Received: 10 July 2019 | Revised: 13 March 2020 | Accepted: 1 April 2020 | Published on: 4 May 2020

DOI: 10.1002/nla.2304

RESEARCH ARTICLE

WILEY

Parameter estimation in the Hermitian and skew-Hermitian splitting method using gradient iterations

Qinmeng Zou[1,2] | **Frédéric Magoulès**[2,3]

[1]School of Science, Beijing University of Posts and Telecommunications, Beijing, China

[2]Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

[3]Faculty of Engineering and Information Technology, University of Pécs, Pécs, Hungary

Correspondence
Frédéric Magoulès, Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France.
Email: frederic.magoules@hotmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR ADOM; Centre National de la Recherche Scientifique, Grant/Award Number: LEFE/INSU

Summary

This article presents enhancement strategies for the Hermitian and skew-Hermitian splitting method based on gradient iterations. The spectral properties are exploited for the parameter estimation, often resulting in a better convergence. In particular, steepest descent with early stopping can generate a rough estimate of the optimal upper bound. This is better than an arbitrary choice since the latter often causes stability problems or slow convergence. In addition, delayed gradient methods are considered as inner solvers for the splitting method. Experiments verify the effectiveness of the proposed estimation strategies and show that delayed gradient methods are competitive with conjugate gradient in low precision.

KEYWORDS

Barzilai-Borwein method, delayed gradient methods, Hermitian and skew-Hermitian splitting, minimal gradient, parameter estimation, steepest descent

1 | INTRODUCTION

We are interested in solving the linear system

$$Ax = b, \quad A \in \mathbb{C}^{N\times N}, \quad b \in \mathbb{C}^N, \tag{1}$$

where A is a non-Hermitian positive definite matrix. Let x_* be the exact solution of (1). It has been observed that splitting methods can be used with success. The traditional alternating direction implicit method[1] has inspired the construction of alternate two-step splittings $A = \mathcal{M}_1 - \mathcal{N}_1$ and $A = \mathcal{M}_2 - \mathcal{N}_2$, and this leads to a method called Hermitian and skew-Hermitian splitting (HSS)[2] in which alternately a shifted Hermitian system and a shifted skew-Hermitian system are solved. HSS has received much attention,[3-9] possibly due to its guaranteed convergence and mathematical beauty.

Let H and S denote the Hermitian and skew-Hermitian parts of A, respectively. Let A^{H} be the conjugate transpose of matrix A. It follows that

$$H = \frac{A + A^{\mathsf{H}}}{2}, \quad S = \frac{A - A^{\mathsf{H}}}{2}.$$

Let I be the identity matrix. In short, the HSS method is defined as follows:

$$\begin{cases} (\gamma I + H)x_{n+\frac{1}{2}} = (\gamma I - S)x_n + b, \\ (\gamma I + S)x_{n+1} = (\gamma I - H)x_{n+\frac{1}{2}} + b, \end{cases} \tag{2}$$

with $\gamma > 0$. It could be regarded as a stationary iterative process

$$x_{n+1} = T(\gamma)x_n + p(\gamma),$$

where x_0 is a given vector. Following the notations of Bai et al.,[2] let us set

$$\mathcal{M}_1(\gamma) = \gamma I + H, \quad \mathcal{N}_1(\gamma) = \gamma I - S, \quad \mathcal{M}_2(\gamma) = \gamma I + S, \quad \mathcal{N}_2(\gamma) = \gamma I - H,$$

where $T(\gamma)$ and $p(\gamma)$ can be expressed as

$$T(\gamma) = \mathcal{M}_2^{-1}(\gamma)\mathcal{N}_2(\gamma)\mathcal{M}_1^{-1}(\gamma)\mathcal{N}_1(\gamma), \quad p(\gamma) = \mathcal{M}_2^{-1}(\gamma)(I + \mathcal{N}_2(\gamma)\mathcal{M}_1^{-1}(\gamma))b.$$

Let $\sigma(\cdot)$ be the spectrum of a matrix and let $\rho(\cdot)$ be the spectral radius. A convergence result for (2) in the non-Hermitian positive definite case was established by Bai et al.,[2] which can be shown as follows:

$$\rho(T(\gamma)) \leq \left\|\mathcal{N}_2(\gamma)\mathcal{M}_1^{-1}(\gamma)\right\| = \max_{\lambda \in \sigma(H)} \frac{|\lambda - \gamma|}{|\lambda + \gamma|} = \omega(\gamma), \tag{3}$$

where $\|\cdot\|$ denotes 2-norm. This shows that the spectral radius of iteration matrix $T(\gamma)$ is less than 1. As a result, HSS has guaranteed convergence and the convergence speed depends only on the Hermitian part H. Let $\lambda_i(\cdot)$ be the ith eigenvalue of a matrix in ascending order. The key observation here is that choosing

$$\gamma = \tilde{\gamma} = \sqrt{\lambda_1(H)\lambda_N(H)} \tag{4}$$

leads to the optimal upper bound

$$\omega(\tilde{\gamma}) = \frac{\sqrt{\lambda_N(H)} - \sqrt{\lambda_1(H)}}{\sqrt{\lambda_N(H)} + \sqrt{\lambda_1(H)}} = \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1}, \tag{5}$$

where $\kappa(\cdot)$ denotes the condition number. It is noteworthy that inequality (5) is similar to the convergence result for conjugate gradient (CG)[10] in terms of A-norm error. As mentioned by Bai et al.,[2] $\tilde{\gamma}$ minimizes indeed the upper bound of $\rho(T(\gamma))$ but not $\rho(T(\gamma))$ itself.

In this article, we exploit spectral properties of gradient iterations for obtaining a rough estimate of $\tilde{\gamma}$ without knowing the extremal eigenvalues of H. In Section 2, we focus on the asymptotic analysis of the steepest descent method. In Section 3, we discuss some strategies for estimating the parameter in HSS based on gradient iterations and give a comparison of delayed gradient methods and the CG method for solving Hermitian positive definite systems in low precision. Numerical results are shown in Section 4 and some concluding remarks are drawn in Section 5.

2 | ASYMPTOTIC ANALYSIS OF STEEPEST DESCENT

In this section, we consider the Hermitian positive definite (HPD) linear system

$$Hx = \hat{b}, \quad H \in \mathbb{C}^{N \times N}, \quad \hat{b} \in \mathbb{C}^N. \tag{6}$$

For $n = 0, 1, \ldots$, the gradient method is of the form

$$x_{n+1} = x_n - \alpha_n g_n, \tag{7}$$

where $g_n = \nabla f(x_n) = Hx_n - \hat{b}$. This gives the updating formula

$$g_{n+1} = g_n - \alpha_n H g_n. \tag{8}$$

The steepest descent (SD) method proposed by Cauchy[11] defines a sequence of steplengths as follows:

$$\alpha_n^{\mathrm{SD}} = \frac{g_n^{\mathsf{H}} g_n}{g_n^{\mathsf{H}} H g_n}, \tag{9}$$

which is the reciprocal of Rayleigh quotient. It minimizes the H-norm error of the system (6) and gives theoretically an optimal result at each step

$$\alpha_n^{\mathrm{SD}} = \arg\min_{\alpha} \|(I - \alpha H)e_n\|_H^2,$$

where $e_n = \hat{x}_* - x_n$. Here, $\hat{x}_*$ denotes the exact solution of (6). This classical method is known to behave badly in practice. The directions tend to asymptotically alternate between two directions resulting in a slow convergence.[12]

The motivation for this article arose during the development of efficient gradient methods.[13] We notice that generally SD converges much slower than CG for HPD systems. However, the spectral properties of the former could be beneficial to parameter estimation. Akaike[12] provided a probability distribution model for the asymptotic analysis of SD, while it appears that standard techniques used in linear algebra are not very helpful in this case. The so-called two-step invariance property led to the work of Nocedal et al.[14] in which further asymptotic results are presented. Let $v_i(\cdot)$ be the eigenvector corresponding to the eigenvalue $\lambda_i(\cdot)$. Relevant properties by Nocedal et al.,[14] which will be exploited in the following text can be briefly described in Lemma 1. Note that a symmetric positive definite real matrix was used by Nocedal et al..[14] Therefore, we extend this result and present a new lemma and its proof in the case of Hermitian positive definite systems.

