Structured condition numbers of structured Tikhonov regularization problem and their estimations

Huai-An Diao\textsuperscript{a}, Yimin Wei\textsuperscript{b}, Sanzheng Qiao\textsuperscript{c,*}

\textsuperscript{a} School of Mathematics and Statistics, Northeast Normal University, No. 5268 Renmin Street, Chang Chun 130024, PR China
\textsuperscript{b} School of Mathematics & Shanghai Key Laboratory of Contemporary Applied Mathematics, Fudan University, Shanghai 200433, PR China
\textsuperscript{c} Department of Computing and Software, McMaster University, Hamilton, Ontario L8S4K1, Canada

**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.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

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

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

where \(\lambda\) 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/© 2016 Elsevier B.V. All rights reserved.
For the regularization problem (1.1), to ensure the uniqueness of the solution for any $\lambda > 0$, we always assume that $\text{rank}(L) = p \leq n \leq m$ and rank $\begin{bmatrix} A \end{bmatrix} = n$ (cf. [2, Section 5]). The regularization problem (1.1) can be rewritten in the matrix form

$$\min_x \left\| A^{\top} A + \lambda^2 L^{\top} L \right\|_F x - \begin{bmatrix} b \\ 0 \end{bmatrix},$$

(1.2)

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

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

(1.3)

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

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

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 rank $(L) = p$ and rank $\begin{bmatrix} A \end{bmatrix} = n$, the GSVD of $(A, L)$ is given by the pair of factorizations

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n-p} \end{bmatrix} RQ^\top$$

and

$$L = V \begin{bmatrix} S & 0 \\ 0 & I_{r} \end{bmatrix} R^Q,$$

(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 an $n$-by-$n$ upper triangular and nonsingular, and $\Sigma$ and $S$ are $p \times p$ diagonal matrices: $\Sigma = \text{Diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ and $S = \text{Diag}(\mu_1, \mu_2, \ldots, \mu_p)$ 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 $\Sigma^2 + S^2 = I_n$. Then the generalized singular values $\gamma_i$ 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]

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

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 $\Sigma^\top$ is the Moore–Penrose inverse of $\Sigma$ [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 quantiles 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 $M \in \mathbb{R}^{m \times n}$ 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 \( B \in \mathbb{R}^{m' \times n} \), then \( [A; B] \) denotes the matrix obtained by putting \( A \) and \( B \) side by side. The symbol \(^T\) denotes matrix transpose, \( \| \cdot \|_2 \) is the spectral norm, \( \| \cdot \|_F \) is the Frobenius norm and \( \| \cdot \|_\infty \) is the infinity norm. The matrix \( \text{Diag}(d) \in \mathbb{R}^{n \times n} \) 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 \( vec(A) \in \mathbb{R}^{mn} \) by \( vec(A) = [a_{11}, a_{12}, \ldots, a_{mn}]^T \), where \( A = (a_{ij}) \) with \( a_{ij} \in \mathbb{R} \), \( i = 1,2,\ldots,n \). The unvec operation is defined as \( A = \text{unvec}(v) \) which sets the entries of \( A \) to \( a_{ij} = v_{(i-1)n+j} \) for \( v = [v_1, v_2, \ldots, v_{mn}] \in \mathbb{R}^{1 \times mn} \). We define a permutation matrix \( \mathcal{P} \) of order \( mn \) so that \( \mathcal{P}(\text{vec}(A)) = \text{vec}(A^\top) \). 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 \( |A| \leq |B| \) means that \( |a_{ij}| \leq |b_{ij}| \). For the Kronecker product, we recall the following properties, which can be found in [44].

\[
(A \otimes B)^\top = A^\top \otimes B^\top, \quad |A \otimes B| = |A| \otimes |B|, \quad \text{vec}(AXB) = (B^\top \otimes A) \text{vec}(X),
\]

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 \( S(a, \varepsilon) = \{ x \in \mathbb{R}^p \mid \| x - a \|_2 \leq \varepsilon \| a \|_2 \} \) and \( T(a, \varepsilon) = \{ x \in \mathbb{R}^p \mid \| x - a \|_2 \leq \varepsilon \} \). For a function \( F : \mathbb{R}^p \to \mathbb{R}^q \), we denote \( \text{Dom}(F) \) 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 \( \text{Dom}(F) \subset \mathbb{R}^p \). Let \( a \in \text{Dom}(F) \) such that \( a \neq 0 \) and \( F(a) \neq 0 \).

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

\[
m(F, a) = \lim_{\varepsilon \to 0} \sup_{x \in S(a, \varepsilon) \setminus \{a\}} \frac{\|F(x) - F(a)\|_2}{\|F(a)\|_2} \frac{1}{\|a\|_2} = \frac{\|DF(a)\|_2}{\|F(a)\|_2}.
\]

where \( DF(a) \) 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 \( f_j(a) \neq 0 \) for \( j = 1,2,\ldots,q \). Then the componentwise condition number of \( F \) at \( a \) is

\[
c(F, a) = \lim_{\varepsilon \to 0} \sup_{x \in S(a, \varepsilon) \setminus \{a\}} \frac{d(F(x), F(a))}{d(x,a)(DF(a))^\top a} = \frac{\|DF(a)\|_2}{\|F(a)\|_2}.
\]

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

\[
\kappa(F, a) = \lim_{\varepsilon \to 0} \sup_{x \in S(a, \varepsilon) \setminus \{a\}} \frac{\|F(x) - F(a)\|_2}{\|F(a)\|_2} \frac{\|a\|_2}{\|x - a\|_2} = \frac{\|DF(a)\|_2}{\|F(a)\|_2} \frac{\|a\|_2}{\|x - a\|_2}.
\]

In the following we assume that \( \Delta A \) and \( \Delta b \) are perturbations to \( A \) and \( b \) respectively, which satisfy \( \text{rank}(\begin{bmatrix} A & \Delta A \\ \lambda L & \lambda I \end{bmatrix}) = n \).

The perturbed counterpart of the problem (1.1) and its normal equations (1.3) are, respectively,

\[
\min_{x,\Delta A} \left\{ \| (A + \Delta A)(x + \Delta x) - (b + \Delta b) \|_2^2 + \lambda \| L(x + \Delta x) \|_2^2 \right\} = 0,
\]

(2.1)
and

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

Then the perturbed Tikhonov regularized solution is given by

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

Denoting

\[ P(A, \lambda) = (A^\top A + \lambda^2 L^\top L)^{-1}. \]

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

\[
m_{\text{Reg}} = \lim_{\epsilon \to 0} \sup_{\|\Delta A\| < \epsilon} \frac{\|\Delta x\|_\infty}{\|x_\lambda\|_\infty} = \left\| \frac{|H(A, b)| \text{vec}(|A|) + |P(A, \lambda)A^\top| |b|}{\|x_\lambda\|_\infty} \right\|_\infty, \tag{2.3}
\]

\[
c_{\text{Reg}} = \lim_{\epsilon \to 0} \sup_{\|\Delta A\| < \epsilon} \left\{ \frac{\|\Delta x\|_\infty}{\|x_\lambda\|_\infty} = \left\| \frac{|H(A, b)| \text{vec}(|A|) + |P(A, \lambda)A^\top| |b|}{\|x_\lambda\|_\infty} \right\|_\infty, \right. \tag{2.4}
\]

\[
\text{cond}_{\text{Reg}} = \lim_{\epsilon \to 0} \sup_{\|\Delta A\| < \epsilon} \left\{ \frac{\|\Delta x\|_2}{\|x_\lambda\|_2} = \left\| \frac{|H(A, b)| \text{vec}(|A|) + |P(A, \lambda)A^\top| |b|}{\|x_\lambda\|_2} \right\|_2 \right. \left\| [A, b] \right\|_F \right\} \tag{2.5}
\]

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

If we define a mapping

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

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

\[
m_{\text{Reg}} := m(\psi, [A, b]), \quad c_{\text{Reg}} := c(\psi, [A, b]), \quad \text{cond}_{\text{Reg}} := \kappa(\psi, [A, b]).
\]

When the coefficient matrix \(A\) in (1.1) has some structures, such as Toeplitz, it is reasonable to assume that the perturbation \(\Delta A\) in (2.1) has the same structure of \(A\). Then \(\Delta A\) 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 \(\delta\) be the set of structured matrices under consideration and \(a\) the vector representing a structured matrix \(A\), then we define a mapping

\[
g : a \in \mathbb{R}^k \mapsto A \in \delta.
\]

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

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

where \(M \in \mathbb{R}^{l \times n}, l \leq n\), is general. In particular, when \(M = e_i^\top\), the \(i\)th column of the identity matrix, then we are interested in some particular component of \(x_\lambda\).

