Comput Optim Appl (2018) 71:73–93
https://doi.org/10.1007/s10589-017-9957-y

Planar methods and grossone for the Conjugate Gradient breakdown in nonlinear programming

Renato De Leone[1] · **Giovanni Fasano**[2] · **Yaroslav D. Sergeyev**[3,4]

Received: 12 March 2017 / Published online: 23 October 2017

Abstract This paper deals with an analysis of the Conjugate Gradient (CG) method (Hestenes and Stiefel in J Res Nat Bur Stand 49:409–436, 1952), in the presence of degenerates on indefinite linear systems. Several approaches have been proposed in the literature to issue the latter drawback in optimization frameworks, including reformulating the original linear system or recurring to approximately solving it. All the proposed alternatives seem to rely on algebraic considerations, and basically pursue the idea of improving numerical efficiency. In this regard, here we sketch two separate analyses for the possible CG degeneracy. First, we start detailing a more standard algebraic viewpoint of the problem, suggested by *planar methods*. Then, another algebraic perspective is detailed, relying on a novel recently proposed theory, which includes an additional number, namely *grossone*. The use of grossone allows to work numerically with infinities and infinitesimals. The results obtained using the two proposed approaches perfectly match, showing that grossone may represent a fruitful and promising tool to be exploited within Nonlinear Programming.

✉ Giovanni Fasano
fasano@unive.it

Renato De Leone
renato.deleone@unicam.it

Yaroslav D. Sergeyev
yaro@dimes.unical.it

1 Scuola di Scienze e Tecnologie, Università di Camerino, Camerino, Italy

2 Dipartimento di Management, Università Ca' Foscari Venezia, Venice, Italy

3 Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica, Università della Calabria, Rende, Italy

4 Department of Software and Supercomputing, Lobachevsky State University, Nizhniy Novgorod, Russia

Keywords Conjugate Gradient (CG) method · Planar-CG methods · Infinities and Infinitesimals · Grossone

1 Introduction

In this paper we deal with the solution of indefinite linear systems, by iterative methods uniquely based on generating conjugate directions. As a consequence, here we will not directly consider Lanczos-based iterations too, unless in their connection with the generation of conjugate vectors.

In particular, we study the specific behaviour of the CG in case of degeneracy, since it has relevant implications when solving Numerical Analysis problems and within Optimization frameworks. In this regard, the current literature of Krylov subspace methods (see e.g. [46] or [13]) provides plenty of applications where the CG is used and it can possibly fail to yield reliable solutions. In addition, both unconstrained and constrained optimization frameworks include problems where the search of stationary points of convex and nonconvex functions is sought, requiring the solution of a positive definite or indefinite symmetric system.

We recall that the CG (see the scheme in Table 1) iteratively generates a sequence of approximate solutions $\{y_k\}$ to the symmetric linear system $Ay = b$, until a stop condition based on the current residual $r_k = b - Ay_k$ is met, so that the current approximate solution y_k is used. Unfortunately, on specific indefinite linear systems, and depending on the choice of the initial iterate y_0, the CG may experience a premature undesired stop. As well known (see also [3] for the consequences in optimization frameworks), when this scenario occurs, an algebraic drawback takes place during the CG iteration: namely a division by a small amount is involved. This situation is usually addressed in the literature as a *pivot breakdown*, and corresponds to the fact that at Step k the search direction p_k yields $p_k^T Ap_k = 0$, i.e. the stepsize α_k along p_k can not be computed. As a consequence, the CG stops beforehand and the current iterate y_k may be far from being a solution of the linear system (equivalently the quantity $\|r_k\|$ might be significantly nonzero). From a different perspective, some comments on the contents of the current paragraph can be found also in [16], where illconditioning for nonlinear programming problems is partially addressed, combining ideas from quasi-Newton methods and preconditioning.

Table 1 The CG algorithm for solving the symmetric linear system $Ay = b$, $A \in \mathbb{R}^{n\times n}$

The Conjugate Gradient (CG) method	
Data:	Set $k = 0$, $y_0 = 0$, $r_0 = b - Ay_0$. If $r_0 = 0$, then STOP. Else, set $p_0 = r_0$
Step k:	Compute $\alpha_k = r_k^T p_k / p_k^T Ap_k$, $y_{k+1} = y_k + \alpha_k p_k$, $r_{k+1} = r_k - \alpha_k Ap_k$
	If $r_{k+1} = 0$, then STOP. Else, set $\beta_k = -r_{k+1}^T Ap_k / p_k^T Ap_k = \|r_{k+1}\|^2/\|r_k\|^2$, and $p_{k+1} = r_{k+1} + \beta_k p_k,\ k = k + 1$
	Go to Step k

In this work we specifically address the pivot breakdown of the CG, from a different novel perspective. Our analysis includes an algebraic approach which encompasses an extension of real numbers. In some sense our analysis can be *unusual* for the CG, since the literature of the last decades has mainly focused on its performance and on the stability of its iterative process in Table 1. Nevertheless, we are convinced that a proper investigation of the ultimate algebraic reasons of CG degeneracy should be fruitfully exploited, in order to prevent pivot breakdowns and further improve it.

The paper is organized as follows. In Sect. 2 we highlight an algebraic perspective for the CG, when applied to solve indefinite linear systems. In Sect. 3, starting from some preliminary considerations, we infer geometric results on the CG degeneracy, in connection with the so called *Planar-CG* methods from the literature. Sect. 4 introduces an extension of the Cartesian space, including advances using the recently introduced numeral *grossone*; this section also contains specific algebraic properties of the resulting extended real space. Then, Sect. 5 reports a novel algebraic perspective for the CG degeneracy, which strongly relies on the use of grossone and the results reported in Sect. 4. Finally, a section of conclusions and an Appendix complete the paper.

As regards the notation, $\|\cdot\|$ indicates the Euclidean norm. With $|\lambda_m(A)|$ and $|\lambda_M(A)|$ we respectively indicate the smallest and the largest modulus of an eigenvalue of matrix $A \in \mathbb{R}^{n\times n}$. Finally, the symbol ① indicates the numeral *grossone*, whose formal properties are better detailed in Sect. 4.

2 An algebraic approach using the CG for indefinite linear systems, in optimization frameworks

In the previous section we remarked the role played by conjugate directions within nonlinear programming frameworks. This has also motivated, in the literature of optimization, the interest for possibly rearranging indefinite linear systems, whose approximate solution by iterative methods may provide suitable gradient-related directions (see [27]), based on conjugate directions. In this regard, the proposals in [14,15] directly aim at using the CG for building a suitable search direction based on conjugacy among vectors. As by product, the proposals in [14,15] indirectly rely on a *modified* linear system, as detailed in the next proposition (which refers to the CG in Table 1).

The next novel result in the literature considers that, on indefinite linear systems, a certain number of CG iterations can be performed before halting. The conjugate directions generated in these iterations are subject to an interesting interpretation. Namely, we can show that these conjugate directions can be suitably combined to yield a solution, of both the indefinite linear system $Ax = b$ and an auxiliary positive definite linear system $\tilde{A}x = b$. The relevant implications of the latter result, in optimization frameworks, are detailed in the end of the present section.

