Linear Algebra and its Applications 624 (2021) 87–102

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Multiplicative perturbation bounds for multivariate multiple linear regression in Schatten p-norms

Jocelyn T. Chi [a,*], Ilse C.F. Ipsen [b]

[a] *Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*
[b] *Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA*

A R T I C L E I N F O

Article history:
Received 12 July 2020
Accepted 30 March 2021
Available online 6 April 2021
Submitted by N.J. Higham

MSC:
15-02

Keywords:
Projector
Multiplicative perturbations
Moore Penrose inverse
Schatten p-norms
Multivariate multiple linear regression

A B S T R A C T

Multivariate multiple linear regression (MMLR), which occurs in a number of practical applications, generalizes traditional least squares (multivariate linear regression) to multiple right-hand sides. We extend recent MLR analyses to sketched MMLR in general Schatten p-norms by interpreting the sketched problem as a multiplicative perturbation. Our work represents an extension of Maher's results on Schatten p-norms. We derive expressions for the exact and perturbed solutions in terms of projectors for easy geometric interpretation. We also present a geometric interpretation of the action of the sketching matrix in terms of relevant subspaces. We show that a key term in assessing the accuracy of the sketched MMLR solution can be viewed as a tangent of a largest principal angle between subspaces under some assumptions. Our results enable additional interpretation of the difference between an orthogonal and oblique projector with the same range.

* Corresponding author.
E-mail addresses: jtchi@ncsu.edu (J.T. Chi), ipsen@ncsu.edu (I.C.F. Ipsen).

https://doi.org/10.1016/j.laa.2021.03.039

1. Introduction

Multivariate multiple linear regression (MMLR)[1] is a natural generalization of traditional least squares regression (multivariate linear regression) to multiple right-hand sides. It is also useful in many large-scale real-world applications including image classification [28,58], quality control monitoring [15,38], genetic association studies [4,27], spatial genetic variation studies [52], climate studies [22], and low-rank tensor factorizations [25] to name a few. In the mathematics literature, least squares problems with multiple right-hand sides occur in the total least squares context, where both the independent and dependent variables may contain errors [18,19,45].

In recent years, randomized approaches have become a popular method of dealing with very large data problems in numerical linear algebra [35,57]. The idea is to utilize random projections, random sampling, or some combination of the two to reduce the problem to a lower dimension while approximately retaining the characteristics of the original problem. Referred to as *sketching*, this has become a popular approach for the fast solution of highly overdetermined or underdetermined regression problems [2,9,12,29,30,36,39,41], where either the number of rows far exceeds the number of columns, or vice versa.

We view row-sketched MMLR as a multiplicative perturbation of MMLR, and derive perturbation bounds that are amenable to geometric interpretation. Following up on our recent work [9], which quantifies the effect of sketching on the geometry of traditional least squares, we extend our analysis to sketched MMLR in general Schatten p-norms, which appear in numerous machine learning problems. In particular, the nuclear ($p = 1$) and Frobenius ($p = 2$) norms appear in penalized regression [55,58], regularized matrix regression [59], matrix completion [6,7], trace approximation [16,48], image feature extraction [14], and image processing and classification [26,53,54].

1.1. Problem setting

We begin with the exact MMLR problem in a Schatten p-norm. Denote the singular values of a matrix $\mathbf{M} \in \mathbb{R}^{m\times d}$ by

$$\sigma_1(\mathbf{M}) \geq \sigma_2(\mathbf{M}) \geq \cdots \geq \sigma_{\min(m,d)}(\mathbf{M}) \geq 0.$$

The Schatten p-norm [23, page 199] of $\mathbf{M}$ is a function of its singular values

$$\|\mathbf{M}\|_{(p)} = \sqrt[p]{\sigma_1(\mathbf{M})^p + \cdots + \sigma_r(\mathbf{M})^p} \quad \text{for} \quad 1 \leq p \leq \infty.$$

Given a pair of matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times d}$ with $\operatorname{rank}(\mathbf{A}) = n$, the goal is to estimate the solution $\widehat{\mathbf{X}} \in \mathbb{R}^{n\times d}$ satisfying

[1] We abbreviate multivariate multiple linear regression as "MMLR" throughout this paper.

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \|\mathbf{AX}-\mathbf{B}\|_{(p)} \quad \text{for } 1\le p\le\infty. \tag{1}$$

Popular Schatten p-norms include the

- $p=1$ nuclear (trace) norm $\|\mathbf{M}\|_* = \sum_{j=1}^{\min(m,d)} \sigma_j(\mathbf{M}) = \|\mathbf{M}\|_{(1)}$,
- $p=2$ Frobenius norm $\|\mathbf{M}\|_{\mathrm{F}} = \sqrt{\sum_{j=1}^{\min(m,d)} \sigma_j(\mathbf{M})^2} = \|\mathbf{M}\|_{(2)}$, and
- $p=\infty$ Euclidean (operator) norm $\|\mathbf{M}\|_2 = \sigma_1(\mathbf{M}) = \|\mathbf{M}\|_{(\infty)}$.

Given a matrix $\mathbf{S}\in\mathbb{R}^{c\times m}$ with $n\le c\le m$, the perturbed MMLR problem in a Schatten p-norm via randomized row-sketching is

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \|\mathbf{S}(\mathbf{AX}-\mathbf{B})\|_{(p)} \quad \text{for } 1\le p\le\infty. \tag{2}$$

Row-sketching can be an effective approach to handling large data in the highly over-constrained case [11,12,30,39,51], where $m\gg n$.

1.2. Existing work

Widely considered to have originated in [41], randomized sketching has become a popular approach to solving large data problems in machine learning and numerical linear algebra [35,57]. In the regression setting, sketching approaches can be broadly classified [46, Section 1] according to whether they achieve row compression [3,11,12,21,29,30,40,51], column compression [2,46], or both [36]. Recent work has improved the theoretical understanding of randomized regression from a statistical [9,29,30,39,55] and geometric perspective [9].

The sketched MMLR problem in (2) can be viewed as a generalization of weighted least squares since $\mathbf{S}$ is not required to be positive definite diagonal [24,42,56]. Additionally, (2) holds more generally for Schatten p-norms with $1\le p\le\infty$ rather than only the Frobenius norm. Perturbation analysis for weighted least squares quantifies the effect of additive perturbations of the weights, $\mathbf{A}$, or both [56]. By contrast, we view the sketched problem in (2) as a multiplicative perturbation of (1).