Let \(\Delta a\) be the perturbation on \(a\), then the structured perturbation matrix \(\Delta A\) 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,

\[
m_{\text{Reg}} (A, b) := m(\phi, [a; b]) = \lim_{\epsilon \to 0} \sup_{\|\Delta a\| < \epsilon} \frac{\|M\Delta x\|_\infty}{\|Mx_\lambda\|_\infty},
\]

\[
c_{\text{Reg}} (A, b) := c(\phi, [a; b]) = \lim_{\epsilon \to 0} \sup_{\|\Delta a\| < \epsilon} \left\{ \frac{1}{\epsilon} \right\} \frac{\|M\Delta x\|_2}{Mx_\lambda},
\]

\[
\kappa_{\text{Reg}} (A, b) := \kappa(\phi, [a; b]) = \lim_{\epsilon \to 0} \sup_{\|\Delta a\| < \epsilon} \frac{\|M\Delta x\|_2}{\|Mx_\lambda\|_2},
\]

where \(\Delta x\) 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, when we restrict δ to be \(\mathbb{R}^{m \times n}\), which are generalizations of (2.3)–(2.5), as follows

\[
m_{\text{reg}}(A, b) = \lim_{\epsilon \to 0} \sup_{\|M\Delta x\|_\infty \leq \epsilon} \frac{1}{\|M\Delta x\|_\infty} \|M\Delta x\|_\infty, \quad c_{\text{reg}}(A, b) = \lim_{\epsilon \to 0} \sup_{\|M\Delta x\|_\infty \leq \epsilon} \frac{1}{\epsilon} \|M\Delta x\|_\infty^2.
\]

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

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

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

\[
f(x + \delta d) - f(x) = \delta (Df(x))^\top d + \mathcal{O}(\delta^2),
\]

where \(\delta\) is a small scalar, \(\|d\|_2 = 1\) and \(Df(x)\) is the Fréchet derivative of \(f\) at \(x\). Note that the quantity \((Df(x))^\top d\) (denoted by \(Df(x; d)\)) 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 \(\delta\),

\[
|f(x + \delta d) - f(x)| \approx \delta Df(x; d),
\]

then the local sensitivity can be measured by \(\|Df(x)\|_2\). The condition numbers of \(f\) at \(x\) are mainly determined by the norm of the gradient \(Df(x)\) [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 \(E(Df(x; d)/\omega_p)\) is \(\|Df(x)\|_2\), where \(\omega_p\) is the Wallis factor. In practice, the Wallis factor can be approximated accurately [37] by

\[
\omega_p \approx \sqrt{\frac{2}{\pi (p - 1)}}.
\]

Therefore, we can use

\[
v = \frac{|Df(x; d)|}{\omega_p}
\]

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

\[
\text{Prob} \left( \frac{\|Df(x)\|_2}{\gamma} \leq v \leq \gamma \|Df(x)\|_2 \right) \geq 1 - \frac{2}{\pi \gamma^2} + \mathcal{O} \left( \frac{1}{\gamma^2} \right),
\]

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

\[
v(k) = \frac{\omega_t}{\omega_p} \sqrt{\|Df(x; d_1)\|^2 + |Df(x; d_2)|^2 + \cdots + |Df(x; d_t)|^2},
\]

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 \(v(2)\) is given by

\[
\text{Prob} \left( \frac{\|\nabla f(x)\|_2}{\gamma} \leq v(2) \leq \gamma \|\nabla f(x)\|_2 \right) \approx 1 - \frac{\pi}{4\gamma^2}, \quad \gamma > 1.
\]

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 \text{vec} and \text{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 $a = [a_1, a_2, \ldots, a_k]^T \in \mathbb{R}^k$ such that

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

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 $D\phi([a; b])$ of function $\phi$ defined in (2.7) is given by

$$D\phi([a; b]) = MP(A, \lambda) [v_1, v_2, \ldots, v_k, A^T],$$

where $v_i = -A^T S_i x_i + S_i^T r_i$ for $i = 1, 2, \ldots, k$.

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

$$A = (A^T A + \lambda^2 L^T L) + (A^T (\Delta A) + (\Delta A)^T A) + (\Delta A)^T (\Delta A)$$

$$= (A^T A + \lambda^2 L^T L) [I_n + P(A, \lambda) (A^T (\Delta A) + (\Delta A)^T A) + P(A, \lambda) ((\Delta A)^T (\Delta A))].$$

If $\|\Delta A\|$ is sufficiently small, then $\|P(A, \lambda) (A^T (\Delta A) + (\Delta A)^T A + (\Delta A)^T (\Delta A))\| < 1$, from Lemma 2, $A$ is nonsingular and its inverse

$$A^{-1} = \left[I_n + P(A, \lambda) (A^T (\Delta A) + (\Delta A)^T A) + P(A, \lambda) ((\Delta A)^T (\Delta A))\right]^{-1} P(A, \lambda)$$

$$= P(A, \lambda) - P(A, \lambda) (A^T (\Delta A) + (\Delta A)^T A) P(A, \lambda) + O(\|\Delta A\|^2),$$

since $\|A^T (\Delta A) + (\Delta A)^T A\| = O(\|\Delta A\|)$ and $\|(\Delta A)^T (\Delta A)\| = O(\|\Delta A\|^2)$. From (2.2), (3.3) and $x_i = P(A, \lambda) A^T b$, after some algebraic manipulation, we have

$$\Delta x = P(A, \lambda) \left[A^T (\Delta b) + (\Delta b)^T (b - Ax_i) - A^T (\Delta A) x_i\right] + O(\|\Delta A\|^2 + O(\|\Delta A\| \|\Delta b\|)).$$

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

$$\Delta x \approx P(A, \lambda) \left[A^T (\Delta b) + (\Delta b)^T (b - Ax_i) - A^T (\Delta A) x_i\right]$$

$$= P(A, \lambda) \left[(-x_i^T \otimes A^T) + (r_i^T \otimes I_n) IT\right] vec(\Delta A) + A^T (\Delta b)$$

$$= P(A, \lambda) \left[(-x_i^T \otimes A^T) + (I_n \otimes r_i^T)\right] vec(\Delta A) + A^T (\Delta b),$$

recalling that $r_i = b - Ax_i$.

Since $\Delta A$ 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 $\Delta a = [\Delta a_1, \Delta a_2, \ldots, \Delta a_k]^T$. From (3.4), we have

$$\phi([a + \Delta a; b + \Delta b]) - \phi([a; b]) \approx MP(A, \lambda) \left[(-x_i^T \otimes A^T) + (I_n \otimes r_i^T)\right] vec(S_1), \ldots, vec(S_k)\Delta a + A^T (\Delta b)$$

$$= MP(A, \lambda) \left[-A^T S_1 x_i, S_1^T r_i, \ldots, -A^T S_k x_i, S_k^T r_i, A^T\right] \Delta v,$$

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 $x_i = (A^T A + \lambda^2 L^T L)^{-1} A^T b = P(A, \lambda) A^T b$ be the Tikhonov regularized solution of (1.1). Then we obtain the structured normwise, componentwise, and mixed condition numbers:

$$m^\text{req}_\mathcal{L}(A, b) = \frac{\sum_{i=1}^{k} |a_i| |MP(A, \lambda) (A^T S_i x_i - S_i^T r_i)| + |MP(A, \lambda) A^T| |b|}{\|Mx_i\|_\infty},$$
Proof. From Lemmas 1 and 3, we have

\[
\kappa^{\text{Reg}}_\ell(A, b) = \frac{\|MP(A, \lambda) [S_i^T r_\lambda - A^T S_i x_\lambda, \ldots, S_k^T r_\lambda - A^T S_k x_\lambda, A^T] \|_2}{\|M x_\lambda\|_2},
\]

where

\[
\kappa^{\text{Reg}}_{\text{struct}}(A, b) = \frac{\|MP(A, \lambda) [S_i^T r_\lambda - A^T S_i x_\lambda, \ldots, S_k^T r_\lambda - A^T S_k x_\lambda, A^T] \|_2}{\|M x_\lambda\|_2}.
\]

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

When \( \{S_i\} \) 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 \( m^{\text{Reg}}(A, b), c^{\text{Reg}}(A, b) \) and \( \kappa^{\text{Reg}}(A, b) \).

