An Analysis of Weighted Least Squares Method and Layered Least Squares Method with the Basis Block Lower Triangular Matrix Form*

Tomonari Kitahara† Takashi Tsuchiya‡

May 2008

Abstract

In this paper, we analyze the limiting behavior of the weighted least squares problem $\min_{x \in \Re^n} \sum_{i=1}^{p} \|D_i(A_i x - b_i)\|^2$, where each D_i is a positive definite diagonal matrix. We consider the situation where the magnitude of the weights are drastically different block-wisely so that $\max(D_1) \geq \min(D_1) \gg \max(D_2) \geq \min(D_2) \gg \max(D_3) \geq \ldots \gg \max(D_{p-1}) \geq \min(D_{p-1}) \gg \max(D_p)$. Here $\max(\cdot)$ and $\min(\cdot)$ represents the maximum and minimum entries of diagonal elements, respectively. Specifically, we consider the case when the gap $g \equiv \min_i 1/(\|D_i^{-1}\|\|D_{i+1}\|)$ is very large or tends to infinity. Vavasis and Ye proved that the limiting solution exists (when the proportion of diagonal elements within each block D_i is unchanged and only the gap g tends to ∞), and showed that the limit is characterized as the solution of a variant of the least squares problem called the *layered least squares (LLS) problem*. We analyze the difference between the solutions of WLS and LLS quantitatively and show that the norm of the difference of the two solutions is bounded above by $O(\chi_A \bar{\chi}_A^{2(p+1)} g^{-2}\|b\|)$ and $O(\bar{\chi}_A^{2p+3} g^{-2}\|b\|)$ in the variable and the residual spaces, respectively, using the two condition numbers $\chi_A \equiv \max_{B \in \mathcal{B}} \|B^{-1}\|$ and $\bar{\chi}_A \equiv \max_{B \in \mathcal{B}} \|B^{-1}A\|$ of A, where $\mathcal{B}$ is the set of all nonsingular $n \times n$ submatrix of A, $A = [A_1; \ldots; A_p]$ and $b = [b_1; \ldots; b_p]$. A remarkable feature of this result is the error bound is represented in terms of A, g (and b) and independent of the weights D_i, $i = 1, \ldots, p$. The analysis is carried out by making the change of variables to convert the matrix A into a basis lower-triangular form and then by applying the Sharmann-Morrison-Woodbury formula.

---

*A part of this research was supported with Grant-in-Aid for Scientific Research (B), 2008, 20340024 from the Japan Society for the Promotion of Science.

†Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 Japan (`kitahara.t.ab@m.titech.ac.jp`).

‡The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569 Japan (`tsuchiya@sun312.ism.ac.jp`).

1 Introduction

Let us consider the following weighted least squares (WLS) problem

$$\min_x \sum_{i=1}^{p} \|D_i(A_i x - b_i)\|^2 \tag{1}$$

where $A_i \in \Re^{m_i \times n}$, $b_i \in \Re^{m_i}$ for $i = 1, \ldots, p$ and each D_i is an $m_i \times m_i$-dimensional positive definite diagonal matrix, that is, $D_i \in \mathcal{D}_{++}^{m_i}$.

Let

$$A = [A_1; \ldots; A_p] = \begin{bmatrix} A_1 \\ \vdots \\ A_p \end{bmatrix}, \; b = [b_1; \ldots; b_p] = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}. \tag{2}$$

(We use analogous "Matlab-like" notations to concatenate matrices throughout the paper.) We assume that the column vectors of A are linearly independent and A does not contain a zero row vector. Then the solution of the weighted least squares problem (1) is unique and is written as follows:

$$x^*_{WLS} = (A^T D^2 A)^{-1} A^T D^2 b,$$

where $D = [D_1; \ldots; D_p]$.

In this paper, we analyze the asymptotic behavior of the solution of (1) when the magnitude of the block weights D_i is drastically different. This means (without loss of generality) that

$$\begin{array}{rcl} \max(D_1) & \geq & \min(D_1) \\ \gg \max(D_2) & \geq & \min(D_2) \\ \gg \max(D_3) & \geq & \min(D_3) \\ & \vdots & \\ \gg \max(D_p) & \geq & \min(D_p) \end{array} \tag{3}$$

holds, where $\min(D)$ and $\max(D)$ denote the minimum and the maximum of the diagonal elements of the matrix D, respectively. We define the gap of the weights as

$$g = \min_{i=1,\ldots,p-1} \min\,(D_i)/\max\,(D_{i+1}) = \min_{i=1,\ldots,p-1} (\|D_{i+1}\|\|D_i^{-1}\|)^{-1}. \tag{4}$$

In terms of g, the purpose of this paper is described as follows: we analyze the asymptotic behavior of (1) when $g \to \infty$.

The simplest type of this problem is the ordinarily weighted least squares problem where each D_i is proportional to the identity matrix. The more general form we consider here arises in the context of interior-point methods for linear programming [2, 14, 17, 18].

Intuitively, when g is large, one may solve this problem approximately by the following heuristic idea. Since the first term in the sum (1) dominates in the quadratic objective function, we first minimize this term ignoring other terms. Then we minimize the second term under the condition that the residual of the first term is fixed. We repeat this procedure to the last term. This procedure is called the layered least squares (LLS) method, and formally described as follows.

The Layered Least Squares Method
(1)Set $i = 1$ and $x^0 = 0$.
(2)Let Δx^* be the solution of the following problem:

$$\begin{array}{ll} \min & \|D_i\{(A_i(x^{i-1} + \Delta x) - b_i\}\|^2 \\ \text{subject to} & A_j \Delta x = 0, \; j = 1, \ldots, i-1, \end{array} \tag{5}$$

where the problem is unconstrained when $i = 1$. Set $x^i := x^{i-1} + \Delta x^*$.
(3)If $i = p$, output x^p. Otherwise, increase i by one and return to (2).

We denote by x^*_{LLS} the layered least squares solution.

LLS was introduced by Vavasis and Ye in their seminal work [17] of the polynomial-time interior-point method whose complexity just depends on the coefficient matrix. Vavasis and Ye proved that the solution of WLS converges to that of LLS when g tends to infinity (when the proportion of diagonal elements within each block D_i is unchanged and only the gap g tends to ∞). In the following few paragraphs, we briefly discuss potential importance of LLS as a model through two examples. The first one is from statistics and the second one is related to optimal design of magnetic shielding [10].

