Journal of Computational Mathematics, Vol.22, No.3, 2004, 427–436.

RELATIONSHIP BETWEEN THE STIFFLY WEIGHTED PSEUDOINVERSE AND MULTI-LEVEL CONSTRAINED PSEUDOINVERSE *1)

Mu-sheng Wei
(*Department of Mathematics, East China Normal University, Shanghai 200062, China*)

Abstract

It is known that for a given matrix A of rank r, and a set $\mathcal{D}$ of positive diagonal matrices, $\sup_{W\in\mathcal{D}} \|(W^{\frac{1}{2}}A)^{\dagger}W^{\frac{1}{2}}\|_2 = (\min_i \sigma_+(A^{(i)}))^{-1}$, in which $(A^{(i)})$ is a submatrix of A formed with $r = (\text{rank}(A))$ rows of A, such that $(A^{(i)})$ has full row rank r. In many practical applications this value is too large to be used.

In this paper we consider the case that both A and $W(\in \mathcal{D})$ are fixed with W severely stiff. We show that in this case the weighted pseudoinverse $(W^{\frac{1}{2}}A)^{\dagger}W^{\frac{1}{2}}$ is close to a multi-level constrained weighted pseudoinverse therefore $\|(W^{\frac{1}{2}}A)^{\dagger}W^{\frac{1}{2}}\|_2$ is uniformly bounded. We also prove that in this case the solution set the stiffly weighted least squares problem is close to that of corresponding multi-level constrained least squares problem.

Mathematics subject classification: 15A09, 15A12, 65F20.
Key words: Weighted least squares, Stiff, Multi-Level constrained pseudoinverse.

1. Introduction

In this paper we are concerned with the stiffly weighted least squares (stiffly WLS) problem

$$\min_x \|W^{\frac{1}{2}}(Ax-b)\|_2 = \min_x \|D(Ax-b)\|_2 \tag{1}$$

and related weighted pseudoinverse $A_W^{\dagger} \equiv (W^{\frac{1}{2}}A)^{\dagger}W^{\frac{1}{2}}$, where $\|\cdot\| \equiv \|\cdot\|_2$ denotes the Euclidean vector norm or subordinate matrix norm, $A \in \mathbf{C}^{m\times n}, b \in \mathbf{C}^m$ are known coefficient matrix and observation vector, respectively,

$$D = \text{diag}(d_1, d_2, \cdots, d_m) = \text{diag}(w_1^{\frac{1}{2}}, w_2^{\frac{1}{2}}, \cdots, w_m^{\frac{1}{2}}) = W^{\frac{1}{2}} \tag{2}$$

is the weight matrix. WLS problem Eq. (1) with extremely ill-conditioned weight matrix W (in this case Björck [3] called W stiff weight matrix), where the scalar factors $w_1, \cdots, w_m$ vary widely in size arise, e.g., in electronic network, certain classes of finite element problems, interior-point method for constrained optimization (e.g., see [8, 15]), and for solving the equality constrained least squares problem by the method of weighting [16, 1, 14].

In the case that W is severely stiff, it is not at all apparent that an accurate numerical solution to Eq. (1) is possible, since ill-conditioning in W presumably means extreme sensitivity to roundoff errors, because in standard numerical analysis, error bounds of the solutions to Eq. (1) have a weighted condition number $\kappa(W^{\frac{1}{2}}A) = \|W^{\frac{1}{2}}A\|\|(W^{\frac{1}{2}}A)^{\dagger}\|$ as a factor so that when W becomes ill-conditioned the condition number would become unbounded.

* Received January 7, 2002; final revised April 17, 2003.
1) The author was supported by NSFC under grant 10371044 and Shanghai Priority Academic Discipline Foundation, Shanghai, China.

On the other hand, one can define a new condition number $\kappa = \|A\|\|A_W^\dagger\|$. If $\|A_W^\dagger\|$ is uniformly bounded, then κ would be uniformly bounded.

Stewart [13] obtained an upper bound of scaled projections when $A \in R^{m\times n}$ has full column rank and weight matrices W range over a set $\mathcal{D}$ of positive diagonal matrices. Liu and Xu [10] then proved that this upper bound for scaled projection is indeed the supremum. Wei [19], Forsgren [6], Wei [20] respectively have obtained the supremum of weighted pseudoinverses when weight matrices W range over $\mathcal{D}$, or a set $\mathcal{P}$ of real symmetric diagonal dominant semi-positive matrices. Forsgren [6] and Wei [20] have also extended the results to constrained weighted pseudoinverses. For more detailed description on this topic, we refer to [21].

In practical applications, the supremum [19, 20]

$$\sup_{W\in\mathcal{D}} \|A_W^\dagger\| = \frac{1}{\min \sigma_+(A^{(i)})} \tag{3}$$

sometimes may be too large to be of practical usefulness. For instance, suppose

$$A = \begin{pmatrix} 1 & 0 \\ \delta & 0 \\ 0 & 1 \end{pmatrix}, \; W_0 = \mathrm{diag}(w_1, w_1, w_3),$$

where $w_1 > w_3 > 0$ are arbitrary, and $0 < \delta \ll 1$. Then

$$\|A_{W_0}^\dagger\| = 1 \quad \text{and} \quad \sup_{W\in\mathcal{D}} \|A_W^\dagger\| = \frac{1}{\delta} \gg 1.$$

This example rises a question: if the weight matrix W is given and is very ill-conditioned, does exist an upper bound of $\|A_W^\dagger\|$ which is of moderate size?

