



---

*Research article*

## Condition numbers of the generalized ridge regression and its statistical estimation

Jing Kong<sup>1</sup> and Shaoxin Wang<sup>2,\*</sup>

<sup>1</sup> Qufu Mingde School, Qufu 273165, China

<sup>2</sup> School of Statistics and Data Science, Qufu Normal University, Qufu 273165, China

\* **Correspondence:** Email: shxwang@qfnu.edu.cn, shxwangmy@163.com.

**Abstract:** In this paper, we considered the condition number theory of a new generalized ridge regression model. The explicit expressions of different types of condition numbers were derived to measure the ill-conditionness of the generalized ridge regression problem with respect to different circumstances. To overcome the computational difficulty of computing the exact value of the condition number, we employed the statistical condition estimation theory to design efficient condition number estimators, and the numerical examples were also given to illustrate its efficiency.

**Keywords:** ridge regression, normwise condition number; mixed and componentwise condition numbers; statistical condition estimation

**Mathematics Subject Classification:** 15A12, 15A60, 65F35

---

### 1. Introduction

Ridge regression [1] is a classical and powerful statistical method to inhibit the multicollinearity among covariates in the following linear regression model:

$$y = X\beta + \epsilon, \tag{1.1}$$

where  $y \in \mathbb{R}^n$  is the observations of response,  $X \in \mathbb{R}^{n \times p}$  is the design matrix of covariates,  $\beta \in \mathbb{R}^p$  is the unknown parameter vector, and  $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$  with  $I_n$  being the identity matrix of order  $n$ . The ridge estimator of model (1.1) is given by

$$\hat{\beta}_r = (X^T X + \lambda I_n)^{-1} X^T y, \tag{1.2}$$

where  $\lambda > 0$  is the ridge parameter. Finding the ridge estimator (1.2) is also equivalent to solving the following optimization problem:

$$\hat{\beta}_r = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2, \quad (1.3)$$

where  $\|\cdot\|_2$  is the Euclidian vector norm. Equation (1.3) is also known as the famous Tikhonov regularization technique dealing with discrete ill-posed problems [2]. Incorporating weights to the observations and generalizing the ridge penalty leads to the following generalized ridge regression (GRR) model:

$$\hat{\beta}_{gr} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} (y - X\beta)^T \Sigma (y - X\beta) + (\beta - \beta_0)^T \Theta (\beta - \beta_0), \quad (1.4)$$

where  $\Sigma \in \mathbb{R}^{n \times n}$  is a diagonal matrix with nonnegative elements on its diagonal,  $\Theta \in \mathbb{R}^{p \times p}$  is a symmetric positive definite matrix and determines the speed and direction of shrinkage, and  $\beta_0$  is a user-specified, non-random target vector [3]. By the Karush-Kuhn-Tucker (KKT) optimality condition, taking the derivative of the objective function in (1.4) with respect to  $\beta$  shows that the generalized ridge estimator (GRE) denoted by  $\hat{\beta}_{gr}$  satisfies the following normal equation:

$$(X^T \Sigma X + \Theta) \beta = X^T \Sigma y + \Theta \beta_0 \quad (1.5)$$

and  $\hat{\beta}_{gr}$  is given by

$$\hat{\beta}_{gr} = (X^T \Sigma X + \Theta)^{-1} (X^T \Sigma y + \Theta \beta_0). \quad (1.6)$$

Though the diagonal matrix  $\Sigma$  can give different weights to the observations, the correlation structure between observations cannot be captured. For this reason, we relax the assumption and allow  $\Sigma$  to be a symmetric positive definite matrix, which makes (1.4) the exact combination of weighted least squares loss and the generalized ridge penalty. By varying the weighting matrices  $\Sigma$  and  $\Theta$ , the generalized Tikhonov regularization [2] can also be derived from (1.4).

The multicollinearity in covariates often leads to a large condition number of  $X^T X$ , which plays a central role in the sensitivity analysis of the least squares estimate of model (1.1) [4, Ch. 5]. Condition number is a powerful tool in measuring the sensitivity of a problem and gives the maximum amplification of the resulting change in solution, with respect to a small perturbation in the input data, and has been studied for too many numerical linear algebra problems to list here. The development of the definition of condition number and their comparison should be referred to [5–9]. In this paper, we employ the following very general definition of condition number given by [6], which covers most of the popular condition numbers in the current literature as its special cases.

**Definition 1.** (projected condition number) [6] Let  $\mathcal{F} : \mathbb{R}^p \rightarrow \mathbb{R}^q$  be a continuous map defined on an open set  $\operatorname{Dom}(\mathcal{F})$ , the domain of definition of  $\mathcal{F}$ , and  $L \in \mathbb{R}^{q \times k}$ , then the projected condition number of  $\mathcal{F}$  at  $x \in \operatorname{Dom}(\mathcal{F})$  with respect to  $L \in \mathbb{R}^{q \times k}$  is defined by

$$\kappa_{L\mathcal{F}}(x) = \lim_{\delta \rightarrow 0} \sup_{\left\| \frac{\Delta x}{\alpha} \right\|_{\mu} \leq \delta} \frac{\left\| \frac{L^T (\mathcal{F}(x+\Delta x) - \mathcal{F}(x))}{\xi_L} \right\|_{\nu}}{\left\| \frac{\Delta x}{\alpha} \right\|_{\mu}}, \quad (1.7)$$

where  $\cdot/\cdot$  is the componentwise division satisfying that a nonzero numerator divided by a zero denominator remains unchanged,  $\xi_L \in \mathbb{R}^k$  and  $\alpha \in \mathbb{R}^p$  are parameters with a requirement that if some element of  $\alpha$  is zero, then the corresponding element of  $\Delta x$  must also be zero, and  $\|\cdot\|_{\mu}$  and  $\|\cdot\|_{\nu}$  are two vector norms defined on  $\mathbb{R}^p$  and  $\mathbb{R}^k$ , respectively.

In Definition 1, the matrix  $L$  can be treated as an operator to transform the image vector of  $\mathcal{F}$ , for example, when we set  $L = e_i$ , the  $i$ -th element of the canonical base of  $\mathbb{R}^q$ , Definition 1 gives the condition number of the  $i$ -th element of the image vector. The two vector norms  $\|\cdot\|_\mu$  and  $\|\cdot\|_\nu$  also enable us to give different measures of the perturbations in the input and output spaces, respectively. When the map  $\mathcal{F}$  is Fréchet differentiable, we can get the following Theorem 1, which largely reduces the difficulty of finding the supremum of (1.7) to establishing the Fréchet derivative of  $\mathcal{F}$ . The rationality of Definition 1 and its relationship with other popular condition numbers has been extensively discussed, and the interested readers are referred to [6–8] and the references therein.

**Theorem 1.** [6] *When the map  $\mathcal{F}$  in Definition 1 is Fréchet differentiable at  $x$ , the explicit expression of the projected condition number  $\kappa_{L\mathcal{F}}(x)$  is given by*

$$\kappa_{L\mathcal{F}}(x) = \left\| \text{Diag}(\xi_L^\ddagger) L^T D\mathcal{F}(x) \text{Diag}(\alpha) \right\|_{\mu\nu},$$

where  $D\mathcal{F}(x)$  is the Fréchet derivative of  $\mathcal{F}$  at  $x$ ,  $\xi_L^\ddagger$  is a vector with  $\xi_{Li}^\ddagger = \begin{cases} 1/\xi_{Li}, & \xi_{Li} \neq 0, \\ 1, & \xi_{Li} = 0, \end{cases}$   $\text{Diag}(\xi_L^\ddagger)$  and  $\text{Diag}(\alpha)$  are diagonal matrices with the elements of the vector on the diagonal, and  $\|\cdot\|_{\mu\nu}$  is the matrix norm induced by the vector norms  $\|\cdot\|_\mu$  and  $\|\cdot\|_\nu$ .

