the accumulative loss ratio over the 4500 turns in Fig. 5.32. Island loss is the fastest growing loss over the first 1000 turns before settling almost asymptotically. This is explained by different modulation frequencies around the period-3 fixed point structures. The closer to the unstable fixed point, the larger the modulation and the slower the modulation frequency. Accordingly, the island modulation is on average faster than the shuriken modulation.

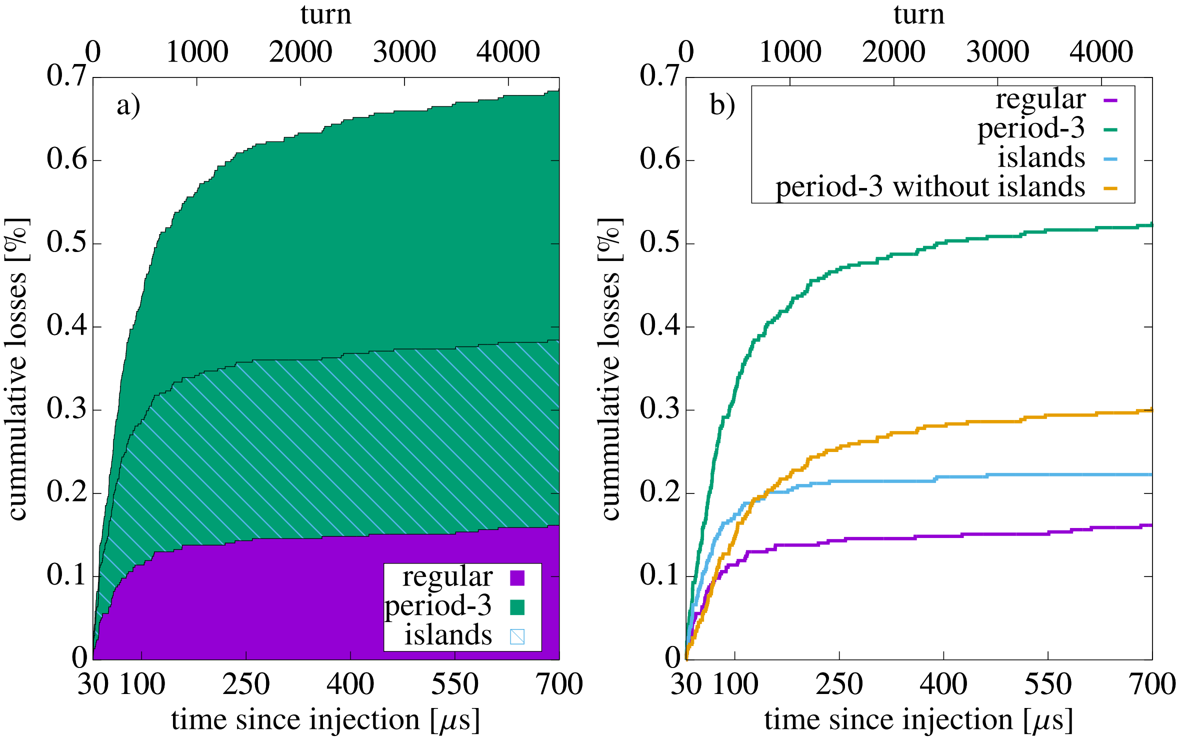

Figure 5.32: a) Shows how the muon loss ratio is composed of regular particles (purple) and particles involved with period-3 fixed point structures (green). Of the period-3 particles (green), the fraction caught in islands structures is indicated by the blue stripe pattern. In b) the loss ratio over time is shown for each subgroup of lost particles to better understand which losses drive to overall loss from plot a). The tracking starts after the initial 30 μs of final beam preparation when data taking is initiated.

5.6 Conclusion

The Poincaré return map description of the storage ring model of the Muon g-2 Experiment [85] and its analysis with DA normal form methods yielded many insightful characteristics of the system. We gained an understanding of the form of the closed orbit within the storage ring as well as details on how it changes with an offset in the momentum δp. Considering that particles oscillate around their corresponding reference orbit, which is the closed orbit of their momentum offset, the radial shift of the closed orbit with momentum offset is particularly critical. This shift brings the equilibrium state of the radial oscillation closer to the collimator boundary, which increases the risk of particles getting lost.

The tune analysis provided a detailed understanding of how the oscillation frequencies of particles depend on their momentum offset and their amplitudes relative to their respective reference orbit. This analysis showed that particles over the entire momentum offset range could cross the vertical $1/3$-resonance frequency for certain vertical and radial amplitude combinations.

The strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential have a significant effect on amplitude and parameter dependent tune shifts. This property manifests itself in the dominating eighth order dependencies in the amplitude and momentum dependent tune shifts and the drastic change in the tune footprint for calculations of order $m > 8$, which include the ninth order terms of the original map.

Further tracking analysis revealed period-3 fixed point structures in the vertical phase space. They are associated with the vertical $1/3$-resonance tune and cause significant vertical amplitude modulations to the particles that are caught around them. We were able to connect all vertical phase space patterns of lost particles either with regular distorted elliptical patterns or with patterns that arise around one or two of these period-3 fixed point structures. Additionally, instabilities caused by large radial amplitudes and/or closeness to the $(17/18)$ resonance point significantly mixed multiple of the known orbit patterns. This only allowed for a limited automatic recognition of patterns, which in turn revealed valuable insights about the effect of these period-3 fixed point structures on the loss rates of muons in the storage ring. Particles associated with period-3 fixed point structures are at a higher risk of getting lost.

CHAPTER 6

VERIFIED CALCULATIONS USING TAYLOR MODELS

In this chapter, we take steps towards making the methods presented above self-verified.

Since many aspects that have to be carefully considered for a rigorous transfer to the verified world lay beyond the scope of this thesis, this chapter will only yield a discussion of the basic principles behind some of them. However, the aspect of verified global optimization and its application to the normal form defect for verified stability estimates will be analyzed in greater detail.

To introduce the concept of verified global optimization using Taylor Models (TM), we first apply it to two well-known example optimization problems. First, in Sec. 6.1, we run a Taylor Model based verified global optimization in different operating modes on the 2D and the generalized Rosenbrock function, as it is one of the most commonly used examples to test global optimization algorithms.

In Sec. 6.2, we discuss the optimization problem of finding minimum energy configurations of particles that have their pairwise interaction energy modeled by the Lennard-Jones potential. It is one of the simplest examples to explain, yet arbitrarily complex to solve depending on the number of particles in the configuration and the dimensionality of the configuration. In comparison to the Rosenbrook example, the setup of the Lennard-Jones optimization problem is more complex especially for configuration in 2D and 3D.

In Sec. 6.3, we discuss the intricacies of using the methods from Chapter 4 and Chapter 5 for a verified stability analysis of those dynamical systems. In particular, we take a detailed look at options of utilizing the normal form defect from Sec. 2.4 as a measure of stability. With the gained understanding of verified global optimization from Sec. 6.1 and Sec. 6.2, we analyze its application to the normal form defect to calculate verified stability estimates for the simulated phase space behavior in the Muon g-2 Storage Ring.

6.1 The Rosenbrock Optimization Problem

6.1.1 The Rosenbrock Function

The Rosenbrock function

$$f(x, y) = (a - x)^2 + b\left(y - x^2\right)^2 \tag{6.1}$$

was introduced by Howard H. Rosenbrock [78]. It is a non-convex function that is commonly used as a test problem for optimization algorithms. The parameters are usually set to $(a, b) = (1, 100)$, and so we will use those parameters here as well. Fig. 6.1 illustrates the Rosenbrock function for those parameters.

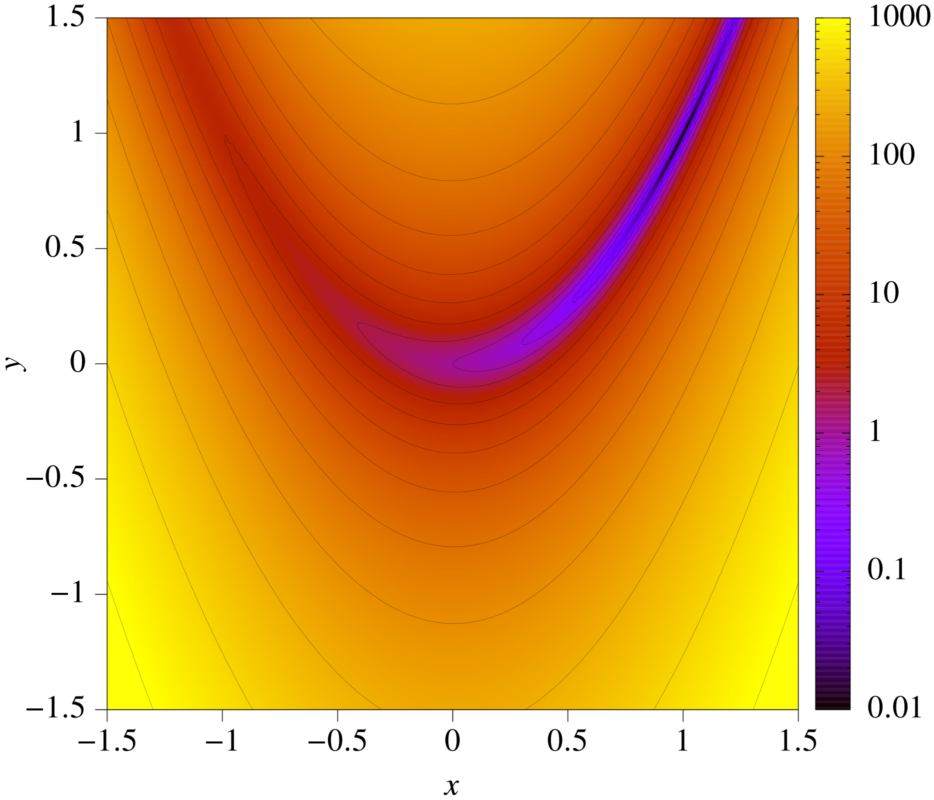

Figure 6.1: A contour plot of the Rosenbrock function with $(a, b) = (1, 100)$.

The Rosenbrock function is also referred to as Rosenbrock's valley function or Rosenbrock's banana function for obvious reasons. It is characterized by a long and deep valley, the floor of which constitutes a shallow valley. This shallowness is one of the aspects that challenges optimizers.

There are various multidimensional generalizations of the Rosenbrock function to compare more advanced optimization algorithms. In this work, we will use the following generalized form

$$f_{n\mathrm{D}}\left(\vec{x}\right) = \sum_{i=1}^{n-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + \left(1 - x_i\right)^2\right], \tag{6.2}$$

where $n \geq 2$ is the dimension and x_i are the optimization variables.

Note that this generalized definition is consistent with the definition of the 2D Rosenbrock function from above and also retains the difficulties of the original problem of a deep valley with a long and shallow valley floor, but with a complexity that increases with n. Unless specified otherwise, we will refer to the generalized Rosenbrock function as the Rosenbrock function or the objective function of the optimization.

The Rosenbrock function is a composition of quadratic expressions. Because of the double squares, the Rosenbrock function is always a fourth order polynomial. Additionally, none of the individual terms in the sum can be negative. Hence, a global minimum would be reached if all individual terms of the sum are zero. The $(1 - x_i)^2$ terms are only zero for $x_i = 1$, which also yields zero for the remaining terms. Accordingly, $\vec{x}^{\star} = (1, 1, ..., 1)$ is the single global minimum of the Rosenbrock function for which every term is zero and therefore the overall objective function is zero.

In Fig. 6.2, the Rosenbrock function is illustrated in multiple 2D projections around its minimum at $\vec{x}^{\star}$. In other words, all x_i are set to one except for the variables shown in the projection.

The Rosenbrock function also has a dependency problem. For the first variable x_1, the following dependent terms appear

$$100\left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2. \tag{6.3}$$

For any of the variables x_i with $1 < i < n$, there is one additional dependent term with

$$100\left(x_{i+1} - x_i^2\right)^2 + \left(1 - x_i\right)^2 + 100\left(x_i - x_{i-1}^2\right)^2. \tag{6.4}$$

The last variable x_n, only appears in one term, namely

$$100\left(x_n - x_{n-1}^2\right)^2. \tag{6.5}$$

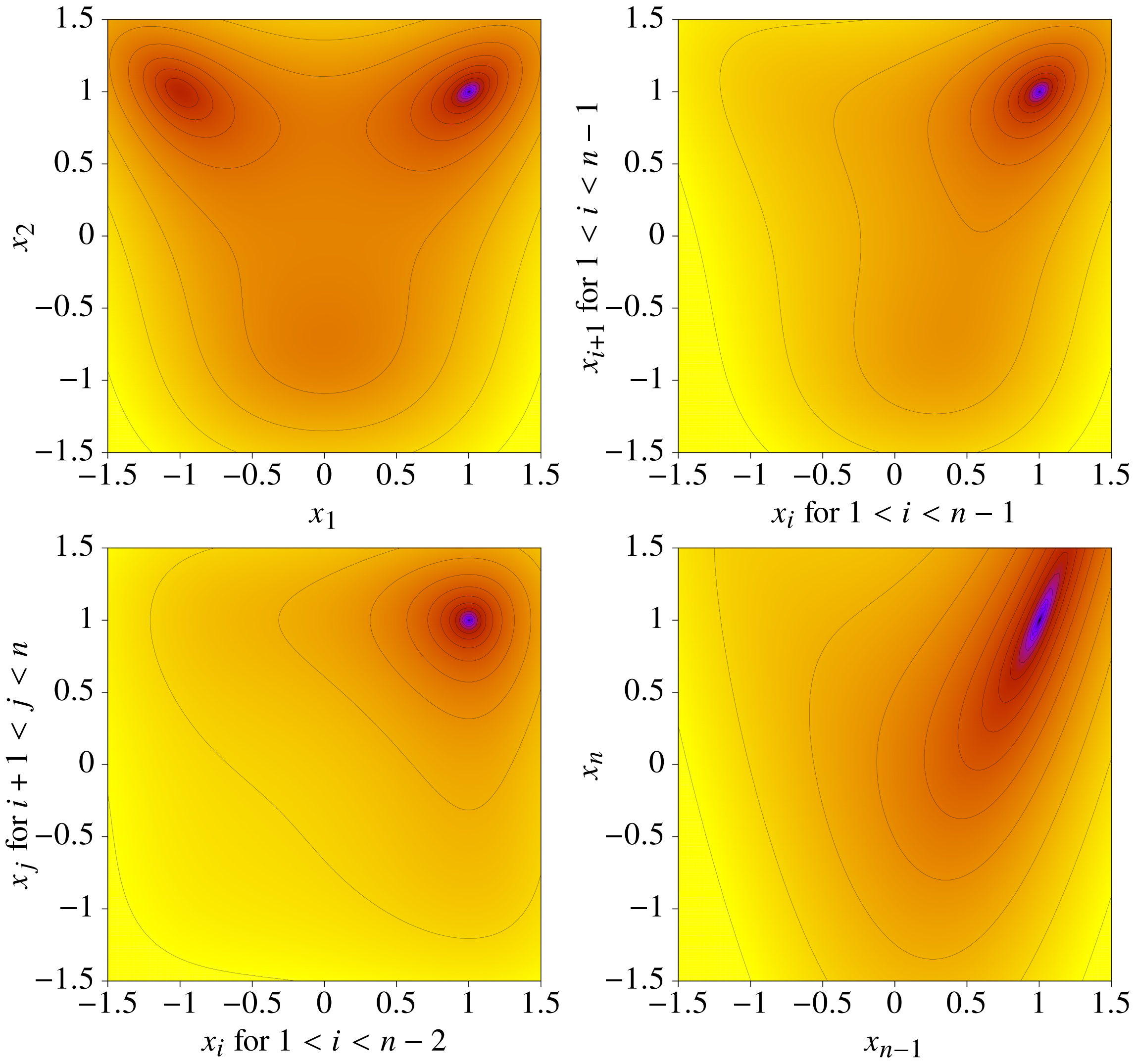

Figure 6.2: Projections of the multidimensional generalizations of the Rosenbrock function (Eq. (6.2)) into 2D-subspaces around minimum at $\vec{x} = (1, 1, ..., 1)$, i.e., all variables are equal one except for the ones shown in the respective plot.

6.1.2 Global Optimization Using COSY-GO

The global optimization is performed using COSY-GO [63, 64, 59]. In the most advanced setting (QFB/LDB), the algorithm uses both of the advanced Taylor Model based bounding methods, namely, the quadratic fast bounder (QFB) and the linear dominated bounder (LDB), which were mentioned in Sec. 2.6 and were introduced in [64]. Additionally, COSY-GO also uses naive Taylor Model bounding and interval evaluations (IN). For comparisons, COSY-GO offers to run an optimization with some of the advanced methods disabled. By ranking the bounding methods in the order: QFB, LDB, naive TM, and IN, the operating mode is denoted by its highest ranking bounding method, e.g., the running mode LDB indicates that LDB, naive TM, and IN are used but not QFB.

Because the global minimum of the Rosenbrock function is already known, we are just interested in the algorithm's performance to narrow down the domain of the minimum and its value. Accordingly, we can choose an arbitrary search domain for the optimization that includes $\vec{x}^{\star}$. We will investigate the Rosenbrock function over the domain $[-1.5, 1.5]^n$.

For the optimization, we evaluate the objective function the way it is written in Eq. (6.2) and not expanded out in a single second, third, and fourth order polynomial terms. We also clarify that the optimization is performed with no additional knowledge about the derivatives of the objective function.

6.1.2.1 Illustration of the Cluster Effect and Dependency Problem using the 2D Rosenbrock Function

In Fig. 6.3, the performance of COSY-GO on the 2D Rosenbrock function is visualized in the form of its splitting pattern. It shows the individual boxes analyzed in the various operation modes. All calculations are performed with fourth order Taylor Models (TM) except for the interval evaluation, which does not use TM.

The significant differences in the splitting patterns are the number of splits, and the way boxes are split. For the operating mode in naive TM and IN, boxes are always split in half, where each of the two modes has its own methods of deciding in which variable domain the box is split, i.e., for 2D either splitting in x or in y. With LDB and QFB, the boxes are decreased in size as the respective method sees fit. Especially close to the minimum this avoids the cluster effect [44, 37]. In Fig. 6.4, the boxing close to the minimum is illustrated, which clearly shows the cluster effect and its avoidance using QFB/LDB.

Another advantage of the Taylor Model based approach is the avoidance of the dependency problem [55]. However, due to the simplicity of the 2D Rosenbrock function and its weak dependency problem in the form from Eq. (6.2), the advantages of the Taylor Model based methods are not so

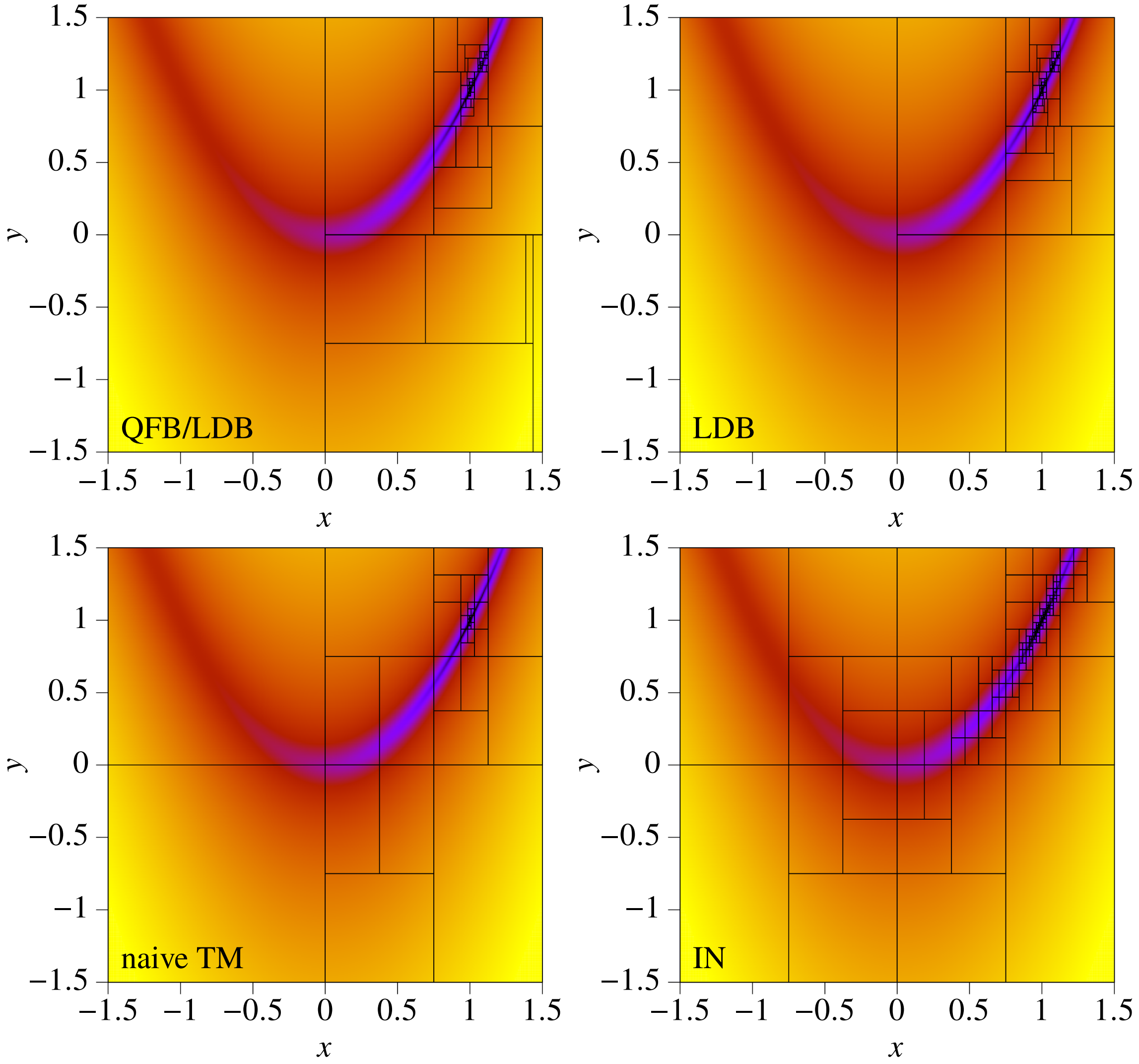

Figure 6.3: Verified global optimization of the 2D Rosenbrock function using COSY-GO in different operation modes with fourth order Taylor Models for all modes except interval evaluations (IN).

prominent relative to the IN evaluations.

To still visually emphasize the advantages of the TM operations over intervals, we artificially increase the dependency problem in the objective function by modifying it to $f = f_{2\mathrm{D}} - f_{2\mathrm{D}} + f_{2\mathrm{D}}$. In Fig. 6.5, the QFB/LDB methods using fourth order Taylor Models are compared to the interval method for the modified objective function.

Even though the fourth order TM representation of the modified objective function only differs from the TM representation of the non-modified objective function by a slightly different remainder bound, the behavior and efficiency of the algorithm with QFB/LDB change more for the modified

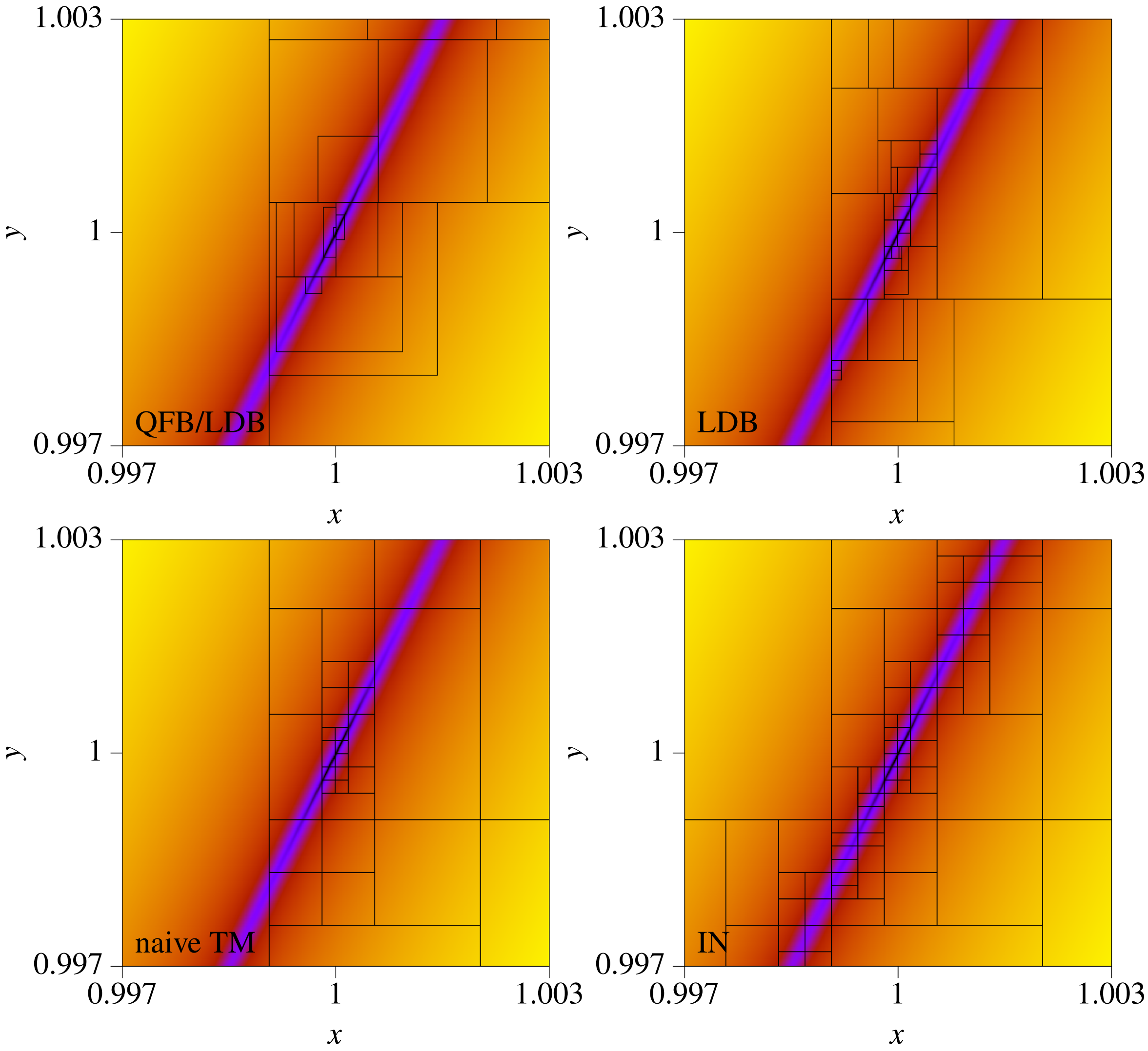

Figure 6.4: No cluster effect for the COSY-GO operating mode QFB/LDB, but a significant cluster effect for the IN evaluation.

objective function than one would initially expect. This is because the algorithm in the QFB/LDB mode also performs intermediate steps with lower order Taylor Models, which are quicker to evaluate but less accurate. Those lower order evaluations are more sensitive to the dependency problem, which explains the effect of those intermediate steps on the splitting decisions.

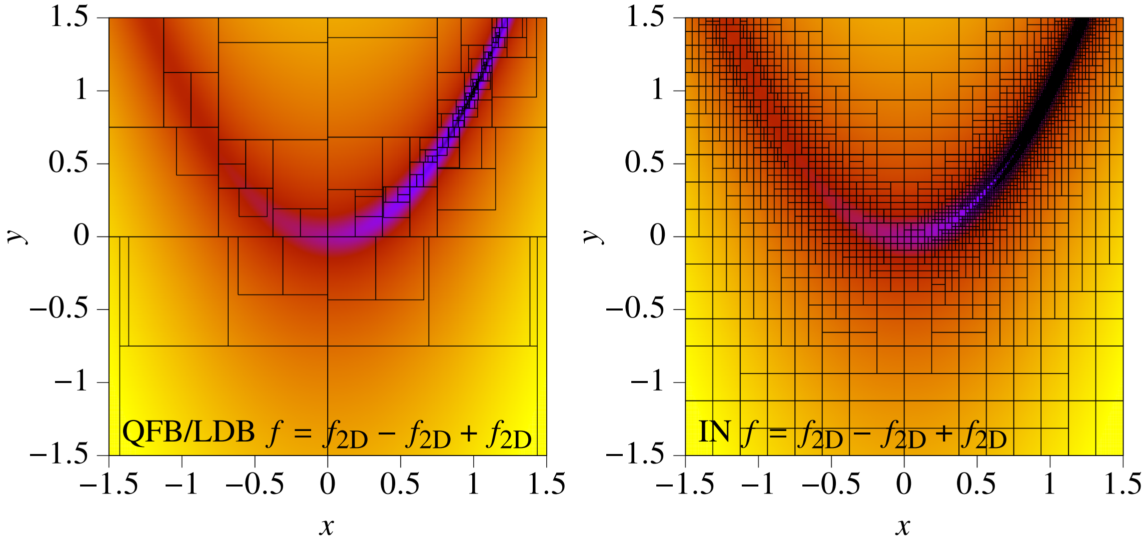

Figure 6.5: Splitting comparison between fourth order Taylor Model approach with QFB/LDB enabled and interval evaluation using the example of the modified 2D Rosenbrock function.

6.1.2.2 Performance of COSY-GO for High Dimensional Rosenbrock Function

Next, we analyze the performance of COSY-GO for the optimization of the higher dimensional Rosenbrock function in the form from Eq. (6.2). The search domain of the optimization is always set to $[-1.5, 1.5]^n$. Accordingly, the search volume increases exponentially with the dimension of the objective function.

As a stopping condition of the algorithm, we require that boxes with a side length $s <$ 1E-6 are not split. Ideally, the optimizer reduces the search volume of 3^n by at least a factor of $3{,}000{,}000^n$ to a single box with a volume smaller than $(1\text{E-}6)^n$. In the most advanced setting (QFB/LDB), which requires a minimum Taylor Model order of two, COSY-GO manages to reduce the search domain to a single box with a side length $s <$ 1E-6 for every dimension n that we tested.

Fig. 6.6 illustrates how the performances of COSY-GO in the evaluation of the generalized Rosenbrock function from Eq. (6.2) varies for different Taylor Model orders in the QFB/LDB mode. For comparison, the performance of using interval evaluations is also shown.

The second order calculation outperforms the higher order calculations in both aspects, regarding the speed and the number of required steps. On the one hand, the evaluation of higher order Taylor

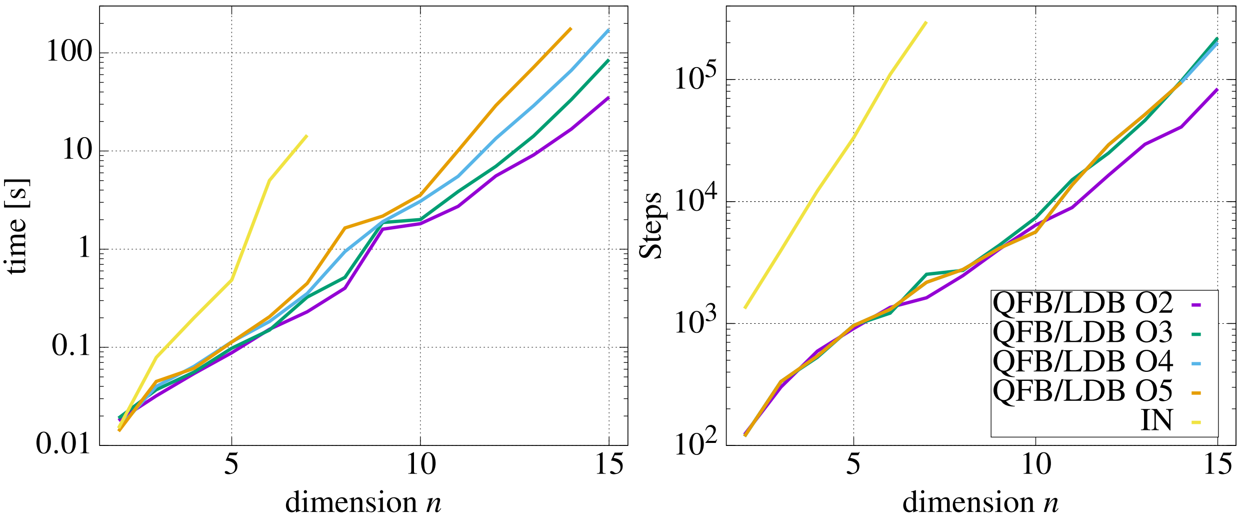

Figure 6.6: Time consumption and number of steps in the optimization of the regular n dimensional Rosenbrock function from Eq. (6.2) at various orders with COSY-GO and QFB/LDB enabled. Additionally, the interval evaluation performance is also shown for comparison.

Models takes longer than the evaluation of lower order Taylor Models. On the other hand, higher order Taylor Models usually provide tighter bounds, which decreases the number of required steps. Regarding the second aspect, the result from Fig. 6.6 on the required steps is rather unusual, because the low order Taylor Model of second order requires fewest steps than the higher order Taylor Models. This phenomenon is specific to certain objective functions, like the Rosenbrock function here, for which the higher order do not bound tighter than the lower order ones.

If we analyze the Rosenbrock function with the artificially increased dependency problem, the second order calculation behaves as one would expect, namely requiring more steps than its higher order counter parts (see Fig. 6.7). The interval evaluation dies of the dependency problem when aiming for $s < 1\text{E-}6$, which is why it is not shown in Fig. 6.7.

For all calculations using QFB/LDB, the global minimum of the generalized Rosenbrock function could be bound to $[-1\text{E}306, 2\text{E-}27]$. This bound is very tight considering the high dimensionality and the required verified computations based on floating point numbers. The optimization variables of all calculations are contained in $[0.999999998, 1.000000002]^n$, which is a box of side length 4E-9 and hence almost three orders of magnitude smaller than the minimum split size. This is because QFB and LDB are not bound to splitting boxes in half, but they can decrease their size as

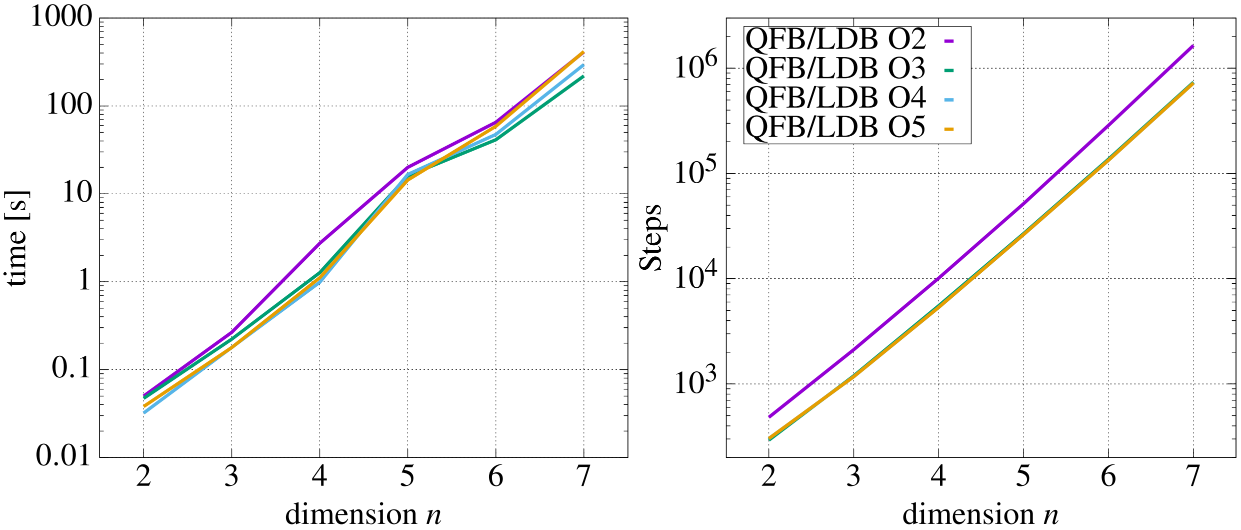

Figure 6.7: Time consumption and number of steps in the optimization of the n dimensional Rosenbrock function with an additional artificial dependency problem $f = f_{2D} - f_{2D} + f_{2D}$ for various Taylor Model orders with COSY-GO and QFB/LDB enabled.

far as their rigorous methods allow them to.

In summary, the example cases of the Rosenbrock functions illustrated that Taylor Model based global optimizers, and COSY-GO in particular, can handle high dimensional objective functions very efficiently. The QFB and LDB avoid the cluster effect, while the Taylor Model evaluation significantly decreases the dependency problem. For the $n = 15$ dimensional non-expanded Rosenbrock function, a reduction of the search volume by a factor of more than 4E157 was accomplished in 84017 steps and less than 36 seconds (see Fig. 6.6) on an Intel®Core$^{\text{TM}}$ i5-7200U CPU 2.5GHz.

6.2 The Lennard-Jones Potential Problem

6.2.1 Introduction

In this section, the capabilities of a Taylor Model based verified global optimizer are demonstrated on the example of finding minimum energy configurations of particles when the well-known Lennard-Jones potential models their pairwise interactions. First, we introduce the Lennard-Jones potential and the principal aspects of the optimization problem. Then, we discuss the setup of the optimization problem for particle configurations in 1D before presenting the results of the verified optimization. Lastly, we discuss the more involved setup of the optimization problem in 2D and 3D and present the associated verified optimization results.

6.2.1.1 The Lennard-Jones Potential U_{LJ}

The 12-6 Lennard-Jones potential

$$U_{\mathrm{LJ}}(r) = 4U_0\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{6.6}$$

is used as a simplified model to describe the interaction between two electrically neutral atoms or molecules with a distance $r > 0$ between them. It was proposed by Lennard-Jones [50] as a specific version of the more general r^{-a}-r^{-b} type potentials he suggested in [49] to model such interactions.

The r^{-12} term represents the strong repulsion of particles at very small distances. The attraction for moderate distances, which quickly decreases with larger distances, is modeled by the r^{-6} term. The parameter U_0 scales the depth of the potential well, which is related to the strength of the interaction between the two particles. The Van-der-Waals radius σ is also referred to as the particle size and indicates where the sign of the potential changes. It represents the distance at which the interaction potential of the two particles assumes the same value as for the configuration where the two particles are infinitely far away from each other.

The potential assumes its single minimum at the equilibrium distance of $r^\star = \sqrt[6]{2}\sigma$. For distances smaller than the equilibrium distance, the potential is strictly monotonically decreasing, and for distances larger than the equilibrium distance, the potential strictly monotonically increasing.

The values σ and U_0 depend on the particles involved in the modeled pairwise interaction. For our analysis, we will only consider one sort of particle corresponding to only one set of values for σ and U_0. To simplify the potential, we consider distances r and σ in units of the equilibrium distance $\sqrt[6]{2}\sigma$, and energy in units of U_0. This yields

$$U_{\mathrm{LJ,lit}} = r^{-12} - 2r^{-6}, \tag{6.7}$$

a form of the Lennard-Jones potential that is often used in literature. However, we offset this potential by one for convenience of the calculations and optimization in this section. So, we define the Lennard-Jones potential of two identical particles with a distance $r > 0$ between them as

$$U_{\mathrm{LJ}}(r) = 1 + r^{-12} - 2r^{-6}, \tag{6.8}$$

so that its single minimum $U^{\star}_{\mathrm{LJ}}$ at $r^{\star}$ yields

$$U^{\star}_{\mathrm{LJ}} = 0 \quad \text{and} \quad r^{\star} = 1. \tag{6.9}$$

In Fig. 6.8, the single pairwise interaction potential between two identical particles from Eq. (6.8) is shown. Note the shallowness of the potential and the large range of function values.

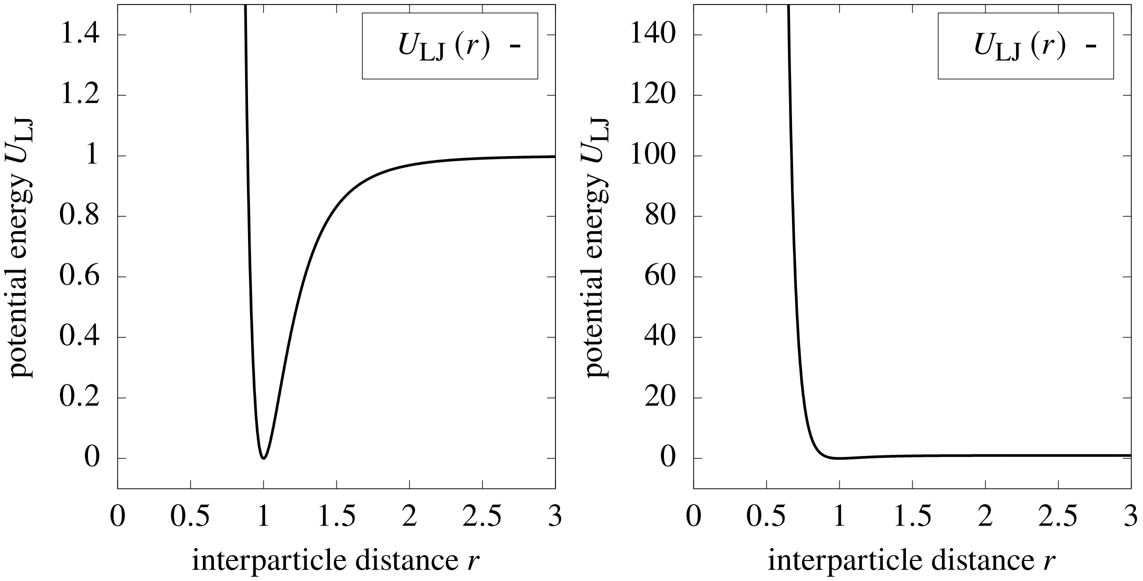

Figure 6.8: The Lennard-Jones potential for a pairwise interaction between two particles as defined in Eq. (6.8). The potential well around the minimum is shown on the left, while the right plot emphasizes its shallowness compared to the steep potential wall for $r < 1$. The potential is offset for the convenience of calculation so that the single minimum $U^{\star}_{\mathrm{LJ}}$ has an energy of zero.