In this paper we will study the above question. Without loss of generality, we make the following notation and assumptions for A and W.

Assumption 1.1. *The matrices A and W in Eq. (1) satisfy the following conditions: $\|A(i,:)\| \equiv \|(a_{i1}, a_{i2}, \cdots, a_{in})\|$ have the same order for $i = 1, \cdots, m$, $w_1 > w_2 > \cdots > w_k > 0$, $m_1 + m_2 + \cdots + m_k = m$, and we denote*

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_k \end{pmatrix} \begin{matrix} m_1 \\ \vdots \\ m_k \end{matrix}, \; C_j = \begin{pmatrix} A_1 \\ \vdots \\ A_j \end{pmatrix}, \quad j = 1, \cdots, k, \tag{4}$$

$$\begin{aligned} & W = \mathrm{diag}(w_1 I_{m_1}, w_2 I_{m_2}, \cdots, w_k I_{m_k}), \\ & 0 < \epsilon_{ij} \equiv \tfrac{w_i}{w_j} \ll 1, \; \textit{for } 1 \le j < i \le k \textit{ so } \epsilon = \max_{1\le j<k}\{\epsilon_{j+1,j}\} \ll 1. \end{aligned} \tag{5}$$

We also set

$$P_0 = I_n, \; P_j = I - C_j^\dagger C_j, \; \mathrm{rank}(C_j) = r_j, \; j = 1, \cdots, k. \tag{6}$$

Vavasis and Ye [17] studied interior-point method for solving linear programming problem, in which the matrices A and W basically satisfy Assumption 1.1.

The paper is organized as follows. In §2 we will derive several equivalent formulas of the stiffly weighted pseudoinverse; in §3 we will derive the multi-level constrained pseudoinverse and corresponding multi-level constrained least squares (MCLS) problem; in §4 we will prove that the stiffly weighted pseudoinverse is indeed close to the multi-level constrained pseudoinverse therefore is uniformly bounded; in §5 we will deduce upper bounds of difference of the solutions between of the stiffly WLS problem and the MCLS problem; finally in §6 we will conclude the paper with some remarks.

2. Equivalent Formulas of the Stiffly Weighted Pseudoinverse

In this section we will derive several equivalent formulas of the stiffly weighted pseudoinverse. We first provide some preliminary results which are necessary for our further discussion.

Lemma 2.1. *[20, 24] For given matrices $A \in C^{m\times n}$ and $W \in \mathcal{D}$, define*

$$A_W = WA(WA)^{\dagger}A, \tag{7}$$

then

$$A_W^{\dagger} = (W^{\frac{1}{2}}A)^{\dagger}W^{\frac{1}{2}}, \tag{8}$$

and

$$A_W^{\dagger}A_W = A_W^{\dagger}A = A^{\dagger}A. \tag{9}$$

If $A = \begin{pmatrix} L \\ K \end{pmatrix}$, then we also have $\operatorname{rank}(A) = \operatorname{rank}(L) + \operatorname{rank}(KP)$ *and*

$$A_W^{\dagger}A_W = A_W^{\dagger}A = A^{\dagger}A = L^{\dagger}L + (KP)^{\dagger}KP$$

with $P = I - L^{\dagger}L$.

Lemma 2.2. *[7] Suppose that $D, E \in \mathbf{C}^{m\times n}$ and rank(D)=rank(E). Then*

$$\begin{aligned} &\|DD^{\dagger} - EE^{\dagger}\| \le \min\{\|(D-E)D^{\dagger}\|, \|(D-E)E^{\dagger}\|, 1\}, \\ &\|DD^{+} - EE^{\dagger}\| \le \min\{\|D^{\dagger}(D-E)\|, \|E^{\dagger}(D-E)\|, 1\}. \end{aligned} \tag{10}$$

Lemma 2.3. *Under the notation of Assumption 1.1,*

$$\begin{aligned} &(A_jP_{j-1})^{\dagger}A_jP_{j-1} = C_j^{\dagger}C_j - C_{j-1}^{\dagger}C_{j-1}, \\ &\operatorname{rank}(A_jP_{j-1}) = \operatorname{rank}(C_j) - \operatorname{rank}(C_{j-1}) = r_j - r_{j-1} \end{aligned} \tag{11}$$

for $j = 2, \cdots, k$. Denote $(A_jP_{j-1})^H = Q_jR_j$ the unitary decomposition of $(A_jP_{j-1})^H$ (A_j^H is the conjugate transpose of the matrix A_j), where $Q_j^HQ_j = I_{r_j - r_{j-1}}$ and R_j has full row rank $r_j - r_{j-1}$. Then for $j = 1, \cdots, k$,

$$(Q_1, \cdots, Q_j)^H(Q_1, \cdots, Q_j) = I_{r_j},\ C_j^{\dagger}C_j = \sum_{l=1}^{j} Q_lQ_l^H, \tag{12}$$

$$A_jP_{j-1} = A_jQ_jQ_j^H,\ (A_jP_{j-1})^{\dagger} = Q_j(A_jQ_j)^{\dagger}. \tag{13}$$