Proposition 2.1 *Consider the nonsingular indefinite linear system $Ay = b$, with $A \in \mathbb{R}^{n\times n}$. Suppose the CG in Table 1 is applied for its solution, and assume it generates up to Step n the (nonzero) conjugate directions $p_1, \ldots, p_n$, satisfying*

$$\begin{cases} p_i^T A p_i \neq 0, & \\ p_i^T A p_j = 0, & 1 \leq i \neq j \leq n, \\ A\left(\sum_{i=1}^{n} \alpha_i p_i\right) = b. & \end{cases}$$

Let us reorder the set $\{p_1, \ldots, p_n\}$ *so that* $n = P + N$ *and* $\{p_1, \ldots, p_n\} = P1 \cup P2$*, where*

$$P1 = \{p_1, \ldots, p_P\} \quad \text{with } p_i^T A p_i > 0, \;\; i = 1, \ldots, P;$$
$$P2 = \{p_{P+1}, \ldots, p_{P+N}\} \quad \text{with } p_i^T A p_i < 0, \;\; i = P+1, \ldots, P+N.$$

Then, there exists a positive definite matrix $\tilde{A} \in \mathbb{R}^{n \times n}$ *such that:*

(i) $\tilde{A}^{-1}$ *is given by*

$$\tilde{A}^{-1} = \sum_{i=1}^{P} \frac{1}{p_i^T A p_i} p_i p_i^T - \sum_{i=P+1}^{P+N} \frac{1}{p_i^T A p_i} p_i p_i^T;$$

(ii) *if* $p_i \in P1$ *then* $\tilde{A}^{-1}(Ap_i) = p_i$*;*
(iii) *if* $p_i \in P2$ *then* $\tilde{A}^{-1}(Ap_i) = -p_i$*;*
(iv) $p_i^T \tilde{A} p_j = 0$*, for any* $1 \leq i \neq j \leq n$*;*
(v) *setting* $d_P = \sum_{i=1}^{P} \alpha_i p_i$ *and* $d_N = -\sum_{i=P+1}^{P+N} \alpha_i p_i$*, then the saddle point* d^* *of the function* $f(d) = 1/2 d^T A d - b^T d$ *is given by* $d^* = d_P - d_N$*, while the minimum point* d^{**} *of the function* $g(d) = 1/2 d^T \tilde{A} d - b^T d$ *is given by* $d^{**} = d_P + d_N$*.*

Proof The proof can be found in the Appendix. □

The latter proposition shows that in practice, the computation of d_P and d_N, i.e. separating the contribution of positive and negative curvature (conjugate) directions, allows equivalently to build Newton's direction (namely $d_P + d_N$) of the positive definite linear system $\tilde{A}d = b$. Thus, there is a correspondence between the stationary point of the quadratic functional $1/2 d^T A d - b^T d$ and the minimum point of the convex auxiliary functional $1/2 d^T \tilde{A} d - b^T d$. The latter property is much appealing in optimization frameworks, when $A = \nabla^2 f(y)$ and $b = -\nabla f(y)$, since the vector $d_P - d_N$ solves Newton's equation

$$\nabla^2 f(y) d + \nabla f(y) = 0$$

but might not be *gradient-related*. On the contrary, $d_P + d_N$ is a gradient-related direction (see also [10]), which easily allows to state global convergence properties for the overall optimization framework.

The observations in this section prove that the use of conjugate directions may be quite useful in both the positive definite and the indefinite case, since they can be suitably combined to provide search directions in Nonlinear Programming schemes. However, we also remark that in case the CG degenerates at iteration k, namely $p_k^T A p_k \approx 0$, then the above analysis fails and the CG stops prematurely, so that the contents in this section are yet unable to fully cope with the degenerate case of the CG.

3 Geometric consequences of CG degeneracy on indefinite linear systems

In this section we briefly analyze some algebraic and geometric implications of possible CG failures, when the CG is applied to an indefinite linear system $Ay = b$. The case when possibly A is positive definite follows as a consequence. In particular, we want to recall some properties satisfied by the CG when at Step k a degenerate or nearly degenerate situation occurs, namely $p_k^T A p_k \approx 0$. The couple of results we report here will be suitably reinterpreted from an alternative standpoint, using *grossone* in Sect. 5.

When the matrix A is positive definite, at any Step k of the CG we have $\lambda_m(A)\|p_k\|^2 \le p_k^T A p_k$, so that the quantity $p_k^T A p_k$ may be suitably bounded from below. Conversely, in case A is indefinite nonsingular, such a bound does not hold, being potentially $p_k^T A p_k = 0$. Nevertheless, we can say (see the analysis in [9] for details) that in case the matrix A is indefinite nonsingular and at Step k we have

$$p_k^T A p_k > \varepsilon_k \|p_k\|^2, \quad \varepsilon_k > 0,$$

with $\|p_k\| < +\infty$ and $\|p_{k+1}\| < +\infty$, then the angle $\widehat{p_k, p_{k+1}}$ between the directions p_k and p_{k+1} satisfies the relations

$$\frac{\pi}{2} - \arccos\left(\frac{\varepsilon_k}{|\lambda_M(A)|}\right) \le |\widehat{p_k, p_{k+1}}| \le \frac{\pi}{2} + \arccos\left(\frac{\varepsilon_k}{|\lambda_m(A)|}\right), \tag{3.1}$$

showing that when ε_k is sufficiently bounded away from zero, then p_k and p_{k+1} may not become parallel. On the contrary, if p_k and p_{k+1} tend to be parallel, from (3.1) we have that $\varepsilon_k \to 0$.

As a second fact, in the next proposition we specifically investigate the norm of the directions generated by the CG, in a nearly degenerate case.

Proposition 3.1 *Consider the indefinite nonsingular linear system $Ay = b$, with $A \in \mathbb{R}^{n\times n}$, suppose the CG is applied for its solution, and let at Step k be $\|r_k\| = \|b - Ay_k\| \ge \varepsilon > 0$, with $0 < \|p_k\| < +\infty$. Then, setting $\gamma = p_k^T A p_k$, we have*

$$\lim_{\gamma\to 0} \|p_{k+1}\| = +\infty.$$

Proof By the hypotheses $\|r_k\| \ge \varepsilon$. Then, using well-known properties of the CG and recalling the expression $\beta_k = -r_{k+1}^T A p_k/\gamma$, we have

$$p_{k+1}^T p_k = (r_{k+1} + \beta_k p_k)^T p_k = \beta_k \|p_k\|^2 = -\frac{(r_k - \alpha_k A p_k)^T A p_k}{\gamma} \|p_k\|^2$$
$$= -\frac{(p_k - \beta_{k-1} p_{k-1} - \alpha_k A p_k)^T A p_k}{\gamma} \|p_k\|^2 = \alpha_k \frac{\|A p_k\|^2}{\gamma} \|p_k\|^2 - \|p_k\|^2.$$

Thus, since $r_k^T p_k = \|r_k\|^2$ and by direct computation $\|p_k\|^2 = \|r_k\|^2 + \beta_{k-1}^2 \|p_{k-1}\|^2$, then $\|r_k\| \leq \|p_k\| < +\infty$ so that

$$\frac{p_{k+1}^T p_k}{\|p_k\|} = \cos(\widehat{p_k, p_{k+1}}) \|p_{k+1}\| = \frac{\|r_k\|^2 \|A p_k\|^2 \|p_k\|}{\gamma^2} - \|p_k\|$$
$$\geq \left[\frac{\varepsilon^2 |\lambda_m(A)|^2 \|p_k\|^2}{\gamma^2} - 1 \right] \|p_k\|,$$

which yields by $\|p_k\| \geq \|r_k\| \geq \varepsilon$

$$\lim_{\gamma \to 0} \frac{|p_{k+1}^T p_k|}{\|p_k\|} = +\infty.$$

Finally, the boundedness of $\cos(\widehat{p_k, p_{k+1}})$ and the latter relation yield $\lim_{\gamma \to 0} \|p_{k+1}\| = +\infty$. □

3.1 Planar methods as a remedy to CG degeneracy

This section briefly reviews some CG-based Krylov-subspace methods from the literature, which have been proposed to cope with the case in which the CG degenerates at Step k, on indefinite nonsingular linear systems. In this section we will not include also the comprehensive analysis by Oren [28], though it proposes a specific planar method for CG degeneracy. Indeed, the proposal in [28] does not purely rely on conjugate directions, but it starts from considering the family of quasi-Newton methods in [19]. Nevertheless, the paper [28] deserves much attention in our opinion, since it also provides to large extent a generalization of some planar methods we are going to analyze.