The condition number theory of some special cases of model (1.4) have been well studied in the literature, like (weighted) linear least squares problem [6, 7, 10, 11] and (generalized) Tikhonov regularization [12–14]. In this paper, we will extend the current results and consider the condition number theory of the GRR model (1.4). To compute the exact value of the condition number is usually expensive, we also propose some efficient statistical condition estimation methods to reduce the computation burden but still give reasonable estimates. The rest of the paper is organized as follows. In Section 2, we present the explicit expressions of condition numbers for the GRR model (1.4). Section 3 contains the statistical condition estimation method for estimating the condition numbers of the GRR model (1.4). Numerical experiments and a conclusion of the whole paper are given as Sections 4 and 5, respectively.

## 2. The condition number of the GRR problem

In order to present the condition number of the GRR model (1.4), we first explicitly define the map  $\mathcal{F}$  in (1.7) as follows:

$$\begin{aligned} \mathcal{F} : \mathbb{R}^n \times \mathbb{R}^{n \times p} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{p \times p} \times \mathbb{R}^p &\rightarrow \mathbb{R}^p, \\ \mathcal{F}(y, X, \Sigma, \Theta, \beta_0) &\rightarrow \hat{\beta}_{gr} = (X^T \Sigma X + \Theta)^{-1} (X^T \Sigma y + \Theta \beta_0). \end{aligned} \quad (2.1)$$

Let  $\tilde{y} = y + \Delta y$ ,  $\tilde{X} = X + \Delta X$ ,  $\tilde{\Sigma} = \Sigma + \Delta \Sigma$ ,  $\tilde{\Theta} = \Theta + \Delta \Theta$  and  $\tilde{\beta}_0 = \beta_0 + \Delta \beta_0$ , then the perturbed GRR model is given by

$$\hat{\beta}_{\tilde{g}r} = \underset{\beta \in \mathbb{R}^p}{\text{argmin}} (\tilde{y} - \tilde{X}\beta)^T \tilde{\Sigma} (\tilde{y} - \tilde{X}\beta) + (\beta - \tilde{\beta}_0)^T \tilde{\Theta} (\beta - \tilde{\beta}_0), \quad (2.2)$$

where  $\hat{\beta}_{\tilde{g}r}$  is the GRE of the perturbed GRR model (2.2) and satisfies the following normal equation:

$$(\tilde{X}^T \tilde{\Sigma} \tilde{X} + \tilde{\Theta}) \hat{\beta}_{\tilde{g}r} = \tilde{X}^T \tilde{\Sigma} \tilde{y} + \tilde{\Theta} \tilde{\beta}_0. \quad (2.3)$$

For simplicity of presentation and with a little abuse of notation, we set

$$\text{vec}(y, X, \Sigma, \Theta, \beta_0) := [y^T, \text{vec}(X)^T, \text{vec}(\Sigma)^T, \text{vec}(\Theta)^T, \beta_0^T]^T$$

then the projected condition number of the GRR model (1.4) can be defined as follows.

**Definition 2.** Considering the perturbed GRR model (2.2), the projected condition number of the GRR model (1.4), with respect to  $L \in \mathbb{R}^{q \times k}$ , is defined by

$$\kappa_{L\mathcal{F}} = \lim_{\delta \rightarrow 0} \sup_{\left\| \text{vec} \left( \frac{\Delta y}{\gamma}, \frac{\Delta X}{\Psi}, \frac{\Delta \Sigma}{\Phi}, \frac{\Delta \Theta}{\Upsilon}, \frac{\Delta \beta_0}{\varpi} \right) \right\|_{\mu} \leq \delta} \frac{\left\| \frac{L^T (\mathcal{F}(\tilde{y}, \tilde{X}, \tilde{\Sigma}, \tilde{\Theta}, \tilde{\beta}_0) - \mathcal{F}(y, X, \Sigma, \Theta, \beta_0))}{\xi_L} \right\|_{\nu}}{\left\| \text{vec} \left( \frac{\Delta y}{\gamma}, \frac{\Delta X}{\Psi}, \frac{\Delta \Sigma}{\Phi}, \frac{\Delta \Theta}{\Upsilon}, \frac{\Delta \beta_0}{\varpi} \right) \right\|_{\mu}},$$

where  $\xi_L \in \mathbb{R}^k$ ,  $\gamma \in \mathbb{R}^n$ ,  $\Psi \in \mathbb{R}^{n \times p}$ ,  $\Phi \in \mathbb{R}^{n \times n}$ ,  $\Upsilon \in \mathbb{R}^{p \times p}$ ,  $\varpi \in \mathbb{R}^p$  are parameters with a requirement that if some element is zero, then the corresponding element in the numerator must also be zero, and  $\|\cdot\|_{\mu}$  and  $\|\cdot\|_{\nu}$  are two vector norms defined on  $\mathbb{R}^{n+np+n^2+p^2+p}$  and  $\mathbb{R}^k$ , respectively.

If we set  $\hat{\beta}_{gr} = \hat{\beta}_{gr} + \Delta\beta$  and further assume that

$$\max \{ \|\Delta y\|, \|\Delta \Sigma\|, \|\Delta X\|, \|\Delta \Theta\|, \|\Delta \beta_0\| \} \leq \epsilon$$

with  $\|\cdot\|$  denoting appropriate norm and  $\epsilon$  being a sufficiently small positive real number, then with some algebra and omitting the higher order terms, we can get the following equality by subtracting (1.5) from (2.3)

$$\begin{aligned} \mathcal{W}\Delta\beta &= X^T \Sigma (\Delta y - \Delta X \hat{\beta}_{gr}) + X^T \Delta \Sigma r + \Delta X^T \Sigma r \\ &\quad + \Theta \Delta \beta_0 + \Delta \Theta (\beta_0 - \hat{\beta}_{gr}) + \mathcal{O}(\epsilon^2), \end{aligned} \quad (2.4)$$

where  $\mathcal{W} = X^T \Sigma X + \Theta$  and  $r = y - X \hat{\beta}_{gr}$ . Backing to Definition 2, we have

$$\mathcal{F}(\tilde{y}, \tilde{X}, \tilde{\Sigma}, \tilde{\Theta}, \tilde{\beta}_0) - \mathcal{F}(y, X, \Sigma, \Theta, \beta_0) = \Delta\beta$$

and

$$\begin{aligned} \Delta\beta &= \mathcal{W}^{-1} X^T \Sigma (\Delta y - \Delta X \hat{\beta}_{gr}) + \mathcal{W}^{-1} X^T \Delta \Sigma r + \mathcal{W}^{-1} \Delta X^T \Sigma r \\ &\quad + \mathcal{W}^{-1} \Theta \Delta \beta_0 + \mathcal{W}^{-1} \Delta \Theta (\beta_0 - \hat{\beta}_{gr}) + \mathcal{O}(\epsilon^2). \end{aligned} \quad (2.5)$$

With (2.5), the first order expansion of  $\Delta\beta$ , we can establish the explicit expression of the projected condition number for the GRR model (1.4).