Proof. The identities in Eq. (11) are just mentioned in Lemma 2.1. For $j = 1$, Eqs. (12)-(13) are true. Suppose that for $1 \le j \le t < k$, Eqs.(12)-(13) are true. Then for $j = t+1$, from Eq. (11) and the definition of $Q_{t+1}, (Q_1, \cdots, Q_t)^HQ_{t+1} = 0$, so

$$(Q_1, \cdots, Q_{t+1})^H(Q_1, \cdots, Q_{t+1}) = I_{r_{t+1}},\ C_{t+1}^{\dagger}C_{t+1} = \sum_{l=1}^{t+1} Q_lQ_l^H,$$

$$A_{t+1}Q_{t+1}Q_{t+1}^H = A_{t+1}(C_t^{\dagger}C_t + P_t)Q_{t+1}Q_{t+1}^H = A_{t+1}P_tQ_{t+1}Q_{t+1}^H = A_{t+1}P_t,$$

and by induction hypothesis, we prove Eqs. (12)-(13).

Lemma 2.4. *[11, 24] Let $A \in \mathbf{C}^{m\times n}$ and $\hat{A} = A + \delta A \in \mathbf{C}^{m\times n}$. Then we have the following results.*

1. *If* $\|\delta A\| \cdot \|A^\dagger\| < 1$, *then* $\mathrm{rank}(\hat{A}) \geq \mathrm{rank}(A)$.
2. *If* $\|\delta A\| \cdot \|A^\dagger\| < 1$, *and* $\mathrm{rank}(\hat{A}) > \mathrm{rank}(A)$, *then* $\|\hat{A}^\dagger\| \geq \frac{1}{\|\delta A\|}$.
3. *If* $\|\delta A\| \cdot \|A^\dagger\| < 1$, *and* $\mathrm{rank}(\hat{A}) = \mathrm{rank}(A)$, *then*

$$\frac{\|A^\dagger\|}{1+\|\delta A\| \cdot \|A^\dagger\|} \leq \|\hat{A}^\dagger\| \leq \frac{\|A^\dagger\|}{1-\|\delta A\| \cdot \|A^\dagger\|}.$$

So $\|\hat{A}^\dagger\|$ *is bounded for all small perturbation* δA *with*

$$\|\delta A\| \cdot \|A^\dagger\| \leq \eta < 1, \text{ if and only if } \mathrm{rank}(\hat{A}) = \mathrm{rank}(A),$$

where $0 \leq \eta < 1$ *is a constant.*

We now present the main result of this section.

Theorem 2.1. *Under the notation in Assumption 1.1,*

$$\begin{aligned} A_W &= B_\epsilon B_\epsilon^\dagger A = A_\epsilon A_\epsilon^\dagger A = (B_\epsilon^\dagger)^H B_\epsilon^H B_1 Q^H, \\ A_W^\dagger &= Q(B_\epsilon^H B_1)^{-1} B_\epsilon^H, \end{aligned}$$

$$B_\epsilon = \begin{pmatrix} A_1Q_1 & 0 & \cdots & 0 \\ \epsilon_{21}A_2Q_1 & A_2Q_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \epsilon_{k1}A_kQ_1 & \epsilon_{k2}A_kQ_2 & \cdots & A_kQ_k \end{pmatrix}, \tag{14}$$

$$A_\epsilon = \begin{pmatrix} A_1 \\ \epsilon_{21}A_2Q_1Q_1^H + A_2Q_2Q_2^H \\ \vdots \\ \epsilon_{k1}A_kQ_1Q_1^H + \epsilon_{k2}Q_2Q_2^H + \cdots + A_kQ_kQ_k^H \end{pmatrix},$$

in which B_ϵ *has full column rank* $r_k = \mathrm{rank}(A) = \mathrm{rank}(A_\epsilon)$, *and* B_1 *is obtained from* B_ϵ *by replacing all* ϵ_{ij} *in* B_ϵ *with ones.*

