Iterative Solution of Linear Systems in the 20-th Century

Yousef Saad* Henk A. van der Vorst†

Abstract

This paper sketches the main research developments in the area of iterative methods for solving linear systems during the 20th century. Although iterative methods for solving linear systems find their origin in the early nineteenth century (work by Gauss), the field has seen an explosion of activity spurred by demand due to extraordinary technological advances in engineering and sciences. The past five decades have been particularly rich in new developments, ending with the availability of large toolbox of specialized algorithms for solving the very large problems which arise in scientific and industrial computational models. As in any other scientific area, research in iterative methods has been a journey characterized by a chain of contributions building on each other. It is the aim of this paper not only to sketch the most significant of these contributions during the past century, but also to relate them to each other.

1 Introduction

Numerical linear algebra is an exciting field of research and much of this research has been triggered by a problem that can be posed simply as: Given $A \in \mathcal{C}^{m\times n}$, $b \in \mathcal{C}^m$, find solution vector (s) $x \in \mathcal{C}^n$ such that $Ax = b$. Many scientific problems lead to the requirement to solve linear systems of equations as part of the computations. From a pure mathematical point of view, this problem can be considered as being solved in the sense that we explicitly know its solution in terms of determinants. The actual computation of the solution (s) may however lead to severe complications, when carried out in finite precision and when each basic arithmetic operation takes finite time. Even the "simple" case when $n = m$ and A is nonsingular, which is a trivial problem from a mathematical point of view, may become very complicated, from a computational point of view, and may even turn out to be impossible.

The traditional way to solve a nonsingular linear system is to employ Gaussian elimination, and, with all enhancements to overcome numerical instabilities, this process can be carried out in $\mathcal{O}(n^3)$ basic floating point operations (additions and multiplications, assuming $n = m$). Many applications lead to linear systems with a large n (where the notion of "large" depends, of course, on the capacity of the available computer), and it became soon evident that one has to exploit specific properties of the A at hand in order to make solution of the system feasible. This has led to variants of Gaussian elimination in which the non-zero structure of A is exploited, so that multiplications with zero result are avoided and that savings in computer storage could be realized.

Another direction of approach was based on the solution of a nearby linear system, with a matrix that admits a computational cheap process (in terms of computing time and computer storage), and to embed this in an iterative process. Both approaches aim at making the impossible possible, and for the novice in this field this may seem to be just a collection of clever programming tricks: "in principle solving the problem is well-understood but one has to be well-organized to make the computational process a little faster". For this novice it will certainly come as a big surprise that a whole, still incomplete, mathematical framework had to be developed with deep and elegant results. As a result, relevant systems could be solved many orders of magnitude faster (and also often more accurate) than by a straightforward Gaussian

*Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA, e-mail saad@cs.umn.edu. Work supported by NSF/CCR and by the Minnesota Supercomputer Institute.

†Department of Mathematics, Utrecht University, Utrecht, The Netherlands, e-mail vorst@math.uu.nl

elimination approach. In this paper, we will sketch the developments and progress that has taken place in the twentieth century with respect to iterative methods alone. As will be clear, this subfield could not evolve in isolation, and the distinction between iterative methods and Gaussian elimination methods is sometimes artificial - and overlap between the two methodologies is significant in many instances. Nevertheless, each of the two has its own dynamics and it may be of interest to follow one of them more closely.

It is likely that future researchers in numerical methods will regard the decade just passed as the beginning of an era in which iterative methods for solving large linear systems of equations started gaining considerable acceptance in real-life industrial applications. In looking at past literature, it is interesting to observe that iterative and direct methods have often been in competition for solving large systems that arise in applications. A particular discovery will promote a given method from one camp only to see another discovery promote a competing method from the other camp. For example, the 50s and 60s saw an enormous interest in relaxation-type methods - prompted by the studies on optimal relaxation and the work by Young, Varga, Southwell, Frankel and others. A little later, sparse direct methods appeared that were very competitive - both from the point of view of robustness and computational cost. To this day, there are still applications dominated by direct solvers and others dominated by iterative solvers. Because of the high memory requirement of direct solvers, it was sometimes thought that these would eventually be replaced by iterative solvers, in all applications. However, the superior robustness of direct solvers prevented this. As computers have become faster, very large problems are routinely solved by methods from both camps.