A thorough analysis of planar methods if beyond the purposes of this paper. However, as a preliminary consideration, observe that planar methods work on the basis of a common similar mechanism, which takes place when $p_k^T A p_k = 0$ or $p_k^T A p_k \approx 0$ (depending on the planar algorithm utilized). In particular, when either of the latter conditions holds at step k (*planar step*), then an additional direction $q_k \in span\{A p_k, p_k, p_{k-1}\}$ is first generated. Then, a *planar* step is performed, so that starting from the current iterate y_k the novel point

$$y_{k+2} = y_k + \alpha_k p_k + \beta_k q_k \tag{3.2}$$

is computed, in such a way that the Ritz–Galerkin conditions

$$(b - Ay_{k+2})^T q_k \; = \; (b - Ay_{k+2})^T p_k \; = \; 0 \tag{3.3}$$

are fulfilled. This indirectly guarantees that the residual $r_{k+2} = b - Ay_{k+2}$ is also orthogonal to all the search directions $p_1, \ldots, p_k, q_k$. The conditions (3.3) are used to compute the coefficients α_k and β_k in (3.2), and require the solution of a 2×2 symmetric linear system with coefficients matrix

$$\begin{pmatrix} p_k^T A p_k & p_k^T A q_k \\ \\ q_k^T A p_k & q_k^T A q_k \end{pmatrix}. \tag{3.4}$$

Depending on the planar method adopted, analytical conditions are ensured so that the above 2×2 matrix is always nonsingular (though possibly illconditioned).

We urge to recall that basically the planar methods differ with respect to a couple of choices:

1. the criterion adopted to check at step k for the condition $p_k^T A p_k \approx 0$;
2. the computation of the search direction q_k at the current k-th planar step.

As regards item 1., the planar methods in [24] and [8] check for the simpler condition $p_k^T A p_k = 0$, in order to decide whether the current step k should be a planar one. This may evidently yield inaccuracies when $p_k^T A p_k \approx 0$ but $p_k^T A p_k \neq 0$, so that in some problems numerical instability may arise. The choice in [24] and [8] also helps simplify the k-th planar step, skipping some computation.
On the contrary, the planar methods in [17] and [9] (see also [11]) adopt a more general criterion in item 1., since they both possibly apply a planar step also in case $p_k^T A p_k$ is nearly zero, preventing numerical instabilities. Note that the latter choice allows more flexibility, but also requires more computation and an additional difficulty to prove the nonsingularity of the matrix in (3.4), using algebraic arguments.

As regards item 2., for computational reasons the choice of q_k in [24] and [8] is respectively done in such a way that

$$q_k^T A q_k = 0, \tag{3.5}$$

$$q_k^T p_k = 0. \tag{3.6}$$

It can be shown, using algebraic considerations, that by the latter choices the pair of vectors (p_k, q_k) at step k identifies a 2-dimensional linear manifold. This in turn is used to prove that the choice of the criterion (see item 1.) yields a nonsingular matrix in (3.4), i.e. equivalently the k-th planar step is well-posed. We strongly remark that [11] possibly provides an appealing geometric viewpoint, which can be straightforwardly used to replace and simplify several algebraic considerations in [24] and [8].
On the other hand, at step k the search direction q_k is computed by the planar methods in [17] and [9] as $q_k = Ap_k + \sigma_{k-1} p_{k-1}$, where $\sigma_{k-1} \in \mathbb{R}$ and is such that $q_k^T A p_{k-1} = 0$ (or $q_k^T A q_{k-1} = 0$, depending on the chance that the previous step was the planar $(k-1)$-th step).

We complete this section by computing the final expression of the iterate y_{k+2} in (3.2), at the end of the k-th planar step, using planar algorithms. For the sake of brevity, and in view of partially anticipating some considerations contained also in Sect.5, we perform the computation considering only the planar algorithms [24] and [8]. In this regard, by [24] we have for the k-th planar step:

$$q_k = Ap_k - \frac{(Ap_k)^T A(Ap_k)}{2\|Ap_k\|^2} p_k, \qquad \alpha_k = -\frac{(Ap_k)^T A(Ap_k)}{2\|Ap_k\|^4} r_k^T p_k, \qquad \beta_k = \frac{r_k^T p_k}{\|Ap_k\|^2},$$

so that by (3.2) we finally obtain

$$y_{k+2} = y_k + \alpha_k p_k + \beta_k p_{k+1} = y_k - \frac{(Ap_k)^T A(Ap_k)}{\|Ap_k\|^4}(r_k^T p_k)p_k + \frac{r_k^T p_k}{\|Ap_k\|^2} Ap_k. \quad (3.7)$$

Similarly, from [8] we have at the k-th planar step (for any $\gamma_k \in \mathbb{R} \setminus \{0\}$)

$$q_k = \gamma_k Ap_k, \qquad \alpha_k = -\frac{r_k^T p_k}{\gamma_k^2 \|Ap_k\|^4} q_k^T Aq_k, \qquad \beta_k = \frac{r_k^T p_k}{\gamma_k \|Ap_k\|^2},$$

so that again (3.2) yields

$$y_{k+2} = y_k - \frac{r_k^T p_k}{\|Ap_k\|^4}(Ap_k)^T A(Ap_k) p_k + \frac{r_k^T p_k}{\|Ap_k\|^2} Ap_k, \quad (3.8)$$

which coincides, as expected, with (3.7) (an analogous result holds using the planar methods in [17] and [9].). The latter fact should not sound surprising, inasmuch as starting from the iterate y_k, both [24] and [8] determine y_{k+2} as the stationary point on the same 2-dimensional manifold, spanned by p_k and Ap_k. We will see how to large extent, the use of *grossone* in Sect. 5 recovers the latter result.

4 Introduction to the algebra of *grossone*

Moving away from the traditional approaches of calculus, a new computational methodology allowing one to work numerically with infinities and infinitesimals was proposed in [31,33,35,39,40]. The method suggests a more accurate lens of observation of the infinite and infinitesimal quantities, and gives the opportunity to execute numerical computations with these numbers in a unique framework with finite quantities. This approach proposes a numeral system that uses *the same numerals* in all the occasions we need infinities and infinitesimals. It is important to emphasize that this numeral system avoids situations like $\infty - 1 = \infty$ and $\infty + 1 = \infty$, providing results ensuring that if a is a numeral written in this system then for any a (i.e., a can be finite, infinite, or infinitesimal) it follows $a - 1 < a$ and $a + 1 > a$. A number of papers connecting the new approach to the historical panorama of ideas dealing with infinities and infinitesimals (see [22,25,43]) has been published, and metamathematical investigations on the new theory and its non-contradictory can be found in [23,42].

This computational methodology has already been successfully applied in optimization and numerical differentiation (see [6,7,36,48]) and in a number of other theoretical and computational research areas such as cellular automata (see [4,5]), percolation (see [20,21,47]), fractals (see for instance [32,34,37,41,47]), Turing machines and supertasks (see [29,43,44]), numerical solution of ordinary differential equations (see [1,26,38], along with [45]).

The new methodology uses an infinite unit of measure expressed by the numeral ① called *grossone*, that is the number of elements of the set, $\mathbb{N}$, of natural numbers. Grossone is introduced by describing its properties (similarly, in order to pass from natural to integer numbers, a new element – zero – is introduced by describing its properties) postulated by the *Infinite Unit Axiom* consisting of three parts: Infinity, Identity, and Divisibility (see below). This axiom is added to axioms for real numbers. Moreover, it is postulated that associative and commutative properties of multiplication and addition, distributive property of multiplication over addition, existence of inverse elements with respect to addition and multiplication hold for grossone, as for finite numbers and for all numbers involving grossone.