1.3. Our contributions

Our results extend the following: 1) Maher's work [31–34] on Schatten p-norms; 2) the analysis in [9] to the sketched MMLR problem in a Schatten p-norm; and 3) the result in [12, Lemma 1] to the $d\ge 1$ case and for Schatten p-norms with $1\le p\le\infty$ under weaker assumptions. We also show that the accuracy of the sketched MMLR solution in a Schatten p-norm depends on a term that captures both 1) how close the sketching matrix $\mathbf{S}$ is to approximately preserving orthogonality [10,37,47] for any rank-preserving $\mathbf{S}$ and 2) how close the vectors in a basis for the sketched subspace are to being orthonormal

(Proposition 3). We present a geometric interpretation of the action of the sketching matrix $\mathbf{S}$ in terms of relevant subspaces. We show that a key term in assessing the accuracy of the sketched MMLR solution can be interpreted as the tangent of a largest principal angle between these subspaces if $\mathbf{S}$ has orthonormal rows (Proposition 4) or if $\mathbf{S}$ preserves rank (Proposition 5). We extend this interpretation to the operator norm difference between an orthogonal and oblique projector with the same range when $\mathbf{S}$ preserves rank (Proposition 6).

1.4. Preliminaries

We begin by setting some notation. Let $\mathbf{I}_n = (\mathbf{e}_1 \quad \mathbf{e}_2 \quad \dots \quad \mathbf{e}_n)$ denote the $n \times n$ identity matrix, and let the superscript T denote the transpose. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a matrix with $\operatorname{rank}(\mathbf{A}) = n$. Then $\mathbf{A}$ has the following full (first equality) and thin (second equality) QR decompositions

$$\mathbf{A} = (\mathbf{Q} \quad \mathbf{Q}_\perp) \begin{pmatrix} \mathbf{R} \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix} = \mathbf{QR}, \tag{3}$$

respectively, where $\mathbf{R} \in \mathbb{R}^{n\times n}$ is nonsingular. Thus, $\mathbf{Q} \in \mathbb{R}^{m\times n}$ and $\mathbf{Q}_\perp \in \mathbb{R}^{m\times(m-n)}$ represent orthonormal bases for $\operatorname{range}(\mathbf{A})$ and $\operatorname{range}(\mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T)$, respectively.

Since $\mathbf{A}$ has full column rank, its Moore-Penrose generalized inverse is

$$\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{R}^{-1}\mathbf{Q}^T.$$

The two-norm condition number of $\mathbf{A}$ with respect to left inversion is

$$\kappa_2(\mathbf{A}) = \|\mathbf{A}\|_2\, \|\mathbf{A}^\dagger\|_2.$$

The following lemma asserts strong multiplicativity for Schatten p-norms and invariance under multiplication by matrices with orthonormal columns (rows) on the left (right).

Lemma 1 *([34, (2.7)]). For $\mathbf{F} \in \mathbb{R}^{m\times n}, \mathbf{G} \in \mathbb{R}^{k\times m}$ and $\mathbf{C} \in \mathbb{R}^{n\times l}$ with $1 \le p \le \infty$, we have*

$$\|\mathbf{GFC}\|_{(p)} \le \|\mathbf{G}\|_2 \|\mathbf{F}\|_2 \|\mathbf{C}\|_{(p)}.$$

This version of Lemma 1 is obtained from a modification of the proof for [34, (2.5)].

2. Multivariate Multiple Linear Regression

We describe the solution and regression residual for the exact and perturbed MMLR problems in a Schatten p-norm in (1) and (2), respectively. The following states that the solutions for (1) are the same, regardless of the choice of $p \ge 1$ [34].

Proposition 1 *([31,32,34]). Let matrices* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{B} \in \mathbb{R}^{m\times d}$ *be given. The MMLR problem in a Schatten p-norm*

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \|\mathbf{AX}-\mathbf{B}\|_{(p)} \quad \textit{for} \quad 1 \le p \le \infty$$

has the minimal Schatten p-norm solution $\widehat{\mathbf{X}} \equiv \mathbf{A}^{\dagger}\mathbf{B}$ *with prediction and regression residual*

$$\widehat{\mathbf{B}} \equiv \mathbf{A}\widehat{\mathbf{X}} \quad \textit{and}$$
$$\widehat{\mathbf{\Gamma}} \equiv \mathbf{B} - \mathbf{A}\widehat{\mathbf{X}} = (\mathbf{I}-\mathbf{AA}^{\dagger})\mathbf{B},$$

respectively. If $\operatorname{rank}(\mathbf{A}) = n$*, then the solution* $\widehat{\mathbf{X}} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{B}$ *is unique with regression residual* $\widehat{\mathbf{\Gamma}} = (\mathbf{I}-\mathbf{QQ}^T)\mathbf{B} = \mathbf{Q}_{\perp}\mathbf{Q}_{\perp}^T\mathbf{B}$.

For a proof that $\widehat{\mathbf{X}}$ is the minimal Schatten p-norm solution to (1), see [31,32,34]. Specifically, [31] shows that $\|\mathbf{AX}-\mathbf{B}\|_{(p)} \ge \|\mathbf{AA}^{\dagger}\mathbf{B}-\mathbf{B}\|_{(p)}$ for $2 \le p < \infty$ and [32] extends the result to $1 \le p < \infty$. Then, [34] extends the inequality to $1 \le p \le \infty$ by showing that $\sigma_j(\mathbf{AX}-\mathbf{B}) \ge \sigma_j(\mathbf{AA}^{\dagger}\mathbf{B}-\mathbf{B})$ for $j = 1, 2, \ldots$ for finite rank operators. Finally, [34, Corollary 3.1] shows that $\widehat{\mathbf{X}}$ has minimal Schatten p-norm. If $\operatorname{rank}(\mathbf{A}) = n$, then $\operatorname{null}(\mathbf{A}) = \{\mathbf{0}\}$ so that the general solution in [34, Corollary 3.1] is also unique.

Let $\mathbf{S} \in \mathbb{R}^{c\times m}$ be a multiplicative perturbation matrix from the left with $n \le c \le m$ and $\operatorname{rank}(\mathbf{SA}) \le \operatorname{rank}(\mathbf{A}) = n$. For example, $\mathbf{S}$ may be a sampling matrix that extracts rows from $\mathbf{A}$ [12,30], a projection matrix [1,41], or a combination of sampling and projection matrices [2,12].

