MINIMUM-VARIANCE PSEUDO-UNBIASED LOW-RANK ESTIMATOR FOR ILL-CONDITIONED INVERSE PROBLEMS

Isao Yamada and *Jamal Elbadraoui*†

Dept. of Communications and Integrated Systems (S3-60),
Tokyo Institute of Technology, Tokyo 152-8552, JAPAN
Email: isao@comm.ss.titech.ac.jp

ABSTRACT

This paper presents a mathematically novel low-rank linear statistical estimator named *minimum-variance pseudo-unbiased low-rank estimator* for applications to ill-conditioned linear inverse problems. Based on a simple fact: 'any low-rank estimator can not be a *(uniformly) unbiased estimator*', we introduce *pseudo-unbiased low-rank estimator*, as an ideal low-rank extension of unbiased estimators. The *minimum-variance pseudo-unbiased low-rank estimator* minimizes the *variance* of estimate among all *pseudo-unbiased low-rank estimators*, hence it is characterized as a solution to a double layered nonconvex optimization problem. The main theorem presents an algebraic structure of the *minimum-variance pseudo-unbiased low-rank estimator* in terms of the *singular value decomposition* of the model matrix in the linear statistical model. The *minimum-variance pseudo-unbiased low-rank estimator* is not only a best low-rank extension of the *minimum-variance unbiased estimator* (i.e., Gauss-Markov estimator) but also a nontrivial generalization of the Marquardt's low-rank estimator (Marquardt 1970).

1. INTRODUCTION

Inherent inadequacy of the *least squares estimator*, for ill-conditioned linear statistical model, is commonly understood through the fact [1]: *the order of the mean squared distance, between the estimate and the value of the parameter, achievable by the least squares estimator, is inversely proportional to the square of the smallest nonzero singular value of the model matrix (the additive noise is often assumed to be white)* [See Eq.(7)]. When the model matrix in the linear statistical model is of full column rank, the least squares estimator, under the above noise model, can alternatively be characterized as the unique *minimum-variance unbiased estimator*. As a result, any unbiased estimator can not remedy the above drawback of the least squares estimator. This fact motivates a variety of *biased estimators* realizing superior performance to the least squares estimator. These include for example, the *ridge regression estimator* [1, 2], essentially based on common idea of so called the *Tikhonov's regularization* [3–6][1], the *minimum-variance conditionally unbiased estimator subject to linear restrictions* [10], and the *rank-reduction estimator* [11, 12].

The rank-reduction estimator [11] is the generalized inverse of the best low-rank approximation to the model matrix of the linear statistical model, where the best low-rank approximation is given by vanishing small singular values in the *singular-value-decomposition* the model matrix. Indeed, Marquardt clearly proves that this simple idea greatly improves the mean squared error attained by the least squares estimator. This idea induced the *shrinkage estimator*[6, 7] and the *rank-shaping estimator*[13] which achieve similar effect by nonnegatively weighting the singular values of the generalized inverse of the model matrix. Chipman [12] recently shed light on the Marquardt's classical invention and presented *"a" generalization* of the Marquardt's idea to the case where the covariance of noise admits general positive definite matrices (See Remark 1).

In spite of the great invention of Marquardt's rank-reduction estimator, it has not yet been clear especially on the following fundamental questions: (i) *Can we characterise in any statistical estimation theoretic sense the Marquardt's rank-reduction estimator over all possible low-rank estimators ?* (ii) *Is the Marquardt's rank-reduction estimator optimal in any statistical estimation theoretic sense ?* (iii) *Is there any connection between the Marquardt's estimator and classical optimal estimator for example the minimum-variance unbiased estimator ?*

The 1st goal of this paper is to propose a novel low-rank linear statistical estimator named *minimum-variance pseudo-unbiased low-rank estimator* for application to ill-conditioned linear inverse problems. The 2nd goal of this paper is to show that this *minimum-variance pseudo-unbiased low-rank estimator* provides a unified statistical estimation theoretic view to understand the function of the *Marquardt's rank-reduction estimator* as well as the *minimum - variance unbiased estimator* (i.e., Gauss-Markov estimator). The main theorem presents an algebraic structure of the *minimum-variance pseudo-unbiased low-rank estimator* in terms of the *singular value decomposition* of the model matrix in the linear statistical model. By this theorem, it is revealed that the *minimum-variance pseudo-unbiased low-rank estimator* is a nontrivial generalization of the Marquardt's rank-reduction estimator. Due to lack of space, all proofs are omitted.