Suppose that we want to solve a least squares problem which includes both accurate data and inaccurate ones. In the notation above, the accurate data and inaccurate ones correspond to A_1 and A_2, respectively. A natural choice avoiding the influence of the bad data would be to consider a WLS problem assigning large weights to accurate data, i.e., $D_1 = d_1 I, \; d_1 \gg 1$ and $D_2 = I$. A difficult problem here is how to choose the weight d_1. Then LLS is considered as a natural alternative because the solution of WLS tends to the solution of LLS when $d_1 \to \infty$ and now we have no need to determine d_1.

We move on to the second example. In [10], we dealt with an optimal design problem of magnetic shielding of MAGLEV train (linear motor car). The problem was formulated as a second-order cone program (SOCP). One of the relevant idea for SOCP formulation is to consider the situation where permeability of shielding material (iron, say) goes to infinity in the Maxwell equation which appears as the constraint of SOCP, and the proper limit was taken based on physical consideration. This model is also derived by using the idea of LLS [9, 16]. If we adopt the variational formulation of the Maxwell equation, we obtain a weighted least squares problem associated with the system where permeability appears as weight. The LLS problem obtained by letting the permeability to infinity imposes the constraint identical to the one obtained initially through physical consideration. This is another example where LLS can be useful as a modeling tool.

In the above examples LLS is considered as an approximation to WLS representing real optimization model. The relation between LLS and WLS can be viewed from the other side. Knowing LLS as a useful model, one may want to consider using LLS in his/her analysis. In such a case, a major difficulty which prevents he/her from using LLS would be its complication; LLS is a bit involved constrained least squares problem, though its interpretation as a model can be easy and simple. A solution to the WLS with a large gap g may well serve as a simple approximation to the LLS solution. Thus, WLS may help in utilizing LLS as a model.

Thus the idea of LLS can be applicable to many fields. Therefore, it is important to analyze theoretical properties of LLS in connection with WLS. In this paper, we study asymptotic behavior of the WLS method quantitatively and provide a bound on difference between the WLS solution and the LLS solution. The main contributions of this paper are summarized as follows.

1. We introduce the basis block lower-triangluar (BBLT) form of A to analyze WLS and LLS. This is the major tool of our analysis.

2. We develop a bound of the difference between the solutions of WLS and LLS. It is shown that the difference is bounded above by $O(\chi_A \bar{\chi}_A^{2(p+1)} g^{-2}\|b\|)$ and $O(\bar{\chi}_A^{2p+3} g^{-2}\|b\|)$ in the variable and the residual spaces, respectively. where χ_A and $\bar{\chi}_A$ are condition numbers associated with A studied by several authors including Khachiyan and Dikin.

An interesting remarkable feature of our result is that the bound is just expressed in terms of the condition numbers of A and the gap g (and b). This is quite anti-intuitive in that generally such an error bound is expected to contain some quantity related to each D_i in addition to g. A special case where every D_i is proportional to the identity matrix is analyzed in [15]; this paper is a generalization of [15].

This paper is organized as follows. In Section 2, we introduce the block lower-triangular (BLT) matrices and the basis block lower-triangular (BBLT) matrices. In Section 3, we introduce two condition numbers associated with A, namely χ_A and $\bar{\chi}_A$, and state their properties. In Section 4, we derive the bound of difference between WLS and LLS based on the BBLT matrix and Sherman-Morrison-Woodbury formula.

2 Block lower-triangular matrices and the layered least squares method

2.1 The Block Lower-Triangular Matrix

Let

$$\bar{A}_i = \begin{bmatrix} A_1 \\ \vdots \\ A_i \end{bmatrix}.$$

Let $\bar{r}_i$ be the rank of $\bar{A}_i$. Note that $\bar{r}_i$ is a monotonically increasing function of i. Also define $r_i = \bar{r}_i - \bar{r}_{i-1}$, where $r_1 = \bar{r}_1$.

We first observe that after an appropriate change of variables $z = Gy$ with a nonsingular matrix G, the least squares problem can be transformed into another one with a block lower triangular (BLT) matrix $T = AG^{-1}$ whose row block corresponds to the partition of layers determined by $A_1, \ldots, A_p$. The BLT form was introduced in [14] in the analysis of the affine-scaling algorithm and then utilized in [6] in the study of a variant of the Vavasis-Ye layered-step interior-point algorithm.

Theorem 2.1 *For the matrix A defined in (2), there exists a nonsingular matrix $G \in \Re^{n\times n}$ satisfying the following properties:*

1. *$T := AG^{-1}$ is a $p \times p$ block lower-triangular (BLT) matrix, i.e.*

$$T = \begin{bmatrix} T_1 \\ \vdots \\ T_p \end{bmatrix} = \begin{bmatrix} T_{11} & O & \cdots & O \\ T_{21} & T_{22} & O & O \\ \vdots & \vdots & \vdots & O \\ T_{p1} & T_{p2} & \cdots & T_{pp} \end{bmatrix} \qquad (6)$$

   where T_{ij} is an $m_i \times r_j$ matrix, for $1 \le i, j \le p$.

2. *For all $i = 1, \ldots, p$, $\mathrm{rank}(T_{ii}) = r_i$, that is, the columns of T_{ii} are linearly independent.*

3. *For $k = 1, \ldots, p$, we have*

$$\mathrm{Im} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix} = \mathrm{Im} \begin{bmatrix} T_{11} & O & \cdots & O \\ T_{21} & T_{22} & O & O \\ \vdots & \vdots & \vdots & O \\ T_{k1} & T_{k2} & \cdots & T_{kk} \end{bmatrix}.$$

Proof. For a matrix C of the same size as T (and hence A), by C_{pq} we mean the (p,q) block of C corresponding to the block determined by T_{pq} above.

First let us choose r_1 column vectors of A_1 which form a basis for $\mathrm{Im}(A_1)$, and perform exchange of columns (they are column elementary transformations) to A to place these vectors becomes the first r_1 columns. Let A' and A'_1 be A and A_1 after this operation. Then we perform another sequence of column elementary transformations to A' to eliminate r_1+1 to $n-r_1$ column vectors of A'_1. After these operations the first row block becomes $[T_{11}\ 0\ \dots\ 0]$. Let $T^{(1)}$ be the resulting matrix.

