International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 7, 2008, pp 41-46

Geometrical Deficiencies of Moore-Penrose Generalized Inverses

M. Asaduzzaman, Ph.D.
Department of Mathematics
University of Rajshahi
Rajshahi-6205, Bangladesh
E-mail: md_asaduzzaman@hotmail.com

[Received July 15, 2007; Revised February 17, 2008; Accepted February 23, 2008]

Abstract

Let A be an $m \times n$ matrix of nonzero row vectors over the field $\mathbf{K}$ ($\mathbf{C}$ or $\mathbf{R}$). Let $X = \mathbf{K}^n$ and $Y = \mathbf{K}^m$ be Euclidean normed spaces of column vectors. Then A defines a linear map from X into Y. Let $\mathbf{y} \in Y$ be given and consider the problem of finding an $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$. It is clear that the problem is solvable only if $\mathbf{y} \in R(A)$, the range of A. If the notion of solution is generalized by finding an $\mathbf{x} \in X$ such that $\|A\mathbf{x} - \mathbf{y}\|$ is minimum over X (we call it a least square solution of $A\mathbf{x} = \mathbf{y}$) then it can be proved that for every $\mathbf{y} \in Y$ the equation $A\mathbf{x} = \mathbf{y}$ has a least square solution. Among the least square solutions of $A\mathbf{x} = \mathbf{y}$ the one with the smallest norm is called the best approximate solution or the Moore-Penrose generalized solution. Geometrically the matrix equation $A\mathbf{x} = \mathbf{y}$ represents a system of linear equations and multiplying some of them by nonzero scalars the system remains unchanged but the Moore-Penrose generalized solution of the resultant system may be different from that of $A\mathbf{x} = \mathbf{y}$. We wish to have the same generalized solution for all geometrically identical system. In this paper we define such a generalized solution and we prove that this solution gives a geometrical meaning. We also prove that elementary operations on a linear system effect the Moore-Penrose generalized solution to move on the subspace $N(A)^{\perp}$ of X.

Keywords and Phrases: Moore-Penrose generalized inverse, least square solution, normalizer, best standard approximate solution.

AMS Classification: Primary 15A04, 15A06, 15A09, 15A29.

1 Introduction

Let A be an $m \times n$ matrix over the field $\mathbf{K}$ ($\mathbf{C}$ or $\mathbf{R}$). Let $X = \mathbf{K}^n$ and $Y = \mathbf{K}^m$ be Euclidean normed spaces of column vectors. Let $\mathbf{y} \in Y$ be given. An element $\mathbf{x}_0 \in X$ satisfying the condition $\|A\mathbf{x}_0 - \mathbf{y}\| = \min_{\mathbf{x}\in X} \|A\mathbf{x} - \mathbf{y}\|$ is called a **least square solution** of the matrix equation $A\mathbf{x} = \mathbf{y}$. The reason behind this terminology is that $\|A\mathbf{x} - \mathbf{y}\|^2 = \sum_{i=1}^{m} |\mathbf{a}_i\mathbf{x} - y_i|^2$, where $\mathbf{a}_i$ is the ith row vector of A and $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$, and $\mathbf{x}_0$ minimizes this sum. If $\mathbf{x}_0$ is a least square solution of $A\mathbf{x} = \mathbf{y}$ then clearly from the definition of least square solution the set $\mathbf{x}_0 + N(A)$, where $N(A)$ is the null space of A, is the set of all least square solutions of $A\mathbf{x} = \mathbf{y}$. Since $N(A)$ is a closed subspace of X, the set $\mathbf{x}_0 + N(A)$ contains an element of smallest norm. The element of $\mathbf{x}_0 + N(A)$ with smallest norm is called the **best approximate solution** or the **Moore-Penrose generalized solution** of $A\mathbf{x} = \mathbf{y}$. The following theorem guarentees the existence of Moore-Penrose generalized solution of any matrix equation.

