An Interval Method for Linear IVPs for ODEs

Ned Nedialkov
Department of Computing and Software
McMaster University, Canada
nedialk@mcmaster.ca

Joint work with Qiang Song
McMaster University

Improved version of the talk given at the
Workshop on Taylor Models
17–20 December 2003, Miami, Florida

The Problem

Enclose the solution of a system of $n \geq 2$ equations IVP

$$y' = A(t)y + g(t), \quad y(0) = y_0 \in [y_0].$$

Idea (Lohner, Nickel)

- Perform $(n+1)$ integrations of points specifying a parallelepiped at t_i and enclose each point solution at t_{i+1}.

  We have $(n+1)$ boxes.

- Find $(n+1)$ points that determine a parallelepiped, which encloses all the parallelepipeds with vertices in these boxes.

- Repeat.

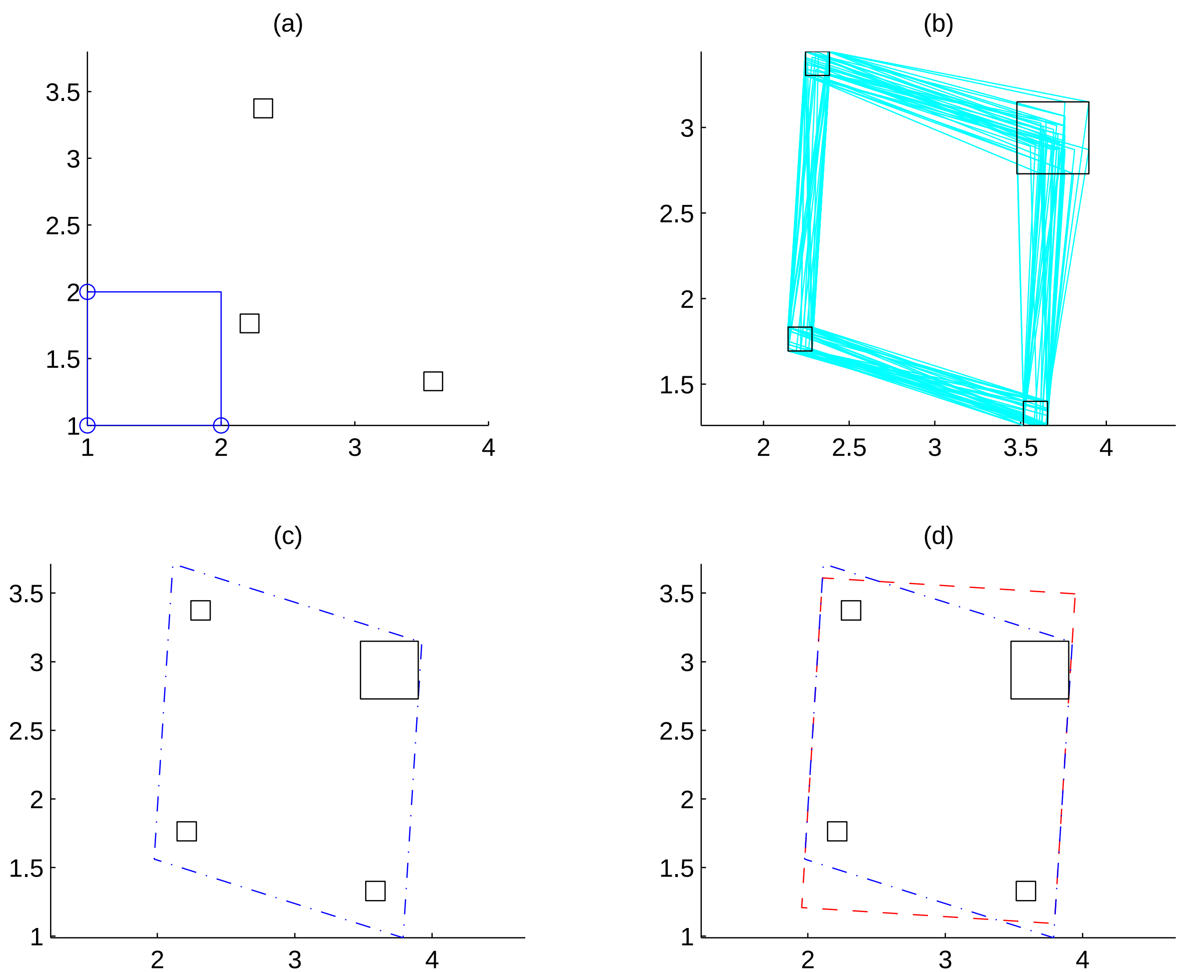

Figure 1: (a) enclosures of point solutions at t_1; (b) some of the parallelepipeds with vertices in these enclosures (boxes); the larger box contains the fourth vertices; (c–d) parallelepipeds enclosing the true solution