Then we continue to the second step. Let T' be the submatrix of $T^{(1)}$ consisting of the $[T_{22}^{(1)}\ \dots, T_{2p}^{(1)}]$ blocks. Performing similar column transformations as in the first step, T' is converted to the form $[*\ O]$, where $*$ is an $m_2\times r_2$ matrix and O is $(n-r_2)\times m_2$ matrix, without affecting the first r_1 columns of $T^{(1)}$. By repeating this procedures, we obtain T satisfying the conditions 1 and 2. Thus there is a nonsingular matrix $G\in\Re^{n\times n}$ such that $T=AG^{-1}$. Finally, the condition 3 is satisfied obviously from the construction. □

With the BLT matrix T, the layered least squares method is described in a simplified manner as follows. Suppose that $T=AG^{-1}$ is a BLT matrix for some nonsingular matrix $G\in\Re^{n\times n}$. Remember that in LLS, the subproblem (5) is to be solved. Let us define $\Delta u=G\Delta x$, $q_i=b_i-A_ix^{i-1}$ and break the variable vector Δu into p layers, that is, $\Delta u=[\Delta u_1;\dots;\Delta u_p]$. Then, the subproblem (5) to be solved in the layered least squares method becomes

$$\begin{array}{ll} \min & \|D_i(q_i-[T_{i1},\dots,T_{ii}][\Delta u_1;\dots;\Delta u_i])\|^2 \\ \text{subject to} & [T_1;\cdots;T_{i-1}][\Delta u_1,\dots,\Delta u_p]=0, \end{array}$$

Notice that the constraint is

$$\begin{bmatrix} T_{11} & O & \cdots & \cdots & \cdots & O \\ T_{21} & T_{22} & O & \cdots & \cdots & O \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{(i-1)1} & \cdots & T_{(i-1)(i-1)} & O & \cdots & O \end{bmatrix} \begin{bmatrix} \Delta u_1 \\ \Delta u_2 \\ \vdots \\ \Delta u_p \end{bmatrix} = 0.$$

From the first block, we have $\Delta u_1=0$ as the columns of T_{11} are linearly independent. Using this equation and the fact that T_{22} is full column rank, it follows that $\Delta u_2=0$. By proceeding similarly, we obtain $[\Delta u_1;\Delta u_2;\dots;\Delta u_{i-1}]=[0;0;\dots,0]$. Then the subproblem turns out to be the normal weighted least squares problem

$$\min\ \|D_i(q_i-T_{ii}\Delta u_i)\|^2,$$

which has a unique solution $\Delta u_i^*=(T_{ii}^TD_i^2D_{ii})^{-1}T_{ii}^TD_i^2q_i$. Thus if we use the BLT matrix, the layered least squares method can be represented as follows.

The Layered Least Squares Method using BLT
Let $T = AG^{-1}$ be a BLT matrix for some nonsingular matrix $G \in \Re^{n\times n}$.
(1)Set $i = 1$ and $q^0 = b$.
(2)Compute $\Delta u_i^* = (T_{ii}^T D_i^2 D_{ii})^{-1} T_{ii}^T D_i^2 q_i^{i-1}$.
(3)Update the residual by $[q_i^i; \cdots; q_p^i] = [q_i^{i-1}; \cdots; q_p^{i-1}] - [T_{ii}; \cdots; T_{pi}]\Delta u_i^*$.
(4)If $i = p$ stop. Output the solution $G^{-1}[u_1^*; \cdots; u_p^*]$. Otherwise increase i by one and return to (2).

We have the following proposition.

Proposition 2.1 *Let $T = AG^{-1}$ be a BLT matrix for some nonsingular matrix $G \in \Re^{n\times n}$. Then the solution of the layered least squares problem can be represented as $x^* = G^{-1}S_L^1 b$, where the matrix S_L^i, $i = 1, \ldots, p$ is defined by the recursive equation*

$$S_L^i = \begin{cases} \begin{bmatrix} (T_{ii}^T D_i^2 T_{ii})^{-1} T_{ii}^T D_i^2 & O \\ -S_L^{i+1} \begin{bmatrix} T_{i+1,i} \\ \vdots \\ T_{p,i} \end{bmatrix} (T_{ii}^T D_i^2 T_{ii})^{-1} T_{ii}^T D_i^2 & S_L^{i+1} \end{bmatrix} & i = 1, \ldots, p-1, \\ (T_{pp}^T D_p^2 T_{pp})^{-1} T_{pp}^T D_p^2 & i = p. \end{cases} \tag{7}$$

The matrix S_L^1 is the linear operator to compute LLS. In the following, S_L^1 is denoted by S_L. Let $T = AG^{-1}$ be a BLT matrix for some matrix $G \in \Re^{n\times n}$. For $D = [D_1; \ldots; D_p]$, define the matrix $S_W = (T^T D^2 T)^{-1} T^T D^2$. As we mentioned in Section 1, $x^* = G^{-1} S_W b$ is a unique solution of the weighted least squares problem (1). Then the difference between the solution x_{WLS}^* of WLS and the solution x_{LLS}^* of LLS is given by

$$x_{WLS}^* - x_{LLS}^* = G^{-1}(S_W - S_L)b.$$

Furthermore, the difference between the residual of the weighted least squares method and that of layered squares method becomes

$$(b - Ax_{WLS}^*) - (b - Ax_{LLS}^*) = -T(S_W - S_L)b.$$

In Section 4, we analyze how these values depend on the gap (4) and condition numbers of A we will discuss in Section 3.

2.2 The Basis Block Lower-Triangular Matrix

Now we introduce the basis block lower triangular (BBLT) matrix [15]. An $m \times n$ matrix T is called a BBLT matrix if the following condition is satisfied.

1. T is a $p \times p$ BLT matrix.
2. T contains, after appropriate row permutations, the n-dimensional identity matrix.
3. Each T_{ii} contains, after appropriate row permutations, the r_i-dimensional identity matrix.

In our setting, there exists a nonsingular matrix G which convert A into a BBLT matrix AG^{-1}.

Theorem 2.2 *There exists a nonsingular matrix $G \in \Re^{n\times n}$ such that $T = AG^{-1}$ is a BBLT matrix.*

Proof. Let $\hat{T} = [\hat{T}_1; \dots; \hat{T}'_p] = AG_1^{-1}$ be a BLT matrix of A. We first deal with the first layer $[\hat{T}_{11}, O, \dots, O]$. Let $\hat{t}_{1k}$ be a nonzero element in the first row of $\hat{T}_{11}$. Multiply the k-th column by $1/\hat{t}_{1k}$ and change the resulting column with the first column. After that, sweep the first row using $(1, 1)$ element (that is, 1). Note these are achieved by column elementary transformations. Repeating the procedure, we can change $\hat{T}$ so that the first layer contains the $r_1 \times r_1$ identity matrix. Let $\hat{T}^{(1)}$ be the resulting matrix.