Theorem 1.1. *Let A, X, Y be defined as above and let $\mathbf{y} \in Y$ be given. Then the matrix equation $A\mathbf{x} = \mathbf{y}$ has a least square solution.*

Proof. Since $R(A)$ is a closed subspace of Y, $Y = R(A) \oplus R(A)^{\perp}$. Hence there exists unique $\mathbf{y}_1 \in R(A)$ and $\mathbf{y}_2 \in R(A)^{\perp}$ such that $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$. It is clear that the matrix equation $A\mathbf{x} = \mathbf{y}_1$ is consistent. We will now prove that all solutions of $A\mathbf{x} = \mathbf{y}_1$ are least square solution of $A\mathbf{x} = \mathbf{y}$. Let $\mathbf{x}_0$ be a solution of $A\mathbf{x} = \mathbf{y}_1$ then $A\mathbf{x}_0 = \mathbf{y}_1$. Since for any $\mathbf{x} \in X$,

$$\begin{aligned} \|A\mathbf{x} - y\|^2 &= \|(A\mathbf{x} - \mathbf{y}_1) - \mathbf{y}_2\|^2 \\ &= \|(A\mathbf{x} - \mathbf{y}_1)\|^2 + \|\mathbf{y}_2\|^2 \\ &\geq \|\mathbf{y}_2\|^2 \end{aligned}$$

and $\|A\mathbf{x}_0 - y\| = \|\mathbf{y}_2\|$. This completes the proof.

We know that if the matrix equation $A\mathbf{x} = \mathbf{y}$ is consistent and if $\mathbf{x}_0$ is a solution of the equation then the set $\mathbf{x}_0 + N(A) = \{\mathbf{x}_0 + \mathbf{x} : \mathbf{x} \in N(A)\}$ is the set of all solutions of the equation. Similar result holds for least square solutions. In fact we have:

Theorem 1.2. *Let $\mathbf{x}_0$ be a least square solution of the matrix equation $A\mathbf{x} = \mathbf{y}$ then the set of all least square solutions of $A\mathbf{x} = \mathbf{y}$ is $\mathbf{x}_0 + N(A)$.*

Proof. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be respectively the projections of $\mathbf{y}$ on $R(A)$ and $R(A)^{\perp}$, and let $\mathbf{x}_1$ be another least square solution of $A\mathbf{x} = \mathbf{y}$. Then according to the argument in the proof of Theorem 1.1, $\mathbf{x}_0$ and $\mathbf{x}_1$ are both solutions of the consistent equation $A\mathbf{x} = \mathbf{y}_1$ and so $A(\mathbf{x}_1 - \mathbf{x}_0) = 0$. Hence $\mathbf{x}_1 \in \mathbf{x}_0 + N(A)$. Conversely, let $\mathbf{x}_1 \in \mathbf{x}_0 + N(A)$ then $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{x}_2$ for some $\mathbf{x}_2 \in N(A)$. Now,

$$\begin{aligned} \|A\mathbf{x}_1 - \mathbf{y}\|^2 &= \|\big(A(\mathbf{x}_0 + \mathbf{x}_2) - \mathbf{y}_1\big) - \mathbf{y}_2\|^2 \\ &= \|(A\mathbf{x}_0 - \mathbf{y}_1) - \mathbf{y}_2\|^2 \\ &= \|\mathbf{y}_2\|^2. \end{aligned}$$

Hence $\mathbf{x}_1$ is a least square solution of $A\mathbf{x} = \mathbf{y}$. The proof is thus complete.

We know that if $A\mathbf{x} = \mathbf{y}$ is a consistent matrix equation and if B is a matrix such that BA is defined then the matrix equation $BA\mathbf{x} = B\mathbf{y}$ is consistent but the converse may not be true. In fact we have:

Theorem 1.3. *Let $A\mathbf{x} = \mathbf{y}$ be a matrix equation consistent or inconsistent. Then the matrix equation $A^*A\mathbf{x} = A^*\mathbf{y}$ is always consistent.*

Proof. The theorem will be proved if we can show that $A^*\mathbf{y} \in R(A^*A)$. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be respectively the projections of $\mathbf{y}$ on $R(A)$ and $R(A)^{\perp}$. Since $R(A)^{\perp} = N(A^*)$, therefore $\mathbf{y}_2 \in N(A^*)$ and so $A^*(\mathbf{y}_2) = 0$. Now,

$$\begin{aligned} A^*\mathbf{y} &= A^*(\mathbf{y}_1 + \mathbf{y}_2) \\ &= A^*\mathbf{y}_1 \in R(A^*A). \end{aligned}$$

We now prove that the solutions of $A^*A\mathbf{x} = A^*\mathbf{y}$ are least square solutions of $A\mathbf{x} = \mathbf{y}$. In fact we have:

Theorem 1.4. *Let $\mathbf{y} \in Y$ be given and let $\mathbf{y}_1$ be its projection on $R(A)$. Then a point $\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{y}_1$ if and only if it is a solution of $A^*A\mathbf{x} = A^*\mathbf{y}$.*

Proof. Let $\mathbf{y}_2$ be the projection of $\mathbf{y}$ on $R(A)^{\perp}$. Then $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$ and $A^*\mathbf{y}_2 = 0$. Let $\mathbf{x}_0$ be a solution of $A\mathbf{x} = \mathbf{y}_1$. Then it is a solution of $A^*A\mathbf{x} = A^*\mathbf{y}_1$. But $A^*\mathbf{y} = A^*(\mathbf{y}_1 + \mathbf{y}_2) = A^*\mathbf{y}_1$. Hence $\mathbf{x}_0$ is a solution of $A^*A\mathbf{x} = A^*\mathbf{y}$. To prove the solution sets for the equations $A\mathbf{x} = \mathbf{y}_1$ and $A^*A\mathbf{x} = A^*\mathbf{y}$ identical, we need to prove that $N(A) = N(A^*A)$. Trivially, $N(A) \subseteq N(A^*A)$. To prove the converse, let $A^*A\mathbf{x} = 0$. Then $A\mathbf{x} \in N(A^*) = R(A)^{\perp}$. But $A\mathbf{x} \in R(A)$. Hence $A\mathbf{x} \in R(A) \cap R(A)^{\perp}$ and so $A\mathbf{x} = 0$. Thus $\mathbf{x} \in N(A)$.

2 Properties of Moore-Penrose generalized solutions

A matrix equation $A\mathbf{x} = \mathbf{y}$ repesents a system of linear equations if and only if none of the row vectors of A are zero. From now on we assume that the row vectors of A are nonzero. By a least square solution of a system of linear equations we mean that for the corresponding matrix equation. It is clear from the definition of least square solutions of a matrix equation $A\mathbf{x} = \mathbf{y}$ that the set of all least square solutions of a system of linear equations remains unchanged due to any permutation among them. So, the Moore-Penrose generalized solution of a system of linear equations is invariant under any permutation among them. It is a nice property. But if we multiply some equations by nonzero scalars then the set of least square solutions of the original system may be different from that for resultant system though geometrically the system remains unchanged. Hence Moore-Penrose generalized solutions of different geometrically identical linear systems may be different. This is a defficiency of Moore-Penrose generalized solution. We give a method to overcome this defficiency and the generalized solutions obtained by our method will be called the **best standard approximate solution**.

Theorem 2.1. *Let $\mathbf{x}_0$ be the Moore-Penrose generalized solution of $A\mathbf{x} = \mathbf{y}$. Then $\mathbf{x}_0 \in N(A)^{\perp}$.*

Proof. Let $\mathbf{x}_1$ be a least square solution of $A\mathbf{x} = \mathbf{y}$. Then by Theorem 1.2, the set of all least square solutions of $A\mathbf{x} = \mathbf{y}$ is $\mathbf{x}_1 + N(A)$. Hence

$$\mathbf{x}_0 - \mathbf{x}_1 \in N(A) \quad \text{and} \quad \|\mathbf{x}_0\| = \inf_{\mathbf{x} \in N(A)} \|\mathbf{x}_1 + \mathbf{x}\|.$$

Let $\mathbf{x} \in N(A)$ with $\mathbf{x} \neq \mathbf{0}$ and let t be any nonzero scalar then $\mathbf{x}_0 - t\mathbf{x} \in \mathbf{x}_1 + N(A)$ and $\mathbf{x}_0 - t\mathbf{x} \neq \mathbf{x}_0$. Hence

$$\begin{aligned} & \|\mathbf{x}_0 - t\mathbf{x}\|^2 > \|\mathbf{x}_0\|^2 \\ \Rightarrow \quad & \langle \mathbf{x}_0 - t\mathbf{x}, \mathbf{x}_0 - t\mathbf{x} \rangle > \|\mathbf{x}_0\|^2 \\ \Rightarrow \quad & |t|^2 \|\mathbf{x}\|^2 - 2\mathrm{Re}(t\langle \mathbf{x}, \mathbf{x}_0 \rangle) > 0. \end{aligned}$$