Corollary 1. *Let matrices* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{B} \in \mathbb{R}^{m\times d}$ *be given. The perturbed MMLR problem in a Schatten p-norm*

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \|\mathbf{S}(\mathbf{AX}-\mathbf{B})\|_{(p)} \quad \textit{for} \quad 1 \le p \le \infty$$

in (2) has the minimal Schatten p-norm solution $\widetilde{\mathbf{X}} = (\mathbf{SA})^{\dagger}\mathbf{SB}$*. If* $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A}) = n$*, then* $\widetilde{\mathbf{X}}$ *is unique.*

Following convention [30,39], we define the prediction and regression residual of the perturbed MMLR problem to be

$$\widetilde{\mathbf{B}} = \mathbf{A}\widetilde{\mathbf{X}} \quad \text{and} \quad \widetilde{\mathbf{\Gamma}} = \mathbf{B} - \mathbf{A}\widetilde{\mathbf{X}}.$$

3. General multiplicative perturbations

We present general multiplicative perturbation bounds for (2) requiring no assumptions on $\mathbf{S}$. To enable geometric interpretation, we express the bounds in terms of orthogonal and oblique projectors onto range($\mathbf{A}$) or a subspace of range($\mathbf{A}$). For a matrix $\mathbf{A}$,

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^{\dagger}$$

denotes the orthogonal projector onto range($\mathbf{A}$) along null($\mathbf{A}^T$) ([44, Theorem III.1.3] and [8,20,50]). For the perturbed MMLR problem in (2),

$$\mathbf{P} \equiv \mathbf{A}(\mathbf{S}\mathbf{A})^{\dagger}\mathbf{S}$$

denotes the corresponding oblique projector onto a subspace of range($\mathbf{A}$). If rank($\mathbf{S}\mathbf{A}$) = rank($\mathbf{A}$), then range($\mathbf{P}$) = range($\mathbf{P}_{\mathbf{A}}$) although null($\mathbf{P}$) = null($\mathbf{A}^T\mathbf{S}^T\mathbf{S}$) [49, Theorem 3.1], and null($\mathbf{A}^T\mathbf{S}^T\mathbf{S}$) $\neq$ null($\mathbf{P}_{\mathbf{A}}$) in general [9, Lemma 3.1]. Oblique projectors appear in [43,49] for constrained least squares, [17] for discrete inverse problems, and [5,42] for weighted least squares. The oblique projector $\mathbf{P}$ can be viewed as an extension of the oblique projector

$$\mathbf{P}_D = \mathbf{A}(\mathbf{A}^T\mathbf{D}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{D}$$

in [42] if $\mathbf{D} = \mathbf{S}^T\mathbf{S}$ is a diagonal matrix with positive elements on the diagonal and $(\mathbf{A}^T\mathbf{D}\mathbf{A})^{-1}$ exists. If $\mathbf{S}$ is a sketching matrix that samples without replacement and $c = m$, then $\mathbf{S}^T\mathbf{S} = \mathbf{I}_m$ satisfies the requirements for $\mathbf{D}$ in [42]. In this case, however, the sketched MMLR problem in (2) becomes the exact MMLR problem in (1). If $d = 1$ and $p = 2$ in (2), the oblique projector $\mathbf{P}$ appears in [39] if rank($\mathbf{S}\mathbf{A}$) = rank($\mathbf{A}$) and in [9, Lemma 3.1] for any sketching matrix $\mathbf{S}$. Oblique projectors also appear in other problems, such as the discrete empirical interpolation method (DEIM) oblique projector $\mathbb{D} = \mathbf{U}_r(\mathbf{S}^T\mathbf{U}_r)^{\dagger}\mathbf{S}^T$ in [13, Section 3.1].

Since $\mathbf{A}^{\dagger}$ is a left inverse of $\mathbf{A}$, the exact and perturbed solutions are $\widehat{\mathbf{X}} = \mathbf{A}^{\dagger}\mathbf{P}_{\mathbf{A}}\mathbf{B}$ and $\widetilde{\mathbf{X}} = \mathbf{A}^{\dagger}\mathbf{P}\mathbf{B}$, respectively [9, Lemma 3.1]. Therefore, the absolute error between the solution and regression residual is

$$\begin{aligned}\widetilde{\mathbf{X}} - \widehat{\mathbf{X}} = [(\mathbf{S}\mathbf{A})^{\dagger}\mathbf{S} - \mathbf{A}^{\dagger}]\mathbf{B} \;&=\; \mathbf{A}^{\dagger}(\mathbf{P} - \mathbf{P}_{\mathbf{A}})\mathbf{B} \quad \text{and}\\ \widetilde{\mathbf{\Gamma}} - \widehat{\mathbf{\Gamma}} = \mathbf{A}[\mathbf{A}^{\dagger} - (\mathbf{S}\mathbf{A})^{\dagger}\mathbf{S}]\mathbf{B} \;&=\; (\mathbf{P}_{\mathbf{A}} - \mathbf{P})\mathbf{B}.\end{aligned}$$

Proposition 2 bounds the absolute error of the perturbed solution and regression residual for the MMLR problem in a Schatten p-norm with $1 \leq p \leq \infty$ in terms of the above projection matrices.

Proposition 2. *For the perturbed MMLR problem in* (2)*, the absolute error bounds on the solution and regression residual in a Schatten p-norm are*

$$\begin{aligned}\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)} &\leq \|\mathbf{A}^{\dagger}\|_2 \,\|\mathbf{P} - \mathbf{P}_{\mathbf{A}}\|_2 \,\|\mathbf{B}\|_{(p)} \quad \textit{and}\\ \|\widetilde{\mathbf{\Gamma}} - \widehat{\mathbf{\Gamma}}\|_{(p)} &\leq \|\mathbf{P} - \mathbf{P}_{\mathbf{A}}\|_2 \,\|\mathbf{B}\|_{(p)}.\end{aligned}$$

If $\mathbf{A}^T\mathbf{B} \neq \mathbf{0}$*, the relative error bound in a Schatten* p*-norm is*

$$\frac{\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)}}{\|\widehat{\mathbf{X}}\|_{(p)}} \leq \kappa_2(\mathbf{A})\,\|\mathbf{P} - \mathbf{P}_{\mathbf{A}}\|_2\,\frac{\|\mathbf{B}\|_{(p)}}{\|\mathbf{A}\|_2\|\widehat{\mathbf{X}}\|_{(p)}}.$$

Proof. Lemma 1 implies the bounds for the absolute error in a Schatten p-norm. □

Proposition 2, which extends [9, Corollary 3.5] to multiple right-hand sides and Schatten p-norms with $1 \leq p \leq \infty$, shows that the accuracy of the sketched solution and regression residual depends on the operator norm projector difference $\|\mathbf{P} - \mathbf{P}_{\mathbf{A}}\|_2$.

4. Multiplicative perturbations that preserve rank

We present multiplicative perturbation bounds for (2) that hold if $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$. We begin by rewriting the difference between $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}$ in terms of an orthonormal basis for the column space of $\mathbf{A}$. Since $\operatorname{rank}(\mathbf{SA}) = n$, $(\mathbf{SA})^\dagger = \mathbf{R}^{-1}(\mathbf{SQ})^\dagger$ so that

$$\mathbf{P}_{\mathbf{A}} - \mathbf{P} = \mathbf{QQ}^T - \mathbf{Q}(\mathbf{SQ})^\dagger\mathbf{S}.$$

Although the results in this section require the additional assumption that $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$, they enable geometric interpretation beyond the difference between the projectors $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{P}$.