Infinity Any finite natural number n is less than grossone, i.e., $n < ①$.
Identity The following relations link ① to identity elements 0 and 1

$$0 \cdot ① = ① \cdot 0 = 0, \quad ① - ① = 0, \quad \frac{①}{①} = 1, \quad ①^0 = 1, \quad 1^{①} = 1, \quad 0^{①} = 0. \tag{4.1}$$

Divisibility For any finite natural number n the sets $\mathbb{N}_{k,n}$, $1 \le k \le n$, being the nth parts of the set $\mathbb{N}$ of natural numbers, have the same number of elements indicated by the numeral $\frac{①}{n}$ where

$$\mathbb{N}_{k,n} = \{k, k+n, k+2n, k+3n, \ldots\}, \quad 1 \le k \le n, \quad \bigcup_{k=1}^{n} \mathbb{N}_{k,n} = \mathbb{N}. \tag{4.2}$$

To express infinite and infinitesimal numbers on a computer a numeral positional system with the infinite base ① is used. A number C in this positional system is represented through groups corresponding to powers of ①:

$$C = c_{p_m} ①^{p_m} + \cdots + c_{p_1} ①^{p_1} + c_{p_0} ①^{p_0} + c_{p_{-1}} ①^{p_{-1}} + \cdots + c_{p_{-k}} ①^{p_{-k}}. \tag{4.3}$$

Then, the record

$$C = c_{p_m} ①^{p_m} \ldots c_{p_1} ①^{p_1} c_{p_0} ①^{p_0} c_{p_{-1}} ①^{p_{-1}} \ldots c_{p_{-k}} ①^{p_{-k}} \tag{4.4}$$

represents the number C, where all numerals $c_i \neq 0$ belong to a traditional numeral system and are called *grossdigits*. They express finite positive or negative numbers, and show how many corresponding units ①^{p_i} should be added or subtracted in order to form the number C. Grossdigits can be expressed by several symbols using positional systems, the form $\frac{Q}{q}$ where Q and q are integer numbers, or in any other finite numeral system.

Numbers p_i in (4.4) called *grosspowers* can be finite, infinite, and infinitesimal (the introduction of infinitesimal numbers will be given soon), they are sorted in the decreasing order

$$p_m > p_{m-1} > \ldots > p_1 > p_0 > p_{-1} > \ldots p_{-(k-1)} > p_{-k}$$

with $p_0 = 0$. In the traditional positional systems with finite bases there exists a convention: a digit a_i shows how many powers b^i are present in the number, and the radix b is not written explicitly. In the record (4.4), we write ①^{p_i} explicitly because in the new numeral positional system the number i in general is not equal to the grosspower p_i. This gives possibility to write, for example, the infinite number $34.7$①$^{36.7}$ $15.1$①$^{8.9}$ having grosspowers $p_2 = 36.7$, $p_1 = 8.9$ and grossdigits $c_{36.7} = 34.7$, $c_{8.9} = 15.1$, without indicating grossdigits equal to zero corresponding to grosspowers less than 36.7 and greater than 8.9. Note also that if a grossdigit $c_{p_i} = 1$ then we often write ①^{p_i} instead of $1$①^{p_i}.

The term having $p_0 = 0$ represents the finite part of C because, due to (4.1), we have $c_0$①$^0 = c_0$. The terms having finite positive grosspowers represent the simplest infinite parts of C. Analogously, terms having negative finite grosspowers represent the simplest infinitesimal parts of C. For instance, the number ①$^{-1} = \frac{1}{①n}$ is infinitesimal. It is the inverse element with respect to multiplication for ①, being

$$①^{-1} \cdot ① = ① \cdot ①^{-1} = 1. \tag{4.5}$$

Note that all infinitesimals are not equal to zero. Particularly, $\frac{1}{①} > 0$ because it is a result of division of two positive numbers.

5 A novel algebraic perspective for CG degeneracy using *grossone*

This section is devoted to investigate potential advances for Krylov-based methods, by adopting the recently defined extension of real numbers using *Grossone* (see [31,35,40] and [39]), and its applications in optimization (see [2,6,7,36,48]). We are indeed persuaded that modeling CG degeneracy by means of grossone, whose properties are detailed in Sect. 4, can in general:

- easily recover the standard CG iteration also in the indefinite case, when a CG degeneracy occurs;
- provide results which perfectly match with the analysis carried on for planar CG methods;
- simplify the conclusions obtained using some planar methods.

On this purpose, we consider again the standard CG scheme in Table 1, where A is possibly indefinite nonsingular. Let us consider the formulae therein, for the computation at Step k of the steplength α_k, the residual $r_k = b - Ax_k$ and the search direction p_k. Then, we consider the following position

$$p_k^T A p_k = s①, \tag{5.1}$$

where we set $s = O(①^{-1})$ if the Step k is a non-degenerate CG step, and we set $s = O(①^{-2})$ if the Step k is a degenerate CG step. Moreover, drawing inspiration from the standard Landau-Lifsitz notation, for example with the symbol $O(①^{-2})$ we indicate *a term containing powers of* ① *at most equal to* -2. Note that in case $O(①^{-2})$ then the finite part of $p_k^T A p_k$ is equal to 0, so that the *Identity* property of Sect. 4 is fulfilled. Comparing (5.1) with the expression of C in (4.3) we immediately realize that (5.1) represents a simplified positional expression. Nevertheless, as revealed by our analysis in the sequel, for our purposes the setting (5.1) seems a (completely) sufficient choice. In particular, we want to show that the axioms and the basic algebra reported in Sect. 4 for grossone are well-suited to detail the behaviour of the CG, in the degenerate case.

We immediately warn the reader about the fact that in practice, the setting (5.1) will not alter the instructions at the k-th iteration of the CG. Thus, a remarkably valuable aspect of using grossone to cope with CG degeneracy is that the CG in Table 1 is faithfully applied 'as is', unlike what happens with planar methods. The only effect of introducing grossone in Table 1 is that, in case of CG degeneracy at Step k, the expressions of the coefficients and vectors at Step k may explicitly depend on ① and/or its powers. In this regard, in the next section we compute the expressions of the search directions p_{k+1} and p_{k+2}, when at Step k of the CG we possibly consider $p_k^T A p_k \approx 0$ along with the position (5.1) and $s = O(①^{-2})$. This will explicitly allow us to compare the use of grossone with the approaches detailed in the previous sections.

5.1 The degenerate Step k of the CG using grossone

Recalling Table 1 and (5.1), since for the CG $p_k^T r_k = \|r_k\|^2$, we immediately have

$$r_{k+1} = r_k - \alpha_k A p_k = r_k - \frac{\|r_k\|^2}{s①} A p_k \tag{5.2}$$

so that

$$\begin{aligned} p_{k+1} = r_{k+1} + \beta_k p_k \; &= \; r_k - \frac{\|r_k\|^2}{s①} A p_k - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k \\ &= r_k - \frac{\|r_k\|^2}{s①} A p_k - \left[r_k - \frac{\|r_k\|^2}{s①} A p_k \right]^T A p_k \frac{p_k}{s①} \\ &= r_k - \frac{\|r_k\|^2}{s①} A p_k - \left[p_k^T A p_k - \frac{\|r_k\|^2}{s①} \|A p_k\|^2 \right] \frac{p_k}{s①} \end{aligned}$$

$$= r_k - \frac{\|r_k\|^2}{s①} A p_k - \left[\frac{s^2①^2 - \|r_k\|^2 \|A p_k\|^2}{s^2①^2} \right] p_k$$

$$= -\beta_{k-1} p_{k-1} - \frac{\|r_k\|^2}{s①} A p_k + \frac{\|r_k\|^2 \|A p_k\|^2}{s^2①^2} p_k. \tag{5.3}$$

