Journal of Computational and Applied Mathematics 308 (2016) 276–300

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Structured condition numbers of structured Tikhonov regularization problem and their estimations

Huai-An Diao [a], Yimin Wei [b], Sanzheng Qiao [c,*]

[a] School of Mathematics and Statistics, Northeast Normal University, No. 5268 Renmin Street, Chang Chun 130024, PR China

[b] School of Mathematics & Shanghai Key Laboratory of Contemporary Applied Mathematics, Fudan University, Shanghai 200433, PR China

[c] Department of Computing and Software, McMaster University, Hamilton, Ontario L8S4K1, Canada

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form 7 April 2016

MSC:
15A09
15A12
65F35

Keywords:
Tikhonov regularization
Structured matrix
Condition number
Componentwise
Structured perturbation
Small sample condition estimation

ABSTRACT

Both structured componentwise and structured normwise perturbation analysis of the Tikhonov regularization are presented. The structured matrices under consideration include: Toeplitz, Hankel, Vandermonde, and Cauchy matrices. Structured normwise, mixed and componentwise condition numbers for the Tikhonov regularization are introduced and their explicit expressions are derived. For the general linear structure, based on the derived expressions, we prove structured condition numbers are smaller than their corresponding unstructured counterparts. By means of the power method and small sample statistical condition estimation (SCE), fast condition estimation algorithms are proposed. Our estimation methods can be integrated into Tikhonov regularization algorithms that use the generalized singular value decomposition (GSVD). For large scale linear structured Tikhonov regularization problems, we show how to incorporate the SCE into the preconditioned conjugate gradient (PCG) method to get the posterior error estimations. The structured condition numbers and perturbation bounds are tested on some numerical examples and compared with their unstructured counterparts. Our numerical examples demonstrate that the structured mixed condition numbers give sharper perturbation bounds than existing ones, and the proposed condition estimation algorithms are reliable. Also, an image restoration example is tested to show the effectiveness of the SCE for large scale linear structured Tikhonov regularization problems.

1. Introduction

For discrete ill-posed problems, the Tikhonov regularization (cf. [1]) reads

$$\min_x \left\{ \|Ax-b\|_2^2 + \lambda^2 \|Lx\|_2^2 \right\}, \quad A \in \mathbb{R}^{m\times n} \text{ and } L \in \mathbb{R}^{p\times n} \tag{1.1}$$

where λ is the regularization parameter, which controls the weight between $\|Lx\|_2$ and the residual $\|Ax-b\|_2$. The matrix L is typically the identity matrix I_n or a discrete approximation to some derivation operator. Tikhonov regularization is also known as *ridge regression* in statistics [2].

* Corresponding author.
E-mail addresses: hadiao@nenu.edu.cn, hadiao78@yahoo.com (H.-A. Diao), ymwei@fudan.edu.cn, yimin.wei@gmail.com (Y. Wei), qiao@mcmaster.ca (S. Qiao).

http://dx.doi.org/10.1016/j.cam.2016.05.023
0377-0427/

For the regularization problem (1.1), to ensure the uniqueness of the solution for any $\lambda > 0$, we always assume that $\mathrm{rank}(L) = p \leq n \leq m$ and $\mathrm{rank}\left(\begin{bmatrix} A \\ L \end{bmatrix}\right) = n$(cf. [2, Section 5]). The regularization problem (1.1) can be rewritten in the matrix form

$$\min_x \left\| \begin{bmatrix} A \\ \lambda L \end{bmatrix} x - \begin{bmatrix} b \\ \mathbf{0} \end{bmatrix} \right\|_2, \tag{1.2}$$

where $\mathbf{0}$ is the zero vector. Since the normal equations corresponding to (1.2) are

$$\left(A^\top A + \lambda^2 L^\top L\right) x = A^\top b, \tag{1.3}$$

we can obtain the following explicit expression for the *Tikhonov regularized solution*:

$$x_\lambda = \left(A^\top A + \lambda^2 L^\top L\right)^{-1} A^\top b.$$

Alternatively, the problem (1.1) can also be solved by the generalized singular value decomposition (GSVD) [3–5]. For rectangular matrices $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$ with $\mathrm{rank}(L) = p$ and $\mathrm{rank}\left(\begin{bmatrix} A \\ L \end{bmatrix}\right) = n$, the GSVDof (A, L) is given by the pair of factorizations

$$A = U \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix} RQ^\top \quad \text{and} \quad L = V \begin{bmatrix} S & \mathbf{0} \end{bmatrix} RQ^\top, \tag{1.4}$$

where $U \in \mathbb{R}^{m\times n}$ has orthonormal columns, $V \in \mathbb{R}^{p\times p}, Q \in \mathbb{R}^{n\times n}$ are orthogonal, R is n-by-n, upper triangular and nonsingular, and Σ and S are $p \times p$ diagonal matrices: $\Sigma = \mathrm{Diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ and $S = \mathrm{Diag}(\mu_1, \mu_2, \ldots, \mu_p)$ with

$$0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_p < 1 \quad \text{and} \quad 1 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_p > 0,$$

satisfying $\Sigma^2 + S^2 = I_p$. Then the *generalized singular values* γ_i of (A, L) are defined by the ratios $\gamma_i = \sigma_i/\mu_i\,(i = 1, 2, \ldots, p)$. Once the GSVD is computed, the Tikhonov regularized solution can be obtained by [4, Chapter 4]

$$x_\lambda = QR^{-1} \begin{bmatrix} F & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix} \begin{bmatrix} \Sigma^\dagger & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix} U^\top b, \quad F = \mathrm{Diag}(f_1, f_2, \ldots, f_p),$$

where $f_i = \gamma_i^2/(\gamma_i^2 + \lambda^2)$ for $i = 1, 2, \ldots, p$, are called the *filter factors* for the Tikhonov regularization [4,6] and $\Sigma^\dagger$ is the Moore–Penrose inverse of Σ [2].

In sensitivity analysis, condition numbers are of great importance because they measure the *worst-case* effect of small changes in the data on the solution. For the perturbation analysis of the linear least squares (LS) problem, the reader is referred to [7–11]. Arioli et al. [7] introduced a *partial condition number* of the LS problem, which can be viewed as a condition number of a *linear functional* of the LS problem. Baboulin et al. [8] have shown that the partial condition numbers of the LS problem represent some quantities in statistics. For the perturbation analysis for the Tikhonov regularization, we refer to [12,13] and references therein. Malyshev [14] adopted a unified theory to study the normwise condition numbers for the Tikhonov regularization. Chu et al. [15] investigated the componentwise perturbation analysis of the Tikhonov regularization problems and derived condition number expressions involving the Kronecker products, which can be of huge dimension even for small problems, preventing us from estimating the condition numbers while solving the Tikhonov regularization problem. In this paper, we consider the structured condition numbers for a linear functional of the Tikhonov regularization. Fast condition number estimation, which is important in practice, is discussed.

Structured matrix computation is a hot research topic; see [16,17] and the references therein. The structured Tikhonov regularization problem was recently studied in [18–20]. Eldén gave a stable efficient algorithm for the Tikhonov regularization with triangular Toeplitz structure [21]. Park and Eldén [22] devised fast algorithms for solving LS with Toeplitz structure, based on the generalization of the classical Schur algorithm, and discussed their stability properties. Also, Park and Eldén studied the stability analysis and fast algorithms for triangularization of rectangular Toeplitz matrices [23]. Hence, it is natural to investigate *structured perturbations* on the structured coefficient matrix, which lead to the structured condition numbers for the structured Tikhonov regularization problem. Structured condition numbers for several categories of structured matrices have been presented in [24–26,10,27–32]. In this paper we derive explicit formulas for the condition numbers of the Tikhonov regularization problem, when perturbations of (A, b) are measured by normwise or componentwise or a mixture of normwise and componentwise. To make our discussion general, we consider the condition number of Mx, i.e., a linear function of the Tikhonov regularized solution, where $M \in \mathbb{R}^{l\times n}$ and $x \in \mathbb{R}^n$, $l \leq n$. The common situations are the special cases, when M is the identity matrix (condition number of the Tikhonov regularized solution) or a canonical vector (condition number of one component of the solution). We obtain the expressions of the structured condition numbers in the absence of the Kronecker product, so that they can be estimated by the power method due to Hager [33] and Higham [34,35], see [36, Chapter 15] for the detail, while solving the Tikhonov regularization problem.

Moreover, in this paper, we adopt the statistical condition estimation (SCE) method [37] for numerically estimating the condition of Tikhonov regularization problem. The SCE for both the small/medium scale and large scale structured Tikhonov regularization problem is considered. The SCE can be used to estimate the componentwise local sensitivity of any differentiable function at a given input data, which is flexible and accommodates a wide range of perturbation types such as structured perturbations. Thus SCE often provides less conservative estimates than the methods that do not exploit structures. The SCE method has been shown to be both reliable and efficient for many problems including linear systems [38], structured linear systems [39], linear least squares problems [40], eigenvalue problems [41,42], matrix functions [37], the roots of polynomials [43], etc.

We follow the convention of representing a point $x \in \mathbb{R}^n$ as a column vector. If $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, then $[x; y]$ is an $m+n$ column vector by stacking x on top of y. If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times q}$, then $[A, B]$ denotes the matrix obtained by putting A and B side by side. The symbol '$\cdot^\top$' denotes matrix transpose, $\|\cdot\|_2$ is the spectral norm, $\|\cdot\|_F$ is the Frobenius norm and $\|\cdot\|_\infty$ is the infinity norm. The matrix $\mathrm{Diag}(d) \in \mathbb{R}^{q\times q}$ denotes a diagonal matrix with the vector d's entries being its corresponding diagonal components. For any points $a, b \in \mathbb{R}^n$, the vector $c = \frac{a}{b}$ is obtained by componentwise division. In particular, $b_i = 0$ assumes $a_i = 0$, and in this case $c_i = 0$. For a matrix $A \in \mathbb{R}^{m\times n}$, we define $\mathrm{vec}(A) \in \mathbb{R}^{mn}$ by $\mathrm{vec}(A) = [a_1^\top, a_2^\top, \ldots, a_n^\top]^\top$, where $A = [a_1, a_2, \ldots, a_n]$ with $a_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, n$. The unvec operation is defined as $A = \mathrm{unvec}(v)$ which sets the entries of A to $a_{ij} = v_{i+(j-1)n}$ for $v = [v_1, v_2, \ldots, v_{mn}] \in \mathbb{R}^{1\times mn}$. We define a permutation matrix Π of order mn so that $\Pi(\mathrm{vec}(A)) = \mathrm{vec}\left(A^\top\right)$. Let '$\otimes$' denote the *Kronecker product* [44], i.e., $A \otimes B = [a_{ij}B] \in \mathbb{R}^{mp\times nq}$ for $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. The notation $|A| \le |B|$ means that $|a_{ij}| \le |b_{ij}|$. For the Kronecker product, we recall the following properties, which can be found in [44],

$$(A\otimes B)^\top = A^\top \otimes B^\top, \qquad |A\otimes B| = |A|\otimes|B|, \qquad \mathrm{vec}(AXB) = \left(B^\top \otimes A\right)\mathrm{vec}(X), \tag{1.5}$$

where $|A| = [|a_{ij}|]$ and a_{ij} is the (i, j)-th entry of A.

This paper is organized as follows. We provide some preliminaries in Section 2, investigate matrices with linear structures in Section 3 and move to matrices with nonlinear structures in Section 4. The SCE-based condition estimation algorithms are proposed in Section 5. In Section 6, we demonstrate test results showing the sharpness of our structured condition numbers and effectiveness of the condition estimation algorithms. Finally, conclusions are drawn in Section 7.

2. Preliminaries

In this section, we first recall the general (unstructured) condition number definitions [27]. Then we consider the structured Tikhonov regularization problems, introduce structured perturbations, and define their structured condition numbers. Finally, we briefly describe the basic ideas of SCE.

2.1. Structured condition numbers for the Tikhonov regularization

For $x,\ a \in \mathbb{R}^p$ and $\varepsilon > 0$ we denote $S(a, \varepsilon) = \{x \in \mathbb{R}^p \mid |x-a| \le \varepsilon|a|\}$ and $T(a, \varepsilon) = \{x \in \mathbb{R}^p \mid \|x-a\|_2 \le \varepsilon\}$. For a function $F : \mathbb{R}^p \to \mathbb{R}^q$, we denote $\mathrm{Dom}(F)$ as its domain. The following lemma defines general (unstructured) condition numbers.

Lemma 1 (*[27]*)**.** *Let $F : \mathbb{R}^p \to \mathbb{R}^q$ be a continuous mapping defined on an open set $\mathrm{Dom}(F) \subset \mathbb{R}^p$. Let $a \in \mathrm{Dom}(F)$ such that $a \ne \mathbf{0}$ and $F(a) \ne \mathbf{0}$.*

(i) *The* mixed condition number *of F at a is defined by*

$$m(F, a) = \lim_{\varepsilon\to 0}\sup_{\substack{x\in S(a,\varepsilon)\\ x\ne a}} \frac{\|F(x)-F(a)\|_\infty}{\|F(a)\|_\infty}\frac{1}{d(x,a)} = \frac{\||\mathbf{D}F(a)|\,|a|\|_\infty}{\|F(a)\|_\infty},$$

where $\mathbf{D}F(a)$ is the Fréchet derivative of F at a and $|a| = (|a_i|)$ with $a = [a_1, a_2, \ldots, a_p]^\top$.

(ii) *Suppose $F(a) = (f_1(a), f_2(a), \ldots, f_q(a))$ is such that $f_j(a) \ne 0$ for $j = 1, 2, \ldots, q$. Then the* componentwise condition number *of F at a is*

$$c(F, a) = \lim_{\varepsilon\to 0}\sup_{\substack{x\in S(a,\varepsilon)\\ x\ne a}} \frac{d(F(x), F(a))}{d(x,a)} = \left\|\frac{|\mathbf{D}F(a)|\,|a|}{|F(a)|}\right\|_\infty.$$

(iii) *The* normwise condition number *of F at a is defined by*

$$\kappa(F, a) = \lim_{\varepsilon\to 0}\sup_{\substack{x\in T(a,\varepsilon)\\ x\ne a}} \frac{\|F(x)-F(a)\|_2}{\|x-a\|_2}\frac{\|a\|_2}{\|F(a)\|_2} = \frac{\|\mathbf{D}F(a)\|_2\|a\|_2}{\|F(a)\|_2}.$$

In the following we assume that ΔA and Δb are perturbations to A and b respectively, which satisfy $\mathrm{rank}\left(\begin{bmatrix} A+\Delta A \\ L\end{bmatrix}\right) = n$. The perturbed counterpart of the problem (1.1) and its normal equations (1.3) are, respectively,

$$\min_{x+\Delta x}\left\{\|(A+\Delta A)(x+\Delta x) - (b+\Delta b)\|_2^2 + \lambda^2\|L(x+\Delta x)\|_2^2\right\}, \tag{2.1}$$

and

$$\left[(A+\Delta A)^\top (A+\Delta A)+\lambda^2 L^\top L\right](x_\lambda+\Delta x)=(A+\Delta A)^\top (b+\Delta b).$$

Then the perturbed Tikhonov regularized solution is given by

$$x_\lambda+\Delta x=\left[(A+\Delta A)^\top (A+\Delta A)+\lambda^2 L^\top L\right]^{-1}(A+\Delta A)^\top (b+\Delta b). \tag{2.2}$$

Denoting

$$P(A,\lambda)=\left(A^\top A+\lambda^2 L^\top L\right)^{-1}.$$

Chu et al. [15] define the *non-structured* mixed, componentwise, and normwise condition numbers for the Tikhonov regularization and obtain respectively

$$\mathrm{m}_{\mathrm{Reg}}=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta A|\le\epsilon|A|\\ |\Delta b|\le\epsilon|b|}}\frac{\|\Delta x\|_\infty}{\epsilon\|x_\lambda\|_\infty}=\frac{\left\|\,|H(A,b)|\mathrm{vec}(|A|)+\left|P(A,\lambda)A^\top\right|\,|b|\,\right\|_\infty}{\|x_\lambda\|_\infty}, \tag{2.3}$$

$$\mathrm{c}_{\mathrm{Reg}}=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta A|\le\epsilon|A|\\ |\Delta b|\le\epsilon|b|}}\frac{1}{\epsilon}\left\|\frac{\Delta x}{x_\lambda}\right\|_\infty=\left\|\frac{|H(A,b)|\mathrm{vec}(|A|)+\left|P(A,\lambda)A^\top\right|\,|b|}{x_\lambda}\right\|_\infty, \tag{2.4}$$

$$\mathrm{cond}^{\mathrm{F}}_{\mathrm{Reg}}=\lim_{\epsilon\to 0}\sup_{\|[\Delta A,\,\Delta b]\|_F\le\epsilon\|[A,\,b]\|_F}\frac{\|\Delta x\|_2}{\epsilon\|x_\lambda\|_2}=\frac{\left\|\left[H(A,b),\ P(A,\lambda)A^\top\right]\right\|_2\|[A,\,b]\|_F}{\|x_\lambda\|_2}, \tag{2.5}$$

where $H(A,b)=-x_\lambda^\top\otimes\left[P(A,\lambda)A^\top\right]+\left[P(A,\lambda)\otimes r_\lambda^\top\right]$ and $r_\lambda=b-Ax_\lambda$.