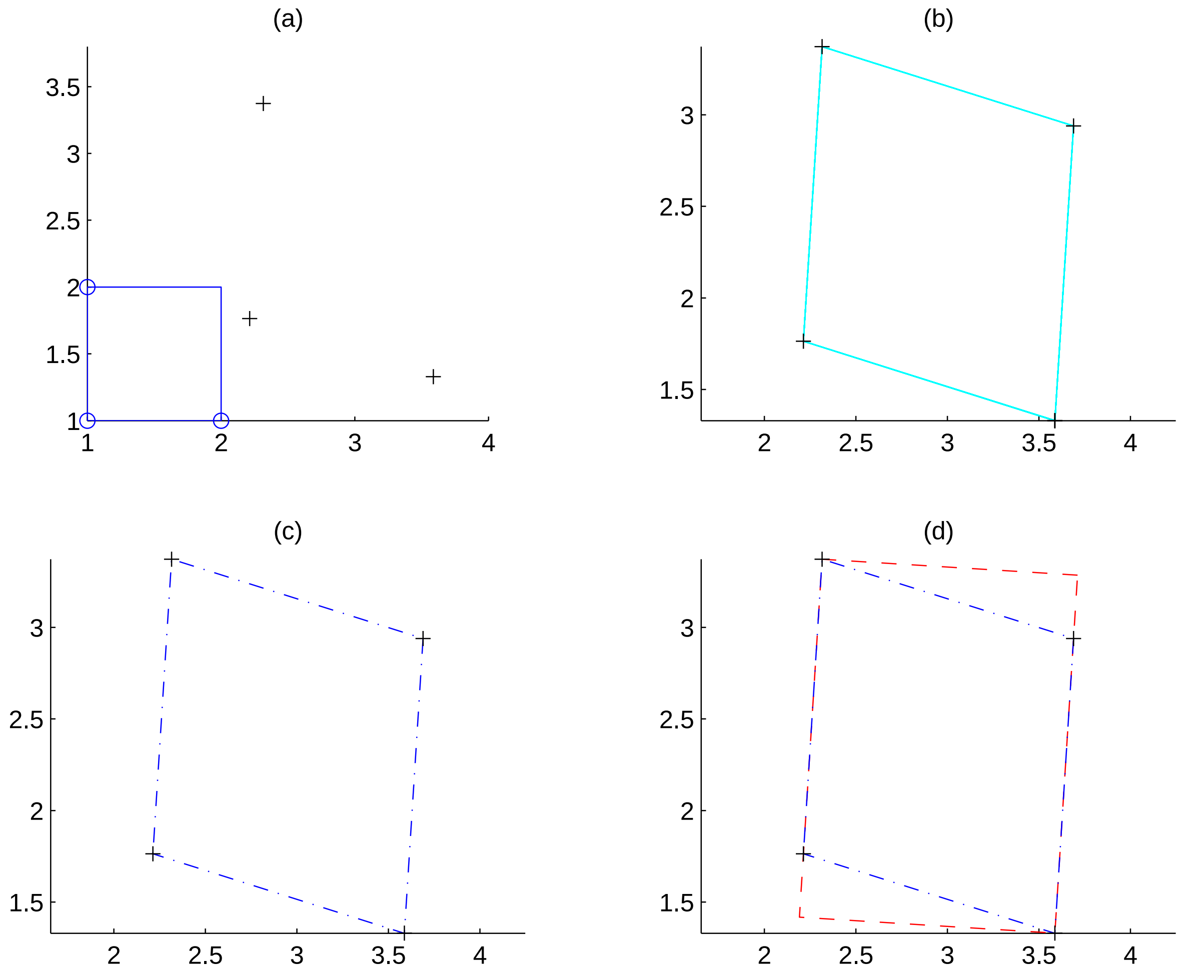

Figure 2: The same computation as in the previous figure, except that the width of each component of the enclosures is 2×10^{-10}. The boxes are denoted by "$+$".

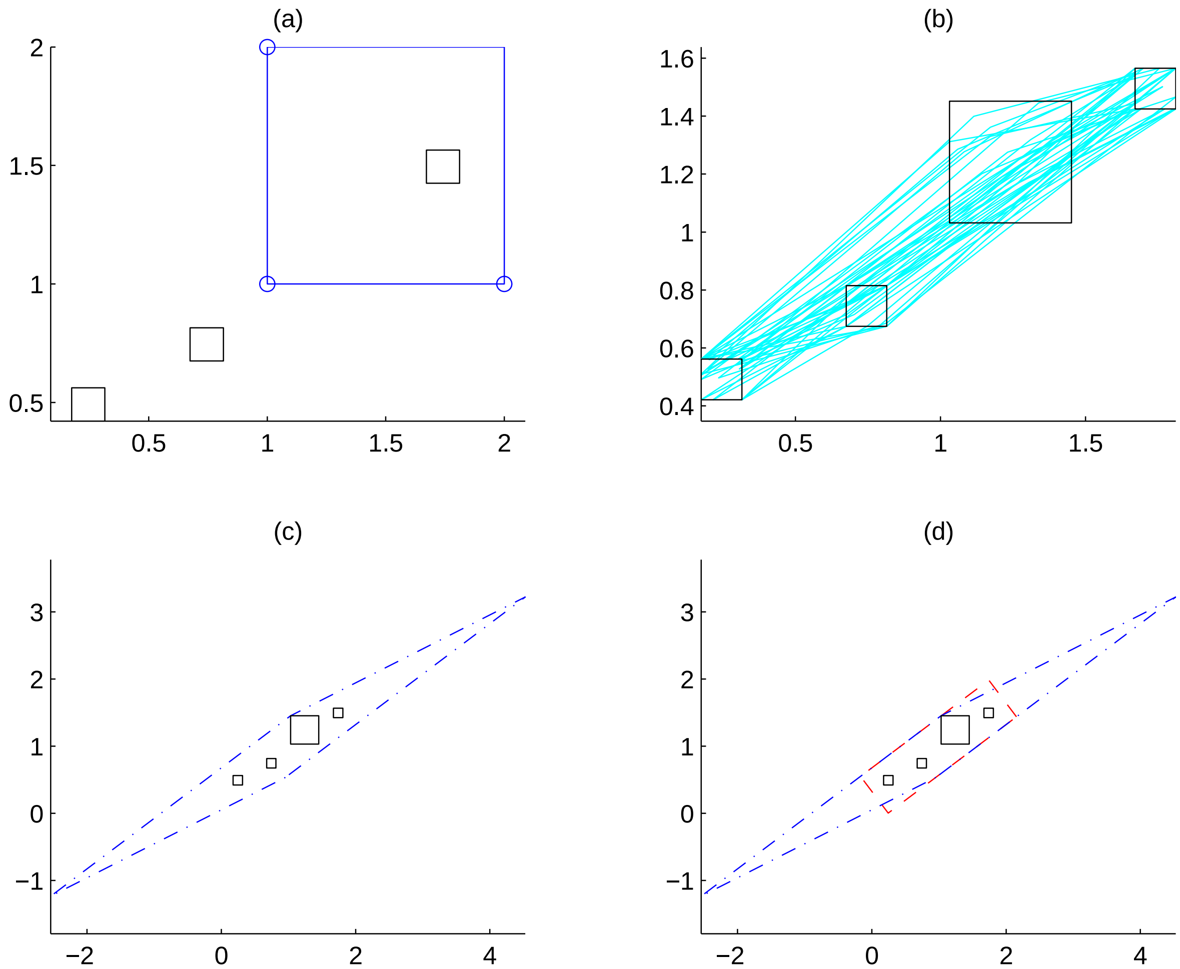

Figure 3: (a) enclosures of point solutions at t_1; (b) some of the parallelepipeds with vertices in these enclosures (boxes); the larger box contains the fourth vertices; (c–d) parallelepipeds enclosing the true solution