Proof. By applying Lemmas 2.1-2.3, we obtain

$$\begin{aligned} WA &= \begin{pmatrix} w_1A_1 \\ w_2A_2 \\ \vdots \\ w_kA_k \end{pmatrix} = \begin{pmatrix} w_1A_1Q_1Q_1^H \\ w_2A_2(Q_1Q_1^H + Q_2Q_2^H) \\ \vdots \\ w_kA_k(Q_1Q_1^H + Q_2Q_2^H + \cdots + Q_kQ_k^H) \end{pmatrix} \\ &= \begin{pmatrix} w_1A_1Q_1 & 0 & \cdots & 0 \\ w_2A_2Q_1 & w_2A_2Q_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ w_kA_kQ_1 & w_kA_kQ_2 & \cdots & w_kA_kQ_k \end{pmatrix} \widetilde{W}^{-1}\widetilde{W}Q^H \\ &= B_\epsilon(\widetilde{W}Q^H), \end{aligned}$$

where $\widetilde{W} = \mathrm{diag}(w_1I_{r_1}, w_2I_{r_2-r_1}, \cdots, w_kI_{r_k-r_{k-1}})$, and both B_ϵ and $Q\widetilde{W}$ have full column rank r_k. Then

$$\begin{aligned} A_W = &\ WA(WA)^\dagger A = B_\epsilon(Q\widetilde{W})^H(B_\epsilon(Q\widetilde{W})^H)^\dagger A \\ = &\ B_\epsilon(Q\widetilde{W})^H(Q\widetilde{W})^{+H}B_\epsilon^\dagger A = B_\epsilon B_\epsilon^\dagger A \\ = &\ B_\epsilon Q^HQB_\epsilon^\dagger A = (B_\epsilon Q^H)(B_\epsilon Q^H)^\dagger A = A_\epsilon A_\epsilon^\dagger A. \end{aligned}$$

Similarly, we have $A = B_1 Q^H$, and it is obvious that

$$B_\epsilon B_\epsilon^\dagger A = (B_\epsilon B_\epsilon^\dagger)^H B_1 Q^H = (B_\epsilon^\dagger)^H (B_\epsilon^H B_1) Q^H,$$

in which both $(B_\epsilon^\dagger)^H$ and Q^H have full row rank, $B_\epsilon^H B_1$ is nonsingular. Therefore we obtain $A_W^\dagger = Q(B_\epsilon^H B_1)^{-1} B_\epsilon^H$, completing the proof of the theorem.

3. Multi-level Constrained Pseudoinverse and MCLS Problem

In this section we first introduce a multi-level constrained pseudoinverse $A_C^\dagger$ which is independent of W, and corresponding MCLS problem. The MCLS problem was first studied by Vavasis and Ye [17] in which the MCLS problem is called the layered least squares problem (LLS).

Theorem 3.1. *Under the notation of Assumption 1.1, define*

$$A_C = \begin{pmatrix} A_1 \\ A_2(A_2P_1)^\dagger A_2 \\ \vdots \\ A_k(A_kP_{k-1})^\dagger A_k \end{pmatrix} = B_0 B_0^\dagger A, \tag{15}$$

then

$$\begin{aligned} A_C^\dagger = \;& (G_k G_{k-1} \cdots G_2 (A_1P_0)^\dagger, G_k G_{k-1} \cdots G_3 (A_2P_1)^\dagger, \\ & \cdots, G_k (A_{k-1}P_{k-2})^\dagger, (A_kP_{k-1})^\dagger) = Q(B_0^H B_1)^{-1} B_0^H \end{aligned} \tag{16}$$

in which B_0 is obtained by setting all ϵ_{ij} in B_ϵ with zeros, and

$$G_j = I_n - (A_jP_{j-1})^\dagger A_j, \quad j = 2, \cdots, k. \tag{17}$$

Proof. Denote the matrix of the middle side in Eq. (16) by F. We need to prove $F = A_C^\dagger$.
Step 1. We first prove by induction that for $l = 2, \cdots, k$,

$$\begin{aligned} C_l^\dagger C_l = \;& G_l \cdots G_2 (A_1P_0)^\dagger A_1 + G_l \cdots G_3 (A_2P_1)^\dagger A_2 \\ & + \cdots + G_l (A_{l-1}P_{l-2})^\dagger A_{l-1} + (A_lP_{l-1})^\dagger A_l. \end{aligned} \tag{18}$$

When $l = 1$, Eq. (18) is trivially true. Suppose that the identity in Eq. (18) holds for $1 \le l \le t < k$. Then by applying Lemma 2.3, we have

$$\begin{aligned} & G_{t+1}G_t \cdots G_2(A_1P_0)^\dagger A_1 + G_{t+1}G_t \cdots G_3(A_2P_1)^\dagger A_2 + \cdots \\ & \quad + G_{t+1}(A_tP_{t-1})^\dagger A_t + (A_{t+1}P_t)^\dagger A_{t+1} \\ & \quad = G_{t+1}[G_t \cdots G_2(A_1P_0)^\dagger A_1 + G_t \cdots G_3(A_2P_1)^\dagger A_2 + \cdots \\ & \quad + (A_tP_{t-1})^\dagger A_t] + (A_{t+1}P_t)^\dagger A_{t+1} \\ & \quad = G_{t+1}C_t^\dagger C_t + (A_{t+1}P_t)^\dagger A_{t+1} \\ & \quad = C_t^\dagger C_t - (A_{t+1}P_t)^\dagger A_{t+1} C_t^\dagger C_t + (A_{t+1}P_t)^\dagger A_{t+1} \\ & \quad = C_t^\dagger C_t + (A_{t+1}P_t)^\dagger A_{t+1} P_t = C_{t+1}^\dagger C_{t+1}. \end{aligned}$$