The above inequality holds for all nonzero scalar t. If $\langle \mathbf{x}, \mathbf{x}_0 \rangle \neq 0$ then we put $t = \langle \mathbf{x}, \mathbf{x}_0 \rangle s$, where s is any positive real, in the above inequality. We have

$$s^2 |\langle \mathbf{x}, \mathbf{x}_0 \rangle|^2 - 2s |\langle \mathbf{x}, \mathbf{x}_0 \rangle|^2 > 0.$$

Thus if we choose $s = 1$ then we have

$$\|\mathbf{x}\|^2 > 2 \quad \text{for any nonzero} \quad \mathbf{x} \in N(A),$$

which is absurd. Hence $\langle \mathbf{x}, \mathbf{x}_0 \rangle = 0$ for all $\mathbf{x} \in N(A)$. Therefore $\mathbf{x}_0 \in N(A)^{\perp}$.

Theorem 2.2. *Moore-Penrose generalized solution of the resultant system of linear equations obtained from elementary operations applied in a linear system whose matrix equation is $A\mathbf{x} = \mathbf{y}$ lies on $N(A)^{\perp}$.*

Proof. Suppose a series of elementary operations is applied on a linear system of equations whose matrix equation is $A\mathbf{x} = \mathbf{y}$ and let the matrix equation of the resultant system be $B\mathbf{x} = \mathbf{z}$. Then there exists a nonsingular matrix P such that $B = PA$ and $\mathbf{z} = P\mathbf{y}$. Let $\mathbf{x}_0$ be a least square solution of $B\mathbf{x} = \mathbf{z}$ then by Theorem 1.2, the set of all least square solution of $B\mathbf{x} = \mathbf{z}$ is $\mathbf{x}_0 + N(B)$. Since P is nonsingular and $B = PA$, therefore $N(B) = N(A)$. Hence the set of all least square solutions of $B\mathbf{x} = z$ is $\mathbf{x}_0 + N(A)$. Thus the Moore-Penrose generalized solution lies on $N(A)^{\perp}$.

Cor. The Moore-Penrose generalized solution of a geometrically identical system to the linear system whose matrix equation is $A\mathbf{x} = y$ lies on $N(A)^{\perp}$.

3 Best standard approximate solution

In order to define a best standard approximate solution of a system of linear equations we need the following definitions:

Definition 1. *For any $m \times n$ matrix A with nonzero row vectors we define a diagonal matrix $D_A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, where λ_i is the reciprocal of the norm of the ith row vector of A, and call it the* **normalizer** *of the matrix A.*

Definition 2. *A square matrix P is called a matrix of* **essentially diagonal** *if each row and each column contain exactly one nonzero element. In other words, P is of essentially diagonal if P is nonsingular and reducible to a diagonal matrix by some permutations among the rows of P.*

Definition 3. *A matrix equation $A\mathbf{x} = \mathbf{y}$ is called geometrically identical to another matrix equation $B\mathbf{x} = \mathbf{z}$ if there exists a matrix P of essentially diagonal such that $B = PA$ and $\mathbf{z} = P\mathbf{y}$.*

Definition 4. *Let $A\mathbf{x} = \mathbf{y}$ be a matrix equation of a system of linear equations. We call the equation $D_A A\mathbf{x} = D_A\mathbf{y}$ as the* **normalized form** *of $A\mathbf{x} = \mathbf{y}$.*

Definition 5. *Let $A\mathbf{x} = \mathbf{y}$ be a matrix equation of a system of linear equations. The Moore-Penrose generalized solution of the normal form $D_A A\mathbf{x} = D_A\mathbf{y}$ of $A\mathbf{x} = \mathbf{y}$ is called the* **best standard approximate** *solution of $A\mathbf{x} = \mathbf{y}$.*

Lemma 3.1. *Let $P = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_m)$ be a nonsingular diagonal matrix and let A be an $m \times n$ matrix with nonzero row vectors then*

$$P^* D_{PA}^2 P = D_A^2.$$

Proof. Let $D_A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, where λ_i's are defined as in Definition 1. Since the ith row of the matrix PA is α_i times that of A, by Definition 1, $D_{PA} = \mathrm{diag}(\frac{\lambda_1}{|\alpha_1|}, \frac{\lambda_2}{|\alpha_2|}, \ldots, \frac{\lambda_m}{|\alpha_m|})$ and so $D_{PA}^2 = \mathrm{diag}(\frac{\lambda_1^2}{|\alpha_1|^2}, \frac{\lambda_2^2}{|\alpha_2|^2}, \ldots, \frac{\lambda_m^2}{|\alpha_m|^2})$. Thus

$$\begin{aligned} P^* D_{PA}^2 P &= \mathrm{diag}(\overline{\alpha_1} \tfrac{\lambda_1^2}{|\alpha_1|^2} \alpha_1, \overline{\alpha_2} \tfrac{\lambda_2^2}{|\alpha_2|^2} \alpha_2, \ldots, \overline{\alpha_m} \tfrac{\lambda_m^2}{|\alpha_m|^2} \alpha_m) \\ &= \mathrm{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_m^2) \\ &= D_A^2. \end{aligned}$$