2. PRELIMINARIES

To clarify the background and the motivation of this study, we start with a brief review on the ill-conditioned linear inverse problem.

A. Estimation in Linear Regression Model

The linear statistical model assumes that we can observe data vector $\boldsymbol{y} \in \mathbb{R}^m$ of the form:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon} = \mathcal{L}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathcal{L} \in \mathbb{R}^{m \times n}$ is a known nonzero *model matrix* (or *design matrix*), $\boldsymbol{\beta} \in \mathbb{R}^n$ is an unknown *parameter vector* to be estimated, and $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ is a random vector with zero mean and with a positive

† This manuscript is made by the 1st author based partially on a joint work when the 2nd author was in Tokyo Institute of Technology (–March 2005). Currently, the 2nd author is working for the SONY Corporation.

[1]The ridge regression estimator requires thus a certain delicate choice, depending on the *unknown estimandum*, of the optimal *regularization parameter*, which has motivated many studies on efficient approximate selections of the optimal *regularization parameter* (See for example [7–9] and references therein).

definite covariance $E(\varepsilon\varepsilon^t) = Q \succ 0$. The vector $\boldsymbol{y}$ can be interpreted as the outcome of inexact measurement of $\boldsymbol{x}$. Indeed, $\boldsymbol{y}$ itself is a random vector because it is the sum of random vector ε and the constant vector $\mathcal{L}\boldsymbol{\beta}$.

Developing well-behaved linear estimator of the form

$$\widehat{\boldsymbol{\beta}} := \Phi\boldsymbol{y}\,(\approx \boldsymbol{\beta}),$$

where $\Phi \in \mathbb{R}^{n\times m}$ is a constant matrix, has been a primitive statistical issue in wide range of mathematical sciences and engineerings including, e.g., communication, ecomonics, signal processing, seismology, and control, More precisely, a major goal of the linear estimator is to find Φ suppressing the *mean square error* of $\Phi\boldsymbol{y}$:

$$\begin{aligned} J(\Phi) &:= E\left(\|\Phi\boldsymbol{y}-\boldsymbol{\beta}\|^2\right) \\ &= \underbrace{E\left(\|\Phi\boldsymbol{y}-E(\Phi\boldsymbol{y})\|^2\right)}_{\text{Variance}} + \underbrace{\|E(\Phi\boldsymbol{y})-\boldsymbol{\beta}\|^2}_{\text{Bias}^2} \\ &= \overbrace{\operatorname{tr}\left(\Phi Q\Phi^t\right)} + \overbrace{\|(\Phi\mathcal{L}-I)\boldsymbol{\beta}\|^2} \qquad (2) \end{aligned}$$

as much as possible, where "E" denotes the *expectation* and $\|\cdot\|$ stands for the *Euclidean norm*.

Since $\boldsymbol{\beta}$ is unknown, it is impossible to minimize $J(\Phi)$ globally over $\Phi \in \mathbb{R}^{n\times m}$, hence we have to optimize some other criteria in practical situations.

B. Least-squares estimation and Minimum-variance unbiased estimation

The *least-squares estimator* is defined as a mapping $\Phi_{ls} : \mathbb{R}^m \to \mathbb{R}^n$ satisfying

$$\Phi_{ls}(\boldsymbol{y}) \in \mathcal{S}_{\mathcal{L}}(\boldsymbol{y}) := \arg\min_{\widehat{\boldsymbol{\beta}}\in\mathbb{R}^n} \|\boldsymbol{y}-\mathcal{L}\widehat{\boldsymbol{\beta}}\|^2, \quad \forall \boldsymbol{y}\in\mathbb{R}^m.$$