Lemma 1. *Assume that $\lambda_1(H) < \ldots < \lambda_N(H)$. Assume that $v_1^{\mathsf{H}}(H)g_0 \neq 0$ and $v_N^{\mathsf{H}}(H)g_0 \neq 0$. Consider the gradient method (7) with steplength (9) being used to solve (6), where H is Hermitian positive definite. Then*

$$\lim_{n\to\infty} \alpha_{2n}^{\mathrm{SD}} = \frac{1 + c^2}{\lambda_1(H)(1 + c^2\kappa(H))}, \tag{10}$$

$$\lim_{n\to\infty} \alpha_{2n+1}^{\mathrm{SD}} = \frac{1 + c^2}{\lambda_1(H)(c^2 + \kappa(H))}, \tag{11}$$

and

$$\lim_{n\to\infty} \frac{\|g_{2n+1}\|^2}{\|g_{2n}\|^2} = \frac{c^2(\kappa(H) - 1)^2}{(1 + c^2\kappa(H))^2}, \tag{12}$$

$$\lim_{n\to\infty} \frac{\|g_{2n+2}\|^2}{\|g_{2n+1}\|^2} = \frac{c^2(\kappa(H) - 1)^2}{(c^2 + \kappa(H))^2}, \tag{13}$$

for some constant c.

Proof. Let $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ be the real and imaginary parts, respectively. The coefficients of system (6) have the following form:

$$H = \mathrm{Re}(H) + \imath\mathrm{Im}(H), \quad x = \mathrm{Re}(x) + \imath\mathrm{Im}(x), \quad \hat{b} = \mathrm{Re}(\hat{b}) + \imath\mathrm{Im}(\hat{b}),$$

where $\imath$ denotes the imaginary unit. It is possible to rewrite system (6) into the real equivalent form

$$\tilde{H}\tilde{x} = \begin{pmatrix} \mathrm{Re}(H) & -\mathrm{Im}(H) \\ \mathrm{Im}(H) & \mathrm{Re}(H) \end{pmatrix} \begin{pmatrix} \mathrm{Re}(x) \\ \mathrm{Im}(x) \end{pmatrix} = \begin{pmatrix} \mathrm{Re}(\hat{b}) \\ \mathrm{Im}(\hat{b}) \end{pmatrix} = \tilde{b}. \tag{14}$$

By lemma 3.3 and theorem 5.1 in Nocedal et al.,[14] it is known that results (10) to (13) hold in the real case. To prove the desired result in the Hermitian case, it suffices to show that SD applied to (14) is equivalent to that for (6), namely, they should yield the same sequences of gradient vectors and steplengths. One finds that

$$g_n = (\mathrm{Re}(H) + \imath\mathrm{Im}(H))(\mathrm{Re}(x_n) + \imath\mathrm{Im}(x_n)) - (\mathrm{Re}(\hat{b}) + \imath\mathrm{Im}(\hat{b})) = \varphi_n + \imath\psi_n, \tag{15}$$

where

$$\begin{aligned} \varphi_n &= \mathrm{Re}(H)\mathrm{Re}(x_n) - \mathrm{Im}(H)\mathrm{Im}(x_n) - \mathrm{Re}(\hat{b}), \\ \psi_n &= \mathrm{Re}(H)\mathrm{Im}(x_n) + \mathrm{Im}(H)\mathrm{Re}(x_n) - \mathrm{Im}(\hat{b}). \end{aligned}$$

Assume that the two blocks in $\tilde{x}_n$ are the same as the real and imaginary parts of x_n, respectively. Then, from (14) one obtains that

$$\tilde{g}_n = \tilde{H}\tilde{x}_n - \tilde{b} = \begin{pmatrix} \varphi_n \\ \psi_n \end{pmatrix}. \tag{16}$$

On the other hand, let

$$\tilde{\alpha}_n = \frac{\tilde{g}_n^{\mathsf{T}}\tilde{g}_n}{\tilde{g}_n^{\mathsf{T}}\tilde{H}\tilde{g}_n}.$$

Combining (15) and (16) implies $g_n^{\mathsf{H}}g_n = \tilde{g}_n^{\mathsf{T}}\tilde{g}_n$. Since $\mathrm{Im}(H)^{\mathsf{T}} = -\mathrm{Im}(H)$, it follows that $u^{\mathsf{T}}\mathrm{Im}(H)u = 0$ for all $u \in \mathbb{R}^N$, from which one obtains that

$$\mathrm{Re}(g_n)^{\mathsf{T}}\mathrm{Im}(H)\mathrm{Re}(g_n) = 0, \quad \mathrm{Im}(g_n)^{\mathsf{T}}\mathrm{Im}(H)\mathrm{Im}(g_n) = 0.$$

Hence, the following result holds:

$$\begin{aligned} g_n^{\mathsf{H}}Hg_n &= (\mathrm{Re}(g_n) + \imath\mathrm{Im}(g_n))^{\mathsf{H}}(\mathrm{Re}(H) + \imath\mathrm{Im}(H))(\mathrm{Re}(g_n) + \imath\mathrm{Im}(g_n)) \\ &= \mathrm{Re}(g_n)^{\mathsf{T}}\mathrm{Re}(H)\mathrm{Re}(g_n) + \mathrm{Im}(g_n)^{\mathsf{T}}\mathrm{Re}(H)\mathrm{Im}(g_n) + 2\mathrm{Im}(g_n)^{\mathsf{T}}\mathrm{Im}(H)\mathrm{Re}(g_n). \end{aligned}$$

Along with (14), this implies that $g_n^{\mathsf{H}}Hg_n = \tilde{g}_n^{\mathsf{T}}\tilde{H}\tilde{g}_n$, according to which one finds that $\tilde{\alpha}_n = \alpha_n$ when the two blocks in $\tilde{g}_n$ are equal to the real and imaginary parts of g_n, respectively. Hence, the SD iteration for Hermitian system (6) and that for 2-by-2 real form yield exactly the same sequence of solutions. Since properties (10) to (13) in the real case have been proved by Nocedal et al.,[14] we arrive at the desired conclusion. ∎

Concerning the assumption used in Lemma 1, if there exist repeated eigenvalues, then we can choose the eigenvectors so that the corresponding gradient components vanish.[15] If $v_1^{\mathsf{H}}(H)g_0 = 0$ or $v_N^{\mathsf{H}}(H)g_0 = 0$, then we can use inner eigenvectors and the corresponding eigenvalues in Lemma 1 with no effect on the theoretical results.

It took some time before the spectral properties described by Nocedal et al.[14] were applied for solving linear systems. De Asmundis et al.[16] proposed an auxiliary steplength

$$\alpha_n^{\mathrm{A}} = \left(\frac{1}{\alpha_{n-1}^{\mathrm{SD}}} + \frac{1}{\alpha_n^{\mathrm{SD}}} \right)^{-1}, \tag{17}$$

which could be used for efficient implementations of gradient methods. The major result is a direct consequence of (10) and (11). We state the lemma without proof (see De Asmundis et al.[16] for further discussion).

Lemma 2. *Under the assumptions of Lemma 1, the following result holds:*

$$\lim_{n\to\infty} \alpha_n^{\mathrm{A}} = \frac{1}{\lambda_1(H) + \lambda_N(H)}. \tag{18}$$

Another direction of approach was based on a delicate derivation by Yuan.[17] Let us write $\alpha_n^{\mathrm{RA}} = \left(\alpha_n^{\mathrm{A}}\right)^{-1}$ and

$$\Gamma_n = \frac{1}{\alpha_{n-1}^{\mathrm{SD}}\alpha_n^{\mathrm{SD}}} - \frac{\|g_n\|^2}{\left(\alpha_{n-1}^{\mathrm{SD}}\right)^2\|g_{n-1}\|^2}. \tag{19}$$

Yuan[17] developed a new auxiliary steplength of the form

$$\alpha_n^{\mathrm{Y}} = \frac{2}{\alpha_n^{\mathrm{RA}} + \sqrt{\left(\alpha_n^{\mathrm{RA}}\right)^2 - 4\Gamma_n}}, \tag{20}$$

which leads to some two-dimensional finite termination methods for solving system (6).[17] Let us now introduce an alternative steplength

$$\alpha_n^{\mathrm{Z}} = \frac{2}{\alpha_n^{\mathrm{RA}} - \sqrt{\left(\alpha_n^{\mathrm{RA}}\right)^2 - 4\Gamma_n}}. \tag{21}$$

Let us write $\alpha_n^{\mathrm{RY}} = \left(\alpha_n^{\mathrm{Y}}\right)^{-1}$ and $\alpha_n^{\mathrm{RZ}} = \left(\alpha_n^{\mathrm{Z}}\right)^{-1}$. It follows that

$$\alpha_n^{\mathrm{RY}} + \alpha_n^{\mathrm{RZ}} = \alpha_n^{\mathrm{RA}}, \quad \alpha_n^{\mathrm{RY}}\alpha_n^{\mathrm{RZ}} = \Gamma_n.$$

The spectral properties of (19), (20), and (21) are shown in Lemma 3. Note that the Equations (22) and (23) have appeared in De Asmundis et al.[18] for the real case.