If we define a mapping

$$\psi:\ [A,\ b]\in\mathbb{R}^{m\times n}\times\mathbb{R}^m\mapsto\left(A^\top A+\lambda^2 L^\top L\right)^{-1}A^\top b\in\mathbb{R}^n \tag{2.6}$$

then it is easy to see that the definitions in Lemma 1 are equivalent to (2.3)–(2.5), that is,

$$\mathrm{m}_{\mathrm{Reg}}:=m(\psi,[A,\ b]),\qquad \mathrm{c}_{\mathrm{Reg}}:=c(\psi,[A,\ b]),\qquad \mathrm{cond}^{\mathrm{F}}_{\mathrm{Reg}}:=\kappa(\psi,[A,\ b]).$$

When the coefficient matrix A in (1.1) has some structures, such as Toeplitz, it is reasonable to assume that the perturbation ΔA in (2.1) has the same structure of A. Then ΔA is called *structured perturbation* [29,30] on A. Usually a structured matrix $A\in\mathbb{R}^{m\times n}$ can be represented by fewer than mn parameters. For example, an $m\times n$ Toeplitz matrix can be represented by its first column and last row, $m+n-1$ parameters. Here we use a mapping to characterize this relationship. Let $\mathcal{S}$ be the set of structured matrices under consideration and a the vector representing a structured matrix A, then we define a mapping

$$g:\ a\in\mathbb{R}^k\mapsto A\in\mathcal{S}.$$

In order to apply Lemma 1 to define the structured condition numbers for the Tikhonov regularization, we construct a mapping

$$\phi:\ [a;\,b]\in\mathbb{R}^{k+m}\mapsto M\left(A^\top A+\lambda^2 L^\top L\right)^{-1}A^\top b\in\mathbb{R}^l, \tag{2.7}$$

where $M\in\mathbb{R}^{l\times n}$, $l\le n$, is general. In particular, when $M=e_i^\top$, the ith column of the identity matrix, then we are interested in some particular component of x_λ.

Let Δa be the perturbation on a, then the structured perturbation matrix ΔA on A in (2.1) is $g(a+\Delta a)-g(a)$. Now we are ready to define the *structured mixed, componentwise and normwise* condition numbers for a linear functional of the structured Tikhonov regularization,

$$m^{\mathrm{Reg}}_{\mathcal{S}}(A,b):=m(\phi,[a;\,b])=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta a|\le\epsilon|a|\\ |\Delta b|\le\epsilon|b|}}\frac{\|M\Delta x\|_\infty}{\epsilon\,\|Mx_\lambda\|_\infty},$$

$$c^{\mathrm{Reg}}_{\mathcal{S}}(A,b):=c(\phi,[a;\,b])=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta a|\le\epsilon|a|\\ |\Delta b|\le\epsilon|b|}}\frac{1}{\epsilon}\left\|\frac{M\Delta x}{Mx_\lambda}\right\|_\infty,$$

$$\kappa^{\mathrm{Reg}}_{\mathcal{S}}(A,b):=\kappa(\phi,[a;\,b])=\lim_{\epsilon\to 0}\sup_{\|[\Delta a;\,\Delta b]\|_2\le\epsilon\|[a;\,b]\|_2}\frac{\|M\Delta x\|_2}{\epsilon\|Mx_\lambda\|_2},$$

where Δx is defined in (2.2).

Remark 1. Note that here g is a general mapping, in that it can represent any structure. When the structure in A is linear, such as symmetric, or Toeplitz, or Hankel, we can choose g a linear mapping, which will be discussed in Section 3. When A has a nonlinear structure such as Vandermonde or Cauchy, we can choose a nonlinear mapping g to define the structured condition numbers. Especially we can define the unstructured linear functional condition number for x_λ when we restrict $\mathcal{S}$ to be $\mathbb{R}^{m\times n}$, which are generalizations of (2.3)–(2.5), as follows

$$m^{\mathrm{Reg}}(A,b)=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta A|\le\epsilon|A|\\|\Delta b|\le\epsilon|b|}}\frac{1}{\epsilon}\frac{\|M\Delta x\|_\infty}{\|Mx_\lambda\|_\infty},\qquad c^{\mathrm{Reg}}(A,b)=\lim_{\epsilon\to 0}\sup_{\substack{|\Delta A|\le\epsilon|A|\\|\Delta b|\le\epsilon|b|}}\frac{1}{\epsilon}\left\|\frac{M\Delta x}{Mx_\lambda}\right\|_\infty,$$

$$\kappa^{\mathrm{Reg}}(A,b)=\lim_{\epsilon\to 0}\sup_{\|[\Delta A,\Delta b]\|_2\le\epsilon\|[A,b]\|_2}\frac{\|M\Delta x\|_2}{\epsilon\|Mx_\lambda\|_2}.$$

When $M=I_n$, the above definitions reduce to (2.3)–(2.5).

Finally, we give the well-known Banach lemma, which will be useful in Section 3.

Lemma 2. *Let $E\in\mathbb{R}^{n\times n}$ and $\|\cdot\|$ be any norm on $\mathbb{R}^{n\times n}$, if $\|E\|<1$, then I_n+E is nonsingular and its inverse can be expressed by*

$$(I_n+E)^{-1}=I_n-E+\mathcal{O}(\|E\|^2).$$

2.2. Statistical condition estimation

In the SCE, a small random perturbation is introduced to the input, and the change in the output, by an appropriate scaling, is measured as a condition estimate. Explicit bounds on the probability of the accuracy of the estimate exist [37]. The idea of SCE can be illustrated by a general real-valued function: $f:\mathbb{R}^p\to\mathbb{R}$, and we are interested in the sensitivity at some input vector x. By the Taylor theorem we have

$$f(x+\delta d)-f(x)=\delta(\mathbf{D}f(x))^\top d+\mathcal{O}(\delta^2),$$

where δ is a small scalar, $\|d\|_2=1$ and $\mathbf{D}f(x)$ is the Fréchet derivative of f at x. Note that the quantity $(\mathbf{D}f(x))^\top d$ (denoted by $\mathbf{D}f(x;d)$) is just the directional derivative of f with respect to x at the direction d. It is easy to see that up to the first order in δ,

$$|f(x+\delta d)-f(x)|\approx\delta\mathbf{D}f(x;d),$$

then the local sensitivity can be measured by $\|\mathbf{D}f(x)\|_2$. The condition numbers of f at x are mainly determined by the norm of the gradient $\mathbf{D}f(x)$ [37]. According to [37], if we select d uniformly and randomly from the unit p-sphere S_{p-1} (denoted $d\in\mathcal{U}(S_{p-1})$), then the expectation $\mathbf{E}(|\mathbf{D}f(x;d)|/\omega_p)$ is $\|\mathbf{D}f(x)\|_2$, where ω_p is the Wallis factor. In practice, the Wallis factor can be approximated accurately [37] by

$$\omega_p\approx\sqrt{\frac{2}{\pi\left(p-\frac{1}{2}\right)}}.$$

Therefore, we can use

$$\nu=\frac{|\mathbf{D}f(x;d)|}{\omega_p}$$

as a condition estimator, which can estimate $\|\mathbf{D}f(x)\|_2$ with high probability for the function f at x (see [37] for details), for example,

$$\mathrm{Prob}\left(\frac{\|\mathbf{D}f(x)\|_2}{\gamma}\le\nu\le\gamma\|\mathbf{D}f(x)\|_2\right)\ge 1-\frac{2}{\pi\gamma}+\mathcal{O}\left(\frac{1}{\gamma^2}\right),$$

for $\gamma>1$. We can use multiple samples of d, denoted d_j, to increase the accuracy [37]. The t-sample condition estimation is given by

$$\nu(k)=\frac{\omega_t}{\omega_p}\sqrt{|\mathbf{D}f(x;d_1)|^2+|\mathbf{D}f(x;d_2)|^2+\cdots+|\mathbf{D}f(x;d_t)|^2},$$

where $[d_1,d_2,\ldots,d_t]$ is orthonormalized after $d_1,d_2,\ldots,d_t$ are selected uniformly and randomly from $\mathcal{U}(S_{p-1})$. The accuracy of $\nu(2)$ is given by

$$\mathrm{Prob}\left(\frac{\|\nabla f(x)\|_2}{\gamma}\le\nu(2)\le\gamma\|\nabla f(x)\|_2\right)\approx 1-\frac{\pi}{4\gamma^2},\quad\gamma>1.$$

Usually, a few samples are sufficient for good accuracy. These results can be conveniently generalized to vector- or matrix-valued functions by viewing f as a map from $\mathbb{R}^p$ to $\mathbb{R}^q$. The operations vec and unvec can be used to convert between matrices and vectors, where each of the q entries of f is a scalar-valued function. Evaluating the matrix function at a slightly perturbed argument yields a local condition estimate for one component of the computed solution.

3. Linear structures

In this section, we consider the classes $\mathcal{L}$ of structured matrices that is a linear subspace of $\mathbb{R}^{m\times n}$. The examples of such class include Toeplitz and Hankel matrices. We first present a structured perturbation analysis and structured condition numbers. Then we propose efficient condition number estimators using the power method.

3.1. Condition numbers

Suppose that $\dim(\mathcal{L}) = k$, and $S_1, S_2, \ldots, S_k$ form a basis for $\mathcal{L}$. Then for $A \in \mathcal{L}$, there is a unique point $a = [a_1, a_2, \ldots, a_k]^\top \in \mathbb{R}^k$ such that

$$A = \sum_{i=1}^{k} a_i S_i. \tag{3.1}$$

We write $A = g(a)$. Since A is determined by a, we consider the perturbation $\Delta a \in \mathbb{R}^k$ on a. Then we denote $\Delta A = g(a + \Delta a) - g(a) = g(\Delta a)$, since g is linear.

Lemma 3. *The Fréchet derivative $\mathbf{D}\phi([a; b])$ of function ϕ defined in (2.7) is given by*

$$\mathbf{D}\phi([a; b]) = MP(A, \lambda)\left[\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_k, A^\top\right], \tag{3.2}$$

where $\mathsf{v}_i = -A^\top S_i x_\lambda + S_i^\top r_\lambda$ for $i = 1, 2, \ldots, k$.

Proof. Let $\Delta A = g(\Delta a)$ and Δb be perturbations on $A = g(a)$ and b respectively. Firstly, denoting $\mathcal{A} = (A + \Delta A)^\top(A + \Delta A) + \lambda^2 L^\top L$ and recalling that $P(A, \lambda) = (A^\top A + \lambda^2 L^\top L)^{-1}$, we have

$$\begin{aligned}\mathcal{A} &= \left(A^\top A + \lambda^2 L^\top L\right) + \left(A^\top(\Delta A) + (\Delta A)^\top A\right) + (\Delta A)^\top(\Delta A)\\ &= \left(A^\top A + \lambda^2 L^\top L\right)\left[I_n + P(A, \lambda)\left(A^\top(\Delta A) + (\Delta A)^\top A\right) + P(A, \lambda)\left((\Delta A)^\top(\Delta A)\right)\right].\end{aligned}$$

If $\|\Delta A\|$ is sufficiently small, then $\left\|P(A, \lambda)\left(A^\top(\Delta A) + (\Delta A)^\top A + (\Delta A)^\top(\Delta A)\right)\right\| < 1$, from Lemma 2, $\mathcal{A}$ is nonsingular and its inverse

$$\begin{aligned}\mathcal{A}^{-1} &= \left[I_n + P(A, \lambda)\left(A^\top(\Delta A) + (\Delta A)^\top A\right) + P(A, \lambda)\left((\Delta A)^\top(\Delta A)\right)\right]^{-1} P(A, \lambda)\\ &= P(A, \lambda) - P(A, \lambda)\left(A^\top(\Delta A) + (\Delta A)^\top A\right)P(A, \lambda) + \mathcal{O}(\|\Delta A\|^2),\end{aligned} \tag{3.3}$$

since $\left\|A^\top(\Delta A) + (\Delta A)^\top A\right\| = \mathcal{O}(\|\Delta A\|)$ and $\left\|(\Delta A)^\top(\Delta A)\right\| = \mathcal{O}(\|\Delta A\|^2)$. From (2.2), (3.3) and $x_\lambda = P(A, \lambda)A^\top b$, after some algebraic manipulation, we have

$$\Delta x = P(A, \lambda)\Big(A^\top(\Delta b) + (\Delta A)^\top(b - Ax_\lambda) - A^\top(\Delta A)x_\lambda\Big) + \mathcal{O}(\|\Delta A\|^2) + \mathcal{O}(\|\Delta A\|\,\|\Delta b\|).$$

Omitting the second and higher order terms and applying the third equation in (1.5), we have

$$\begin{aligned}\Delta x &\approx P(A, \lambda)\Big(A^\top(\Delta b) + (\Delta A)^\top(b - Ax_\lambda) - A^\top(\Delta A)x_\lambda\Big)\\ &= P(A, \lambda)\left[\left(-(x_\lambda^\top \otimes A^\top) + \left(r_\lambda^\top \otimes I_n\right)\Pi\right)\mathsf{vec}(\Delta A) + A^\top(\Delta b)\right]\\ &= P(A, \lambda)\left[\left(-(x_\lambda^\top \otimes A^\top) + (I_n \otimes r_\lambda^\top)\right)\mathsf{vec}(\Delta A) + A^\top(\Delta b)\right],\end{aligned} \tag{3.4}$$

recalling that $r_\lambda = b - Ax_\lambda$.

Since ΔA is a structured perturbation on A, then $\Delta A = g(\Delta a)$, i.e., there exist parameters $\Delta a_1, \Delta a_2, \ldots, \Delta a_k$ such that $\Delta A = \sum_{i=1}^{k} \Delta a_i S_i$. Denote $\Delta a = [\Delta a_1, \Delta a_2, \ldots, \Delta a_k]^\top$. From (3.4), we have

$$\begin{aligned}\phi([a + \Delta a; b + \Delta b]) - \phi([a; b]) &\approx MP(A, \lambda)\left\{\left(-(x_\lambda^\top \otimes A^\top) + (I_n \otimes r_\lambda^\top)\right)[\mathsf{vec}(S_1), \ldots, \mathsf{vec}(S_k)]\Delta a + A^\top(\Delta b)\right\}\\ &= MP(A, \lambda)\left[-A^\top S_1 x_\lambda + S_1^\top r_\lambda, \ldots, -A^\top S_k x_\lambda + S_k^\top r_\lambda, A^\top\right]\Delta v,\end{aligned}$$

where $\Delta v = [\Delta a; \Delta b]$. By the definition of the Fréchet derivative, the lemma then can be proved. □

Theorem 1. *Let $A \in \mathcal{L}$, $b \in \mathbb{R}^m$ and $x_\lambda = \left(A^\top A + \lambda^2 L^\top L\right)^{-1} A^\top b = P(A, \lambda)A^\top b$ be the Tikhonov regularized solution of (1.1). Then we obtain the structured normwise, componentwise, and mixed condition numbers:*

$$m_{\mathcal{L}}^{\mathsf{Reg}}(A, b) = \frac{\left\|\sum_{i=1}^{k} |a_i|\left|MP(A, \lambda)\left(A^\top S_i x_\lambda - S_i^\top r_\lambda\right)\right| + \left|MP(A, \lambda)A^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty},$$

$$c_{\mathcal{L}}^{\mathrm{Reg}}(A,b)=\left\|\frac{\sum_{i=1}^{k}|a_i|\left|MP(A,\lambda)\left(A^\top S_i x_\lambda - S_i^\top r_\lambda\right)\right|+\left|MP(A,\lambda)A^\top\right|\,|b|}{Mx_\lambda}\right\|_\infty,$$

$$\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)=\frac{\left\|MP(A,\lambda)\left[S_1^\top r_\lambda - A^\top S_1 x_\lambda,\ldots,S_k^\top r_\lambda - A^\top S_k x_\lambda,\ A^\top\right]\right\|_2\left\|\begin{bmatrix}a\\b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}.$$

Proof. From Lemmas 1 and 3, we have

$$\begin{aligned}
m_{\mathcal{L}}^{\mathrm{Reg}}(A,b)&=\frac{\left\|\left|\mathbf{D}\phi\left([a;b]\right)\right|\begin{bmatrix}|a|\\|b|\end{bmatrix}\right\|_\infty}{\|x_\lambda\|_\infty}\\
&=\frac{\left\|\left|MP(A,\lambda)\left[\mathrm{v}_1;\mathrm{v}_2;\ldots;\mathrm{v}_k\right]\right|\,|a|+\left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty}\\
&=\frac{\left\|\sum_{i=1}^{k}|a_i|\left|MP(A,\lambda)\left(A^\top S_i x_\lambda - S_i^\top r_\lambda\right)\right|+\left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty}.
\end{aligned}$$

Similarly, we can obtain explicit expressions of the structured componentwise and normwise condition numbers. □

When $\{S_k\}$ is the canonical basis for $\mathbb{R}^{m\times n}$ in Theorem 1, we have the following compact forms of the unstructured condition numbers in Remark 1 for $m^{\mathrm{Reg}}(A,b)$, $c^{\mathrm{Reg}}(A,b)$ and $\kappa^{\mathrm{Reg}}(A,b)$.