The penetration of iterative solvers into applications has been a slow process that is still ongoing. At the time of this writing for example, there are applications in structural engineering as well as in circuit simulation, which are dominated by direct solvers.

This review will attempt to highlight the main developments in iterative methods over the past century. It is clear that a few pages cannot cover an exhaustive survey of 100 years of rich developments. Therefore, we will emphasize the ideas that were successful and had a significant impact.

Among the sources we used for our short survey, we would like to mention just a few that are notable for their completeness or for representing the thinking of a particular era. The books by Varga [170] and Young [186] give a complete treatise of iterative methods as they were used in the 60s and 70s. Varga's book has several excellent historical references. These two masterpieces remained the handbooks used by academics and practitioners alike for three decades. Householder's book [90] contains a fairly good overview of iterative methods - specifically oriented towards projection methods. Among the surveys we note the outstanding booklet published by the National Bureau of Standards in 1959 which contains articles by Rutishauser [137], Engeli [60] and Stiefel [156]. Later Birkhoff [19], who supervised David Young's PhD thesis in the late 1940s, wrote an excellent historical perspective on the use of iterative methods as he experienced them himself from 1930 to 1980. The more recent literature includes the books by Axelsson [6], Brezinski [24], Greenbaum [78], Hackbusch [85], and Saad [144], each of which has a slightly different perspective and emphasis.

2 The quest for fast solvers: a historical perspective

Iterative methods have traditionally been used for the solution of large linear systems with diagonally dominant sparse matrices. For such systems the methods of Gauss-Jacobi and Gauss-Seidel could be used with some success, not so much because of the reduction in computational work, but mainly because of the limited amount of memory that is required. Of course, reduction of the computational work was also a serious concern, and this led Jacobi (1846) to apply plane rotations to the matrix in order to force stronger diagonal dominance, giving up sparsity. Jacobi had to solve many similar systems in the context of eigenvalue computations; his linear systems were rather small, say of the order $10 - 20$.

In this century, simple iterative methods were predominantly applied for solving discretized elliptic self-adjoint partial differential equations, together with a local parameter for accelerating the iteration process. The first and simplest of these methods is Richardson's method [133]. Actually, this method was later viewed as a polynomial method and many authors have sought to optimize it by selecting its parameters so that the iteration polynomials became the Chebyshev polynomials; this was work done in the period 1950-1960 by Young, Lanczos and others. In the second half of this decade it became apparent that using the explicit three-term recurrence relation between Chebyshev polynomials, which led to three-term recurrence iteration methods (rather than the classical methods that are two-term iterations), were numerically superior in terms of stability (Golub and Varga, [77] 1961).

The acceleration of the slightly more difficult to analyze Gauss-Seidel method led to point successive overrelaxation techniques introduced simultaneously by Frankel [69] and by Young [184]. It was shown, for rather simple Dirichlet problems, that a suitably chosen relaxation parameter could lead to drastic improvements in convergence. Young showed that these improvements could be expected for a larger class of matrices, characterized by his *property A* [184]. Successive overrelaxation methods, and numerous variants, became extremely popular and were the method of choice in computer codes for large practical problems, such as nuclear reactor diffusion, oil reservoir modeling and weather prediction. Although their popularity has been overshadowed later, around after 1980, by more powerful techniques, they are still used in some applications either as the main iterative solution method or in combination with recent techniques (e.g. as smoothers for multigrid or as preconditioners for Krylov methods). The Successive Over-Relaxation (SOR) methods made it possible to solve efficiently systems with in the order of $20,000$ unknowns by 1960 (Varga [170]), and by 1965 systems of the order of $100,000$ could be solved in problems related to eigenvalue computations in nuclear diffusion codes. The success of the SOR methods has led to a rich theory for iterative methods; this could be used fruitfully for the analysis of later methods as well. In particular, many methods, including SOR, could be viewed as simple Richardson iterations for specific splittings of the matrix of the linear system.