Lemma 3. *Under the assumptions of Lemma 1, the following limits hold:*

$$\lim_{n\to\infty} \Gamma_n = \lambda_1(H)\lambda_N(H), \tag{22}$$

$$\lim_{n\to\infty} \alpha_n^{\mathrm{Y}} = \frac{1}{\lambda_N(H)}, \tag{23}$$

$$\lim_{n\to\infty} \alpha_n^{\mathrm{Z}} = \frac{1}{\lambda_1(H)}. \tag{24}$$

Proof. Combining (10) and (11) implies

$$\lim_{n\to\infty} \alpha_{n-1}^{\mathrm{SD}}\alpha_n^{\mathrm{SD}} = \frac{(1+c^2)^2}{\lambda_1^2(H)(c^2+\kappa(H))(1+c^2\kappa(H))}. \tag{25}$$

Combining (10) with (13), one could deduce that

$$\lim_{n\to\infty} \frac{1}{(\alpha_{2n}^{\mathrm{SD}})^2} \lim_{n\to\infty} \frac{\|g_{2n+1}\|^2}{\|g_{2n}\|^2} = \lim_{n\to\infty} \frac{1}{(\alpha_{2n+1}^{\mathrm{SD}})^2} \lim_{n\to\infty} \frac{\|g_{2n+2}\|^2}{\|g_{2n+1}\|^2},$$

from which one finds

$$\lim_{n\to\infty} \frac{\|g_n\|^2}{\left(\alpha_{n-1}^{\mathrm{SD}}\right)^2\|g_{n-1}\|^2} = \frac{\lambda_1^2(H)c^2(\kappa(H)-1)^2}{(1+c^2)^2}. \tag{26}$$

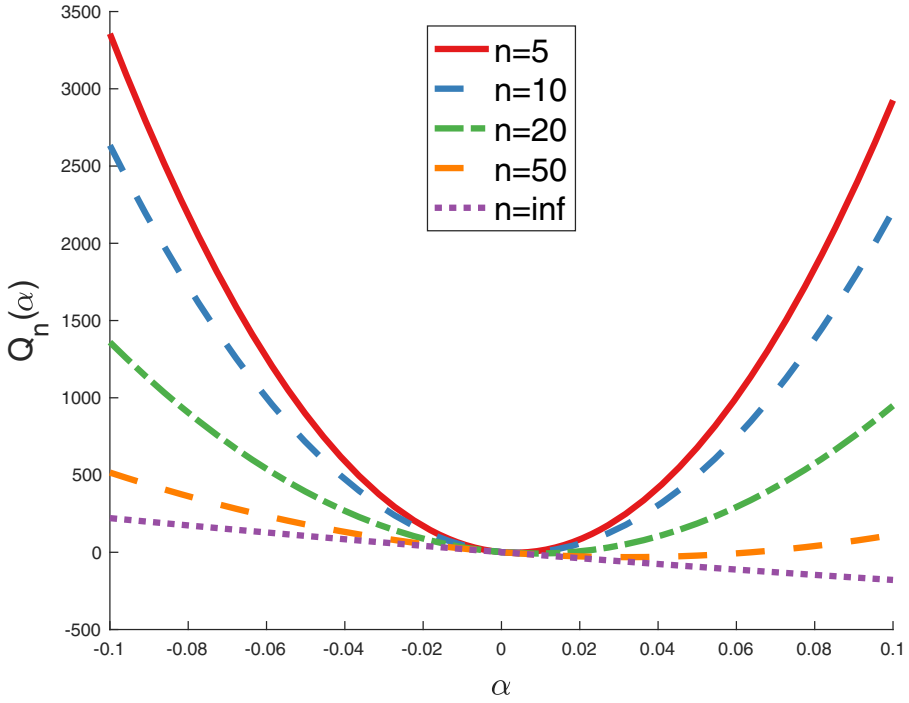

FIGURE 1 Curves of $Q_n(\alpha)$ for a few representative iteration numbers. Steepest descent is used for solving system (6) where H satisfies (28) and $\hat{b}$ is a vector of all ones

The first equation follows by combining (25) and (26). Along with (18), this implies that

$$\lim_{n\to\infty}\left(\left(\alpha_n^{\mathrm{RA}}\right)^2 - 4\Gamma_n\right) = (\lambda_1(H) - \lambda_N(H))^2,$$

which yields the desired limits (23) and (24). ∎

It is noteworthy that steplengths (20) and (21) could be expressed as the roots of a quadratic function

$$Q_n(\alpha) = \Gamma_n\alpha^2 - \alpha_n^{\mathrm{RA}}\alpha + 1, \tag{27}$$

with

$$Q_n(0) = 1, \quad Q_n(\alpha_n^{\mathrm{A}}) = \Gamma_n\left(\alpha_n^{\mathrm{A}}\right)^2,$$

$$Q_n(\alpha_{n-1}^{\mathrm{SD}}) = -\frac{\|g_n\|^2}{\|g_{n-1}\|^2}, \quad Q_n(\alpha_n^{\mathrm{SD}}) = -\frac{\left(\alpha_n^{\mathrm{SD}}\right)^2\|g_n\|^2}{\left(\alpha_{n-1}^{\mathrm{SD}}\right)^2\|g_{n-1}\|^2},$$

and it is known[18] that $\Gamma_n > 0$, leading to the following relation:

$$\alpha_n^{\mathrm{A}} < \alpha_n^{\mathrm{Y}} < \min\{\alpha_{n-1}^{\mathrm{SD}},\ \alpha_n^{\mathrm{SD}}\}.$$

As mentioned by Yuan,[17] a slightly shortened steplength would improve the efficiency of steepest descent. This is one reason why the Yuan steplength could be fruitfully used in alternate gradient methods.[18,19]

As an example, assume that $x_0 = 0$ and

$$H = \mathrm{diag}(1,\ 2,\ 10,\ 20,\ 100,\ 200,\ 1{,}000,\ 2{,}000). \tag{28}$$

Assume that $\hat{b}$ is constructed by $\hat{b} = H\hat{x}_*$ where $\hat{x}_*$ is a vector of all ones. We plot in Figure 1 the curves of (27) for a few representative iteration numbers. This figure shows that the curves of $Q_n(\alpha)$ corresponding to steepest descent converge to the limit, as proved in Lemmas 2 and 3.

3 | APPLICATION TO HSS ITERATIONS

3.1 | Preliminary considerations

In this section, we try to compute estimates for parameter γ in the HSS method. One possible solution is to simply choose $\gamma = 1$ without resorting to special techniques, but experience shows that it often leads to very slow convergence or even divergence, depending on the system being solved. Another approach is based on the observation that γ was introduced to enable the bounded convergence, as seen in (3), and it is possible to express it differently. As an example consider a positive definite diagonal matrix D such that

$$\begin{cases} (D+H)x_{n+\frac{1}{2}} = (D-S)x_n + b, \\ (D+S)x_{n+1} = (D-H)x_{n+\frac{1}{2}} + b. \end{cases} \tag{29}$$

As a result, the iteration matrix is of the form

$$T(D) = (D+S)^{-1}(D-H)(D+H)^{-1}(D-S).$$

Notice that (29) is a special case of preconditioned HSS[20] when choosing $\gamma = 1$ and $P = D$. In particular, the fact that theorem 2.1 in Bertaccini et al.[20] holds for (29) implies $\rho(T(D)) < 1$, yielding the guaranteed convergence. On the basis of similar reasoning as in HSS,[2] the spectral radius is bounded by

$$\rho(T(D)) \leq \left\| (D-H)(D+H)^{-1} \right\|.$$

A natural idea is to seek D so that the upper bound is small. At first glance we may choose D as the diagonal elements of H. Inspired by the diagonal weighted matrix in Freund,[21] the Euclidean norms of column vectors could also be exploited. However, the common experience is that these strategies may lead to a stagnation of convergence and sometimes perform much worse than choosing $\gamma = 1$.