For every $\boldsymbol{y} \in \mathbb{R}^n$, the *orthogonal projection theorem* ensures $\mathcal{S}_{\mathcal{L}}(\boldsymbol{y}) \neq \emptyset$ and the unique existence of minimum norm solution: $\widehat{\boldsymbol{\beta}}_{gi} = \arg\min_{\widehat{\boldsymbol{\beta}}\in\mathcal{S}_{\mathcal{L}}(\boldsymbol{y})} \|\widehat{\boldsymbol{\beta}}\|$. The mapping $\mathcal{L}^\dagger : \boldsymbol{y} \mapsto \widehat{\boldsymbol{\beta}}_{gi}$ is nothing but the *Moore-Penrose generalized inverse* of $\mathcal{L}$, hence $\mathcal{L}^\dagger$ is a realization of the least-squares estimator. Suppose the *singular value decomposition* (SVD) of $\mathcal{L}$ is given by

$$\mathcal{L} = U\Sigma V^t =: \sum_{i=1}^{\min(m,n)} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^t, \qquad (3)$$

where $U = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m] \in \mathbb{R}^{m\times m}$ and $V = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$ are orthogonal (i.e., $U^tU = I_m$ and $V^tV = I_n$) and

$$\begin{aligned} \Sigma &= \operatorname{diag}(\sigma_1, \cdots, \sigma_{\min(m,n)}) \\ &:= \begin{cases} \Sigma_m \in \mathbb{R}^{m\times m} & \text{if } m = n \\ \begin{bmatrix} \Sigma_n \\ \cdots \\ 0 \end{bmatrix} \in \mathbb{R}^{m\times n} & \text{if } m > n \\ \begin{bmatrix} \Sigma_m & \vdots & 0 \end{bmatrix} \in \mathbb{R}^{m\times n} & \text{if } m < n, \end{cases} \end{aligned}$$

with

$$\Sigma_{\min(m,n)} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{\min(m,n)} \end{bmatrix}.$$

The diagonal entry σ_i of $\Sigma_{\min(m,n)}$ is called the (i-th largest) singular value of $\mathcal{L}$ and satisfies $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\operatorname{rank}(\mathcal{L})} > 0$ and $\sigma_i = 0$ if $i > \operatorname{rank}(\mathcal{L})$.

The *Moore-Penrose generalized inverse* $\mathcal{L}^\dagger$ can be expressed as $\mathcal{L}^\dagger = V\Sigma^\dagger U^t$, where $\Sigma^\dagger := \operatorname{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_{\operatorname{rank}(\mathcal{L})}}, 0, \ldots, 0\right) \in \mathbb{R}^{n\times m}$. In particular, when

$$\operatorname{rank}(\mathcal{L}) = n, \qquad (4)$$

the set $\mathcal{S}_{\mathcal{L}}(\boldsymbol{y})$ is singleton, and the unique least-squares estimator is expressed simply by $\Phi_{ls}(\boldsymbol{y}) = \mathcal{L}^\dagger\boldsymbol{y} = \left(\mathcal{L}^t\mathcal{L}\right)^{-1}\mathcal{L}^t\boldsymbol{y}$. In this special case [i.e. (4)], $\Phi := \Phi_{ls}$ is an *unbiased* estimator, i.e., it satisfies

$$\Phi\mathcal{L} = I_n \qquad (5)$$

($\Longleftrightarrow E(\Phi\boldsymbol{y}) = \boldsymbol{\beta}, \forall\boldsymbol{\beta}\in\mathbb{R}^n$ [see (2)]).

Obviously, there exists an unbiased estimator Φ, hence Φ satisfies $\Phi\mathcal{L} = I_n$, if and only if $\mathcal{L}$ satisfies (4). Therefore, under the conditions (4) and $Q \succ 0$, a natural better candidate than Φ_{ls} is the solution to a constrained optimization problem:

Problem 1

$$\left.\begin{array}{ll} \textit{minimize} & \operatorname{tr}\left(\Phi Q\Phi^t\right) \\ \textit{subject to} & \Phi\mathcal{L} = I_n. \end{array}\right\}$$

