



Polarity for quadratic hypersurfaces and conjugate gradient method: Relation between degenerate and nondegenerate cases

Fasano Giovanni, Giove Silvio, and Gusso Riccardo

Citation: [AIP Conference Proceedings](#) **1776**, 090031 (2016); doi: 10.1063/1.4965395

View online: <http://dx.doi.org/10.1063/1.4965395>

View Table of Contents: <http://scitation.aip.org/content/aip/proceeding/aipcp/1776?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[A comparative study of two new conjugate gradient methods](#)

AIP Conf. Proc. **1643**, 616 (2015); 10.1063/1.4907502

[A new nonlinear conjugate gradient method](#)

AIP Conf. Proc. **1643**, 577 (2015); 10.1063/1.4907497

[The global convergence properties of a conjugate gradient method](#)

AIP Conf. Proc. **1602**, 286 (2014); 10.1063/1.4882501

[Observation of Different Transient Absorptions Between Single and Multilayer Graphene from Non-degenerate Pump-probe Spectroscopy](#)

AIP Conf. Proc. **1399**, 781 (2011); 10.1063/1.3666609

[Mode selectivity and misalignment sensitivity in non-degenerate phase-conjugators](#)

AIP Conf. Proc. **146**, 755 (1986); 10.1063/1.35803

Polarity for Quadratic Hypersurfaces and Conjugate Gradient Method: Relation between Degenerate and Nondegenerate Cases

Fasano Giovanni^{1,a)}, Giove Silvio^{2,b)} and Gusso Riccardo^{3,c)}

¹Department of Management, University Ca'Foscari of Venice, S.Giobbe Cannaregio 873, 30121, Venice, Italy

²Department of Economics, University Ca'Foscari of Venice, S.Giobbe Cannaregio 873, 30121, Venice, Italy

³Department of Management, University Ca'Foscari of Venice, S.Giobbe Cannaregio 873, 30121, Venice, Italy

^{a)}Corresponding author: fasano@unive.it, <http://venus.unive.it/~fasano>

^{b)}sgiove@unive.it

^{c)}rgusso@unive.it

Abstract. In this paper we consider a geometric viewpoint to analyze the behaviour of the Conjugate Gradient (CG) method, for the solution of a symmetric linear system, when at current step a pivot breakdown possibly occurs (*degenerate* case). As well known this can occur when the system matrix is indefinite or singular. In the latter case the CG gets stuck, since the steplength along the current search direction cannot be computed. We show here that a simple geometric interpretation can be provided for the degenerate case, as long as some basics on projective geometry in the Euclidean space are considered.

INTRODUCTION

The CG method is a very successful iterative procedure [1] for the solution of the linear system

$$Ay = b, \quad A = A^T. \quad (1)$$

In particular, in optimization frameworks such linear systems arise within a large class of applications, ranging from unconstrained to constrained problems, where convex and nonconvex functions are involved. As well known, the CG generates the sequence of approximate solutions $\{y_k\}$ of (1), and may conversely experience a premature stop in the indefinite case, whenever a pivot breakdown occurs. I.e., in case at step k the current search direction p_k satisfies $p_k^T A p_k = 0$, then the steplength along p_k cannot be computed and the CG halts. This implies that the current approximate solution y_k is possibly far from being a stationary point of the quadratic functional $g(y) = 1/2y^T A y - b^T y$, associated with the linear system (1).

Here we want to briefly study the possible degeneracy of the CG from a geometric standpoint, when considering either a singular or an indefinite nonsingular matrix A in (1). Observe that in case the matrix A is *positive semidefinite*, then all the considerations reported in [2], [3] and [4] hold, including the fact that the sequence $\{y_k\}$ generated by the CG will converge to the solution $y^* = A^+ b$ (where A^+ is the Moore-Penrose generalized inverse of A), provided that the vector $b - A y_1$ (initial residual) has a nonzero projection on all the eigenvectors associated with distinct positive eigenvalues of A . In the latter case, though A is possibly singular, then the standard geometry associated with the CG iterations in the positive definite case applies. Equivalently, at step k of the CG the Ritz-Galerkin condition

$$0 = (b - A y_{k+1})^T p_k = 0 \quad (2)$$

is fulfilled, implying that the hyperplane π_{k+1} of equation

$$\pi_{k+1} : \{y \in \mathbb{R}^n : (b - A y)^T p_k = 0\} \quad (3)$$

is considered, which is both tangent at $y_{k+1} \in \mathbb{R}^n$ to the hypersurface \mathcal{G} given by

$$\mathcal{G} : \left\{ y \in \mathbb{R}^n : \frac{1}{2}y^T Ay - b^T y - \left(\frac{1}{2}y_{k+1}^T Ay_{k+1} - b^T y_{k+1} \right) = 0 \right\}, \quad (4)$$

and has the normal vector

$$n = \frac{b - Ay_{k+1}}{\|b - Ay_{k+1}\|_2}.$$

On the contrary, in case the matrix A is indefinite, we show that some concepts from projective geometry need to be invoked, in order to more precisely address the behaviour of the CG. In this regard, following [5] (see also [6]) we recall that to any n -dimensional real vector $y \in \mathbb{R}^n$ we can associate the $(n+1)$ -tuple $(\rho x^1, \dots, \rho x^n, \rho x^0) \in \mathbb{R}^{n+1}$ of *homogeneous coordinates*, such that

$$y^i = \frac{\rho x^i}{\rho x^0}, \quad i = 1, \dots, n, \quad (5)$$

where $\rho \neq 0$ and $(x^1, \dots, x^n, x^0) \neq 0$. We highlight that homogeneous coordinates allow to use a simple algebra to handle points at infinite. In particular, observe that the line

$$\ell : \{y \in \mathbb{R}^n : y = \bar{y} + \alpha p, \quad \bar{y}, p \in \mathbb{R}^n, \alpha \in \mathbb{R}\},$$

whose directional cosines are proportional to $p = (p^1, \dots, p^n)$, from the point \bar{y} , has the unique point at infinite $(p^1, \dots, p^n, 0)^T \in \mathbb{R}^{n+1}$ in homogeneous coordinates.

We remark that resorting to homogeneous coordinates, in place of Cartesian coordinates, often provides a very powerful tool in computational methods. Examples of applications where homogeneous coordinates are widely adopted to simplify the analysis are given by *Robotics* and *3D graphics*. In particular in robotics, homogeneous coordinates allow to use a single matrix in order to represent both affine and projective transformations. As regards 3D graphics, homogeneous coordinates allow to represent translations with matrices, so that massive matrix operations are easily performed. We recall that in this paper $\text{Ker}(A)$ indicates the null space of matrix A ; moreover, for $v = (v^1, \dots, v^n) \in \mathbb{R}^n$ and $v^0 \in \mathbb{R}$, for the sake of brevity we indicate $(v, v^0)^T \equiv (v^1, \dots, v^n, v^0) \in \mathbb{R}^{n+1}$.

DEFINITIONS, BASICS ON POLARITY AND FIRST RESULTS

Given the symmetric linear system (1), we also consider as a reference the quadratic functional

$$g(y) = \frac{1}{2}y^T Ay - b^T y + c, \quad c \in \mathbb{R}, \quad (6)$$

and replacing (5) in (6) we obtain the functional $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ in homogeneous coordinates defined, for $x^0 \neq 0$, by

$$f(x^1, \dots, x^n, x^0) = g\left(\frac{x}{x^0}\right) = \frac{1}{2}\left(\frac{x}{x^0}\right)^T A \left(\frac{x}{x^0}\right) - b^T \left(\frac{x}{x^0}\right) + c. \quad (7)$$

Finally, the functional $f(x^1, \dots, x^n, x^0)$ can be associated with the *quadratic hypersurface* \mathcal{F} given by

$$\mathcal{F} := \left\{ (x, x^0)^T \in \mathbb{R}^{n+1} : f(x, x^0)(x^0)^2 = 0 \right\} = \left\{ (x, x^0)^T \in \mathbb{R}^{n+1} : x^T Ax - 2(x^0)b^T x + 2c(x^0)^2 = 0 \right\}, \quad (8)$$

where possibly the value $x^0 = 0$ can be considered, corresponding to points at infinite of the hypersurface \mathcal{F} .

Definition 1 Given the quadratic hypersurface (8), in homogeneous coordinates, given the point $P = (\bar{x}^1, \dots, \bar{x}^n, \bar{x}^0) \in \mathbb{R}^{n+1}$, the equation

$$\sum_{i=0}^n \frac{\partial \mathcal{F}(x^1, \dots, x^n, x^0)}{\partial x^i} \bar{x}^i = 0 \quad (9)$$

represents an hyperplane, which is said to be the first polar (or **polar hyperplane**) of the point P with respect to the hypersurface (8), in homogeneous coordinates. Moreover, the point P is the **pole** of (9). Finally, the pole of the hyperplane at infinite $x^0 = 0$, with respect to (8), is the **center** of \mathcal{F} .

We immediately realize that since $\mathcal{F}(x^1, \dots, x^n, x^0) = 0$ is a quadratic hypersurface (i.e. $\partial\mathcal{F}(x^1, \dots, x^n, x^0)/\partial x^i$ is linear) then we have from (9)

$$\sum_{i=0}^n \frac{\partial\mathcal{F}(x^1, \dots, x^n, x^0)}{\partial x^i} \cdot \bar{x}^i = \sum_{i=0}^n \frac{\partial\mathcal{F}(x^1, \dots, x^n, x^0)}{\partial x^i} \Big|_{x^i=\bar{x}^i} \cdot x^i = 0. \quad (10)$$

In other words, the first polar of the point $(\bar{x}^1, \dots, \bar{x}^n, \bar{x}^0)$, with respect to the quadratic hypersurface (8), coincides with the first polar of the point (x^1, \dots, x^n, x^0) , with respect to the same quadratic hypersurface, in the homogeneous coordinates $(\bar{x}^1, \dots, \bar{x}^n, \bar{x}^0)$. As shown in [5], in case A is nonsingular the next result can be proved.

Proposition 2 [Equivalence of center] *Consider the quadratic hypersurface \mathcal{F} , with A nonsingular, $c \neq 1/2b^T A^{-1}b$, and center $(x_*, x_*^0)^T \in \mathbb{R}^{n+1}$. Then $x_*^0 = 1/(4c - 2b^T A^{-1}b)$ and the vector $z_* = (x_*/x_*^0)$ is the unique solution of the linear system (1).*

The latter proposition reveals that, under the nonsingularity assumption (along with the technical condition $c \neq 1/2b^T A^{-1}b$), there is a one-to-one correspondence between the solution of the linear system (1) and the pole of the hyperplane at infinite $x^0 = 0$ with respect to \mathcal{F} . Moreover, in [5] we also proved that the CG iteratively generates in Cartesian coordinates a sequence of points $\{y_k\} \rightarrow y_*$ such that $y_* = x_*/x_*^0$, without explicitly recurring to homogeneous coordinates. Here we want to show that, when A is singular, the use of homogeneous coordinates is mandatory in order to describe the geometry behind CG iterations. To the latter purpose the next generalization to Proposition 2 is given.

Proposition 3 [Equivalence of center, A singular] *Consider the quadratic hypersurface \mathcal{F} , with A singular and $c \neq 1/2b^T(A^+b + v)$, for any $v \in \text{Ker}(A)$. There exists the family of generalized centers $(x_*(v), x_*^0(v))^T$ of \mathcal{F} , for any $v \in \text{Ker}(A)$, such that $x_*^0(v) = 1/[4c - 2b^T(A^+b + v)] \neq 0$, and the vector $z_*(v) = x_*(v)/x_*^0(v)$ is a solution of the linear system (1).*

Proof (sketch)

Since A is now singular, the one-to-one correspondence between points (i.e. poles) and hyperplanes (i.e. polar hyperplanes) with respect to the quadratic hypersurface \mathcal{F} is lost. Nevertheless, for any $v \in \text{Ker}(A)$, applying the Definition 1 we can say that the hyperplane $x^0 = 0$ is the polar hyperplane of $(x_*(v), x_*^0(v))^T$ with respect to \mathcal{F} if

$$\begin{pmatrix} 2A & -2b \\ -2b^T & 4c \end{pmatrix} \begin{pmatrix} x_*(v) \\ x_*^0(v) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (11)$$

Since A is singular and by the hypotheses $c \neq 1/2b^T(A^+b + v)$, for any $v \in \text{Ker}(A)$, then the point of \mathbb{R}^n $z_*(v) = x_*(v)/x_*^0(v) = A^+b + v$ is a solution of (11) with

$$\frac{1}{x_*^0(v)} = 4c - 2b^T [A^+b + v] \neq 0.$$

□

Further Geometric Considerations When a Pivot Breakdown Occurs

Clearly Proposition 3 reduces to Proposition 2 when A is nonsingular (i.e. $\text{Ker}(A) = \{0\}$ and $A^+ = A^{-1}$), showing that the correspondence between the center of the quadratic hypersurface \mathcal{F} and the solution of (11) can be easily generalized. In particular, as reported in the first section, in optimization frameworks when A is singular the CG is often applied for the solution of (1), providing the result $y_* = A^+b$. Thus, according with Proposition 3, when A is singular the CG equivalently provides in \mathbb{R}^n a solution of (11) and indirectly computes one of the possible generalized centers of the quadratic hypersurface \mathcal{F} . Then, using both Propositions 2 and 3, along with relation (2), we are now ready to better detail in homogeneous coordinates the case of a possible pivot breakdown of the CG.

Indeed, suppose that starting from the initial approximate solution $y_1 \in \mathbb{R}^n$ of (1) the CG has already generated the search directions p_1, \dots, p_{k-1} (satisfying the well known conjugacy conditions $p_i^T A p_j = 0, 1 \leq i \neq j \leq k-1$),

along with the corresponding steplengths $\alpha_1, \dots, \alpha_{k-1}$, so that $y_k = y_1 + \sum_{i=1}^{k-1} \alpha_i p_i$. Then, setting $r_k = b - Ay_k$, at step k relation (2) imposes that the steplength α_k satisfies the relation

$$[b - A(y_k + \alpha_k p_k)]^T p_k = r_k^T p_k - \alpha_k p_k^T A p_k = 0 \quad \iff \quad \alpha_k = \frac{r_k^T p_k}{p_k^T A p_k},$$

which might not be well posed in case A is indefinite or if A is positive semidefinite, being possibly $p_k^T A p_k = 0$ and $y_{k+1} = y_k + \alpha_k p_k$ a point at infinite. In the latter case the tangent hyperplane (3) to the quadratic hypersurface (4) at y_{k+1} is not defined, though we can still address in homogeneous coordinates the polar hyperplane of \mathcal{F} in (8). Indeed, in homogeneous coordinates the vector $p_k \in \mathbb{R}^n$ corresponds to the point at infinite $(p_k, 0)^T \in \mathbb{R}^{n+1}$ of the line $\ell_{k+1} : \{y \in \mathbb{R}^n : y = y_k + \alpha p_k, \alpha \in \mathbb{R}\}$, which is also a point satisfying the equation of \mathcal{F} (being $p_k^T A p_k = 0$). At the latter point the polar hyperplane of \mathcal{F} is defined. Thus, we can summarize that though at step k of the CG the pivot breakdown $p_k^T A p_k = 0$ occurs, then

- we can transform the quadratic functional $g(y)$ in (6) (in Cartesian coordinates) into the quadratic hypersurface \mathcal{F} in (8) (in homogeneous coordinates);
- we can compute the polar hyperplane of the point $(p_k, 0)^T \in \mathcal{F}$ with respect to \mathcal{F} , so that it represents a generalization of the tangent hyperplane π_{k+1} (in (3)), at the point at infinite $(p_k, 0)^T \in \mathbb{R}^{n+1}$.

Also observe that, as long as $y_{k+1} \in \mathbb{R}^n$ (i.e. y_{k+1} is a finite point in Cartesian coordinates), it is possible to represent it in homogeneous coordinates as $y_{k+1} = x_{k+1}/x_{k+1}^0$, with $x_{k+1}^0 \neq 0$, and compute both the tangent hyperplane π_{k+1} to \mathcal{G} at y_{k+1} , and the polar hyperplane $\tilde{\pi}_{k+1}$ of \mathcal{F} at $(x_{k+1}, x_{k+1}^0)^T$. It is not difficult (see also [5] and [7]) to verify that, since polar hyperplanes in homogeneous coordinates represent generalizations of tangent hyperplanes in Cartesian coordinates, then if $y_{k+1} \in \mathbb{R}^n$ (i.e. y_{k+1} is a finite point), the equivalent expression in Cartesian coordinates of $\tilde{\pi}_{k+1}$ coincides with π_{k+1} .

ACKNOWLEDGMENTS

G.Fasano thanks the National Research Council - Marine Technology Research Institute (CNR-INSEAN), Italy, for the support received. The work of G. Fasano is partially supported by the Italian Flagship Project RITMARE, coordinated by the Italian National Research Council (CNR) and funded by the Italian Ministry of Education, within the National Research Program 2012-2016.

REFERENCES

- [1] M. Hestenes and E. Stiefel, *J. Research Nat. Bur. Standards* **49**, 409–436 (1952).
- [2] G. Fasano, *J. of Optim. Theory and Applic.* **132**, 267–285 (2007).
- [3] G. Fasano, *J. of Optim. Theory and Applic.* **164**, 883–914 (2015).
- [4] G. Fasano and S. Lucidi, *Optimization Letters* **3**, 521–535 (2009).
- [5] G. Fasano and R. Pesenti, submitted as *Part I* to *J. of Optim. Theory and Applic.* (2016).
- [6] A. Seidenberg, *Lecture in projective geometry* (Dover Publications, Inc., Mineola, NY, 2005).
- [7] G. Fasano and R. Pesenti, submitted as *Part II* to *J. of Optim. Theory and Applic.* (2016).