

qm  
27

qm

quaderni di matematica  
volume 27

Series edited by  
Dipartimento di Matematica  
Seconda Università di Napoli  
Caserta - 2012

Recent Advances in Nonlinear Optimization  
and Equilibrium Problems: a Tribute to Marco D'Apuzzo

**Recent Advances  
in Nonlinear Optimization  
and Equilibrium Problems:  
a Tribute to Marco D'Apuzzo**

edited by  
Valentina De Simone, Daniela di Serafino, and Gerardo Toraldo



V. De Simone  
D. di Serafino  
G. Toraldo



# **quaderni di matematica**

volume 27

edited by

Dipartimento di Matematica  
Seconda Università di Napoli

*Published with the support of  
Seconda Università di Napoli*

## quaderni di matematica

### Published volumes

- 1 - Classical Problems in Mechanics (R. Russo ed.)
- 2 - Recent Developments in Partial Differential Equations (V. A. Solonnikov ed.)
- 3 - Recent Progress in Function Spaces (G. Di Maio and Ľ. Holá eds.)
- 4 - Advances in Fluid Dynamics (P. Maremonti ed.)
- 5 - Methods of Discrete Mathematics (S. Löwe, F. Mazzocca, N. Melone, and U. Ott eds.)
- 6 - Connections between Model Theory and Algebraic and Analytic Geometry (A. Macintyre ed.)
- 7 - Homage to Gaetano Fichera (A. Cialdea ed.)
- 8 - Topics in Infinite Groups (M. Curzio and F. de Giovanni eds.)
- 9 - Selected Topics in Cauchy-Riemann Geometry (S. Dragomir ed.)
- 10 - Topics in Mathematical Fluids Mechanics (G.P. Galdi and R. Rannacher eds.)
- 11 - Model Theory and Applications (L. Belair, Z. Chatzidakis et al. eds.)
- 12 - Topics in Diagram Geometry (A. Pasini ed.)
- 13 - Complexity of Computations and Proofs (J. Krajčiček ed.)
- 14 - Calculus of Variations: Topics from the Mathematical Heritage of E. De Giorgi (D. Palara ed.)
- 15 - Dispersive Nonlinear Problems in Mathematical Physics (P. D'Ancona and V. Georgiev eds.)
- 16 - Kinetic Methods for Nonconservative and Reacting Systems (G. Toscani ed.)
- 17 - Set Theory: Recent Trends and Applications (A. Andretta ed.)
- 18 - Selection Principles and Covering Properties in Topology (Lj.D.R. Kočinac ed.)
- 19 - Trends in Incidence and Galois Geometries: a Tribute to Giuseppe Tallini (F. Mazzocca, N. Melone, and D. Olanda eds.)
- 20 - Mathematical Modelling of Bodies with Complicated Bulk and Boundary Behavior (M. Šilhavý ed.)
- 21 - Vector Bundles and Low Codimensional Subvarieties: State of the Art and Recent Developments (G. Casnati, F. Catanese, and R. Notari eds.)
- 22 - Theory and Applications of Proximity, Nearness and Uniformity (G. Di Maio and S. Naimpally eds.)
- 23 - On the Notions of Solution to Nonlinear Elliptic Problems: Results and Developments (A. Alvino, A. Mercaldo, F. Murat, and I. Peral eds.)
- 24 - Numerical Methods for Balance Laws (G. Puppo and G. Russo eds.)
- 25 - Stochastic Partial Differential Equations and Applications (G. Da Prato and L. Tubaro eds.)
- 26 - Asymmetric Topology and its Applications (J. Rodríguez-López and S. Romaguera eds.)

Recent Advances in  
Nonlinear Optimization  
and Equilibrium Problems:  
a Tribute to Marco D'Apuzzo

edited by Valentina De Simone, Daniela di Serafino, and Gerardo Toraldo

---

Received July 2012

---

© 2012 by Dipartimento di Matematica della Seconda Università di Napoli

---

Photocomposed copy prepared from a  $\LaTeX$  file.

ISBN 978-88-548-5687-5

Authors' addresses

Annabella Astorino  
Istituto di Calcolo e Reti  
ad Alte Prestazioni,  
Consiglio Nazionale delle Ricerche,  
Via P. Bucci, cubo 41C  
I-87036 Rende (CS), Italy  
Email: [astorino@icar.cnr.it](mailto:astorino@icar.cnr.it)

Giancarlo Bigi  
Dipartimento di Informatica  
Università degli Studi di Pisa  
Largo B. Pontecorvo, 3  
I-56127 Pisa, Italy  
Email: [giancarlo.bigidi.unipi.it](mailto:giancarlo.bigidi.unipi.it)

Silvia Bonettini  
Dipartimento di Matematica  
Università degli Studi di Ferrara  
Polo Scientifico Tecnologico, Blocco B  
Via Saragat, 1  
I-44100 Ferrara, Italy  
Email: [silvia.bonettini@unife.it](mailto:silvia.bonettini@unife.it)

Marco Castellani  
Dipartimento di Sistemi ed  
Istituzioni per l'Economia  
Università degli Studi dell'Aquila  
Via G. Falcone, 25  
I-67100 L'Aquila, Italy  
Email: [marco.castellani@univaq.it](mailto:marco.castellani@univaq.it)

Gianni Di Pillo  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale "A. Ruberti"  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [dipillo@dis.uniroma1.it](mailto:dipillo@dis.uniroma1.it)

Stefania Bellavia  
Dipartimento di Energetica "S. Stecco"  
Università degli Studi di Firenze  
Viale G.B. Morgagni, 40  
I-50134 Firenze, Italy  
Email: [stefania.bellavia@unifi.it](mailto:stefania.bellavia@unifi.it)

Immanuel M. Bomze  
Department of Statistics and  
Operations Research  
University of Vienna  
Universitätsstraße 5  
A-1010 Wien, Austria  
Email: [immanuel.bomze@univie.ac.at](mailto:immanuel.bomze@univie.ac.at)

Maria Paula Brito  
Faculdade de Economia &  
LIAAD-INESC Porto LA  
Universidade do Porto  
Rua Dr. Roberto Frias  
4200-464 Porto, Portugal  
Email: [mpbrito@fep.up.pt](mailto:mpbrito@fep.up.pt)

Roberto Cavicchioli  
Dipartimento di Matematica Pura  
ed Applicata  
Università di Modena e Reggio Emilia  
Via Campi, 213/B  
I-41125 Modena, Italy  
Email: [roberto.cavicchioli@unimore.it](mailto:roberto.cavicchioli@unimore.it)

Francisco Facchinei  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale "A. Ruberti"  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [facchinei@dis.uniroma1.it](mailto:facchinei@dis.uniroma1.it)

Giovanni Fasano  
Dipartimento di Management  
Università Ca Foscari Venezia  
S. Giobbe, Cannaregio 873  
I-30121 Venezia, Italy  
Email: [fasano@unive.it](mailto:fasano@unive.it)

Emanuele Galligani  
Dipartimento di Matematica Pura  
ed Applicata  
Università di Modena e Reggio Emilia  
Via Campi, 213/b  
I-41125, Modena, Italy  
Email: [emanuele.galligani@unimore.it](mailto:emanuele.galligani@unimore.it)

Manlio Gaudio  
Dipartimento di Elettronica, Informatica  
e Sistemistica  
Università della Calabria  
Via P. Bucci, cubo 41C  
I-87036 Rende (CS), Italy  
Email: [gaudio@deis.unical.it](mailto:gaudio@deis.unical.it)

Lorenzo Lampariello  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale "A. Ruberti"  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [lampariello@dis.uniroma1.it](mailto:lampariello@dis.uniroma1.it)

Marco Locatelli  
Dipartimento di Ingegneria Informatica  
Università degli Studi di Parma  
Viale G.P. Usberti, 181/A  
I-43100 Parma, Italy  
Email: [locatell@ce.unipr.it](mailto:locatell@ce.unipr.it)

Emanuele Frandi  
Dipartimento di Scienza e  
Alta Tecnologia  
Università degli Studi dell'Insubria  
Via Valleggio, 11  
I-22100 Como, Italy  
Email: [emanuele.frandi@insubria.it](mailto:emanuele.frandi@insubria.it)

Maria Grazia Gasparo  
Dipartimento di Energetica "S. Stecco"  
Università degli Studi di Firenze  
Viale G.B. Morgagni, 40  
I-50134 Firenze, Italy  
Email: [mariagrazia.gasparo@unifi.it](mailto:mariagrazia.gasparo@unifi.it)

Massimiliano Giuli  
Dipartimento di Sistemi ed Istituzioni  
per l'Economia  
Università degli Studi dell'Aquila  
Via G. Falcone, 25  
I-67100 L'Aquila, Italy  
Email: [massimiliano.giuli@univaq.it](mailto:massimiliano.giuli@univaq.it)