Proposition 3. *For the perturbed MMLR problem in* (2)*, if* $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$*, the absolute error bound in a Schatten* p*-norm for* $1 \leq p \leq \infty$ *is*

$$\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)} \leq \|\mathbf{A}^\dagger\|_2\,\|(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\|_2\,\|\widehat{\mathbf{\Gamma}}\|_{(p)}.$$

Proof. Since $\operatorname{rank}(\mathbf{SA}) = n$, we have $(\mathbf{SA})^\dagger = \mathbf{R}^{-1}(\mathbf{SQ})^\dagger$. Thus,

$$\begin{aligned}\widetilde{\mathbf{X}} - \widehat{\mathbf{X}} &= (\mathbf{SA})^\dagger\mathbf{SB} - \mathbf{A}^\dagger\mathbf{B}\\ &= \mathbf{R}^{-1}[(\mathbf{SQ})^\dagger\mathbf{S} - \mathbf{Q}^T]\mathbf{B}.\end{aligned} \tag{4}$$

Multiplying $\mathbf{B}$ on the left by the identity matrix $\mathbf{I} = \mathbf{QQ}^T + \mathbf{Q}_\perp\mathbf{Q}_\perp^T$ and inserting it in (4) gives

$$\begin{aligned}\widetilde{\mathbf{X}} - \widehat{\mathbf{X}} &= \mathbf{R}^{-1}[(\mathbf{SQ})^\dagger\mathbf{S} - \mathbf{Q}^T](\mathbf{QQ}^T + \mathbf{Q}_\perp\mathbf{Q}_\perp^T)\mathbf{B}\\ &= \mathbf{R}^{-1}(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\mathbf{Q}_\perp^T\mathbf{B}.\end{aligned} \tag{5}$$

Lemma 1 implies the following upper bound on the Schatten p-norm of the absolute error difference between the sketched and exact MMLR solutions

$$\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)} \leq \|\mathbf{R}^{-1}\|_2 \, \|(\mathbf{SQ})^\dagger \mathbf{S}\mathbf{Q}_\perp\|_2 \, \|\mathbf{Q}_\perp \mathbf{Q}_\perp^T \mathbf{B}\|_{(p)}.$$

Finally, applying the definition of the exact regression residual $\widehat{\mathbf{\Gamma}} = \mathbf{Q}_\perp \mathbf{Q}_\perp^T \mathbf{B}$ concludes the proof. □

Since $\|\mathbf{A}^\dagger\|_2$ and $\|\widehat{\mathbf{\Gamma}}\|_{(p)}$ are fixed for any pair of $\mathbf{A}$ and $\mathbf{B}$, only $\|(\mathbf{SQ})^\dagger \mathbf{S}\mathbf{Q}_\perp\|_2$ is affected by the choice of the sketching matrix $\mathbf{S}$. We compare this to the approximate isometry term $\|(\mathbf{SQ})^T \mathbf{S}\widehat{\mathbf{\Gamma}}\|_2$ from [12, Equation 9], where $(\mathbf{SQ})^T \mathbf{S}\widehat{\mathbf{\Gamma}}$ is a vector. Notice that we can arrive at the $\|(\mathbf{SQ})^T \mathbf{S}\widehat{\mathbf{\Gamma}}\|_2$ term if we revert to (5) in the above proof and assume that the columns of $\mathbf{SQ}$ are orthonormal so that $(\mathbf{SQ})^\dagger = (\mathbf{SQ})^T$. If we further restrict our analysis to the $d = 1$ and $p = 2$ case, we recover the same normed quantity as in [12, Equation 9]. Thus, we compare Proposition 3 to [12, Lemma 1], where the absolute solution error for the $d = 1$ and $p = 2$ case is

$$\|\widehat{\mathbf{X}} - \widetilde{\mathbf{X}}\|_2 \leq \|\mathbf{A}^\dagger\|_2 \sqrt{\epsilon} \|\widehat{\mathbf{\Gamma}}\|_2 \tag{6}$$

for ϵ and $\mathbf{S}$ satisfying [12, Equations 8 and 9]:

$$\|(\mathbf{SQ})^\dagger\|_2 \leq 2^{\frac{1}{4}} \quad \text{and} \tag{7}$$

$$\|(\mathbf{SQ})^\dagger \mathbf{S}\widehat{\mathbf{\Gamma}}\|_2 \leq \sqrt{\frac{\epsilon}{2}} \|\widehat{\mathbf{\Gamma}}\|_2. \tag{8}$$

Proposition 3 can be viewed as an extension of [12, Lemma 1] in the following ways. First, Proposition 3 extends the result in [12, Lemma 1] for $d \geq 1$ and for Schatten p-norms with $1 \leq p \leq \infty$. Second, [12, Lemma 1] is a special case of Proposition 3 when $d = 1$, $p = 2$, and $\sqrt{\epsilon} = \|(\mathbf{SQ})^\dagger \mathbf{S}\mathbf{Q}_\perp\|_2$. Third, in contrast with [12, Lemma 1], the bound in Proposition 3 holds without requiring the assumptions (7) or (8).

5. Angle between the original and perturbed subspaces

We show that $\|(\mathbf{SQ})^\dagger \mathbf{S}\mathbf{Q}_\perp\|_2$ is the tangent of a largest principal angle under two conditions: if $\mathbf{S}$ has orthonormal rows, or if $\mathbf{S}$ preserves rank. Furthermore we show that if $\mathbf{S}$ preserves rank, then $\|(\mathbf{SQ})^\dagger \mathbf{S}\mathbf{Q}_\perp\|_2$ equals the operator norm difference between the orthogonal projector $\mathbf{P}_\mathbf{A}$ and the oblique projector $\mathbf{P}$. Therefore, if an orthogonal and an oblique projector have the same range, then their operator norm difference can be interpreted in terms of principal angles. We begin with a decomposition of range$(\mathbf{S}^T)$ with respect to range$(\mathbf{Q})$ and range$(\mathbf{Q}_\perp)$.

5.1. *A decomposition of* range$(\mathbf{S}^T)$

The following geometric interpretations depend on a decomposition of $\mathbf{S}$ into three subspaces. Let $\mathcal{Q} \equiv$ range$(\mathbf{Q})$, $\mathcal{Q}^\perp \equiv$ range$(\mathbf{Q}_\perp)$, and $\mathcal{S} \equiv$ range$(\mathbf{S}^T)$. Following the notation in [60, Section 2], we can decompose $\mathcal{S}$ into the direct sum of the following subspaces

$$\mathcal{S}_1 \equiv \mathcal{S} \cap \mathcal{Q}, \quad \mathcal{S}_0 \equiv \mathcal{S} \cap \mathcal{Q}^\perp, \quad \text{and} \quad \mathcal{S}_{10} \equiv \mathcal{S} \cap (\mathcal{Q} \oplus \mathcal{Q}^\perp)^\perp.$$

We summarize and interpret these subspaces of $\mathcal{S}$ as follows. The subspace $\mathcal{S}_1$ contains the directions in $\mathcal{S}$ that are also in $\mathcal{Q}$. Specifically, $\mathcal{S}_1 = \{\mathbf{s} \in \mathcal{S} : \mathbf{s}^T\mathbf{q} = \|\mathbf{s}\|_2\|\mathbf{q}\|_2$ for some $\mathbf{q} \in \mathcal{Q}\}$, where $\|\cdot\|_2$ denotes the Euclidean vector norm.

