Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2011, Article ID 385870, 5 pages
doi:10.1155/2011/385870

Research Article

A Representation of Nonhomogeneous Quadratic Forms with Application to the Least Squares Solution

Czesław Stępniak

Institute of Mathematics, University of Rzeszów, Rejtana 16 A, 35-959 Rzeszów, Poland

Correspondence should be addressed to Czesław Stępniak, stepniak@umcs.lublin.pl

Received 5 June 2011; Revised 16 July 2011; Accepted 16 July 2011

Academic Editor: Piermarco Cannarsa

The least squares problem appears, among others, in linear models, and it refers to inconsistent system of linear equations. A crucial question is how to reduce the least squares solution in such a system to the usual solution in a consistent one. Traditionally, this is reached by differential calculus. We present a purely algebraic approach to this problem based on some identities for nonhomogeneous quadratic forms.

1. Introduction and Notation

The least squares problem appears, among others, in linear models, and it refers to inconsistent system $\mathbf{A}\mathbf{x} = \mathbf{b}$ of linear equations. Formally, it reduces to minimizing the nonhomogeneous quadratic form

$$f(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}). \tag{1.1}$$

Classical approach to the problem, presented in such known books as Scheffé ([1, Chapter 1]), Rao ([2, pages 222 and 223]), Rao and Toutenburg ([3, pages 20–23]), uses differential calculus and leads to the so called normal equation $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, which is consistent. The aim of this note is to present some useful algebraic identities for nonhomogeneous quadratic forms leading directly to normal equation.

Traditional vector-matrix notation will be used. Among others, if $\mathbf{M}$ is a matrix then $\mathbf{M}^T$, $\mathcal{R}(\mathbf{M})$, and $r(\mathbf{M})$ stand for its transposition, range (column space), and rank. Moreover, by R^n will be denoted the n-dimensional euclidean space represented by column vectors.

2. Background

Any system of linear equations may be presented in the vector-matrix form as

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.1}$$

where $\mathbf{A}$ is a given $n \times p$ matrix, $\mathbf{b}$ is a given vector in R^n, while $\mathbf{x} \in R^p$ is unknown vector. It is well known that (2.1) is consistent, if and only if, $\mathbf{b}$ belongs to the range $\mathcal{R}(\mathbf{A})$.

If (2.1) is inconsistent, one can seek for a vector $\mathbf{x}$ minimizing the norm $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ or, equivalently, its square $(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b})$. The *Least Squares Solution (LSS)* of (2.1) is defined as a vector $\mathbf{x}_0 \in R^p$ such that

$$(\mathbf{A}\mathbf{x}_0 - \mathbf{b})^T(\mathbf{A}\mathbf{x}_0 - \mathbf{b}) \leq (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \quad \forall \mathbf{x} \in R^p. \tag{2.2}$$

A crucial problem is how to reduce the *LSS* of the inconsistent equation (2.1) to the usual solution of a consistent one. Formally, the least squares problem deals with minimizing the nonhomogeneous quadratic form $f(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b})$. Traditionally, this problem is solved by differential calculus and leads to the normal equation $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$.

In the next section, we will present some useful algebraic identities for nonhomogeneous quadratic forms. They yield directly the inequality (2.2).

3. Identities and Inequalities for Nonhomogeneous Quadratic Forms

The usual, that is *homogeneous* quadratic form is a real function $f_M(\mathbf{x}) = \mathbf{x}^T\mathbf{M}\mathbf{x}$ defined on R^p. In this note, we shall consider also *nonhomogeneous* quadratic forms of type

$$f_{M,a}(\mathbf{x}) = (\mathbf{x} - \mathbf{a})^T\mathbf{M}(\mathbf{x} - \mathbf{a}), \tag{3.1}$$

where $\mathbf{M}$ is a symmetric $p \times p$ matrix and $\mathbf{a}$ is a vector in R^p.

Some inequalities for nonhomogeneous quadratic forms may be found in Stępniak [4]. Let us recall one of these results, which is very useful in the nonhomogeneous linear estimation.