So the identity in Eq. (18) also holds for $l = t + 1$. Then by the induction process Eq. (18) holds for all $l = 1, \cdots, k$, and finally we obtain

$$FA_C = (FA_C)^H = C_k^\dagger C_k = A^\dagger A. \tag{19}$$

Step 2. We now prove that

$$A_C F = \operatorname{diag}(A_1P_0(A_1P_0)^\dagger, A_2P_1(A_2P_1)^\dagger, \cdots, A_kP_{k-1}(A_kP_{k-1})^\dagger). \tag{20}$$

Because $A_i = A_i(Q_1 Q_1^H + \cdots + Q_i Q_i^H),\ (A_j P_{j-1})^\dagger = Q_j (A_j Q_j)^\dagger$, so

$$A_i(A_iP_{i-1})^\dagger A_i G_l = A_i(A_iP_{i-1})^\dagger A_i(I - Q_j(A_jQ_j)^\dagger A_j) = \begin{cases} A_i(A_iP_{i-1})^\dagger A_i, & i < j, \\ 0, & i = j. \end{cases}$$

With the above observation, we have that for $j \leq k-1$,

$$[A_i(A_iP_{i-1})^\dagger A_i][G_k \cdots G_{j+1}(A_jP_{j-1})^\dagger] = \begin{cases} A_i(A_iP_{i-1})^\dagger, & \text{for } i = j, \\ 0, & \text{for } i \neq j. \end{cases}$$

$$[A_i(A_iP_{i-1})^\dagger A_i](A_kP_{k-1})^\dagger = \begin{cases} 0, & i < k, \\ A_k(A_kP_{k-1})^\dagger, & i = k. \end{cases}$$

therefore

$$A_C F = \text{diag}(A_1(A_1P_0)^\dagger, A_2(A_2P_1)^\dagger, \cdots, A_k(A_kP_{k-1})^\dagger) = (A_CF)^H.$$

Step 3. By applying the identity in Eq. (20) we can easily verify

$$A_CFA_C = (A_CF)A_C = A_C, \quad FA_CF = F(A_CF) = F.$$

Then F satisfies all the four conditions as the unique pseudoinverse of A_C[2]. So $F = A_C^\dagger$. That $A_C^\dagger = Q(B_0^H B_1)^{-1} B_0^H$ results from Theorem 2.1.

The multi-level constrained pseudoinverse $A_C^\dagger$ can be obtained from the following multi-level constrained least squares (MCLS) problem: Let $A_i \in C^{m_i \times n}, b_i \in C^{m_i}$ be given. Define the following sets S_i:

$$S_1 = C^n,\ S_i = \{x \in S_{i-1} : \|A_ix - b_i\| = \min_{y \in S_{i-1}} \|A_iy - b_i\|\}, \tag{21}$$

for $i = 2, \cdots, k$, and let $x \in C^n$ sequentially satisfies the following conditions

$$x \in S_1, x \in S_2, \cdots, x \in S_k. \tag{22}$$

Then we have the following result.

Theorem 3.2. *Suppose that $A \in C^{m\times n}$ satisfies the notation in Assumption 1.1, and A_C is defined in Theorem 3.1. Then any solution $x \in C^n$ of the MCLS problem Eqs. (21)-(22) has the following form:*

$$x = x_k + P_kz_k = A_C^\dagger b + P_kz_k, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}, \tag{23}$$

in which $z_k \in C^n$ is an arbitrary vector.

Proof. We will prove that

$$x = x_l + P_lz_l,$$

with

$$x_l = (G_l \cdots G_2(A_1P_0)^\dagger, \cdots, G_l(A_{l-1}P_{l-2})^\dagger, (A_lP_{l-1})^\dagger) \begin{pmatrix} b_1 \\ \vdots \\ b_l \end{pmatrix} \tag{24}$$

for $l = 1, 2, \cdots, k$. When $l = 1$, it is obvious that

$$x = A_1^\dagger b_1 + (I - A_1^\dagger A_1)z_1 = x_1 + P_1z_1$$

with $z_1 \in C^n$. Suppose that for $1 \le l \le t < k$ the formulas in Eq. (24) is true. Then from $x \in S_{t+1}$, z_t should satisfy

$$\|A_{t+1}(x_t + P_t z_t) - b_{t+1}\| = \min_{z\in C^n} \|A_{t+1}(x_t + P_t z) - b_{t+1}\|,$$

therefore

$$z_t = (A_{t+1}P_t)^\dagger(b_{t+1} - A_{t+1}x_t) + (I - (A_{t+1}P_t)^\dagger A_{t+1}P_t)z_{t+1},$$

and by applying Lemma 2.3,

$$\begin{aligned} x = &\ x_t + P_t z_t = G_{t+1}x_t + (A_{t+1}P_t)^\dagger b_{t+1} + P_{t+1}z_{t+1} \\ = &\ x_{t+1} + P_{t+1}z_{t+1}. \end{aligned}$$

So for $l = t+1$ the assertion is also true. By induction hypothesis Eq. (24) holds for $l = 1, \cdots, k$, and $x = x_k + P_k z_k = A_C^\dagger b + P_k z_k$.