Now we need to compute the quantities $p_{k+1}^T A p_{k+1}$ and $r_{k+1}^T p_{k+1}$. In this regard, from (5.3) and using the relation $p_k = r_k + \beta_{k-1} p_{k-1}$, along with the orthogonality/conjugacy conditions satisfied by the residuals and search directions generated from the algorithm in Table 1

$$r_i^T r_j = 0, \qquad p_i^T A p_j = 0, \qquad\qquad \forall i \neq j,$$

we obtain, after some computation,

$$\begin{aligned} p_{k+1}^T A p_{k+1} &= \left[-\beta_{k-1} p_{k-1} - \frac{\|r_k\|^2}{s①} A p_k + \frac{\|r_k\|^2 \|A p_k\|^2}{s^2①^2} p_k \right]^T A \\ &\quad \cdot \left[-\beta_{k-1} p_{k-1} - \frac{\|r_k\|^2}{s①} A p_k + \frac{\|r_k\|^2 \|A p_k\|^2}{s^2①^2} p_k \right] \\ &= \frac{\|r_k\|^4}{s^2①^2} (A p_k)^T A (A p_k) - \frac{\|r_k\|^4 \|A p_k\|^4}{s^3①^3} + O(①), \end{aligned} \tag{5.4}$$

where the term $O(①)$ in (5.4) just contains terms with powers of ① equal to $+1$ and 0. On the other hand, exploiting the conjugacy between p_{k+1} and p_k we also have

$$\begin{aligned} r_{k+1}^T p_{k+1} &= \left[r_k - \frac{\|r_k\|^2}{s①} A p_k \right]^T p_{k+1} \; = r_k^T p_{k+1} \\ &= r_k^T \left[-\beta_{k-1} p_{k-1} - \frac{\|r_k\|^2}{s①} A p_k + \frac{\|r_k\|^2 \|A p_k\|^2}{s^2①^2} p_k \right] \\ &= -\|r_k\|^2 + \frac{\|r_k\|^4 \|A p_k\|^2}{s^2①^2}. \end{aligned} \tag{5.5}$$

Now, using (5.2), (5.4) and (5.5) we obtain

$$\begin{aligned} r_{k+2} = r_{k+1} - \alpha_{k+1} A p_{k+1} \; &= \; r_k - \frac{\|r_k\|^2}{s①} A p_k - \frac{r_{k+1}^T p_{k+1}}{p_{k+1}^T A p_{k+1}} A p_{k+1} \\ = r_k - \frac{\|r_k\|^2}{s①} A p_k &- \frac{-\|r_k\|^2 + \|r_k\|^4 \|A p_k\|^2 \frac{1}{s^2①^2}}{\frac{\|r_k\|^4}{s^2①^2} (A p_k)^T A (A p_k) - \frac{\|r_k\|^4 \|A p_k\|^4}{s^3①^3} + O(①)} \\ &\cdot \left[-\beta_{k-1} A p_{k-1} - \frac{\|r_k\|^2}{s①} A (A p_k) + \frac{\|r_k\|^2 \|A p_k\|^2}{s^2①^2} A p_k \right]; \end{aligned}$$

when $p_k^T A p_k$ is infinitesimal (i.e. the CG degenerates at Step k) so does $s①$, and the latter relation yields

$$\begin{aligned}
r_{k+2} &= r_k - \frac{\|r_k\|^2}{s①} A p_k - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} A p_{k-1} + \frac{s①}{\|Ap_k\|^2} \left(-\frac{\|r_k\|^2}{s①} A(Ap_k) \right) \\
&\quad + \frac{s①}{\|Ap_k\|^2} \left(\frac{\|r_k\|^2 \|Ap_k\|^2}{s^2 ①^2} A p_k \right) + O(①^{-1}) \\
&= r_k - \frac{\|r_k\|^2}{s①} A p_k - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} A p_{k-1} \\
&\quad - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) + \frac{\|r_k\|^2}{s①} A p_k + O(①^{-1}) \\
&= r_k - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} A p_{k-1} + O(①^{-1}),
\end{aligned} \tag{5.6}$$

being as usual $O(①^{-1})$ a vector with terms containing powers of ① at most equal to -1. Recalling that $p_k^T A p_k = s①$ is infinitesimal, the most significant consequence from (5.2) and (5.6) is that in practice

- the residuals $r_1, \ldots, r_k$ are *independent* of ①,
- r_{k+1} *depends* on ①,
- r_{k+2} is *independent* of negative powers of $s①$,

which implies that applying the standard CG in Table 1, also in case the pivot breakdown $p_k^T A p_k \approx 0$ occurs at Step k, then the sequence of generated residuals $r_1, \ldots, r_{k+2}$, *includes all vectors in* $\mathbb{R}^n$ *apart from* $r_{k+1} \in \hat{\mathbb{R}}^n$, *provided that the terms containing* $s①$ *in* (5.6) *are neglected*. Thus, the algebra related to CG degeneracy at Step k, detailed in the previous sections of the present paper, can be overcome by introducing grossone and neglecting the term with $s①$ in (5.6), leaving unchanged the CG scheme in Table 1.

Furthermore, let us now compute the search direction p_{k+2}, being by (5.2) and (5.6)

$$\begin{aligned}
p_{k+2} = r_{k+2} + \beta_{k+1} p_{k+1} \;&=\; r_k - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) \\
&\quad + \frac{\|r_{k+2}\|^2}{\|r_{k+1}\|^2} p_{k+1} - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} A p_{k-1} + O(①^{-1}) \\
&= r_k - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) + \frac{\|r_{k+2}\|^2}{-\|r_k\|^2 + \frac{\|r_k\|^4 \|Ap_k\|^2}{s^2 ①^2}} p_{k+1} \\
&\quad - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} A p_{k-1} + O(①^{-1}).
\end{aligned} \tag{5.7}$$

Now, recalling (5.3) we can write

$$\begin{aligned} p_{k+2} &= r_k - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} Ap_{k-1} + O(①^{-1}) \\ &\quad + \frac{\|r_{k+2}\|^2}{-\|r_k\|^2 + \frac{\|r_k\|^4 \|Ap_k\|^2}{s^2 ①^2}} \cdot \left[\frac{\|r_k\|^2 \|Ap_k\|^2}{s^2 ①^2} p_k - \frac{\|r_k\|^2}{s①} Ap_k - \beta_{k-1} p_{k-1} \right] \\ &= r_k - \frac{\|r_k\|^2}{\|Ap_k\|^2} A(Ap_k) + \frac{\|r_{k+2}\|^2}{\|r_k\|^2} p_k - \beta_{k-1} \frac{s①}{\|Ap_k\|^2} Ap_{k-1} + O(①^{-1}), \end{aligned} \tag{5.8}$$

which shows that, similarly to r_{k+2}, also p_{k+2} can be viewed as an n-real vector which does not depend on negative powers of ① (i.e. equivalently $\|p_{k+2}\| < ①$). Another remarkable result from (5.8) is that, neglecting the terms which contain powers of $s①$ larger or equal to 1, after a simple computation applying Algorithm CG_Plan of [8] the vector p_{k+2} in (5.8) coincides with the expression of p_{k+2} in [8] (a similar result holds using the algorithm by Luenberger in [24]). This implies that the use of grossone to deal with a CG degeneracy at Step k does not simply allow to generate the mutually conjugate directions $p_1, \ldots, p_k, p_{k+2}$ as planar methods do (see Sect. 3.1), but *it also retrieves the same scaling of the search directions provided by some planar methods*.

After some computations, using standard properties of the CG, it is also easy to verify that the vectors $r_{k+1}, p_{k+1}, r_{k+2}, p_{k+2}$ in (5.2), (5.3), (5.6) and (5.8) satisfy

$$\begin{cases} r_{k+2}^T r_i = 0, & i = 1, \ldots, k+1, \\ p_{k+2}^T A p_i = 0, & i = 1, \ldots, k+1, \end{cases} \tag{5.9}$$

so that recurring to grossone in case of pivot breakdown of the CG allows to retrieve standard CG properties, *even in the degenerate case*.

In addition, since the expression of p_{k+1} in (5.3) explicitly includes negative powers of $s①$, then we have a perfect matching with the results in Proposition 3.1. Indeed, being the k-th CG step degenerate, then vectors whose entries contain negative powers of $s①$ to large extent can be assimilated to vectors with unbounded norm.