In 1955, Peaceman and Rachford [128] suggested a splitting that was motivated by the observation that the matrix for a three-point finite difference stencil for a one-dimensional second order PDE is tridiagonal and this system can easily be solved. Their suggestion was to view the five-point finite difference approximation for a two-dimensional problem as the direct sum of two one-dimensional approximations. This led to an iteration in which alternatingly a tridiagonal associated with one of the two directions was split off, and this became popular as the alternating direction iteration (ADI). With the inclusion of iteration parameters, that steered the inclusion of a diagonal correction to the iteration matrices, the resulting ADI iterations could be tuned into a very effective methods. Varga [170] gives a good overview of the theory for understanding ADI methods. He, as well as Birkhoff [19] mentions that ADI was initially derived as a by-product of numerical methods for parabolic equations (the correction to the diagonal was motivated by the effect of the time derivative in these methods). Sheldon and Wachspress, in 1957, gave an early proof for the convergence of ADI for fixed parameters [174]. Wachspress discusses these ADI methods in his book [175] and considers also other grid-oriented acceleration techniques. One of these techniques exploits approximations obtained on coarser grids and can be viewed as a primitive predecessor to multigrid.

The first half of the century begins also with simple local projection methods, in which one attempts to solve a set of equations by solving each separate equation by a correction that is small in some norm. These methods could be used for over- or underdetermined linear systems, such as those that arise in tomography problems. This has led to the methods of Cimmino [37] and Kaczmarz [96], which were later identified as instances of Gauss-Jacobi and or Gauss-Seidel for related systems with A^TA or AA^T. Modern variants of these methods, under the name of ART and SIRT are very popular, for instance in medical and seismic tomography. ART and SIRT can be related to SOR and Block SOR. Spakman and Nolet [154] report on the solution of $292,451$ by $20,070$ systems related to structures of the upper earth mantle, with these methods (and with LSQR).

The second half of the century was marked by the invention (paper published in 1952) of the Conjugate Gradient method by Hestenes and Stiefel [89] and the Lanczos algorithm for linear systems [106]. This started the era of Krylov iterative methods. These were initially not viewed as iterative techniques, but as direct solution algorithms since they terminated in fewer than n steps, if n is the order of the matrix. A little earlier, papers by Lanczos [104] and by Arnoldi [2] had addressed the issue of transforming a matrix into simpler form for the purpose of diagonalizing it. These four papers together set the foundations of many methods that were developed later.

A famous publication by Engeli et al. [61] considered the method as an iterative process and showed that in rounding precision arithmetic, the Conjugate Gradient method did not terminate in the expected number of iteration steps (equal to at most the order of the matrix). This was shown for a matrix of order 64, a discretized biharmonic problem. Convergence occurred only after a few hundred steps. Notwithstanding this apparent failure, the method appeared later in the famous Wilkinson and Reinsch collection [183] as a kind of memory-friendly direct technique. It was mentioned that actual convergence might occur only after m iterations, where m could be 3 up to five times the order of the matrix. Because of this not well-understood behavior in rounded arithmetic, the method did not make it to the first universal linear algebra package LINPACK (mid seventies). In the early sixties it became clear that the convergence of the Conjugate Gradient method depends on the distribution of the eigenvalues of the matrix, and not so much on the order of the matrix, and it was Reid [132] who suggested to use the Conjugate Gradients method again as an iterative technique, but now for large sparse linear systems arising in the discretization of certain PDEs. Soon after this, the notion of preconditioning (already proposed in the Hestenes and Stiefel paper) became quite popular. Thus, the incomplete Choleski decompositions of Meijerink and van der Vorst [114] led to the ICCG process, which became the de facto iterative solver for SPD systems.