Lemma 3.1. *For any symmetric nonnegative definite matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of order p, the condition*

$$(\mathbf{x} - \mathbf{x}_1)^T\mathbf{M}_1(\mathbf{x} - \mathbf{x}_1) + c_1 \geq (\mathbf{x} - \mathbf{x}_2)^T\mathbf{M}_2(\mathbf{x} - \mathbf{x}_2) + c_2 \tag{3.2}$$

for some $c_1, c_2 \in R$, $\mathbf{x}_1, \mathbf{x}_2 \in R^p$ and all $\mathbf{x} \in R^p$ implies that $\mathbf{M}_1 - \mathbf{M}_2$ is nonnegative definite and $c_1 - c_2 \geq 0$.

Now we will present some identity which may serve as a convenient tool in the *LSS* of (2.1). For convenience, we will start from the case $r(\mathbf{A}) = p$, leaving the singular case $r(\mathbf{A}) < p$ to Section 5.

Proposition 3.2. *For arbitrary $n \times p$ matrix $\mathbf{A}$ of rank p and arbitrary vector $\mathbf{b} \in R^n$*

$$(\mathbf{Ax}-\mathbf{b})^T(\mathbf{Ax}-\mathbf{b}) = \left(\mathbf{A}^T\mathbf{Ax}-\mathbf{A}^T\mathbf{b}\right)^T\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\left(\mathbf{A}^T\mathbf{Ax}-\mathbf{A}^T\mathbf{b}\right) + \mathbf{b}^T\left[\mathbf{I}_n - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\right]\mathbf{b}. \tag{3.3}$$

Proof. Let us start from the trivial identity

$$\mathbf{I}_n = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T + \left[\mathbf{I}_n - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}\right]. \tag{3.4}$$

We only need to premultiply this identity by $(\mathbf{Ax}-\mathbf{b})^T$, postmultiply it by $(\mathbf{Ax}-\mathbf{b})$, and then collect the terms to get (3.3). □

4. Least Squares and Usual Solutions: Nonsingular Case

As above, we consider an inconsistent equation $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is an $n \times p$ matrix of rank p. We are interested in the *LSS* of this equation.

Theorem 4.1. *Vector $\mathbf{x} \in R^p$ is a Least Squares Solution of the inconsistent equation $\mathbf{Ax} = \mathbf{b}$, if and only if, it is usual solution of the consistent equation*

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}. \tag{4.1}$$

Proof. Consistency of (4.1) follows from the fact that $\mathcal{R}(\mathbf{A}^T\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)$.

We note that the second component in the right side of the identity (3.3) does not depend on $\mathbf{x}$. Thus, we only need to minimize the first one. Since $\mathbf{A}^T\mathbf{A}$, and in consequence $(\mathbf{A}^T\mathbf{A})^{-1}$, is positive definite, this component is minimal, if and only if, $\mathbf{A}^T\mathbf{Ax} - \mathbf{A}^t\mathbf{b} = \mathbf{0}$. This completes the proof. □

5. General Case

If $r(\mathbf{A}) < p$ then the matrix $\mathbf{A}^T\mathbf{A}$ is singular, and, therefore, the identity (3.3) is not applicable. However, as we will show, it remains true if we replace $(\mathbf{A}^T\mathbf{A})^{-1}$ by arbitrary generalized inverse $(\mathbf{A}^T\mathbf{A})^{-}$.

There are many papers on generalized inverses and several books; among others Bapat [5] Ben-Israel and Greville [6], Campbell and Meyer [7], Pringle and Rayner [8], and Rao and Mitra [9]. A recent paper by Stępniak [10] may serve as a brief and self-contained introduction to this field.

Let us recall that a given $n \times p$ matrix $\mathbf{A}$, its generalized inverse $\mathbf{A}^{-}$ is defined as an arbitrary $p \times n$ matrix $\mathbf{G}$ satisfying the condition $\mathbf{AGA} = \mathbf{A}$.

A key result in this section is stated as follows.

