J. Math. Study
doi: 10.4208/jms.v50n4.17.04

Vol. **50**, No. 4, pp. 339-357
December 2017

Optimal Parameters for Doubling Algorithms

Tsung-Ming Huang[1], Ren-Cang Li[2,*], Wen-Wei Lin[3]
and Linzhang Lu[4]

[1] *Department of Mathematics, National Taiwan Normal University, Taipei 116, Taiwan*
[2] *Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019, USA*
[3] *Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan*
[4] *School of Mathematics and Computer Science, Guizhou Normal University & School of Mathematical Science, Xiamen University, Xiamen 361005, P. R. China.*

Received May 15, 2017; Accepted August 31, 2017

Abstract. In using the structure-preserving algorithm (SDA) [*Linear Algebra Appl.*, 2005, vol.396, pp.55–80] to solve a continuous-time algebraic Riccati equation, a parameter-dependent linear fractional transformation $z\to(z-\gamma)/(z+\gamma)$ is first performed in order to bring all the eigenvalues of the associated Hamiltonian matrix in the open left half-plane into the open unit disk. The closer the eigenvalues are brought to the origin by the transformation via judiciously selected parameter γ, the faster the convergence of the doubling iteration will be later on. As the first goal of this paper, we consider several common regions that contain the eigenvalues of interest and derive the best γ so that the images of the regions under the transform are closest to the origin. For our second goal, we investigate the same problem arising in solving an M-matrix algebraic Riccati equation by the alternating-directional doubling algorithm (ADDA) [*SIAM J. Matrix Anal. Appl.*, 2012, vol.33, pp.170–194] which uses the product of two linear fractional transformations $(z_1,z_2)\to[(z_2-\gamma_2)/(z_2+\gamma_1)][(z_1-\gamma_1)/(z_1+\gamma_2)]$ that involves two parameters. Illustrative examples are presented to demonstrate the efficiency of our parameter selection strategies.

AMS subject classifications: 15A24, 65F30, 65H10

Key words: Doubling algorithm, SDA, ADDA, optimal parameter, algebraic Riccati equation.

1 Introduction

The following *so-called* continuous-time algebraic Riccati equation (CARE)

$$A^{\mathrm{T}}X+XA-XGX+H=0 \tag{1.1}$$

*Corresponding author. *Email addresses:* `min@ntnu.edu.tw` (T.-M. Huang), `rcli@uta.edu` (R.-C. Li), `wwlin@am.nctu.edu.tw` (W.-W. Lin), `lzlu@xmu.edu.cn` (L. Lu)

 ©2017 Global-Science Press

frequently arises from the continuous-time Linear-Quadratic Gaussian control (LQG) – the $\mathcal{H}_2$-control [12], where $A, G^{\mathrm{T}}=G, H^{\mathrm{T}}=H\in\mathbb{R}^{n\times n}$. Here and in what follows, the superscript T takes the matrix transpose and $\mathbb{R}^{n\times n}$ is the set of all $n\times n$ real matrices. It is well-known that (1.1) is equivalent to

$$\mathscr{H}\begin{bmatrix} I_n \\ X\end{bmatrix} := \begin{bmatrix} A & -G \\ -H & -A^{\mathrm{T}}\end{bmatrix}\begin{bmatrix} I_n \\ X\end{bmatrix} = \begin{bmatrix} I_n \\ X\end{bmatrix}(A-GX), \tag{1.2}$$

where I_n is the $n\times n$ identity matrix. The equation (1.2) implies that for any solution X to (1.1), the column space of $\begin{bmatrix} I_n \\ X\end{bmatrix}$ is an n-dimensional invariant subspace of $\mathscr{H}\in\mathbb{R}^{2n\times 2n}$ associated with its eigenvalues that are those of $A-GX$. More than that, (1.2) leads to

$$\mathscr{H}\begin{bmatrix} I_n & 0 \\ X & I_n\end{bmatrix} = \begin{bmatrix} I_n & 0 \\ X & I_n\end{bmatrix}\begin{bmatrix} A-GX & -G \\ 0 & -(A^{\mathrm{T}}-XG)\end{bmatrix},$$

implying the eigenvalues of $\mathscr{H}$ is the union of the eigenvalues of $A-GX$ and $-(A^{\mathrm{T}}-XG)$.

The matrix $\mathscr{H}$ in (1.2) happens to be real and Hamiltonian, i.e., satisfying

$$\mathscr{H}\mathscr{J}_n = -\mathscr{J}_n\mathscr{H}^{\mathrm{T}} \quad \text{with} \quad \mathscr{J}_n = \begin{bmatrix} 0 & I_n \\ -I_n & 0\end{bmatrix}.$$

As a real Hamiltonian matrix, its eigenvalues come in quadruples $(\lambda,\bar{\lambda},-\lambda,-\bar{\lambda})$, unless λ is purely imaginary in which case, it comes in pairs $(\lambda,-\lambda)$. Under certain conditions from the $\mathcal{H}_2$-optimal control, indeed $\mathscr{H}$ has no eigenvalues on the imaginary axis, and then $\mathscr{H}$ has precisely n eigenvalues in $\mathbb{C}_-$ (the open left half-plane) and n eigenvalues in $\mathbb{C}_+$ (the open right half-plane). The solution to (1.1) of interest is the one for which the eigenvalues of $A-GX$ consist of exactly those of $\mathscr{H}$ in $\mathbb{C}_-$. Denote by Φ this special solution. It can be shown [12] that $\Phi^{\mathrm{T}}=\Phi$, and the eigenvalues of $-(A^{\mathrm{T}}-\Phi G)=-(A-G\Phi)^{\mathrm{T}}$ are precisely the opposites of the ones of $A-G\Phi$.

The doubling algorithms, originally proposed in 1970s for solving CARE and others (see the short survey [3]) but elegantly reformulated in [4] in 2005, turn out to be very much the methods of choice these days to compute the solution Φ for n up to a couple of thousands. Chu, Fan, and Lin [4] also named their reformulation as *structure-preserving doubling algorithm* (SDA) to reflect its structure-preserving feature. For fast convergence of SDA, in the preset-up we have to find an 1-parameter-dependent linear fractional transformation

$$z\in\mathbb{C}\to w(z;\gamma)=\frac{z-\gamma}{z+\gamma} \tag{1.3}$$

that can bring all eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$, i.e., those in $\operatorname{eig}(A-G\Phi)$, the multiset of the eigenvalues of $A-G\Phi$, into the interior of the unit disk, and the asymptotic convergence speed of the doubling iteration is measured by the maximal distance from the transformed eigenvalues to the origin – the smaller the distance the faster the asymptotic convergence. By the knowledge from the complex analysis [2], the mapping $w(z;\gamma)$ maps circles to circles, and, in particular if $\gamma<0$, the imaginary axis $\iota\mathbb{R}$ to the unit circle $\{w:|w|=1\}$, and $\mathbb{C}_-$ to the interior of the open disk $\{w:|w|<1\}$, where ι is the imaginary unit. The question is for what $\gamma<0$, $|w(z;\gamma)|$ is minimized over all $z\in\mathrm{eig}(A-G\Phi)$. It is an important question because the optimal γ will lead to the fastest convergent doubling iteration later on. Roughly speaking, the role of $w(\cdot;\gamma)$ is to suppress the eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$ towards the origin and, as a by-product, expel its eigenvalues in $\mathbb{C}_+$ away from the origin.