Then let us go on to the second layer. Let $\hat{t}^{(1)}_{m_1+1,k'}$ be a nonzero element in the m_1+1-th row. Multiply the k'-th column by $1/\hat{t}^{(1)}_{m+1,k'}$ and change the resulting column with the r_1+1-th column. After that, sweep the m_1+1-th row of $\hat{T}^{(1)}$ using (m_1+1, r_1+1) element. This changes the m_1+1-th row of $\hat{T}^{(1)}$ to the form $(0, \dots, 0, 1, 0, \dots, 0)$. We remark that the first layer is not affected by the series of manipulations. These procedures are repeated to the last row of the second layer, that is, m_1+m_2-th row. It is easy to see that we can obtain a BBLT matrix T by executing similar procedures from the third layer to the p-th layer. □

Note that AG^{-1} is nothing but "a basis matrix" in the context of linear programming.

3 Condition Numbers of Matrix

In this section, we introduce two condition numbers of a matrix and state their properties.

Definition 3.1 *Let $C \in \Re^{m \times n}$ be a matrix whose column vectors are linearly independent. We define the condition numbers χ_C, $\bar{\chi}_C$ of C as follows.*

$$\chi_C := \sup_{D \in \mathcal{D}^m_{++}} \|(C^T D C)^{-1} C^T D\|, \tag{8}$$

$$\bar{\chi}_C := \sup_{D \in \mathcal{D}^m_{++}} \|C(C^T D C)^{-1} C^T D\|. \tag{9}$$

Finiteness of these numbers was first proved by Dikin [2]. Later Stewart [11], Todd [12], Ben-Tal and Teboulle [1] gave other proofs independently.

We summarize in the next proposition several important facts about χ_C and $\bar{\chi}_C$.

Proposition 3.1 *Let $C \in \Re^{m \times n}$ be a column full rank matrix. Then the following statements hold:*

(a) $\chi_C = \max\{\|G^{-1}\| : G \in \mathcal{G}\}$ *where $\mathcal{G}$ denotes the set of all $n \times n$ nonsingular submatrices of C;*

(b) $\bar{\chi}_C = \max\{\|CG^{-1}\| : G \in \mathcal{G}\}$ *where $\mathcal{G}$ denotes the set of all $n \times n$ nonsingular submatrices of C;*

(c) $\bar{\chi}_{CG} = \bar{\chi}_C$ *for any nonsingular matrix $G \in \Re^{m \times m}$;*

(d) Both χ_C and $\bar{\chi}_C$ are invariant under row and/or column permutations.

(e) Suppose that C contains the n-dimensional identity matrix as its submatrix. Then for any submatrix $\tilde{C}$ of C, we have $\|\tilde{C}\| \le \bar{\chi}_C$.

(f) Suppose that C contains the n-dimensional identity matrix as its submatrix. Let $\tilde{C}$ be any submatrix of C, consisting of columns of C. Then we have $\bar{\chi}_{\tilde{C}} \le \bar{\chi}_C$.

(g) If C includes n-dimensional identity matrix as its submatrix, we have $\chi_C \le \bar{\chi}_C$.

(h) Let $\tilde{C}$ be a submatrix of C consisting of its rows. Suppose that $\tilde{C}$ is column full rank. Then we have $\bar{\chi}_{\tilde{C}} \le \bar{\chi}_C$.

Proof. For (a) and (b), see Corollary 2.2 of Forsgren [3] and Theorem 1 of Todd et al. [13], respectively. (c) and (d) are obvious from the definition. For (e), first note that $\|\tilde{C}\| \le \|C\|$. With this and $\|C\| \le \bar{\chi}_C$ from (b), we have the desired result. For (f) refer to Proposition 2.2 of Monteiro and Tsuchiya [7]. Next we show (g). From the assumption, C can be written as

$C = [I; R_C]$ for some matrix R_C and from (a), $\chi_C = \|G^{-1}\|$ for some $n \times n$ nonsingular submatrix of C. Then we have

$$\chi_C = \|G^{-1}\| \leq \|[G^{-1}, R_C G^{-1}]\| = \|CG^{-1}\| \leq \bar{\chi}_C,$$

where the last inequality follows from (b). Finally for (h), first we have $C = [\tilde{C}; Q_C]$ for some matrix Q_C. Let F be any nonsingular submatrix of $\tilde{C}$. Then we have

$$\|\tilde{C}F^{-1}\| \leq \|[\tilde{C}F^{-1}, Q_C F^{-1}] = \|CF^{-1}\| \leq \bar{\chi}_C,$$

which prove the property. □

Khachiyan [5] showed that computation of χ_C is NP-hard. Vavasis and Ye [17] proved that if C is an integer matrix and its input bit length is L_C, both χ_C and $\bar{\chi}_C$ can be bounded above by $2^{O(L_C)}$. Other studies on the condition numbers include, for example, O'Leary [8] and Gonzaga and Lara [4].

4 Analysis of Asymptotic behavior of the layered least squares method

In this section, we analyze the asymptotic behavior of the layered least squares method. To this end, we first deal with the case where the number of the layers is two. Later, we handle general cases based on the result.

4.1 Two layers

Suppose that $A = [A_1; A_2] \in \Re^{m \times n}$ is a two-layer matrix, where $A_i \in \Re^{m_i \times n}$ $(i = 1, 2)$. Assume A is column full rank and define $r_1 = \text{rank}(A_1)$, $r_2 = n - r_1$. Let $T \in \Re^{m \times n}$ be a 2×2 BBLT matrix of A. That is, for some nonsingular matrix $G \in \Re^{n \times n}$,

$$T = AG^{-1} = \begin{bmatrix} T_{11} & O \\ T_{21} & T_{22} \end{bmatrix},$$

where $T_{ij} \in \Re^{m_i \times r_i}$, $1 \leq i, j \leq 2$. Recall T_{11} contains the identity matrix after appropriate row permutations. Therefore we have $T_{21} = VT_{11}$ for some matrix $V \in \Re^{m_2 \times m_1}$, which consists of column vectors of T_{21} and zero vectors. Then we have $\|V\| \leq \|T_{21}\|$, and with V, T is rewritten as

$$T = \begin{bmatrix} T_{11} & O \\ VT_{11} & T_{22} \end{bmatrix},$$

and for $D = [D_1; D_2]$,

$$DT = \begin{bmatrix} D_1 T_{11} & O \\ D_2 T_{21} & D_2 T_{22} \end{bmatrix} = \begin{bmatrix} B & O \\ WB & M \end{bmatrix},$$

where

$$B = D_1 T_{11},\ T_{21} = VT_{11}, W = D_2 V D_1^{-1},\ M = D_2 T_{22}.$$

Since $\|V\| \le \|T_{21}\|$ holds, it is easy to see the following Lemma hold.