**Theorem 2.** For the GRR model (1.4), when  $\Sigma$  and  $\Theta$  are symmetric positive definite matrices, the explicit expression of its projected condition number can be established and given as

$$\kappa_{L\mathcal{F}} = \left\| \text{Diag}(\xi_L^{\ddagger}) L^T \left[ \mathcal{W}^{-1} X^T \Sigma, \mathcal{M}, r^T \otimes (\mathcal{W}^{-1} X^T), \mathcal{N}, \mathcal{W}^{-1} \Theta \right] \text{Diag}(\text{vec}(\gamma, \Psi, \Phi, \Upsilon, \varpi)) \right\|_{\mu\nu},$$

where  $\mathcal{W} = X^T \Sigma X + \Theta$ ,  $r = y - X \hat{\beta}_{gr}$ ,  $\mathcal{M} = \mathcal{W}^{-1} \otimes (\Sigma r)^T - \hat{\beta}_{gr} \otimes (\mathcal{W}^{-1} X^T \Sigma)$  and  $\mathcal{N} = (\beta_0 - \hat{\beta}_{gr})^T \otimes \mathcal{W}^{-1}$ .

*Proof.* Since  $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$ , we apply  $\text{vec}(\cdot)$  operator to  $\Delta\beta$  and get

$$\begin{aligned} \Delta\beta &= \mathcal{W}^{-1} X^T \Sigma \Delta y + \left( \mathcal{W}^{-1} \otimes (\Sigma r)^T - \hat{\beta}_{gr}^T \otimes (\mathcal{W}^{-1} X^T \Sigma) \right) \text{vec}(\Delta X) \\ &\quad + \left( r^T \otimes (\mathcal{W}^{-1} X^T) \right) \text{vec}(\Delta \Sigma) + \left( (\beta_0 - \hat{\beta}_{gr})^T \otimes \mathcal{W}^{-1} \right) \text{vec}(\Delta \Theta) \\ &\quad + \mathcal{W}^{-1} \Theta \Delta \beta_0 + \mathcal{O}(\epsilon^2). \end{aligned} \quad (2.6)$$

Let  $\mathcal{M} = \mathcal{W}^{-1} \otimes (\Sigma r)^T - \hat{\beta}_{gr}^T \otimes (\mathcal{W}^{-1} X^T \Sigma)$  and  $\mathcal{N} = (\beta_0 - \hat{\beta}_{gr})^T \otimes \mathcal{W}^{-1}$ . We can rewrite (2.6) as

$$\Delta\beta = \begin{bmatrix} \mathcal{W}^{-1} X^T \Sigma & \mathcal{M} & r^T \otimes (\mathcal{W}^{-1} X^T) & \mathcal{N} & \mathcal{W}^{-1} \Theta \end{bmatrix} \begin{bmatrix} \Delta y \\ \text{vec}(\Delta X) \\ \text{vec}(\Delta \Sigma) \\ \text{vec}(\Delta \Theta) \\ \Delta \beta_0 \end{bmatrix} + \mathcal{O}(\epsilon^2). \quad (2.7)$$

With (2.7), we can get the following Fréchet derivative of  $\mathcal{F}$

$$D\mathcal{F} = \begin{bmatrix} \mathcal{W}^{-1} X^T \Sigma & \mathcal{M} & r^T \otimes (\mathcal{W}^{-1} X^T) & \mathcal{N} & \mathcal{W}^{-1} \Theta \end{bmatrix} \quad (2.8)$$

and complete the proof by Theorem 1.  $\square$

Theorem 2 presents the generic form of the condition number for the GRR problem (1.4). For practical applications, we need to specify the norms and the parameters with respect to concrete backgrounds. We also note that the explicit expression of the condition number contains specific structures due to the Kronecker product, which enlarges the size of the matrix and increases its computational burden. In the following, we will discuss some specific forms of the condition number and its computational issues.

### 2.1. The normwise condition number

When  $\mu = \nu = 2$ , the parameter vectors and matrices  $\gamma$ ,  $\Psi$ ,  $\Phi$ ,  $\Upsilon$ , and  $\varpi$  reduce to real numbers and equal to  $\|\text{vec}([y, X, \Sigma, \Theta, \beta_0])\|_2$  and  $\xi_L = \|L^T \hat{\beta}_{gr}\|_2 \neq 0$ ,  $\kappa_{L\mathcal{F}}$  gives an overall treatment of the perturbations, then we can obtain the projected relative normwise condition number of the GRR model (1.4) from Theorem 2, which is given as follows.

**Theorem 3.** *When  $\mu = \nu = 2$ ,  $\gamma = \Psi = \Phi = \Upsilon = \varpi = \|\text{vec}(y, X, \Sigma, \Theta, \beta_0)\|_2$ , and  $\xi_L = \|L^T \hat{\beta}_{gr}\|_2 \neq 0$ , the projected relative normwise condition number of the GRR model (1.4) is given by*

$$\kappa_{L\mathcal{F}}^2 = \frac{\left\| L^T \begin{bmatrix} \mathcal{W}^{-1} X^T \Sigma & \mathcal{M} & r^T \otimes (\mathcal{W}^{-1} X^T) & \mathcal{N} & \mathcal{W}^{-1} \Theta \end{bmatrix} \right\|_2 \|\text{vec}(y, X, \Sigma, \Theta, \beta_0)\|_2}{\|L^T \hat{\beta}_{gr}\|_2},$$

where  $\|\cdot\|_2$  denotes the spectral norm of a matrix or the Euclidian norm of a vector,  $\mathcal{W} = X^T \Sigma X + \Theta$ ,  $r = y - X \hat{\beta}_{gr}$ ,  $\mathcal{M} = \mathcal{W}^{-1} \otimes (\Sigma r)^T - \hat{\beta}_{gr}^T \otimes (\mathcal{W}^{-1} X^T \Sigma)$  and  $\mathcal{N} = (\beta_0 - \hat{\beta}_{gr})^T \otimes \mathcal{W}^{-1}$ .

Note that the main difficulty for explicitly computing the value of  $\kappa_{L\mathcal{F}}^2$  lies in the following term

$$\kappa_{L\mathcal{F}}^{2a} := \left\| L^T \begin{bmatrix} \mathcal{W}^{-1} X^T \Sigma & \mathcal{M} & r^T \otimes (\mathcal{W}^{-1} X^T) & \mathcal{N} & \mathcal{W}^{-1} \Theta \end{bmatrix} \right\|_2, \quad (2.9)$$

which contains the Kronecker product and is also called the projected absolute normwise condition number. To remove the Kronecker product in  $\kappa_{L\mathcal{F}}^{2a}$ , the matrix cross-product and Cholesky factorization techniques may be used to establish the compact but equivalent forms of the normwise condition number [6, 7]. However for large problems, there is no need to form the condition number explicitly and a suitable estimate will be enough [15, Ch. 15]. Here, we present a compact form of  $\kappa_{L\mathcal{F}}^{2a}$  that not only removes the Kronecker product, but also provides important support for its estimation procedure described in Section 3.

**Theorem 4.** When  $\mu = \nu = 2$ ,  $\gamma = \Psi = \Phi = \Upsilon = \varpi = \|\text{vec}(y, X, \Sigma, \Theta, \beta_0)\|_2$ , and  $\xi_L = \|L^T \hat{\beta}_{gr}\|_2 \neq 0$ , the projected absolute normwise condition number  $\kappa_{L\mathcal{F}}^{2a}$  of the GRR model (1.4) can also be written as

$$\kappa_{L\mathcal{F}}^{2a} = \left\| L^T \mathcal{W}^{-1} \mathcal{K} \mathcal{W}^{-1} L - L^T \mathcal{W}^{-1} \left( \hat{\beta}_{gr} r^T \Sigma^2 X + X^T \Sigma^2 r \hat{\beta}_{gr}^T \right) \mathcal{W}^{-1} L \right\|_2^{\frac{1}{2}}. \quad (2.10)$$

In particular, when  $L = e_i$ , we can get the projected absolute normwise condition number of the  $i$ -th element in the solution

$$\kappa_{e_i\mathcal{F}}^{2a} = \left\| e_i^T \mathcal{W}^{-1} \mathcal{K} \mathcal{W}^{-1} e_i - e_i^T \mathcal{W}^{-1} \left( \hat{\beta}_{gr} r^T \Sigma^2 X + X^T \Sigma^2 r \hat{\beta}_{gr}^T \right) \mathcal{W}^{-1} e_i \right\|_2^{\frac{1}{2}}, \quad (2.11)$$

where  $\mathcal{W} = X^T \Sigma X + \Theta$ ,  $r = y - X \hat{\beta}_{gr}$  and  $\mathcal{K} = (1 + \|\hat{\beta}_{gr}\|_2^2) X^T \Sigma^2 X + \|r\|_2^2 X^T X + (\|\beta_0 - \hat{\beta}_{gr}\|_2^2 + r^T \Sigma^2 r) I_p + \Theta^2$ .