6.2.1.2 Configurations of Particles $\mathcal{S}_k$

Consider a configuration $\mathcal{S}_k$ of k identical particles that have their pairwise interaction modeled by the Lennard-Jones potential from Eq. (6.8). The overall interaction potential U_k of that configuration is given by

$$U_k = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{\mathrm{LJ}}\left(r_{ij}\right), \tag{6.10}$$

the sum of all pairwise interaction potentials U_{LJ}, where $r_{ij} = r_{ji}$ is the distance between the particles p_i and p_j.

The number of pairwise interactions

$$n_{\mathrm{pairs}} = \frac{k\,(k-1)}{2} \tag{6.11}$$

roughly increases with the square of the number of particles k.

Note that

$$U_{k,\mathrm{lit}} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{\mathrm{LJ,lit}}\left(r_{ij}\right) = U_k - n_{\mathrm{pairs}}, \tag{6.12}$$

allowing a direct calculation of the results in terms of $U_{k,\mathrm{lit}}$ from our results in terms of U_k.

We denote the global minimum of U_k by $U_k^\star$. It corresponds to the lowest energy configurations $\mathcal{S}_k^\star$. Those minimum energy configurations are of practical importance for the formation of molecules, assuming nature is sufficiently described by this model and finds the lowest energy configurations when assembling molecules instead of just a local minimum.

The minimum energy configurations and their associated minimum energy depend on whether the configurations are considered in 1D, 2D, or 3D. Thus, we will discuss those cases of minimum energy configuration in different spatial dimensionality n_{dim} separately.

The following notation is used to distinguish between those cases of one, two, or three spatial dimensions when it is relevant. The overall interaction potential is denoted by $U_{k,n_{\mathrm{dim}}}$ and its global minimum by $U_{k,n_{\mathrm{dim}}}^\star$. The corresponding configurations are denoted by $\mathcal{S}_{k,n_{\mathrm{dim}}}$ and $\mathcal{S}_{k,n_{\mathrm{dim}}}^\star$, respectively.

6.2.1.3 The Lennard-Jones Optimization Problem and its Challenges

The goal of the Lennard-Jones optimization problem is the following: Given $k \geq 2$ identical particles with their pairwise interaction modeled by the Lennard-Jones potential from Eq. (6.8) find the global minimum of the overall interaction energy (Eq. (6.10)) and the corresponding optimal configurations. We will conduct this optimization in a verified fashion for configurations in 1D, 2D, and 3D separately.

This problem is particularly interesting and challenging for global optimization because the objective function is non-convex, highly nonlinear, and potentially high dimensional depending on the number of particles k considered. The function values become exceedingly large when two particles get too close to each other, while the actual resulting local minima are often very shallow. This last aspect is reminiscent of the Rosenbrock function and its long and shallow valley with rapidly rising function values outside the valley.

A further prominent aspect of the system is the strong interdependence – changing the position of a single particle of a k-particle configuration changes $k - 1$ interactions and their contributions to the objective function. Accordingly, the dimensionality and hence the complexity of the optimization problem can be increased as desired by simply increasing the number of particles.

This interdependence also complicates the preparation of the optimization including finding appropriate variables and a tight initial search domain that is also guaranteed to contain all global solutions. Often, some regions of the search space can be excluded by showing that they cannot contain the global minimum, which decreases the initial search volume.

To clearly describe our methods of choosing appropriate optimization variables and their corresponding initial search domains despite this complexity, we build them step by step and start with the analysis of 1D Lennard-Jones configurations ($n_{\text{dim}} = 1$). Based on the understanding of the 1D case, we then adapt the choice of variables and their domains to describe the global optimization of configurations in 2D and 3D.

Note that there will be no detailed discussion the trivial cases when $k \leq n_{\text{dim}} + 1$, since obvious configurations exist where every single pairwise interaction potential of the n_{pairs} pairwise Lennard-Jones interactions is at its minimum $U^{\star}_{\text{LJ}} = 0$. In other words, all distances between particles are optimal with $r_{ij} = r^{\star} = 1$. In particular, the configuration $\mathcal{S}^{\star}_{2}$ is a line segment, $S^{\star}_{3}$ is an equilateral triangle in 2D and 3D, and $S^{\star}_{4}$ is a regular tetrahedron in 3D. All of these trivial configurations are of unit length with $U^{\star} = 0$.

6.2.2 Minimum Energy Lennard-Jones Configurations in 1D

To find minimum energy configurations of k particles in 1D using verified global optimization, we first describe the solution space of all possible minimum energy configurations in terms of a set of optimization variables (see Sec. 6.2.2.1). Then, we determine special characteristics of minimum energy configurations in 1D to reduce the initial search domain. Specifically, we calculate an upper bound on the maximum distance of two adjacent particles denoted by $v_{x,\mathrm{UB}}$, and an upper bound on any r_{ij} of the configuration denoted by $r_{k,\mathrm{UB}}$. Further, we determine an upper bound $U_{k,\mathrm{UB}}$ on the minimum energy, which is then used as an initial cutoff value $\mathcal{C}$ for the verified global optimizer (see Sec. 2.6) and for the calculation of a lower bound on r_{ij} denoted $r_{k,\mathrm{LB}}$. Before we present the results of the verified global optimization, we analytically evaluate the distance of adjacent particles in an infinite equidistant configuration as a reference value for the verified global optimization.

6.2.2.1 Coordinate System, Numbering Scheme, and Variable Definition in 1D

Any configuration $\mathcal{S}_{k,\mathrm{1D}}$ of k particles in 1D can be described by placing it on the x axis with the left most particle at the origin. The particles p_i are numbered from 1 to k according to their x position x_i such that

$$x_i \leq x_j \quad \text{for} \quad i < j \quad \text{with} \quad x_1 = 0. \tag{6.13}$$

Note that p_1 is fixed to the origin and the configuration is forced to extend along the positive x axis.

We denote the distance between two adjacent particles p_i and p_{i+1} by

$$v_{x,i} = x_{i+1} - x_i \geq 0 \quad \text{for} \quad i \in \{1, 2, ..., k-1\}, \tag{6.14}$$

and we choose $v_{x,i}$ as the optimization variables.

The number of optimization variables for 1D configurations is denoted by $n_{\mathrm{1D,var}}$, with

$$n_{\mathrm{1D,var}} = k - 1. \tag{6.15}$$

In Sec. 6.2.2.7, we optimize symmetric configurations in 1D for which the number of variables, denoted by $n_{\text{1D,sym,var}}$, are roughly half of $n_{\text{1D,var}}$, with

$$n_{\text{1D,sym,var}} = \begin{cases} (k-1)/2 & \text{if} \quad k \text{ odd} \\ \\ k/2 & \text{if} \quad k \text{ even} \end{cases} , \tag{6.16}$$

because for symmetric configurations

$$v_{x,i} = v_{x,k-i}. \tag{6.17}$$

The distance r_{ij} between any two particles p_i and p_j with $i < j$ can be expressed in terms of $v_{x,i}$ by

$$r_{ij} = x_j - x_i = \sum_{n=i}^{j-1} v_{x,n}. \tag{6.18}$$

6.2.2.2 Upper Bounds $r_{k,\text{UB}}$ and $v_{x,\text{UB}}$ on Inter-Particle Distances of Minimum Energy Configurations in 1D

Only changing the distance $v_{x,l}$ between the two adjacent particles p_l and p_{l+1} moves the right-side subconfiguration composed of the particles p_j with $j > l$ along the x axis, while leaving the left-side subconfiguration of particles p_i with $i \leq l$ unchanged. We denote all r_{ij} that depend on $v_{x,l}$ by $r_{ij}(v_{x,l})$, which is the cases when p_i belongs to the left-side and p_j belongs to the right-side, satisfying $i \leq l < j$ (see Eq. (6.18)).

If $v_{x,l} > 1$ for any l, all $v_{x,l}$-dependent $r_{ij}(v_{x,l} > 1)$ are at least of length $r_{ij}(v_{x,l} > 1) \geq v_{x,l} > 1$. By setting $v_{x,l} = 1$, all $v_{x,l}$-dependent $r_{ij}(v_{x,l} = 1)$ are shortened such that

$$r_{ij}(v_{x,l} > 1) > r_{ij}(v_{x,l} = 1) \geq 1, \tag{6.19}$$

which monotonically lowers U_{LJ} of every involved particle pair (see monotonicity argument of single Lennard-Jones potential U_{LJ} in Sec. 6.2.1.1), while leaving the uninvolved interaction energies unchanged (see Fig. 6.9).

This monotonically lowers $U_{k,1D}$ such that any configuration $\mathcal{S}_{k,\text{1D}}$ with any $v_{x,i} > 1$ cannot be optimal. Thus, $v_{x,i} = 1$ is the upper bound on any $v_{x,i}$ in optimal configurations $\mathcal{S}^{\star}_{k,\text{1D}}$. We denote

p_1
p_2
p_l
...
p_{l+1}
p_k
x
$v_{x,l}>1$
p_1
p_2
p_l
p_{l+1}
p_k
x
$v_{x,l}=1$

Figure 6.9: The particles p_i are numbered according to their x position. The variable $v_{x,l}$ denotes the distance between particle p_l and p_{l+1}. A configuration with any $v_{x,l} > 1$ (left picture) is never optimal, because the U_k can always be lowered by setting $v_{x,l} = 1$ (right picture).

this upper bound with $v_{x,\mathrm{UB}}$. So,

$$v_{x,\mathrm{UB}} = 1. \tag{6.20}$$

Additionally, this also yields an upper bound $r_{k,\mathrm{UB}}$ for any r_{ij} within a minimum energy configuration of k particles in 1D, with

$$r_{k,\mathrm{UB}} = \sum_{n=1}^{k-1} v_{x,\mathrm{UB}} = k - 1. \tag{6.21}$$

6.2.2.3 The Upper Bound $U_{k,\mathrm{UB}}$ on the Minimum Energy

The potential U_{k,n_dim} of any configuration $\mathcal{S}_{k,n_\mathrm{dim}}$ can serve as an upper bound on the minimum energy $U^\star_{k,n_\mathrm{dim}}$ because by definition U_{k,n_dim} always satisfies

$$U_{k,n_\mathrm{dim}} \geq U^\star_{k,n_\mathrm{dim}}. \tag{6.22}$$

We denote upper bound configurations by $\mathcal{S}_{k,n_\mathrm{dim},\mathrm{UB}}$ and their associated potential by $U_{k,n_\mathrm{dim},\mathrm{UB}}$. A good $\mathcal{S}_{k,n_\mathrm{dim},\mathrm{UB}}$ yields a tight upper bound on the minimum energy, which can then be used as an initial cutoff value $\mathcal{C}$ for the verified global optimizer and for the determination of a good lower bound $r_{k,\mathrm{LB}}$ on any r_{ij} (see Sec. 6.2.2.4). Note that the argumentation so far is not specific to only 1D configurations and will be used later on for the multidimensional cases.

For 1D, a good $\mathcal{S}_{k,\mathrm{1D},\mathrm{UB}}$ is given by mirroring the first half of $\mathcal{S}^\star_{k-1,\mathrm{1D}}$ onto its second half to replace it such that the resulting $\mathcal{S}_{k,\mathrm{1D},\mathrm{UB}}$ is mirror symmetric. The mirror is placed slightly off-center to generate a k-particle upper bound configuration from the optimal $(k-1)$-particle configuration. Specifically, the mirror is placed on particle $p_{(k+1)/2}$ when k is odd, and in the middle between particle $p_{k/2}$ and $p_{k/2+1}$ when k is even. This mirror symmetric configuration yields an upper bound $U_{k,\mathrm{1D},\mathrm{UB}}$ on $U^\star_{k,\mathrm{1D}}$.

6.2.2.4 The Lower Bound $r_{k,\mathrm{LB}}$ on r_{ij}

Determining a lower bound $r_{k,\mathrm{LB}}$ on r_{ij} is critical for the verified optimization. It is essential to formally show that $r_{ij} \geq r_{k,\mathrm{LB}} > 0$ because the objective function is not defined for $r_{ij} = 0$. Additionally, a lower bound often helps reducing the initial search domain of the optimization variables.

To determine $r_{k,\mathrm{LB}}$ on r_{ij}, we first determine the inverse relation between U_{LJ} and r over the two domain sections, where the relation is bijective, namely, $r \leq 1$ and $r \geq 1$, denoted by $r_{\min}$ and $r_{\max}$, respectively. Solving the quadratic equation hidden in $U_{\mathrm{LJ}}(r)$ from Eq. (6.8) yields

$$r(U_{\mathrm{LJ}}) = \begin{cases} r_{\min}(U_{\mathrm{LJ}}) = \left(1 + \sqrt{U_{\mathrm{LJ}}}\right)^{-\frac{1}{6}} & \text{for} \quad 0 \leq U_{\mathrm{LJ}} \\ r_{\max}(U_{\mathrm{LJ}}) = \left(1 - \sqrt{U_{\mathrm{LJ}}}\right)^{-\frac{1}{6}} & \text{for} \quad 0 \leq U_{\mathrm{LJ}} \leq 1 \end{cases}, \tag{6.23}$$

where $r_{\min} \leq 1$ is monotonically decreasing with increasing U_{LJ} and $r_{\max} \geq 1$ is monotonically increasing with increasing U_{LJ}. Fig. 6.10 illustrates $r(U_{\mathrm{LJ}})$.

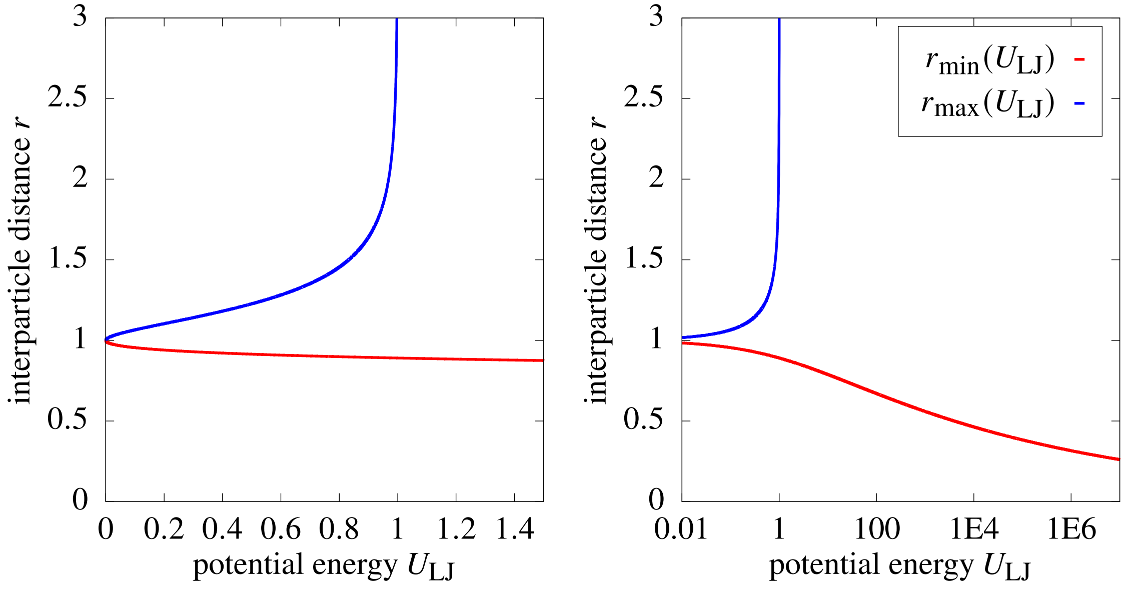

Figure 6.10: The relation between U_{LJ} and the corresponding inter-particle distance(s). Note that $r_{\max}(U_{\mathrm{LJ}})$ is only defined for $U_{\mathrm{LJ}} \leq 1$. $r_{\min}(U_{\mathrm{LJ}})$ is only decreasing very slowly with increasing U_{LJ} as the logarithmic plot on the right shows.

The potential U_k from any configuration $\mathcal{S}_k$ satisfies

$$U_k \geq U_{\mathrm{LJ}}(r_{ij}) \tag{6.24}$$

for any r_{ij} of $\mathcal{S}_k$. In other words, U_k is an upper bound on all pairwise interactions $U_{\mathrm{LJ}}(r_{ij})$ of $\mathcal{S}_k$.

Having the monotonicity property of r_{min}, an upper bound on $U_{\mathrm{LJ}}(r_{ij})$ yields a lower bound on r_{ij} via Eq. (6.23). Thus, $r_{\mathrm{min}}(U_k)$ is a lower bound on all r_{ij} of $\mathcal{S}_k$. More generally, it is a lower bound on all r_{ij} in any $\mathcal{S}'_k$ for which $U'_k \leq U_k$, specifically for $U'_k = U^{\star}_k$. Thus, we use

$$r_{k,\mathrm{LB}} = r_{\mathrm{min}}(U_{k,\mathrm{UB}}), \tag{6.25}$$

with $U_{k,\mathrm{UB}}$ from Sec. 6.2.3.3 as a lower bound on any r_{ij} in any configuration with $U_k \leq U_{k,\mathrm{UB}}$, which includes all configurations considered in the optimization.

Note that this method is independent of n_{dim}. Hence, we will also use it for configurations in 2D and 3D later on to calculate a lower bound on any r_{ij}.

6.2.2.5 The Infinite 1D Equidistant Configuration

Before we investigate finite minimum energy configurations of $k \geq 2$ particles in 1D, we derive the minimum energy state of an infinite equidistant 1D configurations [13, 52]. This one dimensional optimization problem can be solved analytically and shall serve as a reference for the results of verified optimization of finite minimum energy configurations in 1D in Sec. 6.2.2.6 and Sec. 6.2.2.7.

We start our derivation by considering k particles on a line, where the interaction between the particles is modeled by the Lennard-Jones potential U_{LJ} from Eq. (6.8). The distance between any two adjacent particles is a constant value $r > 0$. The overall potential of such a configuration is

$$U_k(r) = \sum_{j=1}^{k-1} (k-j)\, U_{\mathrm{LJ}}(jr), \tag{6.26}$$

where $(k - j)$ indicates how often an inter-particle distance of length jr occurs in the configuration.

Expanding the overall potential yields

$$\begin{aligned} U_k(r) &= \sum_{j=1}^{k-1}(k-j)\left(1 + j^{-12}r^{-12} - 2j^{-6}r^{-6}\right) && (6.27)\\ &= \sum_{j=1}^{k-1}(k-j) + \left(k\sum_{j=1}^{k-1} j^{-12} - \sum_{j=1}^{k-1} j^{-11}\right) r^{-12} - 2\left(k\sum_{j=1}^{k-1} j^{-6} - \sum_{j=1}^{k-1} j^{-5}\right) r^{-6}, && (6.28)\end{aligned}$$

where we denote the summation of the j^{-s} by the function $\zeta_l(s)$ with

$$\zeta_l(s) = \sum_{j=1}^{l} j^{-s} = 1 + 2^{-s} + 3^{-s} + ... + l^{-s}, \quad \text{for} \quad s > 0,\ l \geq 1. \tag{6.29}$$

The function $\zeta_l(s)$ satisfies

$$\zeta_l(s) < \zeta_{l+1}(s) \quad \text{and} \tag{6.30}$$

$$1 < \zeta_l(s_2) < \zeta_l(s_1) \quad \text{for} \quad 0 < s_1 < s_2 \quad \text{and} \quad \forall l > 1. \tag{6.31}$$

For $l = 1$,

$$\zeta_1(s) = \sum_{j=1}^{1} j^{-s} = 1^{-s} = 1. \tag{6.32}$$

Using Eq. (6.29), we rewrite Eq. (6.28) as

$$U_k(r) = \sum_{j=1}^{k-1} (k - j) + \left[k\zeta_{k-1}(12) - \zeta_{k-1}(11)\right] r^{-12} - 2\left[k\zeta_{k-1}(6) - \zeta_{k-1}(5)\right] r^{-6}. \tag{6.33}$$

To find r that minimizes $U_k(r)$, we solve

$$\left.\frac{d\hat{U}_k\,(r)}{dr}\right|_{r=r_\star\in\mathbb{R}^+} = 0 \quad \text{for} \quad r_\star > 0. \tag{6.34}$$

Specifically,

$$\begin{aligned} 0 = \left.\frac{d\hat{U}_k\,(r)}{dr}\right|_{r=r_\star\in\mathbb{R}^+} &= -12r_\star^{-13}\left[k\zeta_{k-1}(12) - \zeta_{k-1}(11)\right] + 12r_\star^{-7}\left[k\zeta_{k-1}(6) - \zeta_{k-1}(5)\right] \\ \Rightarrow \quad r_\star &= \left(\frac{k\zeta_{k-1}(12) - \zeta_{k-1}(11)}{k\zeta_{k-1}(6) - \zeta_{k-1}(5)}\right)^{\frac{1}{6}} = \left(\frac{\zeta_{k-1}(12) - \frac{1}{k}\zeta_{k-1}(11)}{\zeta_{k-1}(6) - \frac{1}{k}\zeta_{k-1}(5)}\right)^{\frac{1}{6}}. \end{aligned} \tag{6.35}$$

As a cross-check, we evaluate $r_\star$ for $k = 2$ using Eq. (6.32) with

$$r_\star = \left(\frac{\zeta_1(12) - \frac{1}{2}\zeta_1(11)}{\zeta_1(6) - \frac{1}{2}\zeta_1(5)}\right)^{\frac{1}{6}} = \left(\frac{1 - \frac{1}{2}}{1 - \frac{1}{2}}\right)^{\frac{1}{6}} = 1, \tag{6.36}$$

which agrees with Eq. (6.9) as expected.

As a second calculation, we evaluate $r_\star$ for $k = 3$ with

$$r_\star = \left(\frac{1 + 2^{-12} - \frac{1}{3} - \frac{1}{3}2^{-11}}{1 + 2^{-6} - \frac{1}{3} - \frac{1}{3}2^{-5}}\right)^{\frac{1}{6}} = \left(\frac{2731}{2752}\right)^{\frac{1}{6}} \in [0.9987241350, 0.9987241351]\,. \tag{6.37}$$

For the limit of $k \to \infty$, we note that $\zeta_\infty(s)$ corresponds to the Riemann zeta function [89]

$$\zeta(s) = \sum_{j=1}^{\infty} \frac{1}{j^s}, \tag{6.38}$$

where

$$\zeta(2) = \frac{\pi^2}{6} \approx 1.644934 \tag{6.39}$$

is known from the Basel problem [7].

With this and under consideration of Eq. (6.31), any $\zeta(s)$ with $s > 2$ converges to values smaller than $\pi^2/6$ but larger than 1. Specifically,

$$\zeta(6) = \frac{\pi^6}{945} \quad \text{and} \quad \zeta(12) = \frac{691\pi^{12}}{638512875}. \tag{6.40}$$

Accordingly, the limit of

$$\begin{aligned} \lim_{k\to\infty} r_\star &= \left(\frac{\zeta_{k-1}(12) - \frac{1}{k}\zeta_{k-1}(11)}{\zeta_{k-1}(6) - \frac{1}{k}\zeta_{k-1}(5)} \right)^{\frac{1}{6}} = \left(\frac{\zeta(12)}{\zeta(6)} \right)^{\frac{1}{6}} \\ &= \pi \cdot \sqrt[6]{\frac{691}{675675}} \in [0.9971792638858069273, 0.9971792638858069274] \,. \end{aligned} \tag{6.41}$$

The upper bound $v_{x,\mathrm{UB}} = 1$ from Sec. 6.2.2.2 already told us that $r_\star \le 1$. Finding $r_\star$ so close to 1 illustrates the steepness of the Lennard-Jones potential for $r < 1$.

6.2.2.6 The Verified Global Optimization Results for Configurations of k Particles in 1D

As discussed above in Sec. 6.2.2.1, place the configuration on the positive x axis and number the particles p_i from 1 to k according to their x position x_i such that

$$x_i \le x_j \quad \text{for} \quad i < j \quad \text{with} \quad x_1 = 0. \tag{6.42}$$

The distances

$$v_{x,i} = x_{i+1} - x_i \ge 0 \quad \text{for} \quad i \in \{1, 2, ..., k-1\}, \tag{6.43}$$

serve as optimization variables, which yields

$$n_{\mathrm{1D,var}} = k - 1 \tag{6.44}$$

optimization variables, as previously noted in Eq. (6.15).

The distances r_{ij} for the objective function are calculated from the optimization variables $v_{x,i}$ according to Eq. (6.18).

The initial search domain of the optimization is determined by the upper and lower bound ($v_{x,\mathrm{UB}}$ and $v_{x,\mathrm{LB}}$) on the distances of two adjacent particles from Sec. 6.2.2.2 and Sec. 6.2.2.4. Specifically,

$$v_{x,i} \in [v_{x,\mathrm{LB}}, v_{x,\mathrm{UB}}] = [r_{k,\mathrm{1D,LB}}, 1] \quad \text{for} \quad i \in \{1, 2, ..., k-1\}. \tag{6.45}$$

While the upper bound $v_{x,\mathrm{UB}} = 1$ remains unchanged for all k, the lower bound $v_{x,\mathrm{LB}} = r_{k,\mathrm{1D,LB}}$ depends on $U^{\star}_{k-1}$ and $U_{k,\mathrm{UB}}$ (see Sec. 6.2.2.3 and Sec. 6.2.2.4).

We start from $k = 2$, which represents a single pair of particles. From Eq. (6.9) we know the solution of this trivial case is

$$U^{\star}_{2,\mathrm{1D}} = 0 \quad \text{and} \quad v^{\star}_{x,1} = 1. \tag{6.46}$$

We follow the method in Sec. 6.2.2.3 to construct $\mathcal{S}_{3,\mathrm{1D,UB}}$. The mirror is placed on the particle p_2 to mirror-copy $S^{\star}_{2,\mathrm{1D}}$ to the right, which yields $\mathcal{S}_{3,\mathrm{1D,UB}}$ with $v_{x,1} = v_{x,2} = 1$. Thus,

$$U_{3,\mathrm{1D,UB}} = U_{\mathrm{LJ}}(v_{x,1}) + U_{\mathrm{LJ}}(v_{x,2}) + U_{\mathrm{LJ}}(v_{x,1} + v_{x,3}) \tag{6.47}$$

$$= 2U_{\mathrm{LJ}}(1) + U_{\mathrm{LJ}}(2) = U_{\mathrm{LJ}}(2) \leq 0.968994140625. \tag{6.48}$$

Using this upper bound in Eq. (6.25) together with the equation for $r_{\min}$ in Eq. (6.23) according to the method in Sec. 6.2.2.4, we have

$$r_{3,\mathrm{1D,LB}} = r_{\min}(U_{3,\mathrm{1D,UB}}) \geq 0.89206405909675. \tag{6.49}$$

Thus, the initial search domain for $k = 3$ is

$$v_{x,i} \in [v_{x,\mathrm{LB}}, v_{x,\mathrm{UB}}] = [0.89206405909675, 1] \quad \text{for} \quad i \in \{1, 2\}. \tag{6.50}$$

Starting with $k = 3$, we iteratively perform the verified optimization for the k particle case and use the result to calculate $U_{k+1,\mathrm{1D,UB}}$ (see Sec. 6.2.2.3), $r_{k+1,\mathrm{1D,LB}}$ (see Sec. 6.2.2.4), and the initial search domain for the $k + 1$ particle case.

The verified global optimization is performed with the Taylor Model based verified global optimizer COSY-GO [63, 64] in its most advanced setting with QFB/LDB enabled. An attempt to run the optimization using Interval evaluations already fails for the simplest non-trivial case of three particles in 1D. Hence, this section will only run COSY-GO in its most advanced setting.

The principle algorithm of the verified global optimizer was outlined in Sec. 2.6. As a stopping condition, we use the threshold length $s_{\text{min}} = 10^{-6}$ for all computation with COSY-GO in this chapter. A box under investigation is split into smaller boxes for further investigation unless the box is too small with all the side-lengths less than s_{min}. The upper bound $U_{k,\text{1D,UB}}$ from Sec. 6.2.2.3 is used as an initial cutoff value $\mathcal{C}$ for the optimizer.

Tab. 6.1 shows the results of the verified optimization. The resulting optimized variables $v^{\star}_{x,i}$ are listed in Tab. 6.2 and shown in Fig. 6.11. Note that the floating point inaccuracies begin to accumulate with increasing k such that the bounding of $U^{\star}_{k,\text{1D}}$ in Tab. 6.1 gets less and less tight.

Table 6.1: Verified global optimization results for $U^{\star}_{k,\text{1D}}$. The k particles form n_{pairs} pairwise interactions. The upper bound $U_{k,\text{1D,UB}}$ on the minimum energy (see Sec. 6.2.2.3) was used to calculate $r_{k\text{1D,UB}}$, which sets the lower bound of the initial search domain (see Sec. 6.2.2.4 and Eq. (6.45)). The optimizer COSY-GO with QFB/LDB enabled was operated with Taylor Models of third order. The number of remaining boxes with all side-lengths $s < s_{\text{min}}$ is denoted by n_{fin}.

k	n_{pairs}	n_{fin}	$r_{k,\text{1D,LB}}$	$U_{k,\text{1D,UB}}$	$U^{\star}_{k,\text{1D}}$
3	3	1	0.89206405909675	0.96899414062500	0.968875869644^{82}_{77}
4	6	1	0.84674764946679	2.93492994152107	2.934863711898^{21}_{13}
5	10	1	0.81433688881131	5.90034265454486	$5.90034204308^{601}_{589}$
6	15	1	0.78913201707003	9.86568839948674	9.865688070463^{48}_{29}
7	21	1	0.76858739727728	14.83099005904041	$14.830990045365^{67}_{39}$
8	28	1	0.75130633425847	20.79627461693932	$20.796274609476^{71}_{30}$
9	36	1	0.73643514933217	27.76155137645473	$27.761551375^{70017}_{69956}$
10	45	1	0.72341340859215	35.72682430087453	$35.72682430044^{963}_{875}$
11	55	1	0.71185356795324	44.69209518551362	$44.69209518543^{888}_{767}$
12	66	1	0.70147671889115	54.65736491976443	$54.6573649197^{2036}_{1862}$
13	78	1	0.69207564364786	65.62263397159721	$65.62263397158^{539}_{307}$
14	91	1	0.68349232451704	77.58790260143020	$77.5879026014^{2233}_{1933}$
15	105	1	0.67560361208721	90.55317096078578	$90.5531709607^{8190}_{7808}$

Table 6.2: Verified global optimization results for configurations of k particles in 1D. The variable $v^{\star}_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, with $1 \leq i < k$, and the mirror symmetry can be observed.

k	i	$v^{\star}_{x,i}$
3	1	0.9987241^{7}_{0}
3	2	0.9987241^{7}_{0}
4	1	0.99864^{309}_{299}
4	2	0.997396^{47}_{38}
4	3	0.99864^{309}_{299}
5	1	0.998632^{38}_{26}
5	2	0.997306^{83}_{71}
5	3	0.997306^{83}_{71}
5	4	0.998632^{38}_{26}
6	1	0.998630^{18}_{03}
6	2	0.997294^{22}_{07}
6	3	0.997215^{18}_{03}
6	4	0.997294^{22}_{07}
6	5	0.998630^{18}_{03}
7	1	0.998629^{59}_{40}
7	2	0.997291^{48}_{30}
7	3	0.99720^{200}_{182}
7	4	0.99720^{200}_{182}
7	5	0.997291^{48}_{30}
7	6	0.998629^{59}_{40}
8	1	0.998629^{40}_{18}
8	2	0.997290^{70}_{48}
8	3	0.99719^{907}_{885}
8	4	0.997188^{63}_{41}
8	5	0.99719^{907}_{885}
8	6	0.997290^{70}_{48}
8	7	0.998629^{40}_{18}
9	1	0.998629^{34}_{07}
9	2	0.997290^{44}_{18}
9	3	0.99719^{822}_{795}
9	4	0.997185^{63}_{36}
9	5	0.997185^{63}_{36}
9	6	0.99719^{822}_{795}
9	7	0.997290^{44}_{18}
9	8	0.998629^{34}_{07}
10	1	0.998629^{33}_{01}
10	2	0.997290^{35}_{04}
10	3	0.997197^{92}_{61}
10	4	0.997184^{74}_{42}
10	5	0.997182^{58}_{27}
10	6	0.997184^{74}_{42}
10	7	0.997197^{92}_{61}
10	8	0.997290^{35}_{04}
10	9	0.998629^{33}_{01}
11	1	0.99862^{934}_{897}
11	2	0.9972^{9033}_{8996}
11	3	0.997197^{82}_{46}
11	4	0.997184^{43}_{07}
11	5	0.997181^{68}_{32}
11	6	0.997181^{68}_{32}
11	7	0.997184^{43}_{07}
11	8	0.997197^{82}_{46}
11	9	0.9972^{9033}_{8996}
11	10	0.99862^{934}_{897}
12	1	0.99862^{937}_{892}
12	2	0.9972^{9034}_{8990}
12	3	0.997197^{80}_{36}
12	4	0.99718^{433}_{389}
12	5	0.99718^{138}_{094}
12	6	0.997180^{79}_{34}
12	7	0.99718^{138}_{094}
12	8	0.99718^{433}_{389}
12	9	0.997197^{80}_{36}
12	10	0.9972^{9034}_{8990}
12	11	0.99862^{937}_{892}
13	1	0.99862^{940}_{889}
13	2	0.9972^{9036}_{8985}
13	3	0.997197^{80}_{30}
13	4	0.99718^{430}_{380}
13	5	0.99718^{127}_{076}
13	6	0.9971^{8047}_{7997}
13	7	0.9971^{8047}_{7997}
13	8	0.99718^{127}_{076}
13	9	0.99718^{430}_{380}
13	10	0.997197^{80}_{30}
13	11	0.9972^{9036}_{8985}
13	12	0.99862^{940}_{889}
14	1	0.99862^{943}_{885}
14	2	0.9972^{9039}_{8981}
14	3	0.997197^{82}_{25}
14	4	0.99718^{430}_{373}
14	5	0.99718^{124}_{066}
14	6	0.9971^{8036}_{7979}
14	7	0.9971^{8016}_{7959}
14	8	0.9971^{8036}_{7979}
14	9	0.99718^{124}_{066}
14	10	0.99718^{430}_{373}
14	11	0.997197^{82}_{25}
14	12	0.9972^{9039}_{8981}
14	13	0.99862^{943}_{885}
15	1	0.99862^{946}_{881}
15	2	0.9972^{9042}_{8977}
15	3	0.997197^{85}_{20}
15	4	0.99718^{432}_{368}
15	5	0.99718^{124}_{060}
15	6	0.9971^{8033}_{7969}
15	7	0.9971^{8005}_{7941}
15	8	0.9971^{8005}_{7941}
15	9	0.9971^{8033}_{7969}
15	10	0.99718^{124}_{060}
15	11	0.99718^{432}_{368}
15	12	0.997197^{85}_{20}
15	13	0.9972^{9042}_{8977}
15	14	0.99862^{946}_{881}

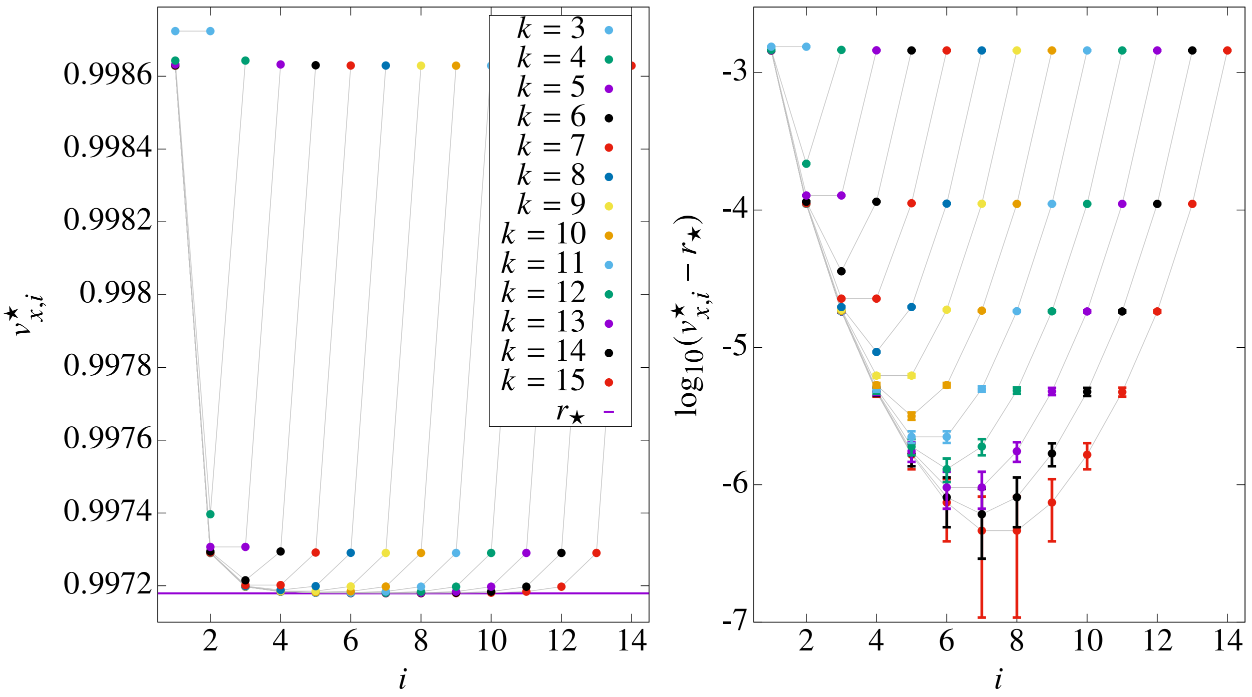

Figure 6.11: The plots show the values of the optimization variables $v^{\star}_{x,i}$ of the minimum energy configuration of k particles in 1D that resulted from the verified global optimization using COSY-GO. The minimum energy configuration is mirror symmetric with the middlemost distances between adjacent particles asymptotically approaching $r_\star \approx 0.998724135$, the solution of the infinite equidistant configuration from Eq. (6.41). The right plot shows the logarithm of the difference between the calculated distances from the verified optimization and $r_\star$. The ranges reflect the side-length of the remaining box.

The left plot of Fig. 6.11 shows that the distance between adjacent particles barley changes in the middle of the configuration. However, the logarithmic plot on the right clearly shows that the distances get shorter towards the center of the configuration.

The ranges in the right plot correspond to the side-length of the remaining box and its position. We observe that the optimal configurations are symmetric and the $v^{\star}_{x,i}$ asymptotically approaches $r_\star$ (Eq. (6.41)), the solution for the infinite equidistant configuration in 1D from Sec. 6.2.2.5.

In Tab. 6.2, we see that the verified bounds for $k = 3$ agree with the calculation of the equidistant configuration in Eq. (6.37).

Tab. 6.3 lists the performance of COSY-GO for different Taylor Model orders. Usually, the higher the order of computation, the tighter the bounding and the lower the required number of steps, which is what we see in Tab. 6.3. At the same time, higher order computations are more time demanding per step. These two factors, the computation time per step and the required number of steps, do not scale the same way with higher orders. For this particular example, calculations of order three (O3) are the most time efficient.

Table 6.3: Performance of verified global optimization using COSY-GO with QFB/LDB enabled on minimum energy search of a 1D configuration of k particles. The Taylor Model orders are denoted by 'O'. Since QFB requires a minimum order of two, order one calculations are not listed.

			Computation time [s]				Steps			
k	$n_{\text{1D,var}}$	n_{pairs}	O2	O3	O4	O5	O2	O3	O4	O5
3	2	3	0.011	0.017	0.010	0.015	14	10	10	8
4	3	6	0.014	0.014	0.013	0.016	24	15	14	14
5	4	10	0.027	0.021	0.021	0.016	36	23	22	18
6	5	15	0.044	0.050	0.026	0.047	57	29	28	25
7	6	21	0.085	0.043	0.034	0.079	83	34	33	33
8	7	28	0.203	0.136	0.058	0.163	130	43	42	41
9	8	36	0.560	0.136	0.101	0.332	236	60	53	52
10	9	45	1.728	0.330	0.179	0.616	475	98	66	65
11	10	55	3.373	0.747	0.438	1.828	925	176	90	81
12	11	66	7.191	0.962	0.969	3.068	1908	294	133	98
13	12	78	17.99	3.796	5.121	7.399	3975	454	292	131
14	13	91	31.96	7.276	9.480	8.928	7690	657	492	160
15	14	105	64.35	8.550	18.25	18.91	15902	994	795	199

Compared to the O2 calculation, the longer computation times of the O3 calculations per step are overcompensated by the tighter bounding and the associated reduction in the number of steps required for the verified optimization. With higher order calculations, the number of steps can be reduced even further, but the computation time per step increases significantly, such that O4 is the second most time efficient and O5 the third most time efficient despite their further reduction of calculation steps.