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

\[
m^{\text{Reg}}(A, b) = \frac{\|D\phi ([a; b]) \|}{\|x_n\|_\infty}.
\]

\[
c^{\text{Reg}}(A, b) = \frac{\|MP(A, \lambda) [\{A^T\} - (x^T \otimes A^T)] \|}{\|M x_\lambda\|_\infty}.
\]

\[
\kappa^{\text{Reg}}(A, b) = \frac{\|MP(A, \lambda) [\{A^T\} - (x^T \otimes A^T)] \|_2}{\|M x_\lambda\|_2}.
\]

**Proof.** For the expression of \( \kappa^{\text{Reg}}_{\ell}(A, b) \) given in Theorem 1, let \( \{S_{ij} = e^{(m)}_j e^{(n)}_i^T\} \) be the canonical basis for \( \mathbb{R}^{m \times n} \), where \( e^{(m)}_j \) is the \( j \)th column of the identity matrix \( I_m, i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \). Then we have the following simplified expression:

\[
-S_{ij}^T r_\lambda + A^T S_i x_\lambda = -e_j^{(m)} r_{\lambda, (i)} + A^T e_i^{(m)} x_{\lambda, (i)},
\]

where \( r_{\lambda, (i)} \) and \( x_{\lambda, (i)} \) are respectively the \( i \)th and \( j \)th components of \( r_\lambda \) and \( x_\lambda \). Now, fixing \( j \), we get

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

which implies that

\[
[-S_{11}^T r_\lambda + A^T S_1 x_\lambda, \ldots, -S_{1n}^T r_\lambda + A^T S_n x_\lambda, -S_{21}^T r_\lambda + A^T S_1 x_\lambda, \ldots, -S_{2n}^T r_\lambda + A^T S_n x_\lambda, \ldots, -S_{nn}^T r_\lambda + A^T S_n x_\lambda]
\]

\[
= \left(-e_1^{(m)} \otimes r_{\lambda} + x_{\lambda, (1)} A^T, -(e_2^{(m)} \otimes r_{\lambda}) + x_{\lambda, (2)} A^T, \ldots, -(e_n^{(m)} \otimes r_{\lambda}) + x_{\lambda, (n)} A^T\right).
\]

Applying the above equation to the expression of \( \kappa^{\text{Reg}}_{\ell}(A, b) \) in Theorem 1, we prove the third statement. The expressions of \( m^{\text{Reg}}(A, b) \) and \( c^{\text{Reg}}(A, b) \) can be obtained similarly.

**Remark 2.** If we choose \( M = I_n \), then \( m^{\text{Reg}}(A, b), c^{\text{Reg}}(A, b) \) and \( \kappa^{\text{Reg}}(A, b) \) respectively reduce to the expressions of \( m_{\text{reg}}, c_{\text{reg}} \) and \( \text{cond}^{\text{reg}}_{\ell} \) in (2.3)–(2.5).
For the structured normwise condition number, when $A$ is a Toeplitz or Hankel matrix, we have

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

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

$$m^r_L(A, b) \leq m^r_{Reg}(A, b) \quad \text{and} \quad c^r_L(A, b) \leq c^r_{Reg}(A, b).$$

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

$$k_{L}^{\text{Reg}}(A, b) \leq \sqrt{2} \max \left\{ \max_{i=1,2,\ldots,k} \|S_i\|_F, 1 \right\} k^r_{\text{Reg}}(A, b).$$

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

$$\left\| \sum_{i=1}^{k} |a_i| \left( |M^r(A, \lambda) (A^T S_i x_i - S_i^T r_i) + |M^r(A, \lambda) A^T | b | \right) \right\|_{\infty}$$

$$= \left\| |M^r(A, \lambda) [A^T S_i x_i - S_i^T r_i, \ldots, A^T S_k x_k - S_k^T r_k] | a | + |M^r(A, \lambda) A^T | b | \right\|_{\infty}$$

$$= \left\| |M^r(A, \lambda) [A^T S_i x_i - S_i^T r_i, \ldots, A^T S_k x_k - S_k^T r_k] | a | \right\|_{\infty}$$

$$\leq \left\| |M^r(A, \lambda) \left( (x^T \otimes A^T) - (I_n \otimes r_i^T) \right) \text{vec}(S_1), \ldots, \text{vec}(S_k), A^T \right| \right\|_{\infty}$$

$$\leq \left\| |M^r(A, \lambda) \left( (x^T \otimes A^T) - (I_n \otimes r_i^T) \right) \text{vec}(S_1), \ldots, \text{vec}(S_k) \right| \right\|_{\infty}$$

$$\leq \left\| |M^r(A, \lambda) \left( (x^T \otimes A^T) - (I_n \otimes r_i^T) \right) | \sum_{i=1}^{k} |a_i| \text{vec}(S_i) \right| + |M^r(A, \lambda) A^T | b | \right\|_{\infty}$$

for the last equality we use the assumption $|A| = \sum_{i=1}^{k} |a_i| |S_i|$.

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 $(B_1, B_2) = trace(B_1^T B_2) = [\text{vec}(B_1)]^T \text{vec}(B_2)$ for $B_1, B_2 \in \mathbb{R}^{m \times n}$. 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 $(B_1, B_2) = trace(B_1^T B_2) = [\text{vec}(B_1)]^T \text{vec}(B_2)$ for $B_1, B_2 \in \mathbb{R}^{m \times n}$. It is easy to prove that

$$\left\| |M^r(A, \lambda) [A^T S_i x_i - S_i^T r_i, \ldots, A^T S_k x_k - S_k^T r_k, A^T] \right\|_2$$

$$= \left\| |M^r(A, \lambda) \left( (x^T \otimes A^T) - (I_n \otimes r_i^T) \right) \text{vec}(S_1), \ldots, \text{vec}(S_k), A^T \right| \right\|_2$$

$$= \left\| |M^r(A, \lambda) \left( (x^T \otimes A^T) - (I_n \otimes r_i^T) \right) \text{vec}(S_1), \ldots, \text{vec}(S_k) \right| \right\|_2 \max \left\{ \max_{i=1,2,\ldots,k} \|S_i\|_F, 1 \right\},$$

If $m^r_{Reg}(A, b)$ is smaller than $m^r_L(A, b)$, 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

$$\|a\|_2 \leq \sqrt{2}\|A\|_F.$$
where for the last equation we used the orthogonality of the basis \( \{ S_i \} \). So from Lemma 4,

\[
\kappa_{\mathcal{L}}^{\text{Reg}}(A, b) \leq \max \left\{ \frac{\max_{i=1,2,\ldots,k} \| S_i \|_F, 1}{\| M x_i \|_2} \left\| \begin{bmatrix} a^	op \\ b \end{bmatrix} \right\|_2 \right\} \left\| \begin{bmatrix} a^	op \\ b \end{bmatrix} \right\|_2 \leq \max \left\{ \frac{\max_{i=1,2,\ldots,k} \| S_i \|_F, 1}{\| M x_i \|_2} \left\| \begin{bmatrix} a^	op \\ b \end{bmatrix} \right\|_2 \right\} \left\| \begin{bmatrix} a^	op \\ b \end{bmatrix} \right\|_2 \sqrt{2 \| A \|_F^2 + \| b \|_2^2},
\]

which completes the proof of this proposition. \( \square \)

**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 \( \kappa_{\mathcal{L}}^{\text{Reg}}(A, b) \), 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 multiplication \( \mathbf{D} \phi([a; b]) h_1 \) and \( \mathbf{D} \phi([a; b]) h_2 \) with \( h_1 \in \mathbb{R}^{m+k} \) and \( h_2 \in \mathbb{R}^l \). To consider \( \mathbf{D} \phi([a; b]) h \), \( h \in \mathbb{R}^l \) and \( v_i \) defined in Lemma 3, denoting \( D = MP(A, \lambda)A^T \) and using \( P(A, \lambda)^T = P(A, \lambda) \), (1.5) and \( \Pi^T = \Pi^{-1} \), we have

\[
\begin{align*}
\nu_i^T P(A, \lambda) M^T h &= r_i^T S_i P(A, \lambda) M^T h - r_i^T S_i A P(A, \lambda) M^T h \\
&= \nu_i^T P(A, \lambda) M^T h - \nu_i^T S_i h_i^T A^T h \\
&= \nu_i^T P(A, \lambda) M^T h - \nu_i^T S_i h_i^T A^T h \\
&= \nu_i^T P(A, \lambda) M^T h - \nu_i^T S_i h_i^T A^T h
\end{align*}
\]

where we applied \( [\nu_1 | \nu_2 | \nu_k] ^T [\nu_1 | \nu_2 | \nu_k] = \text{vec}(A_1^T A_2) \) for the same dimensional matrices \( A_1 \) and \( A_2 \) in the last equality. It follows from (3.5) that

\[
\begin{align*}
\mathbf{D} \phi([a; b])^T h &= \left( MP(A, \lambda) \begin{bmatrix} v_1, \ldots, v_k, A^T \end{bmatrix} \right)^T h \\
&= \begin{bmatrix} \nu_1^T P(A, \lambda) M^T h \\
\vdots \\
\nu_k^T P(A, \lambda) M^T h \\
D^T h \end{bmatrix}
\end{align*}
\]

\[
(3.6)
\]

where \( a(h) = \begin{bmatrix} \text{trace} \left( S_1 (P(A, \lambda) M^T h r_i^T - x_i h_i^T D) \right), \ldots, \text{trace} \left( S_k (P(A, \lambda) M^T h r_i^T - x_i h_i^T D) \right) \end{bmatrix}^T \). It leads to the following proposition.