*Proof.* For the spectral norm, we have  $\|A\|_2 = \|AA^T\|_2^{\frac{1}{2}}$  with  $A \in \mathbb{R}^{m \times n}$ . We apply this equality to  $\kappa_{L\mathcal{F}}^{2a}$  and get its equivalent form

$$\kappa_{L\mathcal{F}}^{2a} := \left\| L^T \begin{bmatrix} \mathcal{W}^{-1} X^T \Sigma & \mathcal{M} & r^T \otimes (\mathcal{W}^{-1} X^T) & \mathcal{N} & \mathcal{W}^{-1} \Theta \end{bmatrix} \begin{bmatrix} \Sigma X \mathcal{W}^{-1} \\ \mathcal{M}^T \\ r \otimes (X \mathcal{W}^{-1}) \\ \mathcal{N}^T \\ \Theta \mathcal{W}^{-1} \end{bmatrix} L \right\|_2^{\frac{1}{2}}.$$

Since

$$\begin{aligned} \mathcal{M} \mathcal{M}^T &= (\mathcal{W}^{-1} \otimes (\Sigma r)^T - \hat{\beta}_{gr}^T \otimes (\mathcal{W}^{-1} X^T \Sigma)) (\mathcal{W}^{-1} \otimes (\Sigma r) - \hat{\beta}_{gr} \otimes (\Sigma X \mathcal{W}^{-1})) \\ &= r^T \Sigma^2 r \mathcal{W}^{-2} - \mathcal{W}^{-1} \hat{\beta}_{gr} r^T \Sigma^2 X \mathcal{W}^{-1} - \mathcal{W}^{-1} X^T \Sigma^2 r \hat{\beta}_{gr}^T \mathcal{W}^{-1} + \|\hat{\beta}_{gr}\|_2^2 \mathcal{W}^{-1} X^T \Sigma^2 X \mathcal{W}^{-1} \end{aligned}$$

and

$$\mathcal{N} \mathcal{N}^T = (\beta_0 - \hat{\beta}_{gr})^T \otimes \mathcal{W}^{-1} (\beta_0 - \hat{\beta}_{gr}) \otimes \mathcal{W}^{-1} = \|\beta_0 - \hat{\beta}_{gr}\|_2^2 \mathcal{W}^{-2},$$

we can easily get

$$\kappa_{L\mathcal{F}}^{2a} = \left\| L^T \mathcal{W}^{-1} \mathcal{K} \mathcal{W}^{-1} L - L^T \mathcal{W}^{-1} \left( \hat{\beta}_{gr} r^T \Sigma^2 X + X^T \Sigma^2 r \hat{\beta}_{gr}^T \right) \mathcal{W}^{-1} L \right\|_2^{\frac{1}{2}}$$

with  $\mathcal{K} = (1 + \|\hat{\beta}_{gr}\|_2^2) X^T \Sigma^2 X + \|r\|_2^2 X^T X + (\|\beta_0 - \hat{\beta}_{gr}\|_2^2 + r^T \Sigma^2 r) I_p + \Theta^2$ .  $\square$

**Remark 1.** Theorem 4 presents a compact but equivalent form  $\kappa_{L\mathcal{F}1}^{2a}$  of the absolute condition number  $\kappa_{L\mathcal{F}}^{2a}$  that does not contain the Kronecker product. Comparing the size of the matrix, we can find that  $\kappa_{L\mathcal{F}1}^{2a}$  requires much less storage memory compared with  $\kappa_{L\mathcal{F}}^{2a}$ . When the exact value of the normwise condition number is computed,  $\kappa_{L\mathcal{F}1}^{2a}$  also needs much less Central Processing Unit (CPU) time on a computer, and this will be illustrated through numerical experiment in Section 4. Here, we also need to point out that the Cholesky factorization technique was not employed to derive the compact form of  $\kappa_{L\mathcal{F}}^{2a}$ . The main reason is that the Cholesky factorization based compact form only gives a moderate size of matrix, which is still larger than  $\kappa_{L\mathcal{F}1}^{2a}$ , though much smaller than  $\kappa_{L\mathcal{F}}^{2a}$ . Thus, considering the economics of storage space, we only derived  $\kappa_{L\mathcal{F}1}^{2a}$ .

## 2.2. The mixed and componentwise condition numbers

Since the normwise condition number gives an overall treatment of all the parameters and ignores the data structure and scaling of the input data, the mixed and componentwise condition numbers are proposed [8,9]. By varying the norms and parameters in Theorem 2, we can also obtain the mixed and componentwise condition numbers of the GRR model (1.4).

**Theorem 5.** When  $\mu = \nu = \infty$ ,  $\gamma = y$ ,  $\Psi = X$ ,  $\Phi = \Sigma$ ,  $\Upsilon = \Theta$ ,  $\varpi = \beta_0$ ,  $\xi_L = \|L^T \hat{\beta}_{gr}\|_\infty$ , and  $L^T \hat{\beta}_{gr}$  in sequel, the projected mixed and componentwise condition numbers of the GRR model (1.4) are given by

$$\kappa_{mL\mathcal{F}}^\infty = \frac{\left\| \left\| L^T \left[ \mathcal{W}^{-1} X^T \Sigma, \mathcal{M}, r^T \otimes (\mathcal{W}^{-1} X^T), \mathcal{N}, \mathcal{W}^{-1} \Theta \right] \right\|_{\text{vec}([y, X, \Sigma, \Theta, \beta_0])} \right\|_\infty}{\|L^T \hat{\beta}_{gr}\|_\infty}$$

and

$$\kappa_{cL\mathcal{F}}^\infty = \left\| \frac{\left\| L^T \left[ \mathcal{W}^{-1} X^T \Sigma, \mathcal{M}, r^T \otimes (\mathcal{W}^{-1} X^T), \mathcal{N}, \mathcal{W}^{-1} \Theta \right] \right\|_{\text{vec}([y, X, \Sigma, \Theta, \beta_0])}}{\|L^T \hat{\beta}_{gr}\|} \right\|_\infty,$$

respectively, where  $\|\cdot\|_\infty$  is the infinite norm giving the largest magnitude among each element of a vector.

*Proof.* The proof is quite easy due to the following fact that for matrix  $A$  and vector  $d$ , the following equalities hold

$$\|A \text{Diag}(d)\|_\infty = \| |A| |\text{Diag}(d)| \|_\infty = \| |A| \|d\|_\infty.$$

Applying the above equalities to Theorem 2 gives the desired results.  $\square$

For the mixed and componentwise condition numbers, the aforementioned matrix-cross product techniques cannot be used to remove its Kronecker product. Here, we present some upper bounds that require little storage space and can be efficiently computed.

