The loss value of multilinear regression

Helmut Kahl

Abstract. Determinant formulas are presented for: a certain positive semi-
definite, hermitian matrix; the loss value of multilinear regression; the mul-
ti-pile linear regression coefficient.

Mathematical Subject Classification (2010). 15A03, 15A15, 62H12.

Keywords. determinant; positive semidefinite hermitian matrix; loss value
of multilinear regression; multiple regression coefficient

1. Euclidean distance by help of determinants

Introduction. We consider the euclidean norm \( x \mapsto \|x\| := \sqrt{x^*x} \) of column
vectors \( x \in \mathbb{C}^m \). For a complex \( m \times n \)-matrix \( A \in \mathbb{C}^{m \times n} \) and a vector \( b \in \mathbb{C}^m \)
we denote by \( (A|b) \) the \( m \times (n+1) \)-matrix \( A \) together with \( b \) as the last column.
The following theorem is well-known in the real case and can be shown by
help of a formula for the volume of an \( n \)-dimensional parallelepiped embedded
in \( \mathbb{R}^m \). Here we offer a less known proof via QR-decomposition of a complex
matrix.

Theorem 1. For the euclidean distance \( \text{dist}(A, b) \) between \( b \in \mathbb{C}^m \) and the
column space of \( A \in \mathbb{C}^{m \times n} \) it holds:

\[
\text{dist}(A, b) \sqrt{\det(A^*A)} = \sqrt{\det((A|b)^*(A|b))}
\]

Proof. For a unitary \( m \times m \)-matrix \( Q \) it holds \( \text{dist}(A, b) = \text{dist}(QA, Qb) \) due to [1], thm. 2.1.4(g). According to [1], thm. 2.1.14(d) there is a unitary \( Q \) s.t.
\( (A|b) = Q(R|c) \) with \( (R|c) \) upper right triangular. Hence \( \delta = \text{dist}(R, c) \) is the
absolute value of the \( (n + 1) \)-th coordinate of \( c \). So we have

\[
\det((A|b)^*(A|b)) = \det((R|c)^*(R|c)) = \delta^2 \det(R^*R) = \delta^2 \det(A^*A).
\]

Since \( \det(A^*A) \geq 0 \) for arbitrary complex matrices \( A \), s. e.g. [1], thm. 4.1.5
& 7.2.7(a), the assertion follows. \( \square \)
Matrix equation for the distance. In case \( A \) has full column rank \( \operatorname{rk}(A) = n \) we have \( \det(A^*A) > 0 \). Then the formula yields \( \operatorname{dist}(A,b) \) as a quotient of the two square root values. And by plugging the minimum point \( x = (A^*A)^{-1}A^*b \) into \( \|Ax - b\|^2 \) we obtain
\[
\frac{\det((A|b)^*(A|b))}{\det(A^*A)} = \operatorname{dist}(A,b)^2 = b^*b - b^*A(A^*A)^{-1}A^*b.
\]

Special determinant equation. Now, for a matrix \( A \in \mathbb{C}^{(n+1)\times n} \) let \( A_i \) denote the matrix \( A \) without its \( i \)-th row. Via developing \( \det((A|a)_{ij}) = 0 \) by the last column for every column \( a_j \) of \( A \) we see that the vector \( b := \left( (-1)^{n+1}\det(A_i) \right)_{i=1,...,n+1} \) is orthogonal to the column space of \( A \). Applying Theorem 1 to \( A \) and \( b \) we obtain

**Corollary 2.** For \( A \in \mathbb{C}^{(n+1)\times n} \) holds the identity of \( n \times n \)-determinants:
\[
\sum_{i=1}^{n+1} |\det(A_i)|^2 = \det(A^*A)
\]

2. Loss value and correlation

Introduction. The task of *multiple linear regression* is the computation of regression coefficients \( \alpha_0, \alpha_1, ..., \alpha_n \) of the fitting hyperplane (in \( \mathbb{R}^{n+1} \))
\[
y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_n x_n
\]
as a function of variables \( x_1, ..., x_n \in \mathbb{R} \) from (empirical) data points \( (x_{11}, ..., x_{1n}, y_1), ..., (x_{m1}, ..., x_{mn}, y_m) \in \mathbb{R}^{n+1}, m \in \mathbb{N} \) s.t. the loss value
\[
\delta := \left( \sum_{i=1}^{m} (\alpha_0 + \alpha_1 x_{i1} + ... + \alpha_n x_{in} - y_i)^2 \right)^{1/2}
\]
is at minimum. For \( a := (\alpha_0, \alpha_1, ..., \alpha_n)^t, \ y := (y_1, ..., y_m)^t \) and the matrix \( (1|X) \) that we obtain from \( X := (x_{ij})_{i\in\mathbb{N},j\in\mathbb{N}} \) by prepending \( (1, ..., 1)^t \in \mathbb{R}^m \) as an extra column (of index 0) we have \( \delta = \|(1|X)a - y\| \). So the minimal value of \( \delta \) is the euclidean distance between \( y \) and the column space of \( (1|X) \).