According to (4), the upper bound $\omega(\gamma)$ can be estimated if the extremal eigenvalues are known. However, this is generally not the case in practice, and thus we want to develop fast and cheap algorithms to accelerate the alternate iteration procedure. In addition to this optimal upper bound established in the original article,[2] there are other publications related to this topic. Consider the optimal parameter

$$\gamma_* = \arg\min_\gamma \rho(T(\gamma)).$$

One may expect that a good estimate of γ_* instead of $\tilde{\gamma}$ can result in better performance. Bai et al.[22] investigated this direction in the two-dimensional case and extended their results to certain two-by-two block matrices. Eventually, they proposed a rough estimation of γ_* for the general case, which resorts to finding positive roots of one of the following equations:

$$(\gamma^2+q^2)^2(\gamma^2-\lambda_N^2(H))(\gamma^2-\lambda_1^2(H)) = (\gamma^2-q^2)^2(\gamma^2-\lambda_1(H)\lambda_N(H))^2, \tag{30a}$$

$$(\gamma^2+q^2)^2(\lambda_N^2(H)-\gamma^2)(\gamma^2-\lambda_1^2(H)) = (\gamma^2-q^2)^2(\gamma^2-\lambda_1(H)\lambda_N(H))^2, \tag{30b}$$

where q is the largest magnitude eigenvalue of the skew-Hermitian part S. Huang[23] argued that choosing $(\gamma I - H)(\gamma I - S) \approx 0$ would lead to fast convergence. By virtue of the Frobenius norm $\|\cdot\|_{\mathrm{F}}$, it follows that minimizing $\|(\gamma I - H)(\gamma I - S)\|_{\mathrm{F}}^2$ might be beneficial, implying that

$$4N\gamma^3 + 3c_1\gamma^2 + 2c_2\gamma + c_3 = 0, \quad c_1 = -2\mathrm{tr}(H), \quad c_2 = \mathrm{tr}(H^2) - \mathrm{tr}(S^2), \quad c_3 = 2\mathrm{tr}(HS^2), \tag{31}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. These two approaches shall be further examined in Section 4.3.

3.2 | Parameter estimation based on gradient iterations

It is observed that (22) leads to a straightforward estimation of parameter $\tilde{\gamma}$ in (4). From Lemma 3, we can deduce that the optimal parameter in HSS could be actually approximated by a direct approach using steepest descent iterations, which is shown in the following theorem.

Theorem 1. *Assume that the matrix H in system (6) is the Hermitian part of A in system (1). If steepest descent is used for solving (6), then the following limit holds:*

$$\lim_{n\to\infty} \sqrt{\Gamma_n} = \tilde{\gamma}. \tag{32}$$

Proof. Combining (4), (22), and the fact that $\Gamma_n > 0$ observed from (27), the desired conclusion follows. ∎

Another approach is to compute the approximation by combining Lemmas 2 and 3 through an indirect mechanism, exploiting $\mathcal{M}_1(\gamma)$ instead of H. This approach is shown in Theorem 2.

Theorem 2. *Assume that the matrix H in system (6) is the Hermitian part of A in system (1). If steepest descent is used for solving $\mathcal{M}_1 x = \hat{b}$, then the following limit holds:*

$$\lim_{n\to\infty} \sqrt{\Gamma_n - \gamma\alpha_n^{\mathrm{RA}} + \gamma} = \tilde{\gamma}. \tag{33}$$

Proof. Recall that $\mathcal{M}_1(\gamma) = \gamma I + H$. Since

$$\lambda_i(\mathcal{M}_1(\gamma)) = \gamma + \lambda_i(H)$$

for $i = 1, \ldots, N$, it follows that

$$\begin{aligned}\tilde{\gamma} &= \sqrt{\lambda_1(H)\lambda_N(H)} = \sqrt{(\lambda_1(\mathcal{M}_1(\gamma)) - \gamma)(\lambda_N(\mathcal{M}_1(\gamma)) - \gamma)} \\ &= \sqrt{\lambda_1(\mathcal{M}_1(\gamma))\lambda_N(\mathcal{M}_1(\gamma)) - \gamma(\lambda_1(\mathcal{M}_1(\gamma)) + \lambda_N(\mathcal{M}_1(\gamma))) + \gamma^2}\end{aligned}$$

Combining (18) and (22) implies

$$\begin{aligned}\tilde{\gamma} &= \sqrt{\lim_{n\to\infty}\Gamma_n - \gamma \lim_{n\to\infty}\alpha_n^{\mathrm{RA}} + \gamma^2} \\ &= \lim_{n\to\infty}\sqrt{\Gamma_n - \gamma\alpha_n^{\mathrm{RA}} + \gamma^2}.\end{aligned}$$

This completes out proof. ∎

Remark 1. Practically, obtaining $\tilde{\gamma}$ by (33) requires a predetermined parameter γ. One could choose $\gamma = 1$ and give an integer k as the maximum number of iterations such that

$$\tilde{\gamma} \approx \sqrt{\Gamma_k - \alpha_k^{\mathrm{RA}} + 1},$$

in which case the HSS algorithm might be executed at reduced costs. Notice that there is no need to compute the solution vector x during the estimating process.

Another direction of approach is based on the minimal gradient (MG) steplength

$$\alpha_n^{\mathrm{MG}} = \frac{g_n^{\mathrm{H}} H g_n}{g_n^{\mathrm{H}} H^2 g_n}, \tag{34}$$

the spectral properties of which have been discussed by the present authors along with several new gradient methods in a separate article.[13] We have the following lemma, which is similar to Lemma 1.

Lemma 4. *Assume that $\lambda_1(H) < \cdots < \lambda_N(H)$. Assume that $v_1^{\mathrm{H}}(H)g_0 \neq 0$ and $v_N^{\mathrm{H}}(H)g_0 \neq 0$. Consider the gradient method (7) with steplength (34) being used to solve (6), where H is Hermitian positive definite. Then*

$$\lim_{n\to\infty} \alpha_{2n}^{\mathrm{MG}} = \frac{1+c^2}{\lambda_1(H)(1+c^2\kappa(H))}, \tag{35}$$

$$\lim_{n\to\infty} \alpha_{2n+1}^{\mathrm{MG}} = \frac{1+c^2}{\lambda_1(H)(c^2+\kappa(H))}, \tag{36}$$

and

$$\lim_{n\to\infty} \frac{g_{2n+1}^{\mathrm{H}} H g_{2n+1}}{g_{2n}^{\mathrm{H}} H g_{2n}} = \frac{c^2(\kappa(H)-1)^2}{(1+c^2\kappa(H))^2}, \tag{37}$$

$$\lim_{n\to\infty} \frac{g_{2n+1}^{\mathrm{H}} H g_{2n+1}}{g_{2n+1}^{\mathrm{H}} H g_{2n+1}} = \frac{c^2(\kappa(H)-1)^2}{(c^2+\kappa(H))^2}, \tag{38}$$

for some constant c.

Proof. The real-valued case was proved in our recent article.[13] The rest of the proof is similar to the proof of Lemma 1. ■

Let us write

$$\alpha_n^{\mathrm{A2}} = \left(\frac{1}{\alpha_{n-1}^{\mathrm{MG}}} + \frac{1}{\alpha_n^{\mathrm{MG}}}\right)^{-1}, \quad \tilde{\Gamma}_n = \frac{1}{\alpha_{n-1}^{\mathrm{MG}}\alpha_n^{\mathrm{MG}}} - \frac{g_n^{\mathrm{H}} H g_n}{\left(\alpha_{n-1}^{\mathrm{MG}}\right)^2 g_{n-1}^{\mathrm{H}} H g_{n-1}},$$

which resemble (17) and (19), respectively. Let $\alpha_n^{\mathrm{RA2}} = \left(\alpha_n^{\mathrm{A2}}\right)^{-1}$. Then we have the following result.