Giampaolo Liuzzi  
Istituto di Analisi dei Sistemi  
ed Informatica "A. Ruberti"  
Consiglio Nazionale delle Ricerche  
Viale Manzoni, 30  
I-00185 Roma, Italy  
Email: [liuzzi@iasi.cnr.it](mailto:liuzzi@iasi.cnr.it)

Stefano Lucidi  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale "A. Ruberti"  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [lucidi@dis.uniroma1.it](mailto:lucidi@dis.uniroma1.it)

Maria Macconi  
Dipartimento di Energetica “S. Stecco”  
Università degli Studi di Firenze  
Viale G.B. Morgagni, 40  
I-50134 Firenze, Italy  
Email: [maria.macconi@unifi.it](mailto:maria.macconi@unifi.it)

Ricardo Ñanculef  
Departamento de Informática,  
Universidad Técnica Federico Santa María  
Avenida España 1680  
CP 110-V, Valparaiso, Chile  
Email: [jnancu@inf.utfsm.cl](mailto:jnancu@inf.utfsm.cl)

Alessandra Papini  
Dipartimento di Energetica “S. Stecco”  
Università degli Studi di Firenze  
Viale G.B. Morgagni, 40  
I-50134 Firenze, Italy  
Email: [alessandra.papini@unifi.it](mailto:alessandra.papini@unifi.it)

Mauro Passacantando  
Dipartimento di Matematica Applicata  
Università degli Studi di Pisa  
Via Buonarroti, 1/c  
I-56127 Pisa, Italy  
Email: [passacantando@dma.unipi.it](mailto:passacantando@dma.unipi.it)

Andrea Prearo  
Dipartimento di Matematica Pura  
ed Applicata  
Università di Modena e Reggio Emilia  
Via Campi, 213/B  
I-41125 Modena, Italy  
Email: [andrea.prearo@unimore.it](mailto:andrea.prearo@unimore.it)

Benedetta Morini  
Dipartimento di Energetica “S. Stecco”  
Università degli Studi di Firenze  
Viale G.B. Morgagni, 40  
I-50134 Firenze, Italy  
Email: [benedetta.morini@unifi.it](mailto:benedetta.morini@unifi.it)

Barbara Panicucci  
Dipartimento di Scienze e Metodi  
dell’Ingegneria  
Università di Modena e Reggio Emilia  
Via Amendola, 2, Pad. Morselli  
I-42122 Reggio Emilia, Italy  
Email: [barbara.panicucci@unimore.it](mailto:barbara.panicucci@unimore.it)

Massimo Pappalardo  
Dipartimento di Matematica Applicata  
Università degli Studi di Pisa  
Via Buonarroti, 1/c  
I-56127 Pisa, Italy  
Email: [pappalardo@dma.unipi.it](mailto:pappalardo@dma.unipi.it)

Veronica Piccialli  
Dipartimento di Ingegneria Civile e  
Ingegneria Informatica  
Università di Roma “Tor Vergata”  
Via del Politecnico, 1  
I-00133 Roma, Italy  
Email: [piccialli@dis.uniroma1.it](mailto:piccialli@dis.uniroma1.it)

Arnaldo Risi  
Clinical & Therapy Development  
Consultant  
Medtronic  
Via Lucrezio Caro, 63  
I-00193 Roma, Italy  
Email: [arnaldo.risi@medtronic.com](mailto:arnaldo.risi@medtronic.com)



Massimo Roma  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale “A. Ruberti”  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [roma@dis.uniroma1.it](mailto:roma@dis.uniroma1.it)

Simone Sagratella  
Dipartimento di Ingegneria Informatica  
Automatica e Gestionale “A. Ruberti”  
Università di Roma La Sapienza  
Via Ariosto, 25  
I-00185 Roma, Italy  
Email: [sagratella@dis.uniroma1.it](mailto:sagratella@dis.uniroma1.it)

Marco Sciandrone  
Dipartimento di Sistemi e Informatica  
Università degli Studi di Firenze  
Via di Santa Marta, 3  
I-50139 Firenze, Italy  
Email: [sciandro@dsi.unifi.it](mailto:sciandro@dsi.unifi.it)

Gaetano Zanghirati  
Dipartimento di Matematica  
Università degli Studi di Ferrara  
Polo Scientifico Tecnologico, Blocco B  
Via Saragat, 1  
I-44122 Ferrara, Italy  
Email: [gaetano.zanghirati@unife.it](mailto:gaetano.zanghirati@unife.it)

Valeria Ruggiero  
Dipartimento di Matematica  
Università degli Studi di Ferrara  
Polo Scientifico Tecnologico, Blocco B  
Via Saragat, 1  
I-44100 Ferrara, Italy  
Email: [rgv@unife.it](mailto:rgv@unife.it)

Fabio Schoen  
Dipartimento di Sistemi e Informatica  
Università degli Studi di Firenze  
Via di Santa Marta, 3  
I-50139 Firenze, Italy  
Email: [fabio.schoen@unifi.it](mailto:fabio.schoen@unifi.it)

Riccardo Zanella  
Laboratorio per le Tecnologie delle  
Terapie Avanzate  
Università degli Studi di Ferrara  
Via Fossato di Mortara, 70  
I-44121 Ferrara, Italy  
Email: [riccardo.zanella@unife.it](mailto:riccardo.zanella@unife.it)

Luca Zanni  
Dipartimento di Matematica Pura  
ed Applicata  
Università di Modena e Reggio Emilia  
Via Campi, 213/B  
I-41125 Modena, Italy  
Email: [luca.zanni@unimore.it](mailto:luca.zanni@unimore.it)

# Contents

## Preface

Two-sphere Separation Procedures via Non-smooth Optimization <i>A. Astorino, I.M. Bomze, M.P. Brito, and M. Gaudioso</i>	1
Diagonally Scaled Dogleg Methods for Nonlinear Systems with Simple Bounds <i>S. Bellavia, M. Macconi, and B. Morini</i>	17
Recent Advances in Equilibrium Problems <i>G. Bigi, M. Castellani, M. Giuli, B. Panucchi, M. Pappalardo, and M. Passacantando</i>	41
Analysis of Interior Point Methods for Edge-preserving Removal of Poisson Noise <i>S. Bonettini and V. Ruggiero</i>	67
Optimization Methods for Digital Image Restoration on MPP Multicore Architectures <i>R. Cavicchioli, A. Prearo, R. Zanella, G. Zanghirati, and L. Zanni</i>	93
An Exact Augmented Lagrangian Algorithm Exploiting Negative Curvature Directions <i>G. Di Pillo, G. Liuzzi, and S. Lucidi</i>	117
Recent Advancements in the Numerical Solution of Generalized Nash Equilibrium Problems <i>F. Facchinei, L. Lampariello, and S. Sagratella</i>	137
Quasi-Newton Updates for Preconditioned Nonlinear Conjugate Gradient Methods <i>G. Fasano and M. Roma</i>	175
Solution of Classification Problems via Computational Geometry Methods <i>E. Frandi, M.G. Gasparo, R. Nanculef, and A. Papini</i>	201
On Solving Optimal Control Problems for Distributed Parameter Systems Using an Inexact Newton Method <i>E. Galligani</i>	227
Partitioning Techniques for Global Optimization <i>G. Liuzzi, S. Lucidi, and V. Piccialli</i>	253

Recent Advances in Global Optimization Methods for Molecular Conformation Problems <i>M. Locatelli and F. Schoen</i>	273
A Convergent Hybrid Decomposition Algorithm Model Using Second Order Information for SVM Training <i>S. Lucidi, L. Palagi, A. Risi, and M. Sciandrone</i>	295

Mathematics subject classification

Two-sphere Separation Procedures via Non-smooth Optimization: 65K05, 62H30.

Diagonally Scaled Dogleg Methods for Nonlinear Systems with Simple Bounds: 65H10.

Recent Advances in Equilibrium Problems: 90C33.

Analysis of Interior Point Methods for Edge-preserving Removal of Poisson Noise: 65K05.

Optimization Methods for Digital Image Restoration on MPP Multicore Architectures: 65K05, 65Y05, 68U10.

An Exact Augmented Lagrangian Algorithm Exploiting Negative Curvature Directions: 90C30, 65K05.

Recent Advancements in the Numerical Solution of Generalized Nash Equilibrium Problems: 90C33, 90C30, 91-02.

Quasi-Newton Updates for Preconditioned Nonlinear Conjugate Gradient Methods: 90C30, 65K05.

Solution of Classification Problems via Computational Geometry Methods: 65K05, 90C20, 49Q10, 68T10.

On Solving Optimal Control Problems for Distributed Parameter Systems Using an Inexact Newton Method: 65K05, 65H10, 49M15, 49M25.

Partitioning Techniques for Global Optimization: 90C26, 65K05.

Recent Advances in Global Optimization Methods for Molecular Conformation Problems: 90C26, 90C06, 90C90, 65K05, 92E10.

A Convergent Hybrid Decomposition Algorithm Model Using Second Order Information for SVM Training: 49M27, 90C30.