Hence, it took about 25 years for the conjugate gradient method to become the method of choice for symmetric positive definite matrices (the incomplete Choleski decompositions were shown to exist for M matrices). A good account of the first 25 years of the history of the CG method was given by Golub and O'Leary [76].

The unsymmetric variants of the Krylov methods required a similar amount of time to mature. The late sixties, early seventies, saw the roots for such methods. Techniques named ORTHODIR, ORTHOMIN, FOM, and others, were introduced but in their original formulations, these methods suffered from breakdowns and numerical instabilities. The GMRES variant, introduced by Saad and Schultz [145], was designed to avoid these undesirable features and became fairly popular. However, it suffered from the disadvantage of requiring increasing computational resources for increasing numbers of iterations. Bi-CG, the unsymmetric variant of conjugate gradients, did not have these disadvantages. The method, based on the unsymmetric Lanczos method (1952), was introduced by Fletcher in 1976 [66], but it is mathematically equivalent to a technique that had already been described in Lanczos' paper. Bi-CG, however, suffered from other practical problems, known as breakdowns of the first and second kind, which prevented early success. Moreover, the occurrence of nonorthogonal transformations led to much suspicion among numerical analysts. Nevertheless, the method became quite popular in a variant known as CGS (Sonneveld, 1984) [152] in which the method, for virtually equal costs could be applied twice, leading often to a twice as fast convergence, but also amplifying the problems of Bi-CG. In the eighties, Parlett and co-authors [127] and later Freund and Nachtigal [72] have shown how to repair the deficiencies in the Bi-CG method so that rather reliable software could be constructed. More recently, we have seen hybrids of the Bi-CG and GMRES approaches, with Bi-CGSTAB [168] as one of the most popular ones.

Originally, the usage of iterative methods was restricted to systems related to elliptic partial differential equations, discretized with finite difference techniques. Such systems came from oil reservoir engineering, weather forecasting, electronic device modeling, etc. For other problems, for instance related to various finite element modeling, practitioners preferred the usage of direct solution techniques, mainly efficient variants of Gaussian elimination, because of the lack of robustness of iterative methods for large classes of matrices. Until the end of the eighties almost none of the big commercial packages for finite element problems included iterative solution techniques. Simon [150] (1989) presented results, obtained for matrices of the order of $55,000$, for direct solution techniques. On the then fastest supercomputers, this required in the order of a few minutes of computing time. He claimed that direct sparse solvers would remain the method of choice for irregularly structured problems. Although this is certainly true if the structure of the matrix allows for an efficient elimination process, it became clear that for many PDE-related problems, the complexity of the elimination process increased too much to make realistic 3 dimensional modeling feasible. Irregularly structured finite element problems of order $1,000,000$, as foreseen by Simon, may be solved by direct methods - given a large enough computer (memory wise) but at tremendous cost and difficulty. However, *some* of them can be solved with iterative techniques, *if* an adequate preconditioning can be constructed. In the last decade of this century, much effort was devoted to the identification of effective preconditioners for classes of matrices. For instance, Pomerell [129] in 1994 reports on successful application of preconditioned Krylov methods for very ill-conditioned unstructured finite element systems of order up to 210,000 that arise in Semiconductor device modeling.

While using iterative methods still requires know-how, skill, and insight, it can be said that enormous progress has been made for their integration in real-life applications. Still, linear systems arising from many relevant problems, for instance large electric and electronic circuits, are not easy to solve in an efficient and reliable manner by iterative methods. Steady progress is made but the field as a whole can still be viewed as being in its infancy.