Lemma 5. *Under the assumptions of Lemma 4, the following limits hold:*

$$\lim_{n\to\infty} \alpha_n^{\mathrm{A2}} = \frac{1}{\lambda_1(H)+\lambda_N(H)}, \tag{39}$$

$$\lim_{n\to\infty} \tilde{\Gamma}_n = \lambda_1(H)\lambda_N(H). \tag{40}$$

Proof. For details, see our recent article.[13] See also the proof of Lemma 3. ■

The following theorems summarize the parameter estimation strategies based on the MG steplength.

Theorem 3. *Assume that the matrix H in system (6) is the Hermitian part of A in system (1). If minimal gradient is used for solving (6), then the following limit holds:*

$$\lim_{n\to\infty} \sqrt{\tilde{\Gamma}_n} = \tilde{\gamma}.$$

Proof. The proof can be obtained similarly as the one in Theorem 1 by exploiting (40). ■

Theorem 4. *Assume that the matrix H in system (6) is the Hermitian part of A in system (1). If minimal gradient is used for solving $\mathcal{M}_1(\gamma)x = \hat{b}$, then the following limit holds:*

$$\lim_{n\to\infty} \sqrt{\tilde{\Gamma}_n - \gamma\alpha_n^{\mathrm{RA2}} + \gamma} = \tilde{\gamma}. \tag{41}$$

Proof. The proof can be obtained similarly as the one in Theorem 2 by exploiting (39) and (40). ∎

3.3 | Delayed gradient methods as inner solvers

Although steepest descent has remarkable spectral properties, as an iterative method, its popularity has been overshadowed by CG. Akaike[12] exploited the fact that the zigzagging behavior nearly always leads to slow convergence, except when initial gradient approaches an eigenvector. This drawback can be cured with a delayed strategy, first proposed by Barzilai and Borwein,[24] which was later called Barzilai-Borwein (BB) method. The idea is to provide a two-point approximation to the quasi-Newton methods, namely,

$$\alpha_n^{\mathrm{BB}} = \arg_\alpha \min \left\| \frac{1}{\alpha} \Delta x - \Delta g \right\|^2,$$

where $\Delta x = x_n - x_{n-1}$ and $\Delta g = g_n - g_{n-1}$, yielding

$$\alpha_n^{\mathrm{BB}} = \frac{g_{n-1}^{\mathsf{H}} g_{n-1}}{g_{n-1}^{\mathsf{H}} H g_{n-1}}.$$

Notice that for any $n \geq 1$ if the gradient vector g_{n-1} generated by BB iterations is equal to that generated by SD iterations, then $\alpha_n^{\mathrm{BB}} = \alpha_{n-1}^{\mathrm{SD}}$. The convergence analysis was given in Raydan[25] and Dai and Liao.[26] A Q-linear result, however, has never been proved due to its nonmonotone convergence. It seems overall that the effect of this irregular behavior is beneficial.

For the HSS method, two inner iterative procedures are needed at each iteration. Since the solution of subproblems in (2) is sometimes as difficult as that of the original system (1), inexact solvers are often considered with rather low precision, especially for ill-conditioned problems. In practice, the first equation of (2) is usually solved by CG, and the second equation of (2) can be solved by CGNE.[10] Friedlander et al[27] made the observation that BB could often be competitive with CG when low precision is required. It is known that CG is sensitive to rounding errors, while delayed gradient methods can remedy this issue[15,28] with less computational costs per iteration. In addition, although BB sometimes suffers from the disadvantage of requiring increasing number of iterations for increasing condition numbers, its low-precision behavior tends to be less sensitive to the ill-conditioning.

A similar method developed by symmetry[24] is of the form

$$\alpha_n^{\mathrm{BB2}} = \frac{g_{n-1}^{\mathsf{H}} H g_{n-1}}{g_{n-1}^{\mathsf{H}} H^2 g_{n-1}},$$

which imposes as well a quasi-Newton property

$$\alpha_n^{\mathrm{BB2}} = \arg_\alpha \min \|\Delta x - \alpha \Delta g\|^2.$$

Similarly, if there exists $n \geq 1$ such that BB2 and MG share the same g_{n-1}, then we have $\alpha_n^{\mathrm{BB2}} = \alpha_{n-1}^{\mathrm{MG}}$. In the last three decades, much effort was devoted to develop new delayed gradient methods, see De Asmundis et al.[18] and the references therein.

An example is illustrated in Figure 2. We solve (6) with different residual thresholds ε, where H is chosen as a diagonal matrix of size 10^3 and $\hat{b}$ is a vector of all ones. The diagonal entries have values logarithmically distributed between 10^{-3} and 1 in ascending order, with the first and the last entries equal to the limits, respectively, such that $\kappa(H) = 10^3$. The plot shows a fairly efficient behavior of BB.

4 | NUMERICAL EXPERIMENTS

In this section, we perform some numerical tests. Assume that iterative algorithms are started from zero vectors. The global stopping criterion in HSS is determined by the threshold $\varepsilon = \|b - Ax_n\| / \|b\|$ with a fixed convergence

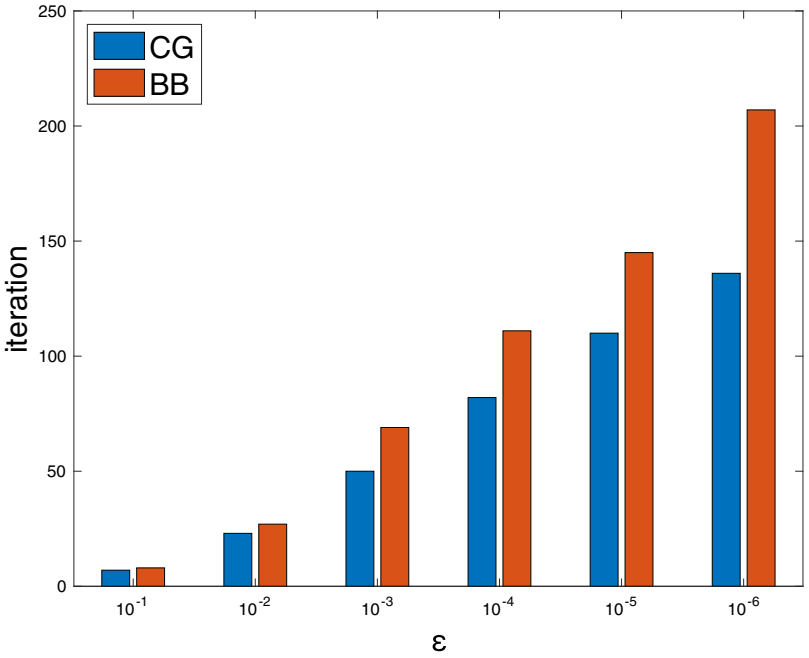

FIGURE 2 Comparison of CG and BB for solving system (6) where H is a diagonal matrix of size 10^3 with $\kappa(H) = 10^3$ and $\hat{b}$ is a vector of all ones

tolerance 10^{-6}. The inner stopping thresholds ε_1 and ε_2 for the two half-steps of (2) are defined in the same way. For gradient iterations applied to system (6), similarly, the stopping criterion is defined by the threshold $\varepsilon = \|\hat{b} - Hx_n\|/\|\hat{b}\|$. All tests are run in double precision on a machine with 2.8 GHz Intel Core i7 CPU and 8 GB memory. The above setting shall be used in Sections 4.1, 4.2, and 4.4. In Section 4.3, we conduct experiments in order to compare our strategies with that proposed in other publications and thus exploit directly the experimental settings therein.

4.1 | Asymptotic results of gradient iterations

The goal of the first experiment is to illustrate how the spectral properties described earlier can be used for providing a rough estimate of parameter $\tilde{\gamma}$. We have implemented steepest descent and minimal gradient iterations for several real matrices of size 1,000 generated by MATLAB routine `sprandsym`. The right-hand side is chosen to be a vector of ones. In Figure 3, parameter γ is plotted versus iteration number, under which a red dotted line marks out the position of $\tilde{\gamma}$. It is clear that γ tends to $\tilde{\gamma}$ asymptotically as expected. As can be seen, steepest descent with limits (32) and (33) turns out to be a better strategy than minimal gradient in all cases. The indirect approximations based on (33) and (41) yield faster convergence for both steepest descent and minimal gradient.