As k increases, the complexity of the problem increases quadratically as n_{pairs} indicates. The time efficiency of the different computation orders can change with the complexity of the objective function.

6.2.2.7 The Verified Global Optimization Results for Symmetric Configurations of k Particles in 1D

Assuming that the Lennard-Jones minimum energy configurations in 1D are indeed symmetric, this section analyzes the associated optimization problem. Considering symmetric 1D configurations roughly reduces the number of optimization variables to describe the configurations to half, since

$$v_{k-i} = v_i. \tag{6.51}$$

This yields

$$n_{\text{1D,sym,var}} = \begin{cases} (k-1)/2 & \text{if} \quad k \text{ odd} \\ k/2 & \text{if} \quad k \text{ even} \end{cases}, \tag{6.52}$$

optimization variables as previously noted in Eq. (6.16).

All other parameters of the optimization like the initial search domain and the method of calculated $U_{k,\text{1D,UB}}$ from Sec. 6.2.2.3 remain unchanged.

As above in Sec. 6.2.2.6, we start with the trivial $\mathcal{S}^{\star}_{2,\text{1D}}$ to determine $U_{3,\text{1D,UB}}$ (see Eq. (6.48)). We then use this upper bound and $U^{\star}_{k,\text{1D}}$ in Eq. (6.25) to determine $r_{3,\text{LB,1D}}$ (see Eq. (6.49)) and with this the initial search domain (see Eq. (6.50)).

Starting with $k = 3$, we perform the verified optimization for the k particle case and use the result to calculate $U_{k+1,\text{1D,UB}}$ (see Sec. 6.2.2.3), $r_{k+1,\text{LB,1D}}$ (see Sec. 6.2.2.4), and the initial search domain for the $k+1$ particle case.

Tab. 6.4 shows the verified results of the optimization using Taylor Models of order three. Note that the floating point inaccuracies begin to accumulate with increasing k such that the bounding of $U^{\star}_{k,\text{1D}}$ in Tab. 6.1 gets less and less tight. For $k \geq 23$, the number of final boxes increases drastically. Due to the high dimensionality, U_k gets so shallow over the $n_{\text{1D,sym,var}}$ dimensional domain that the limit of the floating point accuracy prevents narrowing down the minimum to a single final box of side-lengths $s_{\min} < 10^{-6}$.

The resulting values of the $v^{\star}_{x,i}$ are listed in Tab. 6.5 and Tab. 6.6 below. For $k \geq 23$, all the resulting the final boxes are represented by one big box that contains all of them rigorously. Hence, the presented $v^{\star}_{x,i}$ are the side-lengths of this big box.

Table 6.4: Verified global optimization results on the minimum energy $U^{\star}_{k,1\mathrm{D}}$ of symmetric configurations. The upper bound $U_{k,1\mathrm{D,UB}}$ on the minimum energy (see Sec. 6.2.2.3) was used to calculate $r_{k1\mathrm{D,UB}}$, which sets the lower bound of the initial search domain (see Sec. 6.2.2.4 and Eq. (6.45)). The optimizer COSY-GO with QFB/LDB enabled was operated with Taylor Models of third order. The number of remaining boxes with all side-lengths $s < s_{\min}$ is denoted by n_{fin}.

k	n_{pairs}	n_{fin}	$r_{k,1\mathrm{D,LB}}$	$U_{k,1\mathrm{D,UB}}$	$U^{\star}_{k,1\mathrm{D}}$
3	3	1	0.89206405909675	0.96899414062500	0.968875869644^{82}_{77}
4	6	1	0.84674764946679	2.93492994152101	2.934863711898^{21}_{13}
5	10	1	0.81433688881131	5.90034265454486	$5.90034204308^{601}_{589}$
6	15	1	0.78913201707003	9.86568839948674	9.865688070463^{48}_{29}
7	21	1	0.76858739727728	14.83099005904040	$14.830990045365^{67}_{40}$
8	28	1	0.75130633425847	20.79627461693931	$20.796274609476^{71}_{32}$
9	36	1	0.73643514933217	27.76155137645471	$27.761551375^{70016}_{69958}$
10	45	1	0.72341340859215	35.72682430087450	$35.72682430044^{963}_{878}$
11	55	1	0.71185356795324	44.69209518551359	$44.69209518543^{886}_{772}$
12	66	1	0.70147671889115	54.65736491976435	$54.6573649197^{2037}_{1870}$
13	78	1	0.69207564364786	65.62263397159714	$65.62263397158^{541}_{315}$
14	91	1	0.68349232451704	77.58790260143012	$77.5879026014^{2233}_{1941}$
15	105	1	0.67560361208721	90.55317096078569	$90.5531709607^{8190}_{7818}$
16	120	1	0.66831174278786	104.51843914138416	$104.5184391413^{8084}_{7582}$
17	136	1	0.66153786375283	119.48370720063325	$119.4837072006^{3014}_{2375}$
18	153	1	0.65521749055526	135.44897517554651	$135.4489751755^{4319}_{3522}$
19	171	1	0.64929724579321	152.41424309061438	$152.4142430906^{1040}_{0067}$
20	190	2	0.64373246921844	170.37951096241994	$170.3795109624^{1495}_{0320}$
21	210	1	0.63848543480271	189.34477880243068	$189.3447788024^{1960}_{0544}$
22	231	2	0.63352399921567	209.31004661870341	$209.3100466186^{9639}_{7876}$
23	253	2048	0.62882056259331	230.27531441704119	$230.2753144170^{2488}_{0327}$
24	276	4096	0.62435125909424	252.24058220167566	$252.240582201^{60723}_{58147}$
25	300	4096	0.62009531904927	275.20584997563225	$275.2058499755^{4147}_{1139}$
26	325	8192	0.61603456097294	299.17111774124430	$299.1711177411^{4182}_{0645}$

Table 6.5: Verified global optimization results for symmetric configurations of k particles in 1D for $k = 3$ to $k = 20$. $v^{\star}_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, and p_{k-i-1} and p_{k-i}. The results for $k = 21$ to $k = 26$ are listed in Tab. 6.6.

k	i	$v^{\star}_{x,i}$
3	1	0.9987241^{6}_{1}
4	1	0.9986430^{7}_{0}
4	2	0.997396^{47}_{38}
5	1	0.998632^{36}_{28}
5	2	0.997306^{81}_{72}
6	1	0.998630^{16}_{05}
6	2	0.997294^{20}_{09}
6	3	0.997215^{18}_{03}
7	1	0.998629^{56}_{43}
7	2	0.997291^{45}_{32}
7	3	0.997201^{98}_{85}
8	1	0.998629^{37}_{21}
8	2	0.997290^{67}_{51}
8	3	0.99719^{904}_{888}
8	4	0.997188^{63}_{42}
9	1	0.998629^{30}_{11}
9	2	0.997290^{40}_{21}
9	3	0.99719^{818}_{799}
9	4	0.997185^{59}_{40}
10	1	0.998629^{28}_{06}
10	2	0.997290^{31}_{09}
10	3	0.997197^{88}_{66}
10	4	0.997184^{69}_{47}
10	5	0.997182^{58}_{27}
11	1	0.998629^{29}_{02}
11	2	0.997290^{27}_{02}
11	3	0.997197^{77}_{51}
11	4	0.997184^{38}_{12}
11	5	0.997181^{63}_{37}
12	1	0.99862^{930}_{899}
12	2	0.9972^{9027}_{8997}
12	3	0.997197^{73}_{43}
12	4	0.99718^{427}_{396}
12	5	0.997181^{31}_{01}
12	6	0.997180^{78}_{35}
13	1	0.99862^{932}_{896}
13	2	0.9972^{9028}_{8993}
13	3	0.997197^{73}_{37}
13	4	0.99718^{423}_{387}
13	5	0.99718^{119}_{084}
13	6	0.997180^{40}_{04}
14	1	0.99862^{934}_{894}
14	2	0.9972^{9030}_{8990}
14	3	0.997197^{74}_{33}
14	4	0.99718^{422}_{382}
14	5	0.99718^{115}_{075}
14	6	0.9971^{8028}_{7987}
14	7	0.9971^{8016}_{7959}
15	1	0.99862^{937}_{891}
15	2	0.9972^{9032}_{8987}
15	3	0.997197^{75}_{30}
15	4	0.99718^{423}_{377}
15	5	0.99718^{114}_{069}
15	6	0.9971^{8023}_{7978}
15	7	0.997179^{95}_{50}
16	1	0.99862^{940}_{887}
16	2	0.9972^{9036}_{8983}
16	3	0.997197^{78}_{26}
16	4	0.99718^{425}_{373}
16	5	0.99718^{116}_{064}
16	6	0.9971^{8024}_{7971}
16	7	0.997179^{92}_{40}
16	8	0.997179^{95}_{21}
17	1	0.99862^{944}_{884}
17	2	0.9972^{9039}_{8979}
17	3	0.997197^{81}_{22}
17	4	0.99718^{428}_{369}
17	5	0.99718^{119}_{059}
17	6	0.9971^{8025}_{7966}
17	7	0.997179^{92}_{33}
17	8	0.997179^{81}_{21}
18	1	0.99862^{947}_{880}
18	2	0.9972^{9042}_{8976}
18	3	0.997197^{84}_{19}
18	4	0.99718^{431}_{365}
18	5	0.99718^{121}_{055}
18	6	0.9971^{8027}_{7962}
18	7	0.997179^{93}_{28}
18	8	0.997179^{80}_{15}
18	9	0.99717^{990}_{898}
19	1	0.99862^{950}_{877}
19	2	0.9972^{9046}_{8972}
19	3	0.997197^{88}_{15}
19	4	0.99718^{434}_{361}
19	5	0.99718^{124}_{051}
19	6	0.9971^{8030}_{7957}
19	7	0.997179^{96}_{23}
19	8	0.997179^{82}_{09}
19	9	0.997179^{77}_{04}
20	1	0.99862^{954}_{873}
20	2	0.9972^{9049}_{8969}
20	3	0.997197^{91}_{11}
20	4	0.99718^{438}_{358}
20	5	0.99718^{128}_{048}
20	6	0.9971^{8033}_{7953}
20	7	0.997179^{99}_{19}
20	8	0.997179^{84}_{05}
20	9	0.99717^{978}_{899}
20	10	0.99717^{993}_{881}

Table 6.6: Verified global optimization results for symmetric configurations of k particles in 1D for $k = 21$ to $k = 26$. $v^{\star}_{x,i}$ is the optimal distance between two adjacent particles p_i and p_{i+1}, and p_{k-i-1} and p_{k-i}. The results for $k = 3$ to $k = 20$ are listed in Tab. 6.5.

k	i	$v^{\star}_{x,i}$
21	1	0.99862^{958}_{869}
21	2	0.9972^{9053}_{8965}
21	3	0.997197^{95}_{07}
21	4	0.99718^{441}_{354}
21	5	0.99718^{131}_{044}
21	6	0.9971^{8037}_{7949}
21	7	0.9971^{8002}_{7915}
21	8	0.997179^{88}_{00}
21	9	0.99717^{981}_{894}
21	10	0.99717^{979}_{891}
22	1	0.99862^{963}_{864}
22	2	0.9972^{9058}_{8960}
22	3	0.997197^{100}_{02}
22	4	0.99718^{446}_{349}
22	5	0.99718^{136}_{038}
22	6	0.9971^{8042}_{7944}
22	7	0.9971^{8007}_{7909}
22	8	0.99717^{992}_{894}
22	9	0.99717^{986}_{888}
22	10	0.99717^{983}_{885}
22	11	0.9971^{8002}_{7864}
23	1	0.99862^{968}_{859}
23	2	0.9972^{9062}_{8955}
23	3	0.99719^{805}_{698}
23	4	0.99718^{451}_{344}
23	5	0.99718^{141}_{034}
23	6	0.9971^{8046}_{7939}
23	7	0.9971^{8011}_{7904}
23	8	0.99717^{997}_{890}
23	9	0.99717^{990}_{883}
23	10	0.99717^{987}_{880}
23	11	0.99717^{985}_{878}
24	1	0.99862^{973}_{854}
24	2	0.9972^{9067}_{8951}
24	3	0.99719^{810}_{693}
24	4	0.99718^{456}_{339}
24	5	0.99718^{146}_{029}
24	6	0.9971^{8051}_{7934}
24	7	0.9971^{8016}_{7899}
24	8	0.9971^{8001}_{7884}
24	9	0.99717^{994}_{878}
24	10	0.99717^{991}_{874}
24	11	0.99717^{990}_{873}
24	12	0.9971^{8013}_{7849}
25	1	0.99862^{977}_{850}
25	2	0.9972^{9072}_{8946}
25	3	0.99719^{814}_{688}
25	4	0.99718^{460}_{335}
25	5	0.99718^{150}_{024}
25	6	0.9971^{8056}_{7930}
25	7	0.9971^{8021}_{7895}
25	8	0.9971^{8006}_{7880}
25	9	0.99717^{999}_{873}
25	10	0.99717^{995}_{869}
25	11	0.99717^{994}_{868}
25	12	0.99717^{993}_{867}
26	1	0.99862^{982}_{845}
26	2	0.9972^{9077}_{8941}
26	3	0.99719^{819}_{683}
26	4	0.99718^{465}_{330}
26	5	0.99718^{155}_{019}
26	6	0.9971^{8060}_{7925}
26	7	0.9971^{8025}_{7890}
26	8	0.9971^{8011}_{7875}
26	9	0.9971^{8004}_{7868}
26	10	0.9971^{8000}_{7864}
26	11	0.99717^{998}_{863}
26	12	0.99717^{997}_{862}
26	13	0.9971^{8025}_{7833}

In Fig. 6.12, the results for the distances $v_{x,i}$ are shown. For $k > 19$, the final boxes are summarized, which explains the larger ranges in the right plot. The results for symmetric 1D configurations agree with the previous results presented in Sec. 6.2.2.6, where this symmetry was not assumed.

Tab. 6.7 lists the performance of COSY-GO for different Taylor Model orders, where the orders from two to five are denoted by 'O'. As expected and seen already seen in Tab. 6.3, the number of required steps tends to reduce with higher order Taylor Models due to the tighter bounding capabilities. At the same time, the higher order calculations require more computation time per step, which can increase the overall computation time.

For $k < 23$, the calculations of order three (O3) are the most time efficient just like for the results in Sec. 6.2.2.6. Only for very large k, the verified optimization starts to struggle with the floating

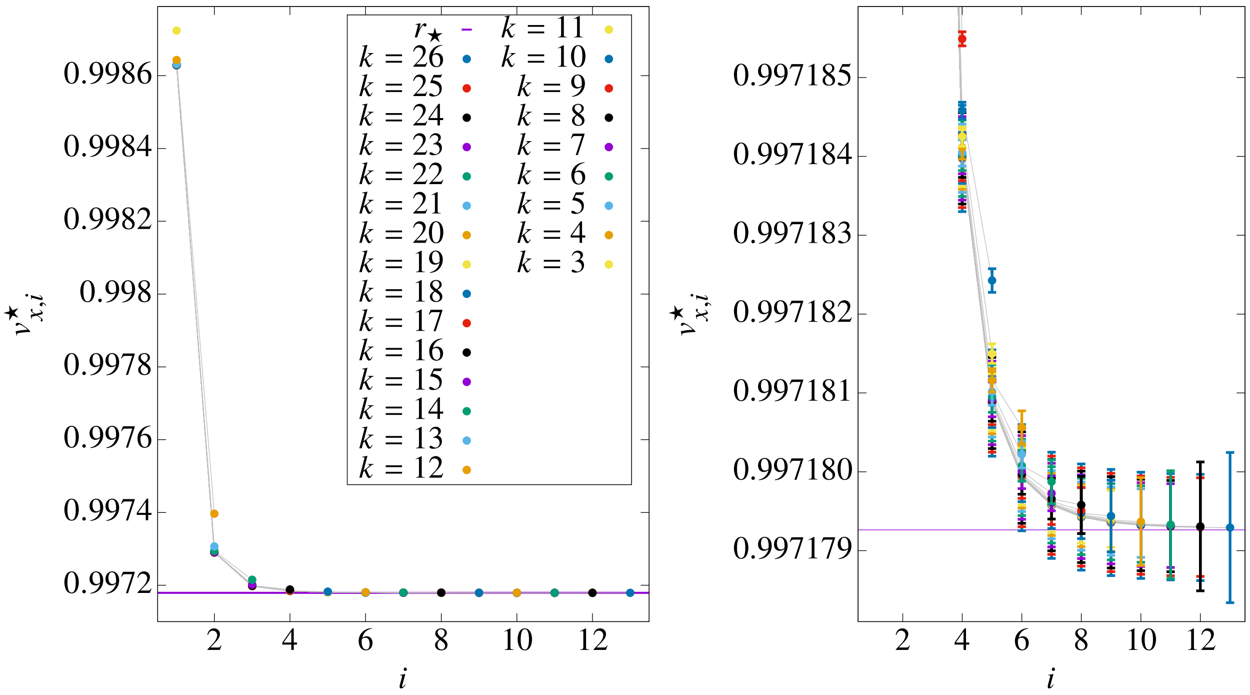

Figure 6.12: The plots show the values for the optimized variables $v^\star_{x,i}$ of the symmetric minimum energy configuration of k particles that resulted from the verified global optimization. Again, the middlemost distances asymptotically approach $r_\star \approx 0.998724135$, the solution of the infinite equidistant configuration from Eq. (6.41). The right plot shows the logarithm of the difference between the calculated distances from the verified optimization and $r_\star$. The ranges reflect the side-length of the remaining box.

Table 6.7: Performance of verified global optimization using COSY-GO with QFB/LDB enabled for the minimum energy search of a 1D symmetric configuration of k particles. The Taylor Model orders are denoted by ‘O’. Note that we use n_{var} as a shorthand notation for $n_{\text{1D,sym,var}}$ in the table.

k	n_{var}	n_{pairs}	Computation time [s]				Steps			
			O2	O3	O4	O5	O2	O3	O4	O5
3	1	3	0.010	0.010	0.016	0.011	8	7	7	5
4	2	6	0.012	0.011	0.016	0.013	17	11	10	10
5	2	10	0.013	0.011	0.016	0.013	18	12	12	10
6	3	15	0.042	0.021	0.031	0.036	28	20	19	16
7	3	21	0.032	0.029	0.016	0.025	28	20	20	18
8	4	28	0.048	0.027	0.047	0.053	40	25	25	24
9	4	36	0.045	0.025	0.047	0.065	42	25	25	24
10	5	45	0.075	0.042	0.078	0.100	59	31	29	29
11	5	55	0.099	0.058	0.094	0.365	62	33	30	29
12	6	66	0.137	0.084	0.140	0.318	95	43	35	34
13	6	78	0.277	0.088	0.187	0.474	101	46	37	34
14	7	91	0.653	0.159	0.343	0.794	171	66	45	38
15	7	105	0.473	0.173	0.406	0.906	188	66	49	39
16	8	120	1.355	0.478	0.843	1.468	313	89	70	43
17	8	136	2.043	1.163	0.641	1.746	317	94	72	45
18	9	153	3.514	1.605	1.893	2.587	573	130	104	53
19	9	171	3.515	1.983	1.991	2.433	571	130	113	59
20	10	190	8.748	3.593	4.860	4.665	1113	164	157	74
21	10	210	8.285	2.517	4.485	6.630	1101	186	170	82
22	11	231	15.47	2.635	8.952	18.40	1959	234	230	118
23	11	253	56.17	68.51	170.2	418.5	6208	4360	4340	4227
24	12	276	118.3	155.1	419.9	1255	12247	8518	8494	8399
25	12	300	130.2	175.5	467.7	1361	12503	8551	8539	8429
26	13	325	276.3	401.5	1215	3601	24527	16832	16839	15332

point accuracy and the associated increase in final boxes.

The reduction of the optimization variables by assuming symmetric configurations, in comparison to the calculations in Sec. 6.2.2.6, significantly reduces the computation time and the number of steps.

In Fig. 6.13, the time efficiency and the number of steps required for the optimization are shown together with the results from the previous section (Sec. 6.2.2.6).

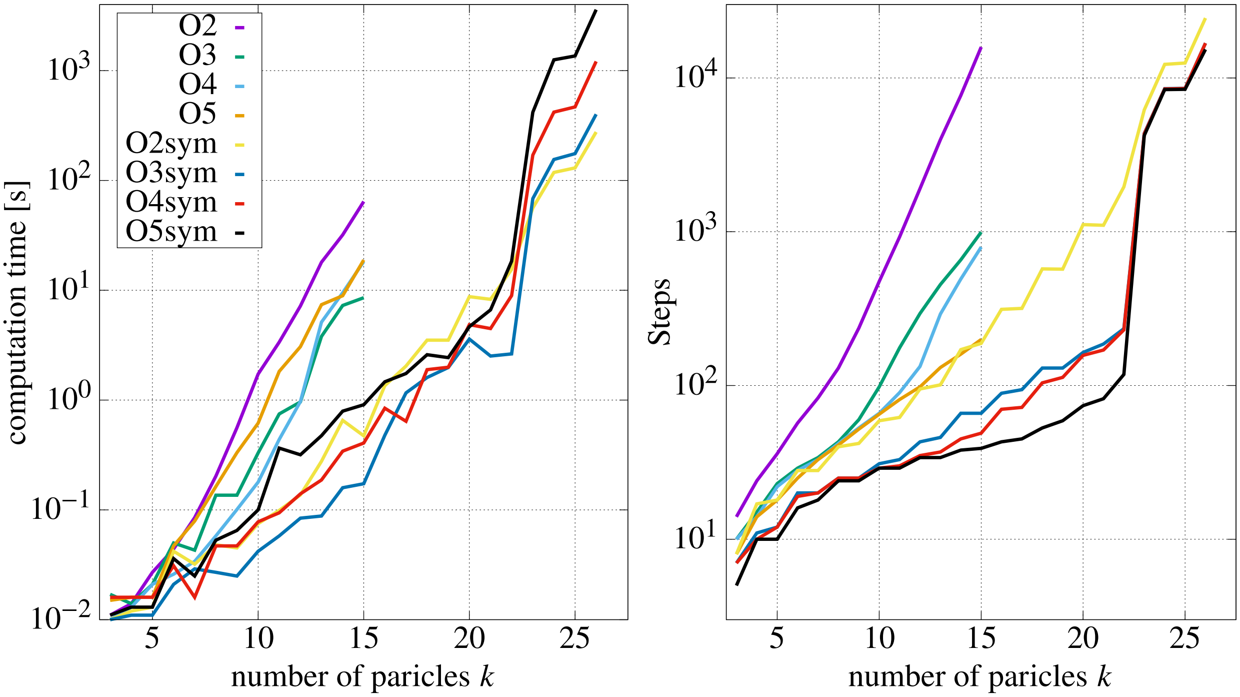

Figure 6.13: Performance of the minimum energy search for configurations of k particles in 1D using COSY-GO with different Taylor Model orders with QFB/LDB enabled. The order of the Taylor Models is denoted by ‘O’. The results from both Sec. 6.2.2.6 and Sec. 6.2.2.7 (’sym’) are shown.

6.2.2.8 Redundancies and Penalty Functions

Note that there are two versions for every nonsymmetric configuration in 1D that are mirror images of each other with regard to the midpoint of the configuration, i.e., for every configuration $(v_{x,1}, v_{x,2}, ..., v_{x,k-1})$, there is a mirror configuration $(v_{x,k-1}, v_{x,k-2}, ..., v_{x,1})$. Both configurations are equivalent for the optimization problem, but are distinct domains in the search space of the optimizer.

Because the solutions of the 1D studies are symmetric, this redundancy of mirror images of the same configuration did not appear in the results. As a preparation for the multidimensional studies below, where the solutions are not always this symmetric, we discuss a method to suppress redundant mirror images of configurations.

A redundancy can be suppressed by finding criteria that distinguish those equivalent configurations and using a penalty function to artificially increases the objective function for the redundant version(s). For mirror symmetries along the x axis, one distinction is the center of mass

$$x_{\text{CM}} = \frac{1}{k} \sum_{i=1}^{k} x_i, \tag{6.53}$$

where $x_i \geq 0$ and $x_1 = 0$.

Without loss of generality, one can require that the optimal configuration satisfies

$$\Delta x_{\text{CM}} = x_{\text{CM}} - \frac{x_k}{2} \leq 0, \tag{6.54}$$

and enforce it by letting the penalty function $b_{x_{\text{CM}}}$ scale with the difference Δx_{CM} if it is positive and be zero otherwise:

$$b_{x_{\text{CM}}} = \begin{cases} \lambda \cdot \Delta x_{\text{CM}} & \text{for } \Delta x_{\text{CM}} > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{6.55}$$

where λ is a large positive number like 10^{10}.

6.2.3 Minimum Energy Lennard-Jones Configurations in 2D and 3D

To find minimum energy configurations of k particles in 2D and 3D using verified global optimization, we are going to build on the methods from the 1D studies. First, we describe the solution space of all possible minimum energy configurations in terms of a set of optimization variables (see Sec. 6.2.3.1). Then, we determine how the bounds on the minimum energy and the inter-particle distances have to be adjusted for configurations in 2D and 3D. Based on those bounds we define the initial search domain for the optimization variables and make sure that the objective function can be evaluated for every point in the initial search domain. Lastly, we present the results of the verified optimization.

6.2.3.1 Coordinate System, Numbering Scheme, and Variable Definition

We use a Cartesian coordinate system (x, y, z) to describe all possible minimum energy configurations of k particles in 2D and 3D. For configurations in 2D, all z coordinates and variables are to be ignored.

The first step is placing the configuration such that the largest distance of the configuration lies on the x axis. The implications for configurations with multiple largest r_{ij} are discussed below. We call the largest inter-particle distance the *major axis* of the configuration, denoted by r_{1k}, because the particles that span the major axis are denoted by p_1 and p_k. The configuration is placed such that

$$\vec{p}_1 = (0, 0, 0) \quad \text{and} \quad \vec{p}_k = (x_k = r_{1k} \geq 0, 0, 0), \tag{6.56}$$

where $\vec{p}_i$ is the position vector of particle p_i. This fixes p_1 to the origin and p_k to the x axis.

To avoid ambiguity regarding which of the two particles of the major axis is p_1, we require that the center of mass of the configuration

$$\vec{p}_{\text{CM}} = (x_{\text{CM}}, y_{\text{CM}}, z_{\text{CM}}) = \frac{1}{k} \sum_{i=1}^{k} \vec{p}_i \tag{6.57}$$

satisfies Eq. (6.54) for the x coordinate, i.e., $x_{\text{CM}} - x_k/2 \leq 0$.

Just like in 1D, we use the x positions of the particles to number them from 1 to k such that

$$x_i \leq x_j \quad \text{for} \quad i < j \quad \text{with} \quad x_1 = 0. \tag{6.58}$$

For configurations in 3D, we require that particle p_2 is in the xy plane, i.e.

$$z_2 = 0 \tag{6.59}$$

without loss of generality.

To determine the orientation of the y (and z) axis we require without loss of generality, that

$$y_2 \leq 0 \quad \text{and} \quad z_3 \leq 0. \tag{6.60}$$

The optimization variables $v_{x,i}$ represent the distance in the x coordinates between the particles p_{i+1} and p_i with

$$v_{x,i} = x_{i+1} - x_i \geq 0 \quad \text{for} \quad i \in \{1, 2, ..., k-1\}. \tag{6.61}$$

The optimization variables $v_{y,i}$ are the y positions of the particles with

$$v_{y,i} = y_i \quad \text{for} \quad i \in \{2, 3, ..., k-1\}. \tag{6.62}$$

For configurations in 3D, we additionally define the optimization variables $v_{z,i}$ as the z positions of the particles with

$$v_{z,i} = z_i \quad \text{for} \quad i \in \{3, 4, ..., k-1\}. \tag{6.63}$$

In total, this yields

$$n_{\text{2D,var}} = 2k - 3 \tag{6.64}$$

optimization variables for configurations of k particles in 2D and

$$n_{\text{3D,var}} = 3k - 6 \tag{6.65}$$

optimization variables for configurations of k particles in 3D.

Note that there is no variable for z_2 because $z_2 = 0$ by definition of the y axis.

The squared distance r_{ij}^2 between any two particles p_i and p_j with $i < j$ can be expressed in terms of $v_{x,i}$, $v_{y,i}$, and $v_{z,i}$ using Eq. (6.18) by

$$r_{ij}^2 = \left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2 + \left(z_j - z_i\right)^2 = \left(\sum_{n=i}^{j-1} v_{x,n}\right)^2 + \left(v_{y,j} - v_{y,i}\right)^2 + \left(v_{z,j} - v_{z,i}\right)^2. \tag{6.66}$$

The center of mass requirement in the x coordinate $\Delta x_{\text{CM}} = x_{\text{CM}} - \frac{x_k}{2} \leq 0$ (from Eq. (6.54)) is enforced using the penalty function from Eq. (6.55):

$$b_{x_{\text{CM}}} = \begin{cases} \lambda \cdot \Delta x_{\text{CM}} & \text{for } \Delta x_{\text{CM}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.67}$$

For the largest distance requirement

$$\Delta r_{ij}^2 = r_{ij}^2 - r_{1k}^2 \leq 0 \quad \forall ij, \tag{6.68}$$

is handled using the sum of individual penalty functions for each inter-particle distance of the configuration, namely,

$$b_{r^2} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} b_{r_{ij}^2} \quad \text{where} \quad b_{r_{ij}^2} = \begin{cases} \lambda \cdot \Delta r_{ij}^2 & \text{for } \Delta r_{ij}^2 > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{6.69}$$

The largest distance requirement reduces the rotational ambiguity in the placement of the coordinate system. Configurations with multiple largest distances often have symmetry properties that relate those largest distances to each other. If this is the case, the representations of the configuration using the coordinate system along the largest distances are identical.

6.2.3.2 Upper Bounds $r_{k,\mathrm{UB}}$ and $v_{x,\mathrm{UB}}$ on Inter-Particle Distances of Minimum Energy Configurations in 2D and 3D

The argument regarding the upper bound on the distance of two adjacent particles from Sec. 6.2.2.2 can be generalized to multidimensional cases. Using the variable and coordinate definition from above, changing a variable $v_{x,l}$ moves the right-side subconfiguration composed of the particles p_j with $j > l$ along the x axis, leaving the left-side subconfiguration of particles p_i with $i \leq l$ unchanged. Following the argumentation in Sec. 6.2.2.2, any configuration with any $v_{x,l} > 1$ is not optimal, thus making

$$v_{x,\mathrm{UB}} = 1 \tag{6.70}$$

an upper bound for all $v_{x,i}$ in the minimum energy configuration. This shows that the result from Eq. (6.20) could be generalized to the multidimensional case.

Fig. 6.14 shows this multidimensional generalization for 2D.

Because the sum of all $v_{x,i}$ yields the length of the major axis, which is the longest distance of the configuration by definition, the upper bound

$$r_{k,\mathrm{UB}} = k - 1 \tag{6.71}$$

on any r_{ij} of the minimum energy configurations is also valid for the multidimensional cases. However, for the multidimensional cases, this upper bound is not as tight as for configurations in 1D. Further advances of the method may yield tighter upper bounds on the minimum for configurations in 2D and 3D, e.g., one may be able to show that the maximum r_{ij} of $\mathcal{S}_k^\star$ in 2D can serve as r_{UB} of $\mathcal{S}_k^\star$ in 3D.

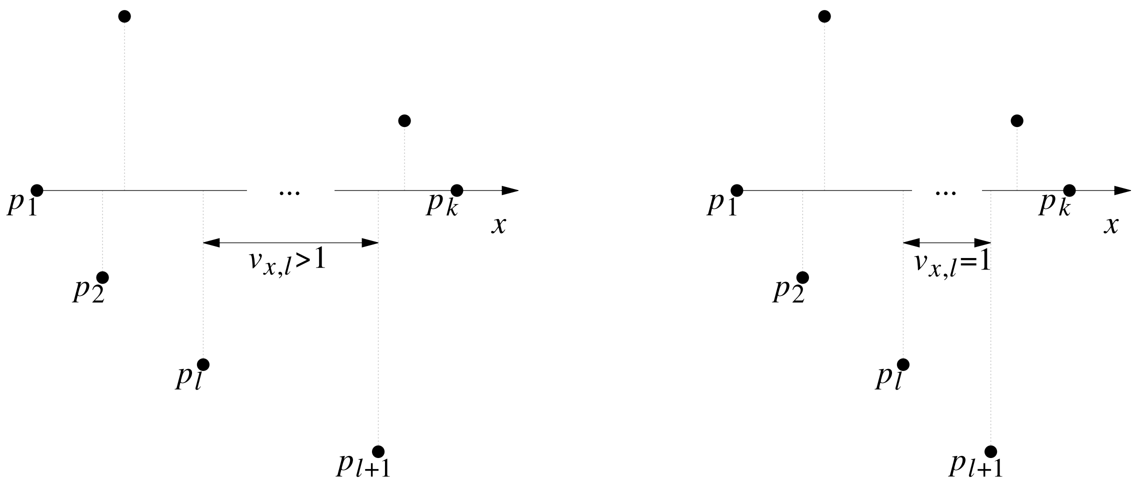

Figure 6.14: The particles p_i are numbered according to their x position. The variable $v_{x,l}$ denotes the x distance between particle p_l and p_{l+1}. A configuration with any $v_{x,l} > 1$ (left picture) is never optimal, because the overall potential can always be lowered by setting $v_{x,l} = 1$ (right picture).

6.2.3.3 The Upper Bound $U_{k,\mathrm{UB}}$ on the Minimum Energy

As previously discussed in Sec. 6.2.2.3, any configuration $\mathcal{S}_{k,n_{\mathrm{dim}}}$ can serve as an upper bound configuration $\mathcal{S}_{k,n_{\mathrm{dim}},\mathrm{UB}}$ providing an upper bound $U_{k,n_{\mathrm{dim}},\mathrm{UB}}$ on the minimum energy $U^{\star}_{k,n_{\mathrm{dim}}}$.

For configurations in 2D and 3D, a good upper bound configuration can be obtained by using $\mathcal{S}^{\star}_{k-1,n_{\mathrm{dim}}}$. We add a kth particle in a small and simple verified global optimization on its own, where only the coordinates of the kth particle are the optimization variables. From Sec. 6.2.3.2, we know that when we place an axis in any orientation in the minimum energy configuration, the distances between adjacent projections onto that axis are bound by 1, i.e., $v_{x,i} \leq 1$. Thus, the initial search domain for the 'upper bound' optimization of the position of the kth particle is determined by the maximum and minimum coordinates of $\mathcal{S}^{\star}_{k-1,n_{\mathrm{dim}}}$ along each orthogonal axis of the n_{dim} space, plus a band of width 1 around it (see Fig. 6.15 for a 2D example of $k = 6$). The resulting upper bound on the overall potential of this optimized upper bound configuration $\mathcal{S}_{k,n_{\mathrm{dim}},\mathrm{UB}}$ is then used is then used as $U_{k,n_{\mathrm{dim}},\mathrm{UB}}$.

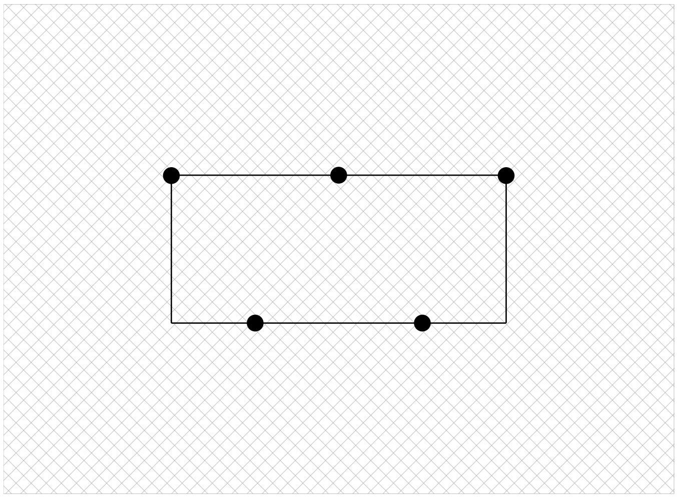

Figure 6.15: The optimal 2D configuration of five particles is denoted by five dots. Enclose all the five particles by a 2D rectangle using the minimum and maximum coordinates in x and y, shown by a solid line rectangle. Surround the resulting rectangle with a band of width 1, and we have a initial search domain for the sixth particle (shaded area).

6.2.3.4 Setup of Initial Search Domain of the Optimization Variables in 2D and 3D

In contrast to the 1D cases, the bounds on the r_{ij} do not directly translate to bounds of Cartesian search domains for configurations in 2D and 3D. As a consequence, the initial search domain covers a larger area to rigorously include all possible minimum energy configurations.

The variables $v_{x,i}$ are bound from the top by $v_{x,\mathrm{UB}} = 1$ (Eq. (6.70)). As for the lower bound, the same argument of the 1D case is not applicable since the particle configuration is not confined on a line anymore, and we are left only with the variable definition itself, i.e., $v_{x,i} \geq 0$ (Eq. (6.61)). Thus, the initial search domain is given by

$$v_{x,i} \in \left[0, v_{x,\mathrm{UB}}\right] = [0,1] \quad \text{for} \quad i \in \{1,2,...,k-1\}. \tag{6.72}$$

We bound the variables $v_{y,i}$ (and $v_{z,i}$) using the requirement that any inter-particle distance r_{ij} is less than or equal to the major axis r_{1k} and the upper bound $r_{\mathrm{UB}} = k-1$ (Eq. (6.71)) on any r_{ij}. Consider any particle p_i of the configuration with $i \notin 1, k$. The distances r_{1i} and r_{ik} must be at most r_{1k}. Fig. 6.16 illustrates this requirement with circles of radius r_{1k} around p_1 and p_k.

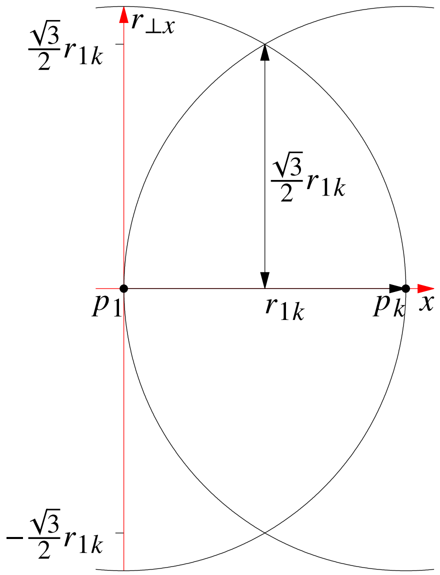
$r_{\perp x}$
$\frac{\sqrt{3}}{2} r_{1k}$
$\frac{\sqrt{3}}{2} r_{1k}$
p_1
r_{1k}
p_k
x
$-\frac{\sqrt{3}}{2} r_{1k}$

Figure 6.16: Schematic illustration of the upper bound on distance perpendicular to the x axis (major axis) due to the requirement of having the longest distance between p_1 and p_k. p_1 at $x_1 = 0$ and p_k at x_k, where the major axis length $r_{1k} = x_k$.

All p_i must be within the overlap of the two circles, which yields an upper bound on the perpendicular distance of p_i to the major axis depending on r_{ik}. The upper bound corresponds to the height of an equilateral triangle of side-length r_{1k}, which is largest for $r_{1k} = r_{\mathrm{UB}}$. Thus, the variables are bound by

$$v_{y,2} \in [-1, 0]\, \frac{\sqrt{3}}{2} r_{k,\mathrm{UB}} \quad \text{and} \quad v_{y,i} \in [-1, 1]\, \frac{\sqrt{3}}{2} r_{k,\mathrm{UB}} \quad \text{for} \quad i \in \{3, 4, ..., k-1\} \tag{6.73}$$

$$v_{z,3} \in [-1, 0]\, \frac{\sqrt{3}}{2} r_{k,\mathrm{UB}} \quad \text{and} \quad v_{z,i} \in [-1, 1]\, \frac{\sqrt{3}}{2} r_{k,\mathrm{UB}} \quad \text{for} \quad i \in \{4, 5, ..., k-1\}, \tag{6.74}$$

where $r_{\mathrm{UB}} = k - 1$ (Eq. (6.71)).

Note that the initial search domains for $v_{y,2}$ and v_{z_3} have an upper bound of zero, because of the definition in Eq. (6.60).

In Fig. 6.17, the $n_{\mathrm{2D,var}}$-dimensional initial search domain box is shown by illustrating the initial search domain of the individual variables and how those variables relate to the position of the particles in the configuration.