4. Differences between $A_W - A_C$ and $A_W^\dagger - A_C^\dagger$

We now prove that when A and W satisfy Assumption 1.1, then $A_W^\dagger$ is close to $A_C^\dagger$.

Theorem 4.1. *Under the notation and conditions of Assumption 1.1,*

$$\|A_W - A_C\| \le \frac{\epsilon}{1-\epsilon}\|A\| \max_{1\le j<i\le k} \|A_i(A_jP_{j-1})^\dagger\| \equiv e_\epsilon. \tag{25}$$

in which $\epsilon = \max_{1\le j<k} \frac{w_{j+1}}{w_j}$. *Therefore when* $e_\epsilon\|A_C^\dagger\| < 1$

$$\|A_W^\dagger\| \le \frac{\|A_C^\dagger\|}{1 - e_\epsilon\|A_C^\dagger\|}, \ \|A_W^\dagger - A_C^\dagger\| \le \sqrt{2}e_\epsilon\|A_C^\dagger\|\|A_W^\dagger\|. \tag{26}$$

Proof. From Eqs. (14)-(15), and Lemma 2.2,

$$\begin{aligned} \|A_W - A_C\| = &\ \|B_\epsilon B_\epsilon^\dagger A - B_0 B_0^\dagger A\| = \|(B_\epsilon B_\epsilon^\dagger - B_0 B_0^\dagger)A\| \\ \le &\ \|(B_\epsilon - B_0)B_0^\dagger\| \cdot \|A\|. \end{aligned}$$

Now

$$\begin{aligned} &(B_\epsilon - B_0)B_0^\dagger \\ &= \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \epsilon_{21}A_2Q_1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \epsilon_{k1}A_kQ_1 & \epsilon_{k2}A_kQ_2 & \cdots & \epsilon_{k,k-1}A_kQ_{k-1} & 0 \end{pmatrix} \mathrm{diag}((A_1Q_1)^\dagger, (A_2Q_2)^\dagger, \cdots, (A_kQ_k)^\dagger) \\ &= \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \epsilon_{21}A_2Q_1(A_1Q_1)^\dagger & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ \epsilon_{k1}A_kQ_1(A_1Q_1)^\dagger & \epsilon_{k2}A_kQ_2(A_2Q_2)^\dagger & & \epsilon_{k,k-1}A_kQ_{k-1}(A_{k-1}Q_{k-1})^\dagger & & 0 \end{pmatrix}, \end{aligned}$$

therefore

$$\begin{aligned} &\|(B_\epsilon - B_0)B_0^\dagger\| \\ &\quad \le \|\mathrm{diag}(\epsilon_{21}A_2Q_1(A_1Q_1)^\dagger, \cdots, \epsilon_{k,k-1}A_kQ_{k-1}(A_{k-1}Q_{k-1})^\dagger\| \\ &\quad +\|\mathrm{diag}(\epsilon_{31}A_3Q_1(A_1Q_1)^\dagger, \cdots, \epsilon_{k,k-2}A_kQ_{k-2}(A_{k-2}Q_{k-2})^\dagger\| \\ &\quad + \cdots + \epsilon_{k1}\|A_kQ_1(A_1Q_1)^\dagger\| \\ &\quad \le \epsilon \max_{1\le j<k} \|A_{j+1}Q_j(A_jQ_j)^\dagger\| + \epsilon^2 \max_{1\le j<k-1} \|A_{j+2}Q_j(A_jQ_j)^\dagger\| \\ &\quad + \cdots + \epsilon^{k-1}\|A_kQ_1(A_1Q_1)^\dagger\| \\ &\quad \le (\epsilon + \epsilon^2 + \cdots + \epsilon^{k-1}) \max_{1\le j<i\le k} \|A_iQ_j(A_jQ_j)^\dagger\| \\ &\quad \le \tfrac{\epsilon}{1-\epsilon} \max_{1\le j<i\le k} \|A_i(A_jP_{j-1})^\dagger\|, \end{aligned}$$

in which we have applying Lemma 2.3 and used the inequality

$$\epsilon_{ij} = \epsilon_{i,i-1} \cdots \epsilon_{j+1,j} \le \epsilon^{i-j}$$

for $1 \le j < i \le k$. Then the inequality in Eq. (25) holds. Also notice that

$$\mathrm{rank}(A_W) = \mathrm{rank}(A_C) = \mathrm{rank}(A) = r_k,$$

in the case $e_\epsilon \|A_C^\dagger\| < 1$, we can apply Lemma 2.4 to obtain

$$\|A_W^\dagger\| \le \frac{\|A_C^\dagger\|}{1 - \|A_W - A_C\| \cdot \|A_C^\dagger\|} \le \frac{\|A_C^\dagger\|}{1 - e_\epsilon \|A_C^\dagger\|}.$$