Theorem 2. *As stated before, we have the following expressions*

$$m^{\mathrm{Reg}}(A,b)=\frac{\left\|\left|MP(A,\lambda)\left[(I_n\otimes r_\lambda^\top)-(x^\top\otimes A^\top)\right]\right|\mathrm{vec}(|A|)+\left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty},$$

$$c^{\mathrm{Reg}}(A,b)=\left\|\frac{\left|MP(A,\lambda)\left[(I_n\otimes r_\lambda^\top)-(x^\top\otimes A^\top)\right]\right|\mathrm{vec}(|A|)+\left|MP(A,\lambda)A^\top\right|\,|b|}{Mx_\lambda}\right\|_\infty,$$

$$\kappa^{\mathrm{Reg}}(A,b)=\frac{\left\|MP(A,\lambda)\left[(I_n\otimes r_\lambda^\top)-(x^\top\otimes A^\top),A^\top\right]\right\|_2\sqrt{\|A\|_F^2+\|b\|_2^2}}{\|Mx_\lambda\|_2}.$$

Proof. For the expression of $\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)$ given in Theorem 1, let $\{S_{ij}=e_i^{(m)}e_j^{(n)\top}\}$ be the canonical basis for $\mathbb{R}^{m\times n}$, where $e_j^{(n)}$ is the jth column of the identity matrix I_n, $i=1,2,\ldots,m$ and $j=1,2,\ldots,n$. Then we have the following simplified expression:

$$-S_{ij}^\top r_\lambda + A^\top S_{ij}x_\lambda = -e_j^{(n)}r_{\lambda,(i)} + A^\top e_i^{(m)}x_{\lambda,(j)},$$

where $r_{\lambda,(i)}$ and $x_{\lambda,(j)}$ are respectively the ith and jth components of r_λ and x_λ. Now, fixing j, we get

$$\left[-e_j^{(n)}r_{\lambda,(1)}+A^\top e_1^{(m)}x_{\lambda,(j)},\ldots,-e_j^{(n)}r_{\lambda,(n)}+A^\top e_n^{(m)}x_{\lambda,(j)}\right]=-e_j^{(n)}r_\lambda^\top + x_{\lambda,(j)}A^\top,$$

which implies that

$$\begin{aligned}
&\left[-S_{11}^\top r_\lambda + A^\top S_{11}x_\lambda,\ldots,-S_{m1}^\top r_\lambda + A^\top S_{m1}x_\lambda,-S_{12}^\top r_\lambda + A^\top S_{12}x_\lambda\ldots,-S_{mn}^\top r_\lambda + A^\top S_{mn}x_\lambda\right]\\
&\quad=\left[-(e_1^{(n)}\otimes r_\lambda^\top)+x_{\lambda,(1)}A^\top,-(e_2^{(n)}\otimes r_\lambda^\top)+x_{\lambda,(2)}A^\top,\ldots,-(e_n^{(n)}\otimes r_\lambda^\top)+x_{\lambda,(n)}A^\top\right]\\
&\quad=\left[-(I_n\otimes r_\lambda^\top)+(x^\top\otimes A^\top)\right].
\end{aligned}$$

Applying the above equation to the expression of $\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)$ in Theorem 1, we prove the third statement. The expressions of $m^{\mathrm{Reg}}(A,b)$ and $c^{\mathrm{Reg}}(A,b)$ can be obtained similarly. □

Remark 2. If we choose $M=I_n$, then $m^{\mathrm{Reg}}(A,b)$, $c^{\mathrm{Reg}}(A,b)$ and $\kappa^{\mathrm{Reg}}(A,b)$ respectively reduce to the expressions of $\mathrm{m}_{\mathrm{Reg}}$, $\mathrm{c}_{\mathrm{Reg}}$ and $\mathrm{cond}_{\mathrm{Reg}}^{\mathrm{F}}$ in (2.3)–(2.5).

How are the structured condition numbers compared to their unstructured counterparts? The cases of nonsingular matrix inversion and linear systems have been investigated in [25,26,29,30,45] and the references therein. In the following proposition, we will show that $m_{\mathcal{L}}^{\mathrm{Reg}}(A, b)$ is smaller than $m^{\mathrm{Reg}}(A, b)$. The same is true for the componentwise and normwise condition numbers. Before that we need the following lemma for rectangular structured matrices. Its proof is omitted since it is similar to that of [29, Lemma 6.3].

Lemma 4. *When A is a Toeplitz or Hankel matrix, and $A = \sum_{i=1}^{k} a_i S_i$, then*

$$\|a\|_2 \le \sqrt{2}\|A\|_F.$$

Proposition 1. *Suppose that the basis $\{S_1, S_2, \dots, S_k\}$ for $\mathcal{L}$ satisfies $|A| = \sum_{i=1}^{k} |a_i|\,|S_i|$ for any $A \in \mathcal{L}$ in (3.1), then*

$$m_{\mathcal{L}}^{\mathrm{Reg}}(A, b) \le m^{\mathrm{Reg}}(A, b) \quad \text{and} \quad c_{\mathcal{L}}^{\mathrm{Reg}}(A, b) \le c^{\mathrm{Reg}}(A, b).$$

For the structured normwise condition number, when A is a Toeplitz or Hankel matrix, we have

$$\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A, b) \le \sqrt{2}\max\left\{\max_{i=1,2,\dots,k}\|S_i\|_F,\, 1\right\}\kappa^{\mathrm{Reg}}(A, b).$$

Proof. From Theorem 1, using the monotonicity of the infinity norm, we have

$$\begin{aligned}
&\left\|\sum_{i=1}^{k}|a_i|\left|MP(A,\lambda)\left(A^\top S_i x_\lambda - S_i^\top r_\lambda\right)\right| + \left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty \\
&= \left\|\left|MP(A,\lambda)\left[A^\top S_1 x_\lambda - S_1^\top r_\lambda, \dots, A^\top S_k x_\lambda - S_k^\top r_\lambda\right]\right|\,|a| + \left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty \\
&= \left\|\left|MP(A,\lambda)\left[A^\top S_1 x_\lambda - S_1^\top r_\lambda, \dots, A^\top S_k x_\lambda - S_k^\top r_\lambda, A^\top\right]\right|\begin{bmatrix}|a|\\|b|\end{bmatrix}\right\|_\infty \\
&= \left\|\left|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right)\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k), A^\top\right]\right|\begin{bmatrix}|a|\\|b|\end{bmatrix}\right\|_\infty \\
&\le \left\|\left[\left|MP(A,\lambda)\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right)\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k)\right|, \left|MP(A,\lambda)A^\top\right|\right]\begin{bmatrix}|a|\\|b|\end{bmatrix}\right\|_\infty \\
&= \left\|\left|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right)\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k), A^\top\right]\right|\begin{bmatrix}|a|\\|b|\end{bmatrix}\right\|_\infty \\
&\le \left\|\left|MP(A,\lambda)\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right)\right|\sum_{i=1}^{k}|a_i|\,|\mathrm{vec}(S_i)| + \left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty \\
&= \left\|\left|MP(A,\lambda)\left[(I_n\otimes r_\lambda^\top) - (x^\top\otimes A^\top)\right]\right|\mathrm{vec}(|A|) + \left|MP(A,\lambda)A^\top\right|\,|b|\right\|_\infty,
\end{aligned}$$

for the last equality we use the assumption $|A| = \sum_{i=1}^{k}|a_i|\,|S_i|$. With the above inequality, and the expressions of $m_{\mathcal{L}}^{\mathrm{Reg}}(A, b)$, $m^{\mathrm{Reg}}(A, b)$, $c_{\mathcal{L}}^{\mathrm{Reg}}(A, b)$, $c^{\mathrm{Reg}}(A, b)$, it is easy to prove the first two inequalities in this proposition.

When A is a Toeplitz or Hankel matrix, the standard basis for the Toeplitz matrix subspace or the Hankel matrix subspace is orthogonal under the inner product $\langle B_1, B_2\rangle = \mathrm{trace}\left(B_1^\top B_2\right) = [\mathrm{vec}(B_1)]^\top \mathrm{vec}(B_2)$ for $B_1,\ B_2 \in \mathbb{R}^{m\times n}$. It is easy to deduce that

$$\begin{aligned}
&\left\|MP(A,\lambda)\left[S_1^\top r_\lambda - A^\top S_1 x_\lambda, \dots, S_k^\top r_\lambda - A^\top S_k x_\lambda,\ A^\top\right]\right\|_2 \\
&= \left\|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right)\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k),\ A^\top\right]\right\|_2 \\
&= \left\|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right),\ A^\top\right]\begin{bmatrix}[\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k)] & 0\\ 0 & I_m\end{bmatrix}\right\|_2 \\
&\le \left\|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right),\ A^\top\right]\right\|_2 \left\|\begin{bmatrix}[\mathrm{vec}(S_1), \dots, \mathrm{vec}(S_k)] & 0\\ 0 & I_m\end{bmatrix}\right\|_2 \\
&= \left\|MP(A,\lambda)\left[\left((x^\top\otimes A^\top) - (I_n\otimes r_\lambda^\top)\right),\ A^\top\right]\right\|_2 \max\left\{\max_{i=1,2,\dots,k}\|S_i\|_F,\, 1\right\},
\end{aligned}$$

where for the last equation we used the orthogonality of the basis $\{S_i\}$. So from Lemma 4,

$$\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b) \le \max\left\{\max_{i=1,2,\ldots,k}\|S_i\|_F,\, 1\right\} \frac{\left\|MP(A,\lambda)\left[\left(x^\top\otimes A^\top\right)-\left(I_n\otimes r_\lambda^\top\right),\, A^\top\right]\right\|_2 \left\|\begin{bmatrix} a\\ b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}$$
$$\le \max\left\{\max_{i=1,2,\ldots,k}\|S_i\|_F,\, 1\right\} \frac{\left\|MP(A,\lambda)\left[\left(x^\top\otimes A^\top\right)-\left(I_n\otimes r_\lambda^\top\right),\, A^\top\right]\right\|_2 \sqrt{2\,\|A\|_F^2+\|b\|_2^2}}{\|Mx_\lambda\|_2},$$

which completes the proof of this proposition. □

Remark 3. Clearly, the assumption $|A|=\sum_{i=1}^k |a_i|\,|S_i|$ in Proposition 1 is satisfied for Toeplitz and Hankel matrices.

3.2. Condition number estimators

Efficiently estimating condition numbers is crucial in practice. The condition number $\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)$, for example, involves the spectral norm of the $l\times(m+k)$ matrix $\mathbf{D}\phi([a;b])$, which can be expensive to compute when m or k is large. The power method can be used for fast condition number estimation [36, page 289]. Its major computation is the matrix–vector multiplications $\mathbf{D}\phi([a;b])h_1$ and $\mathbf{D}\phi([a;b])^\top h_2$ with $h_1\in\mathbb{R}^{m+k}$ and $h_2\in\mathbb{R}^l$. To consider $\mathbf{D}\phi([a;b])^\top h$, for $h\in\mathbb{R}^l$ and v_i defined in Lemma 3, denoting $D=MP(A,\lambda)A^\top$ and using $P(A,\lambda)^\top=P(A,\lambda)$, (1.5) and $\Pi^\top=\Pi^{-1}$, we have

$$\begin{aligned}
\mathrm{v}_i^\top P(A,\lambda)M^\top h &= r_\lambda^\top S_i P(A,\lambda)M^\top h - x_\lambda^\top S_i^\top AP(A,\lambda)M^\top h\\
&= \mathrm{vec}\left[r_\lambda^\top S_i P(A,\lambda)M^\top h - x_\lambda^\top S_i^\top D^\top h\right]\\
&= \left[\left(h^\top MP(A,\lambda)\right)\otimes r_\lambda^\top\right]\mathrm{vec}(S_i) - \left[(h^\top D)\otimes x_\lambda^\top\right]\mathrm{vec}(S_i^\top)\\
&= \left[\left(h^\top MP(A,\lambda)\right)\otimes r_\lambda^\top\right]\mathrm{vec}(S_i) - \left[(h^\top D)\otimes x_\lambda^\top\right]\Pi\,\mathrm{vec}(S_i)\\
&= \mathrm{vec}(S_i)^\top\left[(P(A,\lambda)M^\top h)\otimes r_\lambda - \Pi^\top\left(D^\top h\otimes x_\lambda\right)\right]\\
&= \mathrm{vec}(S_i)^\top \Pi^{-1}\left[\Pi\,\mathrm{vec}(r_\lambda h^\top MP(A,\lambda)) - \mathrm{vec}(x_\lambda h^\top D)\right]\\
&= \left[\Pi\,\mathrm{vec}(S_i)\right]^\top \mathrm{vec}\left[P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D\right]\\
&= \left[\mathrm{vec}(S_i^\top)^\top\right]\mathrm{vec}\left[P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D\right]\\
&= \mathrm{trace}\left[S_i\left(P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D\right)\right],
\end{aligned} \tag{3.5}$$

where we applied $[\mathrm{vec}(A_1)]^\top\mathrm{vec}(A_2)=\mathrm{trace}\left(A_1^\top A_2\right)$ for the same dimensional matrices A_1 and A_2 in the last equality. It follows from (3.5) that

$$\mathbf{D}\phi([a;b])^\top h = \left(MP(A,\lambda)\left[\mathrm{v}_1,\ldots,\mathrm{v}_k,A^\top\right]\right)^\top h = \begin{bmatrix} \mathrm{v}_1^\top P(A,\lambda)M^\top h\\ \vdots\\ \mathrm{v}_k^\top P(A,\lambda)M^\top h\\ D^\top h\end{bmatrix} = \begin{bmatrix} a(h)\\ D^\top h\end{bmatrix}, \tag{3.6}$$

where $a(h)=\left[\mathrm{trace}\left(S_1(P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D)\right),\ldots,\mathrm{trace}\left(S_k(P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D)\right)\right]^\top$. It leads to the following proposition.

Proposition 2. *The adjoint operator of $\mathbf{D}\phi([a;b])$, with the scalar products $a_1^\top a_2+b_1^\top b_2$ and $h^\top h$ in $\mathbb{R}^{k+m}$ and $\mathbb{R}^l$ respectively, is*

$$\mathbf{D}\phi([a;b])^*:\ h\in\mathbb{R}^l\mapsto\left[a(h),\quad D^\top h\right]\in\mathbb{R}^k\times\mathbb{R}^m. \tag{3.7}$$

Furthermore, when $l=1$,

$$\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b) = \frac{\sqrt{\sum_{i=1}^k s_i^2+\|D\|_2^2}\,\left\|\begin{bmatrix} a\\ b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}, \tag{3.8}$$

where $s_i=\mathrm{trace}\Big(S_i\left(P(A,\lambda)M^\top r_\lambda^\top - x_\lambda D\right)\Big)$, $i=1,2,\ldots,k$.

Proof. For any $(\Delta a, \Delta b) \in \mathbb{R}^k \times \mathbb{R}^m$ and $h \in \mathbb{R}^l$, from Lemma 3 and (3.6), we have

$$\begin{aligned} \langle h, \mathbf{D}\phi([a;b]) \cdot (\Delta a, \Delta b)\rangle &= h^\top (\mathbf{D}\phi([a;b]) \cdot (\Delta a, \Delta b)) = h^\top \mathbf{D}\phi([a;b]) \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix} \\ &= (\mathbf{D}\phi([a;b])^\top h)^\top \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix} = a(h)^\top (\Delta a) + (D^\top h)^\top (\Delta b) = \langle \mathbf{D}\phi([a;b])^* \cdot h,\ (\Delta a, \Delta b)\rangle, \end{aligned}$$

which proves the first part. For the second part, noticing that

$$\|\mathbf{D}\phi([a;b])\|_2 = \|\mathbf{D}\phi([a;b])^\top\|_2 = \max_{h\neq 0} \frac{\left\| \left[a(h)^\top,\ (D^\top h)^\top\right]^\top \right\|_2}{\|h\|_2}$$

and using (3.6), where $h \in \mathbb{R}$ since $l = 1$, we can show that

$$\|\mathbf{D}\phi([a;b])\|_2 = \sqrt{\sum_{i=1}^k s_i^2 + \|D\|_2^2},$$

which completes the proof. $\square$

Remark 4. When $l = 1$, we compute the conditioning of the ith component of the solution. In that case M is the ith canonical vector of $\mathbb{R}^{1\times n}$ and, in (3.8), $P(A,\lambda)M^\top$ is the ith column of $P(A,\lambda)$ and D is the ith row of $P(A,\lambda)A^\top$.

Using (3.2) and (3.6), we can now apply the iteration of the power method [36, page 289] in Algorithm 1 to compute the normwise condition number $\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)$. In this algorithm, we assume that x_λ, r_λ and λ are available. When the GSVD (1.4) of (A, L) is available, a compact form of $P(A,\lambda)$ is given by

$$P(A,\lambda) = QR^{-1} \begin{bmatrix} (\Sigma^2 + \lambda^2 S^2)^{-1} & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix} R^{-\top} Q^\top, \tag{3.9}$$

which can be used to reduce the computational cost of the estimators of the normwise, mixed and componentwise condition numbers.

Algorithm 1 The power method for estimating $\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b)$

Select initial vector $h \in \mathbb{R}^l$.
for $p = 1, 2, \ldots$ **do**
  Using (3.9), calculate $P(A,\lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D$. From (3.6), denote $\bar{a}_p = a(h)$ and $\bar{b}_p = D^\top h$.
  Calculate $\nu = \left\|[\bar{a}_p;\ \bar{b}_p]\right\|_2$, let $a_p = \bar{a}_p/\nu$ and $b_p = \bar{b}_p/\nu$.
  Let $A_p = \sum_{i=1}^k a_{p,(i)} S_i$, where $a_{p,(i)}$ is the ith component of a_p.
  Using (3.2) and (3.9), compute $h = MP(A,\lambda)\left(A^\top b_p + A_p^\top r_\lambda - A^\top A_p x_\lambda\right)$.
end for
$\kappa_{\mathcal{L}}^{\mathrm{Reg}}(A,b) = \sqrt{\nu}$.