Note that the initial search domain does not exclude configurations with inter-particle distances $r_{ij} < r_{k,\mathrm{LB}}$, where r_{LB} is given by Eq. (6.25). In particular, it currently includes configurations

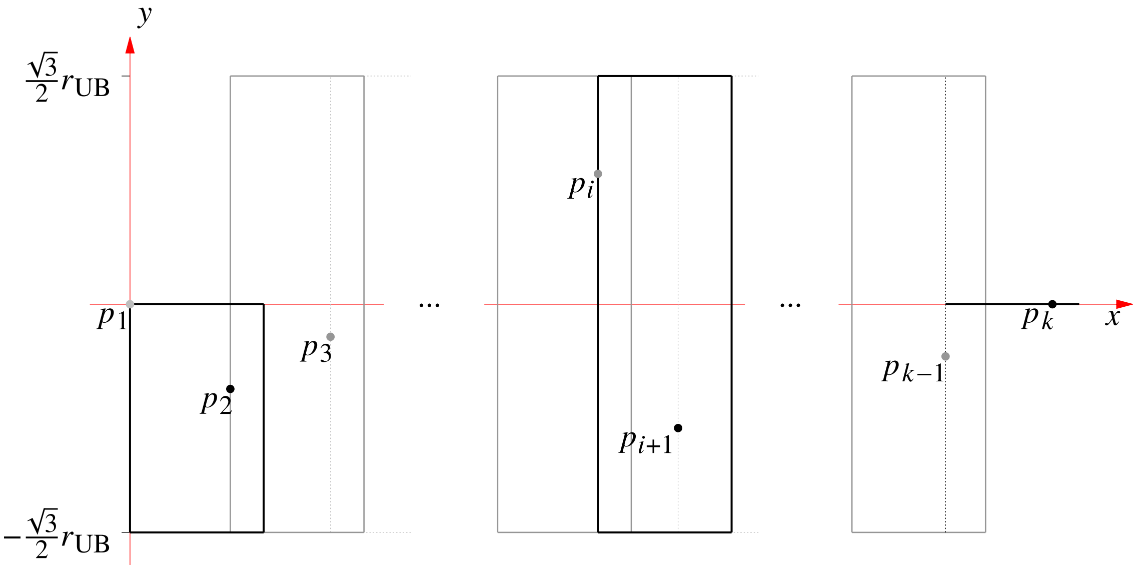
y
$\frac{\sqrt{3}}{2} r_{\mathrm{UB}}$
$-\frac{\sqrt{3}}{2} r_{\mathrm{UB}}$
p_1
p_2
p_3
...
p_i
p_{i+1}
...
p_{k-1}
p_k
x

Figure 6.17: Initial search domain for a configuration of k particles in 2D. Note that the initial domain width in x direction is always 1 (see Eq. (6.72)) and that the x position of particle p_i determines the starting position in x of the domain of particle p_{i+1}. Particle p_1 is fixed to the origin, particle p_2 is bound by $y_2 \leq 0$, and particle p_k has a fixed y value of zero.

with $r_{ij} = 0$, for which the Lennard-Jones potential is not defined. To address this, we define a modified Lennard-Jones potential below without changing the optimization problem.

6.2.3.5 The Evaluation of the Objective Function

We first take a closer look at how to efficiently evaluate the objective function from Eq. (6.10), which is composed of n_{pairs} individual Lennard-Jones interactions

$$U_{\mathrm{LJ}}\left(r_{ij}\right) = 1 + r_{ij}^{-12} - 2r_{ij}^{-6} \tag{6.75}$$

as previously introduced in Eq. (6.8).

Eq. (6.66) yields the squared distances r_{ij}^2. To avoid unnecessarily taking the square-root to compute r_{ij}, we implement a Lennard-Jones potential that takes the squared distance $r_{\mathrm{sqr}} = r_{ij}^2$ as its argument with

$$U_{\mathrm{LJ,sqr}}\left(r_{\mathrm{sqr}}\right) = 1 + r_{\mathrm{sqr}}^{-3}\left(r_{\mathrm{sqr}}^{-3} - 2\right). \tag{6.76}$$

where the squared distance r_{ij}^2 is evaluated from the optimization variables using Eq. (6.66).

To deal with configurations with at least one $r_{ij} = 0$, we remind ourselves that Sec. 6.2.2.4 showed that all configurations with a single r_{ij} below $r_{k,\mathrm{LB}}$ cannot be a minimum energy configuration. This also means that any configuration with at least one $U_{\mathrm{LJ}}(r_{ij})$ larger than $U_{\mathrm{LJ}}(r_{k,\mathrm{LB}})$ is not a minimum energy configuration.

This leads to the idea [13] of modifying the objective function for r_{ij} smaller than $r_{k,\mathrm{LB}}$ such that it can be evaluated for $r_{ij} = 0$ without changing the optimization problem. The only requirement is that the modified Lennard-Jones potential $\tilde{U}_{\mathrm{LJ,sqr}}$ satisfies

$$\tilde{U}_{\mathrm{LJ,sqr}}\left(r_{\mathrm{sqr}}\right) \geq U_{\mathrm{LJ,sqr}}\left(r_{\mathrm{LB}}^2\right) \quad \forall r_{\mathrm{sqr}} < r_{\mathrm{LB}}^2. \tag{6.77}$$

Hence, we define the modified Lennard-Jones potential and compose it of the regular Lennard-Jones potential for $r_{\mathrm{sqr}} \geq r_{\mathrm{LB}}^2$ and the tangential extension at r_{LB}^2 for $r_{\mathrm{sqr}} \leq r_{\mathrm{LB}}^2$. The modified Lennard-Jones potential [13] is then given by

$$\tilde{U}_{\mathrm{LJ,sqr}}\left(r_{\mathrm{sqr}}, r_{\mathrm{LB}}\right) = \begin{cases} U_{\mathrm{LJ,sqr}}\left(r_{\mathrm{sqr}}\right) & \text{for } r_{\mathrm{sqr}} \geq r_{\mathrm{LB}}^2 \\ U'_{\mathrm{LJ,sqr}}\left(r_{\mathrm{LB}}^2\right) \cdot \left(r_{\mathrm{sqr}} - r_{\mathrm{LB}}^2\right) + U_{\mathrm{LJ,sqr}}\left(r_{\mathrm{LB}}^2\right) & \text{for } r_{\mathrm{sqr}} \leq r_{\mathrm{LB}}^2 \end{cases} \tag{6.78}$$

where $U'_{\mathrm{LJ,sqr}}$ is the first derivative of $U_{\mathrm{LJ,sqr}}$ with

$$U'_{\mathrm{LJ,sqr}}\left(r_{\mathrm{sqr}}\right) = 6 r_{\mathrm{sqr}}^{-4}\left(1 - r_{\mathrm{sqr}}^{-3}\right). \tag{6.79}$$

The modified Lennard-Jones potential is shown in Fig. 6.18.

Next we address how to handle such a piecewise function when using Taylor Models.

6.2.3.6 Taylor Model Evaluation of Piecewise Defined Functions

Consider a continuous piecewise defined function

$$f\left(x\right) = \begin{cases} f_{\mathrm{L}}\left(x\right) & \text{for } x \leq x_0, \\ f_{\mathrm{R}}\left(x\right) & \text{for } x \geq x_0, \end{cases} \tag{6.80}$$

with

$$f_{\mathrm{L}}\left(x_0\right) = f_{\mathrm{R}}\left(x_0\right), \tag{6.81}$$

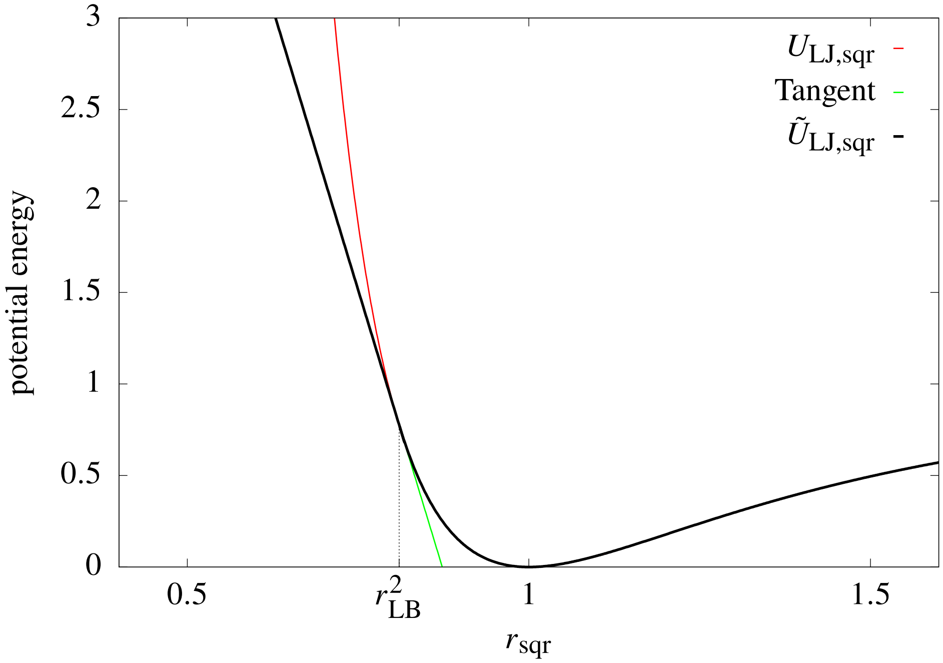

Figure 6.18: Piecewise defined modified Lennard-Jones potential $\tilde{U}_{\mathrm{LJ,sqr}}$ shown by the black curve. The red curves shows the Lennard-Jones potential and the green line shows the tangent of this Lennard-Jones potential of r_{sqr}. The plot shown here is an example case with $r_{\mathrm{LB}}^2 = 0.9^2$.

where $f_{\mathrm{L}}(x)$ and $f_{\mathrm{R}}(x)$ are $m+1$ times differentiable.

We want to find a Taylor Model

$$f_{\mathrm{TM}} = \left(\mathcal{P}_f, \epsilon_f\right) \tag{6.82}$$

that tightly captures $f(x)$ over the domain $\mathbb{D} = [a, b]$ where $x_0 \in \mathbb{D}$, and the subdomains of the function pieces are $\mathbb{D}_{\mathrm{L}} = [a, x_0]$ and $\mathbb{D}_{\mathrm{R}} = [x_0, b]$.

In the first step, we prepare two Taylor Models, $f_{\mathrm{L,TM}}$ and $f_{\mathrm{R,TM}}$, for the function pieces over the respective subdomains $\mathbb{D}_{\mathrm{L}}$ and $\mathbb{D}_{\mathrm{R}}$. Our goal is to find $f_{\mathrm{TM}} = \left(\mathcal{P}_f, \epsilon_f\right)$ such that

$$f_{\mathrm{TM}} \supseteq \begin{cases} f_{\mathrm{L,TM}} & \text{over} \quad \mathbb{D}_{\mathrm{L}}, \\ f_{\mathrm{R,TM}} & \text{over} \quad \mathbb{D}_{\mathrm{R}}, \end{cases} \tag{6.83}$$

which is illustrated in Fig. 6.19.

We start by trying to find a good polynomial $\mathcal{P}_f$ that models $f(x)$ over the domain $\mathbb{D}$ well. Suppose we have polynomials that closely represent f_{L} over $\mathbb{D}_{\mathrm{L}}$ and f_{R} over $\mathbb{D}_{\mathrm{R}}$. We denote those polynomials by $\mathcal{P}_{\mathrm{L}}$ and $\mathcal{P}_{\mathrm{R}}$, respectively. The following weighted average of $\mathcal{P}_{\mathrm{L}}$ and $\mathcal{P}_{\mathrm{R}}$ can be a

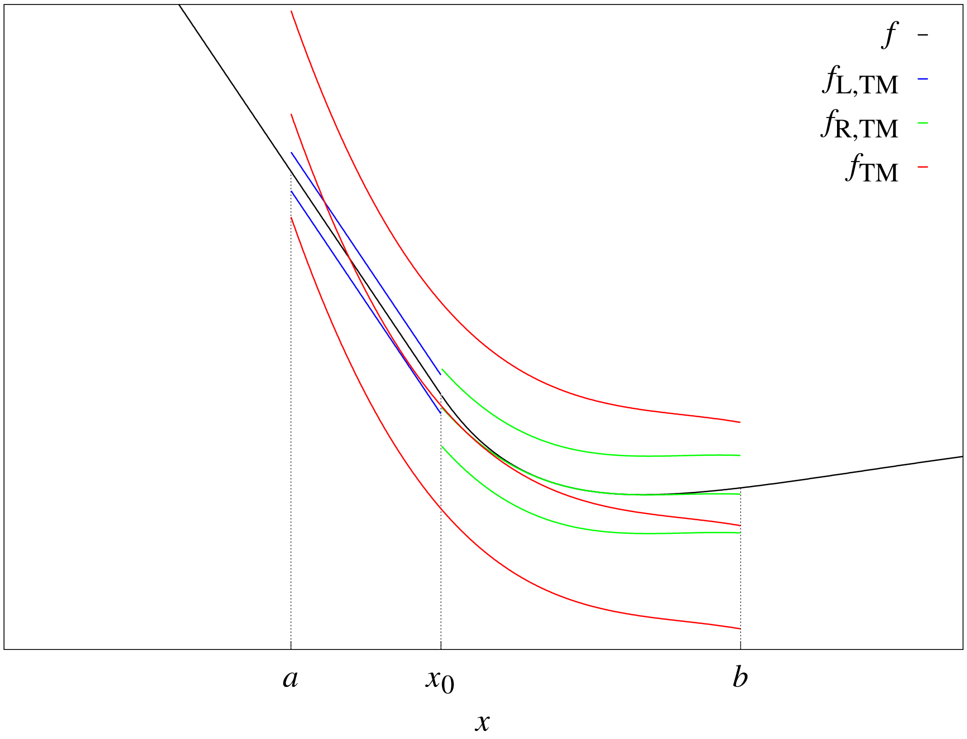

Figure 6.19: Taylor Model description of piecewise defined function. Each Taylor Model is represented by three lines as previously done in Fig. 2.2. The central curve denotes the polynomial part of the Taylor Model, while the curves above and below it indicate the bounds.

good choice for $\mathcal{P}_f$, with

$$\mathcal{P}_f = \frac{w_{\mathrm{L}}\mathcal{P}_{\mathrm{L}} + w_{\mathrm{R}}\mathcal{P}_{\mathrm{R}}}{w_{\mathrm{L}} + w_{\mathrm{R}}}, \tag{6.84}$$

where the widths of the subdomains $w_{\mathrm{L}} = \mathrm{width}\left(\mathbb{D}_{\mathrm{L}}\right)$ and $w_{\mathrm{R}} = \mathrm{width}\left(\mathbb{D}_{\mathrm{R}}\right)$ are used as weights.

To perform the linear combination in Eq. (6.84), it is essential that the two polynomials, $\mathcal{P}_{\mathrm{L}}$ and $\mathcal{P}_{\mathrm{R}}$, are based on the same expansion point and scaling, which are carried to the resulting polynomial $\mathcal{P}_f$. A natural choice is to take the midpoint m and the half width $h = w/2$ of the domain $\mathbb{D}$ as the polynomial expansion point and scaling for both polynomials. We note that the polynomial parts in $f_{\mathrm{L,TM}}$ and $f_{\mathrm{R,TM}}$ in Eq. (6.83) do not necessarily have the same expansion point and scaling discussed here.

Once $\mathcal{P}_f$ is found, a remainder bound ϵ_f can be estimated as follows such that the requirement of Eq. (6.83) is satisfied.

$$\epsilon_f = \max\left(\left|\mathcal{P}_f - f_{\mathrm{L,TM}}\right|_{\mathbb{D}_\mathrm{L}}, \left|\mathcal{P}_f - f_{\mathrm{R,TM}}\right|_{\mathbb{D}_\mathrm{R}}\right), \tag{6.85}$$

where the notation $|\cdot|_{\mathbb{D}}$ indicates a bound over $\mathbb{D}$. As for $\mathcal{P}_f - f_{\mathrm{L,TM}}$, the expansion point and scaling of $\mathcal{P}_f$ and the polynomial part of $f_{\mathrm{L,TM}}$ have to match, and the same applies to R. Typically, either $\mathcal{P}_f$ or the polynomial part of $f_{L,\mathrm{TM}}$ has to be adjusted to have the same expansion point and scaling. Since the latter case requires Taylor Model arithmetic for the adjustment, it would be more economical to make the necessary adjustments to $\mathcal{P}_f$.

6.2.3.7 The Verified Global Optimization Results for Configurations of k Particles in 2D

The coordinate system is defined for the configuration of k particles in 2D according to the description in Sec. 6.2.3.1. The x axis along the major axis (the largest inter-particle distance) of the configuration is used to number the particles from 1 to k according to their x position such that

$$x_i \leq x_j \quad \text{for} \quad i < j. \tag{6.86}$$

The particle p_1 is fixed to the origin with

$$\vec{p}_1 = (0, 0) \tag{6.87}$$

and particle p_k is fixed to the positive x axis with

$$\vec{p}_k = (x_k \geq 0, 0). \tag{6.88}$$

The y axis is orientated such that

$$y_2 \leq 0. \tag{6.89}$$

We describe a configuration of k particles in 2D using the variables

$$v_{x,i} = x_{i+1} - x_i \geq 0 \quad \text{for} \quad i \in \{1, 2, ..., k-1\} \quad \text{and} \tag{6.90}$$

$$v_{y,i} = y_i \quad \text{for} \quad i \in \{2, 3, ..., k-1\}, \tag{6.91}$$

as previously defined in Sec. 6.2.3.1 in Eq. (6.61) and Eq. (6.62), respectively.

This yields a total number of

$$n_{\text{2D,var}} = 2k - 3 \tag{6.92}$$

optimization variables as mentioned in Eq. (6.64).

The variable domains were determined in Eq. (6.72) and Eq. (6.73) in Sec. 6.2.3.4, with

$$v_{x,i} \in [0,1] \quad \text{for} \quad i \in \{1,2,...,k-1\}, \tag{6.93}$$

$$v_{y,2} \in [-1,0]\,\frac{\sqrt{3}}{2} r_{k,\text{UB}}, \quad \text{and} \tag{6.94}$$

$$v_{y,i} \in [-1,1]\,\frac{\sqrt{3}}{2} r_{k,\text{UB}} \quad \text{for} \quad i \in \{3,4,...,k-1\}, \quad \text{with} \quad r_{k,\text{UB}} = k-1, \tag{6.95}$$

where $r_{k,\text{UB}}$ is known from Eq. (6.71) in Sec. 6.2.3.2.

As discussed in Sec. 6.2.3.5, we use the modified Lennard-Jones potential from Eq. (6.78) as the objective function without changing the optimization problem. The lower bound $r_{k,\text{LB}}$ is determined according to Sec. 6.2.2.4. The squared inter-particle distances – the argument of this objective function – are calculated from $v_{x,i}$ and $v_{y,i}$ according to Eq. (6.66) with $v_{z,i} = 0$.

Following the center of mass requirement in the x direction and the major axis requirement from Sec. 6.2.3.1, we use the penalty functions from Eq. (6.55) and Eq. (6.69).

The verified global optimization is performed with the Taylor Model based verified optimizer COSY-GO [63, 64] in its most advanced setting with QFB/LDB enabled (see Sec. 2.6). Unless stated otherwise the optimization is performed with Taylor Models of order three. The threshold length as a stopping condition is $s_{\text{min}} = 10^{-6}$ as mentioned earlier in Sec. 6.2.2.6.

We start from $k = 3$. From Sec. 6.2.1.3, we know that the solution $\mathcal{S}^{\star}_{3,\text{2D}}$ as this trivial case is an equilateral triangle, which can be represented by the particle positions

$$\vec{p}_1 = (0,0)\,,\ \vec{p}_2 = \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right),\ \vec{p}_3 = (1,0) \tag{6.96}$$

with

$$U^{\star}_{3,2D} = 0. \tag{6.97}$$

We follow the procedure in Sec. 6.2.3.3 to determine $U_{4,2\mathrm{D},\mathrm{UB}}$. For this, we optimize the position (x_4, y_4) of a fourth particle p_4 relative to $\mathcal{S}^{\star}_{3,2\mathrm{D}}$. The initial search domain for the fourth particle according to Sec. 6.2.3.3 and Fig. 6.15 is

$$(x_4, y_4) \in [-1, 2] \times \left[-\frac{\sqrt{3}}{2} - 1, 1\right]. \tag{6.98}$$

The optimization yields an upper bound

$$U_{4,2\mathrm{D},\mathrm{UB}} \leq 0.9270161931701777. \tag{6.99}$$

Using this upper bound in Eq. (6.25) together with the equation for $r_{\min}$ in Eq. (6.23) according to the method in Sec. 6.2.2.4, we have

$$r_{4,2\mathrm{D},\mathrm{LB}} = r_{\min}\left(U_{4,2\mathrm{D},\mathrm{UB}}\right) \geq 0.8936896031162850. \tag{6.100}$$

Starting with $k = 4$, we iteratively perform the optimization for the k particle case and use the result to calculate $U_{k+1,2\mathrm{D},\mathrm{UB}}$ and $r_{k+1,2\mathrm{D},\mathrm{LB}}$ for the $k + 1$ particle case.

The minimum energy configurations for four particles in 2D is shown in Fig. 6.20. The overall potential of the minimum energy configurations is bound by

$$U^{\star}_{4,2\mathrm{D}} = 0.926579141537^{22}_{07}. \tag{6.101}$$

The illustration of $\mathcal{S}^{\star}_{4,2\mathrm{D}}$ in Fig. 6.20 appears to be two connected equilateral triangles that are very slightly squished in the horizontal direction. Tab. 6.8 lists the values for the optimal configuration $\mathcal{S}_{4,2\mathrm{D}}$, i.e., the optimal distances between the individual particles, denoted by $r^{\star}_{ij}$, which makes the differences between $\mathcal{S}^{\star}_{2,4\mathrm{D}}$ and the structure of two connected equilateral triangles apparent. Tab. 6.8 also lists the results for the optimization variables. We can observe that $\mathcal{S}^{\star}_{4,2\mathrm{D}}$ has two symmetry axis.

Compared to two equilateral triangles, $\mathcal{S}^{\star}_{4,2\mathrm{D}}$ brings the outermost particles closer together. At the same time the two particles in the middle (p_2 and p_3) are slightly further apart vertically. This horizontal ‘squishing’ of an equidistance structure to yield $\mathcal{S}^{\star}_{4,2\mathrm{D}}$ could already be observed for minimum energy configurations in 1D in Sec. 6.2.2.

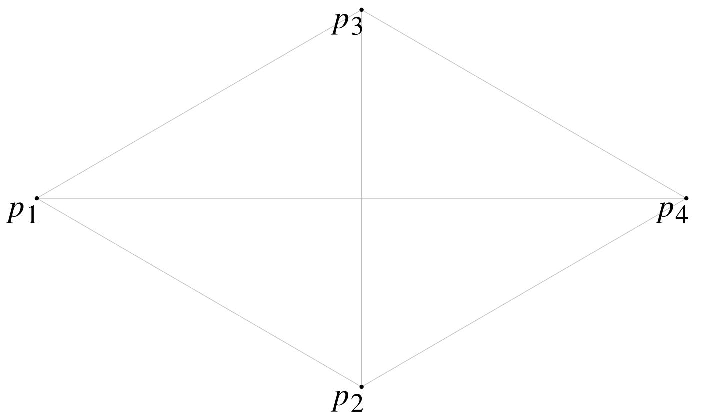

Figure 6.20: Minimum energy configurations of four particles in 2D, $\mathcal{S}^{\star}_{4,2\mathrm{D}}$. Interestingly, the minimum energy configuration is not a square, an obvious symmetric object with fourfold symmetry, but a rhombus with two almost equilateral triangles very slightly squished in the horizontal direction.

Table 6.8: Verified global optimization results for the minimum energy configurations of four particles in 2D, $\mathcal{S}^{\star}_{4,2\mathrm{D}}$. The $r^{\star}_{ij}$ yield the optimal distances between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1}, and $v^{\star}_{y,i}$ is the optimal y position of particle p_i.

k	i	j	$r^{\star}_{ij}$
4	1	2	0.998012^{409}_{230}
4	1	3	0.998012^{470}_{230}
4	1	4	1.726251^{672}_{347}
4	2	3	1.00208^{3071}_{2798}
4	2	4	0.998012^{470}_{230}
4	3	4	0.998012^{409}_{230}

k	†	i	$v^{\star}_{\dagger,i}$
4	x	1	0.863125^{81}_{67}
4	x	2	0.0000000^{+7}_{-1}
4	x	3	0.863125^{81}_{67}
4	y	2	-0.501041^{39}_{54}
4	y	3	0.501041^{54}_{39}

The structure of the two equilateral triangles consists of five distances of 1 and one distance of $\sqrt{3}$. Thus, its potential energy is $U_{\mathrm{LJ}}(\sqrt{(3)}) \approx 0.930041152$. Relative to this, all distances in $\mathcal{S}^{\star}_{4,2\mathrm{D}}$ are slightly smaller except for r_{23}.

Even though a square is intuitively more symmetric than $\mathcal{S}^{\star}_{4,2\mathrm{D}}$, it has r_{14} and r_{23} which are significantly larger than 1 compared to just the large distances r_{14} as the structure of two equilateral triangles forming a rhombus.

As a preparation for the optimization of $k = 5$ particles in 2D, we use the optimal configuration

$\mathcal{S}^{\star}_{4,2\text{D}}$ from above and the method from Sec. 6.2.3.3 to determine

$$U_{5,2\text{D,UB}} \leq 2.822464081988782 \tag{6.102}$$

by optimizing the position (x_5, y_5) of the fifth particle relative to $\mathcal{S}^{\star}_{4,2\text{D}}$.

Using this result in Eq. (6.25) together with the equation for r_{min} in Eq. (6.23), we have

$$r_{5,2\text{D,LB}} = r_{\text{min}}\left(U_{5,2\text{D,UB}}\right) \geq 0.8484840561227015. \tag{6.103}$$

As a result of the verified optimization, the overall potential was bound by

$$U^{\star}_{5,2\text{D}} = 2.821976245492^{24}_{03}. \tag{6.104}$$

The illustration of $\mathcal{S}^{\star}_{5,2\text{D}}$ in Fig. 6.21 is indistinguishable from the formation of three equilateral triangles. Only the distances between the individual particles and the values of the optimized variables provided by Tab. 6.9 can quantify the difference to a structure of equilateral triangles. Note that the values from Tab. 6.9 confirm the existence of the vertical symmetry axis through p_3.

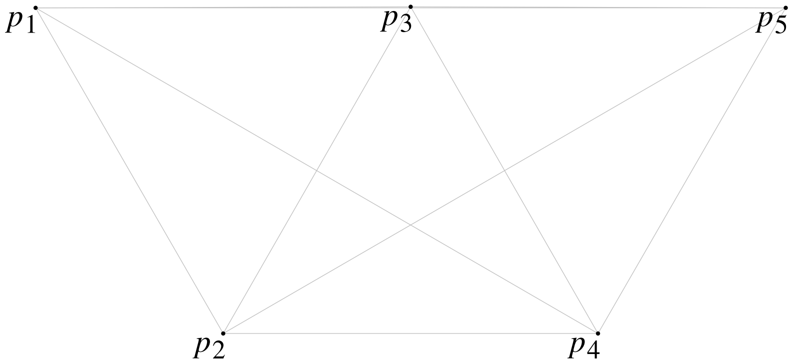

Figure 6.21: Minimum energy configuration of five particles in 2D, $\mathcal{S}^{\star}_{5,2\text{D}}$.

The distance between the major axis particles p_1 and p_5, r_{15}, is slightly below two. Particles p_2 and p_4 are pulled upwards, reducing their distance to each other and their distance to particles p_1 and p_5. Particle p_3 is slightly above the major axis, almost preserving the ideal distance of 1 to the particles p_2 and p_4.

To check that Taylor Models of order three are also the most time efficient calculation order for this Lennard-Jones optimization problem, we compare the performance of COSY-GO for different Taylor Model order for configuration of $k = 4$ and $k = 5$ particles in 2D in Tab. 6.10.

Table 6.9: Verified global optimization results for the minimum energy configurations of five particles in 2D, $\mathcal{S}^{\star}_{5,2\text{D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i.

k	i	j	$r^{\star}_{ij}$
5	1	2	0.99800^{7235}_{6984}
5	1	3	0.996784^{603}_{219}
5	1	4	1.72679^{5219}_{4649}
5	1	5	1.993561^{899}_{133}
5	2	3	1.000010^{593}_{180}
5	2	4	0.99610^{8166}_{7812}
5	2	5	1.72679^{5219}_{4649}
5	3	4	1.000010^{593}_{180}
5	3	5	0.996784^{603}_{219}
5	4	5	0.99800^{7235}_{6984}

k	†	i	$v^{\star}_{\dagger,i}$
5	x	1	0.498726^{87}_{66}
5	x	2	0.49805^{409}_{390}
5	x	3	0.49805^{409}_{390}
5	x	4	0.498726^{87}_{66}
5	y	2	-0.864459^{17}_{35}
5	y	3	0.002698^{85}_{63}
5	y	4	-0.864459^{17}_{35}

Table 6.10: Performance of verified global optimization using COSY-GO with QFB/LDB enabled on minimum energy search of a 2D configuration of k particles. The Taylor Model orders are denoted by 'O'.

		Computation time [s]				Steps			
k	$n_{2\text{D,var}}$	O2	O3	O4	O5	O2	O3	O4	O5
4	5	1.344	0.924	1.327	1.512	5031	3197	2935	2809
5	7	167.9	91.19	108.2	150.4	463758	295767	276720	266745

As observed in the 1D examples in Sec. 6.2.2.6 and Sec. 6.2.2.7, order three turns out to be the most time efficient calculation order. While the number of required steps is slightly improved with higher order calculations, the overall computation time is larger.

For configurations of $k = 6$ particles, the computation times on a single machine are very long. Because COSY-GO is implemented in a way that easily allows for parallel computations using MPI, we used parallel computations for the computationally time intensive cases. A critical aspect of the parallel COSY-GO is the time between processor communication and the associated load balancing. During the communication phase, the processors exchange information like redistributing their remaining domain boxes, as well as sharing their most recent cutoff values. If the time between communication is chosen to be too long, some processors will run idle without work while others still have a lot of boxes to evaluate. If the time is chosen too short, too much time is wasted on communication.

The timing for communication depends on multiple factors. Assume each processor runs the same repetitive code with different content and at some point, in the repetitive process, the code checks if it is time to communicate. If this is the case, the processor gathers all the information it wants to communicate and waits for all the other processors. The exchange of information happens for all involved processors. The more processors there are, the larger the communication time overhead.

To evaluate a good t_{com} for order three calculations, we investigate the optimization for six particles in 2D to determine $\mathcal{S}^{\star}_{6,2\text{D}}$ for various t_{com} with 64 cores (2 Nodes) and 1024 cores (32 Nodes) on Cori at NERSC [72], and the performance results are shown in Tab. 6.11.

Table 6.11: Performance of verified global optimization using parallel COSY-GO with QFB/LDB enabled for minimum energy search of a 2D configuration of six particles, $\mathcal{S}^{\star}_{6,2\text{D}}$. The parallel computations are run on Cori at NERSC using different communication timing t_{com}.

	64 cores (2 Nodes)		1024 cores (32 Nodes)	
t_{com} [s]	wall clock time [s]	steps	wall clock time [s]	steps
1	805.17	107070040	242.49	107504313
2	754.04	107074758	215.63	107302518
4	890.24	106822057	250.40	107350076

Even though the computation power, i.e., the number of cores and nodes, was increased by a factor of 16, the computation speed increased only by a factor of 3.5. While our studies on the parallel computation performance of parallel COSY-GO are still limited, this comparison illustrates some problems associated with communication. A good scheme of load balancing is important and parallel COSY-GO takes this into account.

According to our analysis, $t_{\text{com}} = 2$ s seems efficient, so we will use it for all following parallel computations.

For $k = 6$ particles in 2D, we use $\mathcal{S}^{\star}_{5,2\text{D}}$ from above and the method from Sec. 6.2.3.3 to determine

$$U_{6,2\text{D,UB}} \leq 5.643647992876073 \tag{6.105}$$

by optimizing the position (x_6, y_6) of the sixth particle relative to $\mathcal{S}^{\star}_{5,2\mathrm{D}}$.

Using this result in Eq. (6.25) together with the equation for $r_{\min}$ in Eq. (6.23), we have

$$r_{6,2\mathrm{D,LB}} = r_{\min}\left(U_{6,2\mathrm{D,UB}}\right) \geq 0.8164709262289850. \tag{6.106}$$

As a result of the verified optimization, the overall potential was bound by

$$U^{\star}_{6,2\mathrm{D}} = 5.641725650994^{96}_{65}. \tag{6.107}$$

In Fig. 6.22, $\mathcal{S}^{\star}_{6,2\mathrm{D}}$ is illustrated and Tab. 6.12 lists the distances between the particles and the associated results for the optimized variables. Note that the values from Tab. 6.12 confirm the symmetry axis through p_2 and p_4.

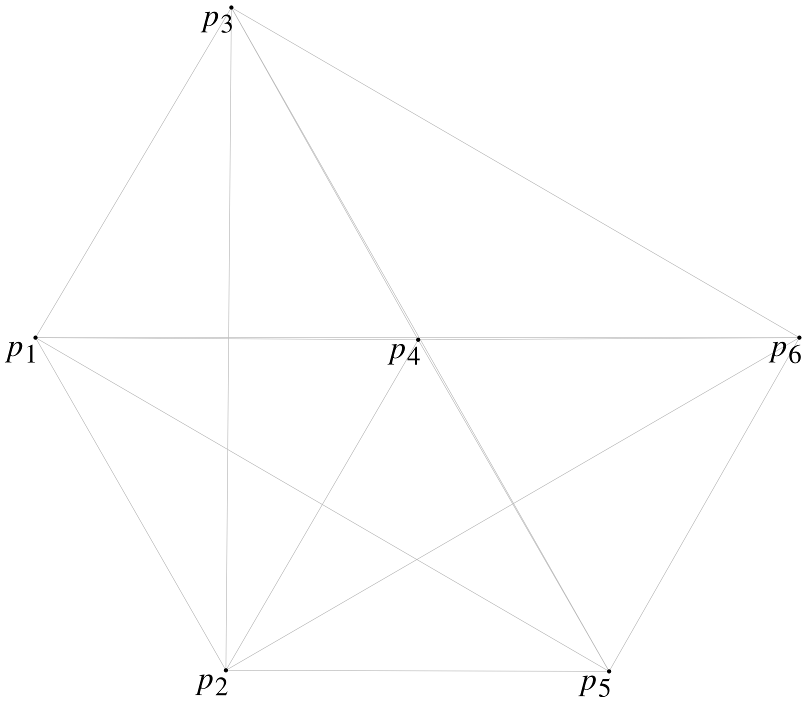

Figure 6.22: Minimum energy configuration of six particles in 2D, $\mathcal{S}^{\star}_{6,2\mathrm{D}}$.

The configuration shown in Fig. 6.22 is composed of four almost equilateral triangles. The connecting lines between the particle positions in Fig. 6.22 show the shape look like an envelope (upside down). In the configuration, there are nine distances close to 1, four distances close to $\sqrt{3}$ from the height of two stacked triangles, and two distances with a length of slightly less than 2. As

Table 6.12: Verified global optimization results for the minimum energy configurations of six particles in 2D, $\mathcal{S}^{\star}_{6,\text{2D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i.

k	i	j	$r^{\star}_{ij}$	k	i	j	$r^{\star}_{ij}$
6	1	2	0.99821^{7170}_{6865}	6	2	6	1.726^{601007}_{599809}
6	1	3	1.000179^{841}_{324}	6	3	4	0.99288^{3479}_{2910}
6	1	4	0.99659^{7164}_{6109}	6	3	5	1.989450^{769}_{132}
6	1	5	1.72842^{2004}_{0806}	6	3	6	1.7111^{30509}_{29697}
6	1	6	1.9894^{51206}_{49694}	6	4	5	0.996596^{889}_{384}
6	2	3	1.726600^{611}_{205}	6	4	6	0.99288^{3425}_{2964}
6	2	4	0.99590^{8556}_{7753}	6	5	6	1.000179^{734}_{430}
6	2	5	0.99821^{7526}_{6508}				

k	†	i	$v^{\star}_{\dagger,i}$
6	x	1	0.495916^{31}_{05}
6	x	2	0.01435^{211}_{166}
6	x	3	0.48631^{410}_{372}
6	x	4	0.497547^{95}_{73}
6	x	5	0.495320^{77}_{51}
6	y	2	-0.866316^{44}_{65}
6	y	3	0.860224^{32}_{11}
6	y	4	-0.00540^{596}_{622}
6	y	5	-0.86891^{684}_{706}

we already saw previously, distances larger than 1 are shorter in $\mathcal{S}^{\star}_{6,\text{2D}}$ compared to a structure of actual equilateral triangles at the cost of the unit distances deviating from 1 to either smaller or larger values. The symmetry of the configuration shown in Fig. 6.22 is also captured by the values in Tab. 6.12.

For the computation of $\mathcal{S}^{\star}_{7,\text{2D}}$, we use $\mathcal{S}^{\star}_{6,\text{2D}}$ from above and the method from Sec. 6.2.3.3 to determine

$$U_{7,\text{2D,UB}} \leq 8.471671506833459 \tag{6.108}$$

by optimizing the position (x_7, y_7) of the seventh particle relative to $\mathcal{S}^{\star}_{6,\text{2D}}$.

Using this result in Eq. (6.25) together with the equation for r_{min} in Eq. (6.23), we have

$$r_{7,\text{2D,LB}} = r_{\text{min}}\left(U_{7,\text{2D,UB}}\right) \geq 0.7966957780184697. \tag{6.109}$$

As a result of the verified optimization, the overall potential was bound by

$$U^{\star}_{7,\text{2D}} = 8.465133482231^{309}_{263}. \tag{6.110}$$

The optimization was computed in parallel on 256 cores (8 Nodes) on Cori at NERSC in 2043 seconds (wall clock time) and 852446890 steps.

As Fig. 6.23 illustrates, the resulting minimum energy configuration $\mathcal{S}^{\star}_{7,2\text{D}}$ is highly symmetric. It is an equilateral hexagon with a side-length of about 0.996434 and an additional particle at its center.

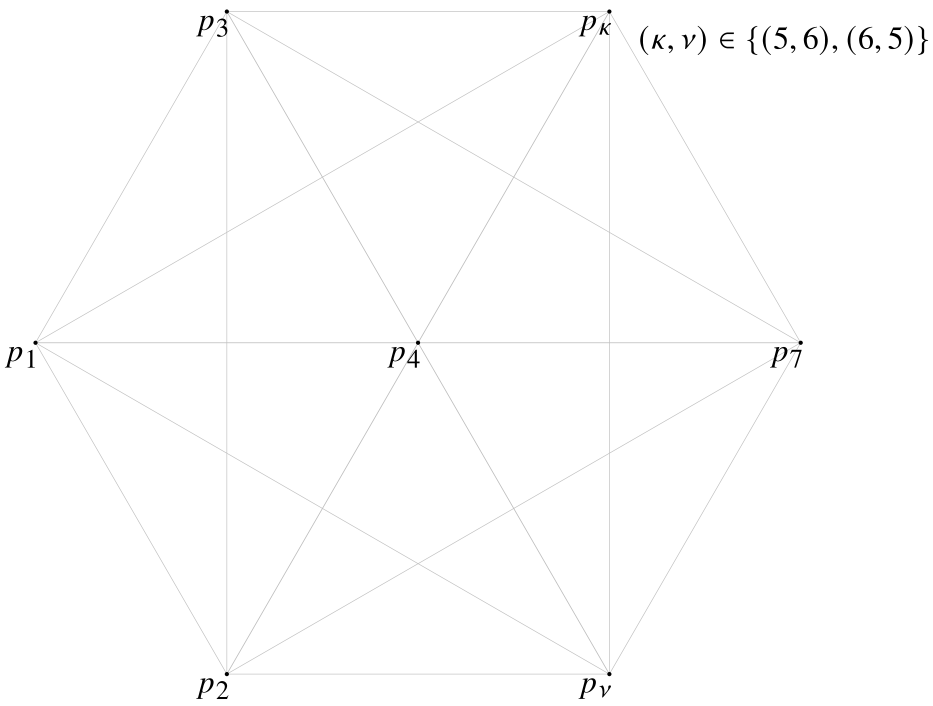

Figure 6.23: Minimum energy configuration of seven particles in 2D, $\mathcal{S}^{\star}_{7,2\text{D}}$. The configuration is represented by two equivalent numbering schemes.

This symmetry is further supported by the values for $r^{\star}_{ij}$ in Tab. 6.13. The table also shows the results for the optimized variables.

Due to the symmetry of the configuration with regard to the x axis, the optimizer finds a configuration for each of the two ambiguous numbering schemes. Specifically, particles p_2 and p_3 have the same x position, just like particles p_5 and p_6. However, because $y_2 \leq 0$, there is only an ambiguity in numbering the particles p_κ and p_ν from Fig. 6.23 either with $(\kappa, \nu) = (5, 6)$ or $(\kappa, \nu) = (6, 5)$. The optimizer yields a result for each of those two numbering schemes. Since both representations are equivalent, Tab. 6.13 lists the distances and variables for $(\kappa, \nu) = (5, 6)$.