## Preface

This volume of quaderni di matematica is devoted to the memory of Marco D'Apuzzo, who died suddenly at the age of 46 in June of 2009.

Marco studied Mathematics at the University of Naples Federico II, where he also received his PhD in Applied Mathematics and Computer Science. His early scientific work dealt mainly with numerical linear algebra and parallel computing, during his years at the Department of Mathematics and its Applications at the University of Naples Federico II.

In 1999 he moved to the Department of Mathematics of the Second University of Naples, first as assistant professor and then as associate professor of numerical analysis, where he found his professional fulfilment. His research turned toward the numerical optimization field, in which Marco brought his passion and care for the computational aspects of numerical analysis, that were a constant throughout his scientific activity. With his colleagues at the Second University of Naples, who shared with him a well-established expertise in numerical linear algebra and numerical software design, he focused on the effective use of numerical linear algebra in the nonlinear optimization field. This activity was motivated by the simple observation that the effectiveness of any code for large-scale optimization depends heavily on the availability of effective linear algebra algorithms and software, since the overall computational effort is dominated by the solution of linear systems or the updating of matrix information. More specifically, Marco was interested in the linear algebra problems arising in interior point methods, and most of his late work was devoted to this topic. One of his last papers, entitled "On mutual impact of numerical linear algebra and large-scale optimization with focus on interior point methods", co-authored by Valentina De Simone and Daniela di Serafino, was posthumously awarded as COAP 2010 Best Paper (Computational Optimization and Applications, vol. 45, pp. 283-310).

His research was carried out within several national projects, with significant roles of responsibility, in which he was appreciated for his scientific and human qualities by his colleagues from all over the country, who demonstrate

their respect and fondness for him with their commitment to this editorial project.

Marco also gave great service to his department, serving on various committees and especially as Math Course Coordinator, showing great kindness, enthusiasm and generosity toward students and colleagues, as well as ability to mediate among different positions.

In conclusion, during his too-short life, Marco proved to be not only a creative and productive scholar and teacher, but also a great person, for his cheerfulness, friendliness and generosity, and this book wants to be an homage from his friends in the Italian optimization community.

This volume includes thirteen invited papers from recognized experts in nonlinear optimization and equilibrium problems, covering a broad range of topics, from computational issues to more theoretical subjects, whose common denominator is the high quality of the contributions as well as the final goal of designing effective algorithms. In this sense, the spirit of this volume perfectly agrees with Marco's belief that the role of applied mathematics is to build effective tools able to provide solutions to problems in a wide range of applications.

The editors are indebted to the authors of the papers in this volume, who took much time from their busy schedules to put together a set of outstanding contributions. A special thanks goes to the eminent referees who gave a remarkable support in the review process, which was thorough and demanding, leading this editorial project to success.

*We really miss the friend, the colleague, and the man.*

Valentina De Simone, Daniela di Serafino, Gerardo Toraldo





# Quasi-Newton Updates for Preconditioned Nonlinear Conjugate Gradient Methods

*Giovanni Fasano and Massimo Roma*

## Contents

1. Introduction (177).
2. Preconditioned Nonlinear Conjugate Gradient algorithm (180).
3. A new Symmetric Rank-2 update (181).
4. A preconditioner using a BFGS-like low-rank quasi-Newton update (187).
5. Preliminary numerical experiences (192).
6. Conclusions and future work (193).



## 1. Introduction

We deal with the large scale unconstrained optimization problem

$$(1.1) \quad \min_{x \in \mathbb{R}^n} f(x),$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a twice continuously differentiable function and  $n$  is large. We assume that for a given  $x_0 \in \mathbb{R}^n$  the level set

$$\Omega_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$$

is compact. The huge number of real world applications which can be modelled as a large scale optimization problem strongly motivates the growing interest for the solution of such problems.

Among the iterative methods for large scale unconstrained optimization, when the Hessian matrix is possibly dense, limited memory quasi-Newton methods are often the methods of choice. As well known (see any textbook, e.g. [11]), they generate a sequence  $\{x_k\}$ , according to the following scheme

$$(1.2) \quad x_{k+1} = x_k + \alpha_k p_k, \quad k = 0, 1, \dots,$$

with

$$p_k = -H_k \nabla f(x_k),$$

where  $H_k$  is an approximation of the inverse of the Hessian matrix  $\nabla^2 f(x_k)$  and  $\alpha_k$  is a steplength. In particular, instead of computing  $H_k$  at each iteration  $k$ , these methods update  $H_k$  in a simple manner, in order to obtain the new approximation  $H_{k+1}$  to be used in the next iteration. Moreover, instead of storing full dense  $n \times n$  approximations, they only save a few vectors of length  $n$ , which allow to represent the approximations implicitly.

Among the quasi-Newton schemes, the L-BFGS method is usually considered one of the most efficient. It is well suited for large scale problems because the amount of storage is limited and controlled by the user. This method is based on the construction of the approximation of the inverse of the Hessian matrix, by exploiting curvature information gained only from the most recent iterations. The inverse of the Hessian matrix is updated at the  $k$ -th iteration by the formula

$$(1.3) \quad H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T$$

where

$$\rho_k = \frac{1}{y_k^T s_k}, \quad V_k = I - \rho_k y_k s_k^T,$$

and

$$(1.4) \quad s_k = x_{k+1} - x_k = \alpha_k p_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Observe that  $H_k$  also satisfies relation

$$\begin{aligned} H_k &= (V_{k-1}^T \cdots V_{k-m}^T) H_k^0 (V_{k-m} \cdots V_{k-1}) \\ &+ \rho_{k-m} (V_{k-1}^T \cdots V_{k-m+1}^T) s_{k-m} s_{k-m}^T (V_{k-m+1} \cdots V_{k-1}) \\ &+ \rho_{k-m+1} (V_{k-1}^T \cdots V_{k-m+2}^T) s_{k-m+1} s_{k-m+1}^T (V_{k-m+2} \cdots V_{k-1}) \\ &+ \cdots \\ &+ \rho_{k-1} s_{k-1} s_{k-1}^T, \end{aligned}$$

where  $m$  is the *memory* of the method and  $H_k^0$  is an initial approximation of the inverse of the Hessian matrix.

The well known reasons for the success of the L-BFGS method can be summarized in the following two points: firstly, even when  $m$  is small,  $H_{k+1}$  is an effective approximation of the inverse of the Hessian matrix, secondly  $H_{k+1}$  is the unique (positive definite) matrix which solves the problem

$$\begin{aligned} \min_H & \|H - H_k\|_F \\ \text{s.t.} & \quad H = H^T, \\ & \quad s_k = H y_k, \end{aligned}$$

where  $\|\cdot\|_F$  is the Frobenius norm. Namely,  $H_{k+1}$  is the positive definite matrix “closest” to the current approximation  $H_k$ , satisfying the *secant equation*  $s_k = H y_k$ . However, L-BFGS method presents some drawbacks, including the slow convergence on ill-conditioned problems, namely when the eigenvalues of

the Hessian matrix are very spread. Moreover, on some applications, the performances of L-BFGS method and the Nonlinear Conjugate Gradient method are comparable.

In this paper we focus on the latter method: the Nonlinear Conjugate Gradient method (NCG). As well known (see any textbook, e.g. [11]) it is a natural extension to general functions of the linear Conjugate Gradient (CG) method for quadratic functions. It generates a sequence  $\{x_k\}$  according to scheme (1.2), with

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1},$$

where  $\beta_k$  is a suitable scalar. Different values of  $\beta_k$  give rise to different algorithms (see [8] for a survey). The most common are the Fletcher and Reeves (FR), the Polak and Ribière (PR) and the Hestenes and Stiefel (HS) algorithms.

Although the NCG methods have been widely studied and are often very efficient when solving large scale problems, a key point for increasing their efficiency is the use of a preconditioning strategy, especially when solving difficult ill-conditioned problems. Defining good preconditioners for NCG methods is currently still considered a challenging research topic. On this guideline, this work is devoted to investigate the use of quasi-Newton updates as preconditioners. In particular, we want to propose preconditioners which possibly inherit the effectiveness of the L-BFGS update. Indeed, here we build preconditioners iteratively defined and based on quasi-Newton updates of the inverse of the Hessian matrix. This represents an attempt to improve the efficiency of the NCG method by conveying information collected from a quasi-Newton method, in a Preconditioned Nonlinear Conjugate Gradient method (PNCG). In particular, we study new symmetric low-rank updates of the inverse of the Hessian matrix, in order to iteratively define preconditioners for PNCG.

It is worth to note that there exists a close connection between BFGS and NCG [10], and on the other hand, NCG algorithms can be viewed as memoryless quasi-Newton methods (see e.g., [13], [12], [11]).

The idea of using a quasi-Newton update as a preconditioner within NCG algorithms is not new. In [2], when storage is available, a preconditioner defined by  $m$  quasi-Newton updates is used within NCG algorithm. In [1] a scaled memoryless BFGS matrix is used as preconditioner in the framework

of NCG. Moreover, an automatic preconditioning strategy based on a limited memory quasi-Newton update for the linear CG is proposed in [9], within Hessian free Newton methods, and is extended to the solution of a sequence of linear systems.