The subspace $\mathcal{S}_0$ contains the directions in $\mathcal{S}$ that are also in $\mathcal{Q}^\perp$. Therefore, these are the directions in $\mathcal{S}$ that are orthogonal to directions in $\mathcal{Q}$. Specifically, $\mathcal{S}_0 = \{\mathbf{s} \in \mathcal{S} : \mathbf{s}^T\mathbf{q} = 0$ for all $\mathbf{q} \in \mathcal{Q}\}$.

The subspace $\mathcal{S}_{10}$ contains the directions in $\mathcal{S}$ that are in neither $\mathcal{Q}$ nor $\mathcal{Q}^\perp$. Therefore, these are the directions in $\mathcal{S}$ that are not orthogonal to $\mathcal{Q}$ but are also not in $\mathcal{Q}$. Specifically, $\mathcal{S}_{10} = \{\mathbf{s} \in \mathcal{S} : 0 < |\mathbf{s}^T\mathbf{q}| < \|\mathbf{s}\|_2\|\mathbf{q}\|_2$ for all $\mathbf{q} \in \mathcal{Q}\}$.

The subspace

$$\mathcal{S}_Q \equiv \mathcal{S}_1 \oplus \mathcal{S}_{10},$$

then comprises the directions in $\mathcal{S}$ that are not orthogonal with directions in $\mathcal{Q}$. Specifically, $\mathcal{S}_Q = \{\mathbf{s} \in \mathcal{S} : 0 < |\mathbf{s}^T\mathbf{q}| \leq \|\mathbf{s}\|_2\|\mathbf{q}\|_2$ for all $\mathbf{q} \in \mathcal{Q}\}$.

Section 5.3.1 presents an illustrative example of these subspaces in the context of Proposition 5. In general, we have

$$\dim(\mathcal{S}_1) \leq \dim(\mathcal{Q}) = n$$

and

$$\dim(\mathcal{S}_1) \leq \dim(\mathcal{S}_Q) \leq \dim(\mathcal{S}) \leq c.$$

If $\operatorname{rank}(\mathbf{SA}) = n$, then we additionally have

$$\dim(\mathcal{S}_1) \leq n \leq \dim(\mathcal{S}_Q) \leq \dim(\mathcal{S}) \leq c.$$

5.2. Interpretation of $\|(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\|_2$ if $\mathbf{S}$ has orthonormal rows

If $\mathbf{S}$ has orthonormal rows, the quantity $\|(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\|_2$ has geometric interpretation even with no additional requirements on $\mathbf{S}$ or $\operatorname{rank}(\mathbf{SA})$. One example is sketching via random sampling without replacement where one row is selected in each sample. The following relies on a key result on the angles between subspaces from [60, Theorem 3.1].

Proposition 4. *For the perturbed MMLR problem in (2) with the subspaces defined in Section 5.1, if $\mathbf{S}$ has orthonormal rows, then*

$$\|(\mathbf{SQ})^{\dagger}(\mathbf{SQ}_{\perp})\|_2 = \tan\theta_1(\mathcal{S}, \mathcal{Q}),$$

where $\theta_1(\mathcal{S}, \mathcal{Q})$ denotes a largest principal angle between $\mathcal{S}$ and $\mathcal{Q}$. The absolute error bound in a Schatten p-norm is

$$\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)} \leq \tan\theta_1(\mathcal{S}, \mathcal{Q})\, \|\mathbf{A}^{\dagger}\|_2\, \|\widehat{\mathbf{\Gamma}}\|_{(p)}.$$

This result follows from [60, Theorem 3.1] using the orthogonal matrix $\begin{pmatrix} \mathbf{Q} & \mathbf{Q}_{\perp} \end{pmatrix}$ and $\mathbf{S}^T$ with $\mathbf{S}$ having orthonormal rows. Thus, the positive singular values of $(\mathbf{SQ})^{\dagger}\mathbf{SQ}_{\perp}$ are the tangents of the principal angles between $\mathcal{S}$ and $\mathcal{Q}$. Therefore, the absolute error in a Schatten p-norm between the sketched and exact MMLR solutions depends on the tangent of a largest principal angle between $\mathcal{S}$ and $\mathcal{Q}$. Notice that without additional assumptions on $\operatorname{rank}(\mathbf{SA})$, the tangent of a principal angle between $\mathcal{S}$ and $\mathcal{Q}$ may be ∞.

5.3. Interpretation of $\|(\mathbf{SQ})^{\dagger}\mathbf{SQ}_{\perp}\|_2$ *if* $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$

If the sketching matrix $\mathbf{S}$ preserves rank so that $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$, the quantity $\|(\mathbf{SQ})^{\dagger}\mathbf{SQ}_{\perp}\|_2$ has geometric interpretation without requiring additional assumptions on $\mathbf{S}$. This interpretation is based on [60, Theorem 3.1 and Remark 3.1].

Proposition 5. *For the perturbed MMLR problem in (2) with the subspaces defined in Section 5.1, if* $\operatorname{rank}(\mathbf{SA}) = \operatorname{rank}(\mathbf{A})$ *and* $\mathbf{Z} \equiv (\mathbf{SQ})^{\dagger}\mathbf{S}$, *then the singular values of* $(\mathbf{SQ})^{\dagger}\mathbf{SQ}_{\perp}$ *represent the tangents of the principal angles between* $\mathcal{Z} \equiv \operatorname{range}(\mathbf{Z}^T)$ *and $\mathcal{Q}$. Therefore,*

$$\|(\mathbf{SQ})^{\dagger}(\mathbf{SQ}_{\perp})\|_2 = \tan\theta_1(\mathcal{Z}, \mathcal{Q}),$$

where $\theta_1(\mathcal{Z}, \mathcal{Q})$ denotes a largest principal angle between $\mathcal{Z}$ and $\mathcal{Q}$. Moreover, $\tan\theta_1(\mathcal{Z}, \mathcal{Q})$ *is strictly less than ∞ and the absolute error bound in a Schatten p-norm is*

$$\|\widetilde{\mathbf{X}} - \widehat{\mathbf{X}}\|_{(p)} \leq \tan\theta_1(\mathcal{Z}, \mathcal{Q})\, \|\mathbf{A}^{\dagger}\|_2\, \|\widehat{\mathbf{\Gamma}}\|_{(p)}.$$

Proof. The proof is adapted from [60, Remark 3.1]. The proof strategy is to construct an orthonormal basis for a subspace of $\mathcal{S}_Q$ and then to apply [60, Theorem 3.1] with the orthonormal basis and $\mathbf{Q}$.