3 Relaxation-based methods

The Gauss-Seidel iteration was the starting point for the successive over-relaxation methods which dominated much of the literature on iterative methods for a big part of the second half of this century. The method was developed in the 19th century, originally by Gauss in the mid 1820s and then later by Seidel in 1874 (see references in [90]). In fact, according to Varga [170], the earliest mention on iterative methods is by Gauss (1823). This method as it was developed in the 19th century was a relaxation technique, in which relaxation was done by "hand". It was therefore natural to eliminate the largest components, see for example Dedekind [47] and Liebmann [107]. This method is referred to as Nekrasov's method in the Russian literature [117]. Referring to the more modern method in which relaxation was done in a cyclic manner, Forsythe is quoted as having stated that "the Gauss-Seidel method was not known to Gauss and not recommended by Seidel", see [90, p. 115].

However, the blossoming of overrelaxation techniques seems to have been initiated by the PhD work of David Young [184]. Young introduced important notions such as *consistent ordering* and *property A*, which he used for the formulation of an elegant theory for the convergence of these methods. Generalizations of Young's results to other relevant classes of matrices were due Varga, who published his book on Matrix Iterative Analysis in 1962. For decades to come this book was the standard text for iterative methods for linear systems. It covered important notions such as *regular splittings*, a rather complete theory on *Stieltjes* and *M-matrices*, and a treatment of semi-iterative methods, including the Chebyshev semi-iteration method. The latter method, analysed by Golub and Varga [77], also became more widely known, especially in the period when inner products were relatively expensive.

The accelerated Gauss-Seidel methods have motivated important developments in the theory of matrix linear algebra. In particular, relevant properties for M-matrices, introduced by Ostrowski in 1937 [122], were uncovered and convergence results for so-called *regular splittings*, introduced by Varga (1960) [171] were established. A cornerstone in the convergence theory was the theorem of Stein-Rosenberg (1948) [155] which proved relations between the asymptotic rates of convergence for the successive overrelaxation methods, including the Gauss-Seidel method, and the Gauss-Jacobi method. The concept of *irreducibility* of a matrix, a natural property for grid-oriented problems, helped to extend results for strongly diagonally dominant matrices to matrices for which the strict diagonal dominance inequality is required to hole only for one single equation at least. Another important notion is the concept of cyclic matrices: an irreducible matrix with k eigenvalues of modulus $\rho(A)$ is said to be of index A. Varga [170] gives a good overview of the relevant theory and the implications of this concept for iterative methods. It has a close relationship with Young's property A [170, p.99], and provides the basis for the convergence theory of the SOR methods. Sufficient conditions for the convergence of the SOR methods were given by theorems of Ostrowski (1954) [123] and Reich (1949) [131]. Lower bounds for the spectral radius of the SOR iteration matrix were derived by Kahan [97] (1958). This together provided the basis for a theory for iterative methods, published in Varga's book [170] from which many new methods emerged. Later, in the seventies, major part of this theory served well in the development of preconditioners for Krylov methods.

The following is a quotation from Varga's book (page 1) "As an example of the magnitude of problems that have been successfully solved on digital computers by cyclic iterative methods, the Bettis Atomic

Power laboratory of the Westinghouse Electric Corporation had in daily use in 1960 a two-dimensional program which would treat as a special case, Laplacean-type matrix equations of order 20,000". So the state of the art in 1960 was a 20,000 $\times$ 20,000 Laplace equation.

In the late 1960s and early 1970s a number of methods appeared in which the order of relaxation was not prescribed or even deterministic. These were appropriately termed "chaotic" or "asynchronous" relaxations. It was established that if a variable is relaxed an infinite number of times, the global method would always converge for any order in which the relaxation takes place. A few of the main contributions were by Chazan and Miranker [34], Miellou [116], Robert [134] and Robert et al. [135]. These methods were motivated by parallelism and were essentially ahead of their time for this reason.