In general, it is not an easy task, if at all possible, to find the optimal γ because this requires that all eigenvalues of $\mathscr{H}$ be known. Fortunately, there is no need to find it exactly either. Experience has shown that the doubling iteration usually converges very fast and in fact it converges quadratically. Usually a not-so but sub-optimal γ would only cost a few extra doubling iterative steps. But still it is important to find a somewhat optimal one. To this end, we will have to relax the problem a little bit since, in particular, it is unlikely that the eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$ are known *a priori*. To compensate such unknown, we assume that some region $\Omega\in\mathbb{C}_-$ that contains these eigenvalues through estimations is available. This is a commonly used practice in numerical analysis such as the convergence analysis of the conjugate gradient method and the GMRES method for linear system [7, pp.203-207]. We will assume such a region is symmetric with respect to the real axis, because the eigenvalues of $\mathscr{H}$ come in quadruples $(\lambda,\bar{\lambda},-\lambda,-\bar{\lambda})$.

In view of our discussion above, we shall consider the following minimax problem: find

$$\gamma_{\rm opt}=\arg\min_{\gamma<0}\max_{z\in\Omega}|w(z;\gamma)|, \tag{1.4}$$

where Ω is a bounded connected region in $\mathbb{C}_-$. In this paper, Ω takes one of the following four shapes:

(a) an interval to the left of the origin,

(b) a disk in $\mathbb{C}_-$ whose intersection with the real axis is its diameter,

(c) an ellipse in $\mathbb{C}_-$ whose intersection with the real axis is its major axis, and

(d) a rectangle in $\mathbb{C}_-$ that is symmetric with respect to the real axis.

The rectangle shape is probably the most practically useful one to have, since presumably being able to bound the real parts and the imaginary parts of the interesting eigenvalues is more likely to be the case than any others. Note the interval shape is a special case of the rectangle shape with zero height.

The rest of this paper is organized as follows. In Section 2, we investigate how to select parameter γ in (1.3) for use in SDA to solve CARE (1.1) by solving the minimax problem (1.4). The doubling algorithm ADDA of [10] for solving MARE uses two parameters, unlike SDA. Their selection strategies are investigated in Section 3. Each section includes an illustrative example. Finally, in Section 4, we present our conclusions.

2 Parameter selections in SDA

Our primary goal in this section is to solve the minimax problem (1.4) for the four aforementioned regions Ω. A closely related problem to (1.4) is

$$\arg\min_{\gamma>0}\max_{z\in\widehat{\Omega}}|w(z;\gamma)|,$$

where $\widehat{\Omega}\subset\mathbb{C}_+$ (the open right half-plane). The two are basically equivalent in the sense that the solution for one leads the solution to the other. In fact, it is not difficult to see

$$\arg\min_{\gamma>0}\max_{z\in\widehat{\Omega}}|w(z;\gamma)|=-\arg\min_{\gamma<0}\max_{z\in-\widehat{\Omega}}|w(z;\gamma)|,$$

where $-\widehat{\Omega}=\{-z:z\in\widehat{\Omega}\}$. The reason that we use the formulation (1.4) is because that is what happens in solving CARE by SDA.

Our main results are summarized in the following theorem.

Theorem 2.1. *Consider the minimax problem* (1.4).

(a) ([9]) *For interval $\Omega=[a,b]\subset\mathbb{R}$ with $a<b<0$, the optimal solution is given by*

$$\gamma_{\rm opt}=-\sqrt{ab},\quad \max_{z\in\Omega}|w(z;\gamma_{\rm opt})|=\frac{1-\sqrt{b/a}}{1+\sqrt{b/a}}. \tag{2.1}$$

(b) *For disk $\Omega=\{z:|z-c|\le r\}$ with $0>c\in\mathbb{R}$ and $c+r<0$, the optimal solution is given by* (2.1) *with $a=c-r$ and $b=c+r$.*

(c) *For ellipse region*

$$\Omega=\{z=x+y\imath:(x-c)^2/R^2+y^2/r^2\le 1\}, \tag{2.2}$$

where $0>c\in\mathbb{R}$ and $c+R<0$ and $0\le r\le R$, the optimal solution is given by (2.1) *with $a=c-R$ and $b=c+R$.*

(d) ([8]) *For rectangle $\Omega=\{z=x+y\imath:a\le x\le b<0,|y|\le r\}$, the optimal $\gamma_{\rm opt}$ is given by*†

$$\gamma_{\rm opt}=\begin{cases}-\sqrt{b^2+r^2}, & \text{if } r^2\ge b(a-b)/2,\\ -\sqrt{ab-r^2}, & \text{if } r^2< b(a-b)/2,\end{cases} \tag{2.3a}$$

$$\max_{z\in\Omega}|w(z;\gamma_{\rm opt})|^2=\begin{cases}\frac{1-\sqrt{b^2/(b^2+r^2)}}{1+\sqrt{b^2/(b^2+r^2)}}, & \text{if } r^2\ge b(a-b)/2,\\ \frac{1-\sqrt{4(ab-r^2)/(a+b)^2}}{1+\sqrt{4(ab-r^2)/(a+b)^2}}, & \text{if } r^2< b(a-b)/2.\end{cases} \tag{2.3b}$$

---

†After the draft of this paper was completed in April 2017, Prof. N. Truhar of University of Osijek, Croatia visited the second author (June 9-18, 2017) and alerted him that (2.3a) had been obtained by Starke [8, Theorem 4.1] in a different but similar context.

The proof of this theorem spread out in the subsections below. A couple of comments are in order.

- The result (2.1) for the interval $\Omega=[a,b]\subset\mathbb{R}$ is not new. In fact, it is a special case of the more general result, involving elliptic functions, on the product of several bilinear functions like $w(z;\gamma)$ [9]. It is included here for its simplicity. Likely it is the first region any one would think of optimizing over. It is also a corollary of special case of case (d), the rectangle with $r=0$.

- The ellipse region (2.2) has its major axis on the real axis. Naturally, we may ask what happens for an ellipse whose major axis is parallel to the imaginary axis‡:

$$\Omega=\{z=x+y\iota : (x-c)^2/r^2+y^2/R^2\le 1\},$$

  where $0>c\in\mathbb{R}$ and $c+r<0$ and $0<r\le R$. While it doesn't seem to have a closed form solution, an estimate can be readily obtained by embedded this ellipse into the rectangle region

$$\{z=x+y\iota : c-r\le x\le c+r<0, |y|\le R\},$$

  and then the result (2.3) applies.

Before we start proving Theorem 2.1 case-by-case, we note that [2]

$$\max_{z\in\Omega}|w(z;\gamma)|=\max_{z\in\partial\Omega}|w(z;\gamma)|,$$

where $\partial\Omega$ is the boundary of Ω. This is because $w(z;\gamma)$ is analytic in Ω for any given $\gamma<0$.

2.1 Disk

Consider case (b): $\Omega=\{z:|z-c|\le r\}$ with $0>c\in\mathbb{R}$ and $c+r<0$, and let $w(\Omega;\gamma)$ be the image of Ω. It is a disk strictly inside the unit circle [2].