We begin with a basis transformation of $\mathbf{S}$ by constructing the orthogonal matrix

$$\mathbf{Q}_B \equiv \begin{pmatrix} \mathbf{Q} & \mathbf{Q}_{\perp} \end{pmatrix} \in \mathbb{R}^{m \times m}.$$

Rewriting $\mathbf{S}$ in terms of $\mathbf{Q}_B$ gives

$$\mathbf{S} = \mathbf{SQ}_B\mathbf{Q}_B^T = \begin{pmatrix} \mathbf{SQ} & \mathbf{SQ}_{\perp} \end{pmatrix} \mathbf{Q}_B^T.$$

Since $\operatorname{rank}(\mathbf{SQ}) = n$, $(\mathbf{SQ})^{\dagger}$ is a left inverse of $\mathbf{SQ}$ and so applying it to $\mathbf{S}$ on the left gives

$$\mathbf{Z} = (\mathbf{SQ})^\dagger \mathbf{S} = \begin{pmatrix} \mathbf{I}_n & (\mathbf{SQ})^\dagger \mathbf{SQ}_\perp \end{pmatrix} \mathbf{Q}_B^T \in \mathbb{R}^{n\times m}.$$

Let $\mathbf{T} \equiv (\mathbf{SQ})^\dagger \mathbf{SQ}_\perp \in \mathbb{R}^{n\times(m-n)}$. We will show that the singular values of $\mathbf{T}$ represent the tangents of the principal angles between $\mathcal{Z}$ and $\mathcal{Q}$.

Notice that the Gram matrix

$$\mathbf{ZZ}^T = \mathbf{I}_n + \mathbf{TT}^T \in \mathbb{R}^{n\times n}$$

is symmetric positive definite. Therefore, its inverse has the unique symmetric positive definite square root $(\mathbf{ZZ}^T)^{-\frac{1}{2}} = (\mathbf{I}_n + \mathbf{TT}^T)^{-\frac{1}{2}}$. Now define

$$\mathbf{Z}_0 \equiv (\mathbf{ZZ}^T)^{-\frac{1}{2}} \mathbf{Z} \in \mathbb{R}^{n\times m}.$$

Then $\mathbf{Z}_0$ has orthonormal rows and the columns of $\mathbf{Z}_0^T$ represent a basis for range$(\mathbf{Z}^T)$. Since rank$(\mathbf{SQ}) = n$, range$(\mathbf{Z}^T) =$ range$(\mathbf{S}^T\mathbf{SQ}) \subseteq$ range$(\mathbf{S}^T) = \mathcal{S}$.

Applying [60, Theorem 3.1] with $\mathbf{Z}_0^T$ and $\mathbf{Q}$ shows that the singular values of $(\mathbf{SQ})^\dagger \mathbf{SQ}_\perp$ are the tangents of the principal angles between $\mathcal{Z} =$ range$(\mathbf{Z}^T)$ and $\mathcal{Q}$. Since $(\mathbf{Z}_0^T)^T\mathbf{Q} = \mathbf{Z}_0\mathbf{Q} = (\mathbf{ZZ})^{-\frac{1}{2}} = (\mathbf{I}_n + \mathbf{TT}^T)^{-\frac{1}{2}}$ is nonsingular, $\mathcal{Z} \subseteq \mathcal{S}_Q$ and the tangents of the principal angles between $\mathcal{Z}$ and $\mathcal{Q}$ are strictly less than ∞. $\square$

Clearly, $\mathcal{Z} \subseteq \mathcal{S}_Q$. One might ask the question: Is $\mathcal{Z} = \mathcal{S}_Q$? Notice that rank$(\mathbf{SA}) = n$ and rank$(\mathbf{S}^T) \leq c$ imply that

$$n \leq \dim(\mathcal{S}_Q) \leq c \quad \text{and} \quad \dim(\mathcal{S}_1) \leq c - n.$$

Although $\mathcal{Z} \neq \mathcal{S}_Q$ in general, if $\dim(\mathcal{S}_Q) = n$, then $\dim(\mathcal{Z}) = n$ implies that $\mathcal{Z} = \mathcal{S}_Q$. Meanwhile, if $\dim(\mathcal{S}_Q) > n$, then $n = \dim(\mathcal{Z}) < \dim(\mathcal{S}_Q)$ so that $\mathcal{Z} \neq \mathcal{S}_Q$. The example in Section 5.3.1 illustrates this concretely.

Propositions 4 and 5 show that if rank$(\mathbf{SA}) =$ rank$(\mathbf{A})$, $\|(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\|_2$ has geometric interpretation as the tangent of a largest principal angle between a subspace of $\mathcal{S}_Q$ and $\mathcal{Q}$. Moreover, the tangents of the principal angles between these two subspaces are bounded. If rank$(\mathbf{SA}) <$ rank$(\mathbf{A})$, then $\|(\mathbf{SQ})^\dagger\mathbf{SQ}_\perp\|_2$ still has geometric interpretation as the tangent of a largest principal angle between $\mathcal{S}$ and $\mathcal{Q}$ if $\mathbf{S}$ has orthonormal rows. Proposition 5 implies that if rank$(\mathbf{SA}) =$ rank$(\mathbf{A})$, then the operator norm difference between $\mathbf{P}$ and $\mathbf{P}_\mathbf{A}$ has the following geometric interpretation.

Proposition 6. *For the perturbed MMLR problem in* (2) *with the subspaces defined in Section 5.1, if* rank$(\mathbf{SA}) =$ rank$(\mathbf{A})$,

$$\|\mathbf{P} - \mathbf{P}_\mathbf{A}\|_2 = \tan\theta_1(\mathcal{Z}, \mathcal{Q}),$$

where $\mathcal{Z}$ is a subspace of $\mathcal{S}_Q$ and $\theta_1(\mathcal{Z}, \mathcal{Q})$ denotes a largest principal angle between $\mathcal{Z}$ and $\mathcal{Q}$. Moreover, $\tan\theta_1(\mathcal{Z}, \mathcal{Q})$ *is strictly less than* ∞.

Proof. We decompose $\mathbf{I}_m$ into the sum of orthogonal projectors and rewrite the operator norm difference between $\mathbf{P_A}$ and $\mathbf{P}$ as the following

$$\mathbf{P_A} - \mathbf{P} = \mathbf{QQ}^T - \mathbf{Q}(\mathbf{SQ})^\dagger \mathbf{S}(\mathbf{QQ}^T + \mathbf{Q}_\perp \mathbf{Q}_\perp^T).$$

After we expand and cancel terms, the result follows from unitary invariance of spectral norms and Proposition 5. $\square$

This result is implied from the absolute error bound in Proposition 5. However, the direct statement of this result ties the interpretation of $\|(\mathbf{SQ})^\dagger \mathbf{SQ}_\perp\|_2$ as the tangent of a largest principal angle between a subspace of $\mathcal{S}_Q$ and $\mathcal{Q}$ to the operator norm difference between $\mathbf{P}$ and $\mathbf{P_A}$. In this way, we have additional geometric interpretation of the difference between an orthogonal and oblique projector with the same range if $\mathbf{S}$ preserves rank.