This test confirms Theorems 1–4. Recall that choosing $\tilde{\gamma}$ as parameter only implies an upper bound of $\rho(T(\gamma))$. Experience shows that this choice may sometimes cause overfitting, resulting in slow convergence or even divergence, especially when $\tilde{\gamma}$ is small. One simple measure is to use early stopping in gradient iterations. In the following, it is assumed that steepest descent is used for parameter estimation in HSS, called preadaptive iterations, and we consider only the direct approach (32).

4.2 | HSS with different parameters

In this test, we generate some matrices obtained from a classical problem in order to understand the convergence behavior of HSS enhanced by steepest descent iterations.

Example 1. Consider system (1) where A arises from the discretization of partial differential equation

$$-\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + \theta\left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z}\right) = f \tag{42}$$

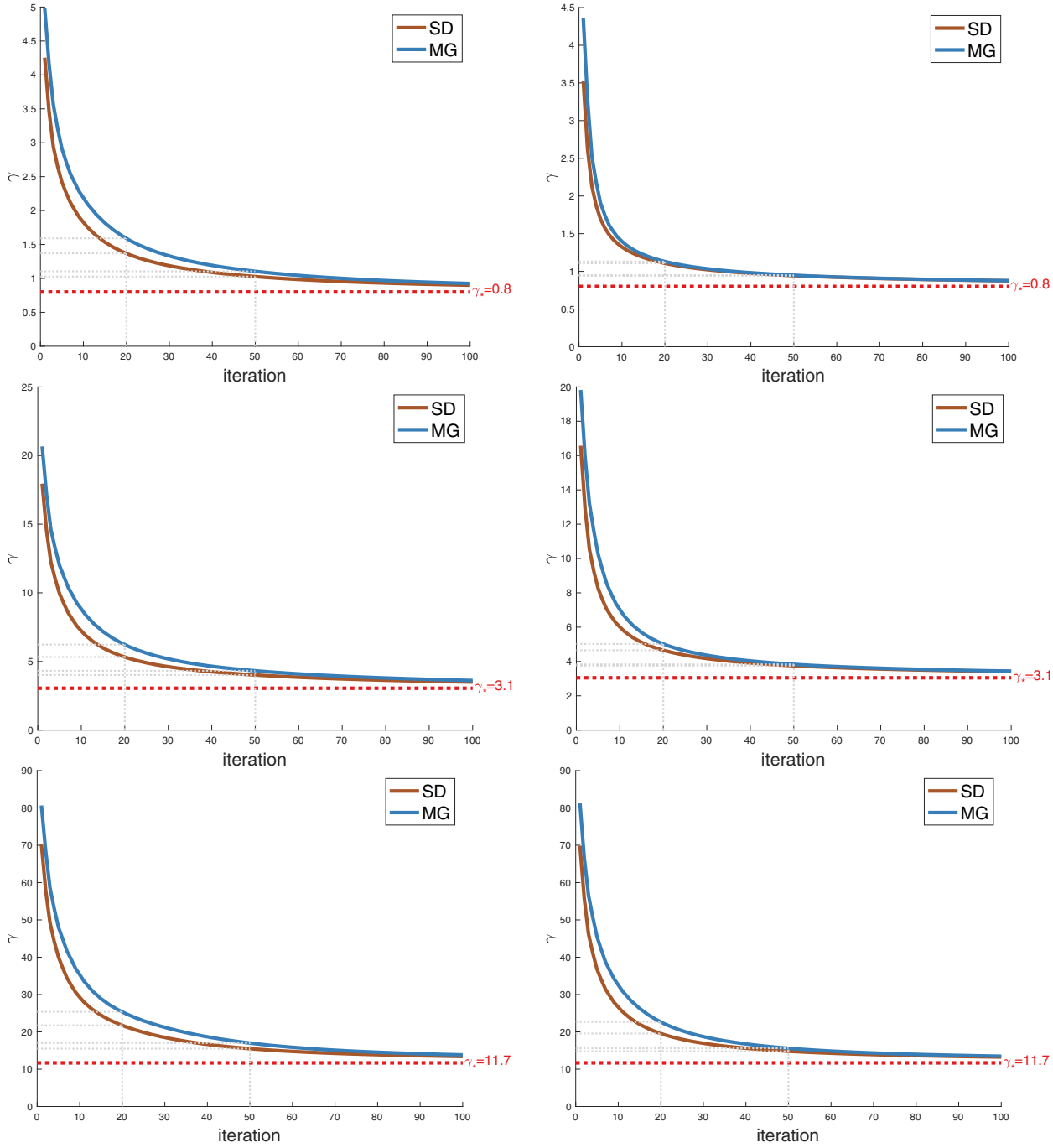

FIGURE 3 Parameter estimation with different matrix H generated randomly by MATLAB: $\tilde{\gamma} = 0.8$ (top), $\tilde{\gamma} = 3.1$ (middle), and $\tilde{\gamma} = 11.7$ (bottom). Parameter γ is computed by two approaches: Theorems 1 and 3 (left) and Theorems 2 and 4 (right)

on the unit cube $\Omega = [0, 1]^3$ with θ a positive constant. In what follows we shall choose $\theta = 1$. Assume that u satisfies homogeneous Dirichlet boundary conditions. The finite difference discretization on a uniform $m \times m \times m$ grid with mesh size $h = 1/(m + 1)$ is applied to the above model yielding a linear system with $N = m^3$.

In what follows we shall use the centered difference scheme for discretization. The right-hand side b is generated with random complex values ranging in $[-10, 10] + \imath[-10, 10]$. As thresholds for inner iterations, $\varepsilon_1 = 10^{-4}$ and $\varepsilon_2 = 10^{-4}$ are chosen. CG is exploited for solving the Hermitian inner system, while CGNE is used for the skew-Hermitian part. Figure 4 shows the convergence behavior of HSS upon different values of the parameter. Here, we set $\gamma \in [0.5, 3.5]$ and $m \in [9, 21]$. The optimal parameters $\tilde{\gamma}$ with $m = 9, 12, 15, 18, 21$ are located by red lines. Notice that a path that zigzags through

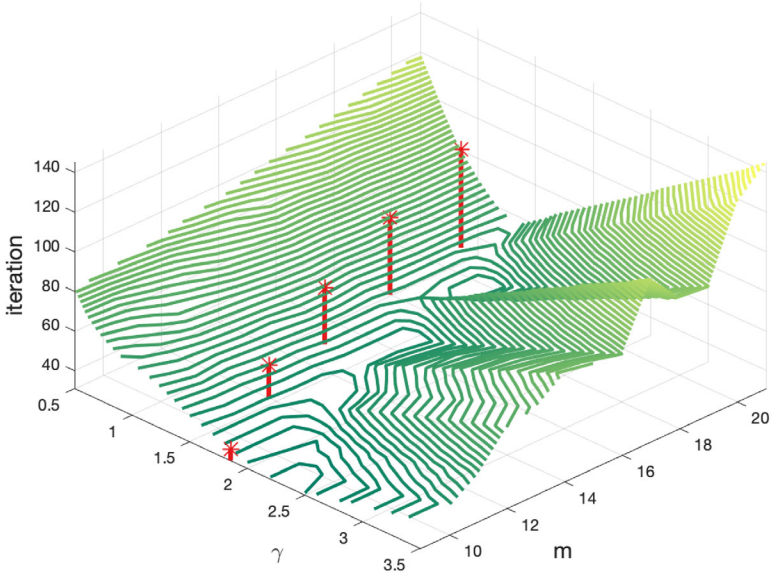

FIGURE 4 Solving problem (42) by HSS with $\gamma \in [0.5,\ 3.5]$ and $m \in [9,\ 21]$. The optimal parameters $\tilde{\gamma}$ are located by red lines

the bottom of the valley corresponds to the optimal parameters. As already noted, the parameter estimates need not be accurate, and thus the red lines are good enough in practice.

Approximating $\tilde{\gamma}$ by inexact steepest descent iterations yields the preadaptive HSS method (PAHSS). Let η denote the number of preadaptive iterations. The convergence behaviors and total computing times are illustrated in Figure 5. The left four plots show the residual curves with several typical choices of η when $m = 80,\ 100,\ 120,\ 140$, namely, $N = 512,000,\ 1,000,000,\ 1,728,000,\ 2,744,000$. Two observations can be made for all dimensions: the first is that larger η yields faster convergence of HSS; the second is that $\eta = 100$ does not lead to significant gains in efficiency compared with $\eta = 50$. The right four plots show total wall-clock times of PAHSS iterations, measured in seconds, upon η ranging from 10 to 160. It can be seen that substantial gains are made in the beginning, following a long period of stagnation. Experience shows that a small number of steepest descent iterations are sufficient, and it is therefore appropriate to use early stopping.