Lemma 4.1 *The following relations hold:*

(1) $\|W\| \le \|D_2\| \|D_1^{-1}\| \|T_{21}\|$.

(2) $\|D_2^{-1} WB\| = \|T_{21}\|$.

(3) $\|WB\| = \|D_2 T_{21}\| \le \|D_2\| \|T_{21}\|$.

(4) $\|D_2^{-1} W\| \le \|D_1^{-1}\| \|T_{21}\|$.

Lemma 4.2 *The following inequality holds:*

$$\|T_{21}\| \le \bar{\chi}_A. \tag{10}$$

Proof. We can prove the lemma using the properties of $\bar{\chi}_A$ (Proposition 3.1) as follows:

$$\begin{aligned} \|T_{21}\| &\le \bar{\chi}_{[T_{11};T_{21}]} \text{ (from (e))} \\ &\le \bar{\chi}_T \text{ (from (f))} \\ &= \bar{\chi}_A \text{ (from (c))} \end{aligned}$$

□

Lemma 4.3 *Let* $T \in \Re^{m \times n}$ *be a BBLT matrix for* $A = [A_1; A_2]$ *and* $D = [D_1; D_2]$ *be a positive definite diagonal matrix. Then we have*

$$\begin{aligned} &(T^T D^2 T)^{-1} T^T D^2 \\ &= \begin{bmatrix} (T_{11}^T D_1^2 T_{11})^{-1} T_{11}^T D_1^2 & O \\ -(T_{22}^T D_2^2 T_{22})^{-1} T_{22}^T D_2^2 T_{21} (T_{11} D_1^2 T_{11})^{-1} T_{11}^T D_1^2 & (T_{22}^T D_2^2 T_{22})^{-1} T_{22}^T D_2^2 \end{bmatrix} \\ &+ \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \end{aligned} \tag{11}$$

where for $t = \|T_{21}\|$ *and* $\tau := \|D_1^{-1}\| \|D_2\|$,

$$\begin{aligned} \|C_{11}\| &\le \chi_{T_{11}}^2 t^2 \tau^2, \\ \|C_{12}\| &\le (\chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2) t \tau^2, \\ \|C_{21}\| &\le \chi_{T_{11}}^2 \chi_{T_{22}} t^3 \tau^2, \\ \|C_{22}\| &\le \chi_{T_{22}} (1 + \chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2) t^2 \tau^2. \end{aligned}$$

Proof. We have

$$T^T D^2 T := \begin{bmatrix} F & G \\ G^T & H \end{bmatrix} = \begin{bmatrix} B^T(I + W^T W)B & B^T W^T M \\ M^T W B & M^T M \end{bmatrix}.$$

From the inversion formula for nonsingular block matrices, it follows that

$$(T^T D^2 T)^{-1} = \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & \tilde{C} \end{bmatrix} = \begin{bmatrix} L^{-1} & -L^{-1}GH^{-1} \\ -H^{-1}G^T L^{-1} & H^{-1} + H^{-1}G^T L^{-1} G H^{-1} \end{bmatrix},$$

where $L = F - GH^{-1}G^T$.

Next transform $\tilde{A}$ of $(T^T D^2 T)^{-1}$ as follows:

$$\begin{aligned} \tilde{A} &= L^{-1} \\ &= (F - GH^{-1}G^T)^{-1} \\ &= (B^T(I + W^T W)B - B^T W^T M(M^T M)^{-1} M^T W B)^{-1}) \\ &= (B^T(I + W^T Q_M W)B)^{-1} \\ &= (B^T B + B^T W^T Q_M Q_M W B)^{-1} \\ &= (B^T B)^{-1} - (B^T B)^{-1} B^T W^T Q_M (I + Q_M W P_B W^T Q_M)^{-1} Q_M W B (B^T B)^{-1}, \end{aligned} \tag{12}$$

where we set $P_B = B(B^T B)^{-1} B^T$, $P_M = M(M^T M)^{-1} M^T$, $Q_M = I - P_M$ and the sixth equality follows from the Sherman-Morrison-Woodbury formula.

Also we have

$$\tilde{B} = -H^{-1} G^T L^{-1} = -(M^T M)^{-1} M^T W B \tilde{A}. \tag{13}$$

By using the equality (12), we further obtain the following:

$$\begin{aligned} \tilde{C} &= H^{-1} + H^{-1} G^T L^{-1} G H^{-1} \\ &= (M^T M^{-1}) + (M^T M)^{-1} M^T W P_B \times \\ &\quad (I - P_B W^T Q_M (I + Q_M W P_B W^T Q_M)^{-1} Q_M W P_B) P_B W^T M^T (M^T M)^{-1} \\ &= (M^T M)^{-1} + (M^T M)^{-1} M^T W P_B (I + P_B W^T Q_M W P_B)^{-1} P_B W^T M (M^T M)^{-1}. \end{aligned}$$

With the above equalities combined, we have

$$\begin{aligned} & (T D^2 T^T)^{-1} T^T D^2 \\ &= \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & \tilde{C} \end{bmatrix} \begin{bmatrix} B^T D_1 & B^T W^T D_2 \\ O & M^T D_2 \end{bmatrix} \\ &= \begin{bmatrix} \tilde{A} B^T D_1 & \tilde{A} B^T W^T D_2 + \tilde{B}^T M^T D_2 \\ \tilde{B} B^T D_1 & \tilde{B} B^T W^T D_2 + \tilde{C} M^T D_2 \end{bmatrix} \\ &= \begin{bmatrix} \tilde{A} B^T D_1 & \tilde{A} B^T D_1 D_1^{-1} W^T D_2 - \tilde{A} B^T D_1 D_1^{-1} W^T M (M^T M)^{-1} M^T D_2 \\ \tilde{B} B^T D_1 & \tilde{B} B^T D_1 D_1^{-1} W^T D_2 + \tilde{C} M^T D_2 \end{bmatrix} \\ &= \begin{bmatrix} \tilde{A} B^T D_1 & \tilde{A} B^T D_1 D_1^{-1} W^T Q_M D_2 \\ \tilde{B} B^T D_1 & \tilde{B} B^T D_1 D_1^{-1} W^T D_2 + \tilde{C} M^T D_2 \end{bmatrix} := \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}. \end{aligned}$$