In this paper, we propose two classes of parameters dependent preconditioners. In particular, in the next section we briefly recall a scheme of a general PNCG method. In Section 3 a new Symmetric Rank-2 update is introduced and its theoretical properties are studied. Section 4 is devoted to describe a new BFGS-like quasi-Newton update. Finally, in Section 5 the results of a preliminary numerical experience are reported, showing a comparison between one of our proposals and an L-BFGS-based preconditioner for PNCG.

## 2. Preconditioned Nonlinear Conjugate Gradient algorithm

In this section we report the scheme of a general *Preconditioned Nonlinear Conjugate Gradient* (PNCG) *algorithm* (see e.g. [12]). In the PNCG scheme  $M_k$  denotes the preconditioner at the iteration  $k$ .

---

### Preconditioned Nonlinear Conjugate Gradient (PNCG) Algorithm

---

**Step 1:** Data  $x_1 \in \mathbb{R}^n$ . Set  $p_1 = -M_1 \nabla f(x_1)$  and  $k = 1$ .

**Step 2:** Compute the steplength  $\alpha_k$  by using a linesearch procedure which guarantees the Wolfe conditions to be satisfied, and set

$$x_{k+1} = x_k + \alpha_k p_k.$$

**Step 3:** If  $\|\nabla f(x_{k+1})\| = 0$  then stop, else compute  $\beta_{k+1}$  and

$$(2.1) \quad p_{k+1} = -M_{k+1} \nabla f(x_{k+1}) + \beta_{k+1} p_k,$$

set  $k = k + 1$  and go to *Step 2*.

---

By setting  $M_k = I$  for any  $k$ , the popular (*unpreconditioned*) Nonlinear Conjugate Gradient (NCG) method is obtained. The parameter  $\beta_{k+1}$  can be

chosen in a variety of ways. For PNCG algorithm the most recurrent choices are the following:

$$(2.2) \quad \beta_{k+1}^{\text{FR}} = \frac{\nabla f(x_{k+1})^T M_k \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)},$$

$$(2.3) \quad \beta_{k+1}^{\text{PR}} = \frac{[\nabla f(x_{k+1}) - \nabla f(x_k)]^T M_k \nabla f(x_{k+1})}{\nabla f(x_k)^T M_k \nabla f(x_k)},$$

$$(2.4) \quad \beta_{k+1}^{\text{HS}} = \frac{[\nabla f(x_{k+1}) - \nabla f(x_k)]^T M_k \nabla f(x_{k+1})}{[\nabla f(x_{k+1}) - \nabla f(x_k)]^T p_k}.$$

We recall that with respect to other gradient methods, a more accurate line-search procedure is required to determine the steplength  $\alpha_k$  in a PNCG algorithm. This is due to the presence of the term  $\beta_{k+1} p_k$  in (2.1). The latter fact motivates the use of the (strong) Wolfe conditions to compute the steplength  $\alpha_k$ , which also guarantee that  $s_k^T y_k > 0$  for any  $k$ .

As already said, preconditioning is applied for increasing the efficiency of the NCG method. In this regard, we remark a noticeable difference between linear CG and NCG. Whenever the linear CG is applied, the Hessian matrix does not change during the iterations of the algorithm. On the contrary, when NCG is applied to a nonlinear function, the Hessian matrix (possibly indefinite) changes at each iteration.

### 3. A new Symmetric Rank-2 update

In this section we study a new quasi-Newton updating formula, by considering the properties of a parameter dependent Symmetric Rank-2 (SR2) update of the inverse of the Hessian matrix. Suppose we generate after  $k$  iterations the sequence of iterates  $\{x_1, \dots, x_{k+1}\}$ . Then our quasi-Newton update  $H_{k+1}$ , which approximates  $[\nabla^2 f(x)]^{-1}$ , satisfies the secant equation along all previous directions; namely it results

$$H_{k+1} y_j = s_j, \quad \text{for all } j \leq k.$$

Observe that the latter appealing property is satisfied by all the updates of the Broyden class, provided that the linesearch adopted is exact (see e.g. [11]). We

would like to recover the motivation underlying the latter class of updates, and by using rank-2 updates we would like to define a preconditioner for PNCG.

On this guideline, in order to build an approximate inverse of the Hessian matrix, we consider the update

$$(3.1) \quad H(\gamma_{k+1}, \omega_{k+1}) = H(\gamma_k, \omega_k) + \Delta_k, \quad \Delta_k \in \mathbb{R}^{n \times n}, \text{ symmetric,}$$

where the sequence  $\{H(\gamma_k, \omega_k)\}$  depends on the parameters  $\gamma_k, \omega_k$  and provides our quasi-Newton updates of  $[\nabla^2 f(x)]^{-1}$ .

It is first our purpose to propose the new update  $H(\gamma_{k+1}, \omega_{k+1})$  such that:

- (0)  $H(\gamma_{k+1}, \omega_{k+1})$  is well-defined and nonsingular,
- (1)  $H(\gamma_{k+1}, \omega_{k+1})$  can be iteratively updated,
- (2)  $H(\gamma_{k+1}, \omega_{k+1})$  collects the information from the iterations  $1, 2, \dots, k$  of a NCG method,
- (3)  $H(\gamma_{k+1}, \omega_{k+1})$  satisfies the secant equation at iterations  $j = 1, 2, \dots, k$ ,
- (4)  $H(\gamma_{k+1}, \omega_{k+1})$  either “tends to preserve” the inertia of the inverse of  $\nabla^2 f(x_{k+1})$ , in case  $f(x)$  is a general quadratic function or, by suitably setting the two parameters, it can be used as a preconditioner for PNCG, i.e.  $M_k = H(\gamma_k, \omega_k)$ .

Observe that the Symmetric Rank-1 (SR1) quasi-Newton update (see Section 6.2 in [11]) satisfies properties (1)-(4) but not the property (0), i.e. it might be possibly not well-defined for a general nonlinear function. The latter result follows from the fact that SR1 update provides only a rank-1 quasi-Newton update, unlike BFGS and DFP. On the other hand, while BFGS and DFP quasi-Newton formulae provide only positive definite updates, the SR1 formula is able to recover the inertia of the Hessian matrix, by generating possibly indefinite updates. Thus, now we want to study an SR2 quasi-Newton update, which satisfies (0)-(4) and where one of the two newest dyads of the update is provided by information from the NCG method. To this aim, assuming that  $H_k = H(\gamma_k, \omega_k)$  is given, we consider the relation (3.1) where we set



(see (1.4))

$$\Delta_k = \gamma_k v_k v_k^T + \omega_k \frac{p_k p_k^T}{y_k^T p_k}, \quad \gamma_k, \omega_k \in \mathbb{R}, \quad v_k \in \mathbb{R}^n,$$

and  $p_k$  is generated at the  $k$ -th iteration of the (unpreconditioned) NCG method. Thus, we will have the new update

$$(3.2) \quad H_{k+1} = H_k + \gamma_k v_k v_k^T + \omega_k \frac{p_k p_k^T}{y_k^T p_k}, \quad \gamma_k, \omega_k \in \mathbb{R}, \quad v_k \in \mathbb{R}^n,$$

and in order to satisfy the secant equation  $H_{k+1} y_k = s_k$  the following equality must hold

$$H_k y_k + \gamma_k (v_k^T y_k) v_k + \omega_k \frac{p_k p_k^T}{y_k^T p_k} y_k = s_k,$$

that is

$$(3.3) \quad \gamma_k (v_k^T y_k) v_k = s_k - H_k y_k - \omega_k p_k.$$

Therefore it results

$$(3.4) \quad v_k = \sigma_k (s_k - H_k y_k - \omega_k p_k)$$

for some scalar  $\sigma_k \in \mathbb{R}$ . By substituting the expression (3.4) of  $v_k$  in (3.3) we have

$$\gamma_k \sigma_k^2 [y_k^T (s_k - H_k y_k - \omega_k p_k)] (s_k - H_k y_k - \omega_k p_k) = s_k - H_k y_k - \omega_k p_k.$$

Thus, the following relation among the parameters  $\gamma_k$ ,  $\sigma_k$  and  $\omega_k$  must hold

$$(3.5) \quad \gamma_k \sigma_k^2 = \frac{1}{s_k^T y_k - y_k^T H_k y_k - \omega_k p_k^T y_k}.$$

Note that from the arbitrariness of  $\gamma_k$ , without loss of generality, we can set  $\sigma_k \in \{-1, 1\}$ .

Now, in the next proposition we first consider the case of quadratic functions, and prove that the update (3.2) satisfies the secant equation, along all previous directions.