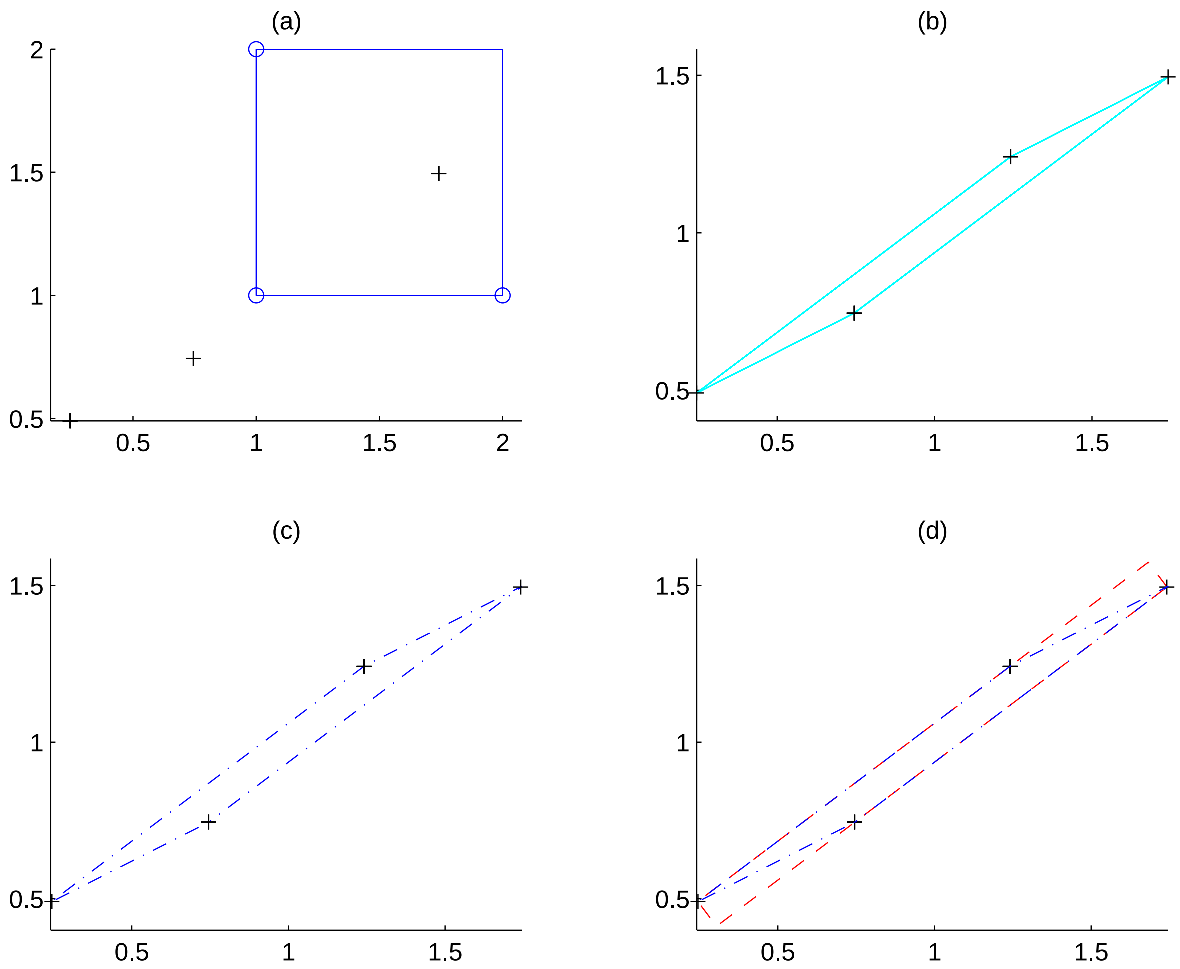

Figure 4: The same computation as in the previous figure, except that the width of each component of the enclosures is 2×10^{-10}. The boxes are denoted by "$+$".

Advantages

- We enclose point solutions:
  Taylor series + remainder term.
- The method does not impose restrictions on the size of the initial box.
- An automatic differentiation package for computing Taylor coefficients for the solution to $Y' = A(t)Y$, $Y(0) = I$ is not needed.

  These coefficients are computed in AWA and VNODE.

Difficulties

- How to compute $(n+1)$ points on each step such that the parallelepiped specified by them encloses the solution set.
- How to achieve small overestimations and reduce the wrapping effect.

Outline

1. Enclosing point solutions
2. Computing a parallelepiped
3. Choice of a transformation matrix
4. Reducing the wrapping effect
5. Concluding remarks

Enclosing Point Solutions

Denote by $f_i(\cdot)$ the ith Taylor coefficient of the solution to

$$y' = A(t)y + g(t). \tag{1}$$

If h and $[\widetilde{y}_0] \ni y_0$ are such that

$$y_0 + \sum_{i=1}^{p-1} t^i f_i(y_0) + t^p f_p([\widetilde{y}_0]) \subseteq [\widetilde{y}_0] \quad \text{for all } t \in [0, h],$$

then (1) with $y(0) = y_0$ has a unique solution in $[0, h]$, and

$$y(t; t_0, y_0) \in [\widetilde{y}_0] \quad \text{for all } t \in [0, h].$$

At $t = h$,

$$y(h; t_0, y_0) \in y_0 + \sum_{i=1}^{p-1} h^i f_i(y_0) + h^p f_p([\widetilde{y}_0]).$$

Assume that at a point t_i, for all $y_0 \in [y_0]$,

$$y(t_i; t_0, y_0) \in \{\, b_0 + B\alpha \mid \alpha \in [0,1]^n \,\},$$

where $B \in \mathbb{R}^{n\times n}$, and $[0,1]^n$ denotes the vector with each component $[0,1]$.

We integrate $v_0 = b_0,\ v_1 = b_0 + b_1, \ldots, v_n = b_0 + b_n$ to compute $[w_0], [w_1], \ldots, [w_n]$.

That is, for each v_j,

$$y\big(t_{i+1}; t_i, v_j\big) \in [w_j] = v_j + \sum_{i=1}^{p-1} h^i f_i(v_j) + h^p f_p([\widetilde{v}_j]),$$

where $v_j \in [\widetilde{v}_j]$.

Computing a Parallelepiped

Denote

$$\begin{aligned} c_j &= \mathrm{mid}([w_j]), \\ [e_j] &= [w_j] - c_j, \quad j = 0, \ldots, n, \\ C & \quad \text{the } n \times n \text{ matrix with } j\text{th column } c_j - c_0, \quad \text{and} \\ [e] &= \sum_{j=1}^{n} [e_j] + (n-1)[e_0]. \end{aligned}$$

For all $y_i \in \{\, b_0 + B\alpha \mid \alpha \in [0,1]^n \,\}$,

$$\begin{aligned} y(t_{i+1}; t_i, y_i) &\in \Big\{ w_0 + \sum_{j=1}^{n} \alpha_j (w_j - w_0) \mid \alpha_j \in [0,1],\, w_j \in [w_j] \Big\} \\ &\subseteq \{\, c_0 + C\alpha + [e] \mid \alpha \in [0,1]^n \,\}, \end{aligned}$$

since for $\alpha \in [0,1]^n$, $w_j \in [w_j]$, and $e_j = w_j - c_j \in [e_j]$ $(j = 0, \ldots, n)$,

$$\begin{aligned}
& w_0 + \sum_{j=1}^{n} \alpha_j (w_j - w_0) \\
& \quad = c_0 + C\alpha + (w_0 - c_0) + \sum_{j=1}^{n} \alpha_j \big(w_j - w_0 - (c_j - c_0)\big) \\
& \quad = c_0 + C\alpha + e_0 + \sum_{j=1}^{n} \alpha_j (e_j - e_0) \\
& \quad = c_0 + C\alpha + \sum_{j=1}^{n} \alpha_j e_j + (1 - \sum_{j=1}^{n} \alpha_j) e_0 \\
& \quad \in \left\{ \, c_0 + C\alpha + \sum_{j=1}^{n} [e_j] + (n-1)[e_0] \mid \alpha \in [0,1]^n \, \right\} \\
& \quad = \left\{ \, c_0 + C\alpha + [e] \mid \alpha \in [0,1]^n \, \right\}.
\end{aligned}$$

(Note that each $[e_j]$ is symmetric.)