4 Richardson and projection methods

Another line of development started with Richardson's methd (1910)

$$x_{k+1} = x_k + \omega r_k = (I - \omega A)x_k + \omega b,$$

which can be viewed as a straight-forward iteration associated with the splitting $A = K - R$, with $K = \frac{1}{\omega}I$, $R = \frac{1}{\omega}I - A$. Here r_k is the residual vector of the current iterate:

$$r_k = b - Ax_k$$

For the residual at the $k+1$-st step, one obtains

$$r_{k+1} = (I - \omega A)^{k+1} r_0 = P_{k+1}(A) r_0,$$

where $P_{k+1}(A)$ is a $k+1$ degree polynomial in A, with $P_{k+1}(t) = (1-t)^{k+1}$. It is easy to see that for symmetric positive definite matrices the process will converge for ω in the open interval $0 < \omega < \lambda_{max}$ where λ_{max} is the largest eigenvalue of A. In addition the best ω is is known to be $2/(\lambda_{min} + \lambda_{max})$, see e.g., [170, 144] for details.

The original Richardson iteration is readily generalized by taking a different $\omega = \omega_k$ for each iteration, which leads to the generalized Richardson iteration

$$x_{k+1} = x_k + \omega_k r_k \tag{1}$$

The sequence of ω_ks can be selected in a number of different ways. Note that the residual vector after step $k+1$ is given by

$$r_{k+1} = (I - \omega_k A)(I - \omega_{k-1} A) \cdots (I - \omega_0 A) r_0, \tag{2}$$

which shows that we can obtain any desired polynomial P_{k+1} with the property that $P_{k+1}(0) = 1$, by selecting its roots as the iteration parameters in (1). This process is referred to as polynomial acceleration of the simple splitting for A that we gave above. It was studied by, among others, Young [185], Shortley [148], and Lanczos (1952). By 1962 it was not considered competitive, since, as quoted from [170, p.159]: "*Richardson's method has the disadvantage of being numerically unstable*".

In fact, the Chebyshev semi-iteration method (Golub and Varga [77]) can, in exact arithmetic, be obtained from these polynomial accelerated methods, by choosing the acceleration parameters in successive Richardson iterations properly, but this approach is unstable. In the Chebyshev semi-iteration method one exploits the three term recurrence relation for Chebyshev polynomials, which leads to a stable three term recurrence iterative method. The main problem with these Chebyshev methods is that one needs fairly accurate information about extremal eigenvalues, since these define the interval for the Chebyshev polynomials.

The method of steepest descent which is attributed to Cauchy (1847) is also of the form (1). Kantorovitch later considered the method in a 1945 paper [100] that appeared in the Russian Doclade Academe Nauke. In this case, the scalar ω_k is selected so as to minimize the quadratic form

$$J(x) = \frac{1}{2}(Ax, x) - (b, x)$$

in the direction of r_k.

In the fifties and sixties other matrix splittings were suggested as a basis for iteration methods. We mentioned before one such splitting, namely the ADI method of Peaceman and Rachford [128]. In 1968, Stone [157] proposed the *Strongly Implicit Procedure*, which is in fact a simple Richardson iteration with a series of splittings of the matrix. The idea, for a five-point finite difference stencil, is to factor the matrix in a lower triangular matrix and an upper triangular matrix each with a three-point stencil (as in incomplete LU with no fill-in). The factors are chosen in such a way that the fill-in is spread over the seven-point stencil of the product matrix. This is different from the Kendall-Dupont-Rachford (1968) [53] decomposition, where the fill-in is compensated by a correction to the diagonal in such a way that the sum of the elements of the error matrix equals zero for eachy row. Convergence in the SIP method is achieved by a set of iteration parameters (up to 18) for different distributions of the elements in the error matrix. The choice of the parameters is motivated by a Fourier analysis. Each value of the iteration parameter kills some components in the error (but may lead to increase in other components). Successful application of SIP requires to apply a special order of the iteration parameters, and for each value the decomposition has to be carried out from top to bottom and next from bottom to top. The SIP method gained quite some popularity in oil reservoir simulation and groundwater simulation problems, but its usage seem to have declined in favor of the Krylov methods. The Dupont-Kendall-Rachford splitting was proposed to be used in combination with Chebyshev polynomial acceleration.