**Theorem 6.** The mixed and componentwise condition numbers of the GRR model (1.4) can be correspondingly bounded as follows:

$$\kappa_{mL\mathcal{F}}^{\infty ubd} = \frac{\|\mathcal{G}^{ubd}\|_\infty}{\|L^T \hat{\beta}_{gr}\|_\infty} \quad \text{and} \quad \kappa_{cL\mathcal{F}}^{\infty ubd} = \left\| \frac{\mathcal{G}^{ubd}}{L^T \hat{\beta}_{gr}} \right\|_\infty,$$

where

$$\begin{aligned} \mathcal{G}^{ubd} = & \left| L^T \mathcal{W}^{-1} X^T \Sigma \right| \left( |y| + |X| |\hat{\beta}_{gr}| \right) + \left| L^T \mathcal{W}^{-1} X^T \right| |\Sigma| |r| + \left| L^T \mathcal{W}^{-1} \right| |X^T| |\Sigma r| \\ & + \left| L^T \mathcal{W}^{-1} \Theta \right| |\beta_0| + \left| L^T \mathcal{W}^{-1} \right| |\Theta| (|\beta_0 - \hat{\beta}_{gr}|). \end{aligned}$$

*Proof.* The explicit expression of  $\mathcal{G}^{ubd}$  is derived with the Lemma 5 in [10], which is straightforward and omitted here.  $\square$

The condition number not only gives a measure of the sensitivity of the problem, but also can be used to give the first order estimate of the forward error. Thus, in many practical applications, the exact value of the condition number is usually not needed and within a 10 factor estimate will be enough. Many techniques have been proposed to estimate the normwise, mixed, and componentwise condition numbers, like the power method [15], probabilistic, and statistical based methods [16]. Considering the computational efficiency and its powerful adaptability, we would like to employ the statistical condition estimation method to estimate the condition numbers of the GRR model.

### 3. The small-sample statistical condition estimation

The small-sample statistical condition estimation (SSCE) theory has been widely applied to estimate the normwise, mixed, and componentwise condition numbers of many numerical linear algebra problems, and examples include the linear system [17], least squares problem [18, 19], matrix factorization [20, 21], eigenvalue problem [22], matrix function [16], Sylvester equation [23], and so on.

To introduce the framework of SSCE, we consider the following differentiable function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ , and its Taylor expansion is given by

$$f(x + \delta z) = f(x) + \delta \nabla f(x)^T z + \mathcal{O}(\delta^2),$$

where  $\delta$  is a small positive real number,  $\nabla f(x)$  is the gradient vector of  $f$  at  $x$ , and  $z$  is a unit 2-norm vector. It is well known that  $\|\nabla f(x)\|_2$  is the absolute condition number and gives an appropriate measure of the local sensitivity of  $f$  at  $x$ . According to [16], if we choose  $z$  uniformly and randomly from a unit sphere  $\mathcal{S}_{p-1}$ , then the expected value of  $|\nabla f(x)^T z|$  is

$$\mathbf{E}(|\nabla f(x)^T z|) = \|\nabla f(x)\|_2 E_p,$$

where  $E_1 = 1$ ,  $E_2 = \frac{\pi}{2}$ , and for  $p > 2$ ,

$$E_p = \begin{cases} \frac{1 \cdot 3 \cdot 5 \cdots (p-2)}{2 \cdot 4 \cdot 6 \cdots (p-1)}, & \text{for } p \text{ odd,} \\ \frac{2}{\pi} \frac{2 \cdot 4 \cdot 6 \cdots (p-2)}{3 \cdot 5 \cdot 7 \cdots (p-1)}, & \text{for } p \text{ even.} \end{cases}$$

$E_p$  is the Wallis factor and can be approximated by  $E_p \approx \sqrt{\frac{2}{\pi(p-\frac{1}{2})}}$  with high accuracy [18], then we can define

$$\eta \equiv \frac{|\nabla f(x)^T z|}{E_p}$$



as the condition estimator, which can give a very reliable estimate of  $\|\nabla f(x)\|_2$ ; specifically, the following probability inequality holds

$$\mathbb{P}\left(\frac{\|\nabla f(x)\|_2}{\tau} \leq \eta \leq \tau\|\nabla f(x)\|_2\right) \geq 1 - \frac{2}{\pi\tau} + \mathcal{O}\left(\frac{1}{\tau^2}\right).$$

The accuracy of condition estimator can be further improved by adding  $k$  samples

$$\eta(k) \equiv \frac{E_k}{E_p} \sqrt{|\nabla f(x)^T z_1|^2 + |\nabla f(x)^T z_2|^2 + \cdots + |\nabla f(x)^T z_k|^2},$$

where  $[z_1, z_2, \dots, z_k]$  is the orthonormalization of  $z_1, z_2, \dots, z_k$  being selected uniformly and randomly from  $\mathcal{S}_{p-1}$ . The  $k$ -sample condition estimator  $\eta(k)$  can achieve a very high accuracy with a small size of samples, for example, when  $k = 3$  and  $\tau = 10$ , we have

$$\mathbb{P}\left(\frac{\|\nabla f(x)\|_2}{10} \leq \eta(3) \leq 10\|\nabla f(x)\|_2\right) \approx 0.9988389,$$

which means the reliability of a condition estimate within a factor 10 can be improved from 0.936338 ( $k = 1$ ) to 0.9988389 by just adding 2 extra samples.

### 3.1. Normwise condition estimation

Note that the SSCE method is very suitable for estimating the condition number of certain elements in the solution. However, for the normwise condition number derived in Section 2.1, what we need to estimate is the spectral norm of a large matrix. This means we need to make some modification on the SSCE method to estimate the normwise condition number of the GRR model (1.4). Here, we employ the strategy proposed by [19] for estimating the normwise condition number of the linear least squares problem.

According to [19], to estimate the normwise condition number of the GRR model (1.4), we first need to estimate the condition number of  $z_i^T x$  with

$$\kappa_i = \left\| z_i^T \mathcal{W}^{-1} \mathcal{K} \mathcal{W}^{-1} z_i - z_i^T \mathcal{W}^{-1} \left( \hat{\beta}_{gr} r^T \Sigma^2 X + X^T \Sigma^2 r \hat{\beta}_{gr}^T \right) \mathcal{W}^{-1} z_i \right\|_2^{\frac{1}{2}},$$

and then the normwise condition number can be estimated by

$$\kappa_{\mathbf{N}} = \frac{E_k}{E_p} \left( \sum_{i=1}^k \kappa_i^2 \right), \quad (3.1)$$

where  $[z_1, \dots, z_k]$  can be obtained via QR factorization [4, Ch. 5] of a random matrix  $Z \in \mathbb{R}^{p \times k}$ . Note that when  $\mathcal{W}$  and  $\mathcal{K}$  are available, the main computational cost for computing  $\kappa_i$  is  $\mathcal{O}(\frac{1}{3}p^3)$  when Cholesky factorization is used to compute  $\mathcal{W}^{-1} z_i$ . If we further take the QR factorization into account, then we can find that the total computational cost for estimating the normwise condition number is  $\mathcal{O}(\frac{k}{3}p^3 + pk^2)$ . We summarize the above procedure as the following Algorithm 1.

---

**Algorithm 1** Absolute normwise condition number estimator.
 

---

- (1) Generate  $k$  vectors  $z_1, \dots, z_k \in \mathbb{R}^p$  with entries from uniform distribution  $\mathcal{U}(0, 1)$ .
- (2) Orthonormalize the vectors  $z_1, \dots, z_k$  with QR factorization.
- (3) Repeat  $k$  times for computing

$$\kappa_i = \left\| z_i^T \mathcal{W}^{-1} \mathcal{K} \mathcal{W}^{-1} z_i - z_i^T \mathcal{W}^{-1} (\hat{\beta}_{gr} r^T \Sigma^2 X + X^T \Sigma^2 r \hat{\beta}_{gr}^T) \mathcal{W}^{-1} z_i \right\|_2^{\frac{1}{2}},$$

with  $i = 1, \dots, k$ , and get  $\kappa_1, \dots, \kappa_k$ .