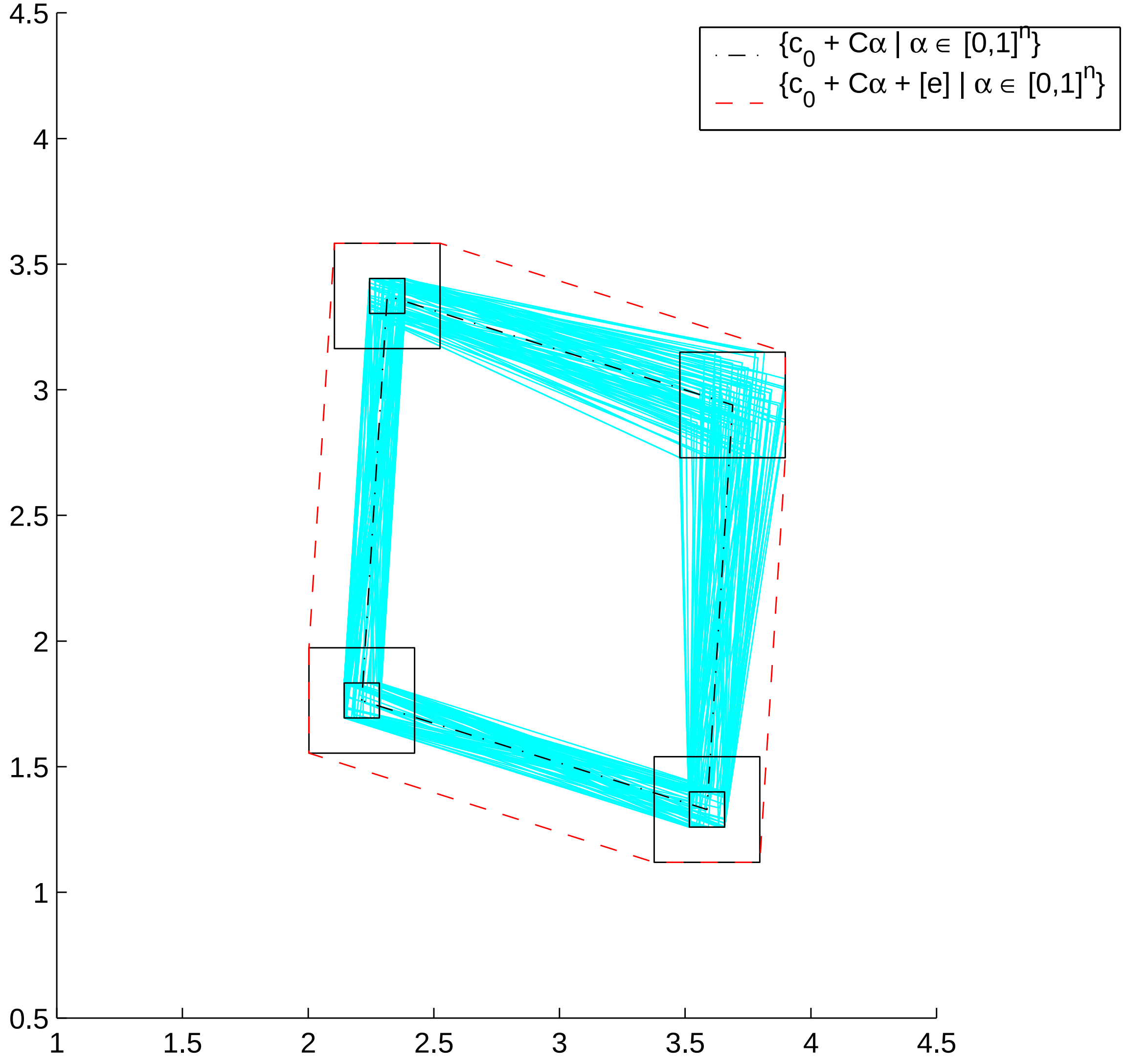

Figure 5: We want to enclose the set $\{\, c_0 + C\alpha + [e] \mid \alpha \in [0,1]^n \,\}$ by a parallelepiped.

We want to find g_0 and G such that

$$\{\, c_0 + C\alpha + e \mid \alpha \in [0,1]^n,\ e \in [e] \,\} \subseteq \{\, g_0 + G\alpha \mid \alpha \in [0,1]^n \,\}.$$

Let $H \in \mathbb{R}^{n\times n}$ be nonsingular.

Denote

$$[r] = (H^{-1}C)[0,1]^n + H^{-1}[e] \quad \text{and} \quad D = \operatorname{diag}\big(\mathrm{w}([r])\big).$$

Then

$$\begin{aligned}
&\big\{c_0 + C\alpha + e \mid \alpha \in [0,1]^n,\ e \in [e]\big\} \\
&\qquad\qquad = \{\, c_0 + H\big((H^{-1}C)\alpha + H^{-1}e\big) \mid \alpha \in [0,1]^n,\ e \in [e] \,\} \\
&\qquad\qquad \subseteq \{\, c_0 + Hr \mid r \in [r] \,\} \\
&\qquad\qquad = \{\, c_0 + H\underline{r} + Hr \mid r \in [0, \overline{r} - \underline{r}] = D[0,1]^n \,\} \\
&\qquad\qquad = \{\, (c_0 + H\underline{r}) + (HD)\alpha \mid \alpha \in [0,1]^n \,\} \\
&\qquad\qquad = \{\, g_0 + G\alpha \mid \alpha \in [0,1]^n \,\}.
\end{aligned}$$

This derivation is by R. Lohner (2001, private communications).

Now, for all $y_i \in \{\, b_0 + B\alpha \mid \alpha \in [0,1]^n \,\}$,

$$y(t_{i+1}; t_i, y_i) \in \{\, g_0 + G\alpha \mid \alpha \in [0,1]^n \,\}.$$

We integrate $g_0,\ (g_0 + g_1), \ldots,\ (g_0 + g_n).$

Subtlety: we compute in floating-point arithmetic $\widetilde{g}_0$ and $\widetilde{G}$ corresponding to g_0 and G.

Is

$$\{\, c_0 + C\alpha + e \mid \alpha \in [0,1]^n,\ e \in [e] \,\} \subseteq \{\, \widetilde{g}_0 + \widetilde{G}\alpha \mid \alpha \in [0,1]^n \,\}\,? \quad (2)$$

If

$$\widetilde{G}^{-1}(c_0 - g_0) + (\widetilde{G}^{-1}C)[0,1]^n + \widetilde{G}^{-1}[e] \subseteq [0,1]^n \qquad (3)$$

then (2) holds.