**Proposition 3.1.** Assume that  $f$  is the quadratic function  $f(x) = \frac{1}{2}x^T Ax + b^T x$ , where  $A \in \mathbb{R}^{n \times n}$  is symmetric and  $b \in \mathbb{R}^n$ . Suppose that  $k$  steps of the (unpreconditioned) CG are performed, in order to detect the stationary point (if any) of the function  $f$ , and that the vectors  $p_1, \dots, p_k$  are generated. Then, the matrix  $H_{k+1}$  in (3.2) satisfies the secant equations

$$(3.6) \quad H_{k+1}y_j = s_j, \quad j = 1, \dots, k,$$

provided that the coefficients  $\gamma_j, \omega_j, j = 1, \dots, k$  are computed such that

$$(3.7) \quad \begin{aligned} \gamma_j &= \frac{1}{s_j^T y_j - y_j^T H_j y_j - \omega_j p_j^T y_j}, & j = 1, \dots, k, \\ \omega_j &\neq \frac{s_j^T y_j - y_j^T H_j y_j}{p_j^T y_j}, & j = 1, \dots, k. \end{aligned}$$

PROOF – The proof proceeds by induction. Equations (3.6) hold for  $k = 1$ , that is  $H_2 y_1 = s_1$ , as long as

$$s_1 = \left[ H_1 + \gamma_1 \sigma_1^2 (s_1 - H_1 y_1 - \omega_1 p_1)(s_1 - H_1 y_1 - \omega_1 p_1)^T + \omega_1 \frac{p_1 p_1^T}{y_1^T p_1} \right] y_1,$$

or equivalently

$$s_1 - H_1 y_1 - \omega_1 p_1 = \gamma_1 (s_1^T y_1 - y_1^T H_1 y_1 - \omega_1 p_1^T y_1) [s_1 - H_1 y_1 - \omega_1 p_1],$$

which is satisfied selecting  $\gamma_1$  and  $\omega_1$  according with (3.7).

Now, suppose that the relations (3.6) hold for the index  $k - 1$ . To complete the induction we need to prove that the relations (3.6) hold for the index  $k$ . Firstly, note that  $H_{k+1} y_k = s_k$  holds. In fact

$$s_k = \left[ H_k + \gamma_k \sigma_k^2 (s_k - H_k y_k - \omega_k p_k)(s_k - H_k y_k - \omega_k p_k)^T + \omega_k \frac{p_k p_k^T}{y_k^T p_k} \right] y_k$$

holds if and only if

$$s_k - H_k y_k - \omega_k p_k = \gamma_k (s_k^T y_k - y_k^T H_k y_k - \omega_k p_k^T y_k)(s_k - H_k y_k - \omega_k p_k),$$

and the latter holds from (3.7) with  $j = k$ . Now, we have to prove that (3.6) hold for any  $j < k$ . For  $j < k$  we have

$$H_{k+1} y_j = H_k y_j + \gamma_k \sigma_k^2 (s_k - H_k y_k - \omega_k p_k)(s_k - H_k y_k - \omega_k p_k)^T y_j + \omega_k \frac{p_k^T y_j}{y_k^T p_k} p_k,$$

where  $H_k y_j = s_j$  by the inductive hypothesis. Moreover,

$$(s_k - H_k y_k)^T y_j = s_k^T y_j - y_k^T H_k y_j = s_k^T y_j - y_k^T s_j = s_k^T A s_j - (A s_k)^T s_j = 0,$$

where the third equality holds since  $y_j = A s_j$ , for any  $j$ , for the quadratic function  $f$ . Finally,

$$\omega_k p_k^T y_j = \omega_k p_k^T A s_j = \omega_k \alpha_j p_k^T A p_j = 0,$$

which follows from the conjugacy of the directions  $\{p_1, \dots, p_k\}$  generated by the CG. Thus, (3.6) hold for any  $j \leq k$  and the induction is complete.  $\square$

As an immediate consequence of the previous proposition, we prove now the finite termination property for a quadratic function, i.e. after at most  $n$  steps  $H_{n+1}$  is the inverse of the Hessian of the quadratic function.

**Corollary 3.1.** *Assume that  $f$  is the quadratic function  $f(x) = \frac{1}{2} x^T A x + b^T x$ , where  $A \in \mathbb{R}^{n \times n}$  is symmetric and  $b \in \mathbb{R}^n$ . Suppose that  $n$  steps of the (unpreconditioned) CG are performed, in order to detect the stationary point of the function  $f$ , and that the vectors  $p_1, \dots, p_n$  are generated. If (3.7) holds, we have  $H_{n+1} = A^{-1}$ .*

PROOF – By applying Proposition 3.1, we have that (3.6) hold for  $k = n$ , i.e.

$$H_{n+1} y_j = s_j, \quad j = 1, \dots, n.$$

Since  $f$  is quadratic then  $y_j = A s_j$ , for any  $j$ , i.e.

$$H_{n+1} A s_j = s_j, \quad j = 1, \dots, n.$$

Now, since  $s_j = \alpha_j p_j$ ,  $j = 1, \dots, n$ , the conjugacy of the vectors  $\{p_1, \dots, p_n\}$  implies that  $H_{n+1} = A^{-1}$ .  $\square$

We highlight that, whenever  $k = n$ , Corollary 3.1 justifies the first part of the statement (4) on page 182. Moreover, later on in the paper we show that for  $k < n$ , the update matrix in (3.2) can be suitably modified to provide a preconditioner.

After analyzing the case of  $f(x)$  quadratic, we turn now to the general case of a nonlinear twice continuously differentiable function. In particular, since we are interested in using the matrix  $H_{k+1}$  in (3.2) as a preconditioner, we need to investigate if there exists a suitable setting of the parameters such that  $H_{k+1}$  is positive definite, provided that (3.7) are satisfied. In the next proposition we prove that if the parameter  $\omega_k$  is below a threshold value, then the matrix  $H_{k+1}$  is almost always positive definite.

**Proposition 3.2.** *Let  $f$  be a nonlinear twice continuously differentiable function. Suppose that the (unpreconditioned) NCG method is used to minimize the function  $f$ . Suppose that (3.7) is satisfied and*

$$(3.8) \quad 0 \leq \omega_k < \frac{s_k^T y_k - y_k^T H_k y_k}{p_k^T y_k},$$

with

$$(3.9) \quad y_k^T s_k + y_k^T H_k y_k \leq 0 \quad \text{or} \quad y_k^T s_k - y_k^T H_k y_k \geq 0,$$

where  $s_j = \alpha_j p_j$ . Then the matrix  $H_{k+1}$  in (3.2) is positive definite.

PROOF – By substituting (3.4) in (3.2), recalling that  $\sigma_k^2 = 1$  we obtain

$$H_{k+1} = \gamma_k \left[ (\alpha_k - \omega_k)^2 p_k p_k^T + (\alpha_k - \omega_k) \left( (H_k y_k) p_k^T + p_k (H_k y_k)^T \right) + (H_k y_k) (H_k y_k)^T \right] + \omega_k \frac{p_k p_k^T}{y_k^T p_k}.$$

Hence  $H_{k+1}$  can be rewritten in the form

$$\begin{pmatrix} p_k & \vdots & H_k y_k \end{pmatrix} \begin{pmatrix} \gamma_k (\alpha_k - \omega_k)^2 + \frac{\omega_k}{y_k^T p_k} & & \gamma_k (\alpha_k - \omega_k) \\ & & \\ & & \gamma_k (\alpha_k - \omega_k) & & \gamma_k \end{pmatrix} \begin{pmatrix} p_k^T \\ \dots \\ (H_k y_k)^T \end{pmatrix}.$$

Therefore  $H_{k+1}$  is positive definite if and only if the following inequalities hold:

$$(3.10) \quad \begin{aligned} \gamma_k (\alpha_k - \omega_k)^2 + \frac{\omega_k}{y_k^T p_k} &> 0, \\ \gamma_k \left( \gamma_k (\alpha_k - \omega_k)^2 + \frac{\omega_k}{y_k^T p_k} \right) - \gamma^2 (\alpha_k - \omega_k)^2 &> 0. \end{aligned}$$

Using the expression of  $\gamma_k$  in (3.5) and recalling that  $y_k^T s_k > 0$  (as a consequence of the Wolfe conditions), (3.10) are equivalent to

$$\begin{aligned} \frac{(\alpha_k - \omega_k)^2 y_k^T p_k}{(\alpha_k - \omega_k) p_k^T y_k - y_k^T H_k y_k} + \omega_k &> 0, \\ \frac{\omega_k}{(\alpha_k - \omega_k) p_k^T y_k - y_k^T H_k y_k} &> 0. \end{aligned}$$

After some computation we obtain that there exist values of the parameter  $\omega_k$  for which the latter inequalities admit solutions, with only one exception. In fact, they are satisfied for any value of  $\omega_k$  such that

$$0 \leq \omega_k < \frac{\alpha_k p_k^T y_k - y_k^T H_k y_k}{p_k^T y_k},$$

but they do not admit solution in case

$$\alpha_k y_k^T p_k + y_k^T H_k y_k > 0 \quad \text{and} \quad \alpha_k y_k^T p_k - y_k^T H_k y_k < 0,$$

i.e. when (3.9) does not hold. □

From Proposition 3.1 and Corollary 3.1, we could use the matrix  $H_{k+1}$  as an approximate inverse of  $\nabla^2 f(x)$ . However, Proposition 3.2 evidences that conditions (3.7) and (3.8) do not suffice to ensure  $H_{k+1}$  positive definite. In fact, whenever (3.9) occurs, additional safeguard is needed since  $H_{k+1}$  is possibly indefinite. Thus, the definition of  $H_{k+1}$  should be possibly modified in order to obtain positive definite updates.