- (4) Compute the absolute normwise condition number estimator by

$$\kappa_N = \frac{E_k}{E_p} \left( \sum_{i=1}^k \kappa_i^2 \right), \text{ with } E_p = \sqrt{\frac{2}{\pi(p - \frac{1}{2})}}.$$


---

### 3.2. Mixed and componentwise condition estimation

From Theorem 5, we note that computing the mixed and componentwise condition numbers is equivalent to finding the largest elements in a vector. The SSCE method can be easily applied to estimate the mixed and componentwise condition numbers by extending the aforementioned SSCE procedure from scalar valued function to vector valued function. We present it in the following Algorithm 2. Note that when the orthonormal vectors are obtained, the computational complexity of SSCE for mixed and componentwise condition numbers also lies in solving the positive definite linear system, which makes it similar to that of Algorithm 1.

## 4. Numerical experiment

To illustrate our theoretical results, we will use a randomly generated GRR model with a known solution. The desired GRR model is constructed as follows. The coefficient matrices are given by

$$\Sigma = U_1 \Lambda_1 U_1^T, \quad \Theta = V_1 \Lambda_2 V_1^T, \quad \text{and} \quad X = U_2 \begin{bmatrix} \Lambda_3 \\ 0 \end{bmatrix} V_2^T, \quad (4.1)$$

where  $U_i$  and  $V_i$  ( $i = 1, 2$ ) are random orthogonal matrices and  $\Lambda_i$  ( $i = 1, 2, 3$ ) are diagonal matrices with  $\lambda_{ij}$  arranged in a descending order on its diagonal. Actually, (4.1) exactly gives the eigenvalue decomposition of  $\Sigma$  and  $\Theta$  and the singular value decomposition of  $X$  [4].  $\Lambda_1$  and  $\Lambda_2$  contain the eigenvalues of  $\Sigma$  and  $\Theta$ , and  $\Lambda_3$  the singular values of  $X$ .  $\beta$  is a random vector generated from the standard normal distribution.  $r$  is a random vector with specified magnitude, that is,  $\|r\|_2$  is given. Then, based on the normal Eq (1.5), we set

$$y = r - X\beta \quad \text{and} \quad \beta_0 = \beta - \Theta^{-1} X^T \Sigma r,$$

and get the random GRR model with specified solution. To compute  $\hat{\beta}_{gr}$ , the preconditioned conjugate gradients (PCG) method will be employed to solve the normal Eq (1.5). With the above setting, we

can control the condition number of the coefficient matrices and easily get

$$\kappa(\Sigma) = \|\Sigma\|_2 \|\Sigma^{-1}\|_2 = \frac{\lambda_{11}}{\lambda_{1n}}, \quad \kappa(\Theta) = \frac{\lambda_{21}}{\lambda_{2p}}, \quad \text{and} \quad \kappa(X) = \frac{\lambda_{31}}{\lambda_{3n}}. \quad (4.2)$$

All the experiments are performed in Matlab R2018a on a PC with Intel i7-10700 CPU @ 2.90 GHz and 16.00 GB random access memory.

---

**Algorithm 2** Mixed and componentwise condition numbers estimator.

---

(1) Generate  $k$  groups of matrices and vectors

$$(\Delta y_1, \Delta X_1, \Delta \Sigma_1, \Delta \Theta_1, \Delta \beta_{01}), \dots, (\Delta y_k, \Delta X_k, \Delta \Sigma_k, \Delta \Theta_k, \Delta \beta_{0k})$$

with elements from the standard normal distribution  $\mathcal{N}(0, 1)$ .

(2) Obtain the orthonormal vectors  $[q_1, \dots, q_k]$  by orthonormalizing the following matrix

$$\begin{bmatrix} \Delta y_1 & \dots & \Delta y_k \\ \text{vec}(\Delta X_1) & \dots & \text{vec}(\Delta X_k) \\ \text{vec}(\Delta \Sigma_1) & \dots & \text{vec}(\Delta \Sigma_k) \\ \text{vec}(\Delta \Theta_1) & \dots & \text{vec}(\Delta \Theta_k) \\ \Delta \beta_{01} & \dots & \Delta \beta_{0k} \end{bmatrix},$$

and reconstruct  $(\Delta y_1, \Delta X_1, \Delta \Sigma_1, \Delta \Theta_1, \Delta \beta_{01}), \dots, (\Delta y_k, \Delta X_k, \Delta \Sigma_k, \Delta \Theta_k, \Delta \beta_{0k})$  with the corresponding orthonormal vectors  $[q_1, \dots, q_k]$ , respectively.

(3) For  $i = 1, \dots, k$ , compute

$$x_i = \mathcal{W}^{-1} X^T \Sigma (\Delta y_i - \Delta X_i \hat{\beta}_{gr}) + \mathcal{W}^{-1} X^T \Delta \Sigma_i r + \mathcal{W}^{-1} \Delta X_i^T \Sigma r + \mathcal{W}^{-1} \Theta \Delta \beta_{0i} + \mathcal{W}^{-1} \Delta \Theta_i (\beta_0 - \hat{\beta}_{gr})$$

(4) Estimate the absolute condition vector with  $k$  samples by

$$C_{\text{abs}}^{\text{GRR}}(k) = \frac{E_k}{E_p} \sqrt{\sum_{i=1}^k |x_i|^2},$$

where the power and absolute operation are performed on the elements of the vectors.

(5) The mixed and componentwise condition number estimators are given by

$$\kappa_{\mathbf{M}} = \frac{\|C_{\text{abs}}^{\text{GRR}}(k)\|_{\infty}}{\|\hat{\beta}_{gr}\|_{\infty}} \quad \text{and} \quad \kappa_{\mathbf{C}} = \left\| \frac{C_{\text{abs}}^{\text{GRR}}(k)}{\hat{\beta}_{gr}} \right\|_{\infty},$$

respectively.