If (3) does not hold in computer arithmetic, inflate $[e]$ and try again.

Choice of a Transformation Matrix

Parallelepiped method

$$H = C,$$
$$[r] = (H^{-1}C)[0,1]^n + H^{-1}[e] = [0,1]^n + C^{-1}[e].$$

This method breaks down when C is close to singular.

QR-factorization method

$$C = QR, \quad H = Q,$$
$$[r] = (H^{-1}C)[0,1]^n + H^{-1}[e] = R[0,1]^n + Q^T[e].$$

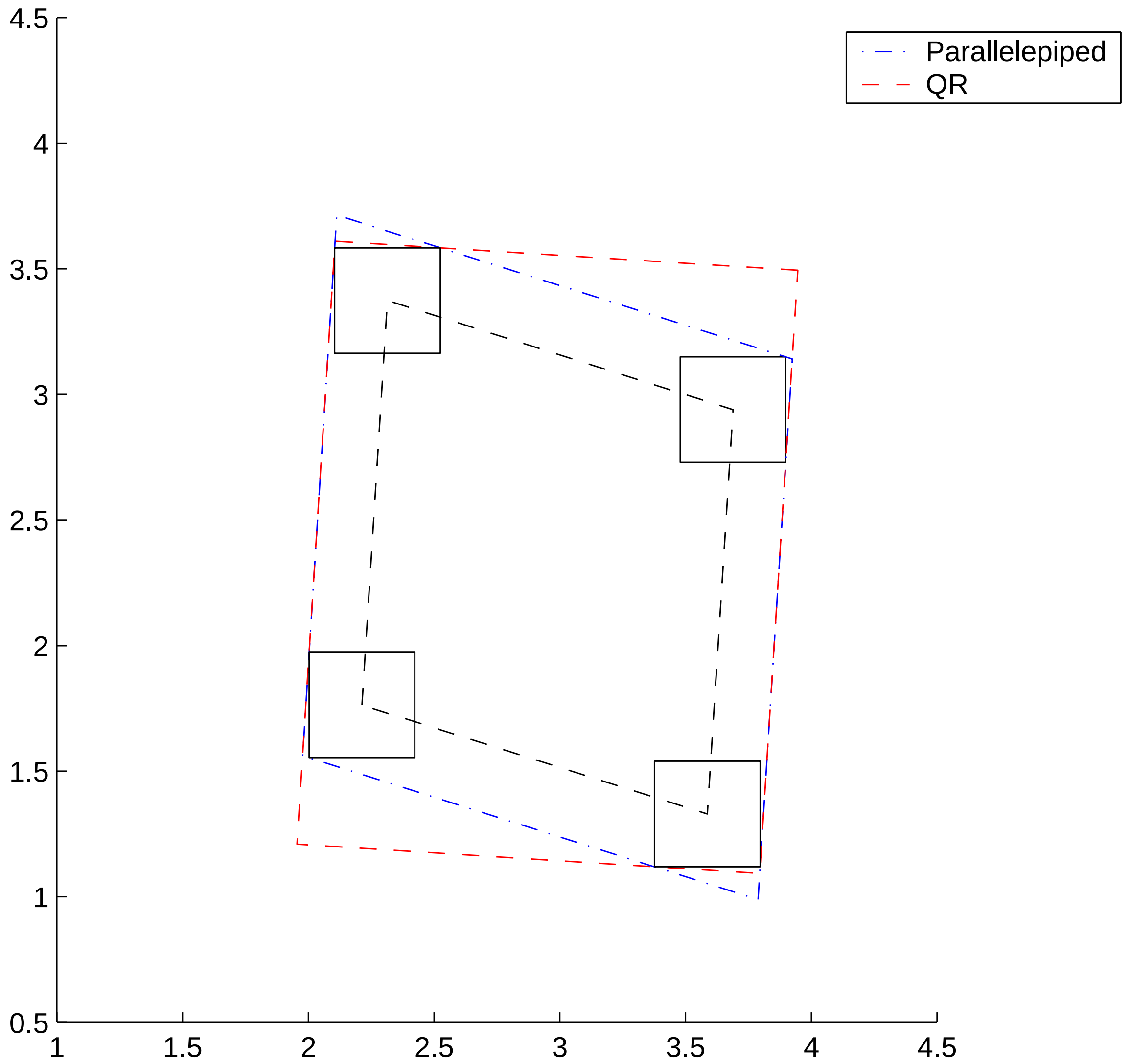

Figure 6: Enclosures obtained by the parallelepiped and QR approaches.

On some problems, with a large initial box, the QR method can produce large overestimations.

Example:

$$y' = \begin{pmatrix} 1 & -2 \\ 3 & -4 \end{pmatrix} y, \quad y(0) \in ([1,2],[1,2])^T.$$

We take $[e] = [-10^{-3}, 10^{-3}]$, $h = 0.2$.

The eigenvalues of $\exp(hA)$ are ≈ 0.8187 and 0.6703.