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

\[
\mathbf{D} \phi([a; b])^* \colon h \in \mathbb{R}^l \mapsto \begin{bmatrix} a(h) \\
D^T h \end{bmatrix} \in \mathbb{R}^k \times \mathbb{R}^m.
\]

Furthermore, when \( l = 1 \),

\[
\kappa_{\mathcal{L}}^{\text{Reg}}(A, b) = \sqrt{\sum_{i=1}^k s_i^2} + \| D \|_2^2 \left\| \begin{bmatrix} a \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} b \end{bmatrix} \right\|_2,
\]

\[
(3.8)
\]

where \( s_i = \text{trace} \left( S_i \left( P(A, \lambda) M^T h r_i^T - x_i h_i^T D \right) \right), \ i = 1, 2, \ldots, k. \)
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

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

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

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

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.

\[ \blacksquare \]

Remark 4. When \(l = 1\), we compute the conditioning of the \(i\)th component of the solution. In that case \(M\) is the \(i\)th canonical vector of \(\mathbb{R}^k \times \mathbb{R}^m\), and in (3.8), \(P(A, \lambda)M^\top\) is the \(i\)th column of \(P(A, \lambda)\) and \(D\) is the \(i\)th 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 \(\kappa^\text{Reg}_\omega(A, b)\). In this algorithm, we assume that \(x_\omega, r_\omega\) and \(\lambda\) are available. When the GSVD (1.4) of \((A, L)\) is available, a compact form of \(P(A, \lambda)\) is given by

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

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 \(\kappa^\text{Reg}_\omega(A, b)\)

Select initial vector \(h \in \mathbb{R}^l\).

for \(p = 1, 2, \ldots\) do

Using (3.9), calculate \(P(A, \lambda)M^\top hr_\omega^\top - x_\omega h^\top D\). From (3.6), denote \(\hat{a}_p = a(h)\) and \(\hat{b}_p = D^\top h\).

Calculate \(v = \|\hat{a}_p; \hat{b}_p\|_2\), let \(a_p = \hat{a}_p/v\) and \(b_p = \hat{b}_p/v\).

Let \(A_p = \sum_{i=1}^k a_{p,i}S_i\), where \(a_{p,i}\) is the \(i\)th component of \(a_p\).

Using (3.2) and (3.9), compute \(h = MP(A, \lambda) \left( A^\top b_p + A_p^\top r_\omega - A^\top A_p x_\omega \right)\).

end for

\(\kappa^\text{Reg}_\omega(A, b) = \sqrt{v}\).

The quantity \(v\) 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 \(v\) and comparing it to a tolerance given by the user.

For the mixed and componentwise condition numbers, we note that

\[
m^\text{Reg}_\omega(A, b) = \frac{\|MP(A, \lambda) [v_1, \ldots, v_k, A^\top] \text{Diag}([a; b])\|_\infty}{\|Mx_\omega\|_\infty} = \frac{\|\mathbf{D}\phi([a; b])\text{Diag}([a; b])\|_\infty}{\|Mx_\omega\|_\infty},
\]

\[
e^\text{Reg}_\omega(A, b) = \frac{\|MP(A, \lambda) [v_1, \ldots, v_k, A^\top] \text{Diag}([a; b])\|_\infty}{\|Mx_\omega\|_\infty} = \frac{\|\mathbf{D}\phi([a; b])\text{Diag}([a; b])\|_\infty}{\|Mx_\omega\|_\infty}.
\]

The above equations show that we only need to estimate the infinity norm of \(\mathbf{D}\phi([a; b])\text{Diag}([a; b])\). 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
Algorithm 2 The power method for estimating \( m_{z}^{\text{Reg}}(A, b) \)

Select initial vector \( h = I^{-1}e \in \mathbb{R}^l \).

for \( p = 1, 2, \ldots \) do
  Using \( (3.9) \), calculate \( P(A, \lambda)M^{\top}hr_{x}^{\top} - x_{h}^{\top}D \). 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 \( \vec{a}_{p} = \text{sign}(\alpha_{p}) \) and \( \vec{b}_{p} = \text{sign}(\beta_{p}) \).
  Compute \( A_{p} = \sum_{i=1}^{k} a_{p,(i)}S_{i} \), where \( a_{p} = [a_{p,(1)}, a_{p,(2)}, \ldots, a_{p,(k)}]^{\top} \).
  Using \( (3.2) \) and \( (3.9) \), compute \( z = MP(A, \lambda)(A^{\top}\vec{b}_{p} + A_{p}^{\top}r_{x} - A_{2}^{\top}A_{2}x_{2}) \).

if \( \|z\|_{\infty} \leq h^{\top}z \) then
  \[ \gamma' = \left\| \begin{bmatrix} \alpha_{p} \\ \beta_{p} \end{bmatrix} \right\|_{1} \]
  quit
endif

end for

\( m_{z}^{\text{Reg}}(A, b) = \gamma. \)

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 \( R' \). 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 \( V = [v_{ij}] \in VdM, \) 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 \( \Delta V \) of \( g(a + \Delta a) - g(a) \).

Lemma 5 \([10, \text{Lemma 6}]\). An explicit expression of \( \Delta V \) is

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

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 \( D\phi([a; b]) \) of function \( \phi \) defined in \((2.7)\) is

\[ D\phi([a; b]) = MP(V, \lambda) \begin{bmatrix} -V^{\top}V_{1}\text{Diag}(x_{2}) + \text{Diag}(y), V^{\top} \end{bmatrix}, \]

where \( y = V_{1}^{\top}r_{h} \) and \( r_{h} = b - Vx_{h} \).
Proof. It follows from (3.4) and Diag($a$)z = Diag($z$)$a$ for vectors $a$ and $z$ of the same dimension,
\[
\phi([a + \Delta a; b + \Delta b]) - \phi([a; b]) \approx MP(V, \lambda) \left( (-x_1^T \otimes V^T + I_n \otimes r_1^T) vec(\Delta V) + V^T(\Delta b) \right)
\]
\[
= MP(V, \lambda) \left( (-x_1^T \otimes V^T + (r_1^T \otimes I_n) \Pi) vec(\Delta V) + V^T(\Delta b) \right)
\]
\[
= MP(V, \lambda) \left( (-V^T(\Delta V)x_1 + (\Delta V)^T r_1) + V^T(\Delta b) \right)
\]
\[
= MP(V, \lambda) \left( (-V^T V_1 Diag(\Delta a)x_1 + Diag(\Delta a)V_1^T r_1) + V^T(\Delta b) \right)
\]
\[
= MP(V, \lambda) \left[ -V^T V_1 Diag(x_1) + Diag(y), V^T \right] \frac{\Delta a}{\Delta b}.
\]
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 $V \in VdM$, $b \in \mathbb{R}^m$ and $x_\lambda = (V^T V + \lambda L^T L)^{-1} V^T b = P(V, \lambda) V^T b$ be the Tikhonov regularized solution of (1.1). Recall that $y = V_1^T r_1$, then the structured condition numbers of the Vandermonde matrix are:

\[
m_{\text{VdM}}^\text{Reg}(V, b) = \frac{\|MP(V, \lambda) \left[ V^T V_1 \text{Diag}(x_\lambda) - \text{Diag}(y) \right] \| a \| + \|P(V, \lambda)V^T \| b \|}{\|Mx_\lambda\|_\infty},
\]
\[
c_{\text{VdM}}^\text{Reg}(V, b) = \frac{\|MP(V, \lambda) \left[ V^T V_1 \text{Diag}(x_\lambda) - \text{Diag}(y) \right] \| a \| + \|P(V, \lambda)V^T \| b \|}{\|Mx_\lambda\|_\infty},
\]
\[
k_{\text{VdM}}^\text{Reg}(V, b) = \frac{\|MP(V, \lambda) \left[ \text{Diag}(y) - V^T V_1 \text{Diag}(x_\lambda), V^T \right] \|_2 \left[ \begin{array}{c} a \\ b \end{array} \right]}{\|Mx_\lambda\|_2}.
\]

In particular, when $l = 1$,
\[
k_{\text{VdM}}^\text{Reg}(V, b) = \frac{\sqrt{\|y \odot (P(V, \lambda)M^T) - x_\lambda \odot (V_1^T D^T V h) \|_2^2 + \|D^T V h\|_2^2}}{\|Mx_\lambda\|_2}.
\]

Analogous to Proposition 2, the adjoint operator of $D\phi([a; b])$, using the scalar products $a^T a_2 + b^T b_2$ and $h^T h$ on $\mathbb{R}^{m+n}$ and $\mathbb{R}$ respectively, is
\[
D\phi([a; b])^* : h \in \mathbb{R}^l \mapsto [\|y \odot (P(V, \lambda)M^T) - x_\lambda \odot (V_1^T D^T V h) \|_2^2 + \|D^T V h\|_2^2] \in \mathbb{R}^n \times \mathbb{R}^m,
\]
where $D^T V = MP(V, \lambda)V^T$. The above expressions can be used to estimate $m_{\text{VdM}}^\text{Reg}(V, b)$, $c_{\text{VdM}}^\text{Reg}(V, b)$ and $k_{\text{VdM}}^\text{Reg}(V, b)$ 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 $C = [c_{ij}] \in$ Cauchy, then there exist $u = [u_1, u_2, \ldots, u_m]^T \in \mathbb{R}^m$ and $v = [v_1, v_2, \ldots, v_n]^T \in \mathbb{R}^n$, with $u_i \neq v_j$ for $i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$ such that, for all $i \leq m$ and $j \leq n$,
\[
c_{ij} = \frac{1}{u_i - v_j}.
\]

If $w = [w; v] \in \mathbb{R}^{m+n}$, then $C = g(w)$. Let $\Delta w = [\Delta u; \Delta v] = [\Delta u_1, \ldots, \Delta u_m, \Delta v_1, \ldots, \Delta v_n]^T \in \mathbb{R}^{m+n}$ be the perturbation on $w$. The first order term $\Delta C$ in $g(w + \Delta w) - g(w)$ is given by [10, Lemma 9]
\[
\Delta C \approx \begin{bmatrix} \frac{\Delta u_i - \Delta v_j}{(u_i - v_j)^2} \end{bmatrix} = \text{Diag}(\Delta u)C_1 - C_1 \text{Diag}(\Delta v) \in \mathbb{R}^{m \times n},
\]
where $C_1 = [1/(u_i - v_j)^2] \in \mathbb{R}^{m \times n}$.