#### 4. A preconditioner using a BFGS-like low-rank quasi-Newton update

In this section we partially address the final remark of Section 3. Indeed, we introduce a new class of preconditioners which are still iteratively constructed by using information from the NCG iterations and, as in the case of BFGS updates, they are always positive definite. On this purpose, the price we pay with respect to (3.2), is that the secant equation is satisfied only at the current iterate, and not necessarily along all the previous iterates.

We draw our inspiration from [4], where a new preconditioner for Newton–Krylov methods is described. In particular, in [4] the set of directions generated by the Krylov subspace method is used to provide an approximate inverse preconditioner, for the solution of Newton’s systems. On this guideline, observe that if  $f(x) = \frac{1}{2}x^T Ax + b^T x$ , where  $A$  is positive definite and  $b \in \mathbb{R}^n$ , then it is well known (see e.g. [6]) that the CG method may generate  $n$  conjugate directions  $\{p_j\}$  such that

$$(4.1) \quad A^{-1} = \sum_{j=1}^n \frac{p_j p_j^T}{p_j^T A p_j}.$$

Now, in order to introduce a class of preconditioners for the NCG, in case of a general twice continuously differentiable function  $f$ , suppose we have performed  $k$  iterations of the (unpreconditioned) NCG, so that the directions  $p_1, \dots, p_k$  are generated. Let us consider the matrix  $M_{k+1}$  defined by

$$(4.2) \quad M_{k+1} = \tau_k C_k + \gamma_k v_k v_k^T + \omega_k \sum_{j=k-m}^k \frac{p_j p_j^T}{p_j^T \nabla^2 f(x_j) p_j},$$

where  $0 \leq m \leq k$ ,  $\gamma_k, \omega_k \geq 0$ ,  $\tau_k > 0$ ,  $C_k \in \mathbb{R}^{n \times n}$  is symmetric positive definite and  $v_k \in \mathbb{R}^n$ . In order to use  $M_{k+1}$  as a preconditioner and to update its expression iteratively, we set  $\tau_k = 1$ ,  $C_k = H(\tau_k, \gamma_k, \omega_k)$  (with  $H(\tau_0, \gamma_0, \omega_0)$  given) and rewrite (4.2) in the form

$$(4.3) \quad H(\tau_{k+1}, \gamma_{k+1}, \omega_{k+1}) = H(\tau_k, \gamma_k, \omega_k) + \gamma_k v_k v_k^T + \omega_k \sum_{j=k-m}^k \frac{p_j p_j^T}{p_j^T \nabla^2 f(x_j) p_j}.$$

$H(\tau_{k+1}, \gamma_{k+1}, \omega_{k+1})$  may be treated as a symmetric quasi–Newton update. However, for simplicity, in the sequel we prefer to use the more general form given by (4.2).

Observe that in the expression of  $M_{k+1}$ ,  $v_k v_k^T$  represents a rank-1 update and from (4.1) the dyads  $p_j p_j^T / p_j^T \nabla^2 f(x_j) p_j$  are aimed to build an approximate inverse. The integer  $m$  can be viewed as a “limited memory” parameter, similarly to the L–BFGS method. Moreover, we can set the vector  $v_k$  and the parameters  $\tau_k, \gamma_k, \omega_k$  such that the class of preconditioners  $M_k$  satisfies, for

any  $k$ , the secant equation

$$(4.4) \quad M_{k+1}y_k = s_k.$$

Indeed, from (4.4) we have

$$\tau_k C_k y_k + \gamma_k (v_k^T y_k) v_k + \omega_k \sum_{j=k-m}^k \frac{p_j^T y_k}{p_j^T \nabla^2 f(x_j) p_j} p_j = s_k;$$

hence, assuming  $\gamma_k (v_k^T y_k) \neq 0$ ,

$$(4.5) \quad v_k = \sigma_k \left[ s_k - \tau_k C_k y_k - \omega_k \sum_{j=k-m}^k \frac{p_j^T y_k}{p_j^T \nabla^2 f(x_j) p_j} p_j \right],$$

for some  $\sigma_k \in \mathbb{R}$ . Using (4.5) in (4.4) we have

$$\begin{aligned} \gamma_k \sigma_k^2 & \left[ s_k^T y_k - \tau_k y_k^T C_k y_k - \omega_k \sum_{j=k-m}^k \frac{(p_j^T y_k)^2}{p_j^T \nabla^2 f(x_j) p_j} \right] \\ & \cdot \left[ s_k - \tau_k C_k y_k - \omega_k \sum_{j=k-m}^k \frac{p_j^T y_k}{p_j^T \nabla^2 f(x_j) p_j} p_j \right] = \\ & s_k - \tau_k C_k y_k - \omega_k \sum_{j=k-m}^k \frac{p_j^T y_k}{p_j^T \nabla^2 f(x_j) p_j} p_j. \end{aligned}$$

Thus, the following relation among the parameters  $\gamma_k, \sigma_k, \tau_k$  and  $\omega_k$  has to be satisfied

$$(4.6) \quad \gamma_k \sigma_k^2 = \frac{1}{-\tau_k y_k^T C_k y_k - \omega_k \sum_{j=k-m}^k \frac{(p_j^T y_k)^2}{p_j^T \nabla^2 f(x_j) p_j} + s_k^T y_k}$$

and without loss of generality we can set  $\sigma_k \in \{+1, -1\}$ . Then, observe that unlike the update proposed in the previous section (namely (3.2)), the matrix  $M_{k+1}$  in (4.4) satisfies the secant equation only at the  $k$ -th iteration (even for quadratic functions), and possibly not along all the previous iterations, as proved in Proposition 3.1 for the update (3.2). As regards the positive

definiteness of  $M_{k+1}$ , the Wolfe conditions used in the linesearch procedure for computing the steplength  $\alpha_k$  ensure that  $s_k^T y_k > 0$ , so that for  $\tau_k > 0$  and  $\omega_k \geq 0$  sufficiently small in (4.6) the matrix  $M_{k+1}$  is positive definite. Indeed, suppose that  $\omega_k \rightarrow 0$ , then  $M_{k+1} \approx \tau_k C_k + \gamma_k v_k v_k^T$ . Now, since  $\tau_k > 0$  and  $C_k$  is positive definite, by (4.6) for  $\tau_k$  sufficiently small we have  $\gamma_k > 0$ , i.e. we definitely have that  $M_{k+1}$  is positive definite.

Finally, observe that the different choices for the parameters  $\tau_k$  and  $\omega_k$  in (4.6) provide a different scaling of the matrices  $C_k$  and  $\sum_{j=k-m}^k \frac{p_j p_j^T}{p_j^T \nabla^2 f(x_j) p_j}$ , in the preconditioners.

Now we note that the quantities  $p_j^T \nabla^2 f(x_j) p_j$ ,  $j = 1, \dots, k$ , in the expression (4.2) of  $M_{k+1}$  are in general unavailable. By considering that the Hessian matrix is not constant at the points in the closed segment  $[x_j, x_{j+1}]$ , then we can use the Mean Value Theorem to estimate the average curvature of  $f$  along the direction  $p_j$ , that is

$$\int_0^1 s_j^T \nabla^2 f[x_j + \beta(x_{j+1} - x_j)] s_j d\beta = s_j^T y_j,$$

and recalling that  $s_j = \alpha_j p_j$ , we can estimate the quantity  $p_j^T \nabla^2 f(x_j) p_j$ , in the expression of  $M_{k+1}$ , by

$$p_j^T \nabla^2 f(x_j) p_j \approx \int_0^1 p_j^T \nabla^2 f[x_j + \beta(x_{j+1} - x_j)] p_j d\beta = \frac{s_j^T y_j}{\alpha_j^2} = \frac{p_j^T y_j}{\alpha_j}.$$

Observe that by the Wolfe conditions used in the linesearch procedure, the latter quantity satisfies the condition

$$\frac{p_j^T y_j}{\alpha_j} > 0.$$

Moreover, in case  $f$  is the quadratic function  $f(x) = \frac{1}{2} x^T A x + b^T x$  then

$$(4.7) \quad \int_0^1 p_j^T \nabla^2 f[x_j + \beta(x_{j+1} - x_j)] p_j d\beta = p_j^T A p_j,$$

i.e. the left hand side of (4.7) may be regarded as a generalization (to the general nonlinear case) of the quantity  $p_j^T \nabla^2 f(x_j) p_j$ .