Lemma 2.1. *$w(\Omega;\gamma)\cap\mathbb{R}$ is a diameter of the disk $w(\Omega;\gamma)$.*

Proof. Let $\mathrm{I{\scriptsize M}}(\cdot)$ extract the imaginary part of a complex number. It suffices to show

$$\max_{|z-c|=r}\mathrm{I{\scriptsize M}}(w(z;\gamma))=-\min_{|z-c|=r}\mathrm{I{\scriptsize M}}(w(z;\gamma)).$$

The circle $|z-c|=r$ can be parameterized as $z=c+re^{\iota\theta}$ for $-\pi\le\theta\le\pi$. We have

$$f(\theta):=\mathrm{I{\scriptsize M}}(w(z;\gamma))=\mathrm{I{\scriptsize M}}\left(\frac{(z-\gamma)(\bar z+\gamma)}{|z+\gamma|^2}\right)=\frac{2\gamma r\sin\theta}{(c+\gamma+r\cos\theta)^2+r^2\sin^2\theta},$$

‡Since any region must be fully in $\mathbb{C}_-$, the major axis is not allowed to be on the imaginary axis.

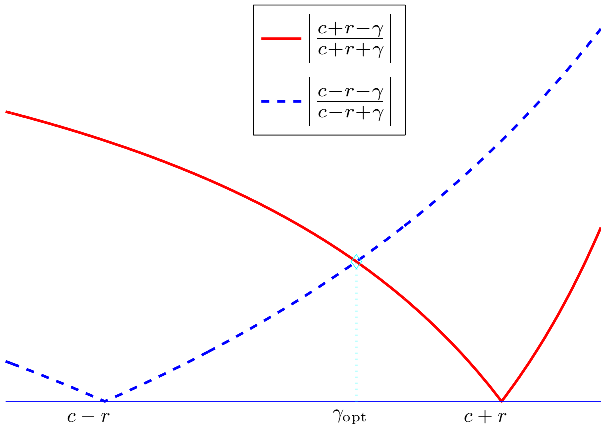

Figure 1: The case of a disk: $\gamma_{\rm opt} = -\sqrt{(c-r)(c+r)}$.

which is an odd function: $f(-\theta) = -f(\theta)$. Therefore

$$\max_{-\pi\le\theta\le\pi} f(\theta) = -\min_{-\pi\le\theta\le\pi} f(\theta),$$

as was to be shown. $\square$

By Lemma 2.1, we conclude that $[w(c-r;\gamma), w(c+r;\gamma)]$ is the diameter of the disk $w(\Omega;\gamma)$ which itself is strictly inside the unit circle, and therefore

$$\max_{|z-c|\le r} |w(z;\gamma)| = \max\{|w(c-r;\gamma)|, |w(c+r;\gamma)|\} \tag{2.4a}$$

$$= \max\left\{\left|\frac{c-r-\gamma}{c-r+\gamma}\right|, \left|\frac{c+r-\gamma}{c+r+\gamma}\right|\right\}. \tag{2.4b}$$

We need to minimize the expression in (2.4b) over $\gamma<0$. To understand how it behaves, we notice that for given $\alpha<0$, the function $(\alpha-\gamma)/(\alpha+\gamma)$ is an increasing function for $\gamma<0$ and takes value 0 at $\gamma=\alpha$ and thus its absolute value $|(\alpha-\gamma)/(\alpha+\gamma)|$, as a function of γ, increases for $0>\gamma>\alpha$ and decreases for $\gamma<\alpha$. This leads to Figure 1. Consequently, the optimal γ is within $(c-r, c+r)$ and satisfies

$$w(c-r;\gamma) = -w(c+r;\gamma) \quad \Rightarrow \quad \frac{c-r-\gamma}{c-r+\gamma} = -\frac{c+r-\gamma}{c+r+\gamma},$$

yielding

$$\gamma_{\rm opt} = -\sqrt{(c-r)(c+r)}. \tag{2.5}$$

Theorem 2.2. *For disk* $\Omega=\{z:|z-c|\le r\}$ *with* $0>c\in\mathbb{R}$ *and* $c+r<0$*, the optimal* $\gamma_{\rm opt}$ *is given by* (2.5).

Corollary 2.1. Let $\widetilde{\Omega}$ be a set of complex numbers such that $\widetilde{\Omega}\subseteq\Omega=\{z:|z-c|\le r\}$ with $0>c\in\mathbb{R}$ and $c+r<0$. If $c\pm r\in\widetilde{\Omega}$, then

$$\arg\min_{\gamma<0}\max_{z\in\widetilde{\Omega}}|w(z;\gamma)|=\arg\min_{\gamma<0}\max_{z\in\Omega}|w(z;\gamma)|=-\sqrt{(c-r)(c+r)}.$$

In particular, the conclusion in case (a): $\Omega=[a,b]\subset\mathbb{R}$ with $a<b<0$ holds.

Proof. We have

$$\max_{z\in\widetilde{\Omega}}|w(z;\gamma)|\le\max_{z\in\Omega}|w(z;\gamma)|=\max\{|w(c-r;\gamma)|,|w(c+r;\gamma)|\}\le\max_{z\in\widetilde{\Omega}}|w(z;\gamma)|,$$

where the first inequality is due to $\widetilde{\Omega}\subseteq\Omega$, the middle equality to (2.4a), and the last inequality to the fact that $c\pm r\in\widetilde{\Omega}$. Therefore,

$$\max_{z\in\widetilde{\Omega}}|w(z;\gamma)|=\max\{|w(c-r;\gamma)|,|w(c+r;\gamma)|\},$$

and thus

$$\arg\min_{\gamma<0}\max_{z\in\widetilde{\Omega}}|w(z;\gamma)|=\arg\min_{\gamma<0}\max\{|w(c-r;\gamma)|,|w(c+r;\gamma)|\}=-\sqrt{(c-r)(c+r)},$$

using the arguments we had for (2.5) from (2.4). $\square$

2.2 Ellipse

Consider the ellipse region (2.2) with $0>c\in\mathbb{R}$ and $c+R<0$ and $0\le r\le R$. It contains in the disk $\widetilde{\Omega}:=\{z:|z-c|\le R\}$ with $c+R<0$, and at the same time $c\pm R\in\widetilde{\Omega}$. A direct application of Corollary 2.1 gives the following theorem:

Theorem 2.3. *For ellipse region* (2.2)*, the optimal* $\gamma_{\rm opt}=-\sqrt{(c-R)(c+R)}$.

2.3 Rectangle

Consider $\Omega=\{z=x+y\iota: a\le x\le b<0, |y|\le r\}$. Necessarily $r\ge 0$. First we will find $\max|w(z;\gamma)|$ over $z\in\partial\Omega$.