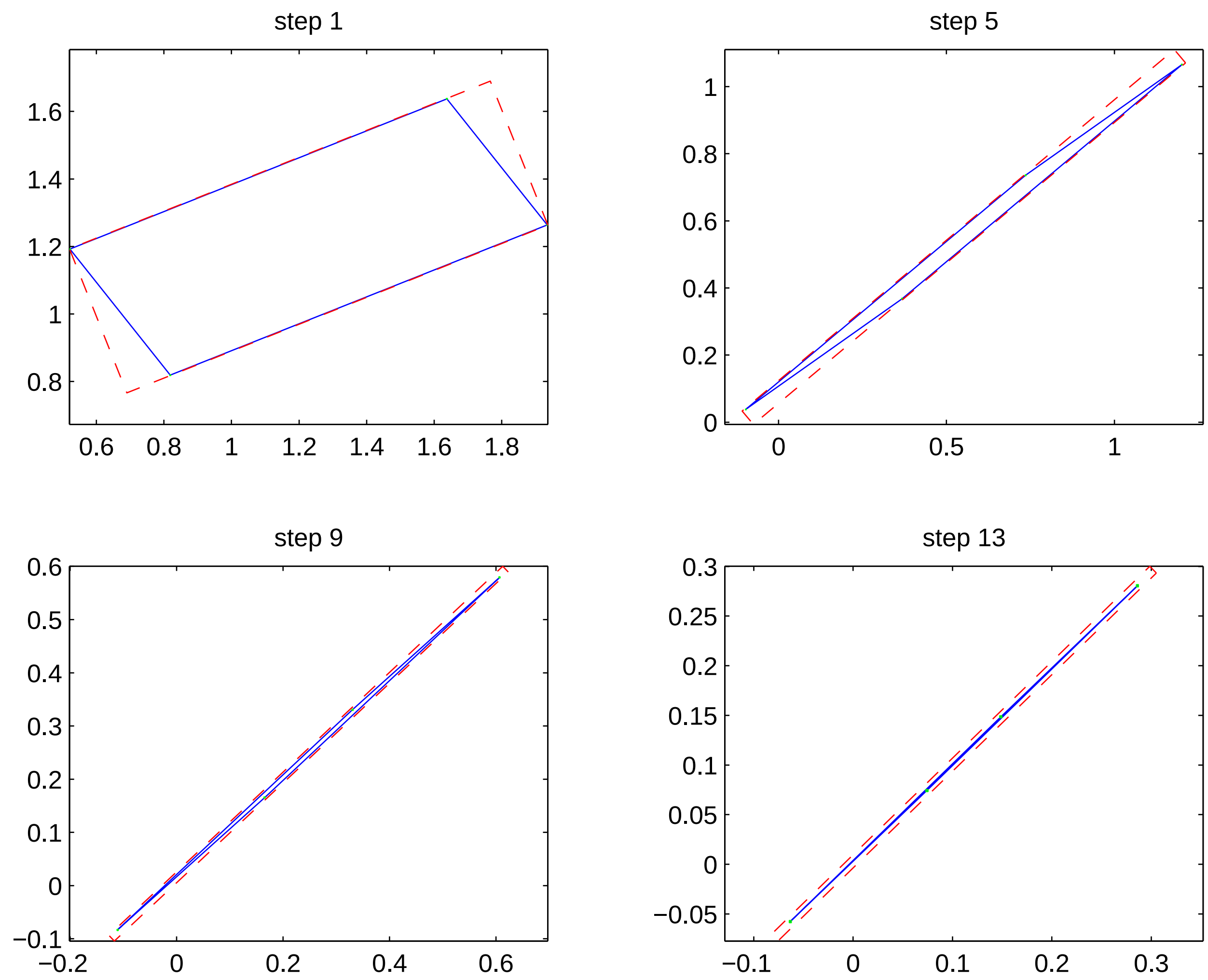

Figure 7: QR; the blue lines denote the true solution set.

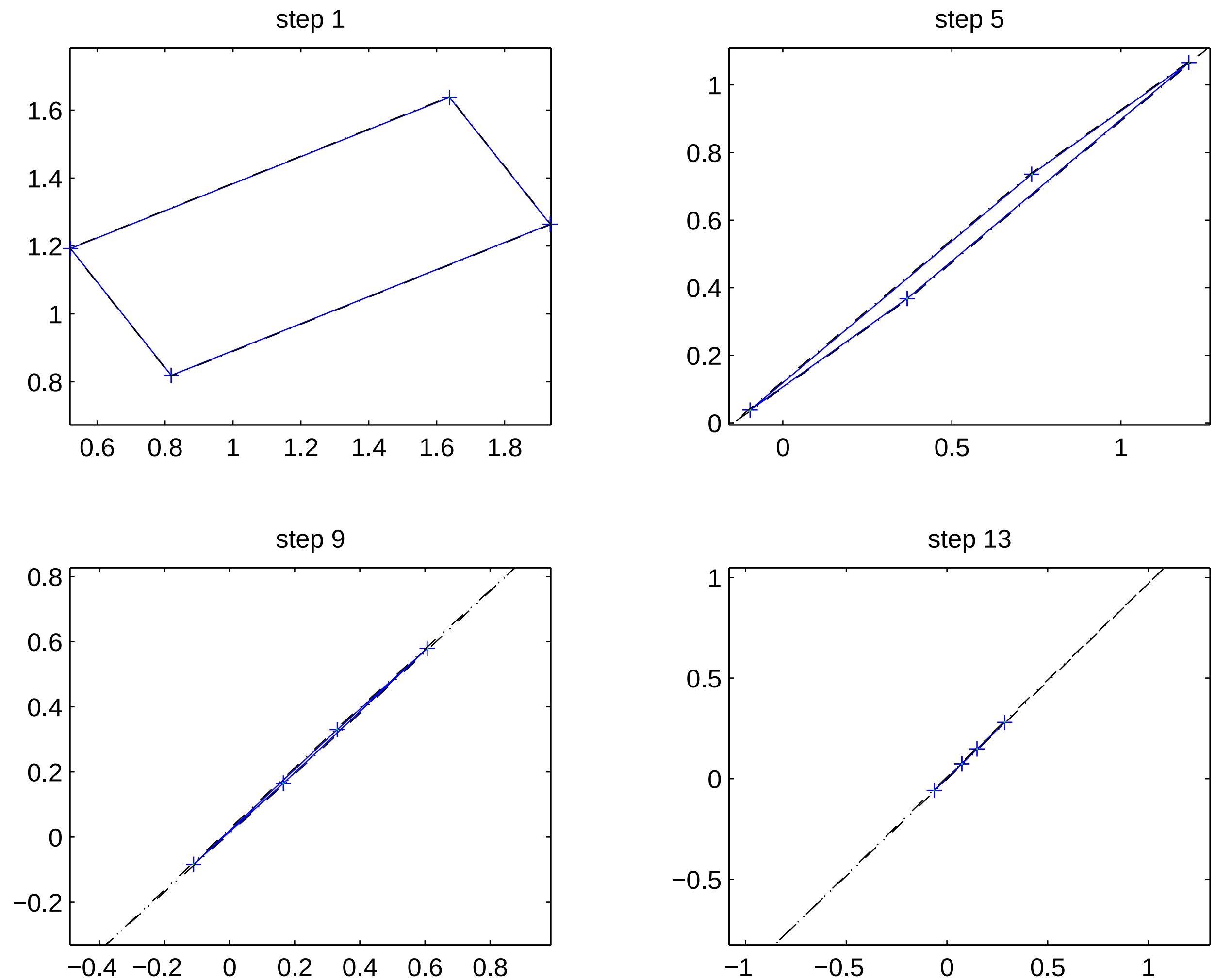

Figure 8: Parallelepiped; the vertices of the true solution are denoted by "+".

Example:

$$y' = \begin{pmatrix} -0.5 & 1 \\ -1 & 0 \end{pmatrix} y, \quad y(0) \in ([1,2],[1,2])^T.$$

We take $[e] = [-10^{-4}, 10^{-4}]$, $h = 0.2$.

The eigenvalues of $\exp(hA)$ are $\approx 0.9334 \pm 0.1831i$, and $\rho\big(\exp(hA)\big) \approx 0.9512$.