The quantity ν computed by Algorithm 1 is an approximation of the largest eigenvalue of $\mathbf{D}\phi([a;b])\mathbf{D}\phi([a;b])^\top$. When there is an estimate of the corresponding dominant eigenvector of $\mathbf{D}\phi([a;b])\mathbf{D}\phi([a;b])^\top$, the initial h can be set to this estimate, but in many implementations h is initialized as a random vector. The algorithm is terminated by a sufficient number of iterations or by evaluating the difference between two consecutive values of ν and comparing it to a tolerance given by the user.

For the mixed and componentwise condition numbers, we note that

$$\begin{aligned} m_{\mathcal{L}}^{\mathrm{Reg}}(A,b) &= \frac{\left\|MP(A,\lambda)\left[\mathsf{v}_1,\ldots,\mathsf{v}_k, A^\top\right]\mathrm{Diag}([a;b])\right\|_\infty}{\|Mx_\lambda\|_\infty} = \frac{\|\mathbf{D}\phi([a;b])\mathrm{Diag}([a;b])\|_\infty}{\|Mx_\lambda\|_\infty}, \\ c_{\mathcal{L}}^{\mathrm{Reg}}(A,b) &= \left\| \frac{MP(A,\lambda)\left[\mathsf{v}_1,\ldots,\mathsf{v}_k, A^\top\right]\mathrm{Diag}([a;b])}{Mx_\lambda} \right\|_\infty = \frac{\|\mathbf{D}\phi([a;b])\mathrm{Diag}([a;b])\|_\infty}{\|Mx_\lambda\|_\infty}. \end{aligned}$$

The above equations show that we only need to estimate the infinity norm of $\mathbf{D}\phi([a;b])\mathrm{Diag}([a;b])$. Since we have the adjoint operator of $\mathbf{D}\phi([a;b])$ in (3.7), the power method for estimating one norm [36, page 292] can be used to estimate

$m_{\mathcal{L}}^{\mathsf{Reg}}(A, b)$ as shown in Algorithm 2. Also, note that, from (3.6), for $h \in \mathbb{R}^l$,

$$(\mathbf{D}\phi([a; b])\mathsf{Diag}([a; b]))^\top h = \mathsf{Diag}([a; b]) \begin{bmatrix} a(h) \\ D^\top h \end{bmatrix} = \begin{bmatrix} a \odot a(h) \\ b \odot (D^\top h) \end{bmatrix}, \tag{3.10}$$

where '$\odot$' denotes the Hadamard (componentwise) product. In Algorithm 2, sign(a) denotes the vector obtained by applying the sign function to each component of the vector a. We can estimate $c_{\mathcal{L}}^{\mathsf{Reg}}(A, b)$ similarly.

Algorithm 2 The power method for estimating $m_{\mathcal{L}}^{\mathsf{Reg}}(A, b)$

Select initial vector $h = l^{-1}\mathbf{e} \in \mathbb{R}^l$.
for $p = 1, 2, \ldots$ **do**
  Using (3.9), calculate $P(A, \lambda)M^\top h r_\lambda^\top - x_\lambda h^\top D$. From (3.7), compute $a(h)$ and $D^\top h$.
  Using (3.10), denote $\alpha_p = a \odot a(h)$ and $\beta_p = b \odot (D^\top h)$.
  Let $\bar{a}_p = \text{sign}\left(\alpha_p\right)$ and $\bar{b}_p = \text{sign}\left(\beta_p\right)$.
  Compute $a_p = a \odot \bar{a}_p,\ b_p = b \odot \bar{b}_p$.
  Form $A_p = \sum_{i=1}^{k} a_{p,(i)} S_i$, where $a_p = [a_{p,(1)}, a_{p,(2)}, \ldots, a_{p,(k)}]^\top$.
  Using (3.2) and (3.9), compute $z = MP(A, \lambda)(A^\top b_p + A_p^\top r_\lambda - A^\top A_p x_\lambda)$.
  if $\|z\|_\infty \le h^\top z$ **then**
    $\gamma = \left\| \begin{bmatrix} \alpha_p \\ \beta_p \end{bmatrix} \right\|_1$
    quit
  end if
  $h = e_j^{(l)}$, where $|z_j| = \|z\|_\infty$ (smallest such j).
end for
$m_{\mathcal{L}}^{\mathsf{Reg}}(A, b) = \gamma$.

The main computational cost of Algorithm 1 or Algorithm 2 is the computation of solving several nonsingular triangular systems with the coefficient matrices R and $R^\top$. If we have the GSVD of (A, L) available, the computational cost is insignificant compared with the cost of solving the Tikhonov regularized problem. Thus, the estimators can be integrated into a GSVD based Tikhonov solver without compromising the overall computational complexity. Our methods can be readily modified for fast unstructured condition number estimation, which is not considered in [15].

4. Nonlinear structures

In this section, we present the structured condition numbers of matrices with nonlinear structures, namely the Vandermonde matrices and the Cauchy matrices.

4.1. Vandermonde matrices

Let VdM be the class of $m \times n$ Vandermonde matrices. If $V = [v_{ij}] \in \mathsf{VdM}$, then there exists $a = [a_0, a_1, \ldots, a_{n-1}]^\top \in \mathbb{R}^n$ such that, for all $i = 0, 1, \ldots, m-1$ and $j = 0, 1, \ldots, n-1$, $v_{ij} = a_j^i$. We write $V = g(a)$. Let $\Delta a = (\Delta a_0, \Delta a_1, \ldots, \Delta a_{n-1})^\top \in \mathbb{R}^n$ be the perturbation on a. Then we define the first order term ΔV of $g(a + \Delta a) - g(a)$.

Lemma 5 *([10, Lemma 6]). An explicit expression of ΔV is*

$$\Delta V = V_1 \mathsf{Diag}(\Delta a), \quad \textit{where } V_1 = \mathsf{Diag}(c) \begin{bmatrix} \mathbf{0} \\ V(1 : m-1, :) \end{bmatrix}, \ c = [0, 1, \ldots, m-1]^\top.$$

Here $V(1 : m-1, :)$ is the $(m-1) \times n$ submatrix of V consisting of the first $m-1$ rows of V.

Lemma 6. *The Fréchet derivative $\mathbf{D}\phi([a; b])$ of function ϕ defined in (2.7) is*

$$\mathbf{D}\phi([a; b]) = MP(V, \lambda)\left[-V^\top V_1 \mathsf{Diag}(x_\lambda) + \mathsf{Diag}(y), V^\top\right],$$

where $y = V_1^\top r_\lambda$ and $r_\lambda = b - Vx_\lambda$.

Proof. It follows from (3.4) and $\mathsf{Diag}(a)z = \mathsf{Diag}(z)a$ for vectors a and z of the same dimension,

$$\begin{aligned}\phi([a+\Delta a; b+\Delta b]) - \phi([a;b]) &\approx MP(V,\lambda)\left[\left(-x_\lambda^\top \otimes V^\top + I_n \otimes r_\lambda^\top\right)\mathsf{vec}(\Delta V) + V^\top(\Delta b)\right] \\ &= MP(V,\lambda)\left[\left(-x_\lambda^\top \otimes V^\top + (r_\lambda^\top \otimes I_n)\Pi\right)\mathsf{vec}(\Delta V) + V^\top(\Delta b)\right] \\ &= MP(V,\lambda)\left[\left(-V^\top(\Delta V)x_\lambda + (\Delta V)^\top r_\lambda\right) + V^\top(\Delta b)\right] \\ &= MP(V,\lambda)\left[\left(-V^\top V_1 \mathsf{Diag}(\Delta a)x_\lambda + \mathsf{Diag}(\Delta a)V_1^\top r_\lambda\right) + V^\top(\Delta b)\right] \\ &= MP(V,\lambda)\left[-V^\top V_1\mathsf{Diag}(x_\lambda) + \mathsf{Diag}(y),\ V^\top\right]\begin{bmatrix}\Delta a\\ \Delta b\end{bmatrix},\end{aligned}$$

which completes the proof of this lemma. □

From Lemmas 1 and 6, we have the following theorem of structured condition numbers of the Vandermonde matrix.

Theorem 3. *Let $V \in \mathsf{VdM}$, $b \in \mathbb{R}^m$ and $x_\lambda = \left(V^\top V + \lambda L^\top L\right)^{-1} V^\top b = P(V,\lambda)V^\top b$ be the Tikhonov regularized solution of (1.1). Recall that $y = V_1^\top r_\lambda$, then the structured condition numbers of the Vandermonde matrix are:*

$$m_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b) = \frac{\left\|\left|MP(V,\lambda)\left[V^\top V_1\mathsf{Diag}(x_\lambda) - \mathsf{Diag}(y)\right]\right||a| + \left|P(V,\lambda)V^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty},$$

$$c_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b) = \left\|\frac{\left|MP(V,\lambda)\left[V^\top V_1\mathsf{Diag}(x_\lambda) - \mathsf{Diag}(y)\right]\right||a| + \left|P(V,\lambda)V^\top\right|\,|b|}{Mx_\lambda}\right\|_\infty,$$

$$\kappa_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b) = \frac{\left\|MP(V,\lambda)\left[\mathsf{Diag}(y) - V^\top V_1\mathsf{Diag}(x_\lambda),\ V^\top\right]\right\|_2 \left\|\begin{bmatrix}a\\ b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}.$$

In particular, when $l = 1$,

$$\kappa_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b) = \frac{\sqrt{\left\|y \odot \left(P(V,\lambda)M^\top\right) - x_\lambda \odot \left(V_1^\top D_V^\top\right)\right\|_2^2 + \|D_V\|_2^2}\,\left\|\begin{bmatrix}a\\ b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}.$$

Analogous to Proposition 2, the adjoint operator of $\mathbf{D}\phi([a;b])$, using the scalar products $a_1^\top a_2 + b_1^\top b_2$ and $h^\top h$ on $\mathbb{R}^{m+n}$ and $\mathbb{R}^l$ respectively, is

$$\mathbf{D}\phi([a;b])^* :\ h \in \mathbb{R}^l \mapsto [y \odot (P(V,\lambda)M^\top h) - x_\lambda \odot (V_1^\top D_V^\top h)\ D_V^\top h] \in \mathbb{R}^n \times \mathbb{R}^m,$$

where $D_V = MP(V,\lambda)V^\top$. The above expressions can be used to estimate $m_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b)$, $c_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b)$ and $\kappa_{\mathsf{VdM}}^{\mathsf{Reg}}(V,b)$ with lower dimensional input. We can devise algorithms similar to Algorithms 1 and 2 for estimating the condition numbers.

4.2. Cauchy matrices

Let Cauchy be the class of $m \times n$ Cauchy matrices. If $C = [c_{ij}] \in \mathsf{Cauchy}$, then there exist $u = [u_1, u_2, \ldots, u_m]^\top \in \mathbb{R}^m$ and $v = [v_1, v_2, \ldots, v_n]^\top \in \mathbb{R}^n$, with $u_i \neq v_j$ for $i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$ such that, for all $i \le m$ and $j \le n$,

$$c_{ij} = \frac{1}{u_i - v_j}.$$

If $w = [u; v] \in \mathbb{R}^{m+n}$, then $C = g(w)$. Let $\Delta w = [\Delta u; \Delta v] = [\Delta u_1, \ldots, \Delta u_m, \Delta v_1, \ldots, \Delta v_n]^\top \in \mathbb{R}^{m+n}$ be the perturbation on w. The first order term ΔC in $g(w + \Delta w) - g(w)$ is given by [10, Lemma 9]

$$\Delta C \approx \left[\frac{\Delta u_i - \Delta v_j}{(u_i - v_j)^2}\right] = \mathsf{Diag}(\Delta u)C_1 - C_1\mathsf{Diag}(\Delta v) \in \mathbb{R}^{m\times n},$$

where $C_1 = [1/(u_i - v_j)^2] \in \mathbb{R}^{m\times n}$.

Lemma 7. *The Fréchet derivative $\mathbf{D}\phi([w;b])$ of function ϕ defined in (2.7) is given by*

$$\mathbf{D}\phi([w;b]) = MP(C,\lambda)\left[C_u, C_v, C^\top\right],$$

where $C_u = C_1^\top\mathsf{Diag}(r_\lambda) - C^\top\mathsf{Diag}(z_1)$, $C_v = C^\top C_1\mathsf{Diag}(x_\lambda) - \mathsf{Diag}(z_2)$, $z_1 = C_1x_\lambda$, $z_2 = C_1^\top r_\lambda$ and $r_\lambda = b - Cx_\lambda$.

Proof. Following the proof of Lemma 6, we can show that

$$\begin{aligned}&\phi([w+\Delta w; b+\Delta b])-\phi([w; b])\\&\quad = MP(C,\lambda)\left[C_1^\top \mathsf{Diag}(r_\lambda)-C^\top\mathsf{Diag}(z_1),\ C^\top C_1\mathsf{Diag}(x_\lambda)-\mathsf{Diag}(z_2),\ C^\top\right]\begin{bmatrix}\Delta w\\ \Delta b\end{bmatrix}.\end{aligned}$$

Then the Fréchet derivative of ϕ at $[w; b]$ is

$$\mathbf{D}\phi[w; b] = MP(C,\lambda)\left[C_1^\top \mathsf{Diag}(r_\lambda)-C^\top\mathsf{Diag}(z_1), C^\top C_1\mathsf{Diag}(x_\lambda)-\mathsf{Diag}(z_2), C^\top\right].$$

Theorem 4. *Let $C \in$ Cauchy, $b \in \mathbb{R}^m$ and $x_\lambda = \left(C^\top C+\lambda L^\top L\right)^{-1}C^\top b = P(C,\lambda)C^\top b$ be the solution of the Tikhonov regularization problem (1.1), then the structured condition numbers are:*

$$m^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b) = \frac{\left\| |MP(C,\lambda)C_u|\,|u| + |MP(C,\lambda)C_v|\,|v| + \left|MP(C,\lambda)C^\top\right|\,|b|\right\|_\infty}{\|Mx_\lambda\|_\infty},$$

$$c^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b) = \left\| \frac{|MP(C,\lambda)C_u|\,|u| + |MP(C,\lambda)C_v|\,|v| + \left|MP(C,\lambda)C^\top\right|\,|b|}{Mx_\lambda}\right\|_\infty,$$

$$\kappa^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b) = \frac{\left\|MP(C,\lambda)\left[C_u, C_v, C^\top\right]\right\|_2 \left\|\begin{bmatrix}w\\ b\end{bmatrix}\right\|_2}{\|Mx_\lambda\|_2}.$$

In particular, when $l=1$,

$$\kappa^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b) = \frac{\sqrt{t^2+s^2+\|D_C\|_2^2}}{\|Mx_\lambda\|_2}\left\|\begin{bmatrix}u\\ v\\ b\end{bmatrix}\right\|_2,$$

where $t = \left\|r_\lambda \odot \left[C_1 P(C,\lambda)M^\top\right] - z_1\odot (D_C^\top)\right\|_2$ and $s = \left\|x_\lambda \odot (C_1^\top D_C^\top) - z_2 \odot \left[P(C,\lambda)M^\top\right]\right\|_2$.

Similar to the case of the Vandermonde matrix, for the Cauchy matrix, the adjoint operator of $\mathbf{D}\phi([w; b])$, using the scalar products $u_1^\top u_2 + v_1^\top v_2 + b_1^\top b_2$ and $h^\top h$ on $\mathbb{R}^{2m+n}$ and $\mathbb{R}^l$ respectively, is

$$\mathbf{D}\phi([w; b])^* :\ h\in\mathbb{R}^l \mapsto \left[u(h)\ v(h)\ D_V^\top h\right] \in \mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^m,$$

where $u(h) = r_\lambda\odot\left[C_1P(C,\lambda)M^\top h\right] - z_1\odot(D_C^\top h)$, $v(h) = x_\lambda\odot(C_1^\top D_C^\top h) - z_2\odot\left[P(C,\lambda)M^\top h\right]$ and $D_C = MP(C,\lambda)C^\top$.

In particular, when $l=1$, we have

$$\kappa^{\mathsf{Reg}}_{\mathsf{Cauchy}}(V,b) = \frac{\sqrt{t^2+s^2+\|D_C\|_2^2}}{\|Mx_\lambda\|_2}\left\|\begin{bmatrix}u\\ v\\ b\end{bmatrix}\right\|_2,$$

where $t = \left\|r_\lambda \odot \left[C_1 P(C,\lambda)M^\top\right] - z_1\odot (D_C^\top)\right\|_2$ and $s = \left\|x_\lambda \odot (C_1^\top D_C^\top) - z_2 \odot \left[P(C,\lambda)M^\top\right]\right\|_2$.

Using the above expressions, the algorithms similar to Algorithms 1 and 2 for estimating $m^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b)$, $c^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b)$ and $\kappa^{\mathsf{Reg}}_{\mathsf{Cauchy}}(C,b)$ can be obtained.

5. SCE for the Tikhonov regularization problem

In this section we use the SCE to devise algorithms for the condition estimations of the structured and unstructured Tikhonov regularization problem, both the normwise and componentwise cases are considered.