Indeed, Problem 1 has the unique solution:

$$\Phi_{gm}(\boldsymbol{y}) = \left(\mathcal{L}^tQ^{-1}\mathcal{L}\right)^{-1}\mathcal{L}^tQ^{-1}\boldsymbol{y}, \quad \forall\boldsymbol{y}\in\mathbb{R}^m, \qquad (6)$$

satisfying obviously $J(\Phi_{gm}) \leq J(\Phi_{ls})$. The estimator Φ_{gm} is called the *minimum-variance unbiased estimator* (or *Gauss-Markov estimator*). From (6), when $Q = \sigma^2 I_m$, $\Phi_{gm} = \Phi_{ls}$ holds under the conditions (4). Moreover, when $Q = \sigma^2 I_m$ and $\operatorname{rank}(\mathcal{L}) \leq n$, it follows

$$J(\mathcal{L}^\dagger) = \underbrace{\sigma^2\left(\sum_{i=1}^{\operatorname{rank}(\mathcal{L})} \frac{1}{\sigma_i^2}\right)}_{\text{Variance}} + \underbrace{\sum_{i=\operatorname{rank}(\mathcal{L})+1}^{n} |\boldsymbol{v}_i^t\boldsymbol{\beta}|^2}_{\text{Bias}^2}. \qquad (7)$$

Eq.(7) implies that the least squares estimator yields inherently the drastic inadequacy when the linear regression model (1) is ill-posed, i.e., $\mathcal{L}$ possesses its singular values near zero [1].

As seen from the simplest case: $n = \operatorname{rank}(\mathcal{L})$ and $Q = \sigma^2 I$, where $\Phi_{gm} = \Phi_{ls}$ holds, any unbiased estimator can not remedy the above drastic inadequacy of the least squares estimator. To help circumvent this difficulty, several biased estimators have been developed, for example, the ridge regression [1, 2], the *minimum-variance conditionally unbiased estimator subject to linear restrictions* [10], the rank reduction estimator [11, 12], rank-shaping estimator [13].

C. Marquardt's idea for ill-conditioned linear regression

To avoid the drastic inadequacy of small singular values $\sigma_i \approx 0$ ($i = r+1, \ldots, \operatorname{rank}(\mathcal{L})$, $r < \operatorname{rank}(\mathcal{L})$) of $\mathcal{L}$, Marquardt [11] proposed, in place of using $\mathcal{L}^\dagger$, to use

$$\widetilde{\mathcal{L}}_r^\dagger \in \mathbb{R}^{m\times n}(r) := \{X \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(X) \leq r\}, \qquad (8)$$

which is the Moore-Penrose generalized-inverse of

$$\widetilde{\mathcal{L}}_r := U\begin{bmatrix} \Sigma_r & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} V^t \in \mathbb{R}^{m\times n}(r),$$

where

$$\Sigma_r = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix}.$$

The elimination of small singular value by $\widetilde{\mathcal{L}}_r$ is a natural choice according to the *Schmidt approximation theorem* (often credited to *Eckart-Young's theorem*):

$$\widetilde{\mathcal{L}}_r \in \arg \min_{X \in \mathbb{R}^{m \times n}(r)} \|X - \mathcal{L}\|_F,$$

where $\|\cdot\|_F$ stands for the Frobenius norm. Indeed, we have

$$J\left(\widetilde{\mathcal{L}}_r^\dagger\right) = \sigma^2 \left(\sum_{i=1}^{r} \frac{1}{\sigma_i^2}\right) + \sum_{i=r+1}^{n} |\boldsymbol{v}_i^t \boldsymbol{\beta}|^2,$$

which yields

$$\sum_{i=r+1}^{\operatorname{rank}(\mathcal{L})} \frac{1}{\sigma_i^2} > \frac{1}{\sigma^2}\|\boldsymbol{\beta}\|^2 \tag{9}$$

$$\Longrightarrow \sum_{i=r+1}^{\operatorname{rank}(\mathcal{L})} \frac{1}{\sigma_i^2} > \frac{1}{\sigma^2} \sum_{i=r+1}^{\operatorname{rank}(\mathcal{L})} |\boldsymbol{v}_i^t \boldsymbol{\beta}|^2 \Leftrightarrow J\left(\widetilde{\mathcal{L}}_r^\dagger\right) < J(\mathcal{L}^\dagger). \tag{10}$$