Proposition 5.1. *For arbitrary $n \times p$ matrix $\mathbf{A}$ and arbitrary vector $\mathbf{b} \in R^n$*

$$(\mathbf{A}\mathbf{x}-\mathbf{b})^T(\mathbf{A}\mathbf{x}-\mathbf{b}) = \left(\mathbf{A}^T\mathbf{A}\mathbf{x}-\mathbf{A}^T\mathbf{b}\right)^T\left(\mathbf{A}^T\mathbf{A}\right)^-\left(\mathbf{A}^T\mathbf{A}\mathbf{x}-\mathbf{A}^T\mathbf{b}\right) + \mathbf{b}^T\left[\mathbf{I}_n - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^-\mathbf{A}^T\right]\mathbf{b}, \tag{5.1}$$

where $^-$ means arbitrary generalized inverse.

Proof. The idea of the proof is the same as in Proposition 3.2. Since $\mathcal{R}(\mathbf{A}^T\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)$ one can replace the vector $\mathbf{A}^T\mathbf{b}$ by $\mathbf{A}^T\mathbf{A}\mathbf{c}$ for some $\mathbf{c}$. □

Now we will apply the identity (5.1) to the least squares problem.

Theorem 5.2. *For arbitrary $n \times p$ matrix $\mathbf{A}$ and arbitrary vector $\mathbf{b} \in R^n$,*

(i) *a vector $\mathbf{x} \in R^p$ is a Least Squares Solution of the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, if and only if, it is the usual solution of the (consistent) equation $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$;*

(ii) *the lower bound of $(\mathbf{A}\mathbf{x}-\mathbf{b})^T(\mathbf{A}\mathbf{x}-\mathbf{b})$ is equal to $\mathbf{b}^T[\mathbf{I}_n - \mathbf{A}(\mathbf{A}^T\mathbf{A})^-\mathbf{A}^T]\mathbf{b}$, and it does not depend on choice of generalized inverse $(\mathbf{A}^T\mathbf{A})^-$.*

Proof. By setting $\mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{A}\mathbf{c}$, the first component in the right side of (5.1) reduces to $(\mathbf{x}-\mathbf{c})^T(\mathbf{A}^T\mathbf{A})(\mathbf{x}-\mathbf{c})$ which is nonnegative and takes zero, if and only if, $(\mathbf{A}^T\mathbf{A})(\mathbf{x}-\mathbf{c}) = \mathbf{0}$ or, equivalently, if $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$. Since the second component does not depend on $\mathbf{x}$, this is just total minimum. The same setting shows that the lower bound does not depend on the choice of generalized inverse. □