Lemma 2.2. *Given* $\alpha<0$ *and* $\gamma<0$*, the function*

$$f(y):=|w(\alpha+y\iota;\gamma)|^2$$

is an even function and increases for $y>0$ *and decreases for* $y<0$*. Consequently,*

$$\max_{|y|\le r}|w(\alpha+y\iota;\gamma)|=|w(\alpha\pm r\iota;\gamma)|.$$

Proof. We have

$$f(y)=\frac{(\alpha-\gamma)^2+y^2}{(\alpha+\gamma)^2+y^2},\quad f'(y)=\frac{8\alpha\gamma y}{[(\alpha+\gamma)^2+y^2]^2}.$$

The conclusions of the lemma are simple consequences by noticing $\alpha\gamma>0$. $\square$

Lemma 2.3. *Given $\beta\neq 0$ and $\gamma<0$, the function*

$$g(x):=|w(x+\beta\iota;\gamma)|^2\equiv|w(x-\beta\iota;\gamma))|^2\quad \textit{for } x<0$$

increases for $0>x>-\sqrt{\gamma^2+\beta^2}$ and decreases for $x<-\sqrt{\gamma^2+\beta^2}$. Consequently,

$$\max_{a\le x\le b}|w(x\pm r\iota;\gamma)|=\max\{|w(a+r\iota;\gamma)|,|w(b+r\iota;\gamma)|\}.$$

Proof. We have

$$g(x)=\frac{(x-\gamma)^2+\beta^2}{(x+\gamma)^2+\beta^2},\quad g'(x)=\frac{4\gamma[x^2-(\gamma^2+\beta^2)]}{[(x+\gamma)^2+\beta^2]^2}.$$

The conclusions of the lemma are simple consequences by noticing $\gamma<0$. $\square$

Lemmas 2.2 and 2.3 together imply that

$$\begin{aligned}\max_{z\in\Omega}|w(z;\gamma)|^2&=\max\{|w(a+r\iota;\gamma)|^2,|w(b+r\iota;\gamma)|^2\}\\&=\max\left\{\frac{(a-\gamma)^2+r^2}{(a+\gamma)^2+r^2},\frac{(b-\gamma)^2+r^2}{(b+\gamma)^2+r^2}\right\}.\end{aligned}\tag{2.6}$$

We have to minimize the expression in (2.6) over $\gamma<0$. For convenience, we define

$$h_\alpha(\gamma)=\frac{(\alpha-\gamma)^2+r^2}{(\alpha+\gamma)^2+r^2}\quad \text{for } \gamma<0.$$

The equation (2.6) becomes

$$\max_{z\in\Omega}|w(z;\gamma)|^2=\max\{h_a(\gamma),h_b(\gamma)\}.$$

Lemma 2.4. *If $ab\le r^2$, then $h_a(\gamma)\le h_b(\gamma)$ for $\gamma<0$. If $ab>r^2$, then*

$$h_a(\gamma)>h_b(\gamma)\quad \textit{for } 0>\gamma>-\sqrt{ab-r^2};\tag{2.7a}$$

$$h_a(\gamma)<h_b(\gamma)\quad \textit{for}\quad \gamma<-\sqrt{ab-r^2}.\tag{2.7b}$$

Proof. It is a consequence of

$$h_a(\gamma)-h_b(\gamma)=\frac{4\gamma(a-b)[(ab-r^2)-\gamma^2]}{[(a+\gamma)^2+r^2][(b+\gamma)^2+r^2]},$$

and $\gamma(a-b)>0$. $\square$

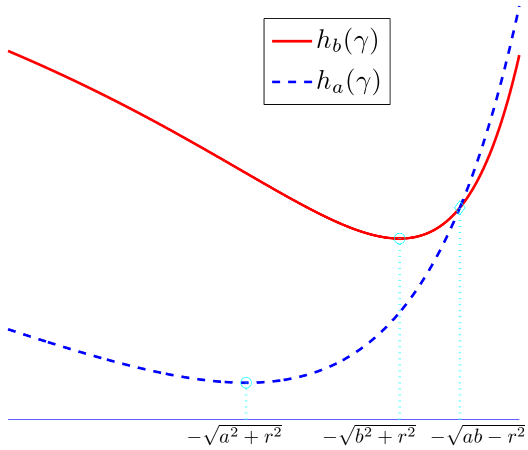

Figure 2: The case of a rectangle for $ab > r^2 > b(a-b)/2$: $\gamma_{\rm opt} = -\sqrt{b^2+r^2}$.

Theorem 2.4. *For rectangular region* $\Omega = \{z = x + y\iota : a \le x \le b < 0, |y| \le r\}$, *the optimal* $\gamma_{\rm opt}$ *is given by* (2.3).

Proof. We will prove the claim according to (1) $r^2 \ge ab$, (2) $a(a-b)/2 \le r^2 < ab$, and (3) $r^2 < a(a-b)/2$.

For the case $r^2 \ge ab$, by Lemma 2.4 we have

$$\arg\min_{\gamma<0}\max_{z\in\Omega}|w(z;\gamma)|^2 = \arg\min_{\gamma<0} h_b(\gamma) = -\sqrt{b^2+r^2},$$

since, similarly to the proof of Lemma 2.3, $h_b(\gamma)$ has one local minimum which is also its global minimum, attained at $\gamma = -\sqrt{b^2+r^2}$.

Consider now $ab > r^2$. We have (2.7). If also $r^2 \ge a(a-b)/2$, then

$$-\sqrt{a^2+r^2} < -\sqrt{b^2+r^2} \le -\sqrt{ab-r^2}.$$

Therefore (for an illustration, see Figure 2)

$$\arg\min_{\gamma<0}\max_{z\in\Omega}|w(z;\gamma)|^2 = \arg\min_{\gamma<-\sqrt{ab-r^2}} h_b(\gamma) = -\sqrt{b^2+r^2}.$$

This proves the case $a(a-b)/2 \le r^2 < ab$. Finally, for the last case $r^2 < a(a-b)/2$,

$$-\sqrt{a^2+r^2} < -\sqrt{ab-r^2} < -\sqrt{b^2+r^2}.$$

The minimum of $\max\{h_a(\gamma), h_b(\gamma)\}$ happens when $h_a(\gamma) = h_b(\gamma)$, giving the optimal $\gamma_{\rm opt}$ as claimed. For an illustration, see Figure 3. $\square$

When $r = 0$, the region $\Omega = \{z = x + y\iota : a \le x \le b < 0, |y| \le r\}$ degenerates to the interval $[a,b]$ with $b < 0$, and Theorem 2.4 yields the well-known result in case (a).

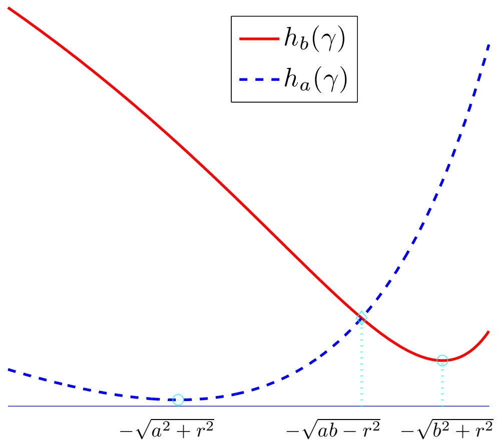

Figure 3: The case of a rectangle for $r^2<b(a-b)/2$: $\gamma_{\rm opt}=-\sqrt{ab-r^2}$.