In the following, we evaluate each element of the right-most matrix. First let us consider H_{11}. We have

$$\begin{aligned} H_{11} &= \tilde{A}B^T D_1 \\ &= (B^T B)^{-1} B^T D_1 \\ &\quad -(B^T B)^{-1} B^T W^T Q_M (I + Q_M W P_B W^T Q_M)^{-1} Q_M W B (B^T B)^{-1} B^T D_1 \\ &= (T_{11}^T D_1^2 T_{11})^{-1} T_{11}^T D_1^2 + C_{11}, \end{aligned}$$

where

$$C_{11} = -(B^T B)^{-1} B^T W^T Q_M (I + Q_M W P_B W^T Q_M)^{-1} Q_M W B (B^T B)^{-1} B^T D_1.$$

As for C_{11}, we can bound its norm as follows:

$$\begin{aligned} \|C_{11}\| &= \|(B^T B)^{-1} B^T W^T Q_M (I + Q_M W P_B W^T Q_M)^{-1} Q_M W B (B^T B)^{-1} B^T D_1\| \\ &\leq \|(B^T B) B^T D_1\| \|D_1^{-1}\| \|W\| \|Q_M\| \|(I + Q_M W P_B W^T Q_M)^{-1}\| \\ &\quad \times \|Q_M\| \|W B\| \|(B^T B)^{-1} B^T D_1\| \\ &\leq \chi_{T_{11}} \|D_1^{-1}\| \|D_1^{-1}\| \|D_2\| \|T_{21}\| \|D_2\| \|T_{21}\| \chi_{T_{11}} \\ &= \chi_{T_{11}}^2 t^2 \tau^2, \end{aligned}$$

where $t := \|T_{21}\|$, $\tau := \|D_1^{-1}\| D_2\|$. For the second inequality, we used Lemma 4.1 and the relation

$$\|(B^T B)^{-1} B^T D_1\| = \|(T_{11}^T D_1^2 T_{11})^{-1} T_{11}^T D_1^2\| \leq \chi_{T_{11}},$$

which follows from the definition of $\chi_{T_{11}}$. Afterall, the norm of H_{11} can be bounded in the following way.

$$\begin{aligned} \|H_{11}\| &= \|(T_{11}^T D_1^2 T_{11})^{-1} T_{11}^T D_1^2 + C_{11}\| \\ &\leq \|(T_{11}^T D_1^2 T_{11})^{-1} T_{11}^T D_1^2\| + \|C_{11}\| \\ &\leq \chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2, \end{aligned}$$

thus we have

$$\|H_{11}\| \leq \chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2. \tag{14}$$

Next we evaluate C_{12}. Observe that

$$C_{12} = H_{12} = \tilde{A}B^T D_1 D_1^{-1} W^T Q_M D_2 = H_{11} D_1^{-1} W^T Q_M D_2.$$

Using the inequality (14), we have

$$\begin{aligned} \|H_{12}\| &= \|H_{11} D_1^{-1} W^T Q_M D_2\| \\ &\leq (\chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2) \|D_1^{-1}\| \|D_1^{-1}\| \|D_2\| \|T_{21}\| \|D_2\| \\ &= (\chi_{T_{11}} + \chi_{T_{11}}^2 t^2 \tau^2) t \tau^2. \end{aligned}$$

Now we derive the bound on C_{21}. We have

$$\begin{aligned} H_{21} &= \tilde{B}B^TD_1 \\ &= -(M^TM)^{-1}M^TWB\tilde{A}B^TD_1 \\ &= -(M^TM)^{-1}M^TWBH_{11} \\ &= -(M^TM)^{-1}M^TWB\{(T_{11}^TD_1^2T_{11})^{-1}T_{11}^TD_1^2 + C_{11}\} \\ &= -(T_{22}^TD_2^2T_{22})^{-1}T_{22}^TD_2^2T_{21}(T_{11}^TD_1^2T_{11})^{-1}T_{11}^TD_1^2 + C_{21} \qquad (15) \\ C_{21} &= -(M^TM)^{-1}M^TWBC_{11}. \end{aligned}$$

The norm of the first term of the righthand side of (15) is bounded from above by

$$\|(T_{22}^TD_2^2T_{22})^{-1}T_{22}^TD_2^2\|\|T_{21}\|\|(T_{11}^TD_1^2T_{11})^{-1}T_{11}^TD_1^2\| \le \chi_{T_{11}}\chi_{T_{22}}t.$$

As to C_{21} we have

$$\begin{aligned} \|C_{21}\| &= \|(M^TM)^{-1}M^TD_2D_2^{-1}WBC_{11}\| \\ &\le \|(T_{22}D_2^2T_{22})^{-1}T_{22}D_2^2\|\|D_2^{-1}WB\|\|C_{11}\| \\ &\le \chi_{T_{22}}\|T_{21}\|\chi_{T_{11}}^2t^2\tau^2 \\ &= \chi_{T_{11}}^2\chi_{T_{22}}t^3\tau^2. \end{aligned}$$

By combining the two bounds, we obtain

$$\|H_{21}\| \le \chi_{T_{11}}\chi_{T_{22}}t(1+\chi_{T_{11}}t^2\tau^2).$$

Finally we deal with C_{22}. First we observe that

$$\begin{aligned} H_{22} &= \tilde{B}B^TD_1D_1^{-1}W^TD_2 + \tilde{C}M^TD_2 \\ &= H_{21}D_1^{-1}W^TD_2 + \tilde{C}M^TD_2 \\ &= \hat{C}_{22} + \tilde{C}M^TD_2, \qquad (16) \end{aligned}$$

where $\hat{C}_{22} = H_{21}D_1^{-1}W^TD$. We rewrite the second term of the righthand side of (16) using the relation

$$\tilde{C} = (M^TM)^{-1}\{I + M^TWP_B(I+P_BW^TQ_MWP_B)^{-1}P_BW^TM(M^TM)^{-1}\}.$$

and we have

$$\tilde{C}M^TD_2 = (T_{22}^TD_2^2T_{22})^{-1}T_{22}^TD_2^2 + \bar{C}_{22},$$

where

$$\bar{C}_{22} = (M^TM)^{-1}M^TWP_B(I+P_BW^TQ_MWP_B)^{-1}P_BW^TM(M^TM)^{-1}M^TD_2.$$

Therefore, we have

$$\begin{aligned} H_{22} &= \hat{C}_{22} + (T_{22}^TD_2^2T_{22})^{-1}T_{22}^TD_2^2 + \bar{C}_{22} \\ &= (T_{22}^TD_2^2T_{22})^{-1}T_{22}^TD_2^2 + C_{22}, \end{aligned}$$

where $C_{22} = \hat{C}_{22} + \bar{C}_{22}$. Now we find bounds on the norm of $\hat{C}_{22}$ and $\bar{C}_{22}$. Using Lemma 4.1, we can bound the norm of $\hat{C}_{22}$ as follows.

