A GEOMETRIC PROPERTY OF THE FUNCTIONAL ASSOCIATED WITH GENERAL SYSTEMS OF ALGEBRAIC EQUATIONS

I.R. Ciric

Department of Electrical and Computer Engineering
The University of Manitoba, Winnipeg, Manitoba, Canada R3T 5V6

Abstract. A functional is associated to general systems of linear algebraic equations such that, for any point defined by an arbitrary set of unknowns, the position along its gradient which is the nearest to the system solution point can be determined directly from the value of the functional and of the gradient. Remarkably, the distance along the gradient to the point closest to the solution point is equal to the ratio of the value of the functional to the magnitude of its gradient at the starting point. By means of simple matrix transformations, this property can be employed to construct independent equations satisfied only by a part of the original system unknowns.

Relationship between Functional, Gradient and Solution

Consider a linear system of \( n \) algebraic equations in \( n \) unknowns

\[ Ax = b \]  

where the nonsingular matrix \( A \) and the vector \( b \) have real entries \( a_{ij} \) and \( b_i \), respectively, \( i, j = 1,2, \ldots, n \).

Let us associate to (1) a quadratic functional

\[ f(x) = (Ax - b)^T (Ax - b) \]

\[ = \sum_{i=1}^{n} (a_{ii}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i)^2 \]  

where \( x^T = (x_1, x_2, \ldots, x_n) \) and \( T \) indicates the transpose. \( f(x) = \text{const} \geq 0 \) represents a family of ellipsoids in an \( n \)-dimensional Euclidean space, with a common centre at the point whose Cartesian coordinates are equal to the unknowns of (1), \( x_j = x_j^{(s)} \), \( j = 1,2, \ldots, n \) (see Fig. 1) [1,2], \( f(x) = 0 \) representing the ellipsoid degenerated to the centre point.

The gradient of \( f \) at an arbitrary point of position vector \( x \) is \( 2g(x) \), where

\[ g(x) = A^T (Ax - b) \]

whose components are

\[ g_k(x) = \frac{1}{2} \frac{\partial f(x)}{\partial x_k} \]

\[ = \sum_{i=1}^{n} a_{ik} (a_{ii}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i), \quad k = 1,2, \ldots, n \]  

A point along the direction of the gradient taken at an arbitrary point and the solution point \( P_0 \) of position vector \( x^{(o)} \) is defined by
\[ x = x^{(o)} + \mu g^{(o)}, \quad g^{(o)} = g(x^{(o)}) \]  

(5)

where \( \mu \) is a scalar. The distance between this point and the solution point \( x^{(S)} \) is given by the Euclidean norm

\[
\| x - x^{(S)} \| = \| x^{(o)} - x^{(S)} + \mu g^{(o)} \|
\]

(6)

Fig. 1. \( n \)-dimensional ellipsoid \( f(x) = f_o \)

The point \( P_m \) for which this distance is minimum is obtained from the derivative of the square of the norm in (6),

\[
\frac{d}{d\mu} \left[ (x^{(o)} - x^{(S)} + \mu g^{(o)})^T (x^{(o)} - x^{(S)} + \mu g^{(o)}) \right] = 0
\]

(7)

which gives

\[
\mu = \frac{\left( x^{(o)} - x^{(S)} \right)^T g^{(o)}}{\| g^{(o)} \|^2}
\]

(8)

whose numerator, with (3) and (1), is just \( f(x^{(o)}) \),

\[
g^{(o)T} (x^{(o)} - x^{(S)}) = (Ax^{(o)} - b)^T A(x^{(o)} - x^{(S)})
\]

\[
= (Ax^{(o)} - b)^T (Ax^{(o)} - b) = f_o, \quad f_o = f(x^{(o)})
\]

(9)

Thus, the point \( P_m \) has a position vector

\[
x^{(m)} = x^{(o)} - \frac{f_o}{\| g^{(o)} \|^2} g^{(o)}
\]

(10)

its components being the coordinates of \( P_m \), i.e.
\[ x^{(m)}_k = x^{(o)}_k - \frac{f^{(o)}}{g^{(o)}} g^{(o)}_k, \quad k = 1, 2, \ldots, n \]  

(11)

The distance between \( P_o \) and \( P_m \) is

\[ \| x^{(m)} - x^{(o)} \| = \frac{f^{(o)}}{g^{(o)}} \]  

(12)

Therefore, the following theorem holds: "for an arbitrary \( x^{(o)} \), the point along the gradient (3) of the functional (2), associated with the linear system (1), which is the closest to the solution of (1) is given by (10)-(11) in terms of only the values of the functional and its gradient at \( x^{(o)} \); the distance between this point and \( x^{(o)} \) is equal to the ratio of the functional to the magnitude of the gradient at \( x^{(o)} \)."