Now, in order to verify to what extent the use of grossone completely recovers the CG iteration also in case of degeneracy at Step k, let us compute the iterate y_{k+2}, similarly to what we have done in (3.7) and (3.8) using planar methods. After a simple computation we first obtain from Table 1

$$y_{k+2} \;=\; y_k + \alpha_k p_k + \alpha_{k+1} p_{k+1}$$

where (5.1) yields

$$\alpha_k \;=\; \frac{\|r_k\|^2}{p_k^T A p_k} \;=\; \frac{\|r_k\|^2}{s①},$$

and by (5.4) along with (5.5)

$$\alpha_{k+1} = \frac{\|r_{k+1}\|^2}{p_{k+1}^T A p_{k+1}} \;=\; \frac{r_{k+1}^T p_{k+1}}{p_{k+1}^T A p_{k+1}}$$

$$
\begin{aligned}
&= \frac{-\|r_k\|^2 + \|r_k\|^4 \|Ap_k\|^2 \frac{1}{s^2 ①^2}}{\frac{\|r_k\|^4}{s^2 ①^2}(Ap_k)^T A(Ap_k) - \frac{\|r_k\|^4 \|Ap_k\|^4}{s^3 ①^3} + O(①)} \\
&= -\frac{s①}{\|Ap_k\|^2} - \frac{s^2 ①^2}{\|Ap_k\|^6}(Ap_k)^T A(Ap_k) \\
&\quad + \frac{\|Ap_k\|^6 - \|r_k\|^2[(Ap_k)^T A(Ap_k)]^2}{\|r_k\|^2 \|Ap_k\|^{10}} s^3 ①^3 + O(①^{-4}) \qquad (5.10)
\end{aligned}
$$

where $O(①^{-4})$ represents the sum of terms containing powers of ① smaller or equal to -4, and the last equality can be verified by direct computation. Then, we obtain from (5.3)

$$
\begin{aligned}
y_{k+2} &= y_k + \frac{\|r_k\|^2}{s①} p_k + \Bigg[-\frac{s①}{\|Ap_k\|^2} - \frac{s^2 ①^2}{\|Ap_k\|^6}(Ap_k)^T A(Ap_k) \\
&\quad + \frac{\|Ap_k\|^6 - \|r_k\|^2[(Ap_k)^T A(Ap_k)]^2}{\|r_k\|^2 \|Ap_k\|^{10}} s^3 ①^3 \\
&\quad + O(①^{-4}) \Bigg] \cdot \left[-\beta_{k-1} p_{k-1} - \frac{\|r_k\|^2}{s①} Ap_k + \frac{\|r_k\|^2 \|Ap_k\|^2}{s^2 ①^2} p_k \right] \\
&= y_k + \frac{\|r_k\|^2}{s①} p_k + \frac{\|r_k\|^2}{\|Ap_k\|^2} Ap_k - \frac{\|r_k\|^2}{s①} p_k \\
&\quad - \frac{\|r_k\|^2}{\|Ap_k\|^4}(Ap_k)^T A(Ap_k) p_k + O(①^{-1}) \\
&= y_k + \frac{\|r_k\|^2}{\|Ap_k\|^2} Ap_k - \frac{\|r_k\|^2}{\|Ap_k\|^4}(Ap_k)^T A(Ap_k) p_k + O(①^{-1}), \qquad (5.11)
\end{aligned}
$$

which coincides exactly with (3.7) and (3.8) as long as $O(①^{-1})$ is neglected. This proves that in case of CG degeneracy at Step k, the use of grossone does not simply allow to compute residuals and search directions as in (5.9), but it also may provide exactly the same iterate y_{k+2} of [24] and [8], *independent of grossone*. The latter result can be summarized in the next proposition, highlighting how introducing grossone can perfectly recover a CG degeneracy, using the algebra of $\hat{\mathbb{R}}^n$ in place of $\mathbb{R}^n$.

Proposition 5.1 *Consider the indefinite linear system $Ay = b$, with $A \in \mathbb{R}^{n\times n}$, and suppose the CG in Table 1 is applied for its solution, where at Step k possibly the finite part of $p_k^T Ap_k$ is equal to 0 and the setting (5.1) is adopted. Then, at Step k and Step $k+1$ the CG preserves in $\hat{\mathbb{R}}^n$ the same properties of the CG (applied in $\mathbb{R}^n$), ignoring any degeneracy. Moreover, y_{k+2} is given by the finite part of (5.11) and p_{k+2} is given by the finite part of (5.8).*

The resulting CG method in $\hat{\mathbb{R}}^n$ (namely CG$_{①}$), introduced in Proposition 5.1, is detailed in Table 2. Note that in Table 2, in case at Step k we have $p_k^T Ap_k \approx 0$, then the test on r_{k+1} is unnecessary, being r_{k+1} computed by (5.2) with $s① \approx 0$. Observe that by comparing (3.7)-(3.8) and (5.11) we can immediately realize the additional

Table 2 The $CG_{①}$ algorithm for solving the symmetric linear system $Ay = b$, $A \in \mathbb{R}^{n\times n}$

The Conjugate Gradient method with grossone ($CG_{①}$)	
Data:	Set $k = 0$, $y_0 = 0$, $r_0 = b - Ay_0$, $\varepsilon > 0$, $s = O(①^{-2})$. If $r_0 = 0$, then STOP. Else, set $p_0 = r_0$

Step k:	If $\|p_k\|$ is finite and $\|p_k^T Ap_k\| \geq \varepsilon \|p_k\|^2$ then
	Compute $\alpha_k = r_k^T p_k / p_k^T Ap_k$, $y_{k+1} = y_k + \alpha_k p_k$, $r_{k+1} = r_k - \alpha_k Ap_k$
	If $r_{k+1} = 0$, then STOP
	Elseif $\|p_k\|$ is finite set $p_k^T Ap_k = s①$ and compute r_{k+1} by (5.2)
	Else compute $\alpha_k = r_k^T p_k / p_k^T Ap_k$, $y_{k+1} = y_k + \alpha_k p_k$, $r_{k+1} = r_k - \alpha_k Ap_k$
	If the finite part of r_{k+1} is zero, then STOP
	Endif
	Set $\beta_k = -r_{k+1}^T Ap_k / p_k^T Ap_k = \|r_{k+1}\|^2 / \|r_k\|^2$, and
	$p_{k+1} = r_{k+1} + \beta_k p_k$, $k = k + 1$
	Go to Step k

contribution given by the use of grossone, with respect to [24] and [8] (indeed, the term $O(①^{-1})$ can strongly affect the final iterate y_{k+2}).

5.2 How the use of grossone can underly CG degeneracy

In this section we informally show how the geometry behind CG degeneracy and the use of grossone, in the nonsingular indefinite case, can also justify the conclusions of Proposition 3.1. This fact is depicted in Fig. 1, where we have considered the three iterates y_m, y_h and y_k generated by the CG, along with the corresponding search directions p_m, p_h and p_k. In Fig. 1 the continuous line represents the level set

$$\{y \in \mathbb{R}^n \ : \ q(y) = \omega\}, \quad \text{where } q(y) = \frac{1}{2} y^T Ay - b^T y, \quad \omega \in \mathbb{R}.$$

At y_m and y_h no degeneracy occurs (i.e. the standard CG method in Table 1 applies), while at y_k we have $p_k^T Ap_k = 0$. In particular at y_m (a similar conclusion holds for the iterate y_h) no degeneracy of the CG is observed, which is evident by the fact that the vector p_{m+1} is geometrically constructed joining y_m and y'_m (the latter point being symmetric of y_m with respect to the point y^*, which satisfies $Ay^* = b$). On the contrary, such a reasoning can not be replicated for the computation of p_{k+1}, because p_k is not tangent at y_k to the continuous line in Fig. 1. Equivalently, the line $y_k + \alpha p_k$, $\alpha \in \mathbb{R}$, is tangent to another level set (dashed and dotted line), in the point at infinity y_{k+1} (for a more rigorous justification of the last statement the reader can refer to [12]). Then, in order to formally compute the next *finite iterate* y_{k+2}, the search direction p_{k+1} satisfying $\|p_{k+1}\| \rightarrow +\infty$ should be provided, as proved in Proposition 3.1. Equivalently, when at Step k of $CG_{①}$ in Table 2 the position (5.1) is adopted, then