As regards the matrix  $C_k$  in (4.2), an obvious choice could be for any  $k$

$$C_k = \varepsilon_k I, \quad \varepsilon_k \in \mathbb{R}.$$

Furthermore,  $\varepsilon_k$  may be computed as the least squares solution of the equation  $(\varepsilon I)y_k - s_k = 0$ , i.e.  $\varepsilon_k$  solves

$$\min_{\varepsilon} \|(\varepsilon I)y_k - s_k\|^2.$$

Hence,

$$\varepsilon_k = \frac{s_k^T y_k}{\|y_k\|^2},$$

so that, since  $s_k^T y_k > 0$  by the Wolfe conditions, the matrix

$$C_k = \frac{s_k^T y_k}{\|y_k\|^2} I$$

is positive definite.

For the sake of clarity we report here the resulting expression of our class of preconditioners (4.2):

$$(4.8) \quad M_{k+1} = \tau_k \frac{s_k^T y_k}{\|y_k\|^2} I + \gamma_k v_k v_k^T + \omega_k \sum_{j=k-m}^k \frac{s_j s_j^T}{y_j^T s_j},$$

where

$$v_k = \sigma_k \left[ s_k - \tau_k \frac{s_k^T y_k}{\|y_k\|^2} y_k - \omega_k \sum_{j=k-m}^k \frac{s_j^T y_k}{y_j^T s_j} s_j \right], \quad \sigma_k \in \{-1, 1\},$$

and

$$\gamma_k \sigma_k^2 = \frac{1}{(1 - \tau_k) s_k^T y_k - \omega_k \sum_{j=k-m}^k \frac{(s_j^T y_k)^2}{y_j^T s_j}}.$$

We conclude this section by highlighting that, interestingly enough, similarly to (4.3) we can construct a class of preconditioners based on DFP-like quasi-Newton updates. Indeed, we can iteratively build matrices

$$B(\tau_{k+1}, \gamma_{k+1}, \omega_{k+1})$$

approximating  $\nabla^2 f(x)$  and not its inverse. Then, by the Sherman-Morrison-Woodbury formula applied to  $B(\tau_{k+1}, \gamma_{k+1}, \omega_{k+1})$  we can compute a class of preconditioners.

## 5. Preliminary numerical experiences

In order to investigate the reliability of the classes of preconditioners we have introduced, we preliminarily performed a numerical testing for the use of the preconditioners defined in (4.8). This choice is motivated by the fact that for this class of preconditioners we can easily guarantee the positive definitiveness, whereas in case of the class of preconditioners given by (3.2) an alternative strategy must be proposed to guarantee the positive definiteness.

Therefore, we embedded the preconditioners (4.8) within the standard CG+ code [5]. We used the same linesearch and the same stopping criterion used by default in CG+ code. Thus we refer to [5] for a complete description of all the details. We tested both the Fletcher and Reeves (FR) and the Polak and Ribiere (PR) versions of the PNCG method at page 180.

As regards the test problems, we selected all the large scale unconstrained test problems in the CUTER collection [7]. The dimension of the test problems is between  $n = 1000$  and  $n = 10000$  (we considered 110 resulting problems). The parameters of the preconditioner (4.8) have been chosen as follows:  $m = 4$ ,  $\sigma_k = 1$  and

$$\tau_k = \omega_k = \frac{\frac{1}{2} s_k^T y_k}{y_k^T C_k y_k + \sum_{j=k-m}^k \frac{(p_j^T y_k)^2}{p_j^T \nabla^2 f(x_j) p_j}}$$

for all  $k$  (this choice ensures that the denominator of (4.6) is equal to  $\frac{1}{2} s_k^T y_k > 0$ ). As preliminary investigation, we considered the results in terms of the number of iterations and the number of function evaluations. We compared the results obtained by (4.8), the unpreconditioned case, and the case where  $M_k$  coincides with the L-BFGS update  $H_{k+1}$  in (1.3). This comparison is reported by using performance profiles [3]. For a fair comparison, we have excluded in each profile all the test problems where the three algorithms converge to different stationary points.

In particular, as regards the FR version, in Figure 1 we report the comparison among the three algorithms in terms of number of iterations. Figure 2 reports the same plot with a different scale. In Figures 3 and 4 the comparison among the three algorithms is reported in terms of number of function evaluations. These profiles show that using the FR algorithm, the preconditioner (4.8) tends to be preferable, both in terms of number of iterations and number of function evaluations.

Now we turn to the PR version of the PNCG algorithm and, in Figure 5 we report the comparison among (4.8), the unpreconditioned algorithm and the L-BFGS based preconditioner in terms of number of iterations. Figure 6 reports the same plot with a different scale. In Figures 7 and 8 the comparison among the three algorithms is reported in terms of number of function evaluations. From the observation of these plots it is easy to ascertain that the situation is reversed with respect to the FR version of the algorithms.

On the overall, even if these preliminary results do not allow us to draw final conclusions, they show that the preconditioning strategies proposed may be reliable and in some cases they are beneficial. In particular, we observe that our proposals are cheaper than the L-BFGS based preconditioner. However, observing the case of PR setting, since in (4.2) we convey only informations from the current iterate, we guess that a more sophisticated choice of the matrix  $\tau_k C_k$  is definitely needed, in order to preserve efficiency.

## 6. Conclusions and future work

In this paper we propose two new classes of quasi-Newton update, aiming at using the update matrix as a preconditioner within NCG method. In the first proposal the satisfaction of the secant equations at each previous iteration is ensured (in the quadratic case), but we can not ensure, in general, that the resulting update is positive definite. In the latter cases, an alternative strategy is needed.

In the second proposal the satisfaction of the secant equation only at the current iteration is ensured but the resulting update is guaranteed to be positive definite. We numerically tested the latter approach both with the un-

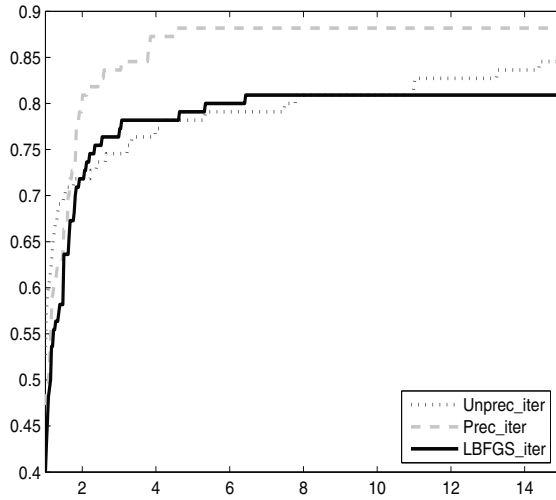


Figure 1 – Comparison of the FR algorithms in terms of number of iterations.

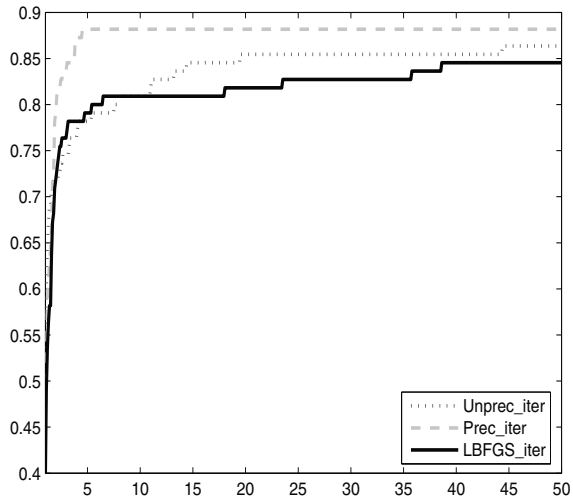


Figure 2 – Comparison of the FR algorithms in terms of number of iterations (expanded).

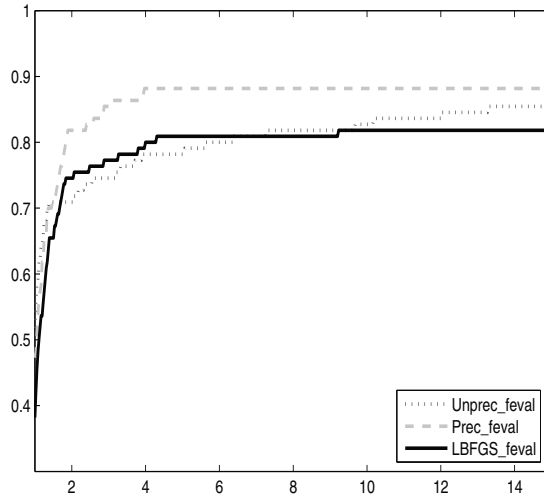


Figure 3 – Comparison of the FR algorithms in terms of number of function evaluations.