Lemma 3.2. *Let P be a nonsingular diagonal matrix and let $A\mathbf{x} = \mathbf{y}$ be the matrix equation of a system of linear equations. Then the best standard approximate solutions of $A\mathbf{x} = \mathbf{y}$ and $PA\mathbf{x} = P\mathbf{y}$ are identical.*

Proof. The lemma will be proved if we can show that the set of least square solutions of $D_A A\mathbf{x} = D_A\mathbf{y}$ and that of $D_{PA} PA\mathbf{x} = D_{PA} P\mathbf{y}$ are identical. Let $\mathbf{x}_0$ be a least square solution of $D_{PA} PA\mathbf{x} = D_{PA} P\mathbf{y}$. Then $\mathbf{x}_0$ satisfies the consistent equation

$$\begin{aligned} & (PA)^* D_{PA}^2 (PA)\mathbf{x} = (PA)^* D_{PA}^2 P\mathbf{y} \\ \Longleftrightarrow\quad & A^* P^* D_{PA}^2 PA\mathbf{x} = A^* P^* D_{PA}^2 P\mathbf{y} \\ \Longleftrightarrow\quad & A^* D_A^2 A\mathbf{x} = A^* D_A^2 \mathbf{y}. \end{aligned}$$

Hence $\mathbf{x}_0$ is a least square solution of $D_A A\mathbf{x} = D_A\mathbf{y}$.

Lemma 3.2 and the fact that the Moore-Penrose generalized solution of a system of linear equations is invariant under permutations among the equations imply the following theorem:

Theorem 3.1. *All geometrically equivalent system of linear equations have the same best standard approximate solution.*

4 Geometrical Interpretation of best standard approximate solutions

Let $\mathbf{K} = \mathbf{R}$ and let $A\mathbf{x} = \mathbf{y}$ be the matrix equation of a system of m linear equations. Let $\mathbf{x}_0$ be the best standard approximate solution of $A\mathbf{x} = \mathbf{y}$. Then by definition $\mathbf{x}_0$ is the Moore-Penrose generalized solution of $D_A A\mathbf{x} = D_A\mathbf{y}$. Hence $\mathbf{x}_0$ minimizes the norm $\|D_A A\mathbf{x} - D_A\mathbf{y}\|^2 = \sum_{i=1}^m \lambda_i^2 |\mathbf{a}_i\mathbf{x} - y_i|^2$, where $\mathbf{a}_i\mathbf{x} = y_i$ is the ith linear equation of the system and $\lambda_i = \frac{1}{\|\mathbf{a}_i\|}$. But

$$\sum_{i=1}^{m} \lambda_i^2 |\mathbf{a}_i\mathbf{x}_0 - y_i|^2 = \sum_{i=1}^{m} \frac{|\mathbf{a}_i\mathbf{x}_0 - y_i|^2}{\|\mathbf{a}_i\|^2},$$

which is the sum of the square distances of the point $\mathbf{x}_0$ from the hyperplanes $\mathbf{a}_i\mathbf{x} = y_i$. Hence best standard approximate solution minimizes the sum of the square distances of a point from the hyperplanes given by the system of linear equations.

Acknowledgments

The author wishes to express his deep thanks to Professor S. Saitoh, Gunma University, Japan for reading carefully this paper and making his valuable comments.

References

[1] C.L. Lawson and R.J. Hanson, (1995). *Solving Least Square Problems*, SIAM, Philadelphia, 1995.

[2] C.R. Rao, (1973). *Linear Statistical Inference and its Applications*, 2^{nd} ed, Wiley Eastern Ltd., 1973.

[3] Frank Natterer, (1986). *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, 1986.

[4] H.W. Engl, M. Hanke and A. Neubauer, (2000). *Regularization of Inverse Problems*, Kluwer Academic Publishers, 2000.