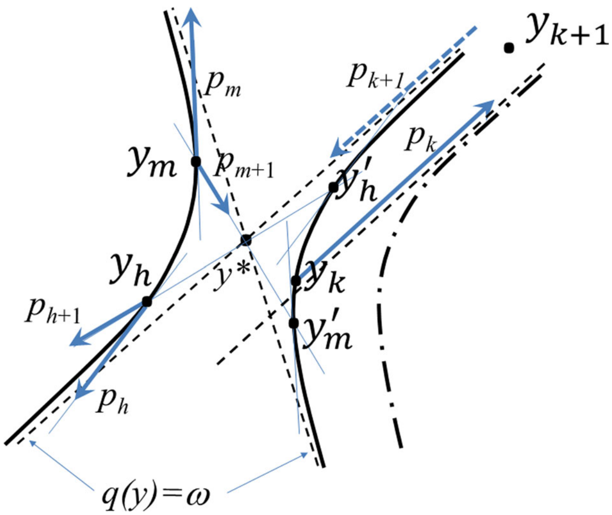

Fig. 1 The level set (continuous line) $\{y \in \mathbb{R}^n \ : \ 1/2y^T Ay - b^T y = \omega\}$, with $\omega \in \mathbb{R}$, being A indefinite nonsingular. The solution point y^* of $Ay = b$ is in the intersection of the asymptotes (dashed lines). At the current Step k we have $p_k^T Ap_k \approx 0$, so that y_{k+1} approaches a point at infinity, according with Proposition 3.1. In order to generate the *finite* point y_{k+2}, the next conjugate direction p_{k+1} needs to satisfy $\|p_{k+1}\| \to +\infty$ (dashed arrow in the figure)

p_{k+1} is computed as in (5.3), so that $s① \approx 0$ again yields the conclusion of Proposition 3.1.

6 Conclusions

In this paper we presented an innovative perspective and implementation of the Conjugate Gradient method, in the case of degeneracy on indefinite linear systems. The proposed approach utilizes the new computational methodology based on ① (namely *grossone*), proposed by Sergeyev and successfully used in Nonlinear Optimization frameworks. Our proposal fits the well known scheme of planar methods for CG degeneracy.

We are persuaded that the analysis detailed in this paper might be also fruitfully adopted to analyze Nonlinear Conjugate Gradient methods. The latter techniques are indeed extensions of the CG to non-quadratic functions, and require specific care when computing the steplength along the current search direction. In this regard, on one hand the use of grossone can be the right tool to handle numerical instabilities; on the other hand, the theory of Polarity (see for instance [30]) might suggest useful extensions of the asymptotic cone (see [12]).

Acknowledgements 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. The research of Ya.D. Sergeyev was supported by the Russian Science Foundation, Project No. 15-11-30022 "Global optimization, supercomputing computations, and applications".

Appendix

Proof of Proposition 2.1 Recalling that $n = P + N$, let us consider for $\tilde{A}^{-1}$ the positive definite matrix

$$\tilde{A}^{-1} = \sum_{i=1}^{P} \frac{1}{p_i^T A p_i} p_i p_i^T - \sum_{i=P+1}^{P+N} \frac{1}{p_i^T A p_i} p_i p_i^T;$$

then (ii)–(iii) follow from the conjugacy of the directions $p_1, \ldots, p_n$ with respect to A. Moreover, we have after a brief arrangement

$$\tilde{A}^{-1} = V \left(\begin{array}{c|c} \mathrm{diag}_{p_i \in P1} \left\{ \frac{1}{p_i^T A p_i} \right\} & \emptyset \\ \hline \emptyset & -\,\mathrm{diag}_{p_i \in P2} \left\{ \frac{1}{p_i^T A p_i} \right\} \end{array} \right) V^T, \tag{6.1}$$

being

$$V = \left(p_1 \vdots \cdots \vdots p_P \vdots p_{P+1} \vdots \cdots \vdots p_{P+N} \right) \in \mathbb{R}^{n \times n}$$

a nonsingular matrix. Then, by (6.1) we obtain

$$\tilde{A} = V^{-T} \left(\begin{array}{c|c} \mathrm{diag}_{p_i \in P1} \{ p_i^T A p_i \} & \emptyset \\ \hline \emptyset & -\,\mathrm{diag}_{p_i \in P2} \{ p_i^T A p_i \} \end{array} \right) V^{-1}$$

hence

$$V^T \tilde{A} V = \left(\begin{array}{c|c} \mathrm{diag}_{p_i \in P1} \{ p_i^T A p_i \} & \emptyset \\ \hline \emptyset & -\,\mathrm{diag}_{p_i \in P2} \{ p_i^T A p_i \} \end{array} \right), \tag{6.2}$$

showing that (iv) holds. Finally, by Table 1 and the properties of the CG, let us consider the expression of the coefficients

$$\alpha_i = \frac{\|r_i\|^2}{p_i^T A p_i} = \frac{r_0^T p_i}{p_i^T A p_i} = \frac{b^T p_i}{p_i^T A p_i}, \qquad i \geq 1.$$

Then, (6.2) implies

$$\begin{cases} p_i^T \tilde{A} p_i = p_i^T A p_i, & \forall p_i \in P1, \\ p_i^T \tilde{A} p_i = -p_i^T A p_i, & \forall p_i \in P2. \end{cases} \tag{6.3}$$

Now, to prove (v) it suffices to show that the condition $\tilde{A}(d_P + d_N) = b$ holds, i.e.

$$\tilde{A}(d_P + d_N) = b \iff \left[b - \tilde{A}(d_P + d_N) \right]^T p_i = 0, \quad i = 1, \ldots, P + N.$$

The result follows by (6.3), being either

$$\left[b - \tilde{A}(d_P + d_N)\right]^T p_i = b^T p_i - \alpha_i p_i^T \tilde{A} p_i = b^T p_i \\ - \frac{b^T p_i}{p_i^T \tilde{A} p_i} p_i^T \tilde{A} p_i = 0, \quad \forall i = 1, \ldots, P,$$

or

$$\left[b - \tilde{A}(d_P + d_N)\right]^T p_i = b^T p_i + \alpha_i p_i^T \tilde{A} p_i = b^T p_i \\ + \frac{b^T p_i}{-p_i^T \tilde{A} p_i} p_i^T \tilde{A} p_i = 0, \quad \forall i = P+1, \ldots, P+N.$$