Centering. In statistics it is common to express empirical values of expectation with help of the arithmetic mean \( \bar{y} := (y_1 + ... + y_m)/m \) of a (sample) vector like \( y \) above. A regression vector \( a \) like described above is defined by the normal equation system
\[
(1|X)^t(1|X)a = (1|X)^ty.
\] (2.1)
After division by \( m \) the equation of row index 0 of equation 2.1 ends in
\[
\bar{y} = \alpha_0 + \alpha_1 \bar{x}_1 + ... + \alpha_n \bar{x}_n
\] (2.2)
The loss value of multilinear regression

where \( x_j \) denotes the \( j \)-th column of \( X \). We denote by \( \hat{y} := (y_1 - \bar{y}, \ldots, y_m - \bar{y})^t \) the centering of \( y \) and by \( \hat{X} \) the \( m \times n \)-matrix obtained from \( X \) by centering all its columns. Then the normal equations of row indices 1 to \( n \) of equation 2.1 are transformed to

\[
\hat{X}^t \hat{X} a_1 = \hat{X}^t \hat{y}, \quad a_1 := (\alpha_1, \ldots, \alpha_n)^t
\]  

(2.3)

by subtracting the \( \bar{x}_i \)-th multiple of equation 2.2 from the \( i \)-th normal equation for \( i = 1, \ldots, n \). This shows \( \text{rk}(1 \mid X) = \text{rk}(\hat{X}) + 1 \).

**Theorem 3.** In case \( \text{rk}(1 \mid X) = n + 1 \) the loss value of the sample matrix \((X \mid y)\) equals

\[
\sqrt{\det \left( \left( \hat{X} \mid \hat{y} \right)^t \left( \hat{X} \mid \hat{y} \right) \right) / \det \left( \left( \hat{X} \right)^t \hat{X} \right)}
\]

Proof. Expressing \( \alpha_0 \) in terms of the other regression coefficients by help of equation 2.2 gives us \( y - \alpha_0 - \alpha_1 x_1 - \ldots - \alpha_n x_n = \hat{y} - \alpha_1 \hat{x}_1 - \ldots - \alpha_n \hat{x}_n \). Hence the loss value is the euclidean distance between \( \hat{y} \) and the column space of \( \hat{X} \). Because \( \hat{X} \) has full rank the formula follows by Theorem 1. \( \square \)

**Correlation.** For the orthogonal projection \( p := (1 \mid X)a \) of \( y \) onto the column space of \((1 \mid X)\) it holds \( \hat{p} = \hat{X} a_1 \). So by equation 2.3 \( \hat{p} \) is the orthogonal projection of \( \hat{y} \) onto the column space of \( \hat{X} \). Hence in case \( \hat{y}, \hat{p} \neq 0 \) the angle between \( \hat{y} \) and \( \hat{p} \) is at most \( \pi/2 \). Therefore the multiple correlation coefficient

\[
\rho(X, y) := \hat{y}^t \hat{p} / \|\hat{p}\| / \|\hat{y}\|
\]

between \( y \) and \( X \) is non-negative. According to the Cauchy-Schwarz inequality it is at most 1. The latter theorem allows the computation of \( \rho(X, y) \) without the computation of \( p \), i.e. without performing the linear regression.

**Corollary 4.** For a sample vector \( y \in \mathbb{R}^m \) with \( \hat{y} \neq 0 \) and a sample matrix \( X \in \mathbb{R}^{m \times n} \) with \( \text{rk}(\hat{X}) = n \) it holds

\[
\rho(X, y) = \sqrt{1 - \det \left( \left( \hat{X} \mid \hat{y} \right)^t \left( \hat{X} \mid \hat{y} \right) \right) / \det \left( \left( \hat{X} \right)^t \hat{X} \right)} \left( \hat{y}^t \hat{y} \right) / (m - 1)
\]

Proof. The assertion follows from Theorem 3 by the Theorem of Pythagoras applied to \( \hat{y} / \|\hat{y}\| \) as the hypotenuse and \( \hat{p} / \|\hat{p}\| \) as a cathetus. \( \square \)

---

1. Then \( (\hat{X}^t \hat{X}) / (m - 1) \) is the sample covariance matrix of the sample matrix \( X \). It serves as an estimator of the covariance matrix of the random vector \((X_1, \ldots, X_n)\) whose \( m \) samples are given by \( X \), row by row. With the additional random variable \( Y \) whose samples are represented by \( y \) the mean squared loss value \( \delta^2 / (m - 1) \) is an estimator of the expected value of the random variable \((Y - \alpha_0 - \alpha_1 X_1 - \ldots - \alpha_n X_n)^2\); s. e.g. [2], Kap. 3.8!

2. The condition \( \hat{y} \neq 0 \) means a non-zero sample variance \( (\hat{y}^t \hat{y}) / (m - 1) \) of \( y \).
References

[1] R. A. Horn/C. R. Johnson, *Matrix Analysis*, 2nd ed. (2013) Cambr. Univ. Press
[2] F. E. Beichelt/D. C. Montgomery (Hrsg.), *Teubner-Taschenbuch der Stochastik* (2003) Teubner Verlag

Helmut Kahl
Munich University of Applied Sciences
Lothstr. 34
D-80335 München
Germany

e-mail: kahl@hm.edu