Table 6.13: Verified global optimization results for the minimum energy configurations of seven particles in 2D, $\mathcal{S}^{\star}_{7,\text{2D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1} and $v^{\star}_{y,i}$ is the optimal y position of particle p_i. The values below are for the configuration $(\kappa, \nu) = (5, 6)$ from Fig. 6.23.

k	i	j	$r^{\star}_{ij}$	k	i	j	$r^{\star}_{ij}$
7	1	2	0.996434^{801}_{474}	7	3	4	0.996434^{879}_{482}
7	1	3	0.996434^{887}_{474}	7	3	5	0.996434^{891}_{384}
7	1	4	0.99643^{5078}_{4369}	7	3	6	1.99286^{9671}_{8964}
7	1	5	1.725875^{962}_{023}	7	3	7	1.725875^{962}_{023}
7	1	6	1.72587^{6111}_{5023}	7	4	5	0.996434^{879}_{482}
7	1	7	1.9928^{70155}_{68738}	7	4	6	0.996434^{879}_{396}
7	2	3	1.725875^{631}_{205}	7	4	7	0.99643^{5078}_{4369}
7	2	4	0.996434^{879}_{396}	7	5	6	1.725875^{631}_{205}
7	2	5	1.99286^{9671}_{8964}	7	5	7	0.996434^{887}_{474}
7	2	6	0.99643^{5235}_{4384}	7	6	7	0.996434^{801}_{474}
7	2	7	1.72587^{6111}_{5023}				

k	†	i	$v^{\star}_{\dagger,i}$
7	x	1	0.498217^{47}_{17}
7	x	2	0.000000^{+18}_{-01}
7	x	3	0.498217^{45}_{19}
7	x	4	0.498217^{45}_{19}
7	x	5	0.000000^{+18}_{-01}
7	x	6	0.498217^{47}_{17}
7	y	2	-0.862937^{60}_{82}
7	y	3	0.862937^{82}_{60}
7	y	4	0.000000^{+1}_{-1}
7	y	5	0.862937^{82}_{60}
7	y	6	-0.862937^{60}_{82}

6.2.3.8 The Verified Global Optimization Results for Configurations of *k* Particles in 3D

The setup for the optimization of configurations of k particles in 3D requires only minor additional definitions to the setup in 2D. However, to be able to read this study without having read the previous studies, we quickly summarize the process.

As discussed in Sec. 6.2.3.1, we use the center of mass of the configuration and its major axis to define the placement in the coordinate system. The x axis along the major axis is used to number the particles from 1 to k according to their x position such that

$$x_i \le x_j \quad \text{for} \quad i < j. \tag{6.111}$$

The particle p_1 is fixed to the origin with

$$\vec{p}_1 = (0, 0, 0) \tag{6.112}$$

and particle p_k is fixed to the positive x axis with

$$\vec{p}_k = (x_k \ge 0, 0, 0). \tag{6.113}$$

The y axis and the z axis are orientated such that

$$\vec{p}_2 = (x_2 \geq 0, y_2 \leq 0, z_2 = 0) \quad \text{and} \tag{6.114}$$

$$\vec{p}_3 = (x_3 \geq 0, y_3, z_3 \leq 0). \tag{6.115}$$

We describe a configuration of k particles in 3D by the variables $v_{x,i}$, $v_{y,i}$, and $v_{z,i}$, with

$$v_{x,i} = x_{i+1} - x_i \geq 0 \text{ for } i \in \{1, 2, ..., k-1\}, \tag{6.116}$$

$$v_{y,i} = y_i \text{ for } i \in \{2, 3, ..., k-1\}, \quad \text{and} \tag{6.117}$$

$$v_{z,i} = z_i \text{ for } i \in \{3, 4, ..., k-1\}, \tag{6.118}$$

as previously defined in Sec. 6.2.3.1 in Eq. (6.61), Eq. (6.62), and Eq. (6.63).

This yields a total number of

$$n_{\text{3D,var}} = 3k - 6 \tag{6.119}$$

optimization variables as mentioned in Eq. (6.65).

The variable domains were determined in Eq. (6.72), Eq. (6.73), and Eq. (6.74) in Sec. 6.2.3.4, with

$$v_{x,i} \in [0, 1] \text{ for} i \in \{1, 2, ..., k-1\}, \tag{6.120}$$

$$v_{y,2}, v_{z,3} \in [-1, 0] \, \frac{\sqrt{3}}{2} r_{k,\text{UB}}, \tag{6.121}$$

$$v_{y,i} \in [-1, 1] \, \frac{\sqrt{3}}{2} r_{\text{UB}} \text{ for} i \in \{3, ..., k-1\}, \quad \text{and} \tag{6.122}$$

$$v_{z,i} \in [-1, 1] \, \frac{\sqrt{3}}{2} r_{\text{UB}} \text{ for} i \in \{4, ..., k-1\} \quad \text{with} \quad r_{\text{UB}} = k - 1, \tag{6.123}$$

where r_{UB} is known from Eq. (6.71) in Sec. 6.2.3.2.

As discussed in Sec. 6.2.3.5, we use the modified Lennard-Jones potential from Eq. (6.78) as the objective function without changing the optimization problem. The lower bound $r_{k,\text{LB}}$ is determined according to Sec. 6.2.2.4. The squared inter-particle distances – the argument of this objective function – are calculated from $v_{x,i}$, $v_{y,i}$, and $v_{z,i}$ according to Eq. (6.66).

Following the center of mass requirement in the x direction and the major axis requirement from Sec. 6.2.3.1, we use the penalty functions from Eq. (6.55) and Eq. (6.69).

The verified global optimization is performed with the Taylor Model based verified optimizer COSY-GO [63, 64] in its most advanced setting with QFB/LDB enabled (see Sec. 2.6). Unless stated otherwise the optimization is performed with Taylor Models of order three. The threshold length as a stopping condition is $s_{\text{min}} = 10^{-6}$ as mentioned earlier in Sec. 6.2.2.6.

We start from $k = 4$. From Sec. 6.2.1.3, we know that the solution $\mathcal{S}^{\star}_{4,\text{3D}}$ as this trivial case is a regular tetrahedron. We note that in literature, the optimization of four particles in 3D is often used as a toy problem, which we discuss in the appendix A.1.

For the computation of $U_{5,\text{3D,UB}}$, we follow the procedure in Sec. 6.2.3.3 and represent $\mathcal{S}^{\star}_{4,\text{3D}}$ by the particle positions

$$\vec{p}_1 = (0,0,0)\,,\ \vec{p}_2 = \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}, 0\right),\ \vec{p}_3 = \left(\frac{1}{2}, -\frac{\sqrt{3}}{6}, -\sqrt{\frac{2}{3}}\right),\ \text{and}\ \vec{p}_4 = (1,0,0) \tag{6.124}$$

with

$$U^{\star}_{4,\text{3D}} = 0. \tag{6.125}$$

Then, we optimize the position (x_5, y_5) of a fifth particle p_5 relative to $\mathcal{S}^{\star}_{4,\text{3D}}$. The initial search domain for p_5 according to Sec. 6.2.3.3 and Fig. 6.15 is

$$\vec{p}_5 = (x_5, y_5, z_5) \in [-1,2] \times \left[-1 - \frac{\sqrt{3}}{2}, 1\right] \times \left[-1 - \sqrt{\frac{2}{3}}, 1\right]. \tag{6.126}$$

The optimization yields an upper bound

$$U_{5,\text{3D,UB}} \le 0.8968457347060826 \tag{6.127}$$

Using this upper bound in Eq. (6.25) together with the equation for r_{min} in Eq. (6.23) according to the method in Sec. 6.2.2.4, we have

$$r_{5,\text{3D,LB}} = r_{\text{min}}\left(U_{5,\text{3D,UB}}\right) \ge 0.8948940496635427. \tag{6.128}$$

As a result of the verified optimization, the overall potential was bound by

$$U^{\star}_{5,\text{3D}} = 0.896147584292^{45}_{18}. \tag{6.129}$$

The optimization was computed in parallel on 64 cores (2 Nodes) on Cori at NERSC in 74.28 seconds (wall clock time) and 2466118 steps.

As Fig. 6.24 illustrates, $\mathcal{S}^{\star}_{5,3\mathrm{D}}$ is very similar to a regular double-tetrahedron. This is further supported by the values of the optimization variables and $r^{\star}_{ij}$ in Tab. 6.14.

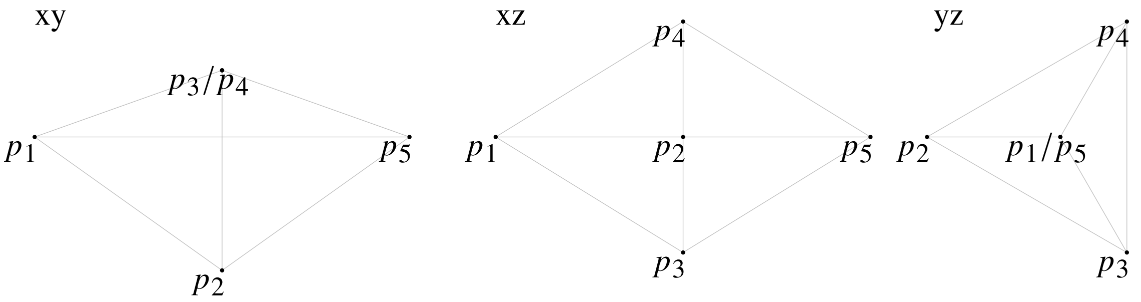

Figure 6.24: Minimum energy configuration of five particles in 3D, $\mathcal{S}^{\star}_{5,3\mathrm{D}}$. The configuration is shown in 2D projections, the xy plane projection (left), the xz plane projection (middle), and the yz plane projection (right). The solution consists of a central equilateral triangle spanned by the particles p_2, p_3, and p_4 in the yz plane, and one particle each centered above and below that triangle. In other words, it is similar to a double tetrahedron, which is slightly squished along the major axis (the x axis) increasing the side-length of the equilateral triangle in the middle to values slightly larger than one. The inter-particle distances are shown in Tab. 6.14.

Table 6.14: Verified global optimization results for the minimum energy configurations of five particles in 3D, $\mathcal{S}^{\star}_{5,3\mathrm{D}}$. The $r^{\star}_{ij}$ yield the optimal distance between particles p_i and p_j. $v^{\star}_{x,i}$ is the optimal x distance between particles p_i and p_{i+1}. $v^{\star}_{y,i}$ and $v^{\star}_{z,i}$ are the optimal y and z positions of particle p_i.

k	i	j	$r^{\star}_{ij}$
5	1	2	0.99790^{7007}_{6764}
5	1	3	0.99790^{7159}_{6694}
5	1	4	0.99790^{7242}_{6694}
5	1	5	1.626671^{944}_{396}
5	2	3	1.001453^{817}_{231}
5	2	4	1.001453^{817}_{231}
5	2	5	0.99790^{7172}_{6764}
5	3	4	1.001453^{816}_{232}
5	3	5	0.99790^{7159}_{6694}
5	4	5	0.99790^{7077}_{6694}

k	†	i	$v^{\star}_{\dagger,i}$
5	x	1	0.813335^{87}_{69}
5	x	2	0.000000^{+11}_{-01}
5	x	3	0.000000^{+11}_{-01}
5	x	4	0.813335^{87}_{69}
5	y	2	-0.578189^{37}_{55}
5	y	3	0.289094^{90}_{56}
5	y	4	0.289094^{90}_{56}
5	z	3	-0.500726^{61}_{91}
5	z	4	0.500726^{91}_{61}

Even though particles p_2, p_3 and p_4 all seem to have the same x coordinate, there is no ambiguous numbering scheme due to the definition of the coordinate system with $y_2 \leq 0$, $z_2 = 0$ and $z_3 \leq 0$.

The regular double tetrahedron consists of nine unit distances and the major axis of length $2\sqrt{2/3}$. $\mathcal{S}^{\star}_{5,3\text{D}}$ is a slightly 'squished' version of the regular double tetrahedron along the major axis. The major axis and the distances in the direction of the major axis are shortened. Only the three inter-particle distances between p_2, p_3, and p_4 are slightly longer than unit length and form an equilateral triangle.

6.2.4 Summary

This section illustrated the many critical aspects of verified global optimization and the capabilities of COSY-GO in its most advanced setting QFB/LDB. Despite the high dimensionality, the strong interdependence, and nonlinearity of the optimization problem, COSY-GO was able to rigorously determine the minimum energy configurations.

In Tab. 6.15, we summarize the results for the global minimum of Lennard-Jones configurations in 2D and 3D and also provide the corresponding values for $U_{k,\text{lit},n^{\star}_{\text{dim}}}$ using Eq. (6.12) for easier comparison with literature.

Table 6.15: Summary of verified global optimization results on the minimum energy of configurations in 2D and 3D.

n_{dim}	k	n_{pairs}	$U^{\star}_{k,n_{\text{dim}}}$	$U^{\star}_{k,\text{lit},n_{\text{dim}}} = U^{\star}_{k,n_{\text{dim}}} - n_{\text{pairs}}$
2	4	6	0.926579141537^{22}_{07}	$-5.073420858462^{78}_{93}$
2	5	10	2.821976245492^{24}_{03}	$-7.178023754507^{76}_{97}$
2	6	15	5.641725650994^{96}_{65}	$-9.358274349005^{04}_{35}$
2	7	21	$8.46513348231^{309}_{263}$	$-12.53486651768^{691}_{737}$
3	5	10	0.896147584292^{45}_{18}	$-9.103852415707^{55}_{82}$

6.3 Verified Stability Analysis of Dynamical Systems

Verified calculations are particularly important for the stability analysis of dynamical systems. With a verified upper bound on the rate of divergence, a system's long term stability can be rigorously estimated. Both of the previously discussed applications in Chapter 4 and Chapter 5 will benefit to different degrees from such verified stability estimates.

6.3.1 The Potential Implications for the Bounded Motion Problem

For the bounded motion orbits under zonal perturbation in the Earth's gravitational field (see Chapter 4), a stability estimate is the maximum rate at which two bounded orbits drift apart. Below we want to list aspects to consider for the calculation of such a verified upper bound on the rate of divergence.

The bounded motion conditions from Sec. 4.2.5 require that the average nodal period $\overline{T}_d$ and the average drift of the ascending node $\overline{\Delta\Omega}$ of two bounded orbits are the same. In other words, two orbits drift apart if those two averaged quantities are not the same for the two orbits. Additionally, each of the orbits might be diverging on its own by slowly increasing or decreasing its distance from the Earth. A verified upper bound on each of those diverging factors must be determined to combine them to an overall verified upper bound on the rate at which the two bounded orbits drift apart.

An upper bound on the radial drift rate of the bounded orbits moving apart is determined by the maximum difference between the individual radial drifts of each of the bounded orbits. The normal form defect of the radial phase space can be used as a measure for this radial drift. However, both the maximum and the minimum normal form defect of each orbit are relevant to determine the worst-case scenario of one of the orbits decreasing its amplitude and one of the orbits increasing its amplitude.

The longitudinal drift rate of the bounded orbits moving apart is determined by the difference in the average revolution frequency of the orbital planes around the symmetry axis. The revolution frequency is proportional to the drift of the ascending node $\overline{\Delta\Omega}$ per nodal period $\overline{T}_d$. Since both of these quantities are oscillating at the same rate, the average revolution frequency can be calculated as the ratio of the average drift of the ascending node $\overline{\Delta\Omega}$ and the average nodal period $\overline{T}_d$.

Even if the orbital planes of the two bounded orbits are not radially or longitudinally drifting apart, the satellites on those orbits might still be drifting apart due to different average nodal periods, which constitutes the third drift factor.

These three factors have to be taken into account and rigorously estimated to calculate an overall maximum drift rate. The combination of the individual factors is not trivial since they are not independent of each other, e.g., the individual radial drifts of the orbits have nonlinear influences on the bounded motion quantities $\overline{\Delta\Omega}$ and $\overline{T}_d$. Verified global optimization of the overall drift rate is required to determine the maximum rate of divergence for any possible combination of the individual radial drift rates.

Given that the overall maximum drift rate is formally defined, we need to determine verified versions of the involved quantities. Accordingly, the starting point of the rigorous calculation of the maximum drift rate is a rigorous map of the system.

The map is based on the equations of motion of the system, which include the zonal coefficients of the Earth's gravitational potential based on measurements. To be rigorous it has to be decided if these coefficients are assumed to be exact or if the uncertainty about these coefficients is considered in the calculation. Given that the approach from Chapter 4 considers the zonal problem, ignoring sectional and tesseral terms, it seems reasonable to consider an idealized system where these coefficients are assumed to be exact.

In the next step, the verified integration of the equations of motion is required to calculate a verified map representation of the system [22, 28]. In our approach (see Chapter 4), we express the vertical momentum component v_z in terms of the other variables and system parameters. This operation includes the calculation of an inverse, which requires special methods to be performed rigorously [41, 20]. For the projection of the transfer map onto the Poincaré surface representing a generalized ascending node state, another rigorous computation of an inverse is required [41, 20]. Additionally, every step of the normal form based averaging procedure from Sec. 4.3.4 for the determination of the averaged quantities $\overline{\Delta\Omega}$ and $\overline{T}_d$ has to be performed rigorously. The approach then calls for another inversion to calculate the constants of motion $\mathcal{H}_z$ and E as a function of the phase space variables such that the averaged bounded motion quantities match between any two orbits in the phase space.

If all those procedures are performed rigorously, one can calculate rigorous bounds on the normal form defect of the system, which can then be used together with the rigorous estimations of the averaged quantities $\overline{\Delta\Omega}$ and $\overline{T}_d$ to calculate the rigorous overall rate of divergences.

In summary, much effort is required to establish a verified upper bound on the maximum rate at which bounded orbits of the zonal problem drift apart. However, the practical implications of such an estimate are limited since the approach does not consider the fully perturbed system. Accordingly, we want to focus our attention on the application of a rigorous stability analysis for the system discussed in Chapter 5.

6.3.2 The Implications for the Stability Analysis of the Muon g-2 Storage Ring

A verified stability estimate of the Muon g-2 Storage Ring can be obtained from the verified maximum rate at which particles escape the storage region of the storage ring. A measure of this rate of divergence is the normal form defect.

In Chapter 5, we saw that the size of the normal form defect that a particle encounters correlates with its likelihood of getting lost. As mentioned before and discussed in [85], the number of lost particles is very important for this high precision experiment, because the losses introduce a systematic bias for the average polarization of the remaining particles, which influences the overall result of the measurement. Below we want to analyze the aspects to consider for the calculation of such a verified upper bound on the rate of divergence in form of the normal form defect using Taylor Model based verified global optimization.

For the fully verified normal form defect analysis, we require a verified phase space map of the storage ring. As already mentioned before, there are many intricacies to consider for a fully rigorous map calculation. A major challenge regarding the verified calculation of the storage ring map is the verified representation of every storage ring component, including all its perturbations, e.g., perturbations from ESQ fringe fields and imperfection in the magnetic field. Because further work is required to generate such a fully verified map of the Muon g-2 Storage Ring, we will proceed with the nonverified tenth order map from Chapter 5 with an ESQ voltage of 18.3 kV. Assuming this map captures all of the relevant dynamics, the difference between using a verified map and a nonverified map is very small. To assess whether our computation order is high enough to capture the relevant dynamics, we estimate inaccuracies in the map by computing maps of various orders and showing that these inaccuracies – the main numerical error which is not based on measurement errors – are sufficiently small and will not affect the analysis result in a meaningful way when using the storage ring map of order ten.

For comparison, we will additionally analyze a storage ring map that considers an ESQ voltage of 17.5 kV instead of 18.3 kV. The tunes of particles under the influence of an ESQ voltage of 17.5 kV are further away from the vertical $1/3$ resonance tune. Accordingly, we expect no period-3 fixed point structures with their unstable fixed points and therefore less diverging behavior for this map compared to the 18.3 kV map.

The goal is to rigorously analyze the stability of the entire five dimensional storage phase space $(x, a, y, b, \delta p)$ of the storage ring maps using verified global optimization of the normal form defect. In Sec. 6.3.3, we specify the normal form defect function as the objective function of the optimization problem. To be able to distinguish the diverging behavior in different areas of the storage region, we divide the five dimensional space into partitions. Each of those partitions is then used as the search domain for the verified global optimizer to find the maximum normal form defect in it. In Sec. 6.3.4, we present the onion layer approach [29, 13], which divides the storage region according to the dynamics in the phase space. Next, we illustrate the complexity and strong nonlinearity of the normal form defect in multiple such onion layers and how it changes for different phase space regions and ESQ voltages (see Sec. 6.3.5). In Sec. 6.3.6, the results of the verified global optimization for the two maps with the different voltages are presented and compared to each other and the results of a nonverified analysis.

6.3.3 The Normal Form Defect as the Objective Function for the Optimization

In Sec. 2.4, the normal form defect for the propagation of a state $\vec{z}$ with a map $\mathcal{M}$ was introduced as the difference between the normal form radius of the mapped state $\mathcal{M}(\vec{z})$ and the normal form radius of the original state $\vec{z}$. If the motion occurs in multiple phase space dimensions, there is some ambiguity to the term 'normal form radius' and the associated normal form defect.

From the definition and algorithms of normal form transformations discussed in Sec. 2.3 and Sec. 2.4, it follows that there is a normal form radius for each normal form phase space. Each of these radii, yields the radius of the circular motion in this particular normal form phase space with

$$r_{\mathrm{NF},i}\left(\vec{z}_0\right) = \sqrt{\left(q_{\mathrm{NF,i}}\left(\vec{z}_0\right)\right)^2 + \left(p_{\mathrm{NF,i}}\left(\vec{z}_0\right)\right)^2}. \tag{6.130}$$

Accordingly, as defined in Sec. 2.4, there is a normal form defect defined for each of those normal form radii, with

$$d_{\mathrm{NF},i}\left(\vec{z}_0\right) = r_{\mathrm{NF},i}\left(\mathcal{M}\left(\vec{z}_0\right)\right) - r_{\mathrm{NF},i}\left(\vec{z}_0\right). \tag{6.131}$$

Additionally, we define the (overall) normal form radius of the motion as the Euclidean distance

$$r_{\mathrm{NF}}\left(\vec{z}_0\right) = \sqrt{\sum_i r^2_{\mathrm{NF},i}\left(\vec{z}_0\right)}. \tag{6.132}$$

This definition of the (overall) normal form radius corresponds to the following definition for the (overall) normal form defect

$$d_{\mathrm{NF}}\left(\vec{z}_0\right) = r_{\mathrm{NF}}\left(\mathcal{M}\left(\vec{z}_0\right)\right) - r_{\mathrm{NF}}\left(\vec{z}_0\right). \tag{6.133}$$

Unless stated otherwise, we will be using and referring to the (overall) normal form radius and the (overall) normal form defect.

6.3.4 The Search Domain in the Form of Onion Layers

The onion layer approach describes a way to partition the phase space regions and determine the associated variables for the verified global optimization. For the partitioning, it is important to consider the dynamics of the system. In Chapter 5, we saw that the main characteristics of the phase space motion in the storage ring are the oscillation amplitudes and the momentum offset δp. Accordingly, we want to calculate the verified stability estimates on the rate of divergence based on partitions categorized by those criteria.

While the partitioning according to the momentum offset δp is straightforward, defining the partitions of different phase space amplitudes is not, because the phase space curve of a particle with a certain amplitude forms a nonlinearly distorted elliptical shape in the original phase space. The onion layer approach (see Fig. 6.25) partitions the phase space along those nonlinearly distorted elliptical phase space curves using the normal form transformation.

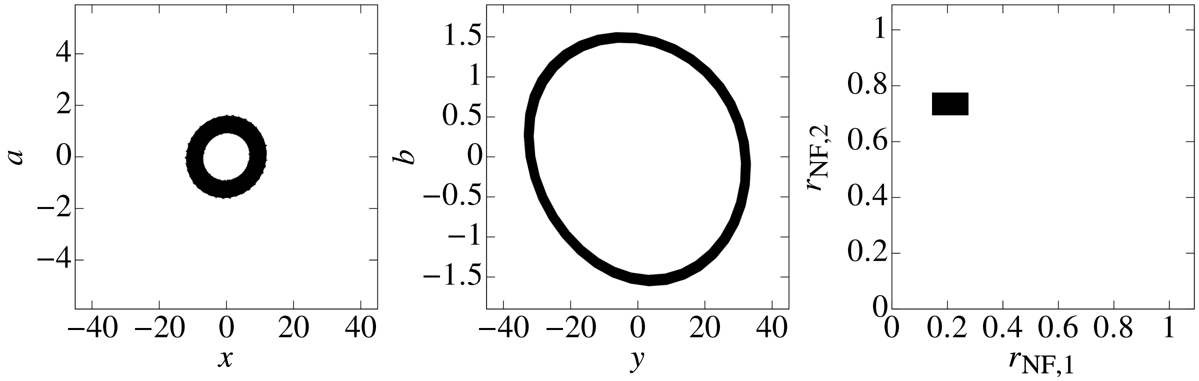

Figure 6.25: The left and the middle plot show the representation of an onion layer (black region) in regular phase space coordinates. The thickness of the onion layer is determined by the range in $r_{\mathrm{NF},1}$ and $r_{\mathrm{NF},2}$ as well as the range in δp. For this particular example, we set δp to a fixed value of $\delta p = 0\%$ instead of a range. The range in the normal form radii is given by $r_{\mathrm{NF},1} \in [0.15, 0.25]$ and $r_{\mathrm{NF},2} \in [0.7, 0.75]$. Note that the thickness in $r_{\mathrm{NF},1}$ is twice the thickness in $r_{\mathrm{NF},2}$. Accordingly, the projection of the onion layer into the radial phase space (x, a) appears roughly twice as thick as the projection into the vertical phase space (y, b).

As illustrated in Fig. 6.25, the normal form coordinates allow us to partition by amplitude. They are our best approximation of mapping the orbital phase space behavior onto circles. Accordingly, we can use the normal form description of the motion to define the onion layers for the global optimization. Specifically, we chose the normal form radii $r_{\mathrm{NF},1}$ and $r_{\mathrm{NF},2}$ as well as the corresponding normal form phase space angles $\phi_{\mathrm{NF},1}$ and $\phi_{\mathrm{NF},2}$ as the optimization variables. Additionally, the momentum offset δp is also considered an optimization variable.

The normal form phase space variables $(q_{\mathrm{NF},1}, p_{\mathrm{NF},1})$ and $(q_{\mathrm{NF},2}, p_{\mathrm{NF},2})$ are expressed in terms of the polar optimization variables with

$$\begin{pmatrix} q_{\mathrm{NF},1} \\ p_{\mathrm{NF},1} \end{pmatrix} = r_{\mathrm{NF},1} \begin{pmatrix} \cos\left(\phi_{\mathrm{NF},1}\right) \\ \sin\left(\phi_{\mathrm{NF},1}\right) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} q_{\mathrm{NF},2} \\ p_{\mathrm{NF},2} \end{pmatrix} = r_{\mathrm{NF},2} \begin{pmatrix} \cos\left(\phi_{\mathrm{NF},2}\right) \\ \sin\left(\phi_{\mathrm{NF},2}\right) \end{pmatrix}. \tag{6.134}$$

The inverse normal form transformation $\mathcal{A}^{-1}$ is then used as a vehicle to express the relevant phase space regions in original phase space (x, a, y, b) in terms of the optimization variables $(r_{\mathrm{NF},1}, \phi_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \phi_{\mathrm{NF},2}, \delta p)$.

Moving along the angles $\phi_{\mathrm{NF},1}$ and $\phi_{\mathrm{NF},2}$ will approximately move along the phase space curve in the original coordinates. Accordingly, the search domain in those optimization variables is always $[-\pi, \pi]$. The domain on the normal form radii and the momentum offset determines the thickness of the onion layer, as illustrated in Fig. 6.25, and is set to 0.04 for normal form radii and to 0.04% in the momentum offset space.

6.3.5 The Complexity and Nonlinearity of the Normal Form Defect Function

In Chapter 5, we analyzed the normal form defect that individual particles encounter during stroboscopic tracking. In other words, we only probed individual phase space points of a particle's orbit for its normal form defect. We found that muons that encounter phase space regions with larger normal form defects are more likely to get lost (see Fig. 5.12). However, the probing only yields an incomplete picture of the normal form defect that a particle can potentially encounter. Fig. 6.26 illustrates how much the normal form defect can vary for fixed normal form amplitudes that approximately represent the normal form defect landscape along the phase space curve of a single particle.

Fig. 6.26 illustrates the normal form defect $d_{\mathrm{NF},1}$ of an onion layer of zero thickness, which is given by a single point in the 3D onion layer thickness space of $r_{\mathrm{NF},1}$, $r_{\mathrm{NF},2}$, and δp. The landscape is characterized by highly nonlinear behavior with many local minima and maxima, which are extreme points of very steep valleys and hills. Accordingly, the stroboscopic normal form defect probing while tracking can significantly underestimate the maximum normal form defect of an orbit

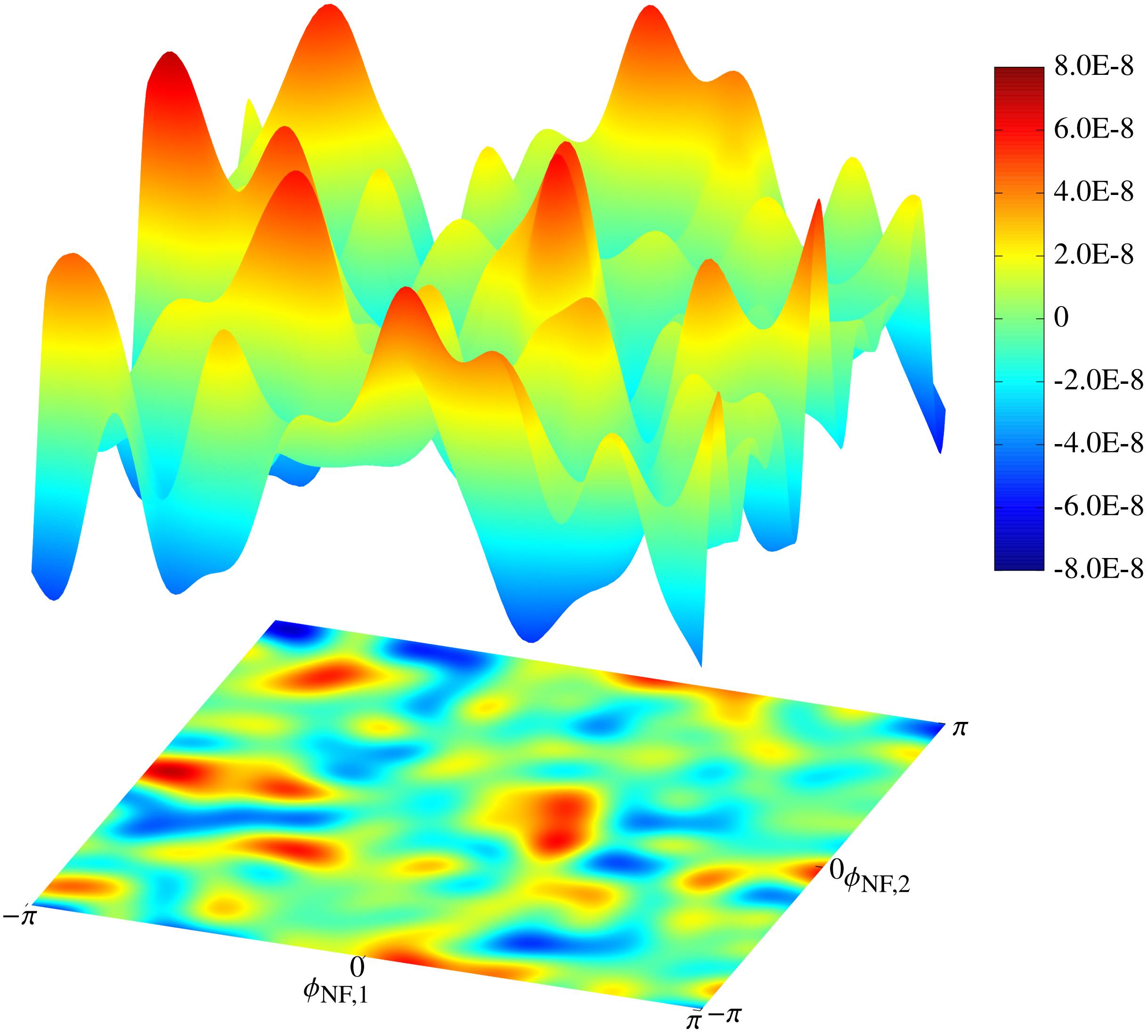

Figure 6.26: Normal form defect landscape of the radial phase space in $\phi_{\mathrm{NF},1}$ and $\phi_{\mathrm{NF},2}$ for fixed normal form amplitudes of $r_{\mathrm{NF},1} = 0.4$ and $r_{\mathrm{NF},2} = 0.4$, and with $\delta p = 0\%$. The underlying map considers an ESQ voltage of 18.3 kV.

in a certain phase space region, which motivates a rigorous analysis of the normal form defect for those phase space regions.

In Fig. 6.27 and Fig. 6.28, the normal form defect landscapes in the vertical and radial direction are shown for maps considering an ESQ voltage of 18.3 kV and 17.5 kV, respectively. The different normal form defect landscapes emphasize how much the landscapes change in shape and magnitude for different normal form phase space points.

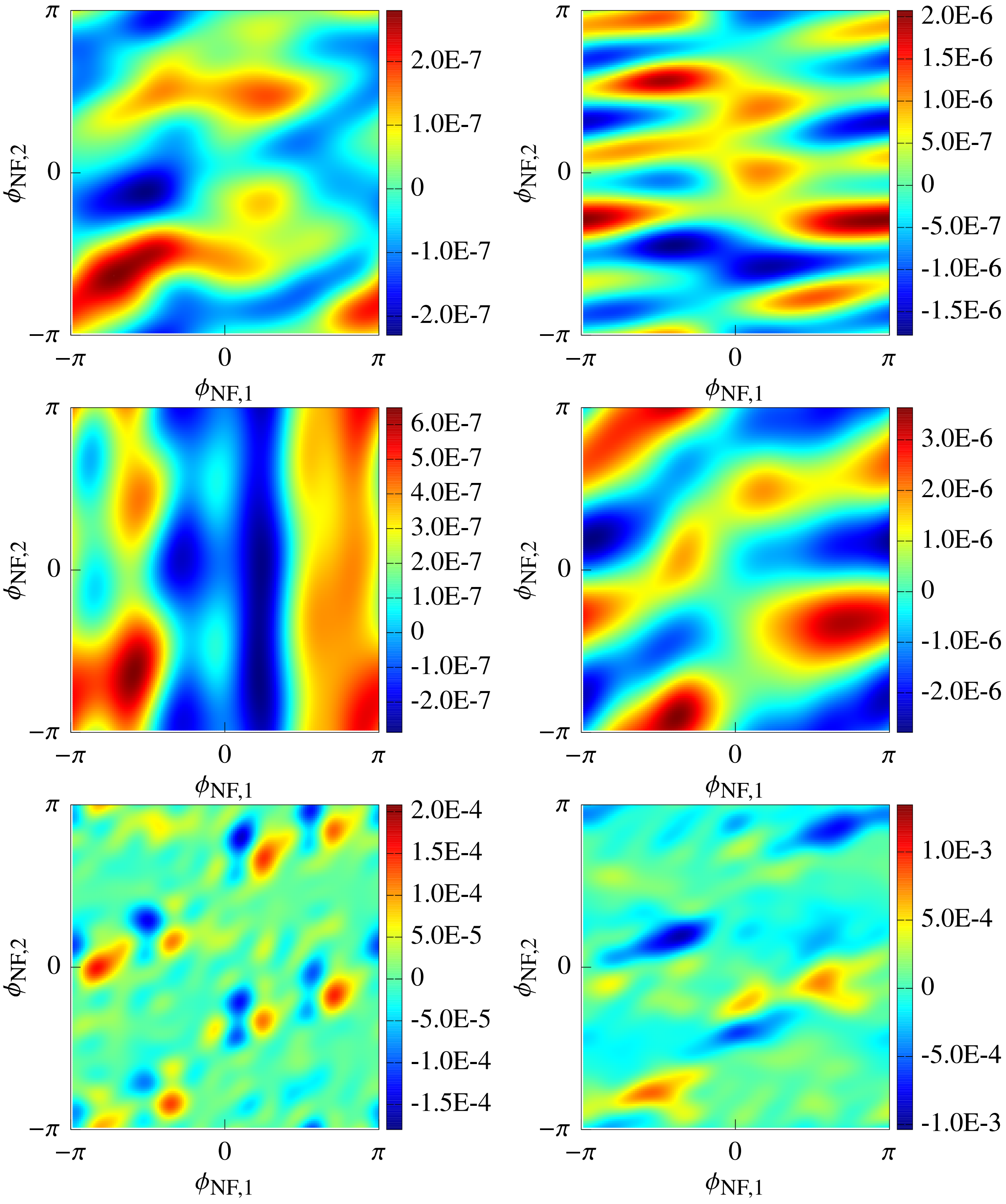

Figure 6.27: The normal form defect landscape of the radial (left side) and vertical (right side) phase space for multiple onion layers of zero thickness, which are characterized by $(r_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \delta p)$. The top row corresponds to $(0.1, 0.2, 0.24\%)$, the middle row corresponds to $(0.2, 0.05, 0.24\%)$, and the bottom row corresponds to $(0.56, 0.72, 0.04\%)$. The underlying map considers an ESQ voltage of **18.3 kV**.

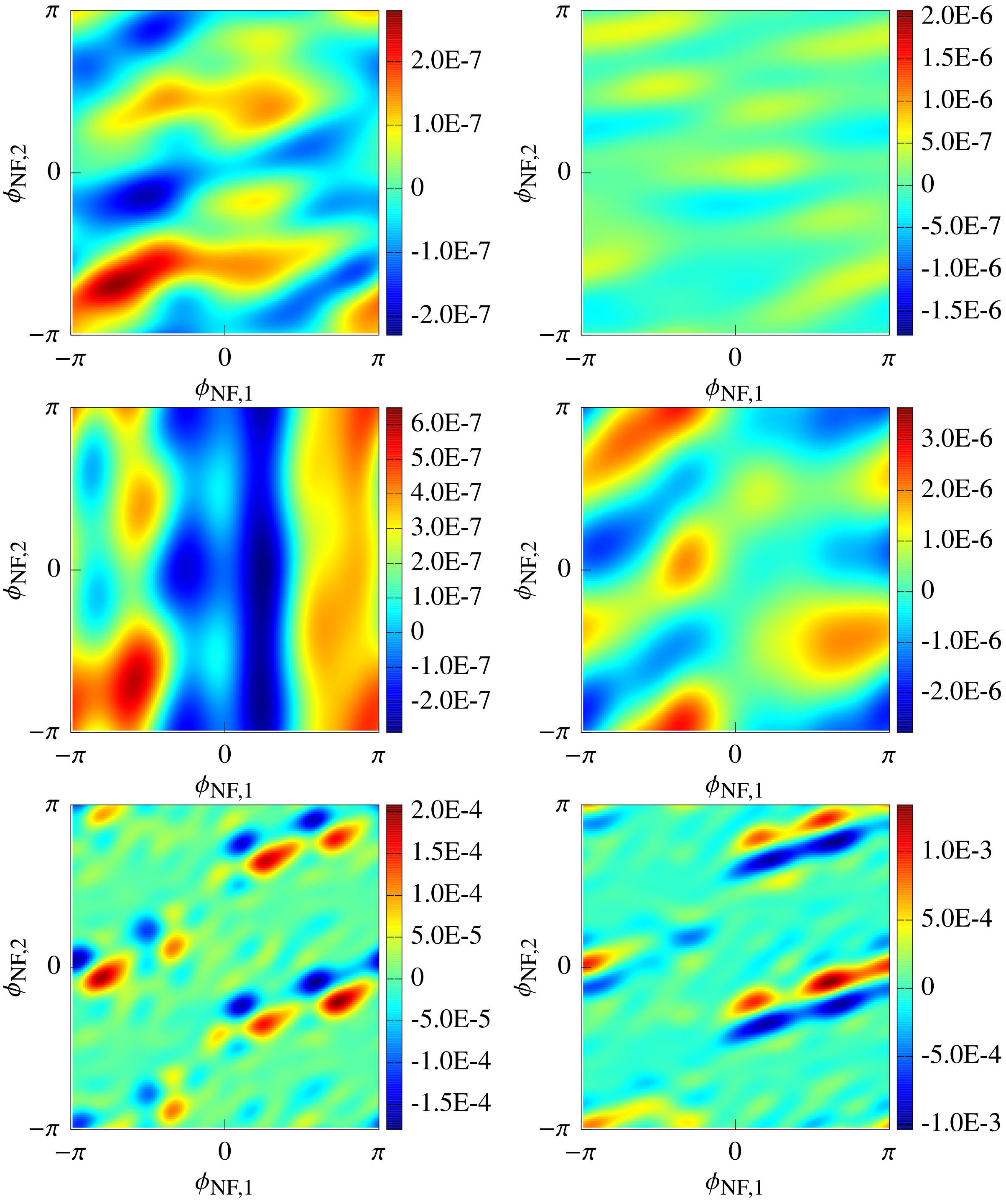

Figure 6.28: The normal form defect landscape of the radial (left side) and vertical (right side) phase space for multiple onion layers of zero thickness, which are characterized by $(r_{\text{NF},1}, r_{\text{NF},2}, \delta p)$. The top row corresponds to $(0.1, 0.2, 0.24\%)$, the middle row corresponds to $(0.2, 0.05, 0.24\%)$, and the bottom row corresponds to $(0.56, 0.72, 0.04\%)$. The underlying map considers an ESQ voltage of **17.5 kV**.

Comparing the normal form defect of the radial and vertical phase space clearly shows the different orders of magnitude at play for those particular onion layers of zero thickness. The normal form defect of the vertical phase space is about 1.5 orders of magnitude larger than the normal form defect of the radial phase space.