5.1. SCE for normwise perturbations

For the unstructured Tikhonov regularization problem, we are interested in the condition estimation for the function $\psi([A,b])$ at the point $[A,b]$ defined in (2.6). Let $[A\ b]$ be perturbed to $[A+\delta E\ \ b+\delta f]$ in the normal equations (1.3), where $\delta\in\mathbb{R}$, $E\in\mathbb{R}^{m\times n}$ and $f\in\mathbb{R}^m$ and $[E\ f]$ has the Frobenius norm equal to one. According to Section 2.2, we first need to evaluate the directional derivative $\mathbf{D}\psi([A,b];[E,f])$ of $\psi([A,b])$ with respect to $[A,b]$ in the direction $[E,f]$. From the proof of Lemma 3, we have

$$\mathbf{D}\psi([A,b];[E,f]) = P(A,\lambda)\Big(A^\top f + E^\top r_\lambda - A^\top E x_\lambda\Big).$$

When we have the GSVD (1.4) of (A, L), it is easy to deduce that

$$\mathbf{D}\psi([A, b]; [E, f]) = QR^{-1}\begin{bmatrix} (\Sigma^2 + \lambda^2 S^2)^{-1} & \mathbf{0} \\ \mathbf{0} & I_{n-p} \end{bmatrix} R^{-\top}Q^{\top}\left(A^{\top}f + E^{\top}r_{\lambda} - A^{\top}Ex_{\lambda}\right). \tag{5.1}$$

With the above result, we now use the results of Section 2.2 to obtain the SCE-based methods for estimating the condition of the Tikhonov regularization problems. Both the normwise and componentwise perturbations are considered. Algorithm 3 computes an estimation of the normwise condition number. Inputs to the method are the matrices $A \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{p\times n}$, the vector $b \in \mathbb{R}^m$, the computed solution x_λ and the parameter λ. The output is an estimation $\kappa_{\mathrm{SCE}}^{(k)}$ of the normwise condition number $\mathrm{cond}_{\mathrm{Reg}}^{\mathrm{F}}$. The method requires the GSVD (1.4) of (A, L), which is generally computed when solving the Tikhonov regularization problem. The integer $k \geq 1$ refers to the number of perturbations of input data. Note that when $k = 1$, there is no need to orthonormalize the set of vectors in Step 1 of the method. In the following the standard normal distribution is denote by $\mathcal{N}(0, 1)$, and for $B = (b_{ij}) \in \mathbb{R}^{p\times q}$, $|B|^2 = (|b_{ij}|^2) \in \mathbb{R}^{p\times q}$ and $\sqrt{|B|} = (\sqrt{|b_{ij}|}) \in \mathbb{R}^{p\times q}$.

Algorithm 3 SCE for the Tikhonov regularization problem under normwise perturbations

1. Generate matrices $[E_1, f_1], [E_2, f_2], \ldots, [E_k, f_k]$ whose entries are random numbers in $\mathcal{N}(0, 1)$, where $E_i \in \mathbb{R}^{m\times n}$, $f_i \in \mathbb{R}^m$. Use Gram–Schmidt orthogonalization process for the matrix

$$\begin{bmatrix} \mathrm{vec}(E_1) & \mathrm{vec}(E_2) & \cdots & \mathrm{vec}(E_k) \\ f_1 & f_2 & \cdots & f_k \end{bmatrix}$$

and form an orthonormal matrix $[q_1, q_2, \ldots, q_k]$. Each q_i can be converted into the desired matrices $[\widetilde{E}_i, \widetilde{f}_i]$ with the unvec operation.
2. Calculate $\mathbf{D}\psi([A, b]; [\widetilde{E}_i, \widetilde{f}_i])$ by (5.1), $i = 1, 2, \ldots, k$.
3. Compute the absolute condition vector

$$\kappa_{\mathrm{abs}}^{(k)} := \frac{\omega_k}{\omega_p}\sqrt{\left|\mathbf{D}\psi([A, b]; [\widetilde{E}_1, \widetilde{f}_1])\right|^2 + \cdots + \left|\mathbf{D}\psi([A, b]; [\widetilde{E}_k, \widetilde{f}_k])\right|^2}.$$

4. Compute the normwise condition estimation:

$$\kappa_{\mathrm{SCE}}^{(k)} := \frac{\left\|\kappa_{\mathrm{abs}}^{(k)}\right\|_2 \|[A,\ b]\|_F}{\|x_\lambda\|_2}.$$

5.2. SCE for componentwise perturbations

Componentwise perturbations are relative to the magnitudes of the corresponding entries in the input arguments (e.g., the perturbation ΔA satisfies $|\Delta A| \leq \epsilon|A|$, see (2.3)). These perturbations may arise from input error or from rounding error, and hence are the most common perturbations encountered in practice. In fact, most of error bounds in LAPACK are componentwise since the perturbations of input data are componentwise in real world computing, see [46, section 4.3.2] for details. We often want to find the condition of a function with respect to componentwise perturbations on inputs. For the function

$$\psi([A,\ b]) = \left(A^{\top}A + \lambda^2 L^{\top}L\right)^{-1} A^{\top}b.$$

SCE is flexible enough to accurately gauge the sensitivity of matrix functions subject to componentwise perturbations. Define the linear function

$$h([B,\ d]) = [B,\ d] \odot [A,\ b], \quad B \in \mathbb{R}^{m\times n},\ d \in \mathbb{R}^m.$$

Let $\mathcal{E} \in \mathbb{R}^{m\times(n+1)}$ be the matrix of all ones, then $h(\mathcal{E}) = [A\, b]$ and

$$h(\mathcal{E} + [E, f]) = [A,\ b] + h([E, f]).$$

We know that $h([E, f])$ is a componentwise perturbation on $[A, b]$, and h converts a general perturbation $\mathcal{E}$ into componentwise perturbations on $[A, b]$. Therefore, to obtain the sensitivity of the solution with respect to relative perturbations, we simply evaluate the Fréchet derivative of

$$\psi([A,\ b]) = \psi(h(\mathcal{E}))$$

with respect to $\mathcal{E}$ in the direction $[E, f]$, which is

$$\mathbf{D}(\psi \circ h)(\mathcal{E}; [E, f]) = \mathbf{D}\psi(h(\mathcal{E}))\mathbf{D}h(\mathcal{E}; [E, f]) = \mathbf{D}\psi([A,\ b])h([E, f]) = \mathbf{D}\psi([A,\ b]; h([E, f])),$$

since h is linear. Thus, to estimate the condition of the Tikhonov regularization solution x_λ when perturbations are componentwise, we first generate the perturbations E and f and multiply them componentwise by the entries of A and b, respectively. The remaining steps are the same as the corresponding steps in Algorithm 3, as shown in Algorithm 4.

Algorithm 4 SCE for the Tikhonov regularization problem under componentwise perturbations

1. Generate matrices $[E_1,\ f_1], [E_2,\ f_2], \ldots, [E_k,\ f_k]$ whose entries are random numbers in $\mathcal{N}(0, 1)$, where $E_i \in \mathbb{R}^{m\times n}$, $f_i \in \mathbb{R}^m$. Use Gram–Schmidt orthogonalization process for the matrix
$$\begin{bmatrix} \mathrm{vec}(E_1) & \mathrm{vec}(E_2) & \cdots & \mathrm{vec}(E_k) \\ f_1 & f_2 & \cdots & f_k \end{bmatrix}$$
to form an orthonormal matrix $[q_1, q_2, \ldots, q_k]$. Each q_i can be converted into the desired matrices $[E_i\, f_i]$ with the unvec operation.
2. For $i = 1, \ldots, k$, set $[\widetilde{E_i},\ \widetilde{f_i}]$ to the componentwise product of $[A,\ b]$ and $[E_i,\ f_i]$.
3. Calculate $\mathbf{D}\psi([A, b]; [\widetilde{E_i}, \widetilde{f_i}])$ by (5.1), $i = 1, 2, \ldots, k$.
4. Compute the absolute condition vector
$$c_{\mathrm{abs}}^{(k)} := \frac{\omega_k}{\omega_p}\sqrt{\left|\mathbf{D}\psi([A, b];\ [\widetilde{E_1}, \widetilde{f_1}])\right|^2 + \cdots + \left|\mathbf{D}\psi([A, b];\ [\widetilde{E_k}, \widetilde{f_k}])\right|^2}.$$
5. The mixed condition estimation $m_{\mathrm{SCE}}^{(k)}$ and componentwise condition estimation $c_{\mathrm{SCE}}^{(k)}$ are:
$$m_{\mathrm{SCE}}^{(k)} := \frac{\left\|c_{\mathrm{abs}}^{(k)}\right\|_\infty}{\|x_\lambda\|_\infty} \quad \text{and} \quad c_{\mathrm{SCE}}^{(k)} := \left\|\frac{c_{\mathrm{abs}}^{(k)}}{x_\lambda}\right\|_\infty .$$

5.3. SCE for structured perturbations

The SCE also is flexible for the condition estimation for structured Tikhonov regularization problem. We are interested in the condition estimation for the function ϕ defined in (2.7), which defines the general function for structured Tikhonov regularization problem. Because SCE can estimate the condition of the each component of x_λ, we only need to choose $M = I_n$ in (2.7).

The key step in the SCE is the computation of the directional derivative $\mathbf{D}\phi([a; b]; [e; f])$ of $\phi([a; b])$ with respect to $[a; b]$ in the direction $[e; f]$, where $e \in \mathbb{R}^k$ and $f \in \mathbb{R}^m$. We have derived the explicit expressions of the Fréchet derivative $\mathbf{D}\phi([a; b])$ in Lemmas 3, 6 and 7 for a general linear structure, Vandermonde or Cauchy matrix. Based on Lemmas 3, 6 and 7, the three directional derivatives $\mathbf{D}\phi([a; b]; [e; f])$ are:

$$\mathbf{D}\phi([a; b]; [e; f]) = P(A, \lambda)\left(A^\top f + E^\top r_\lambda - A^\top E x_\lambda\right), \quad e = (e_i) \in \mathbb{R}^k,\ E = \sum_{i=1}^k e_i S_i,$$

for linear structures,

$$\mathbf{D}\phi([a; b]; [e; f]) = MP(V, \lambda)\left(\mathrm{Diag}(y)e - V^\top V_1 \mathrm{Diag}(x_\lambda)e + V^\top f\right), \quad e \in \mathbb{R}^n,$$

for Vandermonde matrices, and

$$\mathbf{D}\phi([a; b]; [e; f]) = MP(C, \lambda)\left(C_u e_1 + C_v e_2 + C^\top f\right), \quad e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \in \mathbb{R}^{m+n},$$

for Cauchy matrices, where V_1 is defined in Lemma 5, $y = V_1^\top r_\lambda$, C_u and C_v are defined in Lemma 7. Based on those expressions, we can derive algorithms for structured normwise and componentwise condition estimation. The algorithms are similar to those of Algorithms 3 and 4, thus are omitted here.

5.4. SCE for large scale linear structured Tikhonov regularization problems

In this subsection we will incorporate the SCE into the computation of the solution for large scale linear structured Tikhonov regularization problems, which can be used to obtain posterior error estimations for the computed solution. In many applications, such as image restoration, the proposed mathematical model is given as a large scale Tikhonov regularization problem with the structured coefficient matrix. The structured matrix includes block circulant matrix with circulant blocks (BCCB), block Toeplitz matrix with Toeplitz blocks (BTTB), etc., see [16] and references therein. The large scale structured Tikhonov regularization problem (1.1) is often solved by applying the preconditioned conjugate gradient (PCG) algorithm to its normal equations (1.3), see [47,48]. Effective preconditioners for BTTB Tikhonov regularization problems have been proposed [47,49]. The major computational costs of PCG are matrix–vector multiplications. Because of the structure of the matrix, such as BCCB or BTTB, matrix–vector multiplications can be carried out efficiently via the fast Fourier transform (FFT) [16].

As discussed in the previous section, the key step in the SCE for the large scales linear structured Tikhonov regularization problems is the computation of the directional derivative $\mathbf{D}\phi([a; b]; [e; f])$ of $\phi([a; b])$ with respect to $[a; b]$ in the direction

$[e; f]$, where $e \in \mathbb{R}^k$ and $f \in \mathbb{R}^m$. Based on Lemma 3, the directional derivatives $\mathbf{D}\phi([a; b]; [e; f])$ is equivalent to solving the following positive definite linear system

$$\left(A^\top A + \lambda^2 L^\top L\right) v = A^\top f + E^\top r_\lambda - A^\top E x_\lambda, \quad e = (e_i) \in \mathbb{R}^k, \; E = \sum_{i=1}^{k} e_i S_i. \tag{5.2}$$

When using PCG for solving the large scale linear structured Tikhonov regularization problems (1.1), a preconditioner is constructed and used to solve (5.2) via the PCG. In (5.2), when A is BTTB, E is also BTTB. Thus the right side of (5.2) can be computed efficiently via the FFT. The SCE algorithm for large scale linear structured Tikhonov regularization problems is given in Algorithm 5.

Algorithm 5 SCE for the large scales linear structured Tikhonov regularization problem under componentwise perturbations

1. Generate matrices $[E_1, f_1], [E_2, f_2], \ldots, [E_k, f_k]$ whose entries are random numbers in $\mathcal{N}(0, 1)$, where $E_i \in \mathbb{R}^{m\times n}$, $f_i \in \mathbb{R}^m$. Use Gram–Schmidt orthogonalization process for the matrix
$$\begin{bmatrix} \mathrm{vec}(E_1) & \mathrm{vec}(E_2) & \cdots & \mathrm{vec}(E_k) \\ f_1 & f_2 & \cdots & f_k \end{bmatrix}$$
to form an orthonormal matrix $[q_1, q_2, \ldots, q_k]$. Each q_i can be converted into the desired matrices $[E_i\, f_i]$ with the unvec operation.
2. For $i = 1, \ldots, k$, set $[\widetilde{E}_i, \widetilde{f}_i]$ to the componentwise product of $[A, b]$ and $[E_i, f_i]$.
3. Implement PCG to solve
$$\left(A^\top A + \lambda^2 L^\top L\right) v_i = A^\top \widetilde{f} + \widetilde{E}_i^\top r_\lambda - A^\top \widetilde{E}_i x_\lambda, \quad e = (e_i) \in \mathbb{R}^k, \; E = \sum_{i=1}^{k} e_i S_i. \tag{5.3}$$
4. Compute the absolute condition vector
$$c_{\mathrm{abs}}^{(k)} := \frac{\omega_k}{\omega_p}\sqrt{|v_1|^2 + \cdots + |v_k|^2}.$$
5. The mixed condition estimation $m_{\mathrm{SCE}}^{(k)}$ and componentwise condition estimation $c_{\mathrm{SCE}}^{(k)}$ are:
$$m_{\mathrm{SCE}}^{(k)} := \frac{\left\| c_{\mathrm{abs}}^{(k)} \right\|_\infty}{\|x_\lambda\|_\infty} \quad \text{and} \quad c_{\mathrm{SCE}}^{(k)} := \left\| \frac{c_{\mathrm{abs}}^{(k)}}{x_\lambda} \right\|_\infty .$$

6. Numerical examples

In this section, we demonstrate our test results of some numerical examples to illustrate structured condition numbers and condition estimations presented in the previous sections. We performed our numerical experiments on a machine with Intel i5 4590 @3.3 GHz CPU, 8 GB RAM and 1 TB hard driver running Windows 7 professional. All the computations were carried out using MATLAB 8.1 with machine precision about 2.2×10^{-16} and the REGULARIZATION TOOLS package [50].

For a structured matrix A, which is determined by the vector $a \in \mathbb{R}^k$, we generated the perturbed matrix $\widehat{A}$ as follows. For $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^m$, let $[s; f]$ be a random vector whose entries are uniformly distributed in the open interval $(-1, 1)$, where $s \in \mathbb{R}^k$ and $f \in \mathbb{R}^m$, the perturbations on a and b are respectively

$$\Delta a_i = \varepsilon s_i a_i, \qquad \Delta b_j = \varepsilon f_j b_j, \tag{6.1}$$

then $\widehat{A} = g(a + \Delta a)$ and $\widehat{b} = b + \Delta b$. In our experiments, we set $\varepsilon = 10^{-8}$.

REGULARIZATION TOOLS package [50] includes four methods for determining the Tikhonov regularization parameter. For the Tikhonov regularization with continuous regularization parameter, the L-curve is a continuous curve as a parametric plot of the discrete smoothing (semi) norm $\|Lx_\lambda\|_2$ versus the corresponding residual norm $\|Ax_\lambda - b\|_2$, with λ as the parameter. The corner of the L-curve appears for regularization parameters close to the optimal parameter that balances the regularization errors and perturbation errors in x_λ, which is the basis for the L-curve criterion for choosing the regularization parameter. In addition to the L-curve criterion for parameter-choice, a variety of parameter-choice strategies have been proposed, such as the *discrepancy principle* (Discrep. pr.) [51], *generalized cross-validation* (GCV) [52] and the quasi-optimality criterion (Quasi-opt) [51].

The Tikhonov regularization solution x_λ was computed by the Matlab function tikhonov corresponding to A, b in REGULARIZATION TOOLS package with different regularization parameters chosen by the four classical criteria or by a predefined value. The perturbed solution y_λ was obtained in the similar way to x_λ, where y_λ corresponds to $\widehat{A}$ and $\widehat{b}$. Denote the error $\Delta x_\lambda = y_\lambda - x_\lambda$.

Table 1
$L = I_5, M = I_5$.

