Preface

Conjugate direction methods were proposed in the early 1950s. When high speed digital computing machines were developed, attempts were made to lay the foundations for the mathematical aspects of computations which could take advantage of the efficiency of digital computers. The National Bureau of Standards sponsored the Institute for Numerical Analysis, which was established at the University of California in Los Angeles. A seminar held there on numerical methods for linear equations was attended by Magnus Hestenes, Eduard Stiefel and Cornelius Lanczos. This led to the first communication between Lanczos and Hestenes (researchers of the NBS) and Stiefel (of the ETH in Zurich) on the conjugate direction algorithm. The method is attributed to Hestenes and Stiefel who published their joint paper in 1952 [101] in which they presented both the method of conjugate gradient and the conjugate direction methods including conjugate Gram–Schmidt processes. A closely related algorithm was proposed by Lanczos [114] who worked on algorithms for determining eigenvalues of a matrix. His iterative algorithm yields the similarity transformation of a matrix into the tridiagonal form from which eigenvalues can be well approximated. The three-term recurrence relation of the Lanczos procedure can be obtained by eliminating a vector from the conjugate direction algorithm scheme. Initially the conjugate gradient algorithm was called the Hestenes–Stiefel–Lanczos method [86].

According to Hestenes [100] the concept of conjugacy was introduced in his joint paper with G.D. Birkhoff in 1936 in the context of the variational theory. At that time Hestenes also worked on the Gram–Schmidt process for finding conjugate diameters of an ellipsoid. However the latter work was never published; instead the concept of conjugacy was used by Hestenes to develop a general theory of quadratic forms in Hilbert space.

The initial numerical experience with conjugate gradient algorithms was not very encouraging. Although widely used in the 1960s their application to ill-conditioned problems gave rather poor results. At that time preconditioning techniques were not well-understood. They were developed in the 1970s together with methods intended for large sparse linear systems; these methods were prompted by the paper by Reid [184].

Although Hestenes and Stiefel stated their algorithm for sets of linear equations, from the beginning it was viewed as an optimization technique for quadratic functions. In the 1960s conjugate gradient and conjugate direction methods were extended to the optimization of nonquadratic functions. The first algorithm for nonconvex problems was proposed by Feder [64] who suggested using conjugate gradient algorithms for solving some problems in optics. His work referred to the earlier paper of Davidon [50] in which the prime concern was the updating formula for the inverse Hessian approximation. Then the convergence of several versions of a conjugate gradient algorithm for nonquadratic functions were discussed by Fletcher and Reeves [73], by Polak and Ribière [161] and by Polyak [162].

The work by Davidon on a variable metric algorithm was followed by that of Fletcher and Powell [72]. Other variants of these methods were established by Broyden [17], Fletcher [68], Goldfarb [83] and Shanno [190] creating one of the most effective techniques for minimizing a nonquadratic function. The main idea behind variable metric methods is the construction of a sequence of matrices providing improved approximation of the Hessian matrix (or its inverse) by applying rank-one (or rank-two) update formulae. These methods are also conjugate gradient algorithms when applied to a quadratic function and in that case they evaluate the exact Hessian matrix in a finite number of iterates. Thus variable metric methods have a special feature which is not present in other conjugate gradient algorithms. In the monograph we do not pay much attention to this feature of variable metric techniques since our main interests are algorithms for large scale problems. Variable metric approximations to the Hessian matrix are in general dense matrices and for that reason they are not suitable for problems with many variables.

This does not mean that I rule out quasi-Newton methods altogether since one of the main themes of the monograph are limited memory quasi-Newton methods which use a variable metric updating procedure but within the prespecified memory storage. These methods no longer guarantee the exact Hessian reconstruction in a finite number of steps yet they are conjugate gradient algorithms and are intended for large scale problems. The family of these methods was started by the work of Nocedal [144] and one of the most interesting techniques with limited memory feature are limited memory reduced-Hessian algorithms discussed by Gill and Leonard [81].