The comparison between Fig. 6.27 and Fig. 6.28 shows something rather fascinating. Even though the normal form defect landscapes change so drastically for different phase space positions, they are very similar for the two maps at the same normal form positions. The magnitude of the normal form defect is usually higher for the 18.3 kV, but the example in the bottom row shows that there are also normal form phase space regions where it is the other way around.

The top row and middle row of Fig. 6.27 and Fig. 6.28 show phase space points with the same momentum offset and roughly the same overall normal form radius. While the magnitude of the normal form defects in the radial and vertical direction is roughly the same, the shape of the normal form defect landscape differs tremendously. For the global optimization, this means that the objective function looks vastly different for each of the onion layer search domains.

6.3.6 The Results of the Verified Global Optimization of the Normal Form Defect

As mentioned in Sec. 6.3.4, we partition the search space into onion layers of the size $0.04 \times 2\pi \times 0.04 \times 2\pi \times 0.04\%$ in $(r_{\mathrm{NF},1}, \phi_{\mathrm{NF},1}, r_{\mathrm{NF},2}, \phi_{\mathrm{NF},2}, \delta p)$. Based on the δp rage of the realistic particle distribution in Chapter 5, we investigate the δp space from -0.22% to $+0.42\%$ in 16 partitions of size 0.04%.

For each of those 16 pieces, we additionally partition the $(r_{\mathrm{NF},1}, r_{\mathrm{NF},2})$ space into boxes of size 0.04×0.04. To determine which of those boxes represent phase space behavior within the collimator region, we probe the bottom left corner of each box, namely, the point with the lowest normal form amplitudes $(r_{\mathrm{NF},1,\mathrm{min}}, r_{\mathrm{NF},2,\mathrm{min}})$ and check if those lowest amplitudes are already outside the collimator region in the original phase space coordinates. For the probing, we take $30 \times 30 \times 2$ testing points in $\phi_{\mathrm{NF},1}, \phi_{\mathrm{NF},2}, \delta p$ and map them back into the original phase space (x, a, y, b) using the inverse normal form transformation $\mathcal{A}^{-1}$. The two values for δp are the maximum and minimum momentum offset of the onion layer. A box is only analyzed if all of the 1800 probing points satisfy $\sqrt{x^2+y^2} < 0.045$ mm.

To benchmark the verified analysis, we also present a nonverified normal form defect analysis of the same onion layers. The nonverified analysis is based on probes of the top right corner of each box, namely, the point with the largest normal form amplitudes $(r_{\text{NF},1,\text{max}}, r_{\text{NF},2,\text{max}})$. The $30 \times 30 \times 2$ probing points in $\phi_{\text{NF},1}, \phi_{\text{NF},2}, \delta p$ are chosen the same way as above. This probing approach is used in the verified analysis as a method to obtain a good initial cutoff value for the verified global optimizer. Accordingly, the nonverified analysis provides a lower bound on the maximum normal form defect, while the verified analysis constitutes an upper bound.

The results on the following pages (see Fig. 6.29 to Fig. 6.32) are ordered such that the nonverified probing analysis can be compared to the verified global optimization by switching back and forth between pages. Fig. 6.29 and Fig. 6.30 respectively show the nonverified and verified analysis for the map with an ESQ voltage of 17.5 kV, while Fig. 6.31 and Fig. 6.32 respectively show the verified and the nonverified analysis for the map with an ESQ voltage of 18.3 kV. Additionally, the two verified normal form defect analyses in Fig. 6.30 and Fig. 6.31 for the map with an ESQ voltage of 17.5 kV and 18.3 kV, respectively, can be compared the same way.

The color scheme in Fig. 6.29 to Fig. 6.32 indicates the maximum normal form defect in each of the onion layers. Given the 0.04×0.04 box size of the onion layers in normal form space and the maximum normal form defect in the onion layer $d_{\text{NF},\text{max}}$, we can calculate the minimum number of turns N required to cross through each onion layer as a Nekhoroshev-type stability estimate with

$$N = \frac{0.04}{d_{\text{NF},\text{max}}}. \tag{6.135}$$

The inner white onion layers have a maximum normal form defect below 10^{-5}. Accordingly, even in the worst case, it takes at least 4000 turns to cross the respective onion layer. It takes at least 400 turns to cross a yellow onion layer by the same measure and at least 40 turns to cross an orange onion layer. Red onion layers take at least 12 turns to cross, and black onion layers can, in the worst case, be crossed in fewer turns.

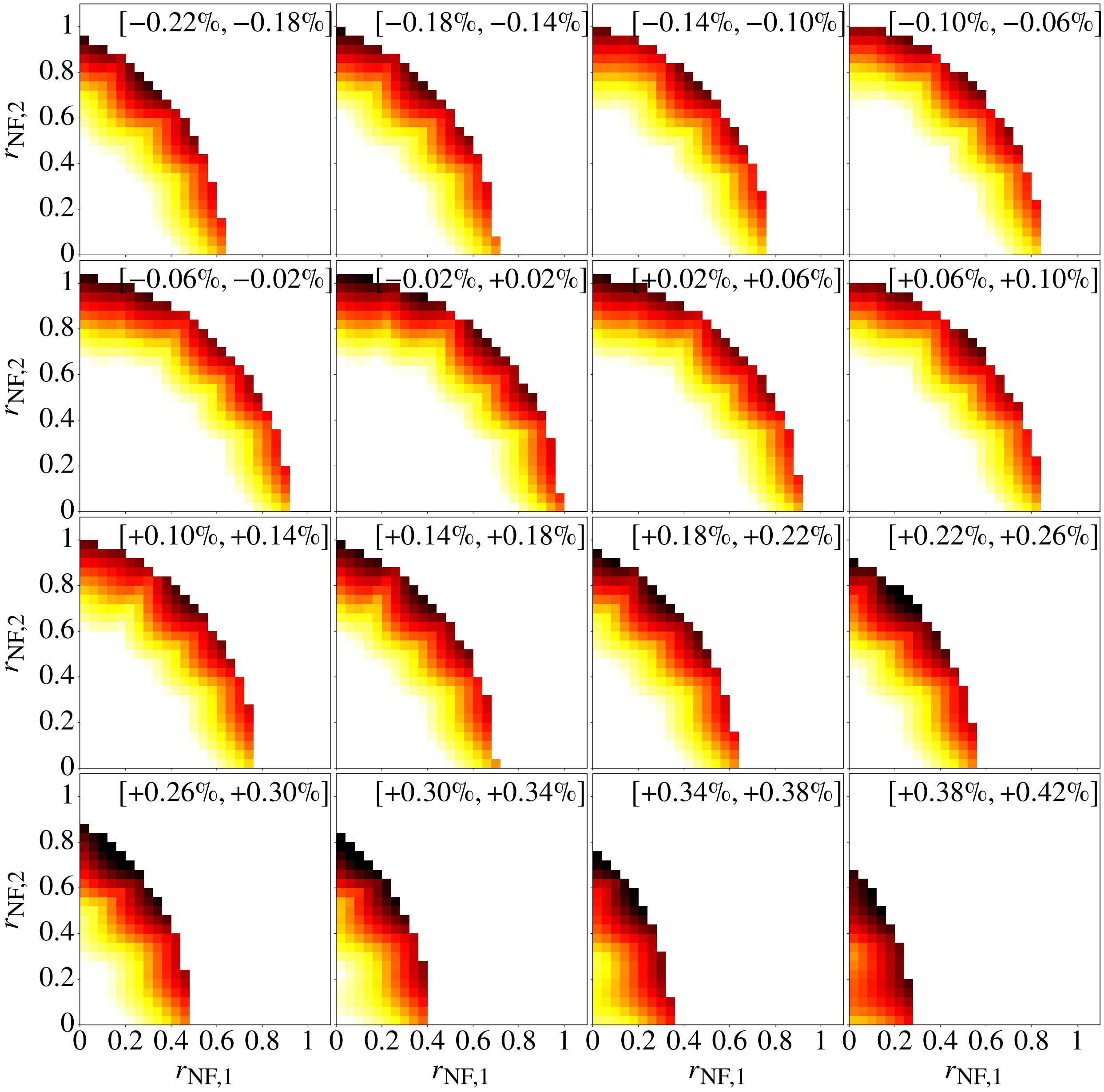

Figure 6.29: **Nonverified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

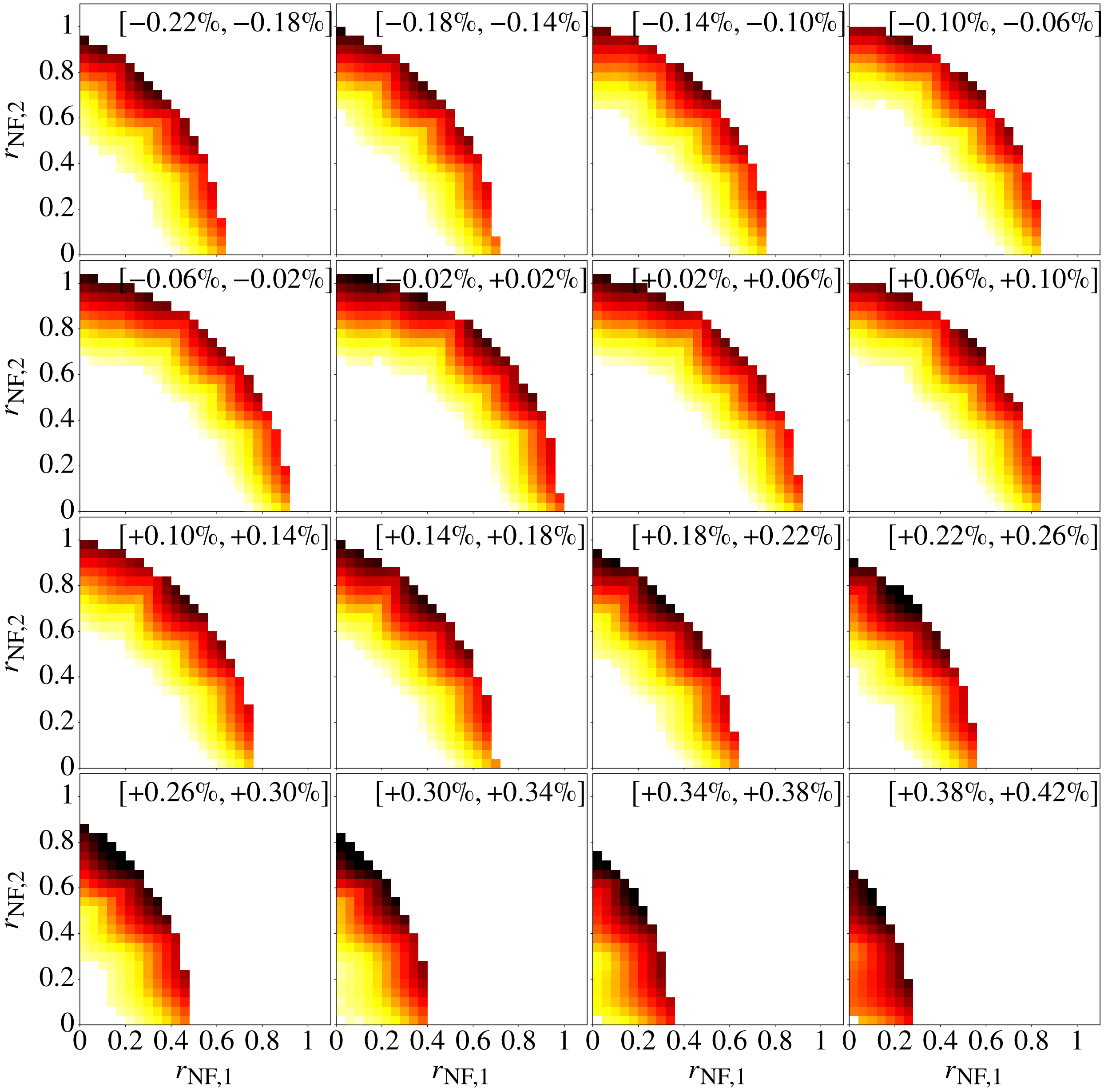

Figure 6.30: **Verified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

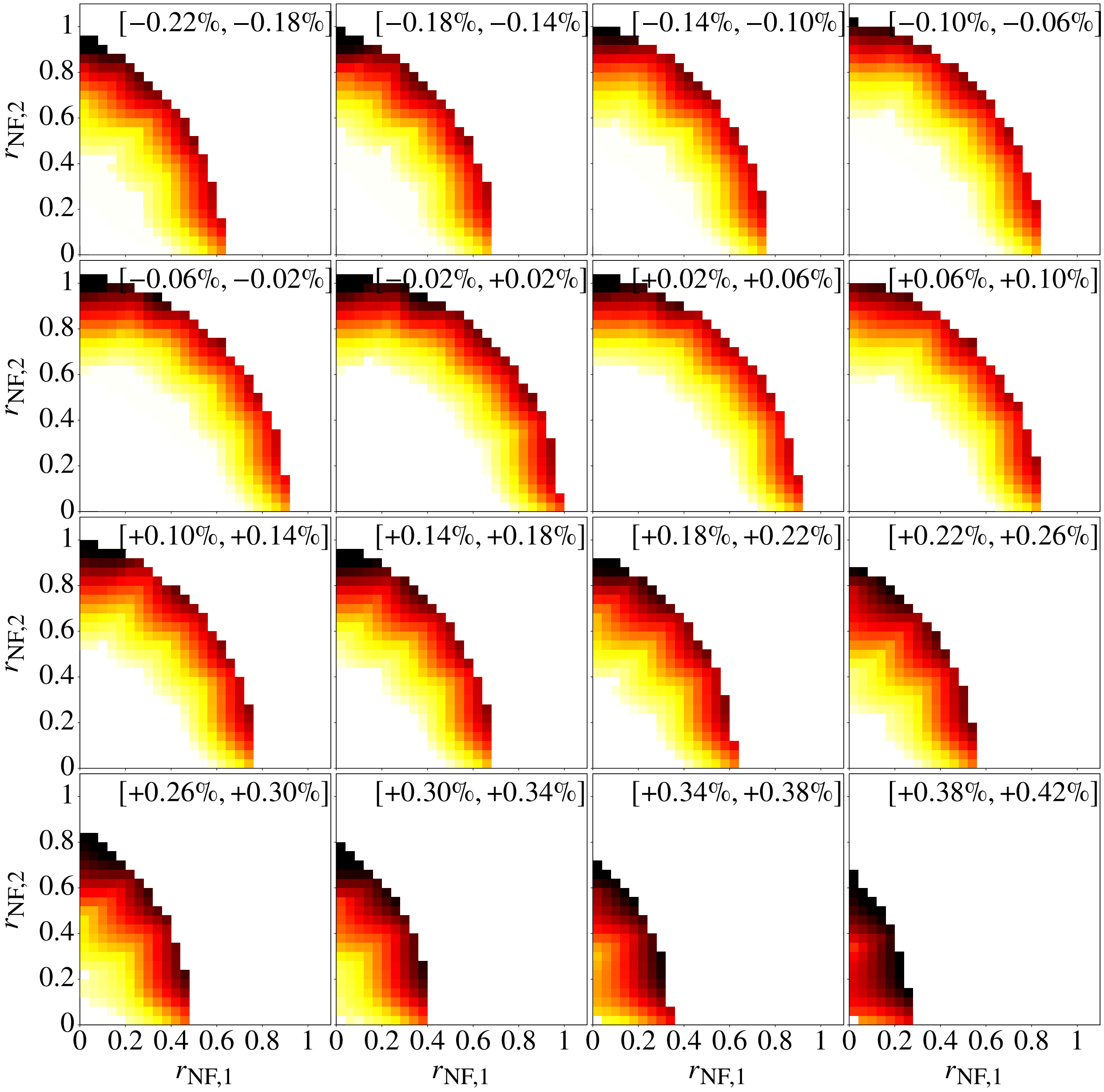

Figure 6.31: **Verified** normal form defect for the phase space storage regions of the Muon g-2 Storage ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

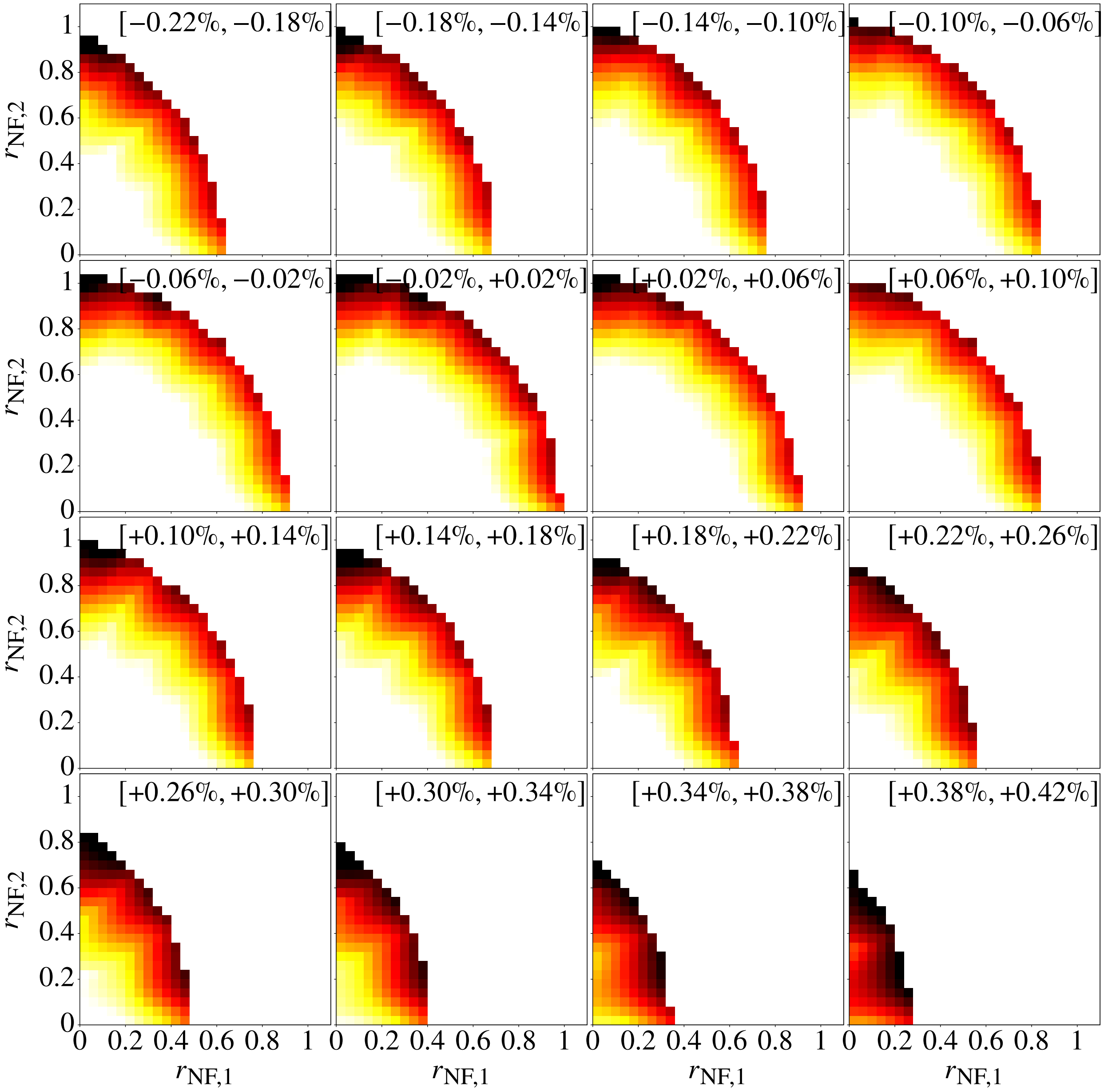

Figure 6.32: **Nonverified** normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

6.3.7 Comparison of Verified Nekhoroshev-type Stability Estimates to Actual Rates of Divergence in Example Island Structure

The minimum turn numbers of the verified Nekhoroshev-type stability estimates are a verified underestimation of the minimum number of turns it takes particles to cross a respective onion layer. The estimation assumes that the maximum normal form defect of the onion layer is encountered in every turn.

To put this underestimation in perspective we take a look at the island patterns from Fig. 5.30 in the storage ring configuration with an ESQ voltage of 18.3 kV. The particles tracked in Fig. 5.30 all have a momentum offset of $\delta p = 0.126\%$ and only differ in their vertical amplitude – the radial amplitude is constant with $r_{\text{NF},1} \approx 0$. We number the five islands from smallest (1) to largest (5).

In Tab. 6.16, the minimal and maximal normal form radii of each island are listed together with the number of turns it takes the islands to get from their lowest normal form amplitude to the largest. The number of turns is directly related to the period at which the vertical amplitude of the particles is modulated due to the island structure. The normal form radius range divided by the number of turns yields the average normal form defect of the particle.

Table 6.16: Analysis of the normal form radius range of the five island particles in Fig. 5.30 and the number of turns it requires to get from the lower end of the range to the upper end. The islands are numbered from smallest (1) to largest (5).

island	$r_{\text{NF},2,\text{min}}$	$r_{\text{NF},2,\text{max}}$	turns	avg. d_{NF}
(1)	0.925	0.932	244	2.8E-5
(2)	0.872	0.983	254	4.4E-4
(3)	0.820	1.028	293	7.1E-4
(4)	0.793	1.051	347	7.4E-4
(5)	0.773	1.066	666	4.4E-4

For the small islands close to the period-3 fixed points, the number of turns required to get from the lowest to the highest normal form amplitude increases only slightly with the size of the island. As a consequence, the average normal form defect increases. For the very large islands, the relation is quite the opposite. The range gets only slightly larger, but the period of modulation increases rapidly such that the average normal form defect even decreases again.

Tab. 6.17 lists the number of turns it takes the various island particles to cross the onion layers $[0.84, 0.88]$, $[0.88, 0.92]$, $[0.92, 0.96]$, and $[0.96, 1]$ in $r_{\text{NF},2}$ together with the predicted minimum number of turns required to cross the onion layer provided by the verified analysis.

Table 6.17: Number of turns required by the islands from Fig. 5.30 to cross the given onion layers in $r_{\text{NF},2}$ direction. The islands are numbered from smallest (1) to largest (5). Additionally, the minimum number of turns required to cross the onion layer determined by the verified analysis is shown.

	$[0.84, 0.88]$	$[0.88, 0.92]$	$[0.92, 0.96]$	$[0.96, 1.00]$
island (1)	-	-	-	-
island (2)	-	81	58	-
island (3)	65	40	32	30
island (4)	59	36	27	23
island (5)	104	36	25	21
Verified Analysis	>15.7	>8.3	>4.5	>2.5

The dynamics in a single onion layer like $[0, 0.04] \times [0.92, 0.96] \times [0.10\%, 0.14\%]$ can vary significantly. Some orbits remain in an onion layer indefinitely, like the smallest island (1). In contrast, others are transported through it with sometimes less than a factor ten between the worst case divergence predicted by the verified normal form defect analysis and the actual rate of divergence. The analysis of the largest island (5) is particularly interesting, because the average rate of divergence varies quite significantly over the normal form radius range.

In short, it is possible to relate the quantitative aspects of the normal form defect analysis to the actual dynamics within the onion layer, and in particular, the potentially worst case dynamics.

6.3.8 Relevance of the ESQ Voltage on the Stability

The global normal form defect analysis is also very powerful for the qualitative stability analysis of different storage ring configurations. A comparison of Fig. 6.30 and Fig. 6.31 yields obvious differences between the verified normal form defect of the map with an ESQ voltage of 17.5 kV and the map with an ESQ voltage of 18.3 kV. There are clearly more diverging regions with a larger maximum normal form defect for 18.3 kV in Fig. 6.31 than there are for the 17.5 kV map in Fig. 6.30.

Note that the individual onion layers of the 18.3 kV map in Fig. 6.31 and the 17.5 kV map in Fig. 6.30 do not necessarily correspond to the same phase space regions in the (x, a, y, b) phase space. The normal form transformation of each map is slightly different such that the representation of the relevant (x, a, y, b) phase space in normal form space can be different for the two maps. However, each of the 16 plots show the exact same viable (x, a, y, b) phase space in the normal form coordinates just with a slightly different scaling in $r_{\mathrm{NF},1}$ and $r_{\mathrm{NF},2}$. Accordingly, comparing the color distributions for each of the 16 plots between the two maps is a valid measure to compare the stability of the two storage ring configurations.

As previously mentioned, this more diverging behavior of the map with an ESQ voltage of 18.3 kV compared to the map with an ESQ voltage of 17.5 kV seen in Fig. 6.31 compared to Fig. 6.30, respectively, is very likely linked to the closeness of low-order resonances and their associated fixed point structures and the resulting amplitude modulations. Specifically, we saw in Chapter 5 that the vertical $1/3$-resonance tune and its associated period-3 fixed point structures for the simulation using an ESQ voltage of 18.3 kV were a major loss and instability factor.

To illustrate the difference in the closeness to low-order resonances, Fig. 6.33 to Fig. 6.35 show the tune shifts of the 17.5 kV map. The tune shifts are of similar magnitude and complexity as the tune shifts of the 18.3 kV map previously shown in Chapter 5 in Fig. 5.7 to Fig. 5.9. However, the absolute values of the tunes for 17.5 kV are in lower vertical tune ranges and therefore further away from the vertical low-order $1/3$ resonance tune.

Even under the combined influence of both the radial and vertical amplitude, as well as the momentum offset, none of the tunes of the 17.5 kV map cross the vertical $1/3$ resonance tune. In contrast, almost for every momentum offset there is a combination of radial and vertical amplitudes that crosses the vertical $1/3$ resonance tune for the 18.3 kV map. This suggests that there are no period-3 fixed point structures within the storage region, which would explain the less diverging onion layer picture in Fig. 6.30 compared to Fig. 6.31.

In summary, both the tune analysis as well as the normal form defect analysis could show that the map with an ESQ voltage of 18.3 kV yields more potential diverging behavior and instability.

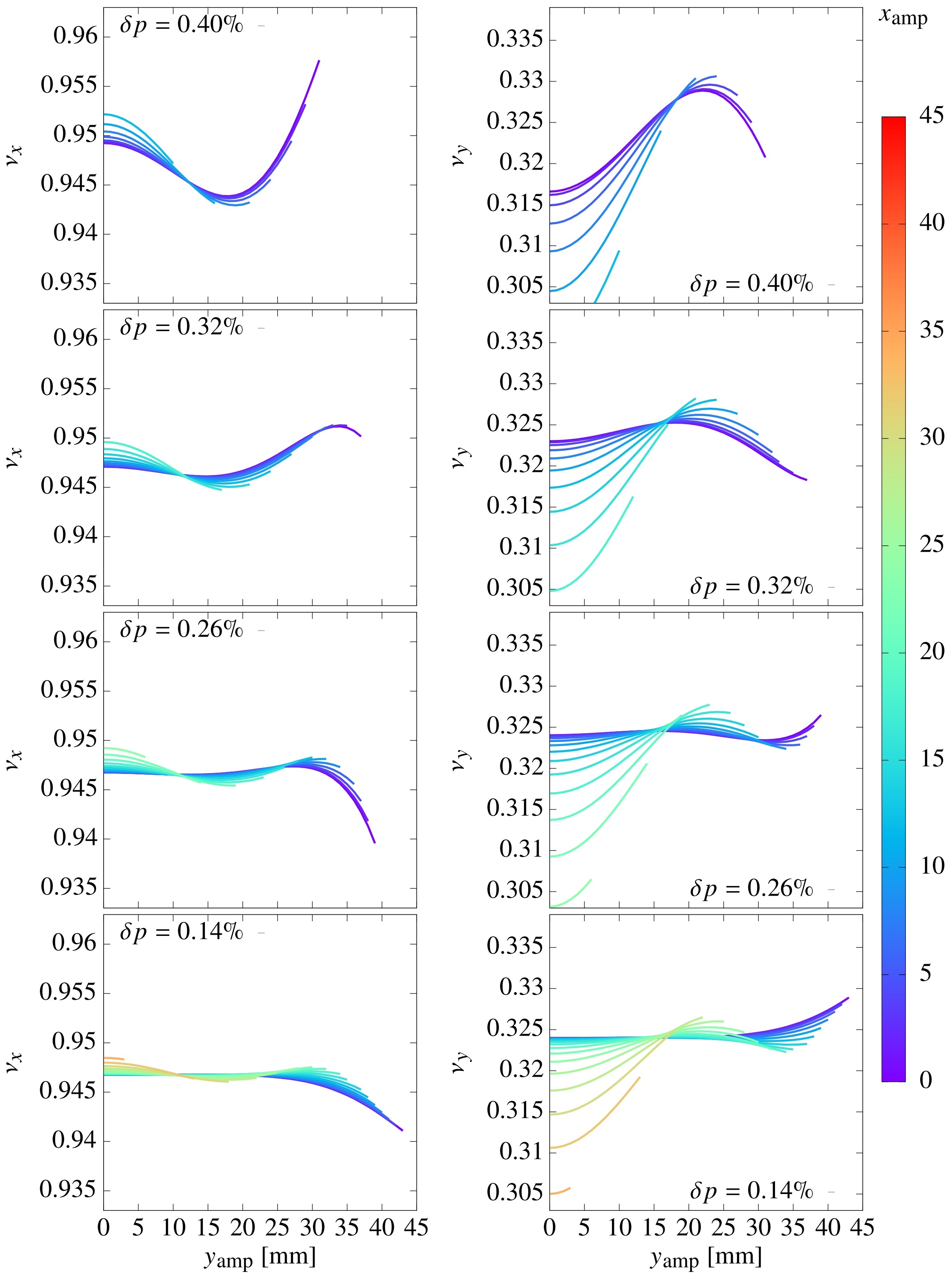

Figure 6.33: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV.

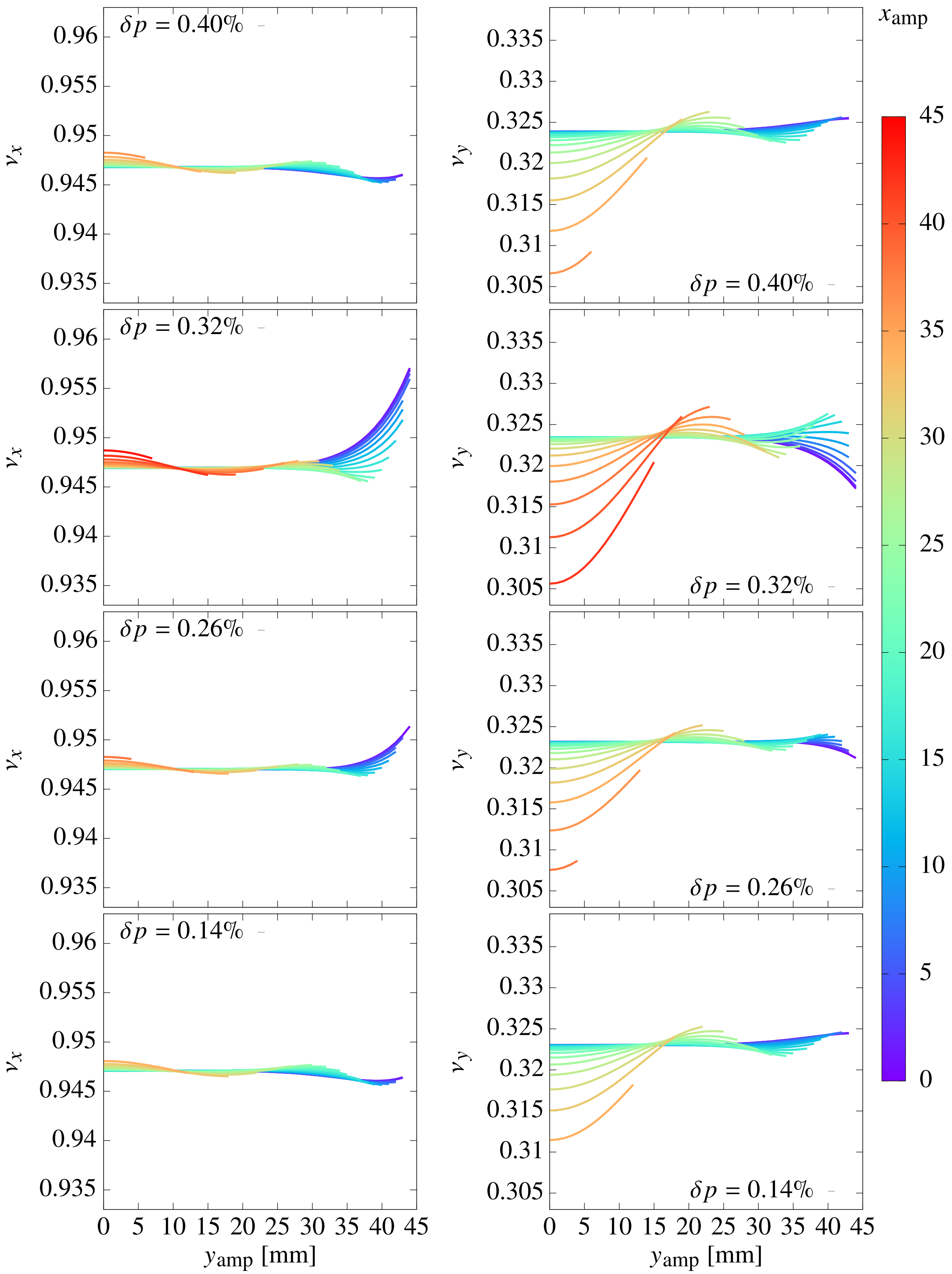

Figure 6.34: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV.

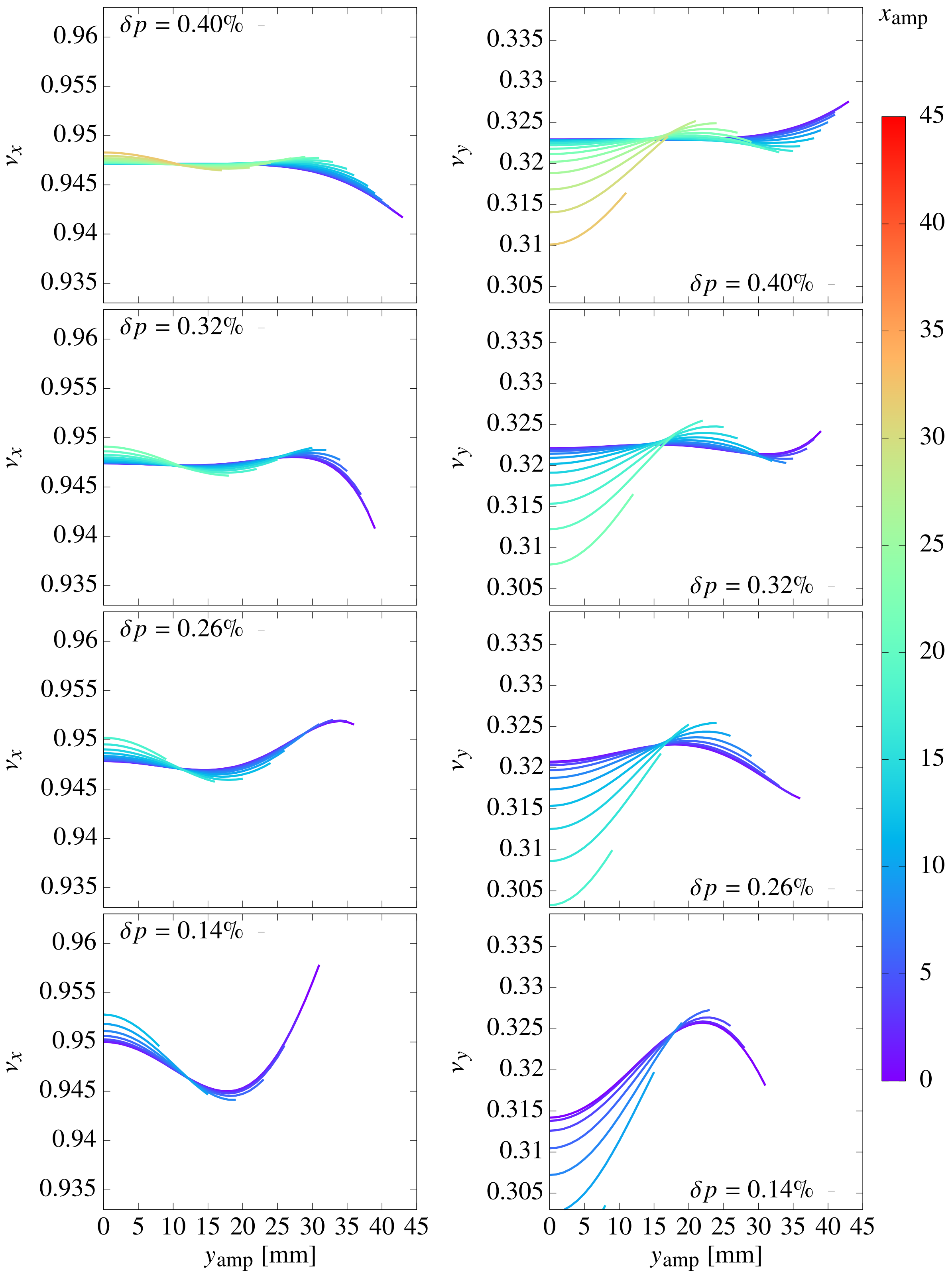

Figure 6.35: Behavior of combined amplitude dependent tune shifts at multiple momentum offsets for an ESQ voltage of 17.5 kV.

6.3.9 Comparison of Nonverified and Verified Normal Form Defect Analysis

The differences between the nonverified and verified computations in Fig. 6.29 and Fig. 6.30 for an ESQ voltage of 17.5 kV, and Fig. 6.31 and Fig. 6.32 for an ESQ voltage of 18.3 kV are small but visible if one switches back and forth between the pages. To emphasize the differences between the verifed and nonverified computations onion layer by onion layer, Fig. 6.36 and Fig. 6.37 illustrate those differences for 17.5 kV and 18.3 kV, respectively. The differences show the importance of a verified method to capture each onion layer's maximum normal form defect, especially for the more diverging regions.