This fact implies that there exists a better estimator, in $\mathbb{R}^{n \times m}(r)$, than $\mathcal{L}^\dagger$ when $\mathcal{L}$ has sufficiently small singular values satisfying (9). However, any $\Phi \in \mathbb{R}^{n \times m}(r)$ can not satisfy (5), i.e., any $\Phi \in \mathbb{R}^{n \times m}(r)$ is no longer unbiased even when $\operatorname{rank}(\mathcal{L}) = n$. This observation induces a natural question: *Is there any better biased estimator, in $\mathbb{R}^{n \times m}(r)$, which can suppress not only its variance but also its bias more than $\mathcal{L}_r^\dagger$ does ?*

Remark 1 *(Chipman's generalization [12]) Marquardt's idea was recently extended to the case where the general linear statistical model (1) admits general $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t) = Q \succ 0$. As a low-rank approximation of $\mathcal{L}$, Chipman employed $\widetilde{\mathcal{L}}_r^{(Q)}$ satisfying*

$$\widetilde{\mathcal{L}}_r^{(Q)} \in \arg \min_{X \in \mathbb{R}^{m \times n}(r)} \|X - \mathcal{L}\|_{Q^{-1}}, \tag{11}$$

where $\|X - \mathcal{L}\|_{Q^{-1}} := \sqrt{\operatorname{tr}\left((X-\mathcal{L})^t Q^{-1} (X - \mathcal{L})\right)}$. As a generalization of Marquardt's rank-reduction estimator, Chipman proposed to use

$$\widetilde{\mathcal{L}}_r^{(Q)\ddagger} : \mathbb{R}^n \ni \boldsymbol{y} \mapsto \widehat{\boldsymbol{\beta}}_{ogi} := \arg \min_{\widehat{\boldsymbol{\beta}} \in \mathcal{S}_{\widetilde{\mathcal{L}}_r^{(Q)}}(\boldsymbol{y})} \|\widehat{\boldsymbol{\beta}}\|_{Q^{-1}},$$

where $\|\widehat{\boldsymbol{\beta}}\|_{Q^{-1}} := \sqrt{\operatorname{tr}\left(\widehat{\boldsymbol{\beta}}^t Q^{-1} \widehat{\boldsymbol{\beta}}\right)}$ and

$$\mathcal{S}_{\widetilde{\mathcal{L}}_r^{(Q)}}(\boldsymbol{y}) := \arg \min_{\widehat{\boldsymbol{\beta}} \in \mathbb{R}^n} \|\boldsymbol{y} - \widetilde{\mathcal{L}}_r^{(Q)} \widehat{\boldsymbol{\beta}}\|^2, \quad \forall \boldsymbol{y} \in \mathbb{R}^m.$$

Some matrix inequalities are derived as a generalization of (9) and (10) [Note: [12, (4.18)] corresponds to the relation (10)] . □

3. MINIMUM-VARIANCE PSEUDO-UNBIASED LOW-RANK ESTIMATION

In this paper, as a natural generalization of problem 1, we consider the following problem.

Problem 2 *(Minimum-variance pseudo-unbiased low-rank estimation (Type 1)) For the linear statistical model (1) and arbitrarily given $r \in \{1, 2, \ldots, \min(m, n)\}$. Then the problem is :*

$$\left.\begin{array}{ll} \textit{Minimize} & \operatorname{tr}\left(\Phi Q \Phi^t\right) \ : \textit{Variance of } \Phi Q \Phi^t \\ \textit{subject to} & \Phi \in \mathcal{S}_1, \\ \textit{where} & \mathcal{S}_1 := \arg \min_{X \in \mathbb{R}^{n \times m}(r)} \|X\mathcal{L} - I_n\|_F . \end{array}\right\}$$