2.4 An illustrative example

As an illustration, we present a CARE (1.1) from [1]. It arises from a mathematical model of position and velocity control for a string of N high-speed vehicles. The size of the system matrices that define (1.1) is $n=2N-1$, and

$$A=\begin{bmatrix} A_{11} & A_{12} & 0 & \cdots & \cdots & 0 \\ 0 & A_{22} & A_{23} & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & A_{N-2,N-2} & A_{N-2,N-1} & 0 \\ \vdots & & & 0 & A_{N-1,N-1} & \begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ 0 & \cdots & \cdots & \cdots & 0 & -1 \end{bmatrix} \in \mathbb{R}^{n\times n},$$

$G=\operatorname{diag}(1,0,1,0,\cdots,1,0,1)$, and $H=\operatorname{diag}(0,10,0,10,\cdots,0,10,0)$, where

$$A_{ii}=\begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{for } 1\le i\le N-1, \quad A_{ii+1}=\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \quad \text{for } 1\le i\le N-2.$$

To solve CARE (1.1), SDA starts by selecting a parameter $\gamma<0$ as we mentioned in Section 1, and then sets initially

$$E_0=I+2\gamma S^{-\mathrm{T}},\ X_0=2\gamma S^{-1}HA_{+\gamma}^{-1},\ Y_0=-2\gamma A_{+\gamma}^{-1}GS^{-1},$$

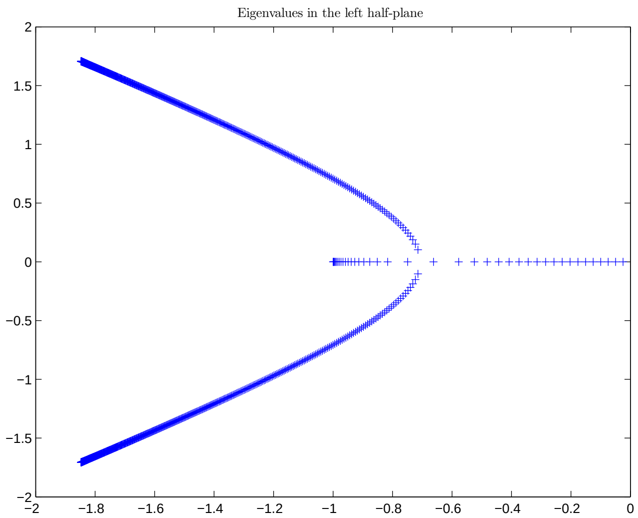

Figure 4: Eigenvalues of $\mathscr{H}$ for $N=400$ lying in $\mathbb{C}_-$.

where $A_{+\gamma}=A+\gamma I_n$, and $S=-A_{+\gamma}^{\mathrm{T}}-HA_{+\gamma}^{-1}G$, and finally for $i=0,1,2,\ldots$ iterates

$$E_{i+1}=E_i(I_m-Y_iX_i)^{-1}E_i, \tag{2.8a}$$

$$X_{i+1}=X_i+E_i^{\mathrm{T}}(I_n-X_iY_i)^{-1}X_iE_i, \tag{2.8b}$$

$$Y_{i+1}=Y_i+E_i(I_n-Y_iX_i)^{-1}Y_iE_i^{\mathrm{T}}, \tag{2.8c}$$

until convergence. At convergence, X_i goes to the desired solution Φ of (1.1) and Y_i goes to a solution of the so-called *dual equation* of (1.1):

$$YA^{\mathrm{T}}+AY-G+YHY=0.$$

Since we are primarily concerned with selecting a good and effective parameter γ, we will not elaborate on any stopping criteria either but simply look at how the normalized residual

$$\mathrm{NRes}_i=\frac{\|A^{\mathrm{T}}X_i+X_iA-X_iGX_i+H\|_{\mathrm{F}}}{\|X_i\|_{\mathrm{F}}(2\|A\|_2+\|X_i\|_2\|G\|_2)+\|H\|_{\mathrm{F}}} \tag{2.9}$$

will behave, where $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ are the matrix spectral norm and Frobenius norm, respectively [5]. But since the matrix spectral norm is not easy to compute, in our tests we substitute it by the ℓ_1-matrix norm $\|\cdot\|_1$.

Consider $N=400$, and thus $n=2N-1=799$. For illustrating purpose, we computed these eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$ and plot them in Figure 4. They are contained in the rectangle $\Omega=\{z=x+y\iota : a\le x\le b<0, |y|\le r\}$ with

$$a=-1.85, \quad b=-0.024, \quad r=1.71. \tag{2.10}$$

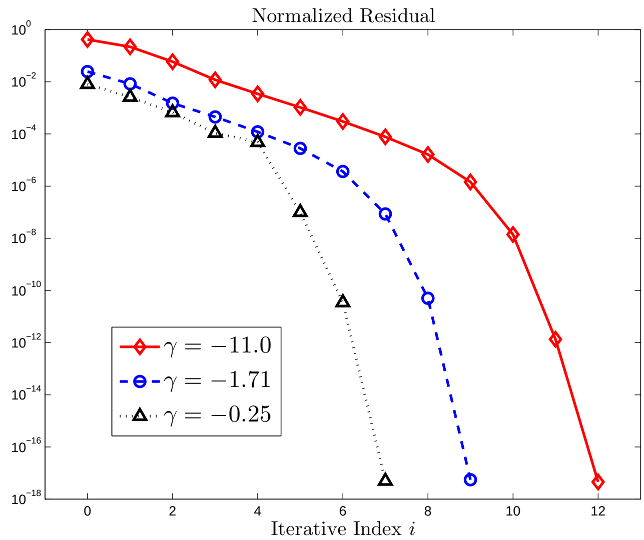

Figure 5: Normalized residuals for $\gamma = -0.25$, -1.71, and -11.0, respectively.

Thus $r^2 = 2.92 > a(a-b)/2 = 1.689$. By Theorem 2.1, the optimal $\gamma = -\sqrt{b^2+r^2} = -1.71$. In Figure 5, we plot NRes_i for three different γ: a more or less randomly selected $\gamma = -11.0$, the optimal $\gamma = -1.71$ subject to knowing that the eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$ are contained in the rectangle Ω, and $\gamma = -0.25$ to which we will return momentarily. The effectiveness of using $\gamma = -1.71$ is evident over using $\gamma = -11.0$, a saving of three doubling iterative steps (2.8).

Conceivably, the more detailed information about the eigenvalues in $\mathbb{C}_-$ we know, the better γ we will be able to find. As an example, suppose that we know, more detailedly, the eigenvalues of $\mathscr{H}$ lying in $\mathbb{C}_-$ are contained in

$$\widehat{\Omega} = \{z = x + y\iota : a \le x \le c < 0, |y| \le r\} \cup [c,b]$$

with a, b, and r are as in (2.10), and $c = -0.7$. We infer from our analysis in Subsections 2.1 – 2.3 that, with the new information, the optimal parameter as defined by (1.4) with Ω replaced by $\widehat{\Omega}$ is given by

$$\gamma_{\rm opt} = \arg\min_{\gamma<0} \max_{1\le j\le 3} h_j(\gamma),$$

where $h_1(\gamma) = |w(a+r\iota;\gamma)|$, $h_2(\gamma) = |w(c+r\iota;\gamma)|$, and $h_3(\gamma) = |w(b;\gamma)|$. This $\gamma_{\rm opt}$ can be computed by setting $h_2(\gamma) = h_3(\gamma)$, giving $\gamma_{\rm opt} = -0.25$ (see Figure 6). This yet another suboptimal γ for the example should be better than the previous $\gamma = -1.71$. Indeed, it saves two doubling iterative steps (2.8) than the previous $\gamma = -1.71$, as shown in Figure 5.