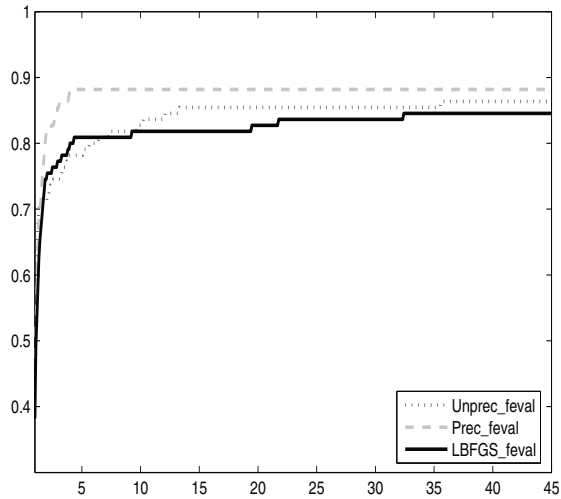


Figure 4 – Comparison of the FR algorithms in terms of number of function evaluations (expanded).

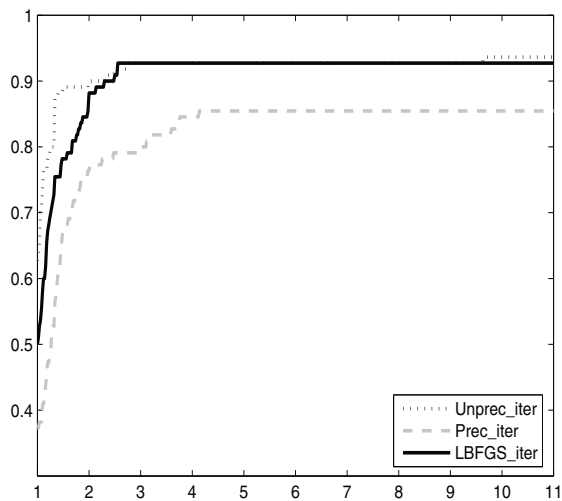


Figure 5 – Comparison of the PR algorithms in terms of number of iterations.

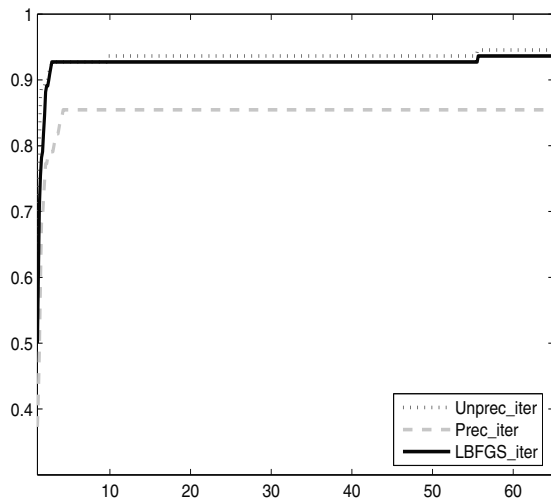


Figure 6 – Comparison of the PR algorithms in terms of number of iterations (expanded).

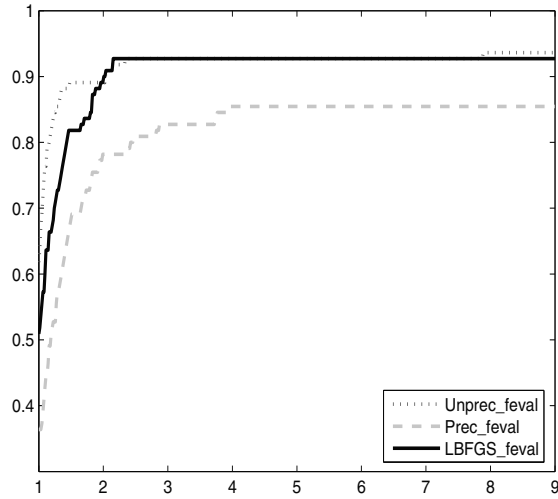


Figure 7 – Comparison of the PR algorithms in terms of number of function evaluations.

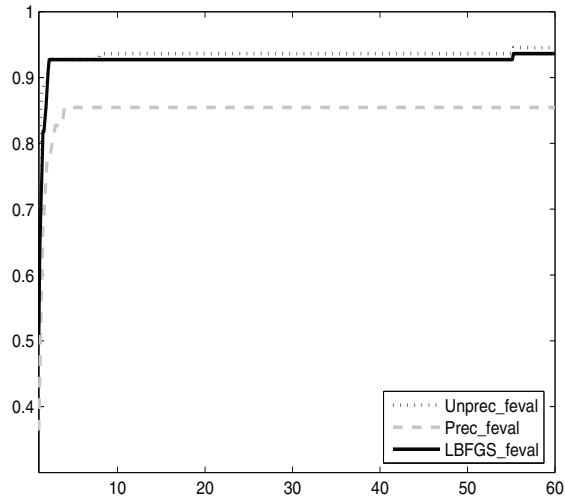


Figure 8 – Comparison of the PR algorithms in terms of number of function evaluations (expanded).

constrained case and L-BFGS based preconditioning approach. The results obtained, though preliminary, showed that it may be promising in some cases, even if non-carefully selected settings of the parameters are chosen.

## Acknowledgements

The authors wish to thank Marco D'Apuzzo, who inspired these afternotes, for his cheerful attitude to life, which greatly highlighted and completed his professionalism.

## References

- [1] N. ANDREI: Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Optimization Methods and Software*, **22**, 2007, pp. 561–571.
- [2] B. BUCKLEY AND A. LENIR: QN-like variable storage conjugate gradients, *Mathematical Programming*, **27**, 1983, pp. 155–175.
- [3] E. D. DOLAN AND J. MORÉ: Benchmarking optimization software with performance profiles, *Mathematical Programming*, **91**, 2002, pp. 201–213.
- [4] G. FASANO AND M. ROMA: Preconditioning Newton–Krylov methods in nonconvex large scale optimization, submitted to *Computational Optimization and Applications*.
- [5] J.C. GILBERT AND J. NOCEDAL: Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal on Optimization*, **2**, 1992, pp. 21–42.
- [6] G.H. GOLUB AND C.F. VAN LOAN: *Matrix Computations*, The John Hopkins Press, Baltimore, third edition, 1996.
- [7] N. I. M. GOULD, D. ORBAN, AND P.L. TOINT: CUTer (and sifdec), a constrained and unconstrained testing environment, revised, *ACM Transaction on Mathematical Software*, **29**, 2003, pp. 373–394.



- [8] W. HAGER AND H. ZHANG: A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, **2**, 2006, pp. 35–58.
- [9] J.L. MORALES AND J. NOCEDAL: Automatic preconditioning by limited memory quasi-Newton updating, *SIAM Journal on Optimization*, **10**, 2000, pp. 1079–1096.
- [10] L. NAZARETH: A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms, *SIAM Journal on Numerical Analysis*, **16**, 1979, pp. 794–800.
- [11] J. NOCEDAL AND S.J. WRIGHT: *Numerical Optimization*, Springer, second edition, 2006.
- [12] R. PYTLAK: *Conjugate Gradient Algorithms in Nonconvex Optimization*, Springer, Berlin, 2009.
- [13] D.F. SHANNO: Conjugate gradient methods with inexact searches, *Mathematics of Operations Research*, **3**, 1978, pp. 244–256.



---

## quaderni di matematica

editor-in-chief

Francesco Mazzocca

editorial board

Paola Bondi - Bruno Carbonaro - Emma D'Aniello  
Paola D'Aquino - Giuseppe Di Maio - Daniela di Serafino  
Adele Ferone - Eva Ferrara Dentice  
Paolo Maremonti - Nicola Melone  
Olga Polverino - Giulio Starita

---

The Series *Quaderni di Matematica* is planned as a sequence of thematic volumes, each presenting papers that deal with notable problems in the same branch of mathematical research. The papers collected in each issue should offer new and meaningful results in their field as well as an up-to-date picture of recent researches about the problems they discuss, together — if possible — with a quick look at the origin and significance of such problems. The volumes of the Series publish only papers that have been invited by one or more Guest Editors appointed by the Editorial Board, composed of the professors at the Department of Mathematics of the Second University of Naples.

*Quaderni di Matematica* publishes, as a rule, two volumes per year. The price of each volume is 42 euro in Italy and 52 euro abroad. Subscription requests and orders should be sent to

Aracne Editrice S.R.L.  
Via Raffaele Garofalo, 133 a-b  
00173 Roma (Italy)  
Tel./fax: +39 0693781065  
[www.aracneeditrice.it](http://www.aracneeditrice.it)  
e-mail: [info@aracneeditrice.it](mailto:info@aracneeditrice.it)

All correspondence concerning the exchange with other publications or any other matter should be sent to

Paola D'Aquino  
Segretario di *Quaderni di Matematica*  
Dipartimento di Matematica  
Seconda Università di Napoli  
Viale Lincoln, 5  
81100 Caserta (Italy)  
Tel.: +39 0823274716, fax: +39 0823274753  
email: [quaderni@unina2.it](mailto:quaderni@unina2.it)