Remark 2

(a) *Problem 2 is a natural generalization of classical minimum-variance unbiased estimator Φ_{gm} [see (6) as the solution to Problem 1] because of (2) and*

$$\textit{Bias}^2 = \|(\Phi\mathcal{L} - I_n)\boldsymbol{\beta}\|^2 \leq \left(\|\Phi\mathcal{L} - I_n\|_2\right)^2 \|\boldsymbol{\beta}\|^2 \tag{12}$$

$$\leq \quad \left(\|\Phi\mathcal{L} - I_n\|_F\right)^2 \|\boldsymbol{\beta}\|^2, \tag{13}$$

where we used expressions, with $\lambda_i\left[(\Phi\mathcal{L} - I_n)^t(\Phi\mathcal{L} - I_n)\right]$ [the i-th eigenvalue of $(\Phi\mathcal{L} - I_n)^t(\Phi\mathcal{L} - I_n)$],

$$\begin{aligned} \|\Phi\mathcal{L} - I_n\|_2 &= \sqrt{\max\{\lambda_i\left[(\Phi\mathcal{L} - I_n)^t(\Phi\mathcal{L} - I_n)\right]\}_{i=1}^n} \\ \|\Phi\mathcal{L} - I_n\|_F &= \sqrt{\sum_{i=1}^n \lambda_i\left[(\Phi\mathcal{L} - I_n)^t(\Phi\mathcal{L} - I_n)\right]}. \end{aligned}$$

(b) *We call any estimator $\Phi \in \mathcal{S}_1$ a pseudo-unbiased low-rank estimator (of Type 1). We call the solution (to Problem 2) $\Phi_r^* \in \arg\min_{\Phi \in \mathcal{S}_1} \operatorname{tr}\left(\Phi Q \Phi^t\right)$ the minimum-variance pseudo-unbiased low-rank estimator (of Type 1). Problem 2 is a meaningful (nontrivial) generalization of Problem 1 because the set $\mathcal{S}$ is not singleton in general (see Lemma 15). This fact is attribute not only to the possible singularity of $\mathcal{L}\mathcal{L}^t \in \mathbb{R}^{m \times m}$ but also to the distribution of singular values of*

$$\mathcal{L}^t \left[\left(\mathcal{L}\mathcal{L}^t\right)^{1/2}\right]^\dagger = V \begin{bmatrix} I_{\operatorname{rank}(\mathcal{L})} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} U^t.$$

(c) *The bounds in (12) and (13) tempt us to formulate, in place of Problem 2, another problem:*

$$\left.\begin{array}{ll} \textit{Minimize} & \operatorname{tr}\left(\Phi Q \Phi^t\right) \ : \textit{Variance of } \Phi Q \Phi^t \\ \textit{subject to} & \Phi \in \mathcal{S}_2, \\ \textit{where} & \mathcal{S}_2 := \arg \min_{X \in \mathbb{R}^{n \times m}(r)} \|X\mathcal{L} - I_n\|_2 . \end{array}\right\} \tag{14}$$

We call any estimator $\Phi \in \mathcal{S}_2$ a pseudo-unbiased low-rank estimator (of Type 2). We call the solution to (14) the minimum-variance pseudo-unbiased low-rank estimator (of Type 2) (The investigation of Type 2 is underway).

(d) *Problem 2 is a double layered constrained optimization problem, where the solution set of the 1st optimization problem is given as $\mathcal{S}_1$ over which $\operatorname{tr}\left(\Phi Q \Phi^t\right)$ must be minimized further. Since the set $\mathcal{S}_1$ is nonconvex in $\mathbb{R}^{n \times m}$ unlike the case discussed in [14], we have to utilize certain algebraic parametrization of $\mathcal{S}_1$ for the 2nd optimization.* □

The next lemma presents an algebraic parametrization of the set $\mathcal{S}_1$ of all *pseudo-unbiased low-rank estimators (of Type 1)*, based on which we reach finally Theorem 1 as the solution of Problem 2.

Lemma 1 *(Parametrization of $\mathcal{S}_1$) Suppose that $\mathcal{L} \in \mathbb{R}^{m \times n}$ is expressed as in a singular value decomposition (3) and $s = \operatorname{rank}(\mathcal{L})$. Then for any $r \in \{1, 2, \ldots, \min(m, n)\}$, the followings are equivalent:*

(a) $\Phi \in \mathcal{S}_1$.