To show that this difference is not an artifact of the bounding range of the global optimizer, Fig. 6.38 and Fig. 6.39 illustrate the difference between the optimized upper and the lower bound on the maximum normal form defect for the 17.5 kV map and 18.3 kV map, respectively. Because both figures only consist of white boxes, those differences are all smaller than 1E-5 and therefore do not alter the calculation of the differences between the nonverified and verified evaluation in Fig. 6.36 and Fig. 6.37

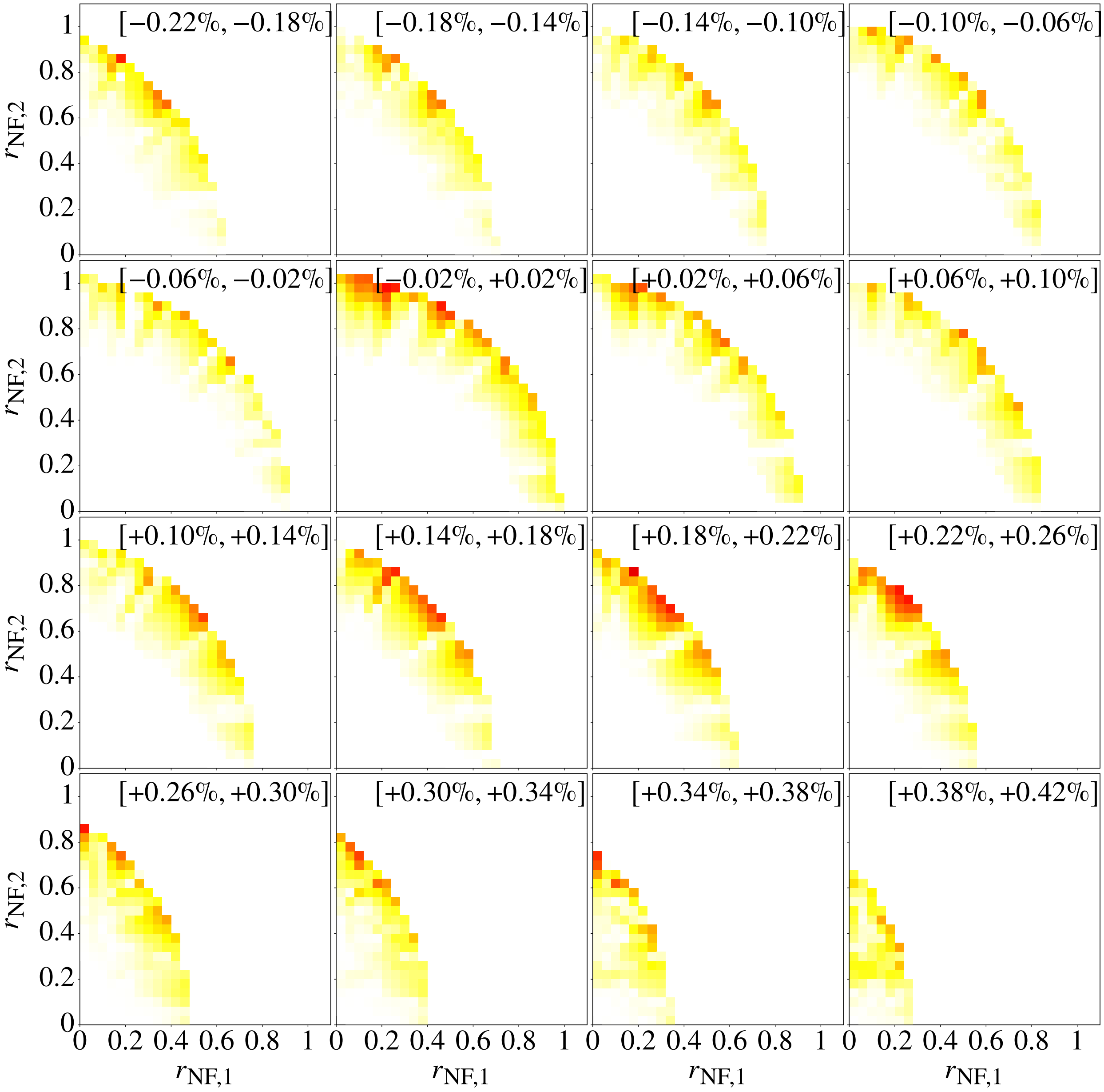

Figure 6.36: Difference between verified normal form defect analysis and nonverified normal form defect analysis for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference of the evaluated normal form defects of the specific onion layer. The white boxes for lower normal form radii indicate differences below 10^{-5}. The yellow boxes denote differences up to 10^{-4}, the orange boxes correspond to differences up to 10^{-3}, the red boxes denote differences up to $10^{-2.5}$ and the black boxes indicate differences larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

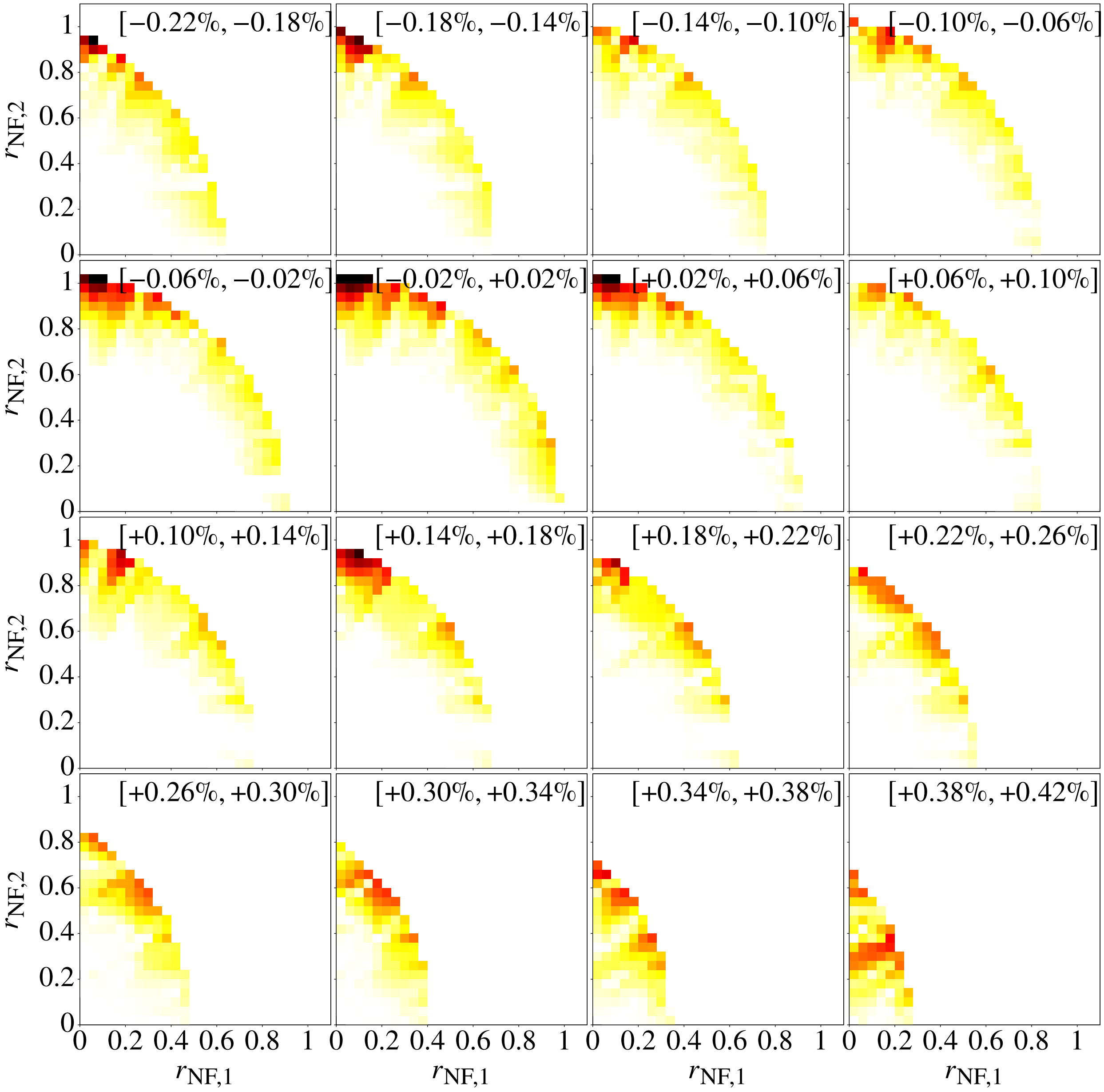

Figure 6.37: Difference between verified normal form defect analysis and nonverified normal form defect analysis for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the difference of the evaluated normal form defects of the specific onion layer. The white boxes for lower normal form radii indicate a difference below 10^{-5}. The yellow boxes denote differences up to 10^{-4}. The orange boxes correspond to differences up to 10^{-3}. The red boxes denote differences up to $10^{-2.5}$, and the black boxes indicate differences larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

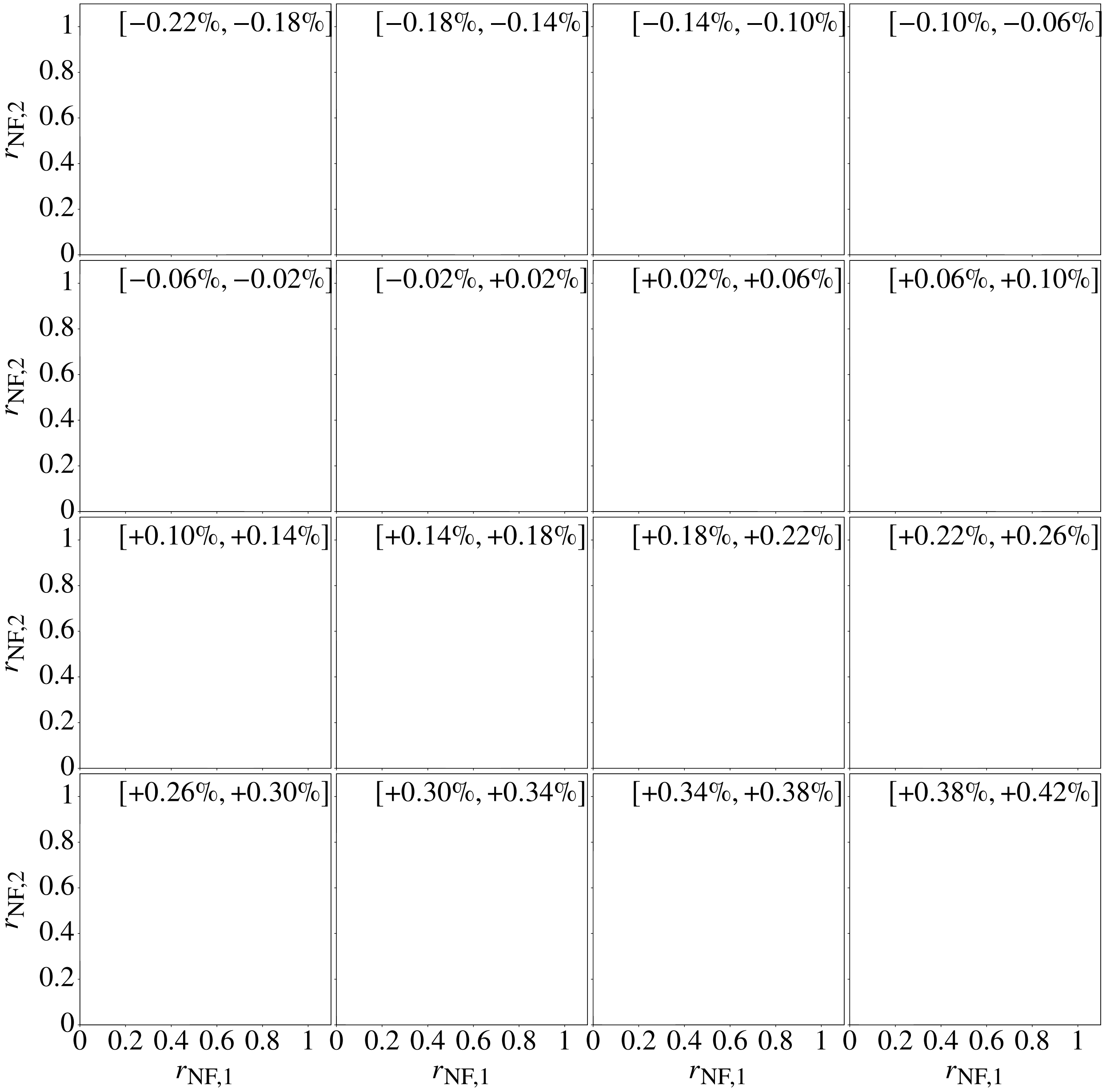

Figure 6.38: Difference between the rigorously guaranteed upper bound and the lower bound of the maximum normal form defect using Taylor Model based verified global optimization. The analysis is for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **17.5 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference between the upper bound and the lower bound of the maximum normal form defect of the specific onion layer. All boxes are white because the difference is below 10^{-5}. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

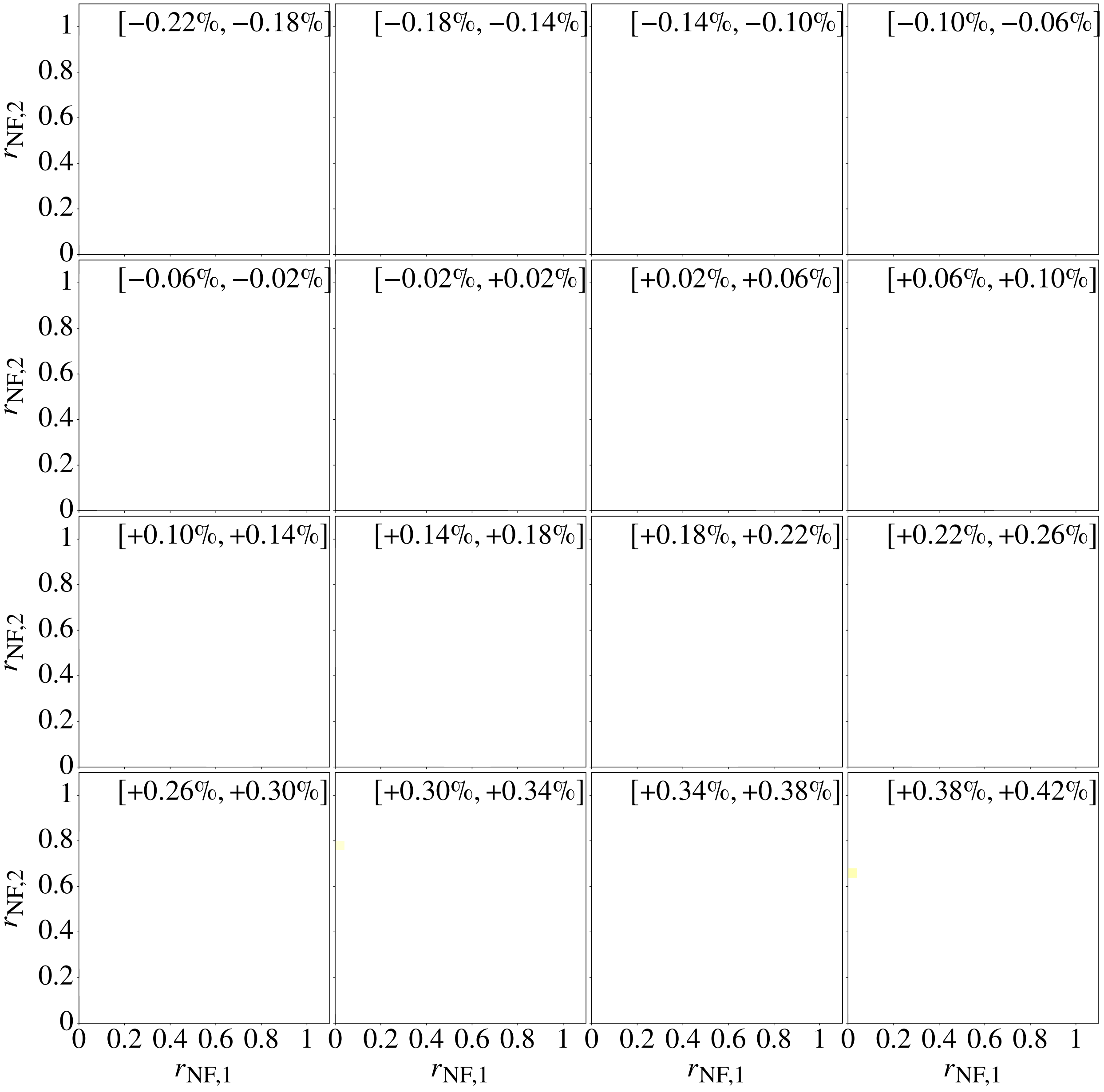

Figure 6.39: Difference between the rigorously guaranteed upper bound and the lower bound of the maximum normal form defect using Taylor Model based verified global optimization. The analysis is for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of **18.3 kV**. The individual plots show different momentum ranges which are clarified by the label at the top of each graph. The color scheme corresponds to the difference between the upper bound and the lower bound of the maximum normal form defect of the specific onion layer. All boxes are white because the difference is below 10^{-5}. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

6.3.10 The Analysis of the Effect of Normal Form Transformations of Different Order on the Normal Form Defect

We use the normal form transformation as a function that provides pseudo-invariants of the motion, i.e., the normal form radii. By using the normal form transformation up to different orders, we can analyze the influence of the respective map orders on the dynamics of the system. In Fig. 6.40 to Fig. 6.49, the nonverified normal form defect analysis is performed for the tenth order map with a ESQ voltage of 18.3 kV using normal form transformations from order one to order ten.

The normal form defect pictures for a normal form transformation of order five, six, and seven look identical even when carefully switching between pages. The largest improvement occurs with the ninth order normal form transformation because it captures large parts of the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential.

To further analyze if the tenth order map does indeed capture most of the relevant dynamics, we produce an eleventh order map and calculate its normal form defect using the tenth order normal form transformation (see Fig. 6.50). This kind of order increasing analysis is known from nonverified integrators with step size control. Compared to the tenth order map evaluation with the tenth order normal form transformation in Fig. 6.49, the eleventh order of the map leads to no visible difference, which is a good sign and suggests that a tenth order map is sufficient to capture the critical dynamics. However, this heuristic approach cannot guarantee that even higher order maps would also not yield a significant change. To capture this uncertainty of unknown higher order terms a verified map is required that includes all higher order errors in its Taylor Model remainder bound.

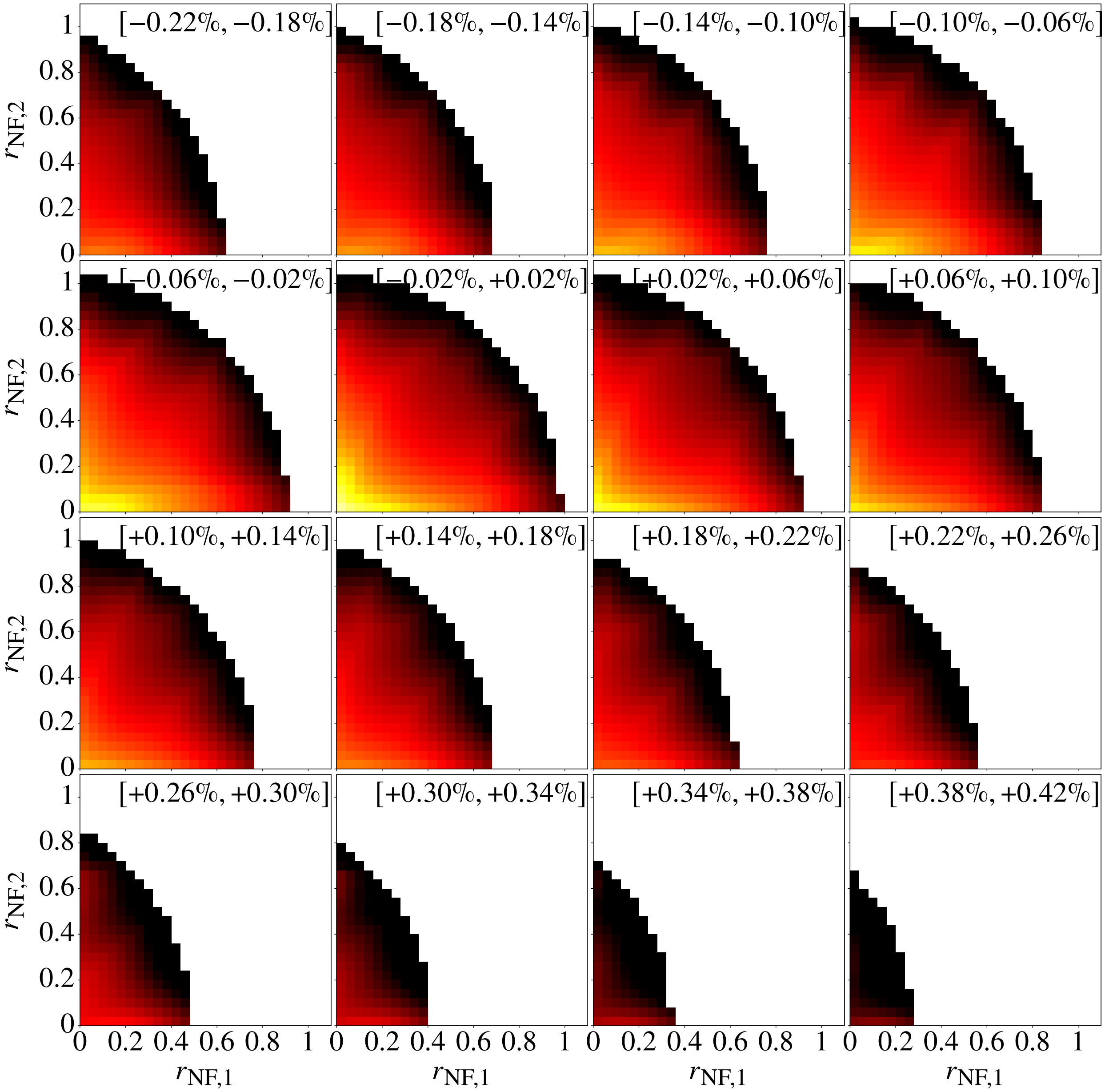

Figure 6.40: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 1** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

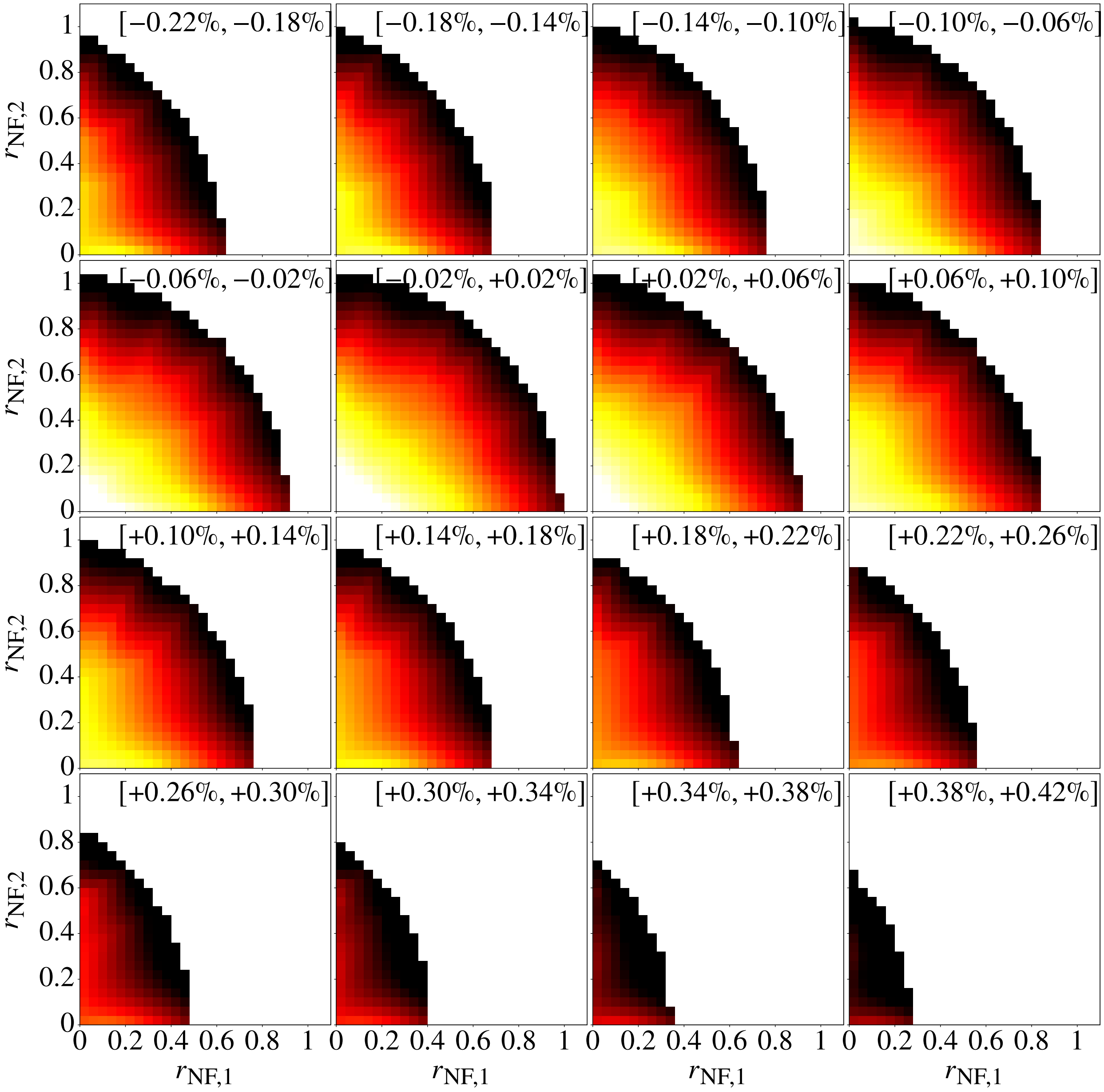

Figure 6.41: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 2** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

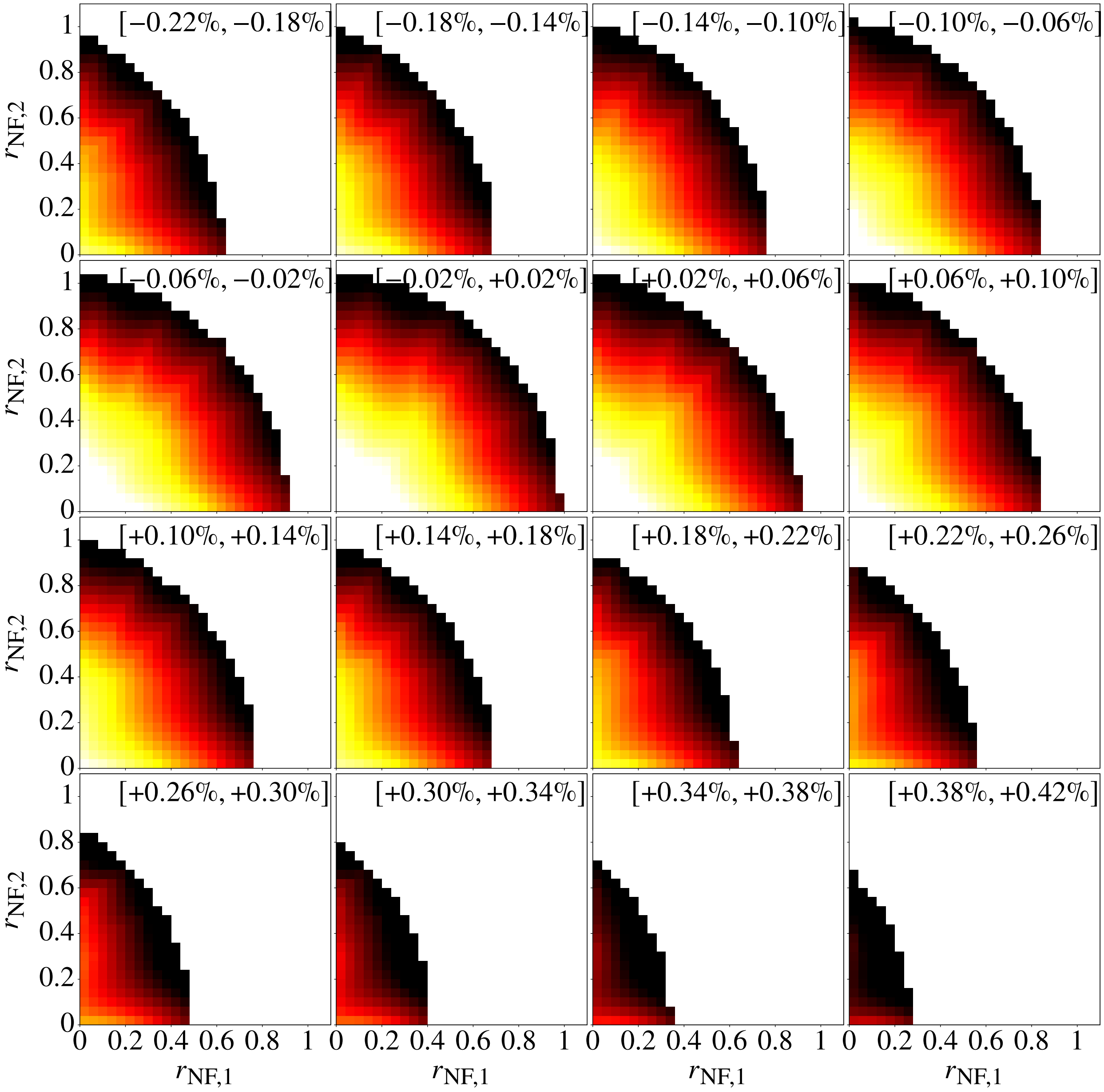

Figure 6.42: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 3** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

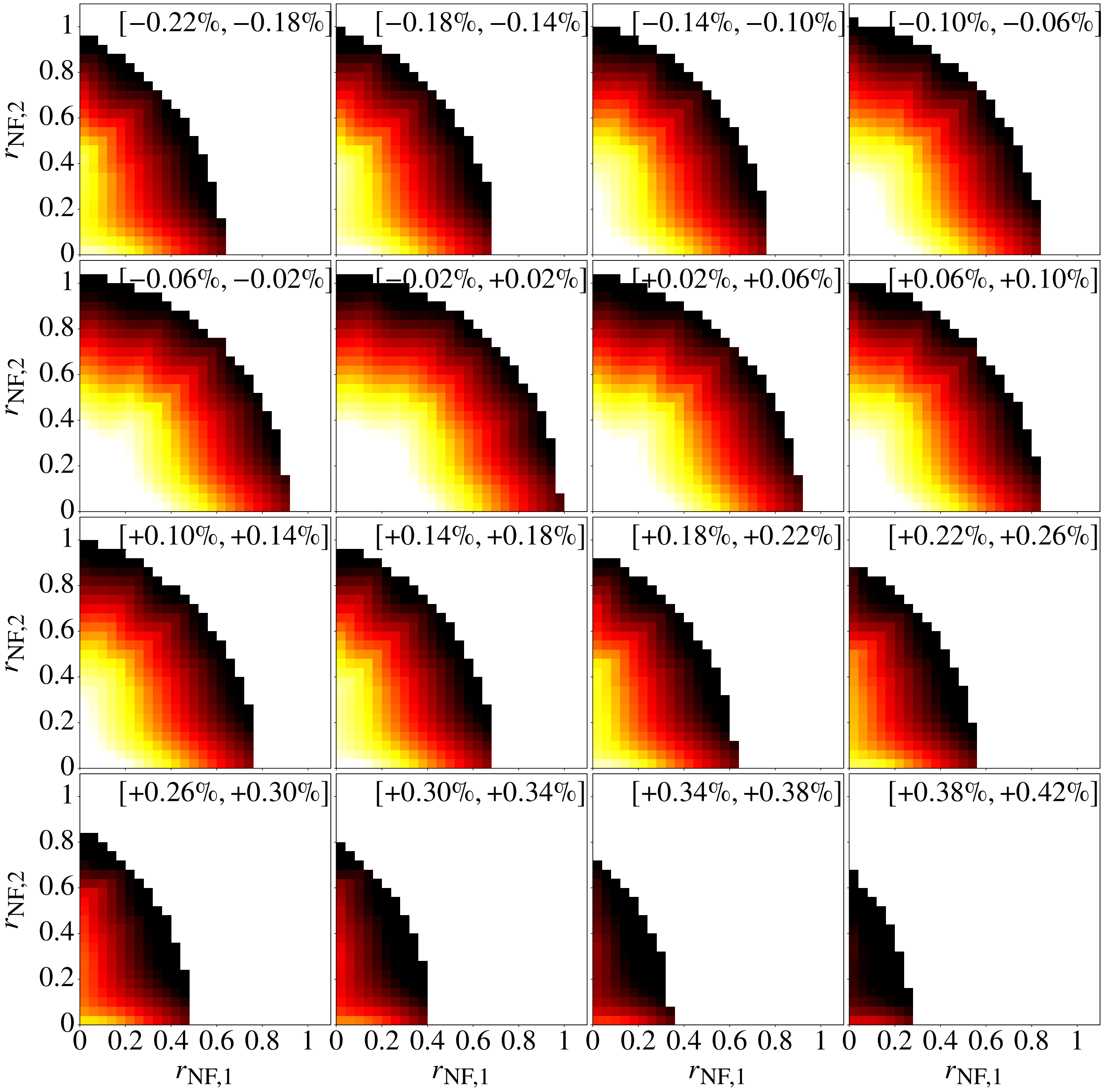

Figure 6.43: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 4** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

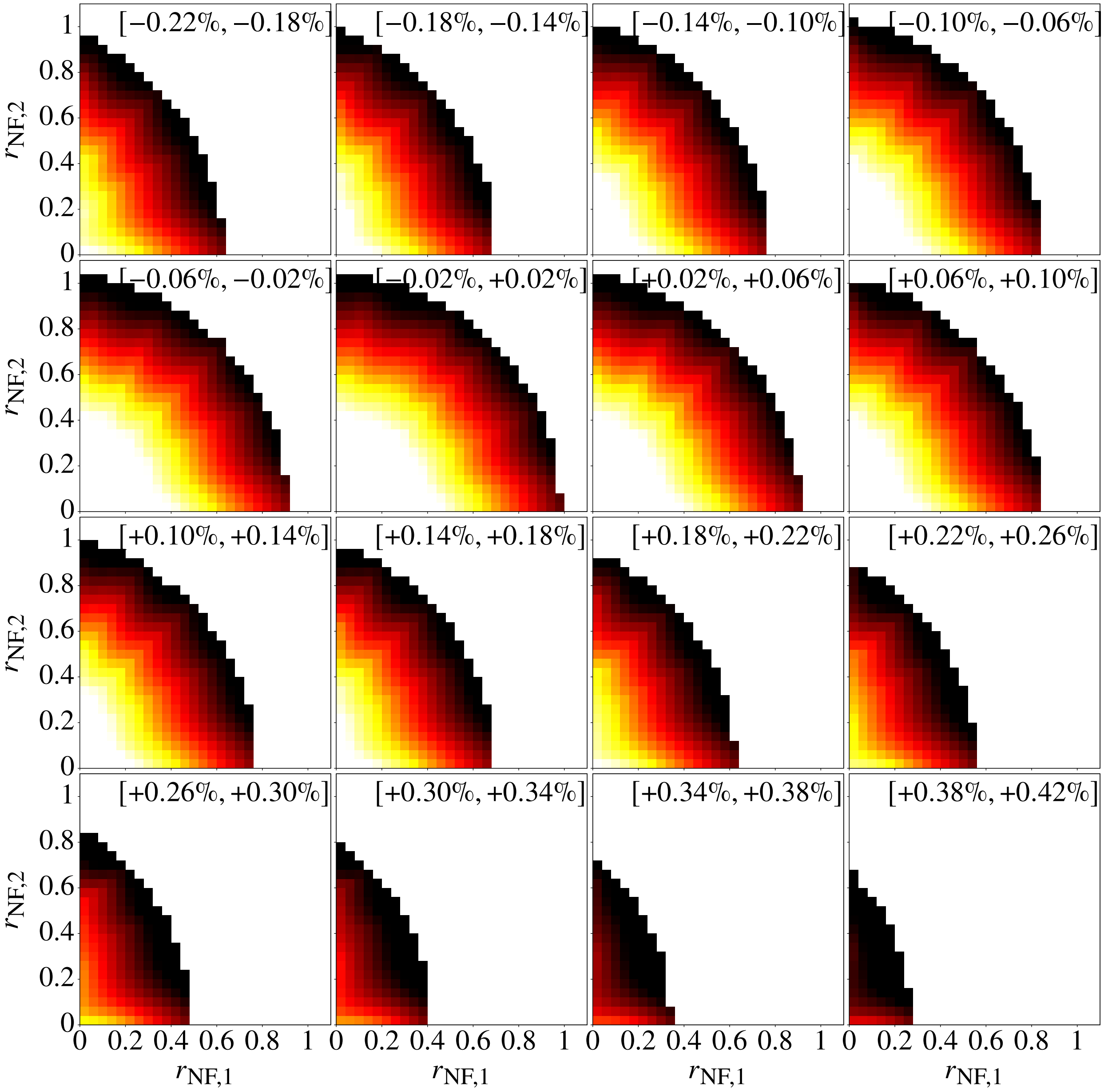

Figure 6.44: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 5** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

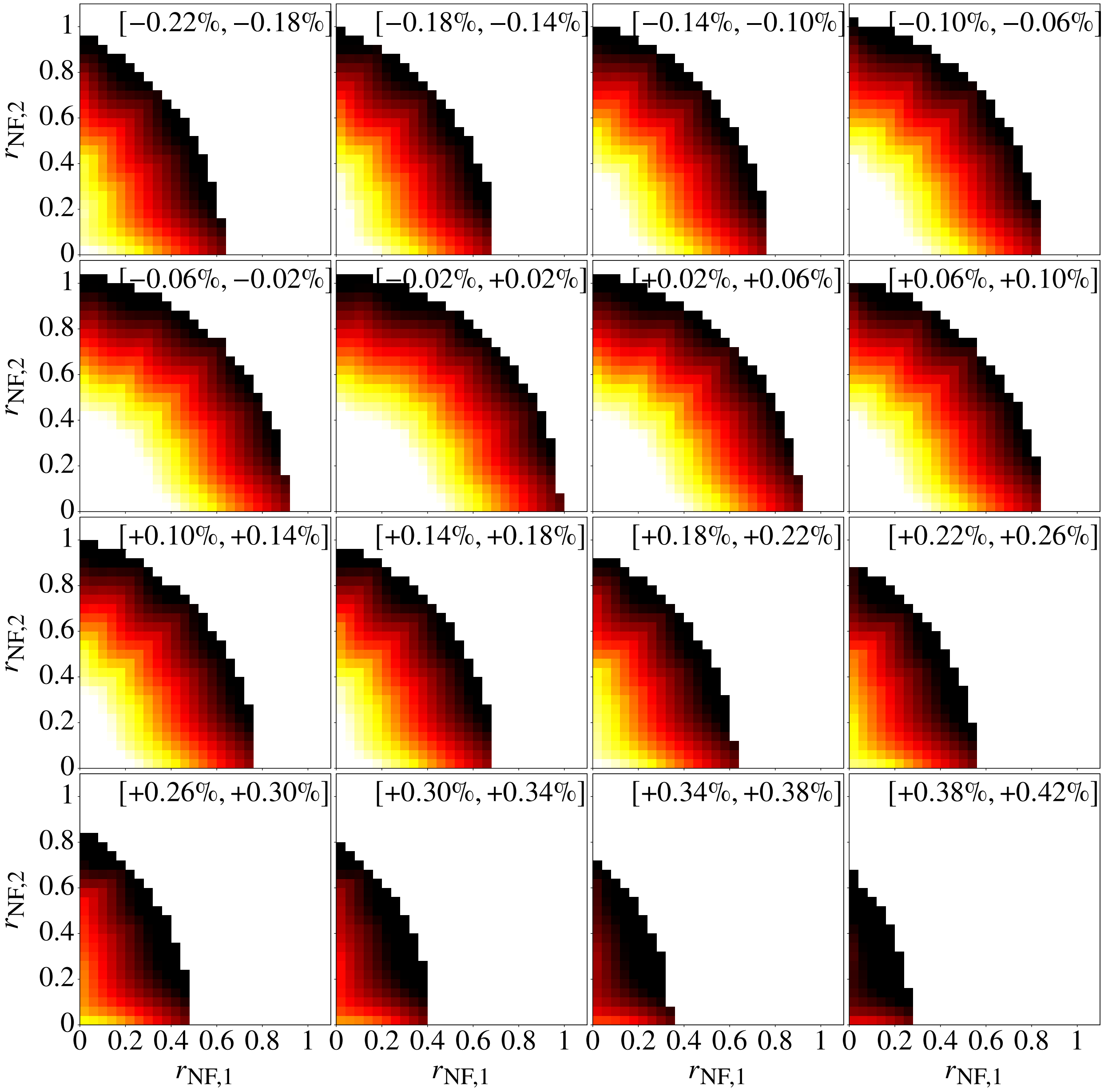

Figure 6.45: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 6** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

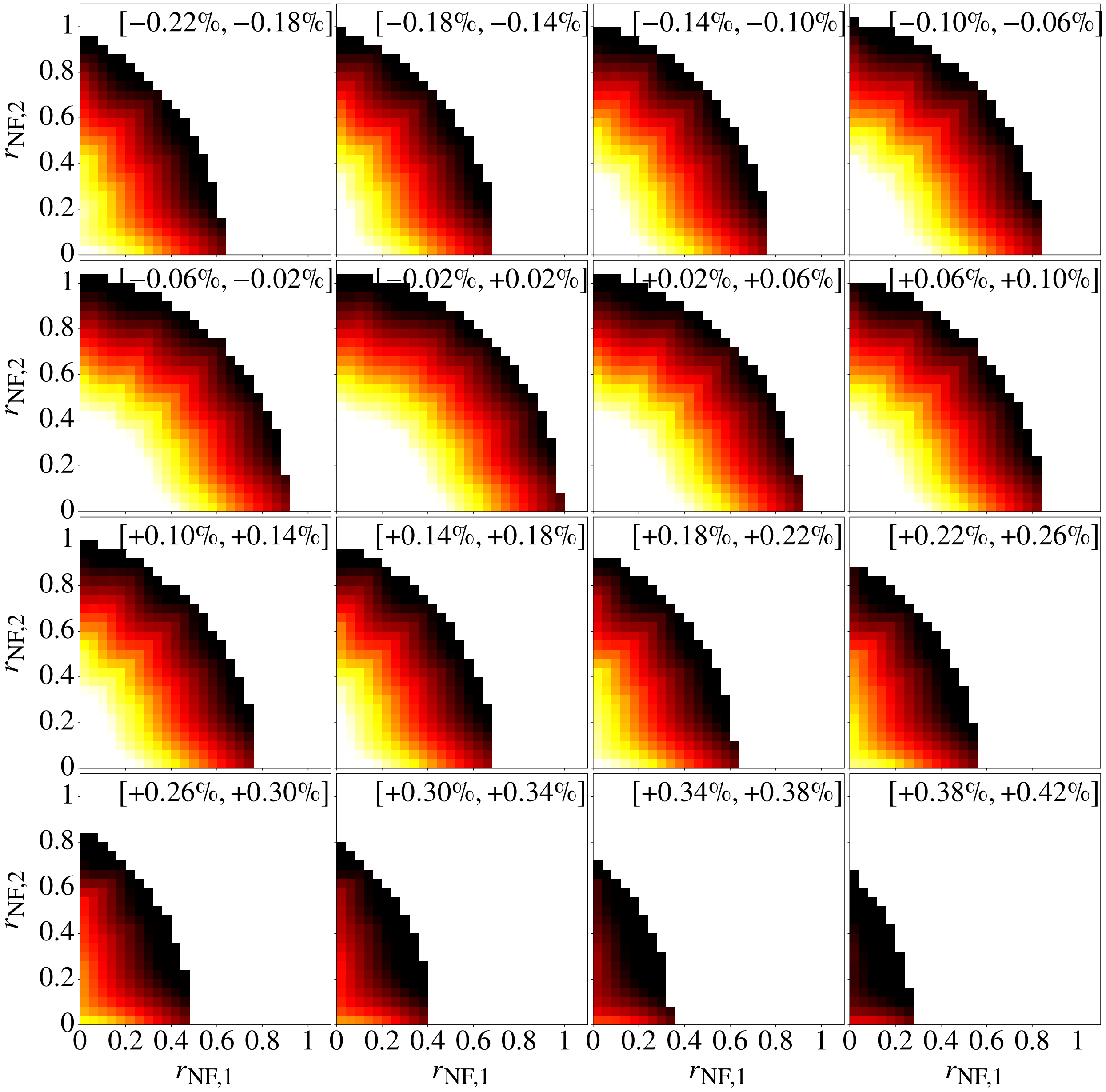

Figure 6.46: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 7** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