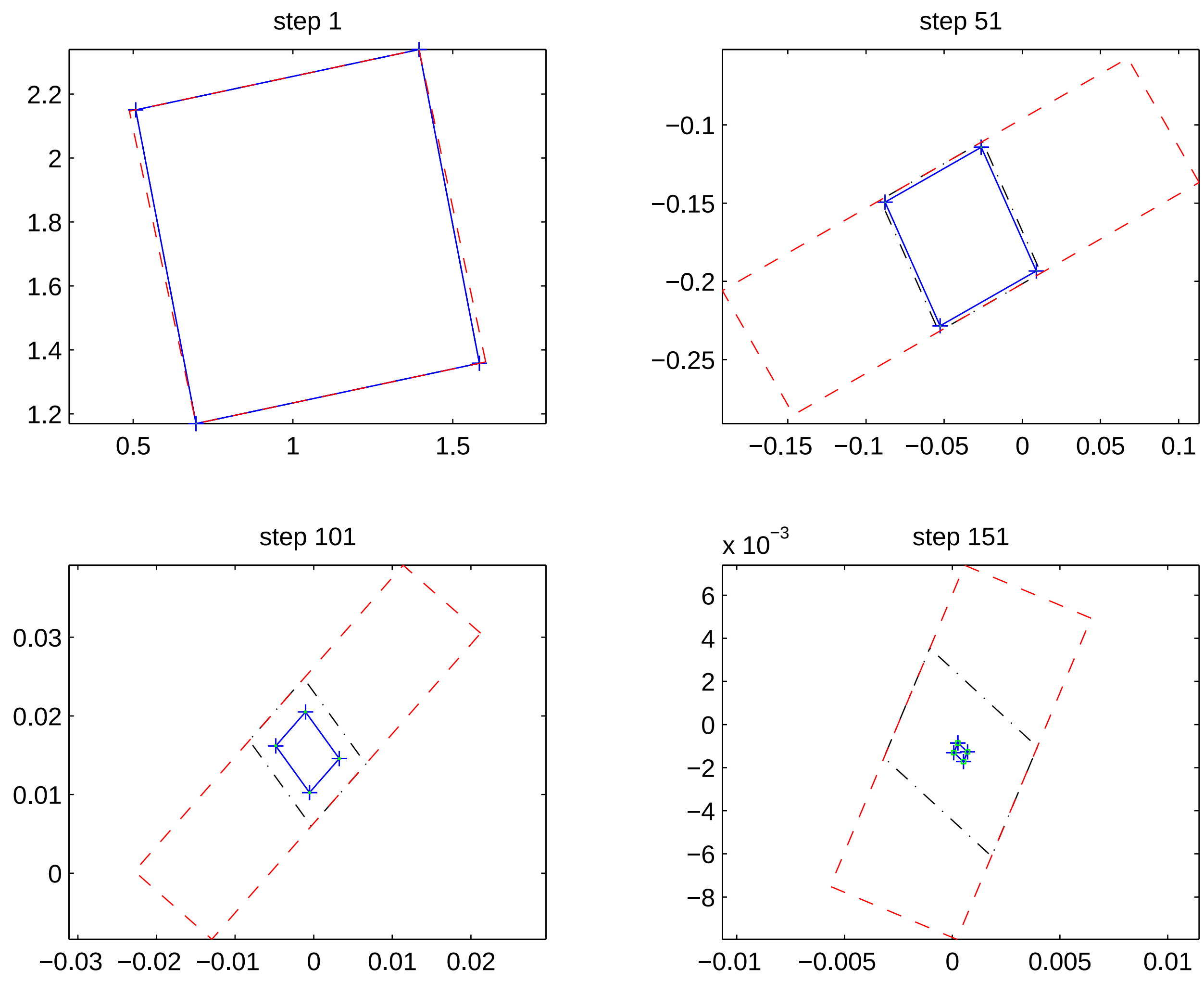

Figure 9: QR and parallelepiped methods

Reducing the Wrapping Effect

The true solution is in

$$\{\, g_0 + G\alpha \mid \alpha \in [0,1]^n \,\} \\ = \{\, c_0 + Hr \mid r \in (H^{-1}C)[0,1]^n + H^{-1}[e] \,\}.$$

Parallelepiped

$$H = C, \quad [r_{\mathrm{P}}] = [0,1]^n + C^{-1}[e].$$

QR factorization

$$C = QR, \quad H = Q, \quad [r_{\mathrm{Q}}] = R[0,1]^n + Q^T[e].$$

Can we combine them, or switch between them at run time?

Two ad-hoc solutions: Approach I and II.

Approach I

We can (roughly) measure the overestimations in the parallelepiped and QR methods by $\left\|\mathrm{w}\big(C[r_{\mathrm{P}}]\big)\right\|$ and $\left\|\mathrm{w}\big(Q[r_{\mathrm{Q}}]\big)\right\|$, respectively.

Select:

if $\left\|\mathrm{w}\big(C[r_{\mathrm{P}}]\big)\right\| \leq \left\|\mathrm{w}\big(Q[r_{\mathrm{Q}}]\big)\right\|$
  $H = C$, $[r] = [r_{\mathrm{P}}]$ (parallelepiped)
else
  $H = Q$, $[r] = [r_{\mathrm{Q}}]$ (QR)

Example:

$$y' = \begin{pmatrix} 1 & -2 \\ 3 & -4 \end{pmatrix} y, \quad y(0) \in ([1,2],[1,2])^T,$$