(b) Φ *is expressed as*

$$\Phi = U_{(r'+\mu)}(I_{r'} \oplus \Gamma)\left\{(I_{r'} \oplus \Gamma)^t (U_{(r'+\mu)})^t \mathcal{L}^t(\mathcal{L}\mathcal{L}^t)^\dagger + Z\left(I_m - (\mathcal{L}\mathcal{L}^t)(\mathcal{L}\mathcal{L}^t)^\dagger\right)\right\},$$

where (i) $(\mu, r') = (s, 0)$ *for* $r \leq s$, $(\mu, r') = (\min(m,n) - s, s)$ *for* $r > s$, *(ii)* $U_{(r'+\mu)} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{r'+\mu}] \in \mathbb{R}^{m\times(r'+\mu)}$, *(iii)* $\Gamma \in \mathbb{R}^{\mu\times(r-r')}$ *satisfying* $\Gamma^t\Gamma = I_{r-r'}$, *(iv)* $Z \in \mathbb{R}^{r\times m}$, *and (v)* $I_{r'} \oplus \Gamma = \begin{bmatrix} I_{r'} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \Gamma \end{bmatrix}$. □

Theorem 1 *(Algebraic expression of the minimum-variance pseudo-unbiased low-rank estimator (Type 1)) Suppose that* $\mathcal{L} \in \mathbb{R}^{m\times n}$ *is expressed as in a singular value decomposition (3) and* $s = \operatorname{rank}(\mathcal{L})$. *Then for any* $r \in \{1, 2, \ldots, \min(m,n)\}$, *a solution* $\Phi_r^* \in \mathbb{R}^{n\times m}(r)$ *to Problem 2 is given by*

$$\Phi_r^* := \begin{cases} V_{(s)}\Gamma^*(\Gamma^*)^t\left(V_{(s)}\right)^t\mathcal{L}^\dagger\left[I_m - Q^{1/2}\left((I_m - \mathcal{L}\mathcal{L}^\dagger)Q^{1/2}\right)^\dagger\right] \\ \quad \text{if } r \leq s, \\ \mathcal{L}^\dagger\left[I_m - Q^{1/2}\left((I_m - \mathcal{L}\mathcal{L}^\dagger)Q^{1/2}\right)^\dagger\right] \\ \quad \text{if } r \geq s, \end{cases} \tag{15}$$

and $\operatorname{tr}(\Phi_r^* Q \Phi_r) = \sum_{1\leq i\leq r}\gamma_i$, *where (i)* $V_{(s)} := [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_s]$, *and (ii)* $(\gamma_i)_{i=1}^s$ *(*$\gamma_1 \leq \cdots \leq \gamma_s$*: non-decreasing order) and* $\Gamma^* := [\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_r] \in \mathbb{R}^{s\times r}$ *satisfying* $(\Gamma^*)^t\Gamma^* = I_r$ *are given by orthogonal matrix* $[\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_s] \in \mathbb{R}^{s\times s}$ *through any eigenvalue decomposition*

$$\sum_{i=1}^s \gamma_i \boldsymbol{\nu}_i \boldsymbol{\nu}_i^t := \left(V_{(s)}\right)^t \mathcal{L}^\dagger Q^{1/2} \left[I_m - \left(Q^{1/2}(I_m - \mathcal{L}\mathcal{L}^\dagger)\right)\left(Q^{1/2}(I_m - \mathcal{L}\mathcal{L}^\dagger)\right)^\dagger\right] Q^{1/2}\left(\mathcal{L}^\dagger\right)^t V_{(s)} \in \mathbb{R}^{s\times s}.$$

In particular, we have Φ_r^* *[in (15)]* $= \widetilde{\mathcal{L}}_r^\dagger$ *[in (8)] if* $Q = \sigma^2 I_m$ *and* $r < s$. □

Remark 3

(a) *For* $r \geq s$, Φ_r^* *does not depend on* r, *hence* $\Phi_r^* \in \mathbb{R}^{n\times m}(s)$ *is guaranteed.*