In a geometric interpretation (see Fig. 1), \( P_o \) is an arbitrary point in space, \( f(x) = f^{(o)} \) is the \( n \)-dimensional ellipsoid passing through \( P_o \), the direction of the gradient is the direction of the normal to the ellipsoid at \( P_o \), and \( P_m \) is the point along the normal whose distance from the ellipsoid centre is minimum.

**Application to the transformation of the system (1)**

Each equation of (1),

\[ a_{ij} x_i + a_{i2} x_2 + \cdots + a_{in} x_n = b_i, \quad i = 1, 2, \ldots, n \]  

(13)

represents an \( n \)-dimensional hyperplane passing through the solution point \( P_S \), of position vector \( x^{(S)} \). A normal direction to the hyperplane is given by a vector of components \( a_{ij}, j = 1, 2, \ldots, n \). From the previous Section, it is obvious that an hyperplane through \( P_m \) perpendicular to the direction \( P_0 P_m \) of the normal to the ellipsoid at \( P_o \) will pass through its centre \( P_S \), i.e. through the solution point of (1). The equation of this hyperplane is written with the components of the gradient in (5) as

\[ g^{(o)}_{i1} x_1 + g^{(o)}_{i2} x_2 + \cdots + g^{(o)}_{in} x_n = \sum_{k=1}^{n} g^{(o)}_{ik} x^{(m)}_k \]  

(14)

Such equations can be used to replace equations in the original system (1). When the gradient \( g^{(o)} \) has some components equal to zero, then the corresponding equations (14) contain only a part of the system unknowns. If the gradient were to have only one component, say \( g^{(o)}_t \) then the equation (14) would give the unknown \( x^{(S)}_t \) of (1). A general transformation to yield the latter situation is not available.

A technique of constructing equivalent equations with a reduced number of unknowns consists, for instance, in the following steps. First, the original system (1) is augmented to an \( (n+1) \)-by-(\( n+1 \)) system in the form

\[
\begin{bmatrix}
A & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
1
\end{bmatrix} =
\begin{bmatrix}
b \\
1
\end{bmatrix}
\]  

(15)

which has the same first \( n \) unknowns as (1). Second, both sides of (15) are multiplied by
\[ Y = \begin{bmatrix} I & y \\ o & 1 \end{bmatrix} \]

where \( I \) is the unit matrix and \( y^{(p)} \) is a vector with \( p \) components different from zero, for instance,

\[ y = \begin{pmatrix} y_1, y_2, \ldots, y_p, 0, 0, \ldots, 0 \end{pmatrix}^T \]

Such that the original system (1) is transformed in

\[ A'x' = b' \]

with

\[ A' = \begin{bmatrix} A & y \\ o & 1 \end{bmatrix}, \quad x' = \begin{bmatrix} x \\ 1 \end{bmatrix}, \quad b' = \begin{bmatrix} b + y \\ 1 \end{bmatrix} \]

Choosing \( x'^{(o)} \) to be the origin, \( x'^{(o)} = 0 \), given the functional and its gradient associated with (19), i.e.

\[ f'_0 = \|b\|^2 \]

and

\[ g'^{(o)} = -A'^T b' = \begin{bmatrix} a_{11}y_1 + a_{21}y_2 + \cdots + a_{p1}y_p + h_1 \\ a_{12}y_1 + a_{22}y_2 + \cdots + a_{p2}y_p + h_2 \\ \vdots \\ a_{1n}y_1 + a_{2n}y_2 + \cdots + a_{pn}y_p + h_n \\ \sum_{k=1}^{p} y_k^2 + \sum_{k=1}^{p} b_k y_k + 1 \end{bmatrix} \]

where

\[ h_j = \sum_{i=1}^{n} a_{ij} b_i, \quad j = 1, 2, \ldots, n \]

In a third step, \( p \) components of \( g'^{(o)} \), for instance the first \( p \), are made zero by solving a \( p \)-by-\( p \) system of equations in \( y_1, y_2, \ldots, y_p \). The other components of \( g'^{(o)} \) are, in general, different from zero and the reduced equation (14) becomes now

\[ g_{p+1}^{(o)} x_{p+1} + g_{p+2}^{(o)} x_{p+2} + \cdots + g_n^{(o)} x_n = -\left( \|b\|^2 + \sum_{k=1}^{p} b_k y_k \right) \]

This is an equation in \( n-p \) unknowns which can replace on of the original system (1). Similarly, by making zero other sets of \( p \) components of (21), one can construct new independent equations in \( n-p \) unknowns to replace other equations in (1). As well, various reduced independent equations of (1) can be generated by changing the position of the \( p \) entries different from zero in \( y \) (see (17)) and by changing \( p \).
Conclusion

A functional property of general linear systems of algebraic equations is identified, which allows the construction of system equations with a reduced number of original unknowns by using simple matrix transformations. Each reduced equation is obtained by solving a smaller system of equations with coefficients which are just those in the original system.

References