□

References

1. Amodio, P., Iavernaro, F., Mazzia, F., Mukhametzhanov, M.S., Sergeyev, Y.D.: A generalized Taylor method of order three for the solution of initial value problems in standard and infinity floating-point arithmetic. Math. Comput. Simulat. **141**, 24–39 (2017)
2. Cococcioni, M., Pappalardo, M., Sergeyev, Y.D.: Towards lexicographic multi-objective linear programming using grossone methodology. In: Sergeyev, Y.D., Kvasov, D.E., Dell'Accio, F., Mukhametzhanov, M.S. (eds.) Proceedings of the 2nd International Conference on "Numerical Computations: Theory and Algorithms", vol. 1776, p. 090040. AIP Publishing, New York (2016)
3. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-Region Methods. MPS–SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia (2000)
4. D'Alotto, L.: Cellular automata using infinite computations. Appl. Math. Comput. **218**(16), 8077–8082 (2012)
5. D'Alotto, L.: A classification of one-dimensional cellular automata using infinite computations. Appl. Math. Comput. **255**, 15–24 (2015)
6. De Cosmis, S., De Leone, R.: The use of grossone in mathematical programming and operations research. Appl. Math. Comput. **218**(16), 8029–8038 (2012)
7. De Leone, R.: Nonlinear programming and grossone: quadratic programming and the role of constraint qualifications. Appl. Math. Comput. **318**, 290–297 (2018)
8. Fasano, G.: Conjugate gradient (CG)-type method for the solution of Newton's equation within optimization frameworks. Optim. Methods Softw. **3–4**, 267–290 (2004)
9. Fasano, G.: Planar conjugate gradient algorithm for large-scale unconstrained optimization, part 1: theory. J. Optim. Theory Appl. **125**, 523–541 (2005a)
10. Fasano, G.: Planar conjugate gradient algorithm for large-scale unconstrained optimization, part 2: application. J. Optim. Theory Appl. **125**, 543–558 (2005b)
11. Fasano, G.: Lanczos-conjugate gradient method and pseudoinverse computation, on indefinite and singular systems. J. Optim. Theory Appl. **132**, 267–285 (2007)
12. Fasano, G., Pesenti, R.: Conjugate direction methods and polarity for quadratic hypersurfaces. J. Optim. Theory Appl. (2018) (**in press**)
13. Greenbaum, A.: Iterative Methods for Solving Linear Systems. SIAM, Philadelphia (1997)
14. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newtons method. SIAM J. Numer. Anal. **23**, 707–716 (1986)
15. Grippo, L., Lampariello, F., Lucidi, S.: A truncated Newton method with nonmonotone line search for unconstrained optimization. J. Optim. Theory Appl. **60**, 401–419 (1989)
16. Hager, W., Zhang, H.: The limited memory conjugate gradient method. SIAM J. Optim. **23**, 2150–2168 (2013)
17. Hestenes, M.R.: Conjugate Direction Methods in Optimization. Springer, New York (1980)

18. Hestenes, M.R., Stiefel, E.L.: Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Stand. **49**, 409–436 (1952)
19. Huang, H.Y.: Unified approach to quadratically convergent algorithms for function minimization. J. Optim. Theory Appl. **5**, 405–423 (1970)
20. Iudin, D.I., Sergeyev, Y.D., Hayakawa, M.: Interpretation of percolation in terms of infinity computations. Appl. Math. Comput. **218**(16), 8099–8111 (2012)
21. Iudin, D.I., Sergeyev, Y.D., Hayakawa, M.: Infinity computations in cellular automaton forest-fire model. Commun. Nonlinear Sci. Numer. Simul. **20**(3), 861–870 (2015)
22. Lolli, G.: Infinitesimals and infinites in the history of mathematics: a brief survey. Appl. Math. Comput. **218**(16), 7979–7988 (2012)
23. Lolli, G.: Metamathematical investigations on the theory of grossone. Appl. Math. Comput. **255**, 3–14 (2015)
24. Luenberger, D.G.: Hyperbolic pairs in the method of conjugate gradients. SIAM J. Appl. Math. **17**, 1263–1267 (1969)
25. Margenstern, M.: Using grossone to count the number of elements of infinite sets and the connection with bijections. p-Adic Numbers Ultrametr. Anal. Appl. **3**(3), 196–204 (2011)
26. Mazzia, F., Sergeyev, Y.D., Iavernaro, F., Amodio, P., Mukhametzhanov, M.S.: Numerical methods for solving ODEs on the Infinity Computer. In: Sergeyev, Y.D., Kvasov, D.E., Dell'Accio, F., Mukhametzhanov, M.S. (eds.) Proceedings of the 2nd International Conference on "Numerical Computations: Theory and Algorithms", vol. 1776, p. 090033. AIP Publishing, New York (2016)
27. Nocedal, J., Wright, S.: Numerical Optimization, 2nd edn. Springer, New York (2006)
28. Oren, S.S.: Planar quasi-Newton algorithms for unconstrained saddlepoint problems. J. Optim. Theory Appl. **43**, 167–204 (1984)
29. Rizza, D.: Supertasks and numeral systems. In: Sergeyev, Y.D., Kvasov, D.E., Dell'Accio, F., Mukhametzhanov, M.S. (eds.) Proceedings of the 2nd International Conference on "Numerical Computations: Theory and Algorithms", vol. 1776, p. 090005. AIP Publishing, New York (2016)
30. Seidenberg, A.: Lecture in Projective Geometry. Dover Publications Inc, Mineola (2005)
31. Sergeyev, Y.D.: Arithmetic of Infinity, 2nd edn. Edizioni Orizzonti Meridionali, Cosenza (2013)
32. Sergeyev, Y.D.: Blinking fractals and their quantitative analysis using infinite and infinitesimal numbers. Chaos Solitons Fractals **33**(1), 50–75 (2007)
33. Sergeyev, Y.D.: A new applied approach for executing computations with infinite and infinitesimal quantities. Informatica **19**(4), 567–596 (2008)
34. Sergeyev, Y.D.: Evaluating the exact infinitesimal values of area of Sierpinski's carpet and volume of Menger's sponge. Chaos Solitons Fractals **42**(5), 3042–3046 (2009)
35. Sergeyev, Y.D.: Lagrange lecture: methodology of numerical computations with infinities and infinitesimals. Rendiconti del Seminario Matematico dell'Università e del Politecnico di Torino **68**(2), 95–113 (2010)
36. Sergeyev, Y.D.: Higher order numerical differentiation on the Infinity Computer. Optim. Lett. **5**(4), 575–585 (2011a)
37. Sergeyev, Y.D.: Using blinking fractals for mathematical modelling of processes of growth in biological systems. Informatica **22**(4), 559–576 (2011b)
38. Sergeyev, Y.D.: Solving ordinary differential equations by working with infinitesimals numerically on the Infinity Computer. Appl. Math. Comput. **219**(22), 10668–10681 (2013)
39. Sergeyev, Y.D.: Computations with grossone-based infinities. In: Calude, C.S., Dinneen, M.J. (eds.) Unconventional Computation and Natural Computation: Proceedings of the 14th International Conference UCNC 2015, vol. LNCS 9252, pp. 89–106. Springer, New York (2015)
40. Sergeyev, Y.D.: Un semplice modo per trattare le grandezze infinite ed infinitesime. Mat. nella Soc. nella Cult. Riv. della Unione Mat. Ital. **8**(1), 111–147 (2015)
41. Sergeyev, Y.D.: The exact (up to infinitesimals) infinite perimeter of the Koch snowflake and its finite area. Commun. Nonlinear Sci. Numer. Simul. **31**(1–3), 21–29 (2016)
42. Sergeyev, Y.D.: Numerical infinities and infinitesimals: methodology, applications, and repercussions on two Hilbert problems. EMS Surveys Math. Sci. (2017) (**in press**)
43. Sergeyev, Y.D., Garro, A.: Observability of turing machines: a refinement of the theory of computation. Informatica **21**(3), 425–454 (2010)
44. Sergeyev, Y.D., Garro, A.: Single-tape and multi-tape turing machines through the lens of the grossone methodology. J. Supercomput. **65**(2), 645–663 (2013)

45. Sergeyev, Y.D., Mukhametzhanov, M.S., Mazzia, F., Iavernaro, F., Amodio, P.: Numerical methods for solving initial value problems on the Infinity Computer. Int. J. Unconv. Comput. **12**(1), 3–23 (2016)
46. Trefethen, L.N., Bau, D.: Numerical Linear Algebra. SIAM, Philadelphia (1997)
47. Vita, M.C., Bartolo, S.D., Fallico, C., Veltri, M.: Usage of infinitesimals in the Menger's Sponge model of porosity. Appl. Math. Comput. **218**(16), 8187–8196 (2012)
48. Žilinskas, A.: On strong homogeneity of two global optimization algorithms based on statistical models of multimodal objective functions. Appl. Math. Comput. **218**(16), 8131–8136 (2012)