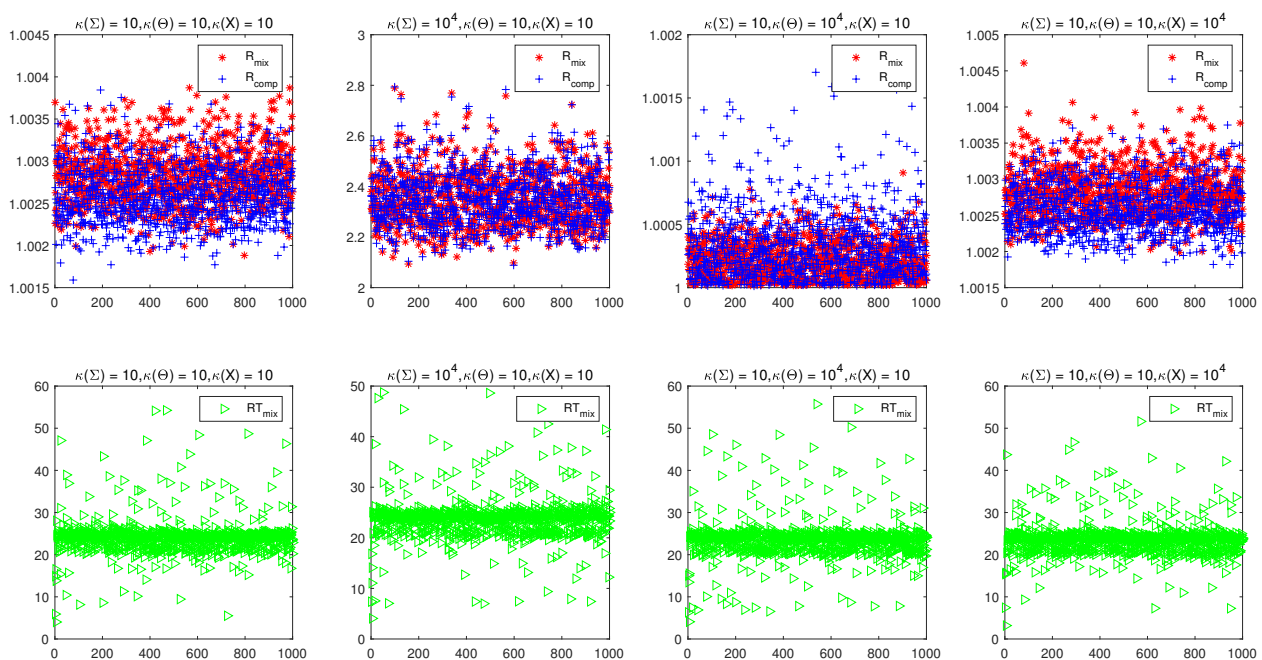
---

**Example 1.** In this example, we will test the tightness and computational efficiency of the upper bounds of the mixed and componentwise condition numbers derived in Theorem 6. For this, we define the

following ratios:

$$R_{\text{mix}} = \frac{\kappa_{\text{mLF}}^{\text{ubd}}}{\kappa_{\text{mLF}}^{\infty}}, R_{\text{comp}} = \frac{\kappa_{\text{cLF}}^{\text{ubd}}}{\kappa_{\text{cLF}}^{\infty}}, RT_{\text{mix}} = \frac{\kappa_{\text{mLF}}^{\text{ubd}} \text{CPU}}{\kappa_{\text{mLF}}^{\infty} \text{CPU}}, \text{ and } RT_{\text{comp}} = \frac{\kappa_{\text{cLF}}^{\text{ubd}} \text{CPU}}{\kappa_{\text{cLF}}^{\infty} \text{CPU}}.$$

$R_{\text{mix}}$  and  $R_{\text{comp}}$  demonstrate that the closer of the ratios are to 1, the tighter the upper bounds are.  $RT_{\text{mix}}$  and  $RT_{\text{comp}}$  measure the computational efficiency of the upper bounds with respect to CPU time and the larger value shows better computational efficiency. For a clear presentation, we present the numerical results in Figure 1, using a red asterisk and blue plus sign to denote  $R_{\text{mix}}$  and  $R_{\text{comp}}$ , correspondingly. Note that there is very little difference in computational complexity between mixed and componentwise condition numbers and its upper bounds. We only present the results of CPU time comparison of the mixed condition number and its upper bound denoted by a green square.



**Figure 1.** Tightness and computational efficiency of the upper bounds for the mixed and componentwise condition numbers.

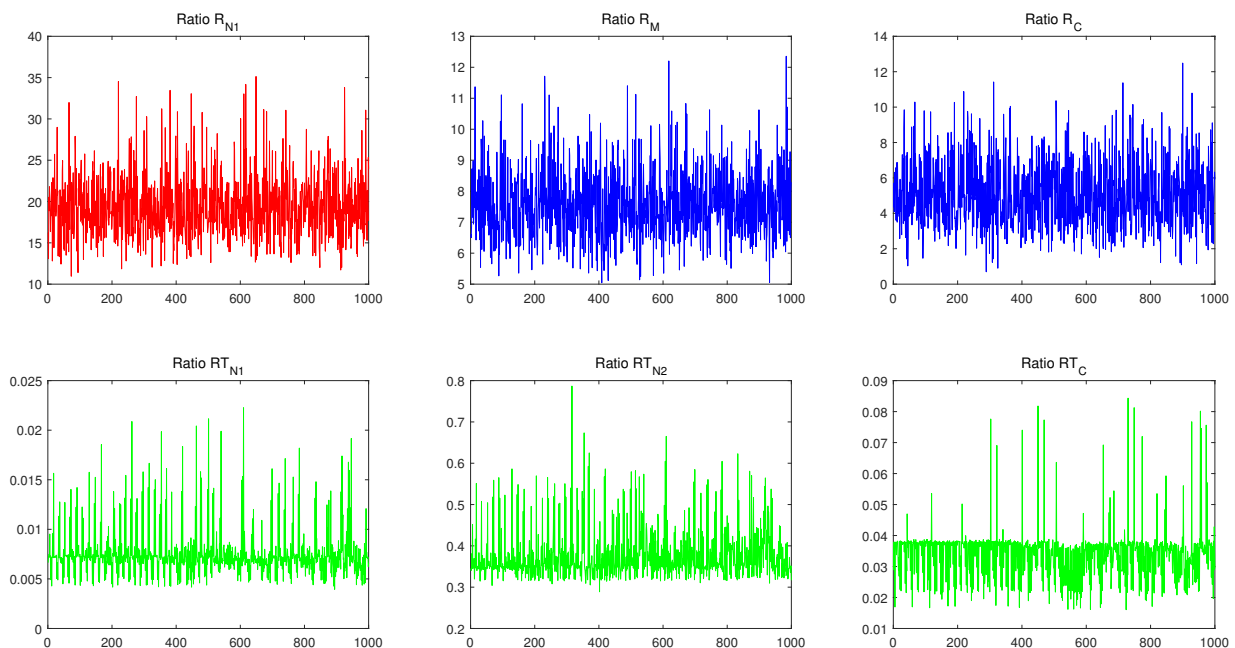
In our computation, we set  $L = I_n$ ,  $n = 100$ ,  $p = 60$ ,  $\|r\|_2 = 10^{-2}$ ,  $\lambda_{ij}$ s in the denominator of (4.2) are equal to 1,  $\lambda_{i1}$  in the numerator are set to accommodate the given condition number, and the rest  $\lambda_{ij}$ s are equally spaced. The normal Eq (1.5) is solved with the Matlab command `pcg` with relative residual smaller than  $10^{-10}$  and maximum iterations smaller than 100. By varying the condition numbers of coefficient matrices (4.2), we repeat 1000 times for each setting and present the numerical results in Figure 1. From the first row of Figure 1, we observe that the majority of the ratios are close to 1. This indicates that the derived upper bounds for the mixed and componentwise condition numbers of the GRR model (1.4) are quite tight. On the other hand, from the second row of Figure 1, we notice that most of the ratios exceed 20. This implies that these upper bounds can be computed efficiently, resulting in a significant improvement in CPU time by a factor of 20. Thus, in practical applications, we can use the upper bounds to measure the illness of the GRR model rather than the exact mixed and componentwise condition numbers.

**Example 2.** In this example, we will check the efficiency of the SSCE of condition numbers of the GRR model (1.4) with respect to accuracy and CPU time. Similar to the former example, we also define some ratios

$$\begin{aligned} R_{N1} &= \frac{\kappa_N}{\kappa_{L\mathcal{F}}^{2a}}, & R_M &= \frac{\kappa_M}{\kappa_{mL\mathcal{F}}^\infty}, & R_C &= \frac{\kappa_C}{\kappa_{cL\mathcal{F}}^\infty}, \\ RT_{N1} &= \frac{\kappa_N \text{CPU}}{\kappa_{L\mathcal{F}}^{2a} \text{CPU}}, & RT_{N2} &= \frac{\kappa_N \text{CPU}}{\kappa_{L\mathcal{F}1}^{2a} \text{CPU}}, & RT_C &= \frac{\kappa_C \text{CPU}}{\kappa_{cL\mathcal{F}}^\infty \text{CPU}}. \end{aligned}$$

The first row of the above ratios measures the accuracy of the SSCE method and the second row measures the efficiency. Specifically, although  $\kappa_{L\mathcal{F}}^{2a}$  in (2.9) and  $\kappa_{L\mathcal{F}1}^{2a}$  in (2.10) give the same values, they have very different storage requirement and computational efficiency. Thus, we only give  $R_{N1}$  to measure the accuracy of the normwise SSCE condition estimator, whereas  $RT_{N1}$  and  $RT_{N2}$  are employed to compare the computational efficiency. Moreover, note that the computational complexities of computing or estimating the mixed and componentwise condition numbers have very little difference, so we only present the CPU time comparison via  $RT_C$ . For a clarity, we also use figures to present our results.