Notice that from Lemma 2.1, $A_W^\dagger A_W = A^\dagger A = A_C^\dagger A_C$, we then have the following identity:

$$\begin{aligned} A_W^\dagger - A_C^\dagger &= -A_W^\dagger (A_W - A_C) A_C^\dagger + A_W^\dagger (I - A_C A_C^\dagger) - (I - A_W^\dagger A_W) A_C^\dagger \\ &= -A_W^\dagger (A_W - A_C) A_C^\dagger + A_W^\dagger (I - A_C A_C^\dagger). \end{aligned} \tag{27}$$

Therefore, for any $z \in C^n$,

$$\begin{aligned} \|z^H (A_W^\dagger - A_C^\dagger)\|^2 &= \|z^H A_W^\dagger (A_W - A_C) A_C^\dagger\|^2 + \|z^H A_W^\dagger (I - A_C A_C^\dagger)\|^2 \\ &\le [\|z\| \|A_W^\dagger\| \|A_W - A_C\| \|A_C^\dagger\|]^2 + [\|z\| \|A_W^\dagger\| \|A_W A_W^\dagger (I - A_C A_C^\dagger)\|]^2 \\ &\le 2[e_\epsilon \|z\| \|A_W^\dagger\| \|A_C^\dagger\|]^2, \end{aligned}$$

from which we prove the second inequality of Eq. (26).

When the matrix A has some special properties, such as A has full row rank, or range(A_j^H) for $j = 1, \cdots, k$ are mutually orthogonal, then $A_W = A_C = A$ and $A_W^\dagger = A_C^\dagger = A^\dagger$, as mentioned in the following corollary.

Corollary 4.1. *Under the notation and conditions of Assumption 1.1, if further more, A has full row rank , or* range(A_j^H) *for $j = 1, \cdots, k$ are mutually orthogonal, then*

$$A_W = A_C = A \text{ and } A_W^\dagger = A_C^\dagger = A^\dagger.$$

Proof. When A has full row rank m, both B_ϵ and B_0 in Eqs. (14)-(15) are nonsingular, and

$$A_W = A_C = A \text{ and } A_W^\dagger = A_C^\dagger = A^\dagger.$$

When range(A_j^H) for $j = 1, \cdots, k$ are mutually orthogonal, then

$$A_j P_{j-1} = A_j = A_j Q_j Q_j^H, \text{ and } A_i Q_j = 0$$

for $i, j = 1, \cdots, k$ and $i \ne j$. Therefore from the formulas of A_W and A_C in Eqs. (14)-(15), we immediately have

$$A_W = A_C = A \text{ and } A_W^\dagger = A_C^\dagger = A^\dagger.$$

5. Difference between the Solution Sets of the Stiffly WLS and WCLS Problems

In this section we provide upper bounds of the difference between the solution sets of the stiffly WLS problem Eq. (1) and the WCLS problem Eqs. (21)-(22).

Theorem 5.1. *Consider the stiffly WLS problem Eq. (1) and the WCLS problem Eqs. (21)-(22), in which the matrices A and W satisfy the conditions mentioned in Theorem 4.1. Then for any solution x_W to the stiffly WLS problem Eq. (1), there exists a solution x_C to the MCLS problem Eqs. (21)-(22), such that*

$$\|x_W - x_C\| \le e_\epsilon \frac{\|A_C^\dagger\|}{1 - e_\epsilon \|A_C^\dagger\|}(\|x_{CLS}\| + \|A_C^\dagger\| \cdot \|r_C\|), \tag{28}$$

in which $x_{CLS} = A_C^\dagger b$ and $r_C = b - A_C x_{CLS}$; and vice versa.

Proof. Notice that from Lemma 2.1 and Theorem 3.1, $A_W^\dagger A_W = A^\dagger A = A_C^\dagger A_C$, so any solution x_W to the stiffly WLS problem Eq. (1) has the form

$$x_W = A_W^\dagger b + (I - A^\dagger A)z$$

for some vector $z \in C^n$. Choose

$$x_C = A_C^\dagger b + (I - A^\dagger A)z,$$

then x_C is a solution to Eqs. (21)-(22). Therefore

$$\begin{aligned} \|x_W - x_C\| &= \|(A_W^\dagger - A_C^\dagger)b\| \\ &\le \|A_W^\dagger\| \cdot [\|A_W - A_C\| \cdot \|x_{CLS}\| + \|A_W A_W^\dagger (I - A_C A_C^\dagger)\| \|r_C\|] \\ &\le \tfrac{\|A_C^\dagger\|}{1 - e_\epsilon \|A_C^\dagger\|}[e_\epsilon \|x_{CLS}\| + e_\epsilon \|A_C^\dagger\| \|r_C\|], \end{aligned}$$

by applying Lemma 2.2, Eq. (27), and the estimate in Theorem 4.1.

When A has full row rank, or range(A_j^H) for $j = 1, \cdots, k$ are mutually orthogonal, $A_W = A_C = A$ and $A_W^\dagger = A_C^\dagger = A^\dagger$, as mentioned in Corollary 4.1, in this case we immediately have

Corollary 5.1. *In Theorem 5.1, if A has full row rank, or* range(A_j^H) *for $j = 1, \cdots, k$ are mutually orthogonal, then the solution sets of the stiffly WLS problem Eq. (1), the MCLS problem Eqs. (21)-(22) and the ordinary least squares problem*

$$\min_x \|Ax - b\|$$

are same.