$$\begin{aligned}\|\hat{C}_{22}\| &= \|H_{21}D_1^{-1}W^T D_2\| \\ &\leq \|H_{21}\|\|D_1^{-1}\|\|W\|\|D_2\| \\ &\leq \chi_{T_{11}}\chi_{T_{22}}t^2\tau^2(1+\chi_{T_{11}}t^2\tau^2).\end{aligned}$$

The norm of $\bar{C}_{22}$ is bounded as follows.

$$\begin{aligned}\|\bar{C}_{22}\| &= \|(M^TM)^{-1}M^TD_2D_2^{-1}WP_B(I+P_BW^TQ_MWP_B)^{-1}P_BW^TP_BD_2\| \\ &\leq \|(M^TM)^{-1}M^TD_2\|\|D_2^{-1}W\|\|P_B\|\|(I+P_BW^TQ_MWP_B)^{-1}\| \\ &\quad \times\|P_B\|\|W^T\|\|P_M\|\|D_2\| \\ &\leq \chi_{T_{22}}t^2\tau^2\end{aligned}$$

Then from the bounds on the norm of $\hat{C}_{22}$ and $\bar{C}_{22}$, we derive

$$\begin{aligned}\|C_{22}\| &= \|\hat{C}_{22}+\bar{C}_{22}\| \\ &\leq \|\hat{C}_{22}\|+\|\bar{C}_{22}\| \\ &\leq \chi_{T_{22}}t^2\tau^2(1+\chi_{T_{11}}+\chi_{T_{11}}^2t^2\tau^2).\end{aligned}$$

Combining the results, the proof of the lemma is completed. □

4.2 General cases

In this subsection, we analyze the asymptotic behavior of the layered least squares method when the number of layers is not necessarily two. For $A = [A_1; \cdots; A_p]$, let

$$T = AG^{-1} = \begin{bmatrix} T_{11} & O & \cdots & \cdots & \cdots & O \\ T_{21} & T_{22} & O\cdots & \cdots & \cdots & O \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{p1} & T_{p2} & \cdots & \cdots & \cdots & T_{pp} \end{bmatrix}$$

be a BBLT matrix of A. Assume that the weight matrices $g \geq 1$, see (4). In the following, we use the following notations:

$$\begin{aligned}\tilde{T}_k &:= \begin{bmatrix} T_{kk} & O & \cdots & \cdots & \cdots & O \\ T_{k+1\ k} & T_{k+1\ k+1} & O\cdots & \cdots & \cdots & O \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{pk} & T_{p\ k+1} & \cdots & \cdots & \cdots & T_{pp} \end{bmatrix},\ k=1,\ldots,p, \\ \tilde{D}_k &= [D_k; D_{k+1}; \ldots; D_p],\ k=1,\ldots,p, \\ S_W^k &= (\tilde{T}_k^T\tilde{D}_k^2\tilde{T}_k)^{-1}\tilde{T}_k^T\tilde{D}_k^2,\ k=1,\ldots,p.\end{aligned}$$

Lemma 4.4 *The following relations hold:*

$$\begin{aligned}&\chi_{T_{kk}} \leq \bar{\chi}_A,\ k=1,\ldots,p, \\ &\chi_{\tilde{T}_k} \leq \bar{\chi}_A,\ k=1,\ldots,p, \\ &\|[T_{k+1,k};\cdots;T_{pk}]\| \leq \bar{\chi}_A,\ k=1,\ldots,p-1.\end{aligned}$$

Proof. From (c), (f), (g) and (h) of Proposition 3.1, we have

$$\begin{aligned}\chi_{T_{kk}} &\le \bar{\chi}_{T_{kk}} \text{ (from (g))}\\ &\le \bar{\chi}_{[T_{kk};T_{k+1\ k};\cdots;T_{pk}]} \text{ (from (h))}\\ &\le \bar{\chi}_T \text{ (from (f))}\\ &= \bar{\chi}_A \text{ (from (c))},\end{aligned}$$

which prove the first relation. The second relation is proved similarly since $\tilde{T}_k$ contains the identity matrix as a submatrix. Finally, we have

$$\|[T_{k+1,k};\cdots;T_{pk}]\| \le \|T\| \le \bar{\chi}_T = \bar{\chi}_A,$$

where the second inequality follows from (e) of Proposition 3.1. Thus the third relation of the lemma holds. □

Lemma 4.5 *Let* S_L^k *be the matrix defined in (7) and define*

$$\tilde{\tau} = g^{-1} = \max\{\|D_1^{-1}\|\|D_2\|, \ldots, \|D_{p-1}^{-1}\|\|D_p\|\}.$$

Then for $k = 1, \ldots, p-1$*, we have the following relations:*

$$S_W^k - S_L^k$$
$$= \begin{bmatrix} O & O \\ -(S_W^{k+1} - S_L^{k+1}) \begin{bmatrix} T_{k+1,k} \\ \vdots \\ T_{pk} \end{bmatrix} (T_{kk}^T D_k^2 T_{kk})^{-1} T_{kk}^T D_k^2 & S_W^{k+1} - S_L^{k+1} \end{bmatrix} + C_k, \tag{17}$$

$$\|C_k\| \le \bar{\chi}_A^6 \tilde{\tau}^2 + 2(1 + \bar{\chi}_A + \bar{\chi}_A^4 \tau^2)\bar{\chi}_A^3 \tau^2, \tag{18}$$

$$\|S_W^k - S_L^k\| \le (1+\bar{\chi}_A^2)\|S_W^{k+1} - S_L^{k+1}\| + \bar{\chi}_A^6 \tilde{\tau}^2 + 2(1+\bar{\chi}_A + \bar{\chi}_A^4 \tilde{\tau}^2)\bar{\chi}_A^3 \tilde{\tau}^2. \tag{19}$$

When $k = p$*, we have* $S_W^p - S_L^p = O$*.*

Proof. From Lemma 4.3, we have

$$S_W^k = \begin{bmatrix} (T_{kk}^T D_k^2 T_{kk})^{-1} T_{kk}^T D_k^2 & O \\ -S_W^{k+1} \begin{bmatrix} T_{k+1,k} \\ \vdots \\ T_{pk} \end{bmatrix} (T_{kk}^T D_k^2 T_{kk})^{-1} T_{kk}^T D_k^2 & S_W^{k+1} \end{bmatrix} + C_k.$$