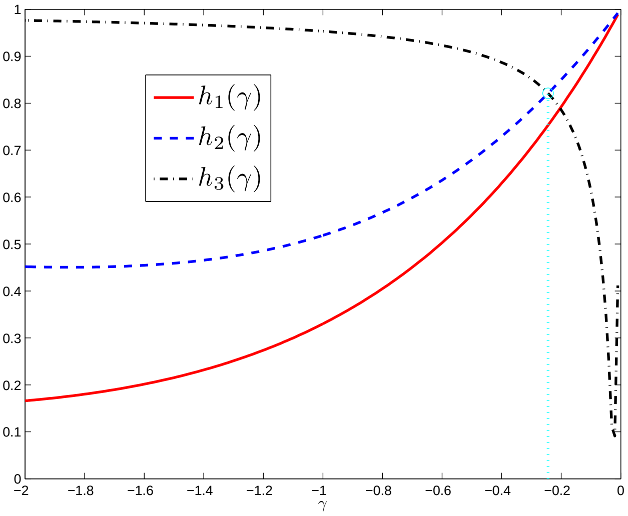

Figure 6: Improved "optimal" $\gamma = -.25$ from more detailed knowledge on eigenvalue containments.

3 Parameter selections in ADDA

The alternating directional doubling algorithm (ADDA) [10], was originally proposed to solve MARE

$$XDX - AX - XB + C = 0, \tag{3.1a}$$

by which we mean

$$W = \begin{array}{c c} & \begin{array}{cc} m & n \end{array} \\ \begin{array}{c} m \\ n \end{array} & \begin{bmatrix} B & -D \\ -C & A \end{bmatrix} \end{array} \tag{3.1b}$$

is a nonsingular M-matrix or an irreducible singular M-matrix. Let $\mathscr{H} = \mathrm{diag}(I_m, -I_n)W$. The equation (3.1a) is equivalent to

$$\mathscr{H} \begin{bmatrix} I_m \\ X \end{bmatrix} \equiv \begin{bmatrix} B & -D \\ C & -A \end{bmatrix} \begin{bmatrix} I_m \\ X \end{bmatrix} = \begin{bmatrix} I_m \\ X \end{bmatrix} (B - DX),$$

which leads to

$$\mathscr{H} \begin{bmatrix} I_m & 0 \\ X & I_n \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ X & I_n \end{bmatrix} \begin{bmatrix} B - DX & -D \\ 0 & -(A - XD) \end{bmatrix}.$$

In particular, this implies that every solution of (3.1a) decouples the eigenvalue problem for $\mathscr{H}$ into two smaller eigenvalue problems for $B - DX$ and $-(A - XD)$.

Define the following 2-parameter-dependent transformation

$$(z_1, z_2) \in \mathbb{C} \times \mathbb{C} \to w((z_1, z_2); (\gamma_1, \gamma_2)) = \frac{z_2 - \gamma_2}{z_2 + \gamma_1} \cdot \frac{z_1 - \gamma_1}{z_1 + \gamma_2}. \tag{3.2}$$

In ADDA, we seek γ_1 and γ_2 to suppress the eigenvalues $B-DX$ and $A-XD$ such that

$$\rho := \max_{\substack{z_1\in\mathrm{eig}(A-XD)\\ z_2\in\mathrm{eig}(B-DX)}} |w((z_1,z_2);(\gamma_1,\gamma_2))| < 1 \tag{3.3}$$

to ensure fast convergence of doubling iterations later on. Generally speaking, ADDA works for any algebraic Riccati equation (3.1a) that may not necessarily be an MARE, so long as (3.1a) has a solution X for which there are parameters γ_1 and γ_2 satisfying (3.3). When such parameters γ_1 and γ_2 are available, ADDA will compute the solution X. An outline of ADDA goes as follows [10].

1. Initialize [11]

$$\begin{array}{cc} & \begin{array}{cc} m & n \end{array} \\ \begin{array}{c} m \\ n \end{array} & \begin{bmatrix} E_0 & Y_0 \\ X_0 & F_0 \end{bmatrix} \end{array} = \begin{bmatrix} \gamma_1^{-1}B+I_m & -\gamma_2^{-1}D \\ -\gamma_1^{-1}C & \gamma_2^{-1}A+I_n \end{bmatrix}^{-1} \begin{bmatrix} I_m-\gamma_2^{-1}B & \gamma_1^{-1}D \\ \gamma_2^{-1}C & I_n-\gamma_1^{-1}A \end{bmatrix}.$$

2. Iterate for $i=0,1\cdots$

$$E_{i+1} = E_i(I_m - Y_iX_i)^{-1}E_i, \tag{3.4a}$$
$$F_{i+1} = F_i(I_n - X_iY_i)^{-1}F_i, \tag{3.4b}$$
$$X_{i+1} = X_i + F_i(I_n - X_iY_i)^{-1}X_iE_i, \tag{3.4c}$$
$$Y_{i+1} = Y_i + E_i(I_m - Y_iX_i)^{-1}Y_iF_i, \tag{3.4d}$$

until convergence.

Aside for possible breakdowns due to the inversions in (3.4), X_i will converge to the particular solution X that makes (3.3) true, and Y_i converges, too, to a solution of the so-called *dual equation* of (3.1a):

$$D - YA - BY + YCY = 0.$$

Moreover, $\|X_i - X\|$ goes to 0 as fast as $O(\rho^{2^i})$. But when (3.1a) is indeed an MARE, it is proved in [10], among many others, that if

$$\gamma_1 \ge \max_i A_{(i,i)}, \quad \gamma_2 \ge \max_j B_{(j,j)}, \tag{3.5}$$

then no breakdown is possible, (3.3) holds, and both X_i and Y_i are nonnegative and monotonically convergent. Usually, the equalities in (3.5) are taken because then ρ is the smallest subject to (3.5) [10].

In view of these discussions, analogously to (1.4), now we need to solve

$$(\gamma_{1;\mathrm{opt}}, \gamma_{2;\mathrm{opt}}) = \arg\min_{\gamma_1>0,\gamma_2>0} \max_{z_1\in\Omega_1, z_2\in\Omega_2} |w((z_1,z_2);(\gamma_1,\gamma_2))|, \tag{3.6}$$

where Ω_i for $i=1,2$ are bounded connected regions. In the case of MARE, both Ω_i should be confined to $\bar{\mathbb{C}}_+$ (the closed right half-plane) and one of them to $\mathbb{C}_+$. This is what we will assume for the rest of this section.

When $\Omega_1=\Omega_2=:\Omega$, we may take $\gamma_1=\gamma_2=:\gamma$ and then, instead of (3.6), we look for

$$\arg\min_{\gamma>0}\max_{z\in\Omega}|w(z;\gamma)|^2=-\arg\min_{\gamma<0}\max_{z\in-\Omega}|w(z;\gamma)|^2,$$

where $w(z;\gamma)$ is as defined in (1.3). The latter is precisely what we investigated in Section 2.

We must point out that having the inequalities in (3.5) will preserve entrywise nonnegativity among all iterative qualities E_i, F_i, X_i, and Y_i. One important outcome of the preservation is the ability to achieve high entrywise relative accuracy [11]. This is remarkable and one should enforce (3.5) with equality when high entrywise relative accuracy is necessary for the underlying application.

In what follows we are targeting at those applications where high relative accuracy in extremely tiny entries is not that important or the magnitudes of the solution entries do not have very wide variations. For such cases, optimal parameters defined by (3.6), or even suboptimal ones, can lead to much faster doubling iterations than those optimal ones suggested in [10] by taking equalities in (3.5).