Lemma 7. The Fréchet derivative $D\phi([w; b])$ of function $\phi$ defined in (2.7) is given by
\[
D\phi([w; b]) = MP(C, \lambda) \left[ C_u, C_r, C^T \right],
\]
where $C_u = C_1^T \text{Diag}(r_\lambda) - C^T \text{Diag}(z_1)$, $C_r = C^T C_1 \text{Diag}(x_\lambda) - \text{Diag}(z_2)$, $z_1 = C_1 x_\lambda$, and $r_\lambda = b - Cx_\lambda$. 


Proof. It follows from (3.4) and Diag($a$)z = Diag($z$)$a$ for vectors $a$ and $z$ of the same dimension,
Proof. Following the proof of Lemma 6, we can show that
\[
\phi((w + \Delta w; b + \Delta b)) - \phi((w; b)) = MP(C, \lambda)\left[C_1^T \text{Diag}(r_2) - C^T \text{Diag}(z_2), C^T C_1 \text{Diag}(x_0) - \text{Diag}(z_2), C^T\right] \Delta w - \Delta b.
\]

Then the Fréchet derivative of \( \phi \) at \([w; b]\) is
\[
D\phi ([w; b]) = MP(C, \lambda) \left[ C_1^T \text{Diag}(r_2) - C^T \text{Diag}(z_2), C^T C_1 \text{Diag}(x_0) - \text{Diag}(z_2), C^T\right].
\]

Theorem 4. Let \([C \in \text{Cauchy}, b \in \mathbb{R}^n\) and \(x_\lambda = (C^T C + \lambda I^T I)^{-1} C^T b = P(C, \lambda)C^T b \) be the solution of the Tikhonov regularization problem (1.1), then the structured condition numbers are:
\[
m_{\text{Reg}}^\text{Cauchy}(C, b) = \frac{\|MP(C, \lambda)C_u\|_2 |u| + |MP(C, \lambda)C_v| |v| + |MP(C, \lambda)C^T| |b|}{\|Mx_\lambda\|}\infty,
\]
\[
c_{\text{Reg}}^\text{Cauchy}(C, b) = \frac{\|MP(C, \lambda)C_u| |u| + |MP(C, \lambda)C_v| |v| + |MP(C, \lambda)C^T| |b|}{\|Mx_\lambda\|}\infty,
\]
\[
k_{\text{Reg}}^\text{Cauchy}(C, b) = \left[ \frac{MP(C, \lambda)\left[ C_u, C_v, C^T\right]}{2}\right]_2 \left[ \frac{u^T}{v^T b}\right]_2.
\]

In particular, when \(l = 1\),
\[
k_{\text{Reg}}^\text{Cauchy}(C, b) = \sqrt{t^2 + s^2 + \|Dc\|_2^2} \left[ \frac{u^T}{v^T b}\right]_2,
\]

where \(t = \|r_\lambda \circ [C_1 P(C, \lambda)M^T] - z_1 \circ (D_c^T)^2\|_2\) and \(s = \|x_\lambda \circ (C_1^T D_c^T) - z_2 \circ [P(C, \lambda)M^T]\|_2\).

Similar to the case of the Vandermonde matrix, for the Cauchy matrix, the adjoint operator of \(D\phi([w; b])\), using the scalar products \(u_1 u_2 + v_1 v_2 + b_1 b_2 + h^T h\) on \(\mathbb{R}^{2m+n}\) and \(\mathbb{R}^n\) respectively, is
\[
D\phi([w; b])^* : h \in \mathbb{R}^l \mapsto [u(h) v(h) D_c^T h] \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m,
\]

where \(u(h) = r_\lambda \circ [C_1 P(C, \lambda)M^T h] - z_1 \circ (D_c^T h)\), \(v(h) = x_\lambda \circ (C_1^T D_c^T h) - z_2 \circ [P(C, \lambda)M^T h]\) and \(D_c = MP(C, \lambda)C^T\).

In particular, when \(l = 1\), we have
\[
k_{\text{Reg}}^\text{Cauchy}(V, b) = \sqrt{t^2 + s^2 + \|Dc\|_2^2} \left[ \frac{u^T}{v^T b}\right]_2,
\]

where \(t = \|r_\lambda \circ [C_1 P(C, \lambda)M^T] - z_1 \circ (D_c^T)^2\|_2\) and \(s = \|x_\lambda \circ (C_1^T D_c^T) - z_2 \circ [P(C, \lambda)M^T]\|_2\).

Using the above expressions, the algorithms similar to Algorithms 1 and 2 for estimating \(m_{\text{Reg}}^\text{Cauchy}(C, b), c_{\text{Reg}}^\text{Cauchy}(C, b)\) and \(k_{\text{Reg}}^\text{Cauchy}(C, b)\) 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 \(f \in \mathbb{R}^n\) and \([E, f]\) has the Frobenius norm equal to one. According to Section 2.2, we first need to evaluate the directional derivative \(D\psi([A, b]; [E, f])\) of \(\psi([A, b])\) with respect to \([A, b]\) in the direction \([E, f]\). From the proof of Lemma 3, we have
\[
D\psi([A, b]; [E, f]) = P(A, \lambda)\left(A^T f + E^T r_\lambda - A^T E x_\lambda\right).
\]
When we have the GSVD (1.4) of \((A, L)\), it is easy to deduce that
\[
D_a(\psi([A, b]; [E, f])) = QR^{-1} \left[ \begin{bmatrix} \Sigma^2 + \lambda^2 \mathbf{S}^2 \end{bmatrix}^{-1} & 0 \\ 0 & I_{p-n} \right] R^{-T} Q^T \left( A^T f + E^T r_b - A^T E x_b \right).
\] (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 \(x_b\), and the parameter \(\lambda\). The output is an estimation \(\kappa_{\text{SCE}}^{(k)}\) of the normwise condition number \(\kappa_{\text{cond}}^{(k)}\). 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 \(N(0, 1)\), and for \(B = (b) \in \mathbb{R}^{p \times q}\), \(|B|^2 = |b|^2\) and \(\sqrt{|B|} = (\sqrt{|b|}) \in \mathbb{R}^{p \times q}\).

**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 \(N(0, 1)\), where \(E_i \in \mathbb{R}^{m \times n}, f_i \in \mathbb{R}^m\). Use Gram–Schmidt orthogonalization process for the matrix
   \[
   \begin{bmatrix}
   \text{vec}(E_1) & \text{vec}(E_2) & \cdots & \text{vec}(E_k)
   \\
   f_1 & f_2 & \cdots & f_k
   \end{bmatrix}
   \]
   and form an orthonormal matrix \([q_1, q_2, \ldots, q_k]\). Each \(q_i\) can be converted into the desired matrices \([\tilde{E}_i, \tilde{f}_i]\) with the \texttt{unvec} operation.
2. Calculate \(D_a(\psi([A, b]; [\tilde{E}_i, \tilde{f}_i]))\) by (5.1), \(i = 1, 2, \ldots, k\).
3. Compute the absolute condition vector
   \[
   \kappa_{\text{abs}}^{(k)} := \frac{\partial}{\partial \psi} \left| D_a(\psi([A, b]; [\tilde{E}_1, \tilde{f}_1])) \right|^2 + \cdots + \left| D_a(\psi([A, b]; [\tilde{E}_k, \tilde{f}_k])) \right|^2.
   \]
4. Compute the normwise condition estimation:
   \[
   \kappa_{\text{SCE}}^{(k)} := \frac{\| \kappa_{\text{abs}}^{(k)} \|_2}{\| \psi([A, b]) \|_2}.
   \]