6. Conclusion

In this paper we have discussed the relationship between the stiffly weighted pseudoinverse and the corresponding multi-level constrained pseudoinverse, and the solution sets of the stiffly WLS and MCLS problems. We have shown that when $e_\epsilon \|A_C^\dagger\| \ll 1$ and $\|A_C^\dagger\| \ll \sup_{W \in \mathcal{D}} \|A_W^\dagger\|$, then

$$\|A_W^\dagger\| \sim \|A_C^\dagger\| \ll \sup_{W \in \mathcal{D}} \|A_W^\dagger\|$$

and the solution sets of the stiffly WLS problem Eq. (1) is very close to that of the MCLS problem Eqs. (21)-(22).

There are still some questions concerning the stiffly weighted pseudoinverse and stiffly WLS problem remaining not answered.

1. For fixed A and W with W severely stiff, under what conditions are the perturbations to the stiffly weighted pseudoinverse and the stiffly WLS problem stable? We study this problem in a separate paper [22].

2. Recently we found that column pivoting and row interchanging/row sorting Householder QRD, MGS column pivoting and Givens QRD are all numerical unstable for solving stiffly WLS problems. In [23, 25] we respectively propose row block Householder QRD and row block MGS, and show that these algorithms are numerically stable.

References

[1] J. L. Barlow and S. L. Handy, The direct solution of weighted and equality constrained least squares problems, *SIAM J. Sci. Statist. Comput.*, **9** (1988), 704-716.

[2] A. Ben- Israel and T. N. E. Greville, Generalized Inverses: Theory and Applications, John Wiley, New York, 1974.

[3] Å. Björck, Numerical methods for least squares problems, SIAM, Philadelphia, 1996.

[4] L. Eldén, Perturbation theory for the least squares problems with equality constraints, *SIAM J. Numer. Anal.*, **17** (1980), 338-350.

[5] L. Eldén, A weighted pseudoinverse, generalized singular values, and constrained least squares problems, *BIT*, **22** (1982), 237-254.

[6] A. Forsgren, On linear least-squares problems with diagonally dominant weight matrices, *SIAM J. Matrix Anal. Appl.*, **17** (1996), 763-788.

[7] G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd Edit., The Johns Hopkins University Press, Baltimore, MD, 1996.

[8] N. Karmarkar, A new polynomial time algorithm for linear programming, Combinatorica, **4** (1984), 373-395.

[9] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[10] X. Liu and X. Xu, Positive answer to an open problem of G. W. Stewart, *J Numer. Math., J. Chinese Univ.*, **17:**1, (1985), 31-35.

[11] G.W. Stewart, On the continuity of the generalized inverse, *SIAM J. Appl. Math.*, **17** (1969), 33-45.

[12] G.W. Stewart, On the perturbation of pseudo-inverses projections and linear least squares problems, *SIAM Rev.*, **19** (1977), 634-662.

[13] G.W. Stewart, On scaled projections and pseudoinverses, *Linear Algebra Appl.*, **112** (1989), 189-193.

[14] G. W. Stewart, On the weighting method for least squares problems with linear equality constraints, *BIT*, **37** (1997), 961-967.

[15] G. Strang, A framework for equilibrium equations, *SIAM Rev.*, **30** (1988), 283-297.

[16] C. F. Van Loan, On the method of weighting for equality-constrained least squares problems, *SIAM J. Numer. Anal.*, **22** (1985), 851-864.

[17] S.A. Vavasis and Y. Ye, A primal-dual interior point method whose running tine depend only on the constraint matrix, *Math. Programming*, **74** (1996), 79-120.

[18] P.-Å. Wedin, Perturbation theory for pseudoinverses, *BIT*, **13** (1973), 217-232.

[19] M. Wei, Upper bound and stability of scaled pseudoinverses, *Numer. Math.*, **72** (1995), 285-293.

[20] M. Wei, Equivalent formulae for the supremum and stability of weighted pseudoinverses, *Math. Comp.*, **66** (1997), No. 220, 1487-1508.

[21] M. Wei, Supremum and Stability of Weighted Pseudoinverses and Weighted Least Squares Problems: Analysis and Computations, Nova Science, New York, 2001.

[22] M. Wei, On stable perturbation of the stiffly weighted pseudoinverse and weighted least squares problems, Preprint.

[23] M. Wei, A numerically stable row block Householder QRD for solving stiff weighted least squares problems, Preprint.

[24] M. Wei and A.R. De Pierro, Upper perturbation bounds of weighted projections, weighted and constrained least squares problems, *SIAM J. Matrix Anal. Appl.*, **21**:3 (2000), 931-951.

[25] M. Wei and Q. Liu, A numerically stable row block modified Gram- Schmidt QRD for solving stiffly weighted least squares problems, Preprint.