Polarity and Conjugacy for Quadratic Hypersurfaces: a unified framework with recent advances

Giovanni Fasano*, Raffaele Pesenti†

Abstract. We aim at completing the analysis in [1] for quadratic hypersurfaces, where the geometric viewpoint suggested by the Polarity theory is considered, in order to recast basic properties of the Conjugate Gradient (CG) method [2]. Here, the focus is on possibly exploiting theoretical advances on nonconvex quadratic hypersurfaces, in order to address guidelines for efficient optimization methods converging to second order stationary points, in large scale settings. We first recall some results from [1], in order to fully analyze the relationship between the CG and the Polarity theory. Then, we specifically address, from a different perspective, the geometric insight of the pivot breakdown, which might occur when solving a nonsingular indefinite Newton's equation applying the CG. Furthermore, we fully exploit some novel theoretical advances of the Polarity theory on nonconvex quadratic hypersurfaces not considered in [1]. Finally, we show that our approach describes a general framework, which also encompasses a class of CG–based methods, namely Planar CG–based methods. The framework we consider intends to emphasize a bridge between the geometry behind stationary points of nonconvex quadratic hypersurfaces and their efficient computation using Krylov–subspace methods.

Keywords: Polarity in homogeneous coordinates, Nonconvex Quadratic hypersurfaces, Conjugate Gradient method, Indefinite linear systems.

1 Introduction

This paper deals with the Polarity theory in homogeneous coordinates [3, 4, 5]. We show that it can be exploited to explain both the standard behavior and a possible pivot breakdown of the Conjugate Gradient (CG) method, on symmetric indefinite linear systems. Our results are based on and complement the ones in [1]. In particular, with respect to [1], we prove that the use of the polarity theory is a fruitful tool to interpret several geometric properties of the class of Krylov-subspace methods known as Planar CG–based methods [6, 7, 8]. By fully exploiting conjugacy among vectors associated to indefinite matrices, we also prove some novel properties of the Polarity theory applied to nonconvex quadratic hypersurfaces (see Section 4).

The CG–based methods, and in general Krylov–subspace methods, play a keynote role in many theoretical and real-world applications of optimization. As well known, these methods are used to

* Department of Management, University Ca' Foscari of Venice, Venice (ITALY), <fasano@unive.it>.
† Department of Management, University Ca' Foscari of Venice, Venice (ITALY), <pesenti@unive.it>.

solve large symmetric, possibly indefinite, linear systems and to gain information on the system matrix eigenvalues.

As a further application, they have been successfully employed also within nonconvex optimization frameworks, in order to iteratively individuate second order stationary points of a given objective function [9]. Specifically, at each iteration, CG–based methods can be used to refine the computation of suitable search directions, namely *negative curvature directions*, which are essential to detect stationary points satisfying second order necessary optimality conditions (see the rigorous approach in [10]). Considering large scale settings, such directions are of difficult computation (see, e.g., [11], since they typically require to (see, e.g., [12])

- *implicitly decompose* the Hessian matrix of the objective function,
- *store large matrices* or perform burdensome calculations, due to some *re–computing*,
- *explicitly compute* an underestimation of the least eigenvalue of the Hessian matrix.

Hence, effective iterative methods which may at once solve Newton's equation and efficiently assess negative curvature directions, in large scale optimization, are definitely of dramatic impact. In this regard, Lanczos–based and CG–based methods are often the Krylov–subspace methods of choice.

In this paper we investigate CG–based methods from the geometric perspective suggested both by Polarity theory and by some recent advances in [13]. Polarity theory commutes from Cartesian coordinates to homogeneous coordinates, in order to deal with points at infinity. We refer the reader to [1], for some basics on homogeneous coordinates and Polarity theory. In the present work, we just recall the *Section Theorem* and *Reciprocity Theorem*, since they are used throughout the paper.

Resorting to homogeneous coordinates may provide a powerful tool in computational methods. Examples of applications are in *Robotics* and in *3D graphics*. In robotics, homogeneous coordinates allow to use a single matrix to represent both affine and projective transformations. Hence, a single matrix is sufficient to define both a rotation and a translation of a vector (see e.g. [14]). In 3D graphics, homogeneous coordinates allow to unify both translations and the division by depth in perspective projections, so that the massive computations that are usually needed in this area can be more efficiently performed [15].

In [1], we point out that the Polarity theory in homogeneous coordinates is helpful to describe a precise relation among algebraic hypersurfaces. In particular, we stress the importance of handling points at infinity, which play a significant role in addressing possible failures of CG–based methods on non-convex quadratics. We also stress that an investigation of CG–based methods in homogeneous coordinates can contribute to achieve an additional insight in their behavior. Here, we reinforce the above results by carrying on a complete analysis on non-convex quadratics.

The paper is organized as follows. Section 2 reports some basics on Polarity for algebraic hypersurfaces. Sections 3 and 4 explicitly analyze how conjugate directions and polar hyperplanes, in both homogeneous and Cartesian coordinates, play a keynote role to provide a unified geometric perspective for the CG–based methods. In particular, geometric insights on a possible pivot breakdown of the CG method in the indefinite case are included in Section 3.1 and in the Appendix. Section 5, together with Sections 3 and 4, encompasses the advances with respect to [1]. It uses the Polarity theory to analyze the Planar CG–based methods, which represent extensions of the CG method, and have been proposed in the literature of Krylov–subspace methods. Finally, a section of Conclusions indicates future guidelines for further investigation.

In this paper we use the following notation. We represent the Euclidean norm with $\| \cdot \|$. We indicate the n-dimensional Cartesian space with $\mathbb{R}^n$ and (to simplify the notation with respect to

[1]) the associated homogeneous coordinates projective space with $\mathbb{R}^{n+1}$. Given the vector $x \in \mathbb{R}^n$ and the scalar $x_0 \in \mathbb{R}$, we indicate with $(x, x_0)^T$ a vector in $\mathbb{R}^{n+1}$. We use lowercase Greek letters to represent hyperplanes, either in Cartesian or homogeneous coordinates. For the sake of brevity, we treat the terms *hyperplane* and *linear manifold* as synonyms, then a linear manifold that includes the origin represents a *linear subspace*. Finally, $A \succ 0$ indicates that the symmetric matrix A is positive definite.

2 Basics on Polarity and on quadratic hypersurfaces

In this section, we report some definitions and two fundamental results of Polarity and quadratic hypersurfaces: we refer the reader to [1] for further details. These definitions and results address the relation between Polarity for quadratic hypersurfaces [3, 5], in homogeneous coordinates, and both the solutions of symmetric nonsingular linear systems and the stationary points of quadratic functionals.

Let $A = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ be a symmetric and nonsingular matrix, $b = (b_1, \ldots, b_n)^T \in \mathbb{R}^n$ be a vector, and $c \in \mathbb{R}$ be a scalar. Then, throughout the paper, we use the linear system

$$Ay = b \tag{1}$$

as a reference problem. In addition, we let $g : \mathbb{R}^n \to \mathbb{R}$

$$g(y) = \frac{1}{2} y^T Ay - b^T y + c = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} y_i y_j - \sum_{i=1}^{n} b_i y_i + c \tag{2}$$

represent the prototype of a quadratic functional in Cartesian coordinates and $f : \mathbb{R}^{n+1} \to \mathbb{R}$ the quadratic functional in homogeneous coordinates associated to g such that, for any $x_0 \neq 0$,

$$f(x, x_0) = g\left(\frac{x}{x_0}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(\frac{x_i}{x_0}\right)\left(\frac{x_j}{x_0}\right) - \sum_{i=1}^{n} b_i \left(\frac{x_i}{x_0}\right) + c.$$

Furthermore, we also use

$$\mathcal{F} \doteq \{(x, x_0)^T \in \mathbb{R}^{n+1} : f(x, x_0)x_0^2 = 0\} \equiv \{(x, x_0)^T \in \mathbb{R}^{n+1} : \frac{1}{2} x^T Ax - b^T x x_0 + c x_0^2 = 0\} \tag{3}$$

to indicate the prototype of a quadratic hypersurface, and

$$\mathcal{C}_\infty \doteq \mathcal{F} \cap \{(x, x_0)^T \in \mathbb{R}^{n+1} \ : \ x_0 = 0\} \equiv \{(x, 0)^T \in \mathbb{R}^{n+1} \ : \ x^T Ax = 0\} \tag{4}$$

to represent the intersection between $\mathcal{F}$ and the hyperplane at infinity. We recall that $x_0 = 0$ represents in homogeneous coordinates the hyperplane at infinity, i.e., the locus of all the points at infinity of an n-dimensional Cartesian space $\mathbb{R}^n$.

Throughout the paper, we also hold true the next assumption.

Assumption 2.1 *Let A, b, c be respectively the matrix, the vector and the scalar that define the quadratic hypersurface $\mathcal{F}$ in* (3)*; then the matrix*

$$\begin{pmatrix} A & -b \\ -b^T & 2c \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

is nonsingular.

Assumption 2.1 is necessary to guarantee that the Polarity theory defines a one-to-one correspondence between points and hyperplanes in homogeneous coordinates, with respect to $\mathcal{F}$ in (3). This assumption also allows us to exploit the properties of $\mathcal{F}$ at points at infinity: this helps to describe several properties of CG–based methods, which are used to detect the stationary point of g in (2).

Next, we introduce the concept of *polar hyperplane* and summarize some fundamental results.