In the rest of this section, we will consider the case when each of Ω_1 and Ω_2 is one of the following three shapes:

(a) an interval

(b) a disk whose intersection with the real axis is its diameter, and

(c) an ellipse whose intersection with the real axis is its major axis,

such that both Ω_i are confined to $\bar{\mathbb{C}}_+$ and one of them to $\mathbb{C}_+$. Notably missing here, compared to the list in Section 1, is the rectangle shape for which we don't have a good way to deal with for the moment and we will explain why later.

Our key technique to deal with different Ω_i is due to W. B. Jordan [9, p.27] who devised two correlated linear fractional transformations that transform two intervals into one. Specifically, given $\Omega_i=[a_i,b_i]$ with $0\le a_i<b_i$ and $a_1+a_2>0$, Jordan seeks constants α, β, η, and δ such that

$$z_1\in\Omega_1\to\hat{z}_1=\frac{-\delta z_1+\beta}{\eta z_1-\alpha}\in\hat{\Omega}_1,\quad z_2\in\Omega_2\to\hat{z}_2=\frac{\delta z_2+\beta}{\eta z_2+\alpha}\in\hat{\Omega}_2 \tag{3.7}$$

map, respectively, Ω_1 and Ω_2 one-one and onto the same interval $\hat{\Omega}_1=\hat{\Omega}_2=[\xi,1]$ with $0<\xi<1$. Detailed derivations of α, β, η, and δ can be found in [9, pp.26–30], and they are as follows:

$$\alpha=b_1\sigma-a_1(1+\xi),\quad \beta=a_1(1+\xi)-b_1\sigma\xi, \tag{3.8a}$$

$$\eta=\sigma-(1+\xi),\qquad \delta=1+\xi-\sigma\xi, \tag{3.8b}$$

where

$$\mu=\frac{2(b_1-a_1)(b_2-a_2)}{(a_1+a_2)(b_1+b_2)},\ \xi=\frac{1}{1+\mu+\sqrt{\mu(\mu+2)}},\ \sigma=\frac{2(a_1+b_2)}{b_1+b_2}. \tag{3.8c}$$

It follows from (3.7) that

$$z_1=\frac{\alpha\hat{z}_1+\beta}{\eta\hat{z}_1+\delta},\quad z_2=\frac{\alpha\hat{z}_2-\beta}{-\eta\hat{z}_2+\delta},$$

and then

$$w((z_1,z_2);(\gamma_1,\gamma_2))=\frac{\hat{z}_1-\hat{\gamma}_1}{\hat{z}_1+\hat{\gamma}_2}\cdot\frac{\hat{z}_2-\hat{\gamma}_2}{\hat{z}_2+\hat{\gamma}_1}=:w((\hat{z}_1,\hat{z}_2);(\hat{\gamma}_1,\hat{\gamma}_2)),$$

where

$$\hat{\gamma}_1=\frac{\delta\gamma_1-\beta}{-\eta\gamma_1+\alpha},\ \hat{\gamma}_2=\frac{\delta\gamma_2+\beta}{\eta\gamma_2+\alpha}\quad\Rightarrow\quad\gamma_1=\frac{\alpha\hat{\gamma}_1+\beta}{\eta\hat{\gamma}_1+\delta},\ \gamma_2=\frac{\alpha\hat{\gamma}_2-\beta}{-\eta\hat{\gamma}_2+\delta}. \tag{3.9}$$

The minimax problem (3.6) becomes

$$\arg\min_{\hat{\gamma}_1>0,\hat{\gamma}_2>0}\max_{\hat{z}_1,\hat{z}_2\in\hat{\Omega}_1}|w((\hat{z}_1,\hat{z}_2);(\hat{\gamma}_1,\hat{\gamma}_2))| \tag{3.10}$$

which, upon considering $\hat{\gamma}_1=\hat{\gamma}_2>0$ only, is simplified to

$$\arg\min_{\hat{\gamma}_1>0}\max_{z_1\in\hat{\Omega}_1}|w(\hat{z}_1;\hat{\gamma}_1))|^2=\sqrt{\xi}. \tag{3.11}$$

Now with $\hat{\gamma}_i=\sqrt{\xi}$, γ_i for $i=1,2$ are readily computed by the formulas in (3.9). We point out that such a (γ_1,γ_2) may not necessarily solve (3.6) exactly for the case because forcing $\hat{\gamma}_1=\hat{\gamma}_2$ in (3.10) translates into forcing a relation between γ_1 and γ_2 in (3.6), while they are supposed to be independent in the first place. Having said that, we believe (γ_1,γ_2) so constructed via (3.9) and (3.11) should be a very good approximation to the optimal pair that does solve (3.6).

Next, consider two disks $\Omega_i=\{z:|z-c_i|\le r_i\}$ with $c_i-r_i\ge 0$ for $i=1,2$ and $\sum_{i=1}^2(c_i-r_i)>0$. Set

$$a_i=c_i-r_i,\quad b_i=c_i+r_i. \tag{3.12}$$

With them, we again can have the transformations (3.7) with involved constants given by (3.8a) and (3.8b). They maps circles to circles, i.e., $\hat{\Omega}_i$ are disks. Similarly to Lemma 2.1, we can prove that $\hat{\Omega}_i\cap\mathbb{R}=[\xi,1]$ is a diameter of $\hat{\Omega}_i$, and thus $\hat{\Omega}_1=\hat{\Omega}_2$. Therefore again the problem (3.6) becomes (3.10) in form and its simplified version (3.11) remains valid.

For the case where one of Ω_i is a disk and the other is an interval, say $\Omega_1=\{z:|z-c_1|\le r_1\}$ with $c_1-r_1\ge 0$ and $\Omega_2=[a_2,b_2]$ with $a_2\ge 0$ and $(c_1-r_1)+a_2>0$. Set a_1 and b_1 as in (3.12). Again set up the transformations (3.7) with involved constants given by (3.8a) and (3.8b). We know $\hat{\Omega}_1$ is a disk and $\hat{\Omega}_1\cap\mathbb{R}=[\xi,1]$ is its diameter, and also $\hat{\Omega}_2=[\xi,1]$. The problem (3.6) becomes

$$\arg\min_{\hat{\gamma}_1>0,\hat{\gamma}_2>0}\max_{\hat{z}_1\in\hat{\Omega}_1,\hat{z}_2\in\hat{\Omega}_2}|w((\hat{z}_1,\hat{z}_2);(\hat{\gamma}_1,\hat{\gamma}_2))| \tag{3.13}$$

which, upon considering $\hat\gamma_1=\hat\gamma_2>0$ only, is simplified to

$$\begin{aligned}&\operatorname*{argmin}_{\hat\gamma_1>0}\left(\max_{\hat z_1\in\hat\Omega_1}|w(\hat z_1;\hat\gamma_1))|\times\max_{\hat z_2\in\hat\Omega_2}|w(\hat z_2;\hat\gamma_1))|\right)\\&\quad=\operatorname*{argmin}_{\hat\gamma_1>0}\max_{\hat z_1\in\hat\Omega_1}|w(\hat z_1;\hat\gamma_1))|^2=\sqrt{\xi},\end{aligned}\tag{3.14}$$

where the equality in (3.14) is due to, similarly, the arguments that lead to (2.4).

A similar treatment works for the cases of ellipses. Now it is also a perfect time to explain why the above treatment does not work for the rectangle case: the images of rectangles under the transformations (3.7) are no longer rectangles.

We summarize the results into the following theorem. The case when both Ω_i are intervals, as indicated above, has already been dealt with in [9, pp.26–30]. All other cases seem to be new.