	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$6.3937 \cdot 10^{-4}$	$4.9988 \cdot 10^{-4}$	$4.9988 \cdot 10^{-4}$	$4.9988 \cdot 10^{-4}$
$\frac{\|\Delta x_\lambda\|_2}{\epsilon\|x_\lambda\|_2}$	$9.1464 \cdot 10^{-1}$	1.4820	1.1081	3.1918
$\frac{\|\Delta x_\lambda\|_\infty}{\epsilon\|x_\lambda\|_\infty}$	$9.3703 \cdot 10^{-1}$	1.6158	1.3755	3.7359
$\epsilon^{-1}\left\|\frac{\Delta x_\lambda}{x_\lambda}\right\|_\infty$	$2.29 \cdot 10^6$	$6.26 \cdot 10^6$	$1.55 \cdot 10^6$	$4.74 \cdot 10^6$
$\mathrm{cond}^{\mathrm{F}}_{\mathrm{Reg}}$ in (2.5)	$3.3961 \cdot 10^3$	$4.4761 \cdot 10^3$	$4.4761 \cdot 10^3$	$4.4761 \cdot 10^3$
$\mathrm{m}_{\mathrm{Reg}}$ in (2.3)	$1.5204 \cdot 10^3$	$2.0035 \cdot 10^3$	$2.0035 \cdot 10^3$	$2.0035 \cdot 10^3$
$\mathrm{c}_{\mathrm{Reg}}$ in (2.4)	$9.8192 \cdot 10^6$	$1.6064 \cdot 10^7$	$1.6064 \cdot 10^7$	$1.6064 \cdot 10^7$
$\kappa_{\mathsf{SymToep}}(A, b)$	$1.0047 \cdot 10^3$	$1.3242 \cdot 10^3$	$1.3242 \cdot 10^3$	$1.3242 \cdot 10^3$
$m_{\mathsf{SymToep}}(A, b)$	4.3765	4.4971	4.4971	4.4971
$c_{\mathsf{SymToep}}(A, b)$	$9.8143 \cdot 10^6$	$1.6056 \cdot 10^7$	$1.6056 \cdot 10^7$	$1.6056 \cdot 10^7$

Table 2
$L = I_5, M = e_3^\top$.

	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$6.39 \cdot 10^{-4}$	$5.00 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$
$\frac{\|M\Delta x_\lambda\|_2}{\|Mx_\lambda\|_2}$	$1.3741 \cdot 10^{-2}$	$2.2481 \cdot 10^{-2}$	$2.1749 \cdot 10^{-2}$	$2.1749 \cdot 10^{-2}$
$\frac{\|M\Delta x_\lambda\|_\infty}{\|Mx_\lambda\|_\infty}$	$1.3741 \cdot 10^{-2}$	$2.2481 \cdot 10^{-2}$	$2.1749 \cdot 10^{-2}$	$2.1749 \cdot 10^{-2}$
$\varepsilon\kappa_{\mathsf{SymToep}}(A, b)$	$1.7780 \cdot 10^{-1}$	$2.9088 \cdot 10^{-1}$	$2.8141 \cdot 10^{-1}$	$2.8141 \cdot 10^{-1}$
$\varepsilon m_{\mathsf{SymToep}}(A, b)$	$4.9096 \cdot 10^{-2}$	$8.0320 \cdot 10^{-2}$	$7.7705 \cdot 10^{-2}$	$7.7705 \cdot 10^{-2}$

We compare the structured condition numbers with unstructured ones for various Tikhonov regularization parameters in the following examples.

Example 1 *([30])*. Let $A = g(a)$ be a 5×5 symmetric Toeplitz matrix defined by

$$A = g(a) = \begin{bmatrix} 0 & 0 & 1+h & -1 & 1 \\ 0 & 0 & 0 & 1+h & -1 \\ 1+h & 0 & 0 & 0 & 1+h \\ -1 & 1+h & 0 & 0 & 0 \\ 1 & -1 & 1+h & 0 & 0 \end{bmatrix}, \quad a = \begin{bmatrix} 0 \\ 0 \\ 1+h \\ -1 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ h \\ 2(1+h) \\ h \\ 0 \end{bmatrix}$$

for $h = 10^{-3}$.

The above matrix A is a square symmetric Toeplitz matrix and $g(a)$ is a square symmetric Toeplitz matrix whose first column is a. We can choose the basis $Z_i = g(e_i)$, $i = 1, 2, \ldots, 5$, so that $T = \sum_{i=1}^5 a_i Z_i$. From Theorem 1, with the Z_i, we can get the expressions for $m_{\mathsf{SymToep}}(A, b)$, $c_{\mathsf{SymToep}}(A, b)$ and $\kappa_{\mathsf{SymToep}}(A, b)$. The relative errors and condition numbers are shown in Table 1, where $M = I_5$.

Table 1 shows that the structured mixed condition numbers are much smaller than the corresponding unstructured ones, which give tight linear perturbation bounds. Both $c_{\mathsf{SymToep}}(A, b)$ and $\kappa_{\mathsf{SymToep}}(A, b)$ are smaller than the corresponding unstructured ones.

In Table 2, we choose $M = e_3^\top$, where e_3 is the third column of I_5. In this case,

$$\frac{\|M\Delta x_\lambda\|_\infty}{\|Mx_\lambda\|_\infty} = \left\|\frac{M\Delta x_\lambda}{Mx_\lambda}\right\|_\infty, \qquad m_{\mathsf{SymToep}}(A, b) = c_{\mathsf{SymToep}}(A, b).$$

We compare the true relative perturbation bounds with the first-order asymptotic perturbation bounds given by $m_{\mathsf{SymToep}}(A, b)$ and $\kappa_{\mathsf{SymToep}}(A, b)$.

From Table 2, we can see the quantities $\varepsilon m_{\mathsf{SymToep}}(A, b)$ give tighter perturbation bounds than $\varepsilon\kappa_{\mathsf{SymToep}}(A, b)$, since they have the same order as that of the true relative perturbation bounds.

Table 3 shows the results from different choices of M, i.e., $M = e_1^\top$ and $M = e_3^\top$. For example, if we choose $M = e_1^\top$, then we are interested in the conditioning of the first component of x_λ. We display the values of $m_{\mathsf{SymToep}}(A, b)$ and $\kappa_{\mathsf{SymToep}}(A, b)$. From Table 3, we can say that the first component of x_λ has better conditioning than the third one.

At the end of this example, we use Algorithms 1 and 2 to illustrate the effectiveness of the power method. We set the maximal number of iterations to 10 in Algorithm 1. The estimated condition numbers in Algorithms 1 and 2 are denoted by $\kappa^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$ and $m^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$ respectively.

From Table 4, we can say that $\kappa^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$ and $m^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$ give good estimations for this specific A, L and b, especially $m^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$ gives better estimation.

Table 3
$L = I_5, M = e_1^\top$ and $M = e_3^\top$.

$M = e_1^\top$	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$6.39 \cdot 10^{-4}$	$5.00 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$
$\kappa_{\mathsf{SymToep}}(A, b)$	$1.5887 \cdot 10^3$	$2.0941 \cdot 10^3$	$2.0594 \cdot 10^3$	$2.0594 \cdot 10^3$
$m_{\mathsf{SymToep}}(A, b)$	$7.6056 \cdot 10^2$	$1.0022 \cdot 10^3$	$9.8567 \cdot 10^2$	$9.8567 \cdot 10^2$
$M = e_3^\top$	Discrep. pr.	L-curve	GCV	Quasi-opt
Value of λ	$6.39 \cdot 10^{-4}$	$5.00 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$
$\kappa_{\mathsf{SymToep}}(A, b)$	$1.7780 \cdot 10^7$	$2.9088 \cdot 10^7$	$2.8141 \cdot 10^7$	$2.8141 \cdot 10^7$
$m_{\mathsf{SymToep}}(A, b)$	$4.9096 \cdot 10^6$	$8.0320 \cdot 10^6$	$7.7705 \cdot 10^6$	$7.7705 \cdot 10^6$

Table 4
$L = I_5, M = I_5$ and $M = e_1^\top$.

$M = I_5$	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$6.39 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$
$\kappa_{\mathsf{SymToep}}(A, b)$	$1.5878783796 \cdot 10^3$	$2.0931466345 \cdot 10^3$	$2.0584876249 \cdot 10^3$	$2.0584876249 \cdot 10^3$
$\kappa^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$	$7.5891802517 \cdot 10^2$	$1.0002500616 \cdot 10^3$	$9.8369583257 \cdot 10^2$	$9.8369583257 \cdot 10^2$
$m_{\mathsf{SymToep}}(A, b)$	$7.6117517197 \cdot 10^2$	$1.0027483753 \cdot 10^3$	$9.8617760333 \cdot 10^2$	$9.8617760333 \cdot 10^2$
$m^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$	$7.6055529470 \cdot 10^2$	$1.0022493745 \cdot 10^3$	$9.8567031098 \cdot 10^2$	$9.8567031098 \cdot 10^2$
$M = e_1^\top$	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$6.39 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$	$5.08 \cdot 10^{-4}$
$\kappa_{\mathsf{SymToep}}(A, b)$	$1.5886924101 \cdot 10^3$	$2.0940875560 \cdot 10^3$	$2.0594198505 \cdot 10^3$	$2.0594198505 \cdot 10^3$
$\kappa^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$	$3.7980426062 \cdot 10^2$	$5.0050048168 \cdot 10^2$	$4.9222129601 \cdot 10^2$	$4.9222129601 \cdot 10^2$
$m_{\mathsf{SymToep}}(A, b)$	$7.6055560483 \cdot 10^2$	$1.0022496243 \cdot 10^3$	$9.8567056493 \cdot 10^2$	$9.8567056493 \cdot 10^2$
$m^{\mathrm{Est}}_{\mathsf{SymToep}}(A, b)$	$7.6055529470 \cdot 10^2$	$1.0022493745 \cdot 10^3$	$9.8567031098 \cdot 10^2$	$9.8567031098 \cdot 10^2$

Table 5
$L = I_6, M = I_6$.

	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$7.5918 \cdot 10^{-4}$	$2.5002 \cdot 10^{-4}$	$2.5002 \cdot 10^{-4}$	0.0017
$\frac{\Vert\Delta x_\lambda\Vert_2}{\epsilon\Vert x_\lambda\Vert_2}$	1.0902	2.9510	2.6014	1.3163
$\frac{\Vert\Delta x_\lambda\Vert_\infty}{\epsilon\Vert x_\lambda\Vert_\infty}$	1.3264	4.2237	3.2460	2.0419
$\epsilon^{-1}\left\Vert \frac{\Delta x_\lambda}{x_\lambda}\right\Vert_\infty$	$4.39 \cdot 10^6$	$2.516 \cdot 10^7$	$2.645 \cdot 10^7$	$1.31 \cdot 10^6$
$\mathrm{cond}^{\mathrm{F}}_{\mathrm{Reg}}$ in (2.5)	$2.2310 \cdot 10^3$	$1.1401 \cdot 10^4$	$1.1401 \cdot 10^4$	$4.6222 \cdot 10^2$
$\mathrm{m}_{\mathrm{Reg}}$ in (2.3)	$7.8426 \cdot 10^2$	$4.0032 \cdot 10^3$	$4.0032 \cdot 10^3$	$1.6347 \cdot 10^2$
$\mathrm{c}_{\mathrm{Reg}}$ in (2.4)	$1.3230 \cdot 10^7$	$1.0238 \cdot 10^8$	$1.0238 \cdot 10^8$	$2.6208 \cdot 10^6$
$\kappa_{\mathsf{Hankel}}(A, b)$	$1.0372 \cdot 10^3$	$5.2922 \cdot 10^3$	$5.2922 \cdot 10^3$	$2.1648 \cdot 10^2$
$m_{\mathsf{Hankel}}(A, b)$	3.4999	5.1247	5.1247	3.5000
$c_{\mathsf{Hankel}}(A, b)$	$1.1576 \cdot 10^7$	$8.9578 \cdot 10^7$	$8.9578 \cdot 10^7$	$2.2931 \cdot 10^6$

Example 2 ([30]). Let $A = g([c; r])$ be the 6×6 Hankel matrix defined by

$$A = g([c;r]) = \begin{bmatrix} h & 1 & 1 & -1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 & 1 & 0 \end{bmatrix}, \quad c = \begin{bmatrix} h \\ 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \; r = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \; b = \begin{bmatrix} h \\ 2 \\ 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}$$

for $h = 10^{-3}$, where c is the first column of A and r is the last row of A.

We can choose the basis $Y_1 = g([e_1; \mathbf{0}]), \ldots, Y_5 = g([e_5; \mathbf{0}]), Y_6 = g([e_6; e_1]), Y_7 = g([\mathbf{0}; e_2]), \ldots, Y_{11} = g([\mathbf{0}; e_6])$, so that $A = \sum_{k=1}^{11} a_k Y_k$. Again, from Theorem 1 with the Y_i, the expressions for $m_{\mathsf{Hankel}}(A, b)$, $c_{\mathsf{Hankel}}(A, b)$ and $\kappa_{\mathsf{Hankel}}(A, b)$ can be obtained.

From Table 5, we conclude that the structured mixed condition numbers can be much smaller than the corresponding unstructured condition numbers. Structured mixed condition numbers also give sharp perturbation bounds. The forward errors obtained by multiplying the structured mixed condition numbers with 10^{-8} are of the same order as that of the exact errors.

Table 6
$L = I_{10}, M = I_{10}$.

	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$1.36 \cdot 10^{-5}$	$6.31 \cdot 10^{-5}$	5.69	5.69
$\frac{\\|\Delta x_\lambda\\|_2}{\epsilon\\|x_\lambda\\|_2}$	2.7484	3.007	1.4131	1.4131
$\frac{\\|\Delta x_\lambda\\|_\infty}{\epsilon\\|x_\lambda\\|_\infty}$	3.4057	2.6762	2.7502	2.7502
$\epsilon^{-1}\left\\|\frac{\Delta x_\lambda}{x_\lambda}\right\\|_\infty$	$2.2796 \cdot 10$	$2.1046 \cdot 10$	5.3054	5.3054
$\mathrm{cond}^{F}_{\mathrm{Reg}}$ in (2.5)	$4.8637 \cdot 10^6$	$1.7445 \cdot 10^6$	$2.6028 \cdot 10$	$2.6028 \cdot 10$
$\mathrm{m}_{\mathrm{Reg}}$ in (2.3)	$5.5816 \cdot 10^4$	$2.3085 \cdot 10^4$	$3.9279 \cdot 10$	$3.9279 \cdot 10$
$\mathrm{c}_{\mathrm{Reg}}$ in (2.4)	$5.0645 \cdot 10^5$	$6.3061 \cdot 10^4$	$8.5328 \cdot 10$	$8.5328 \cdot 10$
$\kappa_{\mathrm{VdM}}(A, b)$	$4.7123 \cdot 10$	$5.2721 \cdot 10$	$1.2499 \cdot 10$	$1.2499 \cdot 10$
$m_{\mathrm{VdM}}(A, b)$	$1.4219 \cdot 10$	$1.4076 \cdot 10$	$2.0828 \cdot 10$	$2.0828 \cdot 10$
$c_{\mathrm{VdM}}(A, b)$	$1.8428 \cdot 10^2$	$6.1557 \cdot 10$	$4.0178 \cdot 10$	$4.0178 \cdot 10$

Table 7
$L = I_8, M = I_8$.

	Discrep. pr.	L-curve	GCV	Quasi-opt
λ	$2.46 \cdot 10^{-10}$	$6.97 \cdot 10^{-7}$	1.72	1.72
$\frac{\\|\Delta x_\lambda\\|_2}{\epsilon\\|x_\lambda\\|_2}$	2.7472	5.5724	3.6995	3.6995
$\frac{\\|\Delta x_\lambda\\|_\infty}{\epsilon\\|x_\lambda\\|_\infty}$	2.5007	6.3752	2.2306	2.2306
$\epsilon^{-1}\left\\|\frac{\Delta x_\lambda}{x_\lambda}\right\\|_\infty$	$1.0489 \cdot 10$	$1.4879 \cdot 10$	$1.4934 \cdot 10^2$	$1.4934 \cdot 10^2$
$\mathrm{cond}^{F}_{\mathrm{Reg}}$ in (2.5)	$2.9150 \cdot 10^8$	$4.5472 \cdot 10^7$	$2.7426 \cdot 10$	$2.7426 \cdot 10$
$\mathrm{m}_{\mathrm{Reg}}$ in (2.3)	$2.8775 \cdot 10^7$	$8.0534 \cdot 10^6$	$1.0465 \cdot 10$	$1.0465 \cdot 10$
$\mathrm{c}_{\mathrm{Reg}}$ in (2.4)	$1.0584 \cdot 10^8$	$4.6471 \cdot 10^7$	$8.7045 \cdot 10^2$	$8.7045 \cdot 10^2$
$\kappa_{\mathrm{Cauchy}}(A, b)$	$3.8644 \cdot 10$	$4.1630 \cdot 10$	$5.4144 \cdot 10$	$5.4144 \cdot 10$
$m_{\mathrm{Cauchy}}(A, b)$	$1.8131 \cdot 10$	$2.2502 \cdot 10$	7.2573	7.2573
$c_{\mathrm{Cauchy}}(A, b)$	$3.4802 \cdot 10^2$	$8.0663 \cdot 10$	$3.9086 \cdot 10^2$	$3.9086 \cdot 10^2$

Example 3 (*[53]*). Let $V = g(a)$ be a 25×10 Vandermonde matrix whose (i, j)-entry is

$$V_{ij} = \left(\frac{j}{10}\right)^{i-1}, \qquad a = \left[\frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10}, 1\right]^\top, \qquad b \in \mathbb{R}^{25} \text{ with } b_{2k-1} = -1, b_{2k} = 1.$$

In Table 6, when λ is small, the problem is ill-conditioned under unstructured perturbations. The structured condition numbers are much smaller than the unstructured ones. The perturbation bounds given by the structured condition numbers coincide with the relative errors from the columns Discrep. pr. and L-curve. When we use GCV and Quasi-opt to compute the regularization parameters λ, which is equal to 5.69 in this example, the problem is well-conditioned. The structured condition numbers have the same order as the unstructured ones. Both of them give sharp perturbation bounds.