(b) *By comparison between Problem 2 and Problem 1, the 2nd expression in (15) presents a nontrivial generalization of the minimum-variance unbiased estimator (i.e., Gauss-Markov estimator)* Φ_{gm} *given in (6). Indeed, by Problem 2, we have a new expression of the Gauss-Markov estimator:* $\Phi_{gm} = \Phi_n^*$ *when the condition (4) holds. Major difference between* Φ_{gm} *and* Φ_n^* *is that* Φ_n^* *is always well-defined while* Φ_{gm} *is defined only when the condition (4) is satisfied. This new expression* Φ_n^* *is regarded a natural extension of* Φ_{gm}. *In this sense, we call* Φ_n^* *the minimum-variance pseudo-unbiased estimator (Type 1) (without rank reduction).*

(c) *For* $r = s$, *two expressions in (15) are same. By the orthogonality of* $\Gamma^* \in \mathbb{R}^{s\times s}$, *this fact is verified as*

$$V_{(s)}\Gamma^*(\Gamma^*)^t\left(V_{(s)}\right)^t\mathcal{L}^\dagger = V_{(s)}\left(V_{(s)}\right)^t\mathcal{L}^\dagger = \mathcal{L}^\dagger.$$

(d) *The last statement of Theorem 1 implies that Marquardt's rank-reduction estimator can be characterized as a special example of the minimum-variance pseudo-unbiased low-rank estimator (Type 1) corresponding to the case* $Q = \sigma^2 I$. □

4. CONCLUSION

As an ideal low-rank extension of the *minimum-variance unbiased estimator* for ill-conditioned linear inverse problems, we proposed a mathematically novel low-rank estimator named *minimum variance pseudo-unbiased low-rank estimator*. Thanks to the algebraic structure of the set of all *pseudo-unbiased low-rank estimators*, an algebraic expression of the proposed estimator is obtained. Finally, it is revealed that the proposed estimator is not only a best low-rank extension of the *minimum-variance unbiased estimator* but also a nontrivial generalization of the Marquardt's low-rank estimator.

Acknowledgement: The authors would like to thank Prof. Kohichi Sakaniwa of Tokyo Institute of Technology for helpful discussions and comments.This work was supported in part by JSPS Grants-in-Aid (C-17500139).

5. REFERENCES

[1] A. E. Horel and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.

[2] A. E. Horel, "Application of ridge analysis to regression problems," *Chemical Engineering Progress*, vol. 58, pp. 54–59, 1962.

[3] J. D. Riley, "Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix," *Math. Tables Aids Comput.*, vol. 9, pp. 96–101, 1955.

[4] D. L. Philips, "A technique for the numerical solution of certain integral equations of the first kind," *J. Assoc. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[5] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet. Math. Doktl.*, vol. 5, pp. 1035–1038, 1963.

[6] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses — Theory and Applications, 2nd ed.*, Springer, 2003.

[7] M. Goldstein and A. F. M. Smith, "Ridge-type estimators for regression analysis," *J. Roy. Statist. Soc. Ser. B*, vol. 36, pp. 284–291, 1974.

[8] R. L. Obenchain, "Good and optimal ridge estimator," *Ann. Statist.*, vol. 6, pp. 1111–1121, 1978.

[9] P. C. Hansen, "The L-curve and its use in the numerical treatment of inverse problems," in *Computational Inverse Problems in Electrocardiology*, P. R. Johnston, Ed., pp. 119–142. WIT Press, 2001.

[10] J. S. Chipman, "On least squares with insufficient observations," *J. Amer. Statist. Assoc.*, vol. 59, pp. 1078–1111, 1964.

[11] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, pp. 591–612, 1970.

[12] J. S. Chipman, "Linear restrictions, rank reduction, and biased estimation in linear regression," *Linear Algebra and its Applications*, vol. 289, pp. 55–74, 1999.

[13] A. J. Thorpe and L. L. Scharf, "Data adaptive rank-shaping methods for solving least squares problems," *IEEEL Trans. on Signal Processing*, vol. 43, pp. 1591–1601, 1995.

[14] I. Yamada, N. Ogura, and N. Shirakawa, "A numerically robust hybrid steepest descent method for the convexly constrained generalized inverse problems," in *Inverse Problems, Image Analysis, and Medical Imaging*, Z. Nashed and O. Scherzer, Eds., vol. Contemporary Mathematics, 313, pp. 269–305. American Mathematical Society, 2002.