### 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 \(\Delta A\) 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
\[
\psi([A, b]) = (A^T A + \lambda^2 L^T L)^{-1} A^T b.
\]
SCE is flexible enough to accurately gauge the sensitivity of matrix functions subject to componentwise perturbations. Define the linear function
\[
h([B, d]) = [B, d] \circ [A, b], \quad B \in \mathbb{R}^{m \times n}, \quad d \in \mathbb{R}^m.
\]
Let \(E \in \mathbb{R}^{m \times (n+1)}\) be the matrix of all ones, then \(h(E) = [A b]\) and
\[
h(E + [E, f]) = [A, b] + h([E, f]).
\]
We know that \(h([E, f])\) is a componentwise perturbation on \([A, b]\), and \(h\) converts a general perturbation \(\varepsilon\) 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
\[
\psi([A, b]) = \psi(h(\varepsilon))
\]
with respect to \(\varepsilon\) in the direction \([E, f]\), which is
\[
D(\psi \circ h)(\varepsilon; [E, f]) = D\psi(h(\varepsilon)) D h(\varepsilon; [E, f]) = D\psi([A, b]) h([E, f]) = D\psi([A, b]; h([E, f])),
\]
since \(h\) is linear. Thus, to estimate the condition of the Tikhonov regularization solution \(x_b\) 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 $E_i \in \mathbb{R}^{m \times n}$, $f_i \in \mathbb{R}^{n}$. Use Gram–Schmidt orthogonalization process for the matrix

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

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 $\text{unvec}$ operation.

2. For $i = 1, \ldots, k$, set $[\tilde{E}_i, \tilde{f}_i]$ to the componentwise product of $[A, b]$ and $[E_i, f_i]$.

3. Calculate $D\psi([A, b]; [\tilde{E}_i, \tilde{f}_i])$ by (5.1), $i = 1, 2, \ldots, k$.

4. Compute the absolute condition vector

$$c_{\text{abs}}^{(k)} := \frac{\omega_k}{\omega_p} \sqrt{\|D\psi([A, b]; [\tilde{E}_i, \tilde{f}_i])\|^2 + \cdots + \|D\psi([A, b]; [\tilde{E}_k, \tilde{f}_k])\|^2}.$$

5. The mixed condition estimation $n_{\text{SCE}}^{(k)}$ and componentwise condition estimation $c_{\text{SCE}}^{(k)}$ are:

$$n_{\text{SCE}}^{(k)} := \left\| \frac{c_{\text{abs}}^{(k)}}{\|x_k\|_\infty} \right\|_\infty \text{ and } c_{\text{SCE}}^{(k)} := \left\| \frac{c_{\text{abs}}^{(k)}}{\|x_k\|_\infty} \right\|_\infty.$$

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 $\phi$ 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 $x_k$, we only need to choose $M = I_n$ in (2.7).

The key step in the SCE is the computation of the directional derivative $D\phi([a; b]; [e; f])$ 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 $D\phi([a; b])$ 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 $D\phi([a; b]; [e; f])$ are:

$$D\phi([a; b]; [e; f]) = P(A, \lambda) \left( A^T f + E^T r_k - A^T E x_k \right), \quad e = (e_i) \in \mathbb{R}^k, \quad E = \sum_{i=1}^{k} e_i s_i,$$

for linear structures,

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

for Vandermonde matrices, and

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

for Cauchy matrices, where $V_1$ is defined in Lemma 5, $y = V_1^T r_{\gamma}$, $C_a$ 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 $D\phi([a; b]; [e; f])$ 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 $D\phi([a; b]; [e; f])$ is equivalent to solving the following positive definite linear system

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

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 BTB, $E$ is also BTB. 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 scale 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

$$\begin{bmatrix}
\text{vec}(E_1) & \text{vec}(E_2) & \cdots & \text{vec}(E_k)
\end{bmatrix}
$$

for forming 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 $\text{unvec}$ operation.

2. For $i = 1, \ldots, k$, set $[\tilde{E}_i, \tilde{f}_i]$ to the componentwise product of $[A, b]$ and $[E_i, f_i]$.

3. Implement PCG to solve

$$(A^T A + \lambda^2 L^T L) v_i = A^T \tilde{f}_i + E_i^T r_\lambda - A^T E_i \lambda^2, \quad e = (e_i) \in \mathbb{R}^k, \quad E = \sum_{i=1}^k e_i S_i. \quad (5.3)$$

4. Compute the absolute condition vector

$$c_{abs}^{(k)} := \frac{\alpha_k}{\alpha_p} \sqrt{|v_1|^2 + \cdots + |v_k|^2}.$$

5. The mixed condition estimation $m_{SCE}^{(k)}$ and componentwise condition estimation $c_{SCE}^{(k)}$ are:

$$m_{SCE}^{(k)} := \frac{\|e_{abs}\|_\infty}{\|x\|_\infty} \quad \text{and} \quad c_{SCE}^{(k)} := \frac{c_{abs}}{\|x\|_\infty}. \quad (6.1)$$

### 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 drive running Windows 7 professional. All the computations were carried out using MATLAB 8.1 with machine precision about $2.2 \times 10^{-16}$ 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 $A$ 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

$$\Delta a_i = \varepsilon s_i a_i, \quad \Delta b_j = \varepsilon f_j b_j, \quad (6.1)$$

then $\hat{A} = g(a + \Delta a)$ and $\hat{b} = b + \Delta b$. 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 $\|Ax_\lambda - b\|_2$ versus the corresponding residual norm $\|Ax_\lambda - b\|_2$, with $\lambda$ 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 $x_\lambda$, 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 $x_\lambda$ 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 $y_\varepsilon$ was obtained in the similar way to $x_\lambda$, where $y_\varepsilon$ corresponds to $A$ and $\hat{b}$. Denote the error $\Delta x_\lambda = y_\varepsilon - x_\lambda$. 

---

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 $5 \times 5$ symmetric Toeplitz matrix defined by

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

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 $m_{\text{SymToep}}(A, b)$, $c_{\text{SymToep}}(A, b)$ and $\kappa_{\text{SymToep}}(A, b)$. 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 $c_{\text{SymToep}}(A, b)$ and $\kappa_{\text{SymToep}}(A, b)$ 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,

$$ \frac{\| M \Delta x_1 \|_{\infty}}{\| M x_1 \|_{\infty}} = \frac{\| M \Delta x_2 \|}{\| M x_2 \|_{\infty}}, \quad m_{\text{SymToep}}(A, b) = c_{\text{SymToep}}(A, b). $$

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

From Table 2, we can see the quantities $\varepsilon m_{\text{SymToep}}(A, b)$ give tighter perturbation bounds than $\varepsilon \kappa_{\text{SymToep}}(A, b)$, 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 $x_2$. We display the values of $m_{\text{SymToep}}(A, b)$ and $\kappa_{\text{SymToep}}(A, b)$. From Table 3, we can say that the first component of $x_1$ 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 $\kappa_{\text{Est}}(A, b)$ and $m_{\text{Est}}(A, b)$ respectively.

From Table 4, we can say that $\kappa_{\text{Est}}(A, b)$ and $m_{\text{Est}}(A, b)$ give good estimations for this specific $A, L$ and $b$, especially $m_{\text{Est}}(A, b)$ gives better estimation.
Table 3

$L = I_5; M = e_1^r$ and $M = e_1^r$.

<table>
<thead>
<tr>
<th>$M = e_1^r$</th>
<th>Discrep. pr.</th>
<th>$L$-curve</th>
<th>GCV</th>
<th>Quasi-opt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>6.39 · $10^{-4}$</td>
<td>5.00 · $10^{-4}$</td>
<td>5.08 · $10^{-4}$</td>
<td>5.08 · $10^{-4}$</td>
</tr>
<tr>
<td>$\kappa_{\text{T}oep}(A, b)$</td>
<td>1.5887 · $10^{3}$</td>
<td>2.0941 · $10^{3}$</td>
<td>2.0594 · $10^{3}$</td>
<td>2.0594 · $10^{3}$</td>
</tr>
<tr>
<td>$m_{\text{T}oep}(A, b)$</td>
<td>7.6056 · $10^{2}$</td>
<td>1.0022 · $10^{3}$</td>
<td>9.8567 · $10^{2}$</td>
<td>9.8567 · $10^{2}$</td>
</tr>
</tbody>
</table>

$M = e_1^r$

| Value of $\lambda$ | 6.39 · $10^{-4}$ | 5.00 · $10^{-4}$ | 5.08 · $10^{-4}$ | 5.08 · $10^{-4}$ |
| $\kappa_{\text{T}oep}(A, b)$ | 1.7780 · $10^{2}$ | 2.9088 · $10^{7}$ | 2.8141 · $10^{7}$ | 2.8141 · $10^{7}$ |
| $m_{\text{T}oep}(A, b)$ | 4.9096 · $10^{6}$ | 8.0320 · $10^{6}$ | 7.7705 · $10^{6}$ | 7.7705 · $10^{6}$ |

Table 4

$L = I_5, M = I_5$ and $M = e_1^r$.