Example 4 (*[10]*). Let $A = g(a)$ be a 10×8 Cauchy matrix whose (i, j)-entry is

$$a_{ij} = \frac{1}{i + j - 1}, \qquad a = [u_1, \ldots, u_{10}, v_1, v_2, \ldots, v_8]^\top, \text{ with } u_i = i,\ v_j = 1 - j,$$
$$b = [1, -1, 1, -1, 1, -1, 1, -1, 1, -1]^\top.$$

Then A is a rectangular Hilbert matrix.

From the second and third columns (Discrep. pr. and L-curve) of Table 7, we can see that the structured condition numbers are much smaller than the unstructured one and they give sharp perturbation bounds. The first-order unstructured asymptotic perturbation bounds severely overestimate the true relative errors in both normwise and componentwise cases for the numerical examples of the discrepancy principle and L-curve methods. As for the last two columns, since the regularization parameter λ is large, the problems are well-conditioned. The structured condition numbers are of the same order as that of the unstructured ones.

In the rest of this section, we will show our test results on the proposed SCE algorithms for the conditioning estimation of the Tikhonov regularization solution. Both the unstructured and structured Tikhonov regularization cases are considered. For the unstructured Tikhonov regularization, the test problems originally come from the discretization of the Fredholm integral equations of the first kind, and they lead to discrete ill-posed problems. We use the test problems included in the REGULARIZATION TOOLS package. In these numerical experiments, a discrete ill-posed problem $Ax = b$ using one of the many built-in test problems is firstly generated; then the white noise is added to the right-hand side with a perturbation e

whose elements are normally distributed with zero mean and standard deviation chosen such that the noise-to-signal ratio $\|e\|_2/\|b\|_2 = 10^{-4}$, thus producing a more 'realistic' problem.

We generated the perturbations $\Delta A = \varepsilon \times (E \odot A)$ and $\Delta b = \varepsilon \times (f \odot b)$, where $\varepsilon = 10^{-8}$, E and f are random matrices whose entries are uniformly distributed in the open interval $(-1, 1)$.

To measure the effectiveness of the estimators, we define the over-estimation ratios

$$r_\kappa := \frac{\kappa_{\mathrm{SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda\|_2/\|x_\lambda\|_2}, \qquad r_m := \frac{m_{\mathrm{SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda\|_\infty/\|x_\lambda\|_\infty}, \qquad r_c := \frac{c_{\mathrm{SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda/x_\lambda\|_\infty},$$

where k is the subspace dimension in Algorithms 3 and 4, $\kappa_{\mathrm{SCE}}^{(k)}$, $m_{\mathrm{SCE}}^{(k)}$ and $c_{\mathrm{SCE}}^{(k)}$ are the outputs from Algorithms 3 and 4. Typically the ratios in $(0.1,\ 10)$ are acceptable [36, Chapter 19].

For unstructured Tikhonov regularization problems, we test the SCE for several classical ill-posed problems included in the REGULARIZATION TOOLS package: deriv2, shaw and wing. Those three examples give square coefficient matrices A and right-hand side vectors b. For the matrix L in (1.1), we chose the identity matrix and

$$L_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}$$

which approximates the first derivative operator. We adopted the following four values of the regularization parameter λ:

$$0.1,\ 6\cdot 10^{-2},\ 1.7\cdot 10^{-3},\ 1.7\cdot 10^{-4}.$$

In Table 8, we report the numerical results on the ratios r_κ, r_m and r_c for examples with various dimensions and choices of L. The table shows that the mixed condition estimation $m_{\mathrm{SCE}}^{(k)}$ reflects the true error bound accurately, while the componentwise condition estimation $c_{\mathrm{SCE}}^{(k)}$ gives accurate error bounds for most cases and the normwise condition estimation fails to reflect the true error bound accurately. Specifically, r_m are between 0.94 and 16.33, implying that the condition estimation $m_{\mathrm{SCE}}^{(k)}$ can be considered reliable [36]. The values of the componentwise condition estimation $c_{\mathrm{SCE}}^{(k)}$ are within $(16.54,\ 88.19)$ except for the case shaw, where $n = 512$, $L = I_n$, $k = 3$ with all choices of λ, the case wing, where $n = 256$, $L = I_n$, $k = 3$, $\lambda = 0.1,\ 6\cdot 10^{-2}$ and the case shaw, where $n = 256$, $L = L_1$, $k = 5$, $\lambda = 1.7\cdot 10^{-3}$, indicating that the componentwise condition estimation $c_{\mathrm{SCE}}^{(k)}$ is effective for most cases. For the normwise condition estimation $\kappa_{\mathrm{SCE}}^{(k)}$, most of the values of r_κ are of order $\mathcal{O}(10^2)$, and even some of them are of order $\mathcal{O}(10^3)$, showing that the normwise condition estimation overly estimates for most of the cases.

For structured Tikhonov regularization cases, we tested the following Toeplitz matrix:

$$A = (a_{i-j}) \in \mathbb{R}^{m\times n}, \quad a_{i-j} = \rho^{|i-j|}.$$

We used the right-hand side $b = \mathbf{e} \in \mathbb{R}^m$ and $\rho = 0.99999$. This Toeplitz matrix is also a symmetric matrix. The Tikhonov regularization parameter is determined by the four classical criteria. As discussed in Section 5.3, similar to Algorithms 3 and 4, we can use the SCE to obtain the structured normwise, mixed and componentwise condition estimations denoted by $\kappa_{\mathrm{SymToep,SCE}}^{(k)}$, $m_{\mathrm{SymToep,SCE}}^{(k)}$ and $c_{\mathrm{SymToep,SCE}}^{(k)}$ respectively. The perturbations Δa on a and Δb on b were generated as in (6.1). As in the previous example, let the overestimate ratios be defined by

$$r_\kappa^{\mathrm{SymToep}} := \frac{\kappa_{\mathrm{SymToep,SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda\|_2/\|x_\lambda\|_2}, \qquad r_m^{\mathrm{SymToep}} := \frac{m_{\mathrm{SymToep,SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda\|_\infty/\|x_\lambda\|_\infty}, \qquad r_c^{\mathrm{SymToep}} := \frac{c_{\mathrm{SymToep,SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda/x_\lambda\|_\infty},$$

which measure the reliability of the condition estimators. In this example, we always set $L = I_n$.

In Table 9, except for two cases, the values of all the three ratios are of order $\mathcal{O}(10)$, implying that the SCE structured condition estimations are reliable.

In the rest of this section, we will show the effectiveness of the SCE for the large scale linear structured Tikhonov regularization problems arising from image restoration. Let an $m \times m$ block Toeplitz matrix with $n \times n$ Toeplitz blocks (BTTB(m, n)) be of the form

$$A = [A_{(i-j)}]_{i,j=1}^m$$

with $A_{(j)} = [a_{j,\gamma-\delta}]_{\gamma,\delta=1}^n$ being Toeplitz matrices of order n, $j = 0, \pm 1, \ldots, \pm(m-1)$. Consider the image restoration problem [47]

$$b = Ax_* + \eta, \tag{6.2}$$

where A is the BTTB(m, n) matrix with the diagonals being given by

$$a_{j,l} = \begin{cases} \exp\{-0.1j^2 - 0.1l^2\}, & -8 \le j, l \le 8, \\ 0, & \text{otherwise}, \end{cases} \tag{6.3}$$

Table 8
SCE for the Tikhonov regularization problem.

	r_κ	r_m	r_c
deriv2, $n = 64,\ L = I_n,\ k = 5$			
$\lambda = 0.1$	$7.7533 \cdot 10^2$	2.0847	$5.1984 \cdot 10$
$\lambda = 6 \cdot 10^{-2}$	$7.6234 \cdot 10^2$	2.1284	$4.7450 \cdot 10$
$\lambda = 1.7 \cdot 10^{-3}$	$6.7427 \cdot 10^2$	1.2970	$2.4927 \cdot 10$
$\lambda = 1.7 \cdot 10^{-4}$	$9.2294 \cdot 10^2$	1.1437	$3.9081 \cdot 10$
deriv2, $n = 64,\ L = L_1,\ k = 5$			
$\lambda = 0.1$	$9.2479 \cdot 10^2$	2.2288	$3.8756 \cdot 10$
$\lambda = 6 \cdot 10^{-2}$	$9.2057 \cdot 10^2$	2.5900	$3.6441 \cdot 10$
$\lambda = 1.7 \cdot 10^{-3}$	$6.2048 \cdot 10^2$	1.0458	$2.9864 \cdot 10$
$\lambda = 1.7 \cdot 10^{-4}$	$9.2686 \cdot 10^2$	1.5548	$3.4564 \cdot 10$
wing, $n = 128,\ L = L_1,\ k = 5$			
$\lambda = 0.1$	$2.3753 \cdot 10^3$	$1.6337 \cdot 10$	$1.6544 \cdot 10$
$\lambda = 6 \cdot 10^{-2}$	$1.4154 \cdot 10^3$	7.5189	$1.5910 \cdot 10$
$\lambda = 1.7 \cdot 10^{-3}$	$1.3068 \cdot 10^3$	1.6301	$2.2538 \cdot 10$
$\lambda = 1.7 \cdot 10^{-4}$	$9.6524 \cdot 10^2$	1.3917	$8.1316 \cdot 10$
wing, $n = 256,\ L = I_n,\ k = 3$			
$\lambda = 0.1$	$6.1606 \cdot 10^2$	1.0112	$2.6130 \cdot 10^2$
$\lambda = 6 \cdot 10^{-2}$	$5.4330 \cdot 10^2$	1.3470	$1.8994 \cdot 10^2$
$\lambda = 1.7 \cdot 10^{-3}$	$1.0126 \cdot 10^3$	1.8223	$6.2581 \cdot 10$
$\lambda = 1.7 \cdot 10^{-4}$	$1.0923 \cdot 10^3$	2.1715	$7.2018 \cdot 10$
shaw, $n = 512,\ L = I_n,\ k = 3$			
$\lambda = 0.1$	$1.6632 \cdot 10^2$	2.4174	$1.7467 \cdot 10^2$
$\lambda = 6 \cdot 10^{-2}$	$1.8554 \cdot 10^2$	2.8948	$3.7301 \cdot 10^2$
$\lambda = 1.7 \cdot 10^{-3}$	$1.0707 \cdot 10^2$	3.1782	$3.7127 \cdot 10^2$
$\lambda = 1.7 \cdot 10^{-4}$	$2.9693 \cdot 10^2$	1.6429	$1.8405 \cdot 10^2$
shaw, $n = 256,\ L = L_1,\ k = 5$			
$\lambda = 0.1$	$2.8617 \cdot 10$	2.9029	$8.8186 \cdot 10$
$\lambda = 6 \cdot 10^{-2}$	$3.1563 \cdot 10$	3.0425	$8.3908 \cdot 10$
$\lambda = 1.7 \cdot 10^{-3}$	$3.0292 \cdot 10$	1.0652	$2.0672 \cdot 10^2$
$\lambda = 1.7 \cdot 10^{-4}$	$1.1406 \cdot 10^2$	$9.4368 \cdot 10^{-1}$	$3.0579 \cdot 10$

Table 9
SCE for the structured Tikhonov regularization problem.

	$r_\kappa^{\mathrm{SymToep}}$	r_m^{SymToep}	r_c^{SymToep}
$m = 100,\ n = 50,\ k = 3$			
Discrep. pr. $\lambda = 2.21$	$1.3427 \cdot 10$	1.2354	$4.0474 \cdot 10$
L-curve $\lambda = 6.19 \cdot 10^{-2}$	$1.8396 \cdot 10$	1.7149	$2.9893 \cdot 10$
GCV $\lambda = 1.35 \cdot 10^{-4}$	$1.8113 \cdot 10$	1.5819	$2.0690 \cdot 10$
Quasi-opt $\lambda = 7.48 \cdot 10^{-1}$	$1.1606 \cdot 10$	$8.6429 \cdot 10^{-1}$	$5.7865 \cdot 10$
$m = 300,\ n = 200,\ k = 3$			
Discrep. pr. $\lambda = 4.71$	$7.6537 \cdot 10$	4.1397	$4.8346 \cdot 10$
L-curve $\lambda = 1.10 \cdot 10^{-1}$	$5.6455 \cdot 10$	2.4765	$7.8375 \cdot 10$
GCV $\lambda = 3.22 \cdot 10^{-4}$	$3.2584 \cdot 10$	1.4440	$8.8256 \cdot 10$
Quasi-opt $\lambda = 4.49$	$7.1689 \cdot 10$	3.6997	$5.4007 \cdot 10$
$m = 500,\ n = 300,\ k = 3$			
Discrep. pr. $\lambda = 1.49 \cdot 10^{-2}$	$5.8732 \cdot 10$	1.9115	$1.0526 \cdot 10^2$
L-curve $\lambda = 1.03$	$7.8779 \cdot 10$	3.2648	$9.2462 \cdot 10$
GCV $\lambda = 5.66 \cdot 10^{-4}$	$4.3191 \cdot 10$	1.6125	$1.2357 \cdot 10$
Quasi-opt $\lambda = 9.25$	$1.0777 \cdot 10^2$	4.3658	$6.8813 \cdot 10$

x_* is the $mn \times 1$ vector obtained from the true image by row ordering, b presents the blurred, noisy image, and η presents noise. The simulation image is of sizes 256×256, see Fig. 1. The observed image b is constructed by $b = Ax + \eta$, where the noise vector η is a vector with random entries chosen from normal distribution with mean 0 (η is scaled so that $\|\eta\|_2/\|Ax_*\|_2 = 10^{-3}$). Lin et al. [49] proposed BTTB preconditioner for the following normal equations which also appeared in [47]:

$$(A^\top A + \lambda^2 L^\top L)x = A^\top b, \tag{6.4}$$

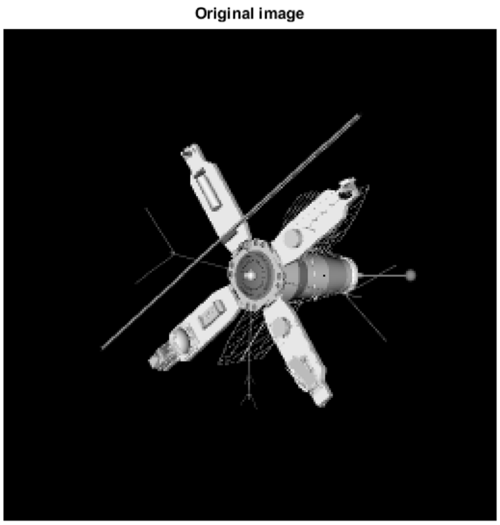

Fig. 1. Original image.

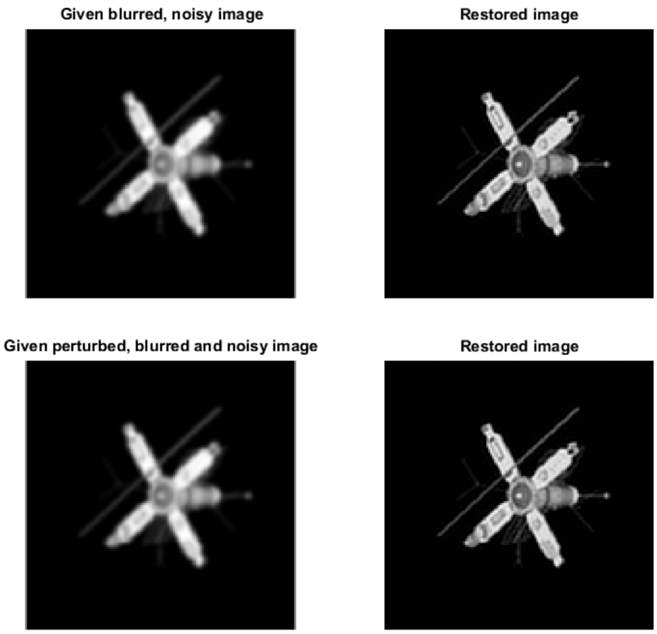

Fig. 2. Perturbed, blurred and noisy images, restored images.

where $L = I_{mn}$ and $\lambda = 0.1$. Lin et al. adopted the PCG with the initial guess being set to the zero vector and the stopping criterion being set to $\|r_\lambda^{(q)}\|_2 / \|r_\lambda^{(0)}\|_2 \leq 10^{-7}$, where $r_\lambda^{(q)}$ is the normal equation residual vector at the qth iteration in the PCG. The perturbations Δa on a and Δb on b were generated as in (6.1). Thus, $\Delta A = g(\Delta a)$ is also BTTB. The top row of Fig. 2 displays the blurred/noisy image and the restored image obtained by using the PCG to solve (6.4). After adding the structured perturbations to A and b, we display the perturbed, blurred and noisy image on the bottom-left of Fig. 2. The restored image shown in the bottom-right of Fig. 2 was computed by solving the following perturbed normal equations via the PCG using the BTTB preconditioner [49],

$$((A + \Delta A)^\top (A + \Delta A)) + \lambda^2 L^\top L(x_\lambda + \Delta x_\lambda) = (A + \Delta A)^\top (b + \Delta b).$$

We use Algorithm 5 to estimate the structured mixed and componentwise condition numbers for (6.4). In the third step of Algorithm 5, we adopt the PCG with the same preconditioner proposed in [49] to compute the directional derivative with respect to the structured perturbation on A and b. The initial guess and stopping criterion are the same as the PCG in [49].