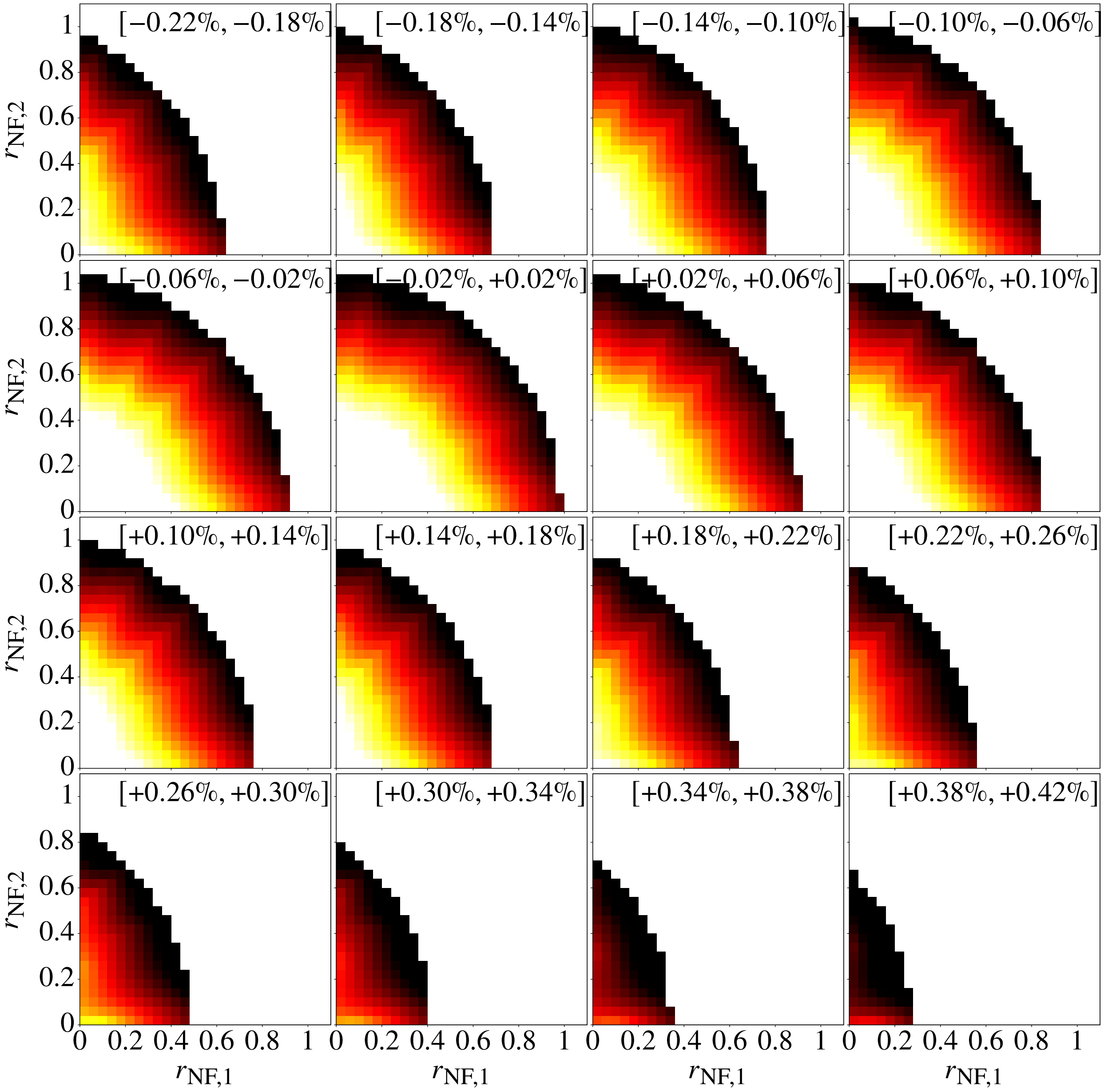

Figure 6.47: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 8** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

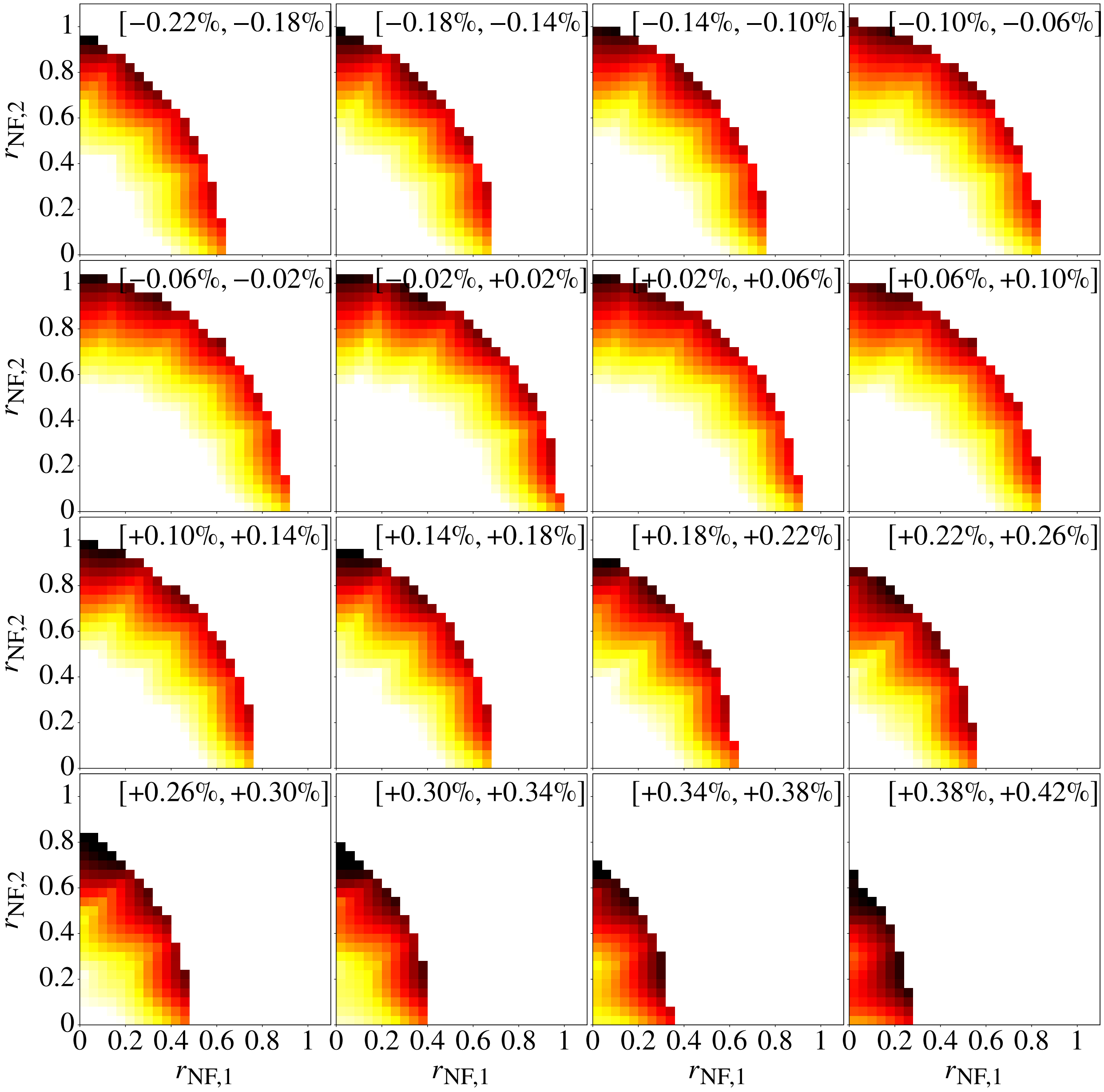

Figure 6.48: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to **order 9** instead of the full tenth order. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

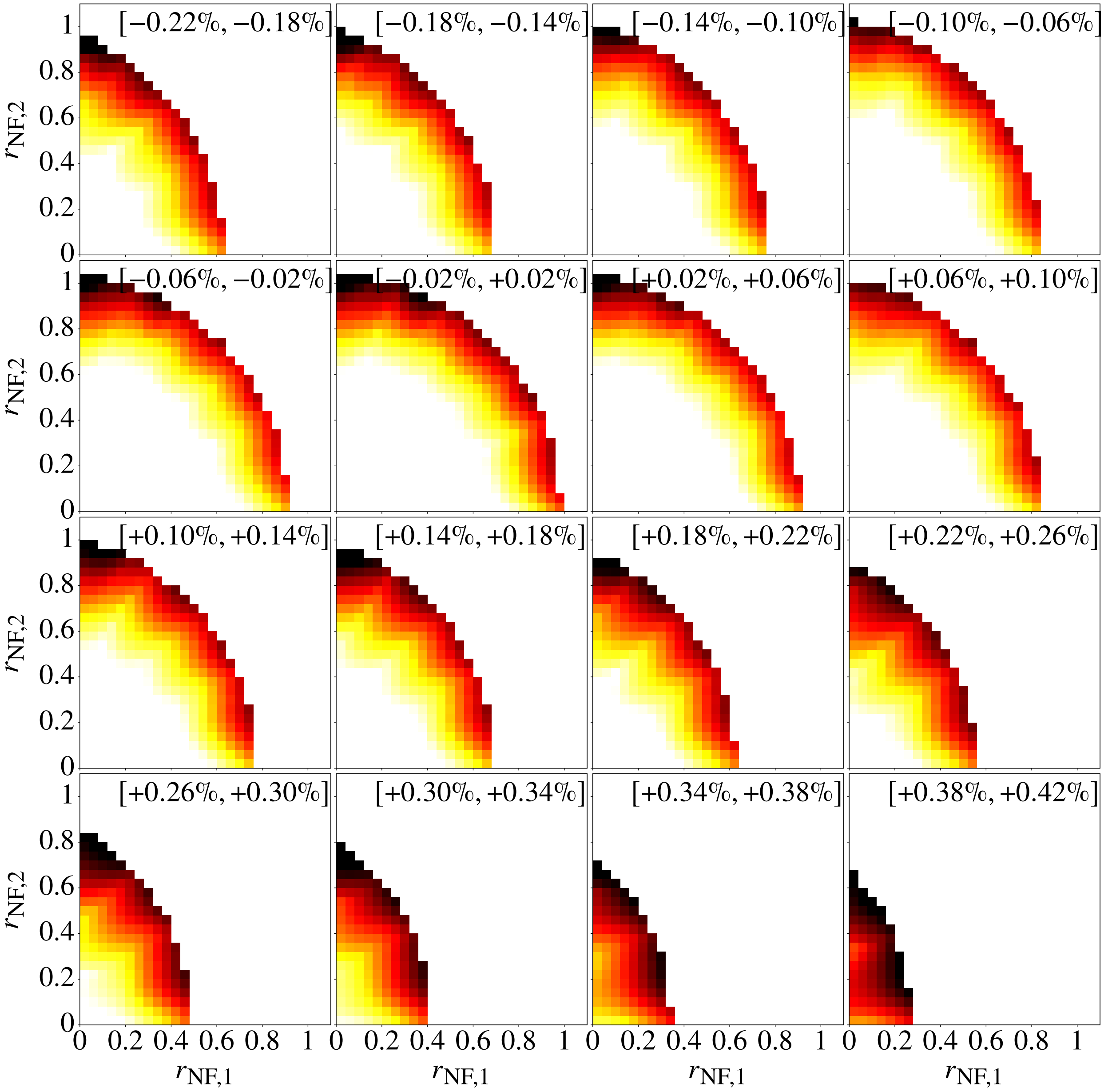

Figure 6.49: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using the normal form transformation up to the **full tenth order**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

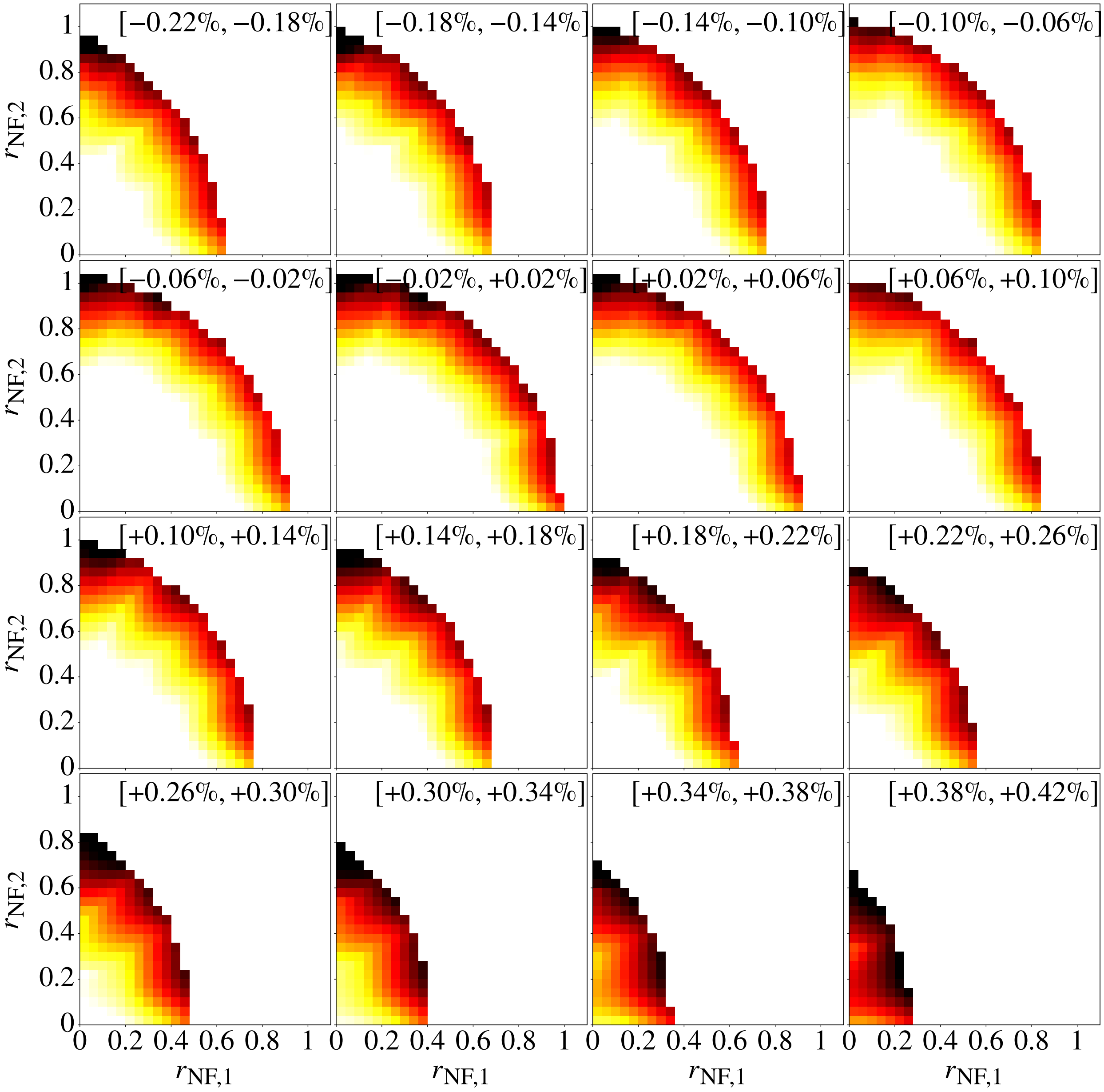

Figure 6.50: Nonverified normal form defect for the phase space storage regions of the Muon g-2 Storage Ring simulation with an ESQ voltage of 18.3 kV using an **eleventh order map** and its normal form transformation up to **tenth order**. The individual plots show different momentum ranges, clarified by the label at the top of each graph. The color scheme corresponds to the normal form defect of the specific onion layer. The white boxes for lower normal form radii indicate a normal form defect below 10^{-5}. The yellow boxes denote normal form defects up to 10^{-4}. The orange boxes correspond to normal form defects up to 10^{-3}. The red boxes denote normal form defects up to $10^{-2.5}$, and the black boxes indicate normal form defects larger than that. Each onion layer corresponds to a 0.04×0.04 box in normal form space with a thickness of 0.04% in δp.

CHAPTER 7

CONCLUSION

We investigated a diverse set of nonlinear systems using normal forms and rigorous differential algebra methods. The differential algebra framework implemented in COSY INFINITY served as the backbone of all the methods and techniques in this thesis. It allowed us to establish algorithms and solutions up to arbitrary order and with floating point accuracy.

The basis of our analysis constituted map representations of the various systems based on the underlying equations of motion. These stroboscopic descriptions of the dynamics were expanded around a fixed point corresponding to an equilibrium state of the motion. Using Poincaré projections, the dimensionality of the system was reduced to the essential components of the system's dynamics.

For the bounded motion problem in the zonal gravitational field of the Earth in Chapter 4, the motion was considered within a four dimensional Poincaré surface capturing all ascending node states. In Chapter 5, the dynamics within the Muon g-2 Storage Ring were analyzed in transverse cross sections of the storage ring at multiple azimuthal locations.

The origin preserving maps were then analyzed using high order normal forms to calculate a description of the phase space dynamics that is rotationally invariant up to calculation order. In Chapter 3, the normal form algorithm was discussed in full detail using the illustrative example of the centrifugal governor. In this particular case, the normal form radii, which constitute the (pseudo-)invariants of the motion up to calculation order produced by the normal form algorithm, were directly related to the energy of the system up to calculation order. Additionally, the normal form produced high order functional descriptions of the period of oscillation of the centrifugal governor arms around their equilibrium angle depending on the amplitude of oscillation and changes in the rotation frequency of the governor.

For the bounded motion problem, this rotational invariant representation of the phase space motion provided by the normal form was used to transform the system into action-angle like coordinates. This allowed us to average the bounded motion quantities while maintaining their functional dependence on the constants of motion. DA inversion methods were then used to enforce the bounded motion conditions and produce parameterized descriptions of the constants of motion, which yielded entire continuous sets of bounded motion orbits. We illustrated that the resulting sets of orbits remained bounded for decades and far beyond the practically relevant distances of formation flying missions.

Our approach can possibly be advanced to the fully gravitationally perturbed case. However, the associated break of the rotational symmetry makes this already complex system even more complex. The introduced longitudinal dependence and the loss of the angular momentum component as a constant of motion increase the dimensionality of the problem by two. Accordingly, pseudo-circular orbits of the full state are required to expand the fixed point map around. Only further research can answer if and how the approach can be adjusted to compensate for the loss of a known constant of motion and the increase in dimensionality.

In our analysis of the dynamics in the Muon g-2 Storage Ring in Chapter 5, we studied the oscillation frequencies of particles in the radial and vertical transverse direction also known as the betatron tunes. The normal form transformation allowed us to calculate the functional dependence of the tunes on the momentum offset of the particles and their amplitude of oscillation. A major insight of this investigation was that particles over the entire momentum offset range could cross the vertical $1/3$-resonance frequency for certain vertical and radial amplitude combinations.

This closeness to the low order resonance triggered intensive lost muon tracking studies, which revealed period-3 fixed point structures in the vertical phase space. Particles caught around those period-3 fixed points experienced significant vertical amplitude modulations, which drastically increased their risk of hitting a collimator and getting lost in the process.

Throughout the analysis, the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential were prominent. They could be found as eighth order dependencies in the amplitude and momentum dependent tune shifts and be visualized by the drastic change in the tune footprint when comparing eighth order to tenth order results.

To further assess the stability of the Muon g-2 Storage Ring rigorously, we utilized Taylor

Model based verified global optimization in Chapter 6. The abilities of Taylor Model based global optimization was presented using the objective functions of different example problems. The generalized Rosenbrock function served as an example to illustrate different effects that can sometimes influence the optimization including the dependency problem and the cluster effect. We illustrated that Taylor Models and their associated advanced bounding techniques could drastically suppress those effects compared to other commonly used approaches.

The Lennard-Jones problem was used to illustrate the many intricacies that have to be solved for rigorous global optimization of some complex systems. While the Lennard-Jones problem is easily formulated, its formal description with optimization variables and bounding to a rigorous initial search domain are far from trivial. Our discussion of the problem also illustrated the struggle associated with not being able to exclude manifolds from the search domain for which the objective function is not defined.

For the rigorous stability analysis of the Muon g-2 Storage Ring, we calculated verified upper bounds on the rate at which particles can escape the storage region. To get a detailed understanding of the stability properties of the storage ring, we partitioned the five dimensional storage region into more than 8000 sections using the onion layer approach. We used Taylor Model based verified global optimization to calculate the maximum rate of divergence in the form of the normal form defect for each one of those partitions. The verified normal form defect results from the map with the closeness to the vertical $1/3$-resonance from Chapter 5 were compared to the results of a map with a different ESQ voltage, which yielded tunes further away from this vertical low order resonance. The comparison illustrated significant differences in the stability of phase space regions close to the collimators, confirming that the low order resonance noticeably impairs the system's long-term stability. The normal form defect analysis was also able to identify the strong ninth order nonlinearities of the map caused by the 20th-pole of the ESQ potential.

APPENDIX

A.1 Toy Example for Verified Optimization of Four Particles in 3D

In literature, the trivial case of four particles in 3D is often discussed as a toy problem, which we run below to provide our results for comparison. In [59], the variable definitions were chosen similarly to our choice in Sec. 6.2.3.8. The only relevant difference is that [59] used the x positions as variables with

$$v'_{x,i} = x_{i+1} \tag{1}$$

instead of $v_{x,i} = x_{i+1} - x_i$ as defined in Eq. (6.116). Note that the first particle is fixed at the origin ($x_1 = 0$) also in [59].

We run the global optimization with Taylor Models of order 5, without providing an initial cutoff value and with a threshold for the smallest boxes of $s_{\min} = 10^{-6}$, just like in [59]. The initial variable search domains of [59] and initial variable search domains of our optimization are listed in Tab. A.1. Tab. A.1 also shows our results for the optimized variables. Note that we used the same size for the initial search volume as in [59], which is 2.88×10^{-4}. COSY-GO reduced the search domain to 7 remaining boxes with a total volume of 7.5×10^{-41} in 2.102 seconds and 2794 steps. The minimum was bound by

$$[-3.115103401910087\text{E-}307, 8.060219158778647\text{E-}14]. \tag{2}$$

Table A.1: The left columns list the variables of [59], denoted by $v'_{\cdot,i}$, and their respective initial search domains. The middle columns list the variables of our optimization and their respective initial search domains. The right columns show the optimized variables of our optimization.

Initial domain		Initial domain		Optimized result	
$v'_{x,1}$	$[0.4, 0.6]$	$v_{x,1}$	$[0.4, 0.6]$	$v^\star_{x,1}$	$[0.499999879, 0.500000183]$
$v'_{x,2}$	$[0.4, 0.6]$	$v_{x,2}$	$[0, 0.2]$	$v^\star_{x,2}$	$[-0.445014773\text{E-}307, 0.229985454\text{E-}6]$
$v'_{x,3}$	$[0.8, 1.2]$	$v_{x,3}$	$[0.4, 0.8]$	$v^\star_{x,3}$	$[0.499999879, 0.500000183]$
$v'_{y,2}$	$[0.7, 1.0]$	$v_{y,2}$	$[0.7, 1.0]$	$v^\star_{y,2}$	$[0.866025282, 0.866025501]$
$v'_{y,3}$	$[0.2, 0.4]$	$v_{y,3}$	$[0.2, 0.4]$	$v^\star_{y,3}$	$[0.288675006, 0.288675372]$
$v'_{z,3}$	$[0.7, 1.0]$	$v_{z,3}$	$[0.7, 1.0]$	$v^\star_{z,3}$	$[0.816496487, 0.816496660]$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Babak Abi et al. (Muon g-2 Collaboration). Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Physical Review Letters*, 126:141801, 2021.

[2] Tareq Albahri et al. (Muon g-2 Collaboration). Beam dynamics corrections to the run-1 measurement of the muon anomalous magnetic moment at Fermilab. *Physical Review Accelerators and Beams*, 24(4):044002, 2021.

[3] Tareq Albahri et al. (Muon g-2 Collaboration). Magnetic-field measurement and analysis for the Muon g-2 Experiment at Fermilab. *Physical Review A*, 103(4):042208, 2021.

[4] Tareq Albahri et al. (Muon g-2 Collaboration). Measurement of the anomalous precession frequency of the muon in the Fermilab Muon g-2 Experiment. *Physical Review D*, 103(7):072002, 2021.

[5] Kyle T. Alfriend, Srinivas R. Vadali, Pini Gurfil, Jonathan P. How, and Louis S. Breger. *Spacecraft Formation Flying: Dynamics, Control and Navigation*. Batterworth-Heinemann, 2010.

[6] Tatsumi Aoyama, Nils Asmussen, Maurice Benayoun, Johan Bijnens, Thomas C. Blum, et al. The anomalous magnetic moment of the muon in the Standard Model. *Physics reports*, 2020.

[7] Raymond Ayoub. Euler and the zeta function. *The American Mathematical Monthly*, 81(10):1067–1086, 1974.

[8] Nicola Baresi, Zubin P. Olikara, and Daniel J. Scheeres. Fully numerical methods for continuing families of quasi-periodic invariant tori in astrodynamics. *The Journal of the Astronautical Sciences*, 65(2):157–182, 2018.

[9] Nicola Baresi and Daniel J. Scheeres. Bounded relative motion under zonal harmonics perturbations. *Celestial Mechanics and Dynamical Astronomy*, 127(4):527–548, 2017.

[10] Nicola Baresi and Daniel J. Scheeres. Design of bounded relative trajectories in the Earth zonal problem. *Journal of Guidance, Control, and Dynamics*, 40(12):3075–3087, 2017.

[11] Gerald W. Bennett et al. (Muon g-2 Collaboration). Final report of the E821 muon anomalous magnetic moment measurement at BNL. *Physical Review D*, 73(7):072003, 2006.

[12] Sonja Berner. Parallel methods for verified global optimization practice and theory. *Journal of Global Optimization*, 9(1):1–22, 1996.

[13] Martin Berz. Private communication.

[14] Martin Berz. The method of power series tracking for the mathematical description of beam dynamics. *Nuclear Instruments and Methods A*, 258(3):431–436, 1987.

[15] Martin Berz. Differential algebraic description of beam dynamics to very high orders. *Part. Accel.*, 24(SSC-152):109–124, 1988.

[16] Martin Berz. High-order computation and normal form analysis of repetitive systems. In *AIP Conference Proceedings*, volume 249, pages 456–489, 1992.

[17] Martin Berz. Differential algebraic formulation of normal form theory. In *Conference series - Institute of Physics*, volume 131, pages 77–77. IOP Publishing LTD, 1993.

[18] Martin Berz. Differential algebraic description and analysis of spin dynamics. *AIP CP*, 343, 1995.

[19] Martin Berz. *Modern Map Methods in Particle Beam Physics*. Academic Press, 1999.

[20] Martin Berz and Jens Hoefkens. Verified high-order inversion of functional dependencies and superconvergent interval Newton methods. *Reliable Computing*, 7(5):379–398, 2001.

[21] Martin Berz and Georg Hoffstätter. Computation and application of Taylor polynomials with interval remainder bounds. *Reliable Computing*, 4(1):83–97, 1998.

[22] Martin Berz and Kyoko Makino. Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliable Computing*, 4(4):361–369, 1998.

[23] Martin Berz and Kyoko Makino. Constructive generation and verification of Lyapunov functions around fixed points of nonlinear dynamical systems. *International Journal of Computer Research*, 12(2):235–244, 2003.

[24] Martin Berz and Kyoko Makino. Suppression of the wrapping effect by Taylor model-based verified integrators: Long-term stabilization by shrink wrapping. *International Journal of Differential Equations and Applications*, 10(4):385–403, 2005.

[25] Martin Berz and Kyoko Makino. COSY INFINITY Version 9.2 programmer's manual. Technical Report MSUHEP-151102, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2015. See also http://cosyinfinity.org.

[26] Martin Berz and Kyoko Makino. COSY INFINITY Version 10.0 beam physics manual. Technical Report MSUHEP-151103-rev, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2017. See also http://cosyinfinity.org.

[27] Martin Berz and Kyoko Makino. COSY INFINITY Version 10.0 programmer's manual. Technical Report MSUHEP-151102-rev, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2017. See also http://cosyinfinity.org.

[28] Martin Berz, Kyoko Makino, and Jens Hoefkens. Verified integration of dynamics in the solar system. *Nonlinear Analysis*, 47:179–190, 2001.

[29] Martin Berz, Kyoko Makino, and Youn-Kyung Kim. Long-term stability of the Tevatron by validated global optimization. *Nuclear Instruments and Methods*, 558(1):1–10, 2006.

[30] Roger A. Broucke. Numerical integration of periodic orbits in the main problem of artificial satellite theory. *Celestial Mechanics and Dynamical Astronomy*, 58(2):99–123, 1994.

[31] Owen Brown and Paul Eremenko. Fractionated space architectures: A vision for responsive space. Technical report, Defense Advanced Research Projects Agency Arlington VA, 2006.

[32] Ernest D. Courant and Hartland S. Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 3(1):1 – 48, 1958.

[33] Simone D'Amico, Jean Sebastien Ardaens, and Robin Larsson. Spaceborne autonomous formation flying experiment on the prisma mission. *Journal of Guidance, Control and Dynamics*, 35(3):834–850, 2012.

[34] Simone D'Amico and Oliver Montenbruck. Proximity operations of formation-flying spacecraft using an eccentricity/inclination vector separation. *Journal of Guidance Control and Dynamics*, 29(3):554–563, 2006.

[35] Paul A. M. Dirac. The quantum theory of the electron. *Proc. R. Soc. Lond. A*, 117(778):610–624, 1928.

[36] Paul A. M. Dirac. The quantum theory of the electron. part II. *Proc. R. Soc. Lond. A*, 118(779):351–361, 1928.

[37] Kaisheng Du and Ralph B. Kearfott. The cluster problem in multivariate global optimization. *J. Global Optim.*, 5:253–265, 1994.

[38] Etienne Forest, John Irwin, and Martin Berz. Normal form methods for complicated periodic systems. *Part. Accel.*, 24:91–107, 1989.

[39] Joe M. Grange et al. (Muon g-2 Collaboration). Muon (g-2) technical design report. Fermilab Technical Publications FERMILAB-FN-0992-E, Fermi National Accelerator Laboratory (FNAL), 2015.

[40] Johannes Grote, Martin Berz, and Kyoko Makino. High-order representation of Poincaré maps. *Nuclear Instruments and Methods A*, 558(1):106–111, 2006.

[41] Johannes Grote, Kyoko Makino, and Martin Berz. Verified computation of high-order Poincaré maps. *Transactions on Systems*, 4(11):1986–1992, 2005.

[42] Yanchao He, Roberto Armellin, and Ming Xu. Bounded relative orbits in the zonal problem via high-order Poincaré maps. *Journal of Guidance, Control, and Dynamics*, 42(1):91–108, 2018.

[43] Ralph B. Kearfott. *Rigorous Global Search: Continuous Problems*. Kluwer, Dordrecht, 1996.

[44] Ralph B. Kearfott and Kaisheng Du. The cluster problem in global optimization: The univariate case. In Stetter H.J. Albrecht R., Alefeld G., editor, *Validation Numerics. Computing Supplementum*, volume 9, pages 117–127. Springer, Vienna, 1992.

[45] Ellis R. Kolchin. *Differential algebra & algebraic groups*. Academic press, 1973.

[46] Wang Sang Koon, Jerrold E. Marsden, Richard M. Murray, and Josep Masdemont. J_2 dynamics and formation flight. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Montreal, Canada, 2001. AIAA.

[47] Ulrich W. Kulisch and Willard L. Miranker. *Computer Arithmetic in Theory and Practice*. Academic Press, New York, 1981.

[48] Vangipuram Lakshmikantham, Vladimir M. Matrosov, and Seenith Sivasundaram. *Vector Lyapunov Functions and Stability Analysis of Nonlinear Systems*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1991.

[49] John E. Lennard-Jones. On the determination of molecular fields. ii. from the equation of state of gas. *Proceedings of the Royal Society of London A*, 106:463–477, 1924.

[50] John E. Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43(5):461, 1931.

[51] Aleksandr M. Lyapunov. *The General Problem of the Stability of Motion*. Taylor and Francis, London, 1992.

[52] Kyoko Makino. Private communication.

[53] Kyoko Makino. *Rigorous Analysis of Nonlinear Motion in Particle Accelerators*. PhD thesis, Michigan State University, East Lansing, Michigan, USA, 1998. Also MSUCL-1093.

[54] Kyoko Makino and Martin Berz. Remainder differential algebras and their applications. In M. Berz, C. Bischof, G. Corliss, and A. Griewank, editors, *Computational Differentiation: Techniques, Applications, and Tools*, pages 63–74. SIAM, 1996.

[55] Kyoko Makino and Martin Berz. Efficient control of the dependency problem based on Taylor model methods. *Reliable Computing*, 5(1):3–12, 1999.

[56] Kyoko Makino and Martin Berz. Effects of kinematic correction on the dynamics in muon rings. *AIP CP*, 530:217–227, 2000.

[57] Kyoko Makino and Martin Berz. Verified global optimization with Taylor model methods. In N. Mastorakis, editor, *Problems in Modern Applied Mathematics*, pages 253–258. World Scientific and Engineering Society Press, 2000.

[58] Kyoko Makino and Martin Berz. Taylor models and other validated functional inclusion methods. *International Journal of Pure and Applied Mathematics*, 6(3):239–316, 2003.

[59] Kyoko Makino and Martin Berz. Range bounding for global optimization with Taylor models. *Transactions on Computers*, 4(11):1611–1618, 2005.

[60] Kyoko Makino and Martin Berz. Suppression of the wrapping effect by Taylor model-based verified integrators: Long-term stabilization by preconditioning. *International Journal of Differential Equations and Applications*, 10(4):353–384, 2005.

[61] Kyoko Makino and Martin Berz. COSY INFINITY version 9. *Nuclear Instruments and Methods*, 558(1):346–350, 2006.

[62] Kyoko Makino and Martin Berz. Suppression of the wrapping effect by Taylor model-based verified integrators: The single step. *International Journal of Pure and Applied Mathematics*, 36(2):175–196, 2007.

[63] Kyoko Makino and Martin Berz. Optimal correction and design parameter search by modern methods of rigorous global optimization. *Nuclear Instruments and Methods*, 645(1):332–337, 2011.

[64] Kyoko Makino and Martin Berz. The LDB, QDB, and QFB bounders. Technical Report MSUHEP-40617, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, June 2004.

[65] Vladimir Martinusi and Pini Gurfil. Closed-form solutions for satellite relative motion in an axially-symmetric gravitational field. *Advances in the Astronautical Sciences*, 140:1525–1544, 2011.

[66] Oliver Montenbruck, Michael Kirschner, Simone D'Amico, and Srinivas Bettadpur. E/I-vector separation for safe switching of the grace formation. *Aerospace Science and Technology*, 10(7):628–635, 2006.

[67] Ramon E. Moore. *Interval Analysis*, volume 4. Prentice-Hall Englewood Cliffs, 1966.

[68] Ramon E. Moore. *Methods and Applications of Interval Analysis*. SIAM, 1979.

[69] Ramon E. Moore, Eldon Hansen, and Anthony Leclerc. Rigorous methods for global optimization. In *Recent Advances in Global Optimization (Princeton, NJ, 1991)*, Princeton Ser. Comput. Sci., pages 321–342. Princeton Univ. Press, 1992.

[70] John E. Nafe, Edward B. Nelson, and Isidor I. Rabi. The hyperfine structure of atomic hydrogen and deuterium. *Physical Review*, 71(12):914, 1947.

[71] Darragh E. Nagle, Renne S. Julian, and Jerrold R. Zacharias. The hyperfine structure of atomic hydrogen and deuterium. *Physical Review*, 72(10):971, 1947.

[72] National Energy Research Scientific Computing (NERSC). https://www.nersc.gov/.

[73] Nikolai N. Nekhoroschev. An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Uspekhi Mat. Nauk 32:6, English translation Russ. Math. Surv.*, 32(6):5–66, 1977.

[74] Henri Poincaré. *Les méhodes nouvelles de la mécanique céleste*, volume I-III. Gauthier-Villars it fils, 1892, 1893, 1899.

[75] Nathalie Revol, Kyoko Makino, and Martin Berz. Taylor models and floating-point arithmetic: Proof that arithmetic operations are validated in COSY. *Journal of Logic and Algebraic Programming*, 64(1):135–154, 2004.

[76] Joseph F. Ritt. *Differential equations from the algebraic standpoint*, volume 14. American Mathematical Soc., Washington, D.C., 1932.

[77] Joseph F. Ritt and Joseph Liouville. *Integration in finite terms: Liouville's theory of elementary methods*. Columbia Univ. Press, 1948.

[78] Howard H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 1960.

[79] Hanspeter Schaub and Kyle T. Alfriend. J_2 invariant relative orbits for spacecraft formations. *Celestial Mechanics and Dynamical Astronomy*, 79(2):77–95, 2001.

[80] Julian Schwinger. Quantum electrodynamics. I. A covariant formulation. *Physical Review*, 74(10):1439, 1948.

[81] Julian Schwinger. Quantum electrodynamics. III. The electromagnetic properties of the electron—radiative corrections to scattering. *Physical Review*, 76(6):790, 1949.

[82] Yannis K. Semertzidis, Gerald Bennett, Efstratios Efstathiadis, Frank Krienen, Richard Larsen, et al. The Brookhaven Muon (g-2) Storage Ring high voltage quadrupoles. *Nuclear Instruments and Methods A*, 503(3):458–484, 2003.

[83] Diktys Stratakis, Mary E. Convery, Carol Johnstone, John Johnstone, James P. Morgan, et al. Accelerator performance analysis of the Fermilab Muon Campus. *Physical Review Accelerators and Beams*, 20(11):111003, 2017.

[84] Michael J. Syphers. Long-term muon loss rates and an estimate of ω_a systematic uncertainty. Technical report, Muon g-2 Collaboration, Fermi National Accelerator Laboratory, 2020.

[85] David Tarazona. *Beam dynamics characterization and uncertainties in the Muon g-2 Experiment at Fermilab*. PhD thesis, Michigan State University, East Lansing, Michigan, USA, 2021.

[86] David Tarazona, Martin Berz, and Kyoko Makino. Muon loss rates from betatron resonances at the Muon g-2 Storage Ring at Fermilab. *International Journal of Modern Physics A*, 34(36):1942008, 2019.

[87] David Tarazona, Martin Berz, Kyoko Makino, Diktys Stratakis, and Michael J. Syphers. Dynamical simulations of the Muon Campus at Fermilab. *International Journal of Modern Physics A*, 34(36):1942033, 2019.

[88] David Tarazona, Eremey Valetov, Adrian Weisskopf, Martin Berz, and Kyoko Makino. E989 note 265: Lost-muon studies. Technical report, Fermilab Muon g-2, 2021.

[89] Edward C. Titchmarsh. *The theory of the Riemann zeta-function*. Oxford University Press, 2nd edition revised by d. r. heath-brown edition, 1986.

[90] Srinivas R. Vadali, Hanspeter Schaub, and Kyle T. Alfriend. Initial conditions and fuel-optimal control for formation flying of satellites. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Portland, OR, 1999.

[91] Eremey Valetov, Martin Berz, and Kyoko Makino. Validation of transfer map calculation for electrostatic deflectors in the code COSY INFINITY. *International Journal of Modern Physics A*, 34(36):1942010, 2019.

[92] Graziano Venanzoni (on behalf of the Fermilab E989 Collaboration). The new Muon g-2 Experiment at Fermilab. *Nuclear and Particle Physics Proceedings*, 273:584–588, 2016.

[93] Adrian Weisskopf. Applications of the DA based normal form algorithm on parameter-dependent perturbations. Master's thesis, Michigan State University, East Lansing, Michigan, USA, 2016.

[94] Adrian Weisskopf. Introduction to the differential algebra normal form algorithm using the centrifugal governor as an example. Technical Report MSUHEP-190617, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2019. arXiv:1906.10758[physics.class-ph].

[95] Adrian Weisskopf, Roberto Armellin, and Martin Berz. Bounded motion design in the Earth zonal problem using differential algebra based normal form methods. *Celestial Mechanics and Dynamical Astronomy*, 132(14), 2020.

[96] Adrian Weisskopf, David Tarazona, and Martin Berz. Computation and consequences of high order amplitude- and parameter-dependent tune shifts in storage rings for high precision measurements. *International Journal of Modern Physics A*, 34(36):1942011, 2019.

[97] Bong Wie. *Space vehicle dynamics and control*. American Institute of Aeronautics and Astronautics, 2008.

[98] Alexander Wittig. *Rigorous High-Precision Enclosures of Fixed Points and Their Invariant Manifolds*. PhD thesis, Michigan State University, East Lansing, Michigan, USA, 2011.

[99] Alexander Wittig and Martin Berz. High period fixed points, stable and unstable manifolds, and chaos in accelerator transfer maps. *Vestnik Mathematica*, 10(2):93–110, 2014.

[100] Ming Xu, Yue Wang, and Shijie Xu. On the existence of J_2 invariant relative orbits from the dynamical system point of view. *Celestial Mechanics and Dynamical Astronomy*, 112(4):427–444, 2012.