My interest in conjugate gradient algorithms started during my work on my doctoral thesis which focused on minimax optimal control problems. There is substantial evidence of application of conjugate gradient algorithms to solving control problems [7], [116], [155], [198]. At the time of my PhD work, we observed spectacular development of nondifferentiable optimization. Beginning from the influential paper by Clarke [32] on a subdifferential of a locally Lipschitzian function, efforts were undertaken to use a new concept of differentiability with the aim of developing algorithms for problems which do not have continuously differentiable functions. Lemaréchal [118] and Wolfe [209] proposed such algorithms for convex functions. They observed that their methods are conjugate direction methods when applied to quadratic functions. Therefore, it was natural to place the Lemaréchal–Wolfe algorithm among other versions of conjugate gradient algorithms. This was done in my

PhD thesis [172] and later presented in a paper [174]. Since at that time my interest in conjugate gradient algorithms was related to optimal control problems the early application of new versions of Lemaréchal–Wolfe algorithm were reported in the context of minimax optimal control problems [173, 178].

I showed that the Lemaréchal–Wolfe method is equivalent to the Fletcher–Reeves version of a conjugate gradient algorithm provided that the exact line search is employed [174]. Later, Day and Yuan referred to the algorithm as the method of shortest residuals [43] pointing out that under that name the algorithm for quadratic functions was also discussed in the monograph by Hestenes [100].

This book reflects my propensity for the method of shortest residuals. It gives the comparison of the original method stated in [174] to other conjugate gradient algorithms making the monograph the overview of standard conjugate gradient algorithms. However, the monograph goes beyond the treatment of techniques which can be regarded as the extension of the method originally proposed by Hestenes, Lanczos and Stiefel. It draws much attention to preconditioned versions of the method, and since limited memory quasi-Newton algorithms are preconditioned conjugate gradient algorithms when applied to quadratics, these variable metric techniques are also discussed.

Many optimization problems are formulated with bounds on the variables. Such problems cannot be treated directly by conjugate gradient, or variable metric algorithms. There were several proposed modifications to these methods to cope efficiently with these constraints. It appears that the way the bounds are treated by optimization techniques has a significant impact on their efficiency. In the monograph, conjugate gradient, preconditioned conjugate gradient and limited memory quasi-Newton algorithms with modifications aimed at simple constraints are presented.

I made efforts to compare numerically the discussed algorithms; however in some cases the changes to a code required adapting it to the used testing environment, were not so obvious to me. In these cases I decided not to present the code numerical behavior as it was not observed on the common set of testing problems. Nevertheless, I hope that the included numerical comparisons give an insight into the efficiency of different classes of conjugate gradient and quasi-Newton algorithms intended for large scale problems. If so, one of the aims of my work has been achieved. But, undoubtedly, this work has come into existence due to my fascination for the optimization technique which I regard to be at the core of optimization.

I am indebted to various writers on numerical methods for optimization. I have been particularly influenced by the writing of Fletcher [71], Nocedal and Wright [146] and by Hestenes [100] on the particular subject considered in the monograph. Part of the work was carried out during my PhD studies at the Institute of Automatic Control of Warsaw University of Technology; the other at the Center for Processes Systems Engineering at Imperial College in London. Both places provided a research-stimulating atmosphere without which the work reported in the monograph would not have been possible.

Several people have influenced this book in various ways. No attempt has been made to give individual credit except in special instances. I am most grateful to

Lucien Polak for several discussions we had during his visit to Imperial College, to Mehiddin Al-Baali for pointing out his joint work with Robert Fletcher on preconditioned conjugate gradient algorithms, and to my PhD student Tomasz Tarnawski for his coding of some of the algorithms presented in this book.

Warsaw, July 2008 *Radosław Pytlak*

Springer

http://www.springer.com/978-3-540-85633-7

Conjugate Gradient Algorithms in Nonconvex Optimization
Pytlak, R.
2009, XXVI, 478 p., Hardcover
ISBN: 978-3-540-85633-7