Theorem 3.1. *Let each of Ω_i for $i=1,2$ be one of*

(a) *intervals $[a_i,b_i]$,*

(b) *disks $\{z:|z-c_i|\le r\}$ (and for the case $a_i=c_i-r_i$, $b_i=c_i+r_i$),*

(c) *ellipses $\{z=x+y\iota:(x-c_i)^2/R_i^2+y^2/r_i^2\le 1\}$ with $0\le r_i\le R_i$ (for the case $a_i=c_i-R_i$, $b_i=c_i+R_i$).*

Assume $a_i\ge 0$ for $i=1,2$ and $a_1+a_2>0$, and define μ, ξ, σ, α, β, η, and δ as in (3.8). *A suboptimal solution to the problem* (3.6) *is given by*

$$\hat\gamma=\sqrt{\xi},\ \ \gamma_1=\frac{\alpha\hat\gamma+\beta}{\eta\hat\gamma+\delta},\ \ \gamma_2=\frac{\alpha\hat\gamma-\beta}{-\eta\hat\gamma+\delta}$$

for which

$$\max_{z_1\in\Omega_1,z_2\in\Omega_2}|w((z_1,z_2);(\gamma_1,\gamma_2))|=\left(\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}\right)^2.$$

3.1 An illustrative example

we will use [6, Example 5.1] (see also [11, Examples 4.1 and 6.4]) to numerically illustrate the superiority of suboptimally selected parameters as far as the speed of convergence is concerned. In this example, $m=n=3$, and

$$A=\begin{bmatrix}4&0&0\\0&15+\delta&-5\\0&-5&15\end{bmatrix},\quad B=\frac{1}{1.001}\begin{bmatrix}15&-5&0\\-5&15&0\\0&0&5\end{bmatrix},$$
$$C=\begin{bmatrix}0&0&4\\5&5&\delta\\5&5&0\end{bmatrix},\quad D=\frac{1}{1.001}\begin{bmatrix}0&5&5\\0&5&5\\4&1&0\end{bmatrix},$$

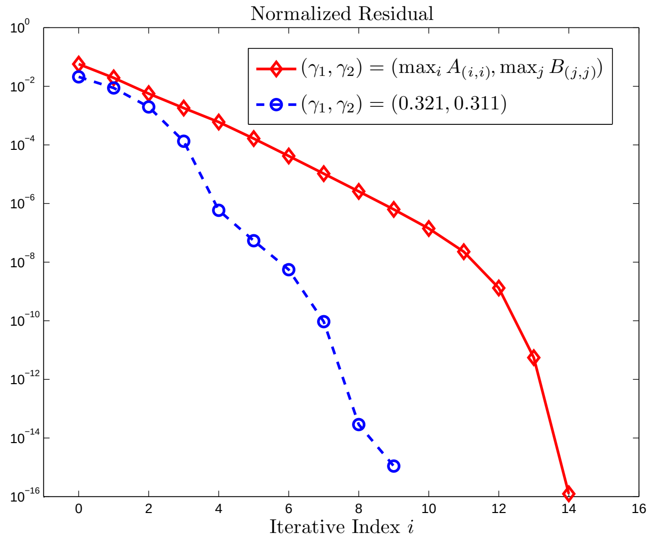

Figure 7: Normalized residuals for two sets of parameters, respectively.

where $\delta = 10^{-8}$. We will test two sets of parameters. The first set is the one suggested in [10], i.e., taking equalities in (3.5). The second set is computed, according to Theorem 3.1, to get $\gamma_1 = 0.321$ and $\gamma_2 = 0.311$, assuming we somehow obtained information on $\mathrm{eig}(A - XD)$ and $\mathrm{eig}(B - DX)$ for the solution X of interest, i.e., $\mathrm{eig}(A - XD) \subset [10^{-2}, 20]$ and $\mathrm{eig}(B - DX) \subset [0, 20]$. In Figure 7, we plot NRes_i (similarly defined to (2.9)) for the two different sets. We see a saving of five doubling iterative steps (3.4). We shall point out that while we indeed achieve substantial savings in computational work, there are losses. Namely, with the suboptimal parameters according to Theorem 3.1, we have no guarantee that there is no breakdown in excuting doubling iteration (3.4), and we can no longer expect all entries in the computed solution have similar entrywise relative accuracy.

4 Conclusion

In this paper, we have investigated how to select good parameters for use in the doubling algorithms to solve continuous-time algebraic Riccati equations (CARE) and M-matrix algebraic Riccati equations (MARE). In general, finding optimal parameters requires complete eigenvalue information of the associated matrix be known, and that is not practical. But sometimes (possibly through computations), regions that confine relevant eigenvalues can be made available. We obtained optimal parameters subject to the confinements in one of the following regions: intervals, circles, ellipses, and rectangles. Illustrative examples show that these optimal parameters subject to the given constraints can reduce the number of doubling iterative steps. Usually, the more detailed information we know, the more effective parameters can be found.

Acknowledgments

Huang was supported in part by the Ministry of Science and Technology (MOST)(105-2115-M-003-009-MY3), National Center of Theoretical Sciences (NCTS) of Taiwan. Li was supported in part by National Natural Science Foundation of China (DMS-1317330, CCF-1527104, DMS-1719620), and National Natural Science Foundation of China (11428104). Lin was supported in part by the MOST, the NCTS, and the ST Yau Centre at the National Chiao Tung University. Lu was supported in part by National Natural Science Foundation of China (11671105, 11428104).

References

[1] J. Abels, P. Benner. CAREX - a collection of benchmark examples for continuous-time algebraic Riccati equations. SLICOT Working Note 1999-16 (Version 2.0), December 1999. Available online at `slicot.org/working-notes/`.

[2] L. V. Ahlfors. Complex Analysis. McGraw-Hill, New York, 3 ed., 1979.

[3] B. D. O. Anderson. Second-order convergent algorithms for the steady-state Riccati equation. Int. J. Control, 1978, 28(2): 295–306.

[4] E. K.-W. Chu, H.-Y. Fan, W.-W. Lin. A structure-preserving doubling algorithm for continuous-time algebraic Riccati equations. Linear Algebra Appl., 2005, 396: 55–80.

[5] J. Demmel. Applied Numerical Linear Algebra. SIAM, Philadelphia, PA, 1997.

[6] G. T. Nguyen, F. Poloni. Componentwise accurate fluid queue computations using doubling algorithms. Numer. Math., 2015, 130(4): 763–792.

[7] Y. Saad. Iterative Methods for Sparse Linear Systems. SIAM, Philadelphia, 2nd edition, 2003.

[8] G. Starke. Optimal alternating direction implicit parameters for nonsymmetric systems of linear equations. SIAM J. Numer. Anal., 1991, 28(5): 1431–1445.

[9] E. L. Wachspress. The ADI Model Problem. Springer, New York, NY, 2013.

[10] W.-G. Wang, W.-C. Wang, R.-C. Li. Alternating-directional doubling algorithm for M-matrix algebraic Riccati equations. SIAM J. Matrix Anal. Appl., 2012, 33(1): 170–194.

[11] J. Xue, R.-C. Li. Highly accurate doubling algorithms for M-matrix algebraic Riccati equations. Numer. Math., 2017, 135: 733–767.

[12] K. Zhou, J. C. Doyle, K. Glover. Robust and Optimal Control. Prentice Hall, Upper Saddle River, New Jersey, 1995.