4.3 | Comparison with other techniques

It is noteworthy that there are many other possible choices for solving large-scale eigenvalue problems. The most commonly used techniques are based on projection methods onto Krylov subspaces.[29] For the Hermitian case, the Lanczos method can be employed to generate a real symmetric tridiagonal matrix, which is similar to the original one. This process requires two dot products, two vector updates (SAXPY), and one sparse matrix-vector multiplication (SpMV), while the SD method requires one less SAXPY operation. Furthermore, the Lanczos method requires more effort to determine the eigenvalues from the tridiagonal matrix and find the extremal ones; see Golub and Van Loan[29] for more details. Our experience is that although Krylov subspace methods converge faster than gradient methods, for estimation of extremal eigenvalues, the gradient iteration exploiting spectral properties is much cheaper, generally more than two times faster, than the Lanczos method in terms of total time consumption.

As mentioned in Section 3.1, there are other ideas that have been developed in response to the parameter estimation problem. The following example has appeared in these publications.[22,23]

Example 2. Consider system (1) where A arises from the discretization of partial differential equation

$$-\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \theta\left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\right) = f \tag{43}$$

on the unit square $\Omega = [0,\ 1]^2$ with θ a positive constant. Assume that u satisfies homogeneous Dirichlet boundary conditions. The finite difference discretization on a uniform $m \times m$ grid with mesh size $h = 1/(m+1)$ is applied to the above model yielding a linear system with $N = m^2$.

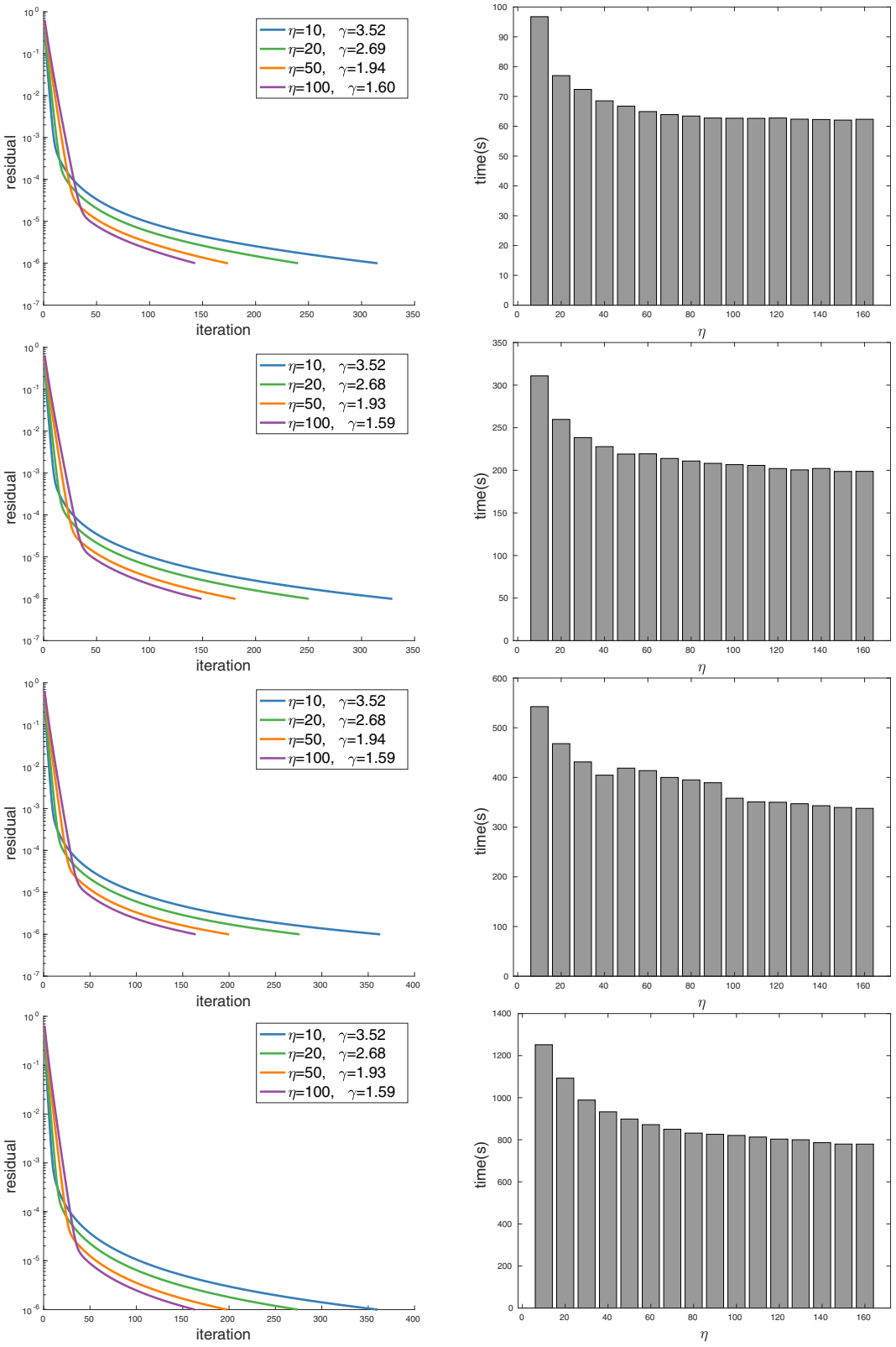

FIGURE 5 Parameter estimation for problem (42) with different mesh densities: $m = 80$ (first), $m = 100$ (second), $m = 120$ (third), $m = 140$ (fourth), namely, $N = 512,000,\ 1,000,000,\ 1,728,000,\ 2,744,000$. Left: convergence curves for different η. Right: average wall-clock times for different η including that of steepest descent iterations

TABLE 1 Test results for problem (43) with $m = 32$

θ	10	50	100	500	1000
γ_1	0.5967	2.7084	5.1536	10.2948	15.0075
$\rho(T(\gamma_1))$	0.8055	0.4582	0.4771	0.6374	0.7179
IT	68	45	46	56	74
Time (s)	0.2821	0.0867	0.1304	0.2624	0.3611
γ_2	0.0180	0.5536	3.2621	3.9358	3.9830
$\rho(T(\gamma_2))$	0.9912	0.8200	0.4790	0.7345	0.8191
IT	> 1000	74	35	49	66
Time (s)	\	0.4659	0.1439	0.4166	0.7239
$\tilde{\gamma}$	0.3802	0.3802	0.3802	0.3802	0.3802
$\rho(T(\tilde{\gamma}))$	0.8312	0.8702	0.8839	0.8999	0.9030
IT	84	106	111	105	99
Time (s)	0.3902	0.8129	1.3905	2.7752	2.7337
γ_{sug}	1.3133	1.3133	1.3133	1.3133	1.3133
$\rho(T(\gamma_{\text{sug}}))$	0.9069	0.6469	0.7027	0.8187	0.8688
IT	148	31	40	60	72
Time (s)	0.4677	0.1103	0.1460	0.8910	1.5130

Note: γ_1 is obtained by (30a) and (30b), γ_2 obtained by (31), $\tilde{\gamma}$ obtained by (4), γ_{sug} the proposed strategy with steepest descent iterations. IT denotes the outer number of iterations.

TABLE 2 Results of different methods for problem (42) with $\epsilon_1 = 10^{-1}$, $\epsilon_2 = 10^{-4}$, and $\gamma = 1$

	HSS-CG				HSS-BB			
N	4,096	32,768	262,144	2,097,152	4,096	32,768	262,144	2,097,152
IT_{total}	380	488	592	614	304	606	857	908
Time (s)	2.63	4.86	31.83	267.48	1.78	4.35	35.69	300.92

We show the experimental results with different θ in Table 1. Here, we choose $m = 32$ and let IT denote the outer number of iterations. γ_1 denotes the optimal parameter obtained by (30a) and (30b), γ_2 obtained by (31), $\tilde{\gamma}$ obtained by (4), γ_{sug} the proposed approach based on steepest descent with $\eta = 50$. We select the right-hand side b such that x_* is a vector of all ones. All algorithms start from the zero vector. This ensures that our experimental setting is the same as the reported results by Bai et al[22] and Huang.[23] We can see from Table 1 that our strategy gives better results compared with $\tilde{\gamma}$ and competitive results compared with γ_2. Although γ_1 tends to generate the best results, it requires an estimate of three unknown parameters, which are seldom available in practice.