$$[e] = [-10^{-3}, 10^{-3}], \quad h = 0.2.$$

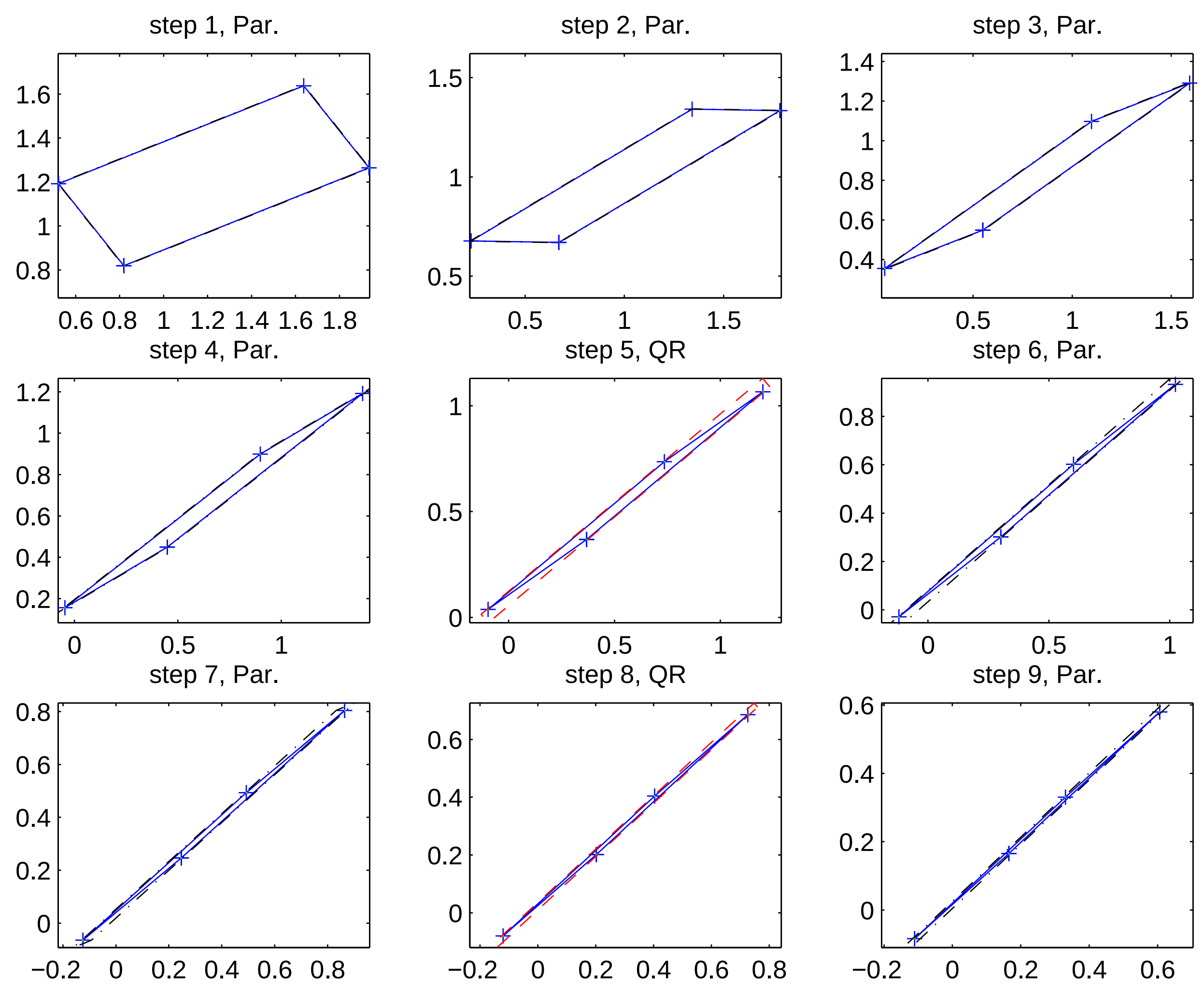

Figure 10: Approach I

Approach II

Let $\beta_{\max}$ be the largest angle among the angles between every two columns of C.

Let $\beta_{\min}$ be the smallest such angle.

Let θ, $0 < \theta \ll \pi$, be a constant.

Select:

if $\beta_{\min} > \theta$ and $\beta_{\max} < \pi - \theta$

$H = C$, $[r] = [r_{\mathrm{P}}]$ (parallelepiped)

else

$H = Q$, $[r] = [r_{\mathrm{Q}}]$ (QR)

Example:

$$y' = \begin{pmatrix} 1 & -2 \\ 3 & -4 \end{pmatrix} y, \quad y(0) \in ([1,2],[1,2])^T,$$

$$[e] = [-10^{-3}, 10^{-3}], \quad h = 0.2,$$

$$\theta = 10^o = \pi/18.$$

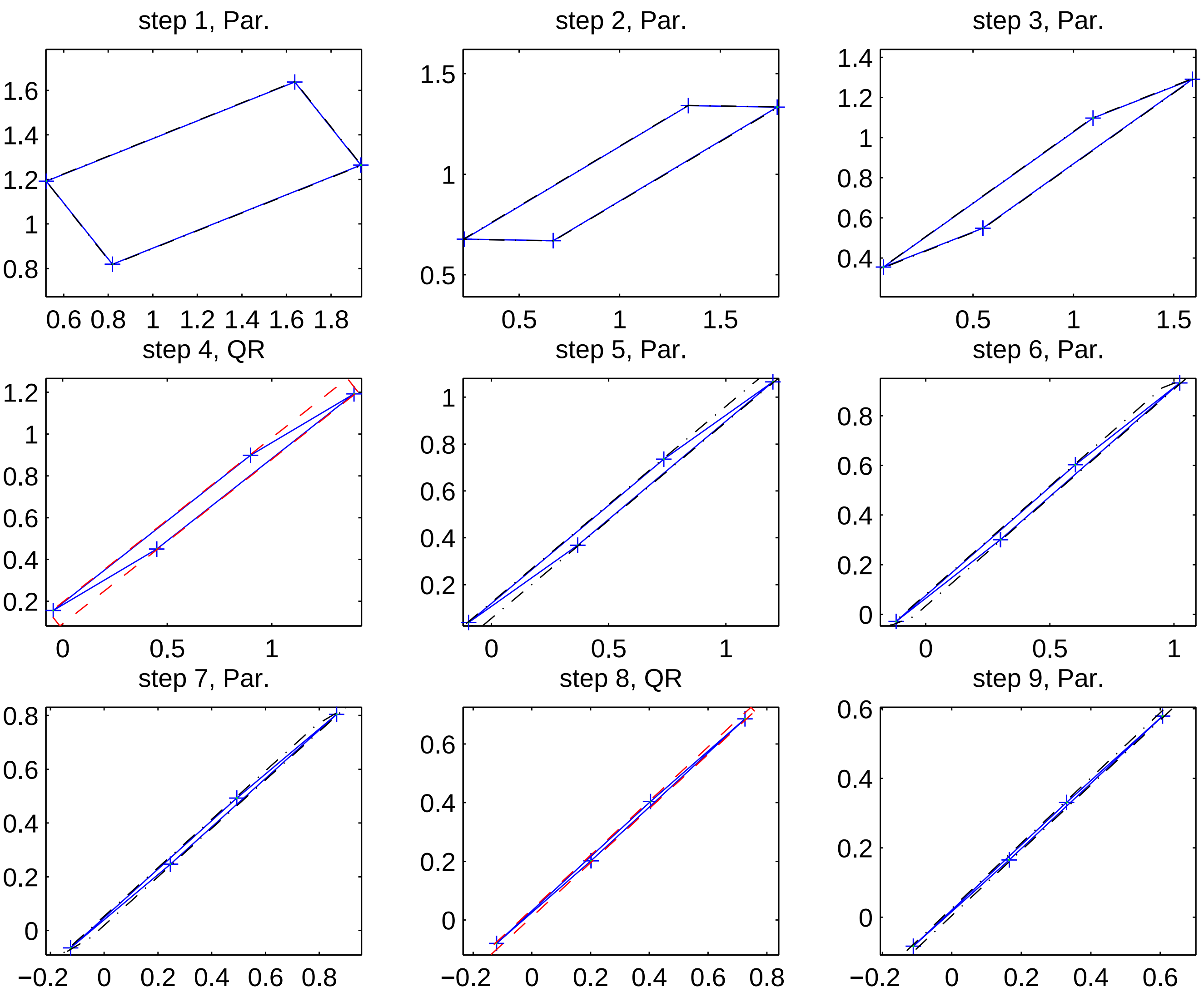

Figure 11: Approach II

Concluding Remarks

- To reduce the wrapping effect when propagating larger sets, a combination of the parallelepiped and QR-factorization methods may be necessary.
- When to switch from one method to the other?
- An eigenvalue, or stability type analysis of a combined approach may be necessary.