In 1937 two papers were published on methods that can be termed 'row-projection methods' (or column projection methods). These methods proposed by Kaczmarz [96] and Cimmino [37] were also based on one-dimensional corrections:

$$x_{k+1} = x_k + \omega_k a_{i,.}, \tag{3}$$

where $a_{i,.} = A^T e_i$ is the i-th row vector of A. These rows are cycled through from 1 to n. Here ω_k is selected so that the i-th component of r_{k+1} becomes zero. Because $r_{k+1} = r_k - \omega_k AA^T e_i$, it is easy to see that this method is mathematically equivalent to the Gauss-Seidel iteration applied to the normal equations

$$AA^T y = b, \quad x = A^T y$$

The method proposed by Cimmino was the Jacobi equivalent of this approach. It is also possible to define similarly a Gauss-Seidel approach for the normal equations

$$A^T Ax = b$$

as was noted by Björk and Elfving [20], and this would correspond to taking directions d_k along the columns of the matrix. This class of methods regained interest in the 70s and 80s with the work of Bjork and Elfving [20], Tanabe [159], and later Kamath and Sameh [98] and Bramley and Sameh [21].

However, one-dimensional projections methods of a different type, based on very general definitions of norms were very popular in the late 50s. Here, we mention the work of Gastinel among others. Gastinel's approach [74] consisted of defining generating vectors for norms. Consider an arbitrary vector norm ϕ (for example the norm $||.||_1$). Gastinel defines the vector v which realizes the norm of a vector r in the sense

$$(v, r) = \phi(r).$$

For example, for the 1-norm, the components of v can be defined as $v_i = \text{sign}(e_i^T r)$, where $r = b - Ax$ is the current residual vector. This vector v is chosen to do an orthogonal projection step. The method can be shown to converge for symmetric positive definite matrices.

5 Second-order and polynomial acceleration

An important observation regarding all acceleration methods of the form (1) is that their residuals take the form (2), so there is room for improvement to the scheme if successive iterates are considered.

In 1950 Frankel [69] proposed an acceleration method which used a three-term recurrence of the form

$$x_{k+1} = x_k + \beta_k \delta_k \quad \delta_k = r_k - \alpha_k \delta_{k-1}$$

This "second order Richardson process" is initialized by $\delta_{-1} = r_{-1} = 0$. Frankel's method [69] uses constant coefficients and results in a residual polynomial which is a combination of Chebyshev polynomials of the first and second kind.

Naturally Chebyshev polynomials should give rise to optimal-like behavior and a number of authors discovered, rediscovered, or enhanced the method at various times. The paper by Flanders and Shortley [65] showed how to use Chebyshev polynomials for eigenvalue calculations. Later Shortley [148] adapted this method for the solution of linear systems. In the Russian literature, Gavurin [75] also introduced the idea independently in 1950. In 1954 Young [185] proposed a method which amounted to compounding (or restarted) Chebyshev iterations. However the method was in the form (1) – which is unstable. Young gave some remedies to the process which consisted of reordering the roots ω_k before applying the polynomial.

In the particular case where A is symmetric positive definite the eigenvalues are located in an interval $[\lambda_{min}, \lambda_{max}]$. The best residual polynomial $1 - \lambda s(\lambda)$ in this case is a shifted and scaled Chebyshev polynomial of the first kind, and its three-term recurrence results in a simple three-term recurrence for the approximate solution, see, e.g., [144].