5.3.1. Illustrative example of the subspaces in Proposition 5

We provide an example illustrating the subspaces of Section 5.1 in the context of Proposition 5. Let

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_\perp = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{S}^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Then $\mathcal{S}$ has the following subspaces

$$\mathcal{S}_1 = \text{range} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathcal{S}_{10} = \text{range} \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad \mathcal{S}_Q = \text{range} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This example illustrates how $\mathcal{S}_1$ contains directions in $\mathcal{S}$ that are in $\mathcal{Q}$, and $\mathcal{S}_{10}$ contains directions in $\mathcal{S}$ that cannot be represented solely by directions in $\mathcal{Q}$ or directions in $\mathcal{Q}^\perp$. This is because vectors in $\mathcal{S}_{10}$ are obtained from a non-trivial linear combination of vectors in $\mathcal{Q}$ with vectors in $\mathcal{Q}^\perp$. Thus, for any $\mathbf{v} \in \mathcal{S}_{10}$ and any $\mathbf{q} \in \mathcal{Q}$, we have $\mathbf{v}^T\mathbf{q} \neq 0$. However, $\mathbf{v} \notin \mathcal{Q}$ and $\mathbf{v} \notin \mathcal{Q}^\perp$.

Notice that in this example, there are no non-zero directions in $\mathcal{S}$ that are also in $\mathcal{Q}^\perp$. Since $\operatorname{rank}(\mathbf{SA}) = n$ and $\operatorname{rank}(\mathbf{S}^T) \le c$ require that $\dim(\mathcal{S}_Q) \ge n$ and $\dim(\mathcal{S}_1) \le c - n$, $\mathcal{S}_0 = \{\mathbf{0}\}$ is an artifact of this example.

Proceeding with the example, we have

$$\mathbf{SQ} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{Z} = (\mathbf{SQ})^{\dagger}\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix},$$

where $\mathbf{SQ}$ has full column rank. This gives us

$$\mathbf{ZZ}^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{4} & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_0 = (\mathbf{ZZ}^T)^{-\frac{1}{2}}\mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2\sqrt{5}}{5} & 0 & 0 & \frac{\sqrt{5}}{5} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

Thus, $\mathbf{Z}_0^T$ has orthonormal columns and

$$\mathbf{Z}_0\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2\sqrt{5}}{5} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}$$

is nonsingular so that $\mathcal{Z} \subseteq \mathcal{S}_Q$ since all three directions in $\mathcal{Z}$ are not orthogonal with directions in $\mathcal{Q}$. However, $\dim(\mathcal{Z}) = 3 = n$ while $\dim(\mathcal{S}_Q) = 4 = c$ so that $\mathcal{Z} \neq \mathcal{S}_Q$.

6. Conclusion

This paper extends recent sketched least squares analyses [9,12] and Maher's results on Schatten p-norms [31–34] to sketched MMLR in general Schatten p-norms by interpreting the sketched problem as a multiplicative perturbation. Our expressions for the exact and perturbed solutions in terms of projectors enable geometric interpretations of: 1) the action of the sketching matrix in terms of relevant subspaces, and 2) the difference between an orthogonal and oblique projector with the same range. As the results in the paper focus on general sketching matrices, we leave as future work investigating their implications for specific sketching algorithms.

Declaration of competing interest

The authors declare that they have no competing interests.

Funding

The work was supported in part by NSF grants DGE-1633587, DMS-1760374, and DMS-1745654.

References

[1] N. Ailon, B. Chazelle, The fast Johnson Lindenstrauss transform and approximate nearest neighbors, SIAM J. Sci. Comput. 39 (2009) 302–322.

[2] H. Avron, P. Maymounkov, S. Toledo, Blendenpik: supercharging LAPACK's least-squares solver, SIAM J. Sci. Comput. 32 (2010) 1217–1236.

[3] C. Boutsidis, P. Drineas, Random projections for the nonnegative least-squares problem, Linear Algebra Appl. 431 (2009) 760–771.

[4] L. Breiman, J.H. Friedman, Predicting multivariate responses in multiple linear regression, J. R. Stat. Soc., Ser. B, Stat. Methodol. 59 (1997) 3–54.

[5] J.J. Brust, R.F. Marcia, C.G. Petra, Computationally efficient decompositions of oblique projection matrices, SIAM J. Matrix Anal. Appl. 41 (2020) 852–870.

[6] E.J. Candes, Y. Plan, Matrix completion with noise, Proc. IEEE 98 (2010) 925–936.

[7] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (2009) 717.

[8] S. Chatterjee, A.S. Hadi, Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (1986) 379–416, With discussion.

[9] J.T. Chi, I.C.F. Ipsen, A geometric analysis of model- and algorithm-induced uncertainties for randomized least squares regression, arXiv:1808.05924, 2019.

[10] J. Chmieliński, Linear mappings approximately preserving orthogonality, J. Math. Anal. Appl. 304 (2005) 158–169.

[11] P. Drineas, M.W. Mahoney, S. Muthukrishnan, Sampling algorithms for l_2 regression and applications, in: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2006, pp. 1127–1136.

[12] P. Drineas, M.W. Mahoney, S. Muthukrishnan, T. Sarlós, Faster least squares approximation, Numer. Math. 117 (2011) 219–249.

[13] Z. Drmač, A.K. Saibaba, The discrete empirical interpolation method: canonical structure and formulation in weighted inner product spaces, SIAM J. Matrix Anal. Appl. 39 (2018) 1152–1180, https://doi.org/10.1137/17M1129635.

[14] H. Du, Z. Zhao, S. Wang, Q. Hu, Two-dimensional discriminant analysis based on Schatten p-norm for image feature extraction, J. Vis. Commun. Image Represent. 45 (2017) 87–94.

[15] M. Eyvazian, R. Noorossana, A. Saghaei, A. Amiri, Phase II monitoring of multivariate multiple linear regression profiles, Qual. Reliab. Eng. Int. 27 (2011) 281–296.

[16] I. Han, D. Malioutov, H. Avron, J. Shin, Approximating spectral sums of large-scale matrices using stochastic Chebyshev approximations, SIAM J. Sci. Comput. 39 (2017) A1558–A1585.

[17] P.C. Hansen, Oblique projections and standard-form transformations for discrete inverse problems, Numer. Linear Algebra Appl. 20 (2013) 250–258, https://doi.org/10.1002/nla.802.

[18] I. Hnětynková, M. Plešinger, D.M. Sima, Z. Strakoš, S. Van Huffel, The total least squares problem in $AX \approx B$: a new classification with the relationship to the classical works, SIAM J. Matrix Anal. Appl. 32 (2011) 748–770.

[19] I. Hnětynková, M. Plešinger, Z. Strakoš, The core problem within a linear approximation problem $AX \approx B$ with multiple right-hand sides, SIAM J. Matrix Anal. Appl. 34 (2013) 917–931.