Acknowledgment

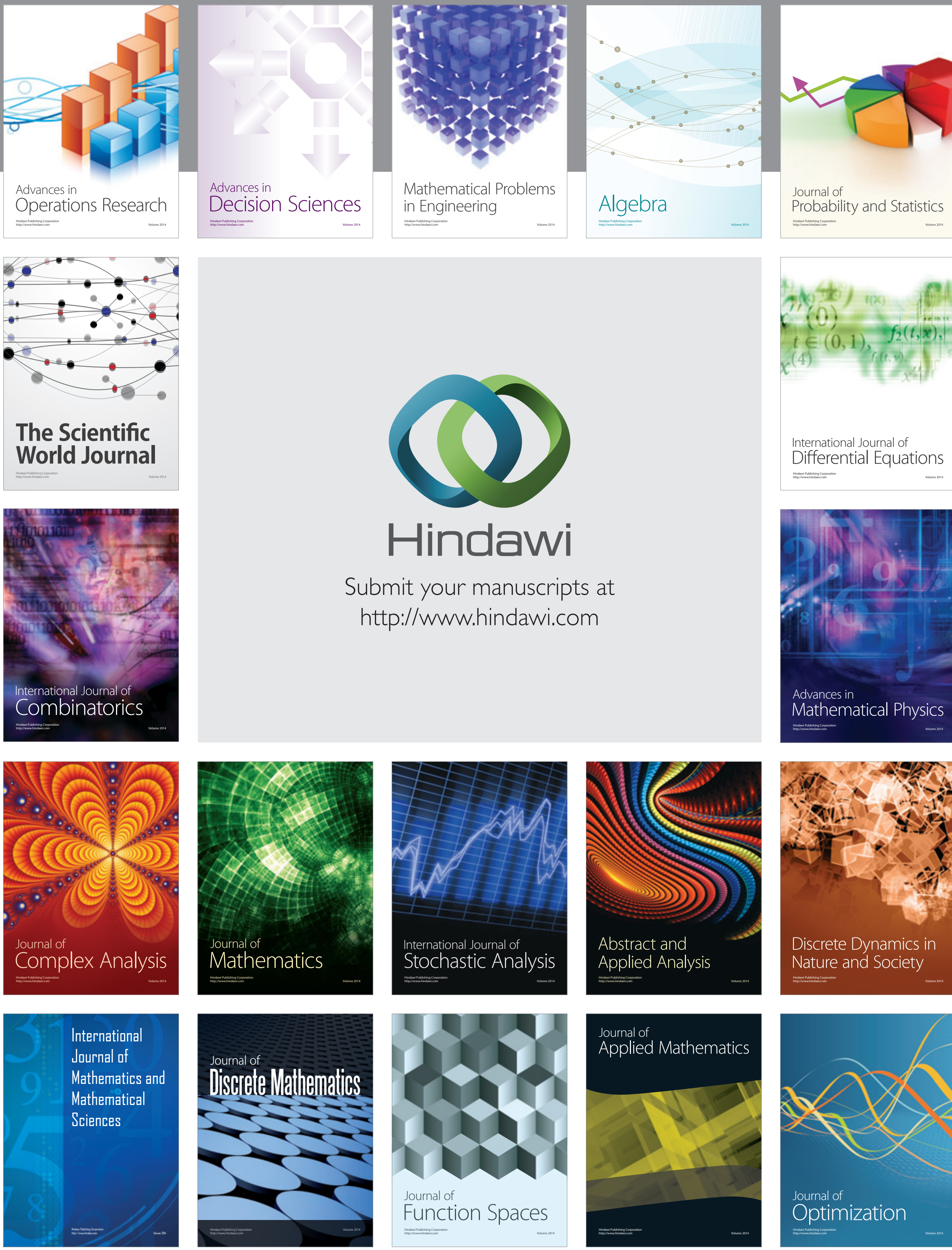
The author thanks a referee for his (or her) useful suggestions concerning presentation of this paper.

References

[1] H. Scheffé, *The Analysis of Variance*, John Wiley & Sons, New York, NY, USA, 1959.
[2] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley Series in Probability and Mathematical Statistic, John Wiley & Sons, New York, NY, USA, 2nd edition, 1973.
[3] C. R. Rao and H. Toutenburg, *Linear Models: Least Squares and Alternatives*, Springer Series in Statistics, Springer, New York, NY, USA, 1995.
[4] C. Stępniak, “Admissible linear estimators in mixed linear models,” *Journal of Multivariate Analysis*, vol. 31, no. 1, pp. 90–106, 1989.
[5] R. B. Bapat, *Linear Algebra and Linear Models*, Universitext, Springer, New York, NY, USA, 2nd edition, 2000.
[6] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 15, Springer, New York, NY, USA, 2nd edition, 2003.
[7] S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, Dover, New York, NY, USA, 1991.
[8] R. M. Pringle and A. A. Rayner, *Generalized Inverses of Matrices with Applications in Statistics*, Griffin, London, UK, 1971.

[9] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, New York, NY, USA, 1971.

[10] C. Stępniak, "Through a generalized inverse," *Demonstratio Mathematica*, vol. 41, no. 2, pp. 291–296, 2008.


Advances in Operations Research
Advances in Decision Sciences
Mathematical Problems in Engineering
Algebra
Journal of Probability and Statistics
The Scientific World Journal
International Journal of Differential Equations
Hindawi
Submit your manuscripts at http://www.hindawi.com
International Journal of Combinatorics
Advances in Mathematical Physics
Journal of Complex Analysis
Journal of Mathematics
International Journal of Stochastic Analysis
Abstract and Applied Analysis
Discrete Dynamics in Nature and Society
International Journal of Mathematics and Mathematical Sciences
Journal of Discrete Mathematics
Journal of Function Spaces
Journal of Applied Mathematics
Journal of Optimization