The relation (17) follows from the above relation and the definition of S_L^k. It also follows from Lemma 4.3 that

$$C_k = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} \\ \tilde{C}_{21} & \tilde{C}_{22} \end{bmatrix},$$

where for $\tilde{t} := \|[T_{k+1,k}; \cdots ; T_{pk}]\|$ and $\hat{\tau} := \|D_k^{-1}\|\|\tilde{D}_{k+1}\| = \|D_k^{-1}\|\|D_{k+1}\|$,

$$
\begin{aligned}
\|\tilde{C}_{11}\| &\leq \chi_{T_{kk}}^2 \tilde{t}^2 \hat{\tau}^2, \\
\|\tilde{C}_{12}\| &\leq (\chi_{T_{kk}} + \chi_{T_{kk}}^2 \tilde{t}^2 \hat{\tau}^2) \tilde{t} \hat{\tau}^2, \\
\|C_{21}\| &\leq \chi_{T_{kk}}^2 \chi_{\tilde{T}_{k+1}} \tilde{t}^3 \hat{\tau}^2, \\
\|C_{22}\| &\leq \chi_{\tilde{T}_{k+1}} (1 + \chi_{T_{kk}} + \chi_{T_{kk}}^2 \tilde{t}^2 \hat{\tau}^2) \tilde{t}^2 \hat{\tau}^2.
\end{aligned}
$$

Using Lemma 4.4 and the fact that $\bar{\chi}_A \geq 1$, we obtain the second relation (18). From this relation and Lemma 4.4, it is easy to see that the third relation (19) holds. In the case where $k = p$, $S_W^p - S_L^p = O$ easily follows as there is only one layer. □

Now we are ready to prove the main theorem of this paper.

Theorem 4.1 *The difference between the solution x_{WLS}^* of the WLS and the solution x_{LLS}^* of LLS is given by*

$$x_{WLS}^* - x_{LLS}^* = G^{-1}(S_W^1 - S_L^1)b, \tag{20}$$

and its norm can be bounded as follows:

$$
\begin{aligned}
& \|G^{-1}(S_W^1 - S_L^1)b\| \\
\leq\ & \chi_A \{(1 + \bar{\chi}_A^2)^{p-1} - 1\} \{\bar{\chi}_A^4 \tau^2 + 2\bar{\chi}_A \tau^2 (1 + \bar{\chi}_A^2 + \bar{\chi}_A^4 \tau^2)\} \|b\|.
\end{aligned}
\tag{21}
$$

Also, the difference between the residual of WLS and that of LLS is given by

$$(b - Ax_{WLS}^*) - (b - Ax_{LLS}^*) = -T(S_W - S_L)b \tag{22}$$

and we have

$$
\begin{aligned}
& \|G^{-1}(S_W^1 - S_L^1)b\| \\
\leq\ & \{(1 + \bar{\chi}_A^2)^{p-1} - 1\} \{\bar{\chi}_A^5 \tau^2 + 2\bar{\chi}_A^2 \tau^2 (1 + \bar{\chi}_A^2 + \bar{\chi}_A^4 \tau^2)\} \|b\|.
\end{aligned}
\tag{23}
$$

Proof. For the relation (20), see Section 2.1. The bound (21) can be obtained by recursive uses of (19) in Lemma 4.5 and the fact $\|G^{-1}\| \leq \chi_A$ (see (a) and (d) of Proposition 3.1).

The relation (22) is discussed in Section 2.1. The bound (23) is derived from (19) and $\|T\| \leq \bar{\chi}_A$ from Lemma 4.4. □

References

[1] A. Ben-Tal and M.A. Teboulle: A geometric property of the least squares solution of linear equations. *Linear Algebra and Its Applications*, **139** (1990), 165–170.

[2] I.I. Dikin: On the speed of an iterative process (in Russian). *Upravlyaemye Sistemi*, **12** (1974), 54-60.

[3] A. Forsgren: On linear least-squares problems with diagonally dominant weight matrices. *SIAM Journal on Matrix Analysis and Applications*, **17** (1996), 763–788.

[4] C.C. Gonzaga and H.J. Lara: A note on properties of condition numbers: *Linear Algebra and Its Applications*, **261**, 269–273.

[5] L. Khachiyan: On the complexity of approximating extremal determinants in matrices. *Journal of Complexity*, **11** (1995), 138–153.

[6] N. Megiddo, S. Mizuno and T. Tsuchiya: A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, **82** (1998), 339–355.

[7] R.D.C. Monteiro and T. Tsuchiya: A variant of the Vavasis-Ye layered-step interior point algorithm for linear programming. *SIAM Journal on Optimization*, **13** (2002), 1054–1079.

[8] D.P. O'Leary: On bounds for scaled projections and pseudoinverses. *Linear Algebra and its Applications*, **132** (1990), 115–117.

[9] Takashi Sasakawa: *Optimal Design of DC Magnetic Shielding with Second-order Cone Programming* (In Japanese). PhD Thesis, Graduate School of Information Science and Technology, the University of Tokyo, 2005.

[10] T. Sasakawa and T. Tsuchiya: Optimal magnetic shield design with second-order cone programming. *SIAM Journal on Scientific Computing*, **24** (2003), 1930–1950.

[11] G.W. Stewart: On scaled projections and pseudoinvereses. *Linear Algebra and Its Applications*, **112** (1989), 189–193.

[12] M.J. Todd: A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, **38** (1990), 1006-1018.

[13] M.J. Todd, L. Tuncel, and Y. Ye: Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Mathematical Programming*, **90** (2001), 59–70.

[14] T. Tsuchiya: Global convergence property of the affine scaling method for primal degenerate linear programming problems. *Mathematics of Operations Research*, **17** (1992), 527–557.

[15] T. Tsuchiya: Layered least squares method — limit of the weighted least squares method (in Japanese). *Proceedings of the Institute of Statistical Mathematics*, **53** (2005), 391 – 404.

[16] T. Tsuchiya and T. Sasakawa: Robust optimization of magnetic shielding with second-order cone programming (in Japanese). *The Proceedings of The Institute of Statistical Mathematics*, **55** (2005), 297–315.

[17] S. Vavasis and Y. Ye: A primal-dual accelerated interior point method whose running time depends only on A. *Mathematical Programming*, **74** (1996), 79–120.

[18] S. J. Wright: *Primal-Dual Interior-Point Methods.* SIAM Publications, 1997.