<table>
<thead>
<tr>
<th>$M = I_5$</th>
<th>Discrep. pr.</th>
<th>$L$-curve</th>
<th>GCV</th>
<th>Quasi-opt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>6.39 · $10^{-4}$</td>
<td>5 · $10^{-4}$</td>
<td>5.08 · $10^{-4}$</td>
<td>5.08 · $10^{-4}$</td>
</tr>
<tr>
<td>$\kappa_{\text{T}oep}(A, b)$</td>
<td>1.588783796 · $10^{3}$</td>
<td>2.0931466345 · $10^{3}$</td>
<td>2.0584876249 · $10^{3}$</td>
<td>2.0584876249 · $10^{3}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>7.581802517 · $10^{2}$</td>
<td>1.0002500616 · $10^{2}$</td>
<td>9.8369583257 · $10^{2}$</td>
<td>9.8369583257 · $10^{2}$</td>
</tr>
<tr>
<td>$m_{\text{T}oep}(A, b)$</td>
<td>7.6117517197 · $10^{2}$</td>
<td>1.0027483753 · $10^{3}$</td>
<td>9.8617760333 · $10^{2}$</td>
<td>9.8617760333 · $10^{2}$</td>
</tr>
<tr>
<td>$m_{\text{H}ank}(A, b)$</td>
<td>7.6055529470 · $10^{2}$</td>
<td>1.0022493745 · $10^{3}$</td>
<td>9.8567031098 · $10^{2}$</td>
<td>9.8567031098 · $10^{2}$</td>
</tr>
</tbody>
</table>

$M = e_1^r$

| Value of $\lambda$ | 6.39 · $10^{-4}$ | 5.00 · $10^{-4}$ | 5.08 · $10^{-4}$ | 5.08 · $10^{-4}$ |
| $\kappa_{\text{T}oep}(A, b)$ | 1.5886924101 · $10^{3}$ | 2.0940875560 · $10^{3}$ | 2.0594198505 · $10^{3}$ | 2.0594198505 · $10^{3}$ |
| $\kappa_{\text{H}ank}(A, b)$ | 3.790426062 · $10^{2}$ | 5.0050048168 · $10^{2}$ | 4.9222129601 · $10^{2}$ | 4.9222129601 · $10^{2}$ |
| $m_{\text{T}oep}(A, b)$ | 7.6055660483 · $10^{2}$ | 1.002496243 · $10^{3}$ | 9.8567056493 · $10^{2}$ | 9.8567056493 · $10^{2}$ |
| $m_{\text{H}ank}(A, b)$ | 7.6055529470 · $10^{2}$ | 1.0022493745 · $10^{3}$ | 9.8567031098 · $10^{2}$ | 9.8567031098 · $10^{2}$ |

Table 5

$L = I_5, M = I_5$.

<table>
<thead>
<tr>
<th>$M = I_5$</th>
<th>Discrep. pr.</th>
<th>$L$-curve</th>
<th>GCV</th>
<th>Quasi-opt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>7.5918 · $10^{-4}$</td>
<td>2.5002 · $10^{-4}$</td>
<td>2.5002 · $10^{-4}$</td>
<td>0.0017</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>1.0902</td>
<td>2.9510</td>
<td>2.6014</td>
<td>1.3163</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>1.3264</td>
<td>4.2237</td>
<td>3.2460</td>
<td>2.0419</td>
</tr>
<tr>
<td>$\epsilon = \frac{b^T \cdot \Delta}{g} \cdot \infty$</td>
<td>$4.39 \cdot 10^{6}$</td>
<td>$2.16 \cdot 10^{7}$</td>
<td>$2.645 \cdot 10^{7}$</td>
<td>$1.31 \cdot 10^{6}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>2.2310 · $10^{3}$</td>
<td>1.1401 · $10^{4}$</td>
<td>1.1401 · $10^{4}$</td>
<td>4.6222 · $10^{2}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>7.8426 · $10^{2}$</td>
<td>4.0032 · $10^{3}$</td>
<td>4.0032 · $10^{3}$</td>
<td>1.6347 · $10^{2}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>1.3230 · $10^{3}$</td>
<td>1.0238 · $10^{4}$</td>
<td>1.0238 · $10^{4}$</td>
<td>2.6208 · $10^{2}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>1.0372 · $10^{3}$</td>
<td>5.2922 · $10^{3}$</td>
<td>5.2922 · $10^{3}$</td>
<td>2.1648 · $10^{2}$</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>3.4999</td>
<td>5.1247</td>
<td>5.1247</td>
<td>3.5000</td>
</tr>
<tr>
<td>$\kappa_{\text{H}ank}(A, b)$</td>
<td>1.1576 · $10^{7}$</td>
<td>8.9578 · $10^{7}$</td>
<td>8.9578 · $10^{7}$</td>
<td>2.2931 · $10^{2}$</td>
</tr>
</tbody>
</table>

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

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

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 $Y_1 = g([e_1; 0]), \ldots, Y_5 = g([e_5; 0]), Y_6 = g([e_6; e_1]), Y_7 = g([0; e_2]), \ldots, Y_{11} = g([0; e_6])$, so that $A = \sum_{k=1}^{11} q_k Y_k$. Again, from Theorem 1 with the $Y_i$, the expressions for $m_{\text{H}ank}(A, b), c_{\text{H}ank}(A, b)$ and $\kappa_{\text{H}ank}(A, b)$ 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.
Both of them give sharper perturbation bounds. The perturbation bounds given by the structured condition numbers are much smaller than the unstructured ones. Therefore, the structured condition numbers are more useful in practice.

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 (J10). Let A = g(a) be a 10 × 8 Cauchy matrix whose (i, j)-entry is

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

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
where their 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

\[
\begin{align*}
  r_k &:= \frac{\kappa_{SCE}^{(k)} \cdot E}{\| \Delta x_k \|_2 / \| x_k \|_2}, \\
  r_m &:= \frac{m_{SCE}^{(k)} \cdot E}{\| \Delta x_k \|_\infty / \| x_k \|_\infty}, \\
  r_c &:= \frac{c_{SCE}^{(k)} \cdot E}{\| \Delta x_k \|_\infty / \| x_k \|_\infty},
\end{align*}
\]

where \( k \) is the subproblem dimension in Algorithms 3 and 4, \( \kappa_{SCE}^{(k)} \), \( m_{SCE}^{(k)} \) and \( c_{SCE}^{(k)} \) 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

\[
L_1 = \begin{bmatrix}
1 & -1 & \cdots & \cdots & -1 \\
1 & -1 & \cdots & \cdots & -1 \\
\end{bmatrix} \in \mathbb{R}^{(n-1) \times n}
\]

which approximates the first derivative operator. We adopted the following four values of the regularization parameter \( \lambda \):

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

In Table 8, we report the numerical results on the ratios \( r_k \), \( r_m \) and \( r_c \) for examples with various dimensions and choices of \( L \). The table shows that the mixed condition estimation \( m_{SCE}^{(k)} \) reflects the true error bound accurately, while the componentwise condition estimation \( c_{SCE}^{(k)} \) 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 \( m_{SCE}^{(k)} \) can be considered reliable [36]. The values of the componentwise condition estimation \( c_{SCE}^{(k)} \) are within (16.54, 88.19) except for the case shaw, where \( n = 512, L = I_n, k = 3 \) with all choices of \( \lambda \), the case wing, where \( n = 256, L = I_n, k = 3, \lambda = 0.1 \cdot 6 \cdot 10^{-2} \) and the case shaw, where \( n = 256, L = I_n, k = 5, \lambda = 1.7 \cdot 10^{-3} \), indicating that the componentwise condition estimation \( c_{SCE}^{(k)} \) is effective for most cases. For the normwise condition estimation \( \kappa_{SCE}^{(k)} \), most of the values of \( r_k \) are of order \( O(10^5) \), and even some of them are of order \( O(10^3) \), showing that the normwise condition estimation overestimates 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 \( b \) of \( E \in \mathbb{R}^n \) 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 \( \kappa_{SCE}^{(k)} \), \( m_{SCE}^{(k)} \) and \( c_{SCE}^{(k)} \), respectively. The perturbations \( \Delta a \) on \( a \) and \( \Delta b \) on \( b \) were generated as in (6.1). As in the previous example, let the overestimate ratios be defined by

\[
\begin{align*}
  r_k^{\text{SymToep}} &:= \frac{\kappa_{\text{SymToep, SCE}}^{(k)} \cdot E}{\| \Delta x_k \|_2 / \| x_k \|_2}, \\
  r_m^{\text{SymToep}} &:= \frac{m_{\text{SymToep, SCE}}^{(k)} \cdot E}{\| \Delta x_k \|_\infty / \| x_k \|_\infty}, \\
  r_c^{\text{SymToep}} &:= \frac{c_{\text{SymToep, SCE}}^{(k)} \cdot E}{\| \Delta x_k \|_\infty / \| x_k \|_\infty},
\end{align*}
\]

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 \( O(10) \), 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 (BTB \((m, n)\)) be of the form

\[
A = [A_{i-j}]_{i,j=1}^m
\]

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

\[
b = Ax + \eta,
\]

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

\[
a_{ij} = \begin{cases} 
\exp(-0.1 j^2 - 0.1 l^2), & -8 \leq j, l \leq 8, \\
0, & \text{otherwise},
\end{cases}
\]
Table 8
SCE for the Tikhonov regularization problem.