In our experiment, we set  $L = I_n$ ,  $n = 200$ ,  $p = 100$ ,  $\|r\|_2 = 10^{-1}$ ,  $\kappa(\Sigma) = 10$ ,  $\kappa(\Theta) = 10$ ,  $\kappa(X) = 10$ , and the other settings are the same as that in Example 1. The numerical results are reported in Figure 2. From the first row in Figure 2, we can find that the SSCE method can give reliable estimates of the mixed and componentwise condition numbers for its ratios within a factor of 10. For the normwise condition number, the SSCE method may give slight overestimates, which coincides with the phenomenon in [19]. From the second row in Figure 2, we note that the ratios are much smaller than 1, which means, in general, the SSCE condition number estimates require much less CPU time compared with the explicit computation of the condition numbers. Comparing the first and second panels in the second row, we can also find that the compact form of the normwise condition number gives much gain of computational efficiency compared with its original form.



**Figure 2.** Efficiency and accuracy of the SSCE of condition numbers.

## 5. Conclusions

In this paper, we extended the GRR model given in [3] by allowing the serial dependence of the observations and investigated the condition number theory of the new model. We first established the generic expression of the condition number of the GRR model (1.4). By varying the norms and parameters in the generic expression, the popular normwise, mixed, and componentwise condition numbers can be obtained as its special cases. Considering the computational difficulty in calculating the exact value of the condition number, we provided the compact form of the normwise condition number and the upper bounds of the mixed and componentwise condition numbers. We also proposed the SSCE condition number estimators for the normwise, mixed, and componentwise condition numbers. Numerical experiments were given to show the tightness and efficiency of the upper bounds and the proposed condition estimators.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

The work was supported by Shandong Provincial Natural Science Foundation (Grant No. ZR2020QA034). WSX was also supported by Shandong Province Higher Education Youth Innovation and Technology Support Program (Grant No. 2023KJ199).

The authors would like to give their sincere thanks to the anonymous five reviewers for their detailed and helpful comments, which led to a much better presentation of their work.

### Conflict of interest

The authors affirm that they have no conflicts of interest to disclose.

### References

1. A. Hoerl, R. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
2. P. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Philadelphia: Society for Industrial and Applied Mathematics, 1998. <https://doi.org/10.1137/1.9780898719697>
3. W. van Wieringen, Lecture notes on ridge regression, arXiv: 1509.09169.
4. G. Golub, C. van Loan, *Matrix computations*, 4 Eds., Baltimore: Johns Hopkins University Press, 2013.
5. J. Rice, A theory of condition, *SIAM J. Numer. Anal.*, **3** (1966), 287–310. <https://doi.org/10.1137/0703023>

6. S. Wang, L. Meng, A contribution to the conditioning theory of the indefinite least squares problems, *Appl. Numer. Math.*, **177** (2022), 137–159. <https://doi.org/10.1016/j.apnum.2022.03.012>
7. S. Wang, H. Yang, Conditioning theory of the equality constrained quadratic programming and its applications, *Linear Multilinear A.*, **69** (2021), 1161–1183. <https://doi.org/10.1080/03081087.2019.1623858>
8. Z. Xie, W. Li, X. Jin, On condition numbers for the canonical generalized polar decomposition of real matrices, *Electron. J. Linear Al.*, **26** (2013), 842–857. <https://doi.org/10.13001/1081-3810.1691>
9. I. Gohberg, I. Koltracht, Mixed, componentwise, and structured condition numbers, *SIAM J. Matrix Anal. Appl.*, **14** (1993), 688–704. <https://doi.org/10.1137/0614049>
10. F. Cucker, H. Diao, Y. Wei, On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems, *Math. Comp.*, **76** (2007), 947–963. <https://doi.org/10.1090/S0025-5718-06-01913-2>
11. Y. Wei, D. Wang, Condition numbers and perturbation of the weighted Moore-Penrose inverse and weighted linear least squares problem, *Appl. Math. Comput.*, **145** (2003), 45–58. [https://doi.org/10.1016/S0096-3003\(02\)00437-X](https://doi.org/10.1016/S0096-3003(02)00437-X)
12. D. Chu, L. Lin, R. Tan, Y. Wei, Condition numbers and perturbation analysis for the Tikhonov regularization of discrete ill-posed problems, *Numer. Linear Algebra*, **18** (2011), 87–103. <https://doi.org/10.1002/nla.702>
13. H. Diao, Y. Wei, S. Qiao, Structured condition numbers of structured Tikhonov regularization problem and their estimations, *J. Comput. Appl. Math.*, **308** (2016), 276–300. <https://doi.org/10.1016/j.cam.2016.05.023>
14. L. Meng, B. Zheng, Structured condition numbers for the Tikhonov regularization of discrete ill-posed problems, *J. Comput. Math.*, **35** (2017), 169–186. <https://doi.org/10.4208/jcm.1608-m2015-0279>
15. N. Higham, *Accuracy and stability of numerical algorithms*, 2Eds., Philadelphia: Society for Industrial and Applied Mathematics, 2002. <https://doi.org/10.1137/1.9780898718027>
16. C. Kenney, A. Laub, Small-sample statistical condition estimates for general matrix functions, *SIAM J. Sci. Comput.*, **15** (1994), 36–61. <https://doi.org/10.1137/0915003>
17. A. Laub, J. Xia, Applications of statistical condition estimation to the solution of linear systems, *Numer. Linear Algebra*, **15** (2008), 489–513. <https://doi.org/10.1002/nla.570>
18. C. Kenney, A. Laub, M. Reese, Statistical condition estimation for linear least squares, *SIAM J. Matrix Anal. Appl.*, **19** (1998), 906–923. <https://doi.org/10.1137/S0895479895291935>
19. M. Baboulin, S. Gratton, R. Lacroix, A. Laub, Statistical estimates for the conditioning of linear least squares problems, In: *Parallel processing and applied mathematics*, Berlin: Springer, 2014, 124–133. [https://doi.org/10.1007/978-3-642-55224\\_13](https://doi.org/10.1007/978-3-642-55224_13)
20. A. Farooq, M. Samar, Sensitivity analysis for the generalized Cholesky block downdating problem, *Linear Multilinear A.*, **70** (2022), 997–1022. <https://doi.org/10.1080/03081087.2020.1751033>

21. A. Farooq, M. Samar, H. Li, C. Mu, Sensitivity analysis for the block Cholesky downdating problem, *Int. J. Comput. Math.*, **97** (2020), 1234–1253. <https://doi.org/10.1080/00207160.2019.1613528>
22. A. Laub, J. Xia, Fast condition estimation for a class of structured eigenvalue problems, *SIAM J. Matrix Anal. Appl.*, **30** (2009), 1658–1676. <https://doi.org/10.1137/070707713>
23. H. Diao, H. Xiang, Y. Wei, Mixed, componentwise condition numbers and small sample statistical condition estimation of Sylvester equations, *Numer. Linear Algebra*, **19** (2012), 639–654. <https://doi.org/10.1002/nla.790>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)