4.4 | CG and BB as low-precision inner solvers

In order to verify that BB can be an efficient alternative to CG as low-precision inner solver for HSS, some tests proceed along the same lines as above but consider both CG and BB as inner solvers for the Hermitian part. Here, we use the same experimental setting as in Section 4.2. Numbers of total iterations IT_{total} and wall-clock times, measured in seconds, are shown in Table 2. Since some of the optimal parameters $\tilde{\gamma}$ for (42) with $m = 16, 32, 64, 128$, namely, $N = 4,096,\ 32,768,\ 262,144,\ 2,097,152$, are less than 1, which may result in stability problem, we choose $\gamma = 1$ for all tests. We conduct 10 repeated experiments and print only the average computation times. A comparison of costs is shown in Table 3. As expected, BB is competitive with CG in terms of computation times, especially for small problems, and shows a clear advantage for storage requirements and resistance to perturbation.[15,28]

TABLE 3 Summary of operations for iteration i and storage requirements

Method	Dot products	Vector updates	Matrix-vector	Storage
CG	2	3	1	$4N$
BB	2	2	1	$3N$
CGNE	2	3	2	$3N$
HSS	Solver	Solver+2	Solver+1	Two matrices+N

Note: In the HSS row, the term "solver" represents an inner solver like CG, BB, or CGNE, while the storage requires Hermitian and skew-Hermitian parts of A and the residual vector.

5 | CONCLUSION

Gradient iterations provide a versatile tool in linear algebra. Apart from parameter estimates related to the spectral properties, steepest descent variants have also been tried recently with success as iterative methods.[16,18,30,31] This article extends the spectral properties of gradient iterations and gives an application in the Hermitian and skew-Hermitian splitting method. Note that this approach can be extended to some other splitting methods. For example, the modified HSS method[32] has the same optimal parameter as that established in Bai et al.[2] for the HSS method; see Bai et al.[32] for more details (see also Wu[7] for another instance). Our experiments confirm that the gradient-enhanced HSS method can be an attractive alternative to the original one.

ACKNOWLEDGMENT

This work was supported by the French national programme LEFE/INSU and the project ADOM (Méthodes de décomposition de domaine asynchrones) of the French National Research Agency (ANR).

Conflict of Interest

The authors declare no potential conflict of interests.

ORCID

Qinmeng Zou https://orcid.org/0000-0001-7752-8100
Frédéric Magoulès https://orcid.org/0000-0002-1198-7539

REFERENCES

1. Peaceman DW, Rachford HH. The numerical solution of parabolic and elliptic differential equations. J Soc Ind Appl Math. 1955;3(1):28–41.
2. Bai Z-Z, Golub GH, Ng MK. Hermitian and Skew-Hermitian splitting methods for non-Hermitian positive definite linear systems. SIAM J Matrix Anal Appl. 2003;24(3):603–626.
3. Bai Z-Z, Golub GH, Ng MK. On successive-overrelaxation acceleration of the Hermitian and skew-Hermitian splitting iterations. Numer Linear Algebra Appl. 2007;14(4):319–335.
4. Bai Z-Z. Several splittings for non-Hermitian linear systems. Sci Chin Ser A. 2008;51(8):1339–1348.
5. Benzi M. A generalization of the Hermitian and skew-Hermitian splitting iteration. SIAM J Matrix Anal Appl. 2009;31(2):360–374.
6. Salkuyeh DK, Hezari D, Edalatpour V. Generalized successive overrelaxation iterative method for a class of complex symmetric linear system of equations. Int J Comput Math. 2015;92(4):802–815.
7. Wu S-L. Several variants of the Hermitian and skew-Hermitian splitting method for a class of complex symmetric linear systems. Numer Linear Algebra Appl. 2015;22(2):338–356.
8. Wu S-L, Li C-X. Modified complex-symmetric and skew-Hermitian splitting iteration method for a class of complex-symmetric indefinite linear systems. Numer Algorithms 2017;76(1):93–107.
9. Pourbagher M, Salkuyeh DK. On the solution of a class of complex symmetric linear systems. Appl Math Lett. 2018;76:14–20.
10. Saad Y. Iterative methods for sparse linear systems. 2nd ed. Philadelphia, PA: SIAM, 2003.
11. Cauchy AL. Méthode Générale pour la Résolution des Systèmes d'Équations Simultanées. Comp Rend Sci. 1847;25(1):536–538.
12. Akaike H. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. Ann Inst Stat Math 1959;11(1):1–16.
13. Zou Qinmeng, Magoulès Frédéric. Fast gradient methods with alignment for symmetric linear systems without using Cauchy step; 2019. preprint available at arXiv:1909.01479.
14. Nocedal J, Sartenaer A, Zhu C. On the behavior of the gradient norm in the steepest descent method. Comput Optim Appl. 2002;22(1):5–35.

15. Fletcher R. On the Barzilai-Borwein method. Boston, MA: Springer, 2005;p. 235–256.
16. De Asmundis R, di Serafino D, Riccio F, Toraldo G. On spectral properties of steepest descent methods. IMA J Numer Anal. 2013;33(4):1416–1435.
17. Yuan Y-X. A new stepsize for the steepest descent method. J Comput Math. 2006;24(2):149–156.
18. De Asmundis R, Di Serafino D, Hager WW, Toraldo G, Zhang H. An efficient gradient method using the Yuan steplength. Comput Optim Appl. 2014;59(3):541–563.
19. Dai Y-H, Yuan Y-X. Analysis of monotone gradient methods. J Ind Manag Optim. 2005;1(2):181–192.
20. Bertaccini D, Golub GH, Capizzano SS, Possio CT. Preconditioned HSS methods for the solution of non-Hermitian positive definite linear systems and applications to the discrete convection-diffusion equation. Numer Math. 2005;99(3):441–484.
21. Freund RW. Conjugate gradient-type methods for linear systems with complex symmetric coefficient matrices. SIAM J Sci Stat Comput. 1992;13(1):425–448.
22. Bai Z-Z, Golub GH, Li C-K. Optimal parameter in hermitian and skew-hermitian splitting method for certain two-by-two block matrices. SIAM J Sci Comput. 2006;28(2):583–603.
23. Huang Y-M. A practical formula for computing optimal parameters in the HSS iteration methods. J Comput Appl Math. 2014;255:142–149.
24. Barzilai J, Borwein JM. Two-point step size gradient methods. IMA J Numer Anal. 1988;8(1):141–148.
25. Raydan M. On the Barzilai and Borwein choice of steplength for the gradient method. IMA J Numer Anal. 1993;13(3):321–326.
26. Dai Y-H, Liao L-Z. R-linear convergence of the Barzilai and Borwein gradient method. IMA J Numer Anal. 2002;22(1):1–10.
27. Friedlander A, Martínez JM, Molina B, Raydan M. Gradient method with retards and generalizations. SIAM J Numer Anal. 1999;36(1):275–289.
28. Vanden Doel K, Ascher U. The chaotic nature of faster gradient descent methods. J Sci Comput. 2012;51(3):560–581.
29. Golub GH, Van Loan CF. Matrix computations. 4th ed. Baltimore, MD: Johns Hopkins University Press, 2013.
30. Zou Q, Magoulès F. A new cyclic gradient method adapted to large-scale linear systems. Paper presented at: Proceedings of the 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science; 2018:196-199; IEEE; Wuxi, China.
31. Gonzaga CC, Schneider RM. On the steepest descent algorithm for quadratic functions. Comput Optim Appl. 2016;63(2):523–542.
32. Bai Z-Z, Benzi M, Chen F. Modified HSS iteration methods for a class of complex symmetric linear systems. Computing. 2010;87(3):93–111.

How to cite this article: Zou Q, Magoulès F. Parameter estimation in the Hermitian and skew-Hermitian splitting method using gradient iterations. *Numer Linear Algebra Appl*. 2020;27:e2304. https://doi.org/10.1002/nla.2304