[20] D.C. Hoaglin, R.E. Welsch, The hat matrix in regression and ANOVA, Am. Stat. 32 (1978) 17–22.

[21] I.C. Ipsen, T. Wentworth, The effect of coherence on sampling from matrices with orthonormal columns, and preconditioned least squares problems, SIAM J. Matrix Anal. Appl. 35 (2014) 1490–1520.

[22] D.I. Jeong, A. St-Hilaire, T.B. Ouarda, P. Gachon, Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator, Clim. Change 114 (2012) 567–591.

[23] C.R. Johnson, R.A. Horn, Topics in Matrix Analysis, Cambridge University Press, 1985.

[24] T. Kitahara, T. Tsuchiya, Proximity of weighted and layered least squares solutions, SIAM J. Matrix Anal. Appl. 31 (2009) 1172–1186, https://doi.org/10.1137/080725787.

[25] B.W. Larsen, T.G. Kolda, Practical leverage-based sampling for low-rank tensor decomposition, arXiv preprint, arXiv:2006.16438, 2020.

[26] S. Lefkimmiatis, J.P. Ward, M. Unser, Hessian Schatten-norm regularization for linear inverse problems, IEEE Trans. Image Process. 22 (2013) 1873–1888.

[27] Y. Li, B. Nan, J. Zhu, Multivariate sparse group Lasso for the multivariate multiple linear regression with an arbitrary group structure, Biometrics 71 (2015) 354–363.

[28] L. Luo, J. Yang, J. Chen, Y. Gao, Schatten p-norm based matrix regression model for image classification, in: Chinese Conference on Pattern Recognition, Springer, 2014, pp. 140–150.
[29] P. Ma, M.W. Mahoney, B. Yu, A statistical perspective on algorithmic leveraging, in: Proceedings of the 31st International Conference on International Conference on Machine Learning, 2014, pp. I-91–I-99.
[30] P. Ma, M.W. Mahoney, B. Yu, A statistical perspective on algorithmic leveraging, J. Mach. Learn. Res. 16 (2015) 861–911.
[31] P.J. Maher, Some operator inequalities concerning generalized inverses, Ill. J. Math. 34 (1990) 503–514.
[32] P.J. Maher, Some norm inequalities concerning generalized inverses, Linear Algebra Appl. 174 (1992) 99–110.
[33] P.J. Maher, Some norm inequalities concerning generalized inverses, 2, Linear Algebra Appl. 420 (2007) 517–525.
[34] P.J. Maher, Some singular values, and unitarily invariant norm inequalities concerning generalized inverses, Filomat 21 (2007) 99–111.
[35] M.W. Mahoney, Randomized Algorithms for Matrices and Data, Foundations and Trends® in Machine Learning, vol. 3, 2011, pp. 123–224.
[36] X. Meng, M.A. Saunders, M.W. Mahoney, LSRN: a parallel iterative solver for strongly over- or underdetermined systems, SIAM J. Sci. Comput. 36 (2014) C95–C118.
[37] B. Mojškerc, A. Turnšek, Mappings approximately preserving orthogonality in normed spaces, Nonlinear Anal., Theory Methods Appl. 73 (2010) 3821–3831.
[38] R. Noorossana, M. Eyvazian, A. Amiri, M.A. Mahmoud, Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application, Qual. Reliab. Eng. Int. 26 (2010) 291–303.
[39] G. Raskutti, M.W. Mahoney, A statistical perspective on randomized sketching for ordinary least-squares, J. Mach. Learn. Res. 17 (2016) 214.
[40] V. Rokhlin, M. Tygert, A fast randomized algorithm for overdetermined linear least-squares regression, Proc. Natl. Acad. Sci. USA 105 (2008) 13212–13217.
[41] T. Sarlós, Improved approximation algorithms for large matrices via random projections, in: 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06), IEEE, 2006, pp. 143–152.
[42] G.W. Stewart, On scaled projections and pseudoinverses, Linear Algebra Appl. 112 (1989) 189–193, https://doi.org/10.1016/0024-3795(89)90594-6.
[43] G.W. Stewart, On the numerical analysis of oblique projectors, SIAM J. Matrix Anal. Appl. 32 (2011) 309–348.
[44] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory. Computer Science and Scientific Computing, Academic Press, Inc., Boston, MA, 1990.
[45] J.G. Sun, Optimal backward perturbation bounds for the linear least-squares problem with multiple right-hand sides, IMA J. Numer. Anal. 16 (1996) 1–11.
[46] G.A. Thanei, C. Heinze, N. Meinshausen, Random projections for large-scale regression, arXiv:1701.05325, 2017.
[47] A. Turnšek, On mappings approximately preserving orthogonality, J. Math. Anal. Appl. 336 (2007) 625–631.
[48] S. Ubaru, J. Chen, Y. Saad, Fast estimation of $\mathtt{tr}(f(A))$ via stochastic Lanczos quadrature, SIAM J. Matrix Anal. Appl. 38 (2017) 1075–1099.
[49] A. Černý, Characterization of the oblique projector $U(VU)^\dagger V$ with application to constrained least squares, Linear Algebra Appl. 431 (2009) 1564–1570, https://doi.org/10.1016/j.laa.2009.05.025.
[50] P.F. Velleman, R.E. Welsch, Efficient computing of regression diagnostics, Am. Stat. 35 (1981) 234–242.
[51] H. Wang, R. Zhu, P. Ma, Optimal subsampling for large scale logistic regression, J. Am. Stat. Assoc. 113 (2018) 829–844.
[52] I.J. Wang, Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation, Evolution 67 (2013) 3403–3411.
[53] Q. Wang, F. Chen, Q. Gao, X. Gao, F. Nie, On the Schatten norm for matrix based subspace learning and classification, Neurocomputing 216 (2016) 192–199.
[54] Q. Wang, Q. Gao, X. Gao, F. Nie, Optimal mean two-dimensional principal component analysis with F-norm minimization, Pattern Recognit. 68 (2017) 286–294.

[55] S. Wang, A. Gittens, M.W. Mahoney, Sketched ridge regression: optimization perspective, statistical perspective, and model averaging, J. Mach. Learn. Res. 18 (2017) 8039–8088.
[56] M. Wei, A.R. De Pierro, Upper perturbation bounds of weighted projections, weighted and constrained least squares problems, SIAM J. Matrix Anal. Appl. 21 (2000) 931–951, https://doi.org/10.1137/S0895479898336306.
[57] D.P. Woodruff, et al., Sketching as a tool for numerical linear algebra, in: Foundations and Trends® in Theoretical Computer Science, vol. 10, 2014, pp. 1–157.
[58] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2016) 156–171.
[59] H. Zhou, L. Li, Regularized matrix regression, J. R. Stat. Soc., Ser. B, Stat. Methodol. 76 (2014) 463–483.
[60] P. Zhu, A.V. Knyazev, Angles between subspaces and their tangents, J. Numer. Math. 21 (2013) 325–340.