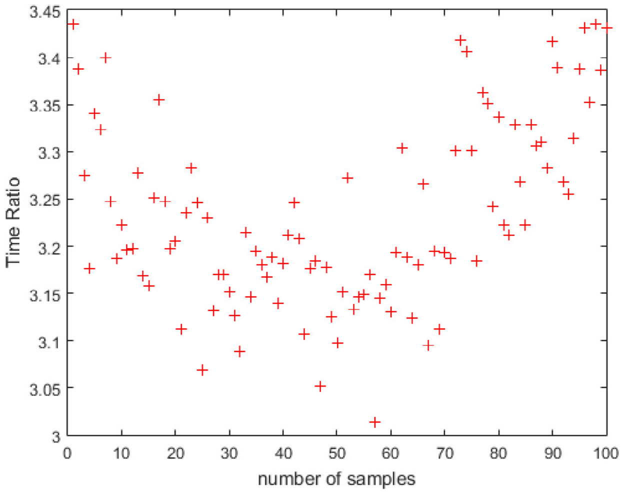

Fig. 3. Time ratios over 100 samples.

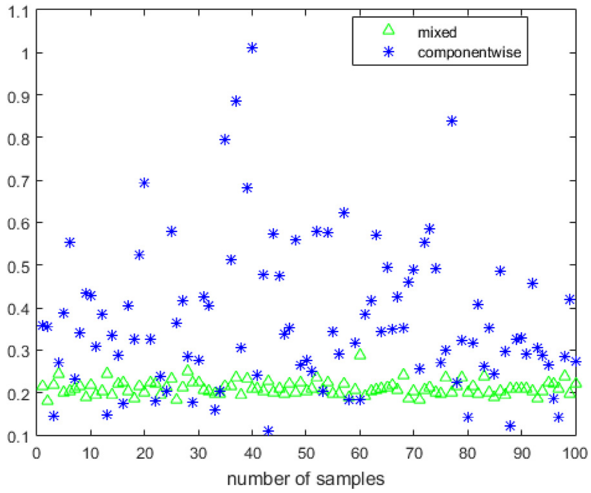

Fig. 4. Overestimations over 100 samples.

We always choose $k = 3$ in Algorithm 5. As in the previous example, let the overestimate ratios be defined by

$$r_m^{\mathrm{BTTB}} := \frac{m_{\mathrm{BTTB,SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda\|_\infty / \|x_\lambda\|_\infty}, \qquad r_c^{\mathrm{BTTB}} := \frac{c_{\mathrm{BTTB,SCE}}^{(k)} \cdot \varepsilon}{\|\Delta x_\lambda / x_\lambda\|_\infty},$$

which measure the reliability of the condition estimators. The CPU time ratio is defined by

$$r_{time} = \frac{t_2}{t_1},$$

where t_1 is the CPU time for solving (6.4) using the PCG in [49] and t_2 is the CPU time for Algorithm 5. We generated 100 samples of Δa and Δb and plotted r_{time}, r_m^{BTTB} and r_c^{BTTB} in Figs. 3 and 4, respectively. From Fig. 3, we can see that, the time ratios r_{time} are around 3.25 since $k = 3$, that is, we need solve (5.3) via the PCG for three times and extra flops to obtain the orthogonal bases for the random matrix through the Gram–Schmidt orthogonalization process in the first step in Algorithm 5. By observing Fig. 4, it is easy to see that r_m^{BTTB} and r_c^{BTTB} are effective because they are in $(0.1,\ 1)$, and they can give the posterior error estimations for the restored image.

Now we consider testing the SCE with a different L in (6.4), where L is given in [49, Example 3]. Let $X_* = [x_{i,j}^*]_{i=1,\ldots,m}^{j=1,\ldots,n}$ be the image to be restored and L be defined as

$$\|LX_*\|_2^2 = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(|x_{i+1,j}^* - x_{i,j}^*|^2 + |x_{i,j+1}^* - x_{i,j}^*|^2) + \sum_{j=1}^{n-1}|x_{m,j+1}^* - x_{m,j}^*|^2 + \sum_{i=1}^{m-1}|x_{i+1,n}^* - x_{i,n}^*|^2. \tag{6.5}$$

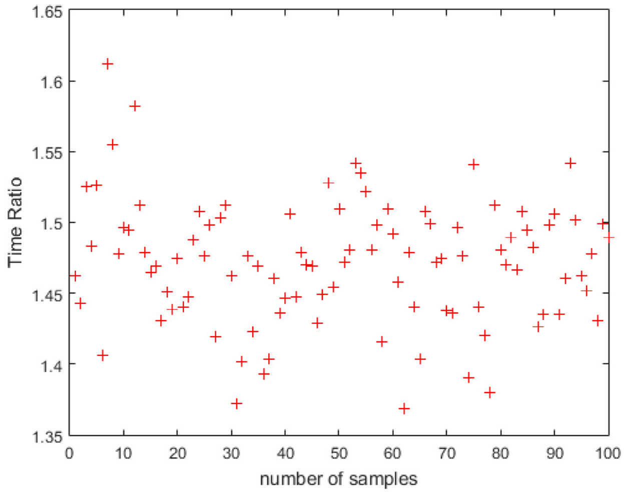

Fig. 5. Time ratios over 100 samples.

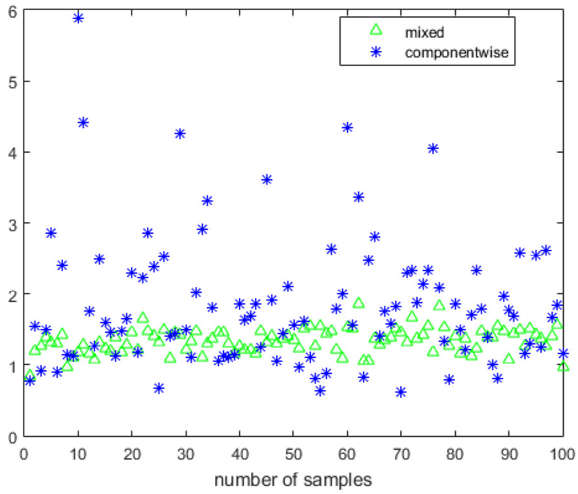

Fig. 6. Overestimations over 100 samples.

The regularization parameter is set to $\lambda = 0.09$ as in [49]. Again, we generated 100 samples of the perturbations Δa and Δb and computed r_{time}, r_m^{BTTB} and r_c^{BTTB}. From Fig. 5, the time ratios are around 1.5, because when we use the PCG to solve (5.3) in Algorithm 5 it usually needs only half of the number of the steps for solving (6.4) via the PCG. Since we choose $k = 3$ in Algorithm 5, the total times for Algorithm 5 are around 1.5 times of the CPU time for solving (6.4) via the PCG. Similarly from Fig. 6, we can see the error estimations based the SCE are reliable.

7. Concluding remarks

In this paper, we introduce the structured condition numbers for the structured Tikhonov regularization problem and derive their exact expressions without the Kronecker product. The structures considered include linear structures, such as Toeplitz and Hankel, and nonlinear structures, such as Vandermonde and Cauchy. We show that our structured condition numbers are smaller than unstructured condition numbers for Toeplitz and Hankel structures. Applying the power method, we devise fast algorithms for estimating the unstructured and structured condition number under normwise and componentwise perturbations, that can be integrated into a GSVD based Tikhonov regularization solver. We also investigate the SCE for estimating structured condition numbers for small/medium scale Tikhonov regularization problems. The SCE for large scale structured Tikhonov regularization problems in image restoration is also considered. The numerical examples show that our structured mixed condition numbers give tight error bounds and the proposed condition estimations are reliable and efficient. A possible future research topic is to study the ratio between the structured and unstructured condition numbers for the structured Tikhonov regularization problem.

Acknowledgments

We thank Principal Editor Prof. Michael Ng and the referees' constructive comments which help us to improve this paper greatly, especially we thank one referee for pointing out that we could consider the SCE for large scale BTTB Tikhonov regularization problems in image restoration. Moreover, the first author thanks Prof. Fu-Rong Lin for sharing the MATLAB codes in [49] to test the effectiveness of the SCE in image restoration. The first author is supported by the National Natural Science Foundation of China under grant 11001045. Part of this work was finished when the author visited Shanghai Key Laboratory of Contemporary Applied Mathematics and McMaster University in 2014. The second author is supported by the National Natural Science Foundation of China under grant 11271084. The third author is partially supported by Natural Science and Engineering Council (NSERC) of Canada under grant RGPIN-2014-04252 and Shanghai Key Laboratory of Contemporary Applied Mathematics.

References

[1] A.N. Tikhonov, Regularization of incorrectly posed problems, Sov. Math. Dokl. 4 (1963) 1624–1627.
[2] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
[3] P.C. Hansen, Regularization GSVD and truncated GSVD, BIT 29 (1989) 491–504.
[4] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1997.
[5] C.F. Van Loan, Generalizing the singular value decomposition, SIAM J. Numer. Anal. 13 (1976) 76–83.
[6] P.C. Hansen, Discrete Inverse Problems: Insight and Algorithms, SIAM, Philadelphia, 2010.
[7] M. Arioli, M. Baboulin, S. Gratton, Partial condition number for linear least squares problems, SIAM J. Matrix Anal. Appl. 29 (2007) 413–433.
[8] M. Baboulin, J. Dongarra, S. Gratton, J. Langou, Computing the conditioning of the components of a linear least-squares solution, Numer. Linear Algebra Appl. 16 (2009) 517–533.
[9] M. Baboulin, S. Gratton, Using dual techniques to derive componentwise and mixed condition numbers for a linear function of a linear least squares solution, BIT 49 (2009) 3–19.
[10] F. Cucker, H. Diao, Mixed and componentwise condition numbers for rectangular structured matrices, Calcolo 44 (2007) 89–115.
[11] F. Cucker, H. Diao, Y. Wei, On mixed and componentwise condition numbers for Moore–Penrose inverse and linear least squares problems, Math. Comp. 76 (2007) 947–963.
[12] M. Gulliksson, P. Wedin, Y. Wei, Perturbation identities for regularized Tikhonov inverses and weighted pseudoinverses, BIT 40 (2000) 513–523.
[13] P.C. Hansen, Perturbation bounds for discrete Tikhonov regularization, Inverse Problems 5 (1989) 41–44.
[14] A.N. Malyshev, A unified theory of conditioning for linear least squares and Tikhonov regularization solutions, SIAM J. Matrix Anal. Appl. 24 (2003) 1186–1196.
[15] D. Chu, L. Lin, R.C.E. Tan, Y. Wei, Condition numbers and perturbation analysis for the Tikhonov regularization of discrete ill-posed problems, Numer. Linear Algebra Appl. 18 (2011) 87–103.
[16] R. Chan, X. Jin, An Introduction to Iterative Toeplitz Solvers, SIAM, Philadelphia, PA, 2007.
[17] M. Ng, Iterative Methods for Toeplitz Systems, Oxford University Press, 2004.
[18] M. Benzi, M.K. Ng, Preconditioned iterative methods for weighted Toeplitz least squares problems, SIAM J. Matrix Anal. Appl. 27 (2006) 1106–1124.
[19] R. Chan, J.G. Nagy, R.J. Plemmons, Circulant preconditioned Toeplitz least squares iterations, SIAM J. Matrix Anal. Appl. 15 (1994) 80–97.
[20] P.C. Hansen, Deconvolution and regularization with Toeplitz matrices, Numer. Algorithms 29 (2002) 323–378.
[21] L. Eldén, An efficient algorithm for the regularization of ill-conditioned least squares problems with triangular Toeplitz matrix, SIAM J. Sci. Stat. Comput. 5 (1984) 229–236.
[22] H. Park, L. Eldén, Schur-type methods for solving least squares problems with Toeplitz structure, SIAM J. Sci. Comput. 22 (2000) 406–430.
[23] H. Park, L. Eldén, Stability analysis and fast algorithms for triangularization of Toeplitz matrices, Numer. Math. 76 (1997) 383–402.
[24] S. Bartels, D.J. Higham, The structured sensitivity of Vandermonde-like systems, Numer. Math. 62 (1992) 17–33.
[25] A. Böttcher, S. Grudsky, Structured condition numbers of large Toeplitz matrices are rarely better than usual condition numbers, Numer. Linear Algebra Appl. 12 (2005) 95–102.
[26] E. Bozzo, D. Fasino, A priori estimates on the structured conditioning of Cauchy and Vandermonde matrices, in: Numerical Methods for Structured Matrices and Applications, in: Oper. Theory Adv. Appl., vol. 199, Birkhäuser Verlag, Basel, 2010, pp. 203–220.
[27] I. Gohberg, I. Koltracht, Mixed, componentwise, and structured condition numbers, SIAM J. Matrix Anal. Appl. 14 (1993) 688–704.
[28] D.J. Higham, N.J. Higham, Backward error and condition of structured linear systems, SIAM J. Matrix Anal. Appl. 13 (1992) 162–175.
[29] S.M. Rump, Structured perturbations. Part I: Normwise distances, SIAM J. Matrix Anal. Appl. 25 (2003) 1–30.
[30] S.M. Rump, Structured perturbations. Part II: Componentwise distances, SIAM J. Matrix Anal. Appl. 25 (2003) 31–56.
[31] H. Xiang, Y. Wei, Structured mixed and componentwise condition numbers of some structured matrices, J. Comput. Appl. Math. 202 (2007) 217–229.
[32] W. Xu, Y. Wei, S. Qiao, Condition numbers for structured least squares, BIT 46 (2006) 203–225.
[33] W.W. Hager, Condition estimates, SIAM J. Sci. Stat. Comput. 5 (1984) 311–316.
[34] N.J. Higham, FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation (Algorithm 674), ACM Trans. Math. Software 14 (1988) 381–396.
[35] N.J. Higham, Experience with a matrix norm estimator, SIAM J. Sci. Stat. Comput. 10 (1990) 804–809.
[36] N.J. Higham, Accuracy and Stability of Numerical Algorithms, second ed., SIAM, Philadelphia, 2002.
[37] C.S. Kenney, A.J. Laub, Small-sample statistical condition estimates for general matrix functions, SIAM J. Sci. Comput. 15 (1994) 36–61.
[38] C.S. Kenney, A.J. Laub, M.S. Reese, Statistical condition estimation for linear systems, SIAM J. Sci. Comput. 19 (2) (1998) 566–583.
[39] A.J. Laub, J. Xia, Applications of statistical condition estimation to the solution of linear systems, Numer. Linear Algebra Appl. 15 (2008) 489–513.
[40] C.S. Kenney, A.J. Laub, M.S. Reese, Statistical condition estimation for the linear least squares, SIAM J. Matrix Anal. Appl. 19 (1998) 906–923.
[41] T. Gudmundsson, C.S. Kenney, A.J. Laub, Small-sample statistical estimates for the sensitivity of eigenvalue problems, SIAM J. Matrix Anal. Appl. 18 (4) (1997) 868–886.
[42] A.J. Laub, J. Xia, Fast condition estimation for a class of structured eigenvalue problems, SIAM J. Matrix Anal. Appl. 30 (2009) 1658–1676.
[43] A.J. Laub, J. Xia, Statistical condition estimation for the roots of polynomials, SIAM J. Sci. Comput. 31 (2008) 624–643.
[44] A. Graham, Kronecker Products and Matrix Calculus with Application, Wiley, New York, 1981.
[45] S.M. Rump, H. Sekigawa, The ratio between the Toeplitz and the unstructured condition number, Oper. Theory Adv. Appl. 199 (2009) 397–419.
[46] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen (Eds.), LAPACK Users' Guide, third ed., SIAM, 1999.
[47] R. Chan, J.G. Nagy, R.J. Plemmons, FFT-based preconditioners for Toeplitz–Block least squares problems, SIAM J. Numer. Anal. 30 (1993) 1740–1768.
[48] R. Chan, K.P. Ng, Conjugate gradient method for Toeplitz systems, SIAM Rev. 38 (1996) 427–482.
[49] F. Lin, D. Zhang, BTTB preconditioners for BTTB least squares problems, Linear Algebra Appl. 434 (11) (2001) 2285–2295.
[50] P.C. Hansen, Regularization tools version 4.0 for matlab 7.3, Numer. Algorithms 46 (2007) 189–194.
[51] V.A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer, New York, 1984.
[52] G. Wahba, Spline Models for Observational Data, in: CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59, SIAM, Philadelphia, 1990.
[53] W. Gautschi, How (un)stable are Vandermonde systems? in: R. Wong (Ed.), Asymptotic and Computational Analysis, in: Lecture Notes in Pure and Applied Mathematics, vol. 124, Marcel Dekker, New York, 1990, pp. 193–210.