Using a different approach altogether, Lanczos in a 1952 paper [105] discusses the use of certain polynomials that are optimal in a least-squares sense. This paper, which was by and large overlooked by researchers, made a number of contributions. Its focus is on symmetric positive definite matrices - for general matrices the author resorts to the normal equations. One of the main ideas proposed is to consider the problem of approximating the delta function in the interval $[0, 1]$ which contains all eigenvalues (after the matrix is scaled by its largest eigenvalue). He then transforms the variable using the change of variables $x = (1 - cos\theta)/2$. Now a least-squares approximation to the delta function is sought in the trigonometric basis. This leads to the so-called Dirichlet Kernel whose solution is well-known

$$\begin{aligned} P_k(\theta) &= \frac{\sin(k+\frac{1}{2})\theta}{(k+\frac{1}{2})\sin\theta} \\ &= \frac{1}{k+\frac{1}{2}}\left[\frac{1}{2} + \cos\theta + \cos 2\theta + \cdots + \cos k\theta\right]. \end{aligned}$$

To avoid the high oscillations around discontinuities, the so-called Gibbs phenomenon, Lanczos suggested a strategy due to Fejer.

Later, a remarkable paper by Stiefel gave a fairly complete view on similar ideas revolving around least-squares polynomials [156]. The above paper by Lanczos was not referenced by Stiefel. It is only in 1983 that the idea of using least-squares polynomials resurfaced in force again, motivated essentially by parallelism and vector processing. Earlier in 1979 a paper by by Dubois, Greenbaum, and Rodrigue [51]. suggested using simple Neumann series expansion of the matrix. In 1976 Axelsson addressed the problem of computing good polynomials when the spectrum is located in two intervals, and he was followed later in 1980 by deBoor and Rice [46] who showed how to compute the best min-max polynomial in this situation and the more general situation of multiple intervals. The least-squares alternative considered by Johnson et al. in [94] was for a single interval, assuming that A is symmetric positive definite. In other words we need to solve

$$\begin{aligned} &\text{Find } s \in \Pi_k \text{ that minimizes:} \\ &\|1 - \lambda s(\lambda)\|_w, \end{aligned} \tag{4}$$

where w is some weight function on the interval $(\lambda_{min}, \lambda_{max})$, and $\|.\|_w$ is the L_2-norm associated with the corresponding inner product. Because the distribution of eigenvalues matters more than condition numbers for the preconditioned conjugate gradient method, the authors observed in [94] that least-squares polynomials tend to perform better than those based on the uniform norm, because they result in a better clustering of the spectrum. Moreover, Lanczos [105] and Rutishauser [137] already noted that the eigenvalue estimates need not be accurate: in fact it suffices to use the simple bounds that are provided by Gershgorin's theorem. Further experiments in [138] did confirm that in some cases the least squares polynomial over the Gershgorin interval, may perform as well as the infinity norm polynomial over $[\lambda_{min}, \lambda_{max}]$. Note that this is only a minor advantage of least-squares polynomials since effective adaptive procedures exist to compute $\lambda_{min}, \lambda_{max}$; see [86] for symmetric problems and [112, 59] for nonsymmetric problems. We should add that the observations made in [94] and in [138], and the simplicity of a method that bypasses eigenvalue estimates, have made least-squares polynomials more popular for polynomial preconditionings.

In the more general nonsymmetric case the interval (or union of intervals) that contains the spectrum is to be replaced by a more complex continuum E in $\mathbb{C}$, which ideally would contain the eigenvalues of the matrix A. Several choices have been used for E. The first idea, proposed by Manteuffel in 1977-78 [111, 112], is to use an ellipse E that encloses an approximate convex hull of the spectrum, or more precisely, the field of values of A. Then the shifted and scaled Chebyshev polynomials are optimal or nearly optimal and the use of these polynomials leads again to an attractive three-term recurrence. He exploited the fact that an unaccurate guess of extremal eigenvalues leads to either divergence or very slow convergence, in which the eigenvectors corresponding to the unidentified extremal eigenvalues play a dominant role. After a few iterations these directions can be identified and the parameters for the Chebyshev iteration polynomials can be adjusted. Although superseded by the parameter-free Krylov iteration methods, the Chebyshev methods are still of interest on computer platforms where the inner products are