<table>
<thead>
<tr>
<th></th>
<th>$r_e$</th>
<th>$r_m$</th>
<th>$r_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>deriv2, $n = 64, L = L_1, k = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>7.7533 · 10^2</td>
<td>2.0847</td>
<td>5.1984 · 10</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>7.6234 · 10^2</td>
<td>2.1284</td>
<td>4.7450 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>6.7427 · 10^2</td>
<td>1.2970</td>
<td>2.4927 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>9.2294 · 10^2</td>
<td>1.1437</td>
<td>3.9081 · 10</td>
</tr>
<tr>
<td>deriv2, $n = 64, L = L_1, k = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>9.2479 · 10^2</td>
<td>2.2288</td>
<td>3.8756 · 10</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>9.2057 · 10^2</td>
<td>2.5900</td>
<td>3.6441 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>6.2048 · 10^2</td>
<td>1.0458</td>
<td>2.9864 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>9.2686 · 10^2</td>
<td>1.5548</td>
<td>3.4564 · 10</td>
</tr>
<tr>
<td>wing, $n = 128, L = L_1, k = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>2.3753 · 10^3</td>
<td>1.6337 · 10^1</td>
<td>1.6544 · 10</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>1.4154 · 10^3</td>
<td>7.5189</td>
<td>1.5910 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>1.3068 · 10^3</td>
<td>1.6301</td>
<td>2.2538 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>9.6524 · 10^2</td>
<td>1.3917</td>
<td>8.1316 · 10</td>
</tr>
<tr>
<td>wing, $n = 256, L = L_1, k = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>6.1606 · 10^2</td>
<td>1.0112</td>
<td>2.6130 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>5.4330 · 10^2</td>
<td>1.3470</td>
<td>1.8994 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>1.0126 · 10^3</td>
<td>1.8223</td>
<td>6.2581 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>1.0923 · 10^3</td>
<td>2.1715</td>
<td>7.2018 · 10</td>
</tr>
<tr>
<td>shaw, $n = 512, L = L_1, k = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>1.6632 · 10^2</td>
<td>2.4174</td>
<td>1.7467 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>1.8554 · 10^2</td>
<td>2.8948</td>
<td>3.7301 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>1.0707 · 10^2</td>
<td>3.1782</td>
<td>3.7127 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>2.9693 · 10^2</td>
<td>1.6429</td>
<td>1.8405 · 10^2</td>
</tr>
<tr>
<td>shaw, $n = 256, L = L_1, k = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.1$</td>
<td>2.8617 · 10^2</td>
<td>2.9029</td>
<td>8.1886 · 10</td>
</tr>
<tr>
<td>$\lambda = 6 · 10^{-2}$</td>
<td>3.1563 · 10^2</td>
<td>3.0425</td>
<td>8.3968 · 10</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-3}$</td>
<td>3.0292 · 10^2</td>
<td>1.0652</td>
<td>2.0672 · 10^2</td>
</tr>
<tr>
<td>$\lambda = 1.7 · 10^{-4}$</td>
<td>1.1406 · 10^2</td>
<td>9.4368 · 10^1</td>
<td>3.0579 · 10</td>
</tr>
</tbody>
</table>

Table 9
SCE for the structured Tikhonov regularization problem.

<table>
<thead>
<tr>
<th></th>
<th>$r_e^{SymToep}$</th>
<th>$r_m^{SymToep}$</th>
<th>$r_c^{SymToep}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 100, n = 50, k = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrep. pr. $\lambda = 2.21$</td>
<td>1.3427 · 10</td>
<td>1.2354</td>
<td>4.0474 · 10</td>
</tr>
<tr>
<td>$L$-curve $\lambda = 6.19 · 10^{-2}$</td>
<td>1.8396 · 10</td>
<td>1.7149</td>
<td>2.9893 · 10</td>
</tr>
<tr>
<td>GCV $\lambda = 1.35 · 10^{-4}$</td>
<td>1.8113 · 10</td>
<td>1.5819</td>
<td>2.0690 · 10</td>
</tr>
<tr>
<td>Quasi-opt $\lambda = 7.48 · 10^{-1}$</td>
<td>1.1606 · 10</td>
<td>8.6429 · 10^{-1}</td>
<td>5.7865 · 10</td>
</tr>
<tr>
<td>$m = 300, n = 200, k = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrep. pr. $\lambda = 4.71$</td>
<td>7.6537 · 10</td>
<td>4.1397</td>
<td>4.8346 · 10</td>
</tr>
<tr>
<td>$L$-curve $\lambda = 1.10 · 10^{-1}$</td>
<td>5.6455 · 10</td>
<td>2.4765</td>
<td>7.8375 · 10</td>
</tr>
<tr>
<td>GCV $\lambda = 3.22 · 10^{-4}$</td>
<td>3.2584 · 10</td>
<td>1.4440</td>
<td>8.8256 · 10</td>
</tr>
<tr>
<td>Quasi-opt $\lambda = 4.49$</td>
<td>7.1689 · 10</td>
<td>3.6997</td>
<td>5.4007 · 10</td>
</tr>
<tr>
<td>$m = 500, n = 300, k = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrep. pr. $\lambda = 1.19 · 10^{-2}$</td>
<td>5.8732 · 10</td>
<td>1.9115</td>
<td>1.0526 · 10</td>
</tr>
<tr>
<td>$L$-curve $\lambda = 1.03$</td>
<td>7.8779 · 10</td>
<td>3.2648</td>
<td>9.2462 · 10</td>
</tr>
<tr>
<td>GCV $\lambda = 5.66 · 10^{-4}$</td>
<td>4.3191 · 10</td>
<td>1.6125</td>
<td>1.2357 · 10</td>
</tr>
<tr>
<td>Quasi-opt $\lambda = 9.25$</td>
<td>1.0777 · 10</td>
<td>4.3658</td>
<td>6.8813 · 10</td>
</tr>
</tbody>
</table>

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

\[(A^T A + \lambda^2 L^T L)x = A^T b,\]  

(6.4)
where \( L = l_{mn} \) 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 \( \| r_2^{(q)} \|_2 / \| r_2^{(0)} \|_2 \leq 10^{-7} \), where \( r_2^{(q)} \) is the normal equation residual vector at the \( q \)-th iteration in the PCG. The perturbations \( \Delta a \) on \( a \) and \( \Delta b \) 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].
We always choose $k = 3$ in Algorithm 5. As in the previous example, let the overestimate ratios be defined by

\[
\begin{align*}
    r_m^{\text{BTTB}.\text{SCE}} &:= \frac{m^{(k)}_{\text{BTTB}.\text{SCE}} \cdot \epsilon}{\| \Delta x_i \|_\infty / \| x_i \|_\infty}, \\
    r_c^{\text{BTTB}} &:= \frac{c^{(k)}_{\text{BTTB}.\text{SCE}} \cdot \epsilon}{\| \Delta x_i / x_i \|_\infty},
\end{align*}
\]

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

\[
    r_{\text{time}} = \frac{t_2}{t_1},
\]

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 $\Delta a$ and $\Delta b$ and plotted $r_{\text{time}}, r_m^{\text{BTTB}}$ and $r_c^{\text{BTTB}}$ in Figs. 3 and 4, respectively. From Fig. 3, we can see that, the time ratios $r_{\text{time}}$ 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 $r_m^{\text{BTTB}}$ and $r_c^{\text{BTTB}}$ 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 $X_\ast = [x_{i,j}^\ast]_{i=1, \ldots, m}^{j=1, \ldots, n}$ be the image to be restored and $L$ be defined as

\[
    \| LX_\ast \|_2^2 = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} |x_{i+1,j}^\ast - x_{i,j}^\ast|^2 + |x_{i,j+1}^\ast - x_{i,j}^\ast|^2 + \sum_{j=1}^{n-1} |x_{m,j+1}^\ast - x_{m,j}^\ast|^2 + \sum_{i=1}^{m-1} |x_{i+1,n}^\ast - x_{i,n}^\ast|^2.
\]
There is one equation in the text:
\[ \lambda = 0.09 \] as in [49].

The regularization parameter is set to \( \lambda = 0.09 \) as in [49]. Again, we generated 100 samples of the perturbations \( \Delta a \) and \( \Delta b \) and computed \( r_{\text{time}}, r_{\text{BTTB}}^a \), and \( r_{\text{BTTB}}^b \). 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 National Science and Engineering Council (NSERC) of Canada under grant RGPIN-2014-04252 and Shanghai Key Laboratory of Contemporary Applied Mathematics.

References