Definition 2.1 *Consider a quadratic hypersurface $\mathcal{F}$ and a point* (pole) $P = (\bar{x}, \bar{x}_0)^T \in \mathbb{R}^{n+1}$. *The hyperplane with equation*

$$\sum_{i=0}^{n} \frac{\partial f(x_1, \ldots, x_n, x_0)}{\partial x_i} \bar{x}_i = 0 \tag{5}$$

is said to be the (first) polar hyperplane *of the point P with respect to $\mathcal{F}$, in homogeneous coordinates.*

In particular, the above definition implies that if a finite point P belongs to $\mathcal{F}$, then the polar hyperplane of P with respect to $\mathcal{F}$ coincides with the *tangent hyperplane* of $\mathcal{F}$ in P.

Theorem 2.1 [Reciprocity Theorem] *Consider a quadratic hypersurface $\mathcal{F}$ and points P, $Q \in \mathbb{R}^{n+1}$. If the polar hyperplane of P with respect to $\mathcal{F}$ includes Q, then the polar hyperplane of Q with respect to $\mathcal{F}$ includes P.*

An application of the Reciprocity Theorem is given in Figure 1, both for the case when the matrix A characterizing the quadratic hypersurface $\mathcal{F}$ is positive definite and indefinite.

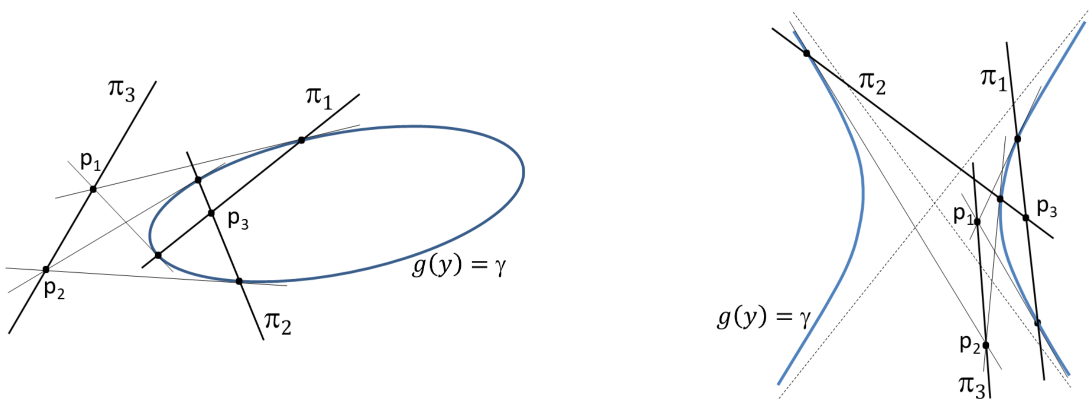

Figure 1: The point p_i is the pole (in Cartesian coordinates) of the polar hyperplane π_i, with respect to the quadratic hypersurface $g(y) = \gamma$. (*left*) Case in which $A \succ 0$, (*right*) case in which A is indefinite. Dashed lines represent the so-called asymptotic cone, see Definition 2.5.

Theorem 2.2 [Section Theorem] *Consider a quadratic hypersurface $\mathcal{F}$, a subspace $V_d \in \mathbb{R}^{n+1}$ of dimension $d \leq n$, and let $\bar{\mathcal{F}} = \mathcal{F} \cap V_d$ be the intersection between $\mathcal{F}$ and V_d. If V_d is not included in $\mathcal{F}$, i.e., $\bar{\mathcal{F}} \neq V_d$, then for any point $P \in V_d$, the intersection of V_d with the polar hyperplane of P with respect to $\mathcal{F}$ coincides with the polar hyperplane of P with respect to $\bar{\mathcal{F}}$.*

An application of the Section Theorem is given in Figure 2.

Definition 2.2 *Consider a quadratic hypersurface $\mathcal{F}$ and a point $P \in \mathbb{R}^{n+1}$. Point P is* self–conjugate *with respect to $\mathcal{F}$ if P belongs to its own polar hyperplane with respect to $\mathcal{F}$.*

It can be easily proven that the following result holds.

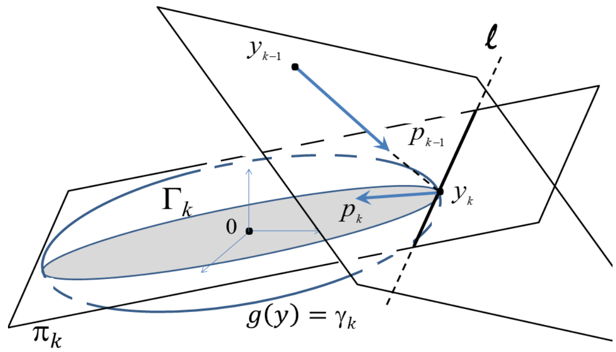

Figure 2: The *Section Theorem* in Cartesian coordinates, for the family of quadratic hypersurfaces (8). The line ℓ is the polar hyperplane of y_k with respect to Γ_k, being Γ_k the intersection between the hypersurface $g(y) = \gamma_k$ and the subspace π_k. The direction p_{k-1} is conjugate to the hyperplane π_k.

Proposition 2.1 *Consider a quadratic hypersurface $\mathcal{F}$ and a point $P \in \mathbb{R}^{n+1}$. Then, the next three conditions are equivalent.*

- *The point P is self–conjugate.*
- *The polar hyperplane of P coincides with the tangent hyperplane of $\mathcal{F}$ at P.*
- *The point P belongs to $\mathcal{F}$.*

The following definitions conclude this section.

Definition 2.3 *Consider a quadratic hypersurface $\mathcal{F}$ and a point $(x^*, x_0^*)^T \in \mathbb{R}^{n+1}$.*

- *The point $(x^*, x_0^*)^T$ is the* center *of $\mathcal{F}$ if it is the pole of the polar hyperplane at infinity $x_0 = 0$ with respect of $\mathcal{F}$.*
- *The hyperplanes through the center $(x^*, x_0^*)^T$ of $\mathcal{F}$ are the* diametral hyperplanes *of $\mathcal{F}$.*
- *The lines through the center $(x^*, x_0^*)^T$ of $\mathcal{F}$ are the* diameters *of $\mathcal{F}$.*

Observe that by the Reciprocity Theorem, any diametral hyperplane of $\mathcal{F}$ is the polar hyperplane of a point at infinity $(x, 0)^T \in \mathbb{R}^{n+1}$, for any $x \in \mathbb{R}^n$.

Definition 2.4 *Consider a quadratic hypersurface $\mathcal{F}$.*

- *Two diametral hyperplanes π_1 and π_2 are* conjugate *if π_i contains the pole of π_j, for $i, j \in \{1, 2\}$, $i \neq j$.*
- *Two diameters ℓ_1 and ℓ_2 are* conjugate *if the point at infinity $(d_i, 0)^T$ of ℓ_i is the pole of a diametral hyperplane which contains ℓ_j, for $i, j \in \{1, 2\}$, $i \neq j$.*
- *Two lines ℓ_1 and ℓ_2 are* conjugate *if they are respectively parallel to conjugate diameters.*
- *A diametral hyperplane π_1 is* conjugate *to a diameter ℓ_2 if any line ℓ_1 contained in π_1 and ℓ_2 are conjugate.*

- *A hyperplane π_1 is conjugate to a line ℓ_2 if π_1 and ℓ_2 are respectively parallel to a diametral hyperplane and a diameter that are conjugate.*

Definition 2.5 *Consider a quadratic hypersurface $\mathcal{F}$ and assume that Assumption 2.1 holds. The* asymptotic cone *of $\mathcal{F}$ is the set of all the lines connecting the center of $\mathcal{F}$ and any point of $\mathcal{C}_\infty$.*

3 The CG and Polarity theory: advances

Given the preliminary results in [1], in this section we further provide a perspective on the CG method, from the point of view of the Polarity theory.

Hereinafter, for the sake of simplicity, in place of considering the linear system (1), the quadratic functional (2) and the hypersurface (3), we respectively address the linear system $Ay = 0$, the quadratic functional

$$g(y) = \frac{1}{2} y^T A y, \tag{6}$$

and the hypersurface in homogeneous coordinates

$$\mathcal{F}_\gamma := \left\{ (x, x_0)^T \in \mathbb{R}^{n+1} : \frac{1}{2} x^T A x - \gamma x_0^2 = 0 \right\}. \tag{7}$$

In this setting, as A is nonsingular, $y^* = 0$ is the unique stationary point of $g(y)$; $(x^*, x_0^*)^T \equiv (0, -1/(4\gamma))^T$ is the center of $\mathcal{F}_\gamma$ and hence $y^* = 0 = x^*/x_0^*$ is also the center of the family of quadratic hypersurfaces

$$\frac{1}{2} y^T A y = \gamma, \qquad \gamma > 0. \tag{8}$$

The last positions do not introduce any loss of generality. Indeed, given the vector $\tilde{y} = -A^{-1}b$, we set

$$y = \hat{y} - \tilde{y}, \tag{9}$$

and observe that system (1) is equivalent to system $A\hat{y} = 0$. In addition, (2) becomes $\hat{g}(\hat{y}) = \frac{1}{2}\hat{y}^T A \hat{y} + \left(c - \frac{1}{2} b^T A^{-1} b\right)$, whose stationary point coincides with the one of (6) and whose associated quadratic hypersurfaces (2) correspond to (7) when $\gamma = \frac{1}{2} b^T A^{-1} b - c$.

In Table 1 we recall a general scheme for the CG method, in case the quadratic functional $g(y)$ in (6) is considered. We incidentally note (see also [1]) that the search directions p_i and p_j, with $i \neq j$, generated by the CG in Table 1, are such that the lines $x_i + \alpha p_i$ and $x_j + \beta p_j$, with $\alpha, \beta \in \mathbb{R}$, are conjugate as in Definition 2.4. Then, we report the next results from [1], because of their relevance for our analysis.

Proposition 3.1 [CG - Polar Hyperplane 1] *Let the CG method in Table 1 perform m steps to solve the linear system $Ay = 0$, with $A \succ 0$. Then, for every $k < m$, the linear manifold*

$$y_{k+1} + span\{p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_m\}$$

represents in Cartesian coordinates a diametral hyperplane of the hypersurface $\mathcal{F}_\gamma$ in (7), for any $\gamma > 0$. This diametral hyperplane is the polar hyperplane of the pole $(p_k, 0)^T$, with respect to $\mathcal{F}_\gamma$, and can be written as

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

□

The Conjugate Gradient (CG) method		
Description: Iterative method for solving equation $Ay = 0$ **Input:** Set an initial solution $y_0 \in \mathbb{R}^n$		
Step 0:	Compute the residual $r_0 = -Ay_0$, set $k = 0$. If $r_0 = 0$, then STOP. Else, set the search direction $p_0 = r_0$. Set $k = k + 1$.	
Step k:	Compute $\alpha_{k-1} = r_{k-1}^T p_{k-1} / p_{k-1}^T A p_{k-1}$. Update the point $y_k = y_{k-1} + \alpha_{k-1} p_{k-1}$, along with the residual $r_k = r_{k-1} - \alpha_{k-1} A p_{k-1}$. If $r_k = 0$, then STOP. Else, set $\beta_{k-1} = \|r_k\|^2 / \|r_{k-1}\|^2$. Update the search direction $p_k = r_k + \beta_{k-1} p_{k-1}$. Set $k = k + 1$, go to **Step** k.	

Table 1: The CG method for solving the symmetric linear system $Ay = 0$.

Proposition 3.2 [CG - Polar Hyperplane 2] *Let the CG method in Table 1 perform m steps to solve the linear system $Ay = 0$, with $A \succ 0$. Then, at Step $k < m$, the CG method generates a hyperplane in Cartesian coordinates equivalent to the polar hyperplane of the point y_k, with respect to the quadratic hypersurface $\mathcal{F}_{\gamma_k}$ in (7), with $\gamma_k = 1/2 y_k^T A y_k$. This hyperplane has equation*

$$\tilde{\pi}_k \;:=\; \left\{ y \in \mathbb{R}^n \;:\; (Ay_k)^T (y - y_k) = 0 \right\} \tag{10}$$

and contains the line $y_{k-1} + \xi p_{k-1}$, $\xi \in \mathbb{R}$. □

3.1 A geometric viewpoint for CG failure

A geometric interpretation for the possible CG failure in the indefinite case requires to consider again Figure 2. The hyperplane π_k represents in $\mathbb{R}^n$ the polar hyperplane of $(p_{k-1}, 0)^T$, with equation

$$\pi_k \;:=\; \{ y \in \mathbb{R}^n \;:\; (Ap_{k-1})^T y = 0 \}. \tag{11}$$

We note that π_k contains the center of the hypersurface $g(y) = \gamma_k$ and is therefore both a diametral hyperplane in $\mathbb{R}^n$ and a subspace. Now, let us consider the linear manifold ℓ, obtained as the intersection between π_k and the tangent hyperplane $\tilde{\pi}_k$ to the quadratic hypersurface $g(y) = 1/2 y_k^T A y_k$ at $y_k = x_k / x_{k0}$. The linear manifold ℓ, by the *Section Theorem*, is equivalent in $\mathbb{R}^n$ to the polar hyperplane of the point y_k with respect to the hypersurface Γ_k, being

$$\Gamma_k \;:\; \begin{cases} g(y) = \gamma_k, & \text{with} \quad \gamma_k = \frac{1}{2} y_k^T A y_k, \\ \\ y \in \pi_k. \end{cases} \tag{12}$$

These observations show that at the k-th iteration of the CG method we also have

$$\pi_k := y_k + span\{p_1, \ldots, p_{k-2}, p_k, \ldots, p_m\},$$

where the vectors $p_1, \ldots, p_{k-2}, p_k, \ldots, p_m$ are conjugate to p_{k-1}.

On the overall, we can conclude that at the $(k+1)$-th iteration of the CG method the analysis can be limited to consider the hyperplane π_k in (11) and the hypersurface Γ_k in (12), as summarized in the next result.

Corollary 3.1 *Let the CG method in Table 1 perform m steps to solve the linear system $Ay = 0$, with $A \succ 0$. Then, at Step $k \le m$, the polar hyperplane of $(p_{k-1}, 0)^T$, with respect to $g(y) = \gamma_k$, includes the point y_k, and is conjugate to p_{k-1}.* □

The next proposition gives a geometric motivation for the CG failure, in case the matrix A is indefinite. In particular, the next result explains the failure to generate a further search direction p_{k+1}, when a direction p_k is in the asymptotic cone of $\mathcal{F}_\gamma$.

Proposition 3.3 [CG - Failure] *Let the CG method in Table 1 perform m steps to solve the linear system $Ay = 0$, where A is indefinite nonsingular. Suppose the CG computes the vector p_m satisfying $p_m^T A p_m = 0$ (i.e. the point at infinity $(p_m, 0)^T$ belongs to the asymptotic cone of $\mathcal{F}_\gamma$ in (7), for some $\gamma > 0$). Then, the CG fails to generate the direction p_{m+1} and*

(1) the point $(p_m, 0)^T$ belongs to its own polar hyperplane with respect to $\mathcal{F}_\gamma$, and is self-conjugate with respect to $\mathcal{F}_\gamma$;

(2) p_m belongs to the span of all the directions which are conjugate to p_m. □

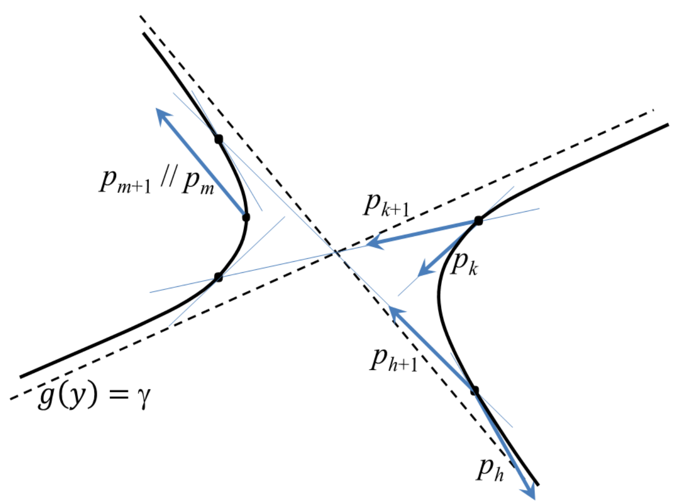

Figure 3: Possible failure of the CG method when the matrix A is indefinite, in Cartesian coordinates. The dashed lines represent the asymptotic cone. At step m the CG gets stuck as by Proposition 3.3.

Figure 3 sketches the results in Proposition 3.3. In case the direction p_m is in the asymptotic cone of $\mathcal{F}_\gamma$ (i.e., $p_m^T A p_m = 0$), then the CG method fails to generate the direction p_{m+1}, since it would be parallel to p_m. On the contrary, in case p_h or p_k is generated, with $h, k < m$, then the CG method can easily compute p_{h+1} or p_{k+1}, respectively. If A represents the indefinite Hessian matrix in Newton's equation $\nabla^2 g(z_p) d = -\nabla f(z_p)$ and, at Step m of the CG method, we have $p_m^T A p_m \approx 0$, then the CG method stops prematurely, which is a serious drawback in the light of preserving global convergence properties for the sequence $\{z_p\}$ (see also [16]). The last considerations may have a dramatic impact (for instance) on truncated–Newton methods for nonconvex programming, where the solution of a sequence of indefinite Newton's equations is sought. Here, an issue about the generation of gradient–related search directions arises. Indeed, a premature stop of the CG method may represent a frequent unexpected event to cope with (see, e.g., [17] and [18]).

We complete this section highlighting that a further viewpoint to study CG failure in the indefinite case may be analyzed, starting from some recent advances reported in the Appendix.

4 Different quadratic hypersurfaces and CG iterations

In this and in the next section we present the main advances with respect to [1]. Similarly to Section 3, here we again consider the linear transformation (9), so that in place of the linear system $Ay = b$, we can limit our analysis to the system $Ay = 0$, i.e. to the functional $g(y)$ in (6). Hence, all the results obtained in the previous section hold. In particular, we recall that Step k of the CG method in Table 1 uses the vector p_{k-1} in $\mathbb{R}^n$, in order to generate the $(n-1)$-dimensional hyperplane π_k in (11) and Figure 2. This hyperplane, in $\mathbb{R}^{n+1}$, is the polar hyperplane of the point at infinity $(p_{k-1}, 0)^T$, with respect to $\mathcal{F}_{\gamma_k}$ in (7), being $\gamma_k = 1/2 y_k^T A y_k$, and in $\mathbb{R}^n$ has the equation in (11).

At the end of the k-th iteration the directions $\{p_1, \ldots, p_k\}$ correspond to the points at infinity in homogeneous coordinates $\{(p_1, 0)^T, \ldots, (p_k, 0)^T\}$, associated with lines $\ell_1, \ldots, \ell_k$. Since $p_1, \ldots, p_k$ are mutually conjugate directions, by Propositions 4.1 and 5.1 of [1] the lines $\ell_1, \ldots, \ell_k$ are also conjugate and linearly independent. Equivalently, at Step k the direction p_k must satisfy the $k-1$ orthogonality conditions $p_k \perp Ap_1, \ldots, p_k \perp Ap_{k-1}$, i.e. the direction p_k belongs to the $[n-(k-1)]$-dimensional hyperplane $\pi(k)$ defined as

$$\pi(k) \ := \ \left\{ p \in \mathbb{R}^n \ : \ (Ap_1)^T p = 0 \ ; \ \ldots \ ; \ (Ap_{k-1})^T p = 0 \right\},$$

and not simply to the $(n-1)$-dimensional hyperplane $\{p \in \mathbb{R}^n : (Ap_{k-1})^T p = 0\}$ in (11) (i.e. the hypersurface π_k in Figure 2). Let us here stress that, similarly to the previous section, the last result is accomplished by iteratively invoking $(k-1)$ times the *Section Theorem* in homogeneous coordinates, then resorting to Cartesian coordinates. Thus, the standard algebraic arguments used by the CG are inessential here, being the Polarity theory sufficient to yield the same results.

In this section we show that results from the Polarity theory can be used to interpret, in a broader sense, the CG method in Table 1. Specifically, we focus on how this method moves from the hypersurface $g(y) = \gamma_k$ to the hypersurface $g(y) = \gamma_{k+1}$, $\gamma_k \neq \gamma_{k+1}$, in consecutive steps.

In the following, we show specific geometric properties of directions and hyperplanes, generated at three consecutive steps of the CG, in the light of the Polarity theory. In this context, the reader may use Figure 4 (case $A \succ 0$) and Figure 5 (case A indefinite) as possible reference examples, where Figure 5 sketches the k-th iteration of the CG, with specific reference to the hyperplane π_k in (11). Let us consider the points y_{r+1} and the directions p_{r+1} computed at the Steps $r = k-2$, $r = k-1$ and $r = k$ of the standard CG method. We use the following notation in Cartesian coordinates, for $r \in \{k-2, k-1, k\}$:

- γ_r is the scalar value $1/2 y_r^T A y_r$;
- π_{r+1} is the polar hyperplane of $(p_r, 0)^T$, with respect to $g(y) = \gamma_{r+1}$,
- $\tilde{\pi}_r$ is the polar hyperplane of y_r, with respect to $g(y) = \gamma_r$;
- $\Sigma_k = \pi_k \cap \tilde{\pi}_k$ is an $(n-2)$-dimensional hyperplane of $\mathbb{R}^n$;
- $\Sigma_{k+1} = \pi_k \cap \tilde{\pi}_{k+1}$ is an $(n-2)$-dimensional hyperplane of $\mathbb{R}^n$;

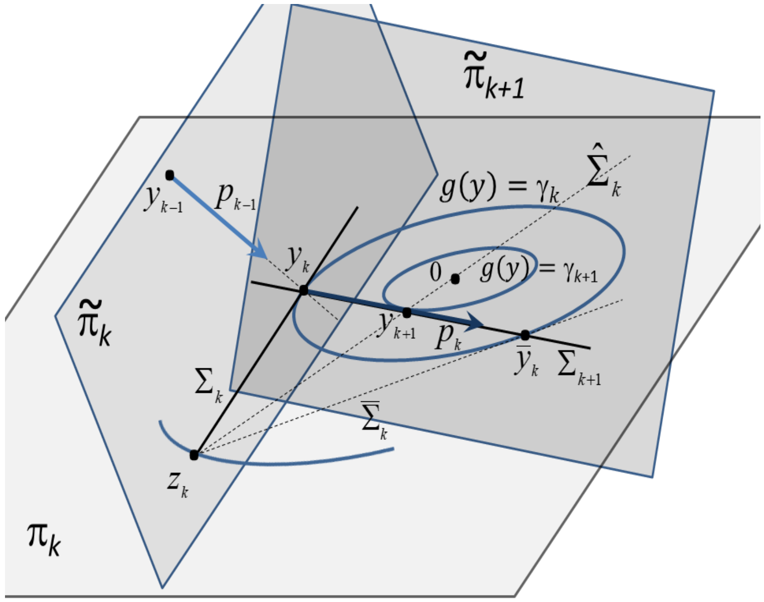

Figure 4: How the CG iteratively generates conjugate directions, exploiting the hypersurfaces $g(y) = \gamma_k$ and $g(y) = \gamma_{k+1}$. The point y_{k-1} does not belong to the hyperplane π_k (see also Figure 2), the vectors $(y_k - y_{k-1})$ and $(y_{k+1} - y_k)$ are conjugate with respect to $g(y) = \gamma_k$. Both y_{k-1} and z_k belong to the polar hyperplane $\tilde{\pi}_k$ of y_k with respect to the hypersurface $g(y) = \gamma_k$, and Σ_k is the intersection between π_k and the latter polar hyperplane of y_k.

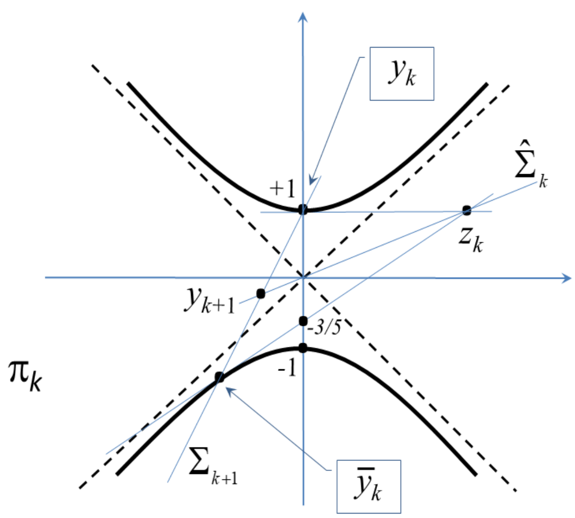

Figure 5: Case in which the quadratic hypersurface in π_k (i.e. $(y_2)^2 - (y_1)^2 = 1$) has an indefinite Hessian matrix, and we have $z_k = (2, 1)$. The segments joining the point y_{k+1} with y_k and $\bar{y}_k$, respectively, have equal length.

- $\hat{\Sigma}_r = \pi_r \cap \pi_{r+1}$ is an $(n-2)$-dimensional hyperplane of $\mathbb{R}^n$;
- $\bar{y}_r \in \mathbb{R}^n$ is the point such that

$$y_{r+1} = \frac{y_r + \bar{y}_r}{2};$$

- $\bar{\Sigma}_k$ is the polar hyperplane of $\bar{y}_r$, with respect to the hypersurface

$$\begin{cases} g(y) = \gamma_r \\ y \in \pi_r, \end{cases} \tag{13}$$

  i.e. $\bar{\Sigma}_k$ is the intersection between the polar hyperplane of $\bar{y}_k$, with respect to $g(y) = \gamma_k$, and the hyperplane π_k;
- z_k is the pole of Σ_{k+1}, with respect to (13).

We observe that $\gamma_{k+1} < \gamma_k$ as long as $A \succ 0$ and y_k is not the solution of the linear system, since it can be easily proved that the direction p_k used to define y_{k+1} is a *descent direction* for the function $g(y)$ in (6). We also recall that by Proposition 5.1 of [1] and Corollary 3.1 the hyperplane π_k contains the points y_j, $j = k, k+1, \ldots$. On the other hand, Proposition 5.2 of [1] implies that both $\tilde{\pi}_k$ contains the line $y_{k-1} + \alpha p_{k-1}$, $\alpha \in \mathbb{R}$, and $\tilde{\pi}_{k+1}$ contains the line $y_k + \alpha p_k$, $\alpha \in \mathbb{R}$. We finally remark that Σ_{k+1} is both the first polar of y_{k+1} with respect to the hypersurface

$$\begin{cases} g(y) = \gamma_{k+1} \\ y \in \pi_k, \end{cases} \tag{14}$$

and the first polar of z_k with respect to the hypersurface Γ_k in (12).

In Figure 4 the point z_k satisfies relation $g(z_k) \neq \gamma_k$. Indeed, $\bar{y}_k$ is on the line joining y_k and y_{k+1}, then the hyperplane Σ_{k+1} passes through both the points y_k and $\bar{y}_k$. Hence, the *Reciprocity Theorem* guarantees that both Σ_k and $\bar{\Sigma}_k$ pass through the pole of Σ_{k+1} with respect to (12). This latter fact and the *Section Theorem* imply that z_k is the pole of Σ_{k+1} with respect to the hypersurface Γ_k in (12). Figure 5 portraits an analogous situation when A is indefinite.

In the light of the previous considerations, the next proposition summarizes the main geometric properties of the CG method in view of Polarity, including a relevant extension to the indefinite case. In particular, the next result shows how the CG method exploits different quadratic hypersurfaces at current Step k, including the case when A is indefinite.

Proposition 4.1 [Conjugate Midpoints] *Let us consider the quadratic hypersurface $g(y) = \gamma$ in (8), with $g(y) = 1/2y^TAy$ and A possibly indefinite. Assume the CG has performed $m \leq n$ iterations*

(i) $g(\bar{y}_k) = \gamma_k$, *i.e.* $\bar{y}_k$ *belongs to the hypersurface* $g(y) = \gamma_k$;

(ii) the vectors z_k *and* $(y_k - \bar{y}_k)$ *are conjugate with respect to* $g(y) = \gamma$, *for any* $\gamma > 0$;

(iii) y_{k+1} *is the tangent point of* Σ_{k+1} *to the hypersurface* $g(y) = \gamma_{k+1}$;

(iv) $\hat{\Sigma}_k$ *contains the point*

$$y_{k+1} = \frac{y_k + \bar{y}_k}{2}; \tag{15}$$

(v) $\tilde{\pi}_r$ *and* π_r *are conjugate hyperplanes with respect to* $g(y) = \gamma_r$, *for* $r = k, k+1$.

Proof

(i) Recalling that $y_{k+1} = y_k + \alpha_k p_k$ then we have $\bar{y}_k = y_k + 2\alpha_k p_k$, which trivially implies $g(\bar{y}_k) = g(y_k + 2\alpha_k p_k)$, for any $k = 0, \dots, m-1$. Hence from Table 1

$$\begin{aligned} g(\bar{y}_k) &= g(y_k + 2\alpha_k p_k) = g(y_k) + \nabla g(y_k)^T (2\alpha_k p_k) + \frac{1}{2}(2\alpha_k p_k)^T \nabla^2 g(y_k)(2\alpha_k p_k) \\ &= g(y_k) - 2\alpha_k^2 (p_k^T A p_k) + 2\alpha_k^2 (p_k^T A p_k) = g(y_k). \end{aligned}$$

(ii) We first observe that z_k is the pole of Σ_{k+1} with respect to the hypersurface

$$\begin{cases} g(y) = \gamma_k \\ y \in \pi_k. \end{cases}$$

Analogously, y_k is the pole of Σ_k and $\bar{y}_k$ is the pole of $\bar{\Sigma}_k$ with respect to the same hypersurface. Thus, since Σ_{k+1} is the intersection between π_k and the polar hyperplane of z_k with respect to $g(y) = \gamma_k$, then Σ_{k+1} has the following equation in $\mathbb{R}^n$ (see Lemma 7.1 of [1])

$$\Sigma_{k+1} := \pi_k \cap \left\{ y \in \mathbb{R}^n \ : \ (z_k^T A z_k - 2\gamma_k) + (Az_k)^T (y - z_k) = 0 \right\}. \tag{16}$$

Moreover, since $y_k, \bar{y}_k \in \Sigma_{k+1}$ we have by (16)

$$\begin{cases} (Az_k)^T (y_k - z_k) = y_k^T A y_k - z_k^T A z_k \\ \\ (Az_k)^T (\bar{y}_k - z_k) = y_k^T A y_k - z_k^T A z_k. \end{cases} \tag{17}$$

Summing the two relations (17) we obtain

$$(Az_k)^T \left[\frac{(y_k + \bar{y}_k)}{2} - z_k \right] = y_k^T A y_k - z_k^T A z_k,$$

which states that also the midpoint in the segment joining y_k and $\bar{y}_k$ belongs to Σ_{k+1}. Finally, by subtracting relations (17) we obtain

$$(y_k - \bar{y}_k)^T A z_k = 0, \tag{18}$$

which shows that the direction $(z_k - 0)$ is conjugate to the direction $(y_k - \bar{y}_k)$.

(iii) We observe that the polar hyperplane of the point y_{k+1}, with respect to the hypersurface $g(y) = \gamma_{k+1}$, has the following equation in $\mathbb{R}^n$ (see Lemma 7.1 [1])

$$2\left(g(y_{k+1}) - \gamma_{k+1}\right) + \left(Ay_{k+1}\right)^T \left(y - y_{k+1}\right) = 0. \tag{19}$$

On the other hand, the tangent hyperplane to $g(y) = \gamma_{k+1}$ at y_{k+1} is simply given by

$$(Ay_{k+1})^T (y - y_{k+1}) = 0. \tag{20}$$

Since $g(y)$ is a regular function the last hyperplane is unique then, by Corollary 7.1 in [1], relations (19) and (20) coincide if and only if $g(y_{k+1}) = \gamma_{k+1}$, i.e. if and only if y_{k+1} is the point where Σ_{k+1} is tangent to the hypersurface $g(y) = \gamma_{k+1}$.

(iv) We first note that by *(iii)* the direction $(y_{k+1} - 0)$ is conjugate to $(y_k - \bar{y}_k)$. Moreover, observe that z_k is conjugate to $(y_k - \bar{y}_k)$ by (18). Finally, since y_{k+1} and z_k must belong to the same

manifold, then the directions $(y_{k+1} - 0)$ and $(z_k - 0)$ are parallel, i.e. they belong to the same diameter containing the origin, proving that $\hat{\Sigma}_k$ contains y_{k+1}.

(v) We observe that y_k is the pole of $\tilde{\pi}_k$ with respect to $g(y) = \gamma_k$, with $y_k \in \pi_k$. Thus, the *Reciprocity Theorem* guarantees that $\tilde{\pi}_k$ includes the pole $(p_{k-1}, 0)^T$ of π_k, proving that $\tilde{\pi}_k$ and π_k are conjugate with respect to $g(y) = \gamma_k$. A similar reasoning yields that $\tilde{\pi}_{k+1}$ and π_{k+1} are conjugate with respect to $g(y) = \gamma_{k+1}$. □

The results in the above proposition are represented in Figure 4 and Figure 5, where Σ_k and Σ_{k+1} are respectively defined by the following relations (in Cartesian coordinates)

$$\Sigma_k : \begin{cases} y_k^T A y = 2\gamma_k \quad \Longleftrightarrow \quad (Ay_k)^T (y - y_k) = 0 \\ \\ y \in \pi_k \end{cases} \tag{21}$$

$$\Sigma_{k+1} : \begin{cases} (z_k^T A z_k - 2\gamma_k) + (Az_k)^T (y - z_k) = 0 \quad \Longleftrightarrow \quad (Az_k)^T y - y_k^T A y_k = 0 \\ \\ y \in \pi_k \\ \\ \text{or equivalently} \\ \\ y_{k+1}^T A y = 2\gamma_{k+1} \quad \Longleftrightarrow \quad (Ay_{k+1})^T (y - y_{k+1}) = 0 \\ \\ y \in \pi_k \end{cases} \tag{22}$$

being $\gamma_k = 1/2 y_k^T A y_k$ and $\gamma_{k+1} = 1/2 y_{k+1}^T A y_{k+1}$. We remark that relations (21) highlight that Σ_{k+1} is the polar hyperplane of z_k, with respect to the hypersurface (12). On the contrary, relations (22) denote that Σ_{k+1} is the polar hyperplane of y_{k+1}, with respect to (14). Thus, since y_k is self-conjugate (in Cartesian coordinates) with respect to the hypersurface $g(y) = \gamma_k$, and since $y_k \in \Sigma_{k+1}$, we also have that $z_k \in \tilde{\pi}_k$ satisfies relation $(Ay_k)^T (z_k - y_k) = 0$, i.e. $z_k^T A y_k = 2\gamma_k$ by the *Reciprocity Theorem*. The latter relation yields therefore

$$0 = y_k^T A y_k - z_k^T A y_k = (y_k - z_k)^T A y_k = \nabla g(y_k)^T (y_k - z_k), \tag{23}$$

proving also that the directions $d_1 = (y_k - z_k)$ and $d_2 = (y_k - 0)$ are conjugate.

5 Polarity and Planar CG–based methods for indefinite linear systems

In Section 5 of [1] we have presented the relation between Polarity and the CG method, for solving the linear system $Ay = b$, with $A \succ 0$, or equivalently for minimizing $g(y)$ in (2). Here we recast a similar analysis for the case of A indefinite. In particular, we apply our results to a class of CG–based Krylov–subspace methods, the so called Planar CG–based methods.

The case of A indefinite requires a specific analysis, as the mere application of the CG method may be inadequate (see Proposition 3.3). In particular, when the linear system $Ay = b$ represents the Newton's equation $\nabla^2 g(z_p) d = -\nabla g(z_p)$, if the Hessian matrix $\nabla^2 g(z_p)$ is indefinite, the step k of a CG–based method may generate a direction p_k that is not *gradient-related*, and it may occur that $\nabla g(z_p)^T p_k \geq 0$.

5.1 Preliminaries on Planar CG–based methods

The Planar CG–based methods for indefinite linear systems [19, 6, 8, 20, 7] are substantially Krylov–subspace methods with a structure similar to the one summarized in Table 2, regardless of different taxonomies adopted to describe them.

In this section, we discuss the relation between Planar CG–based methods and Polarity. To this end and making reference to Table 2, we remark that the sequence of search directions $p_1, \ldots, p_m$ that these methods generate is such that the lines $\ell_1, \ldots, \ell_m$, with points at infinity $(p_1, 0)^T, \ldots, (p_m, 0)^T$, satisfy Proposition 4.2 in [1], i.e. p_i and p_j are conjugate for any $1 \leq i \neq j \leq m$.

Consider a generic Planar CG–based method in Table 2. It checks a criterion CR_k at each Step k. We remark that each Planar CG–based method may include a different criterion CR_k, whose importance deserves a specific attention. Indeed, a Planar CG–based method may show a completely different progress depending on the values returned by CR_k. In case CR_k =*TRUE*, the Planar CG–based method performs a 1×1 pivot, which is justified by the fact that the pivot element $p_{k-1}^T A p_{k-1}$ is large enough (i.e. equivalently a standard CG iteration is performed at Step k_A, inasmuch as $p_{k-1}^T A p_{k-1}$ is sufficiently bounded away from zero). On the other hand, when CR_k =*FALSE*, then the quantity $p_{k-1}^T A p_{k-1}$ is *relatively small*, so that a 2×2 pivot is necessary at Step k_B, in order to overcome a degeneracy. This mechanism allows switching from Step k_A to Step k_B and yields an algorithm which is always well–posed, so that degeneracy is eliminated. A detailed choice for the criterion CR_k is reported in Section 5.3, depending on the Planar CG–based method in hand.

We also note that, if CR_k is satisfied, the direction p_{k-1} is not auto-conjugate (see Definition 4.2 of [1]) and does not belong to the asymptotic cone of $\mathcal{F}$ in (3), i.e. $p_{k-1}^T A p_{k-1} \neq 0$. On the other hand, if CR_k is not satisfied, one of the following two situations occurs depending on the Planar CG–based method currently implemented: p_{k-1} is auto-conjugate and $q_{k-1}^T A p_{k-1} \neq 0$ (as in [6, 20]), i.e.

$$p_{k-1}^T A p_{k-1} = 0, \qquad q_{k-1}^T A p_{k-1} \neq 0, \tag{24}$$

or (see [19, 7])

$$\det \left[(p_{k-1} \quad q_{k-1})^T A (p_{k-1} \quad q_{k-1}) \right] \neq 0. \tag{25}$$

We will show, from the geometric perspective of the Polarity theory, that the conditions (24) - (25) allow the construction of the direction p_{k+1} at Step k_B of Table 2. To this purpose, we observe that (24) uses the fact that $Ap_{k-1} \perp p_{k-1}$, while by the boundedness of k relation (25) exploits the fact that p_{k-1} and q_{k-1} have an angle which is sufficiently bounded away from zero (see Proposition 2.3 in [7] for a proof).

The generic Planar CG–based method in Table 2, at step k, also indirectly performs first an iterative reduction of matrix A to a tridiagonal factorization [12]. Then, it decomposes the resulting tridiagonal matrix in order to provide search directions and obtain a new current approximate solution y_k.

A similar approach is followed also by other Krylov–subspace methods for the solution of symmetric indefinite linear systems. However, methods like SYMMLQ [21] and SYMMBK [22, 23] rely on the generation of a set of orthogonal vectors, namely the Lanczos vectors. Thus, these methods basically do not rely on conjugacy, and may be hardly analyzed in the light of the Polarity theory.

5.2 Polarity and Planar CG–based methods: basics

Detailing the exact algorithmic differences among the Planar CG–based methods in the literature goes beyond the interests of the current paper. Nevertheless, we prove that these methods match

General scheme of Planar CG–based methods

Description: Iterative method for solving equation $Ay = 0$
Input: Set an initial solution $y_0 \in \mathbb{R}^n$

Step 0: Initialization as for the CG method in Table 1

Step k**:** Using the direction p_{k-1} compute $p_{k-1}^T A p_{k-1}$ and check for the criterion CR_k: if CR_k is satisfied goto Step k_A, else goto Step k_B

Step k_A**:**

- Compute the residual $r_k = -Ay_k$, at the point $y_k = y_{k-1} + \alpha_{k-1} p_{k-1}$, with $\alpha_{k-1} \in \mathbb{R}$
- Check for a stopping rule
- Compute the novel direction p_k using r_k
- (Possibly) compute the vector $q_k \in span\{Ap_k, p_{k-1}\}$, such that $p_k^T A q_k \neq 0$
- Set $k = k + 1$. Goto **Step** k

Step k_B**:**

- Compute the residual $r_{k+1} = -Ay_{k+1}$, at the point $y_{k+1} = y_{k-1} + \alpha_{k-1} p_{k-1} + \alpha_k q_{k-1}$, with $\alpha_{k-1}, \alpha_k \in \mathbb{R}$
- Check for a stopping rule
- Compute the novel direction p_{k+1} (indirectly) using r_{k+1}
- (Possibly) compute the vector $q_{k+1} \in span\{Ap_{k+1}, p_{k-1}, q_{k-1}\}$, such that $p_{k+1}^T A q_{k+1} \neq 0$
- Set $k = k + 2$. Goto **Step** k

Table 2: A general scheme for the Planar CG–based methods, when the solution of the indefinite linear system $Ay = 0$ is sought. If Step k_B is never performed, then Planar CG–based methods essentially reduce to the CG. Similarly to the CG, the computation of p_k and possibly q_k, at Step k_A / Step k_B of Planar CG–based methods, complies with Lemma 3.2 in [1].

the geometry presented in [1] and in the previous sections.

The results in Section 2 (see also Section 2 of [1]) perfectly apply, since they are actually independent of the inertia of matrix A, and rely on non-singularity of A.

Propositions 4.1 and 4.2 of [1], regarding the mutual conjugacy of the search directions generated by Planar CG–based methods, apply too. In particular, Proposition 4.1 is at the basis of both Step k_A and Step k_B in Table 2, since it guarantees that at least $n-1$ mutually conjugate directions $\{p_i\}$ can be determined in the indefinite case (being $p_i^T A p_i \neq 0$, for at least $n-1$ values of the index i). Equivalently, in exact arithmetics, when A is indefinite and nonsingular, if p_{k-1} is auto-conjugate then $p_{k-1}^T A p_{k-1} = 0$ and the Step k_B of a Planar CG–based method can be performed at most once. To better grasp the relevance of this latter result, first observe that the criterion CR_k at Step k of Table 2 essentially checks whether the quantity $p_{k-1}^T A p_{k-1}$ is sufficiently bounded away from zero. Then, suppose by contradiction that we have both $p_{k-1}^T A p_{k-1} = p_{\hat{k}-1}^T A p_{\hat{k}-1} = 0$ at Steps $k, \hat{k}$, with $k \neq \hat{k}$, and p_{k-1}, $p_{\hat{k}-1}$ are conjugate. In this case, we would have that both the conjugate directions p_{k-1} and $p_{\hat{k}-1}$ are auto-conjugate with respect to $\mathcal{F}$ in (3), which contradicts Proposition 4.1 of [1]. The previous considerations justify the next conclusion.

Remark 5.1 *Let the matrix A be indefinite and nonsingular, and consider the Planar–GC method in Table 2. Let us generate the conjugate directions $\{p_k\}$: the situation $p_k^T A p_k = 0$ at Step k may occur only for one value of the index k.*

Now observe that the residual r_k at Step k_A of Table 2 is exactly $-\nabla g(y_k)$, being g as in (2), and the point y_k is computed in order to satisfy the relation $\nabla g(y_k)^T (y_{k-1} - y_k) = 0$. The latter equality identifies exactly the $(n-1)$-dimensional hyperplane $\tilde{\pi}_k$ in (10), in accordance with Proposition 3.2, and recalling that Step k_A in Table 2 is essentially coincident with Step k in Table 1.

In order to better justify the Step k_B in view of Polarity, we preliminarily need the next propositions, which extend the results in Propositions 3.1 and 3.2. The next results refer to a set of directions $\{t_i\}$, defined in the following way. Consider the Step k of a Planar CG–based method, with $k < m \leq n$; then we define

(i) $t_{k-1} = p_{k-1}$, if CR_k is satisfied, being $t_{k-1}^T A t_j = 0$, for any $j < k-1$;

(ii) $t_{k-1} = p_{k-1}$ and $t_k = q_{k-1}$, if CR_k is *not* satisfied, being $t_{k-1}^T A t_j = t_k^T A t_j = 0$, for any $j < k-1$.

Proposition 5.1 [Planar CG Polar Hyperplane 1] *Consider any Planar CG–based method as in Table 2. Suppose it performs m steps to solve the linear system $Ay = 0$, with A indefinite and nonsingular. Then, at Step k, $k < m$:*

(a) if CR_k is satisfied, the $(n-1)$-dimensional manifold

$$\pi_k : y_k + span\{t_1, \ldots, t_{k-2}, t_k, \ldots, t_m\}$$

represents (in Cartesian coordinates) a diametral hyperplane of the homogeneous hypersurface $\mathcal{F}_\gamma$ in (7), for any $\gamma > 0$. This diametral hyperplane π_k is the polar hyperplane of the pole $(t_{k-1}, 0)^T$, and has the expression

$$\pi_k \ := \ \{y \in \mathbb{R}^n : (A t_{k-1})^T y = 0\};$$

(b) if CR_k is not satisfied, the $(n-2)$-dimensional manifold

$$\pi_{k+1} : y_{k+1} + span\{t_1, \ldots, t_{k-2}, t_{k+1}, \ldots, t_m\}$$

represents (in Cartesian coordinates) a diametral hyperplane of the homogeneous hypersurface $\mathcal{F}_\gamma$ *in (7), for any* $\gamma > 0$. *This diametral hyperplane* π_{k+1} *is the intersection between two polar hyperplanes, whose respective poles are* $(t_{k-1}, 0)^T$ *and* $(t_k, 0)^T$, *and has the expression*

$$\pi_{k+1} \ := \ \{y \in \mathbb{R}^n : (At_{k-1})^T y = 0, \ (At_k)^T y = 0\}.$$

Proof

The proof of (a) coincides with that of Proposition 3.1. On the other hand, the proof of (b) follows by first observing that, regardless of the current Planar CG–based method, α_{k-1} and α_k at Step k_B are computed so that

$$\nabla g(y_{k+1})^T(\alpha p_{k-1} + \beta q_{k-1}) = 0, \qquad \forall \alpha, \beta \in \mathbb{R}.$$

In other words, we have the relations $\nabla g(y_{k+1})^T p_{k-1} = (Ay_{k+1})^T p_{k-1} = (Ay_{k+1})^T t_{k-1} = 0$, and similarly $\nabla g(y_{k+1})^T q_{k-1} = (Ay_{k+1})^T q_{k-1} = (Ay_{k+1})^T t_k = 0$. Hence, by (i)-(ii)

$$(At_{k-1})^T y = (At_{k-1})^T [y_{k+1} + span\{t_1, \ldots, t_{k-2}, t_{k+1}, \ldots, t_m\}] = 0,$$

$$(At_k)^T y = (At_k)^T [y_{k+1} + span\{t_1, \ldots, t_{k-2}, t_{k+1}, \ldots, t_m\}] = 0,$$

which proves the result. □

Proposition 5.2 [Planar CG Polar Hyperplane 2] *Consider any Planar CG–based method as in Table 2. Suppose it performs* m *steps to solve the linear system* $Ay = 0$, *with* A *indefinite and nonsingular. Then, at Step* k, $k < m$:

- *(a) if* CR_k *is satisfied, the Planar CG–based method equivalently generates (in Cartesian coordinates) the polar hyperplane of the point* y_k, *with respect to the quadratic hypersurface* $\mathcal{F}_{\gamma_k}$ *in (7), being* $\gamma_k = 1/2 y_k^T A y_k$. *This hyperplane has the equation* $(Ay_k)^T(y - y_k) = 0$ *and contains the line* $y_{k-1} + \alpha p_{k-1}$, $\alpha \in \mathbb{R}$;

- *(b) if* CR_k *is not satisfied, the Planar CG–based method equivalently generates (in Cartesian coordinates) the polar hyperplane of the point* y_{k+1}, *with respect to the quadratic hypersurface* $\mathcal{F}_{\gamma_{k+1}}$ *in (7), being* $\gamma_{k+1} = 1/2 y_{k+1}^T A y_{k+1}$. *This hyperplane has the equation* $(Ay_{k+1})^T(y - y_{k+1}) = 0$ *and includes the linear manifold* $y_{k-1} + span\{p_{k-1}, q_{k-1}\}$.

Proof (sketch)

In case CR_k is satisfied, then the proof of (a) essentially coincides with the one of Proposition 3.2, recalling that at Step k_A, for any Planar CG–based method $r_k = -\nabla g(y_k) = r_{k-1} - \alpha_{k-1} A p_{k-1}$, and α_{k-1} is such that $r_k^T p_{k-1} = 0$, so that $(Ay_k)^T(y - y_k) = 0$ is satisfied with $y = y_{k-1} + \alpha p_{k-1}$, $\alpha \in \mathbb{R}$.

On the other hand, using a similar reasoning, when CR_k is not satisfied, at Step k_B the vector r_{k+1} is computed, being $r_{k+1} = -\nabla g(y_{k+1})$. Moreover, since we are solving $Ay = 0$, by definition we have

$$r_{k+1} = -\nabla g(y_{k+1}) = r_0 - \sum_{i=1}^{k+1} \alpha_{i-1} A t_{i-1} = r_{k-1} - \alpha_{k-1} A p_{k-1} - \alpha_k A q_{k-1}, \tag{26}$$

and the parameters α_{k-1}, α_k are computed imposing the conditions $r_{k+1}^T p_{k-1} = 0$ and $r_{k+1}^T q_{k-1} = 0$, or equivalently $r_{k+1}^T(y - y_{k-1}) = 0$, with $y \in y_{k-1} + span\{p_{k-1}, q_{k-1}\}$. By $r_{k+1} = -Ay_{k+1}$ the latter relations coincide with $(Ay_{k+1})^T(y-y_{k+1})=0$, as long as $y \in y_{k-1}+span\{p_{k-1},q_{k-1}\}$. Since y_{k+1} satisfies the equation $g(y)=\gamma_{k+1}$, then the relation $(Ay_{k+1})^T(y-y_{k+1})=0$ represents in Cartesian coordinates the polar hyperplane of y_{k+1} with respect to $g(y)=\gamma_{k+1}$.

Now, it remains to show that when CR_k is not satisfied then the computation of α_{k-1} and α_k is well-posed. On this guideline observe that by (26), imposing the conditions $r_{k+1}^T p_{k-1} = r_{k+1}^T q_{k-1} = 0$ is equivalent to solve the 2×2 linear system

$$\begin{pmatrix} p_{k-1}^T A p_{k-1} & p_{k-1}^T A q_{k-1} \\ \\ q_{k-1}^T A p_{k-1} & q_{k-1}^T A q_{k-1} \end{pmatrix} \begin{pmatrix} \alpha_{k-1} \\ \\ \alpha_k \end{pmatrix} = \begin{pmatrix} r_{k-1}^T p_{k-1} \\ \\ r_{k-1}^T q_{k-1} \end{pmatrix}.$$

Recalling that at Step k_B either condition (24) or condition (25) holds, depending on the Planar CG–based method adopted, then the matrix in the last linear system is always nonsingular. □

Moreover, also for Planar CG–based methods results similar to Corollary 3.1 and Proposition 4.1 hold, with simple modifications.

5.3 Planar CG–based methods: further results

This section proposes some advances on Step k_B in Table 2, so that we can better motivate the role of the theory of Polarity when Planar CG–methods are considered, in view of Proposition 4.2 of [1]. We first note that, roughly speaking, the Step k_B in Table 2 is equivalent to a double CG step. Then we observe that the different Planar CG–based methods proposed in the literature substantially differ at Step k_B. In particular, they differ in the way they generate the vector q_{k-1} (see also the vector d_{i+1} considered in the proof of Proposition 4.2 of [1]).

The first method that we consider is the one proposed by Luenberger in 1969 [20]: to the best of our knowledge, it is the first method for solving indefinite linear systems, entirely based on conjugate directions. This method performs Step k_B in Table 2 when an auto-conjugate direction with respect to $\mathcal{F}$ in (3) is detected. In other words, as criterion CR_k this method tests whether the current direction p_{k-1} is auto-conjugate. In case CR_k is fulfilled then Step k_B is executed, so that this method generates a new direction p_{k+1} using directions p_{k-1} and q_{k-1}, being q_{k-1} a direction in the asymptotic cone of $\mathcal{F}$ (i.e. $q_{k-1}^T A q_{k-1} = 0$). Thus, by Proposition 4.1 of [1] the vectors p_{k-1} and q_{k-1} cannot be conjugate (as also indicated in Table 2). Then, the directions $p_1,\dots,p_{k-1},q_{k-1}$, computed up to Step k of Luenberger's method, enjoy the following properties (as they satisfy Proposition 4.2 of [1])

1. $p_1,\dots,p_{k-1},q_{k-1}$ are linearly independent;
2. q_{k-1} is conjugate to p_i, with $i \in \{1,\dots,k-2\}$.

The choice of q_{k-1} in [20] perfectly matches both the conclusions in Proposition 4.2 and Corollary 4.1 of [1] (setting respectively $d_1 = p_{k-1}$ and $d_2 = q_{k-1}$). The fact that q_{k-1} is self-conjugate simplifies the computation of some coefficients in Step k_B [20]. However, testing within CR_k the analytical condition $p_{k-1}^T A p_{k-1} = 0$ might yield a numerically unstable procedure, which explains why in the literature Luenberger's method is rarely applied (an analogous issue holds also for the method proposed in [6]).

A different criterion CR_k is adopted in the Planar CG–based methods proposed in [19] and in [8, 7]. In [19] the quantity $\Delta_{k-1} = (p_{k-1}^T A p_{k-1})(q_{k-1}^T A q_{k-1}) - (p_{k-1}^T A q_{k-1})^2$ is tested at Step k, where p_{k-1} and q_{k-1} are directions computed either at Step $(k-1)_A$ or Step $(k-2)_B$. As proved in [8], when Δ_{k-1} is sufficiently large the angle between p_{k-1} and q_{k-1} is sufficiently bounded away from zero, so that p_{k-1} and q_{k-1} are linearly independent. Thus, when Δ_{k-1} is large enough, then Step k_B is performed, otherwise the Step k_B is not well-posed and Step k_A is applied.

Note that in [7] the criterion CR_k simply tests the quantity $p_{k-1}^T A p_{k-1}$, in place of Δ_{k-1}. If $p_{k-1}^T A p_{k-1}$ is *relatively* large, then p_{k-1} is surely not an auto-conjugate direction, and Step k_A is performed. Otherwise, the Step k_B is preferred. We remark that both in [19] and [7], at Step k_B a direction q_{k-1} satisfying 1.–2. is used. Thus, these methods are numerically more stable and reliable (though a bit more expensive) than [6] and [20], as also showed in [8].

We conclude this section by recalling that Step k of the standard CG method determines the steplength α_{k-1} so that the functional $g(y)$ in (6) has a stationary point along the direction p_{k-1}. Similarly, the four Planar CG–based methods reported above, at Step k_B (see Table 2) compute the steplengths α_{k-1} (along p_{k-1}) and α_k (along q_{k-1}) so that the Ritz-Galerkin condition

$$\nabla g(y_{k+1})^T (y - y_{k+1}) = 0, \qquad \forall y \in y_{k-1} + span\{p_{k-1}, q_{k-1}\} \tag{27}$$

holds. This guarantees that $g(y)$ has the stationary point y_{k+1} on the 2-dimensional manifold

$$y_{k-1} + span\{p_{k-1}, q_{k-1}\},$$

as item (b) of Proposition 5.2 suggests. The latter computation of α_{k-1}, α_k is well-posed thanks to Proposition 5.2-(b) and Corollary 4.1 of [1], which ensure that p_{k-1} and q_{k-1} are linearly independent (though not conjugate).

To sum up, the general theoretical results in Section 4 of [1] have an algorithmic counterpart in the Planar CG–based methods. These methods, as well as the CG method, exploit the Polarity theory but they do not need to resort explicitly to homogeneous coordinates. In the literature their recurrences only refer to vectors in Cartesian coordinates. However, as we proved in Proposition 5.2, Planar CG–based methods strongly rely on Polarity for quadratic hypersurfaces. This latter fact proves that the Polarity theory might have potentialities to suggest novel methods even when indefinite linear systems are considered. These new methods could be based on a different criterion CR_k at Step k_B in Table 2. Equivalently, by the taxonomy in the proof of Proposition 4.2 of [1], they could choose a novel vector d_{i+1} (or similarly a novel vector q_{k-1} at Step k_B), satisfying possible additional properties. In particular, the perspective from projective geometry, as well as both the novel approach suggested in Section 3.1 using the grossone and the stability issues exploited in SYMMBK (see [23]), when performing a 2×2 pivot step, might all be useful ingredients to design an effective and stable novel Planar CG–based method.

6 Conclusions

In this paper, we have investigated in depth the strong relationship between CG–based methods and the Polarity theory in homogeneous coordinates. In particular, we focused on quadratic hypersurfaces whose Hessian matrix is indefinite, recalling the geometric motivation behind a possible CG failure.

Then, following the guidelines in [1], we have also showed how the so called Planar CG–based methods can be recast skipping homogeneous coordinates, and simply providing their recurrences in Cartesian coordinates. The Polarity theory has been used to give evidence on how, from a geometric standpoint, the Planar CG–based methods are provably well-posed. The last result was naturally accomplished in the indefinite case, resorting to the asymptotic cone.

We think that the Polarity theory may be useful to suggest further generalizations, considering possible extensions to general algebraic hypersurfaces, not necessarily quadratic, of the Planar

CG–based methods. In particular, possible issues of interest for further research are presented as follows.

- New, possibly inexact, linesearch procedures along p_k could be conceived, in case the matrix A is indefinite and p_k is in the asymptotic cone of $\mathcal{F}$ in (3). These procedures should not be simply based on Ritz-Galerkin condition (27), and should take a specific care about preserving convergence properties, in the light of Proposition 3.3 and its proof.

- By the proof of Proposition 3.3, observe that if in (3) the matrix A is indefinite, then the polar hyperplane of a point at infinity P in the asymptotic cone is well-defined. On the contrary, the latter hyperplane has not at P formally a counterpart in Cartesian coordinates, being P a point at infinity. This suggests that a failure of the CG method, in case A is indefinite, might be possibly recovered by suitably alternating homogeneous coordinates and Cartesian coordinates in CG iterations. Indeed, by Proposition 3.3 a CG failure occurs in case at Step k we have $p_k^T A p_k = 0$, i.e. p_k is auto-conjugate with respect to the quadratic hypersurface. Thus, from the definition of asymptotic cone, p_k is in principle still a possible search direction, in order to detect the center of the hypersurface, by computing a suitable steplength. Of course, since possibly y_k in Table 2 is not in the asymptotic cone of $\mathcal{F}$, in general in Cartesian coordinates the line $y_k + \lambda p_k$, $\lambda \in \mathbb{R}$, does not include the center y^* of $\mathcal{F}$. Unfortunately, the CG method is unable to compute a finite steplength along p_k (see also Figure 5), so that it stops prematurely. Nevertheless, an *ad hoc* inexact linesearch procedure along p_k could be investigated.

- A comment similar to the previous one also applies to the Nonlinear Conjugate Gradient method, when used to detect a stationary point of a general polynomial function $\Psi(x)$. In the latter case, a point at infinity $(p_k, 0)^T$ in the asymptotic cone of the hypersurface associated to $\Psi(x)$ can possibly play a significant role.

- The case when the CG method detects a *nearly* auto-conjugate direction p_k (i.e., such that $p_k^T A p_k \approx 0$), represents another intriguing scenario to theoretically investigate, from a geometric standpoint. In fact, the latter case makes the CG method well-posed but possibly numerically unstable (see e.g. [24]).

We conclude this section by observing that, as suggested in Section 3.1 and in the Appendix, we may give an alternative interpretation for a CG failure, by adopting a recent novel paradigm for infinite and infinitesimal computing (see also [13, 25, 26]) that gives a completely different perspective.

Appendix

In this appendix, we argue that the failure of the CG method may be studied exploiting a recent computational methodology that allows to easily manipulate infinities and infinitesimals (see [27, 28]) using a standard algebra.

In particular, by Proposition 3.1 in [13] we know that in case the matrix A is indefinite, a failure to perform the Step k for the CG method in Table 1 implies

$$\lim_{\gamma \to 0} \|p_k\| = +\infty,$$

being $\gamma_k = p_{k-1}^T A p_{k-1}$; hence the CG method stops beforehand, yielding a *degeneracy.* This drawback can be skipped by introducing the numeral *grossone* as in [27, 28], which allows an elementary manipulation of the quantity $\|p_k\|$ in case of degeneracy at Step k. We strongly remark that the analysis using grossone uses a completely different standpoint with respect to the *Nonstandard Analysis* in calculus (see e.g. [29]). Nevertheless, the use of grossone seems yet inadequate to explain the geometry behind a CG degeneracy, as the Polarity theory straightforwardly yields (see also Section 5.2 of [13]).

Acknowledgments. G. Fasano thanks the National Research Council - Marine Technology Research Institute (CNR-INSEAN), Italy, along with the working group GNCS of INδAM (Istituto Nazionale di Alta Matematica), Italy, for the support received.

References

[1] G. Fasano and R. Pesenti, "Conjugate direction methods and polarity for quadratic hypersurfaces," *Journal of Optimization Theory and Applications*, vol. 175(3), pp. 764–794, 2017.

[2] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–435, 1952.

[3] M. Beltrametti, E. Carletti, D. Gallarati, and G. Monti Bragadin, *Lectures on Curves, Surfaces and Projective Varieties - A Classical View of Algebraic Geometry.* European Mathematical Society, 2009.

[4] J. Casey, *A Treatise on the Analytical Geometry of the point, line, circle and conic sections.* Dublin University Press, 1885.

[5] O. Schreirer and E. Sperner, *Projective geometry of n dimensions.* Chelsea Publishing Company, NY, 1961.

[6] G. Fasano, "Conjugate Gradient (CG)-type Method for the Solution of Newton's Equation within Optimization Frameworks," *Optimization Methods and Software*, vol. 19, pp. 267–290, 2004.

[7] G. Fasano, "Planar-Conjugate Gradient algorithm for Large Scale Unconstrained Optimization, Part 1: Theory," *Journal of Optimization Theory and Applications*, vol. 125, pp. 523–541, 2005.

[8] G. Fasano, "Planar-Conjugate Gradient algorithm for Large Scale Unconstrained Optimization, Part 2: Application," *Journal of Optimization Theory and Applications*, vol. 125, pp. 543–558, 2005.

[9] F. Curtis and D. Robinson, "Exploiting negative curvature in deterministic and stochastic optimization," *Mathematical Programming*, vol. 176, pp. 69–94, 2019.

[10] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1724–1732, JMLR.org, 2017.

[11] J. Moré and D. Sorensen, "On the use of directions of negative curvature in a modified Newton method," *Mathematical Programming*, vol. 16, pp. 1–20, 1979.

[12] G. Fasano and M. Roma, "Iterative computation of negative curvature directions in large scale optimization," *Computational Optimization and Applications*, vol. 38, pp. 81–104, 2007.

[13] R. De Leone, G. Fasano, and Y. D. Sergeyev, "Planar methods and grossone for the conjugate gradient breakdown in nonlinear programming," *Computational Optimization and Applications*, vol. 71, pp. 73–93, 2018.

[14] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control.* Hoboken, NJ: John Wiley & Sons, Inc., 2020.

[15] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer graphics. Principles and Practice.* Addison-Wesley Publishing Company, 1990.

[16] R. Dembo, S. Eisenstat, and T. Steihaug, "Inexact Newton methods," *SIAM Journal on Numerical Analysis*, vol. 19, pp. 400–408, 1982.

[17] S. Nash, "A survey of truncated-Newton methods," *Journal of Computational and Applied Mathematics*, vol. 124, pp. 45–59, 2000.

[18] A. Caliciotti, G. Fasano, S. Nash, and M. Roma, "An adaptive truncation criterion, for linesearch-based truncated Newton methods in large scale nonconvex optimization," *Operations Research Letters*, vol. 46, pp. 7–12, 2018.

[19] M. Hestenes, *Conjugate Direction Methods in Optimization.* Springer Verlag, New York, Heidelberg, Berlin, 1980.

[20] D. Luenberger, "Hyperbolic pairs in the method of conjugate gradients," *SIAM Journal on Applied Mathematics*, vol. 17, p. 1969, 1263–1267.

[21] C. Paige and M. Saunders, "Solution of sparse indefinite systems of linear equations," *SIAM Journal on Numerical Analysis*, vol. 12, pp. 617–629, 1975.

[22] R. Chandra, *Conjugate gradient methods for partial differential equations.* PhD thesis, Yale University, New Haven, 1978. Research Report 129.

[23] J. Bunch and L. Kaufman, "Some stable methods for calculating inertia and solving symmetric linear equations," *Mathematics of Computations*, vol. 31, pp. 163–179, 1977.

[24] A. Greenbaum, *Iterative methods for solving linear systems.* Philadelphia, PA: SIAM, 1997.

[25] R. De Leone, N. Egidi, and L. Fatone, "The use of grossone in elastic net regularization and sparse support vector machines," *Soft Computing, https://doi.org/10.1007/s00500-020-05185-z*, 2020.

[26] R. De Leone, G. Fasano, M. Roma, and Y. D. Sergeyev, "Iterative grossone-based computation of negative curvature directions in large-scale optimization," *Journal of Optimization Theory and Applications*, vol. 186, pp. 554–589, 2020.

[27] Y. D. Sergeyev, "Numerical infinities and infinitesimals: Methodology, applications, and repercussions on two Hilbert problems," *EMS Surveys in Mathematical Sciences*, vol. 4(2), pp. 219–320, 2017.

[28] Y. D. Sergeyev, "Higher order numerical differentiation on the infinity computer," *Optimization Letters*, vol. 5(4), pp. 575–585, 2011.

[29] Y. D. Sergeyev, "Independence of the Grossone-Based Infinity Methodology from Non-standard Analysis and Comments upon Logical Fallacies in Some Texts Asserting the Opposite," *Foundations of Science*, vol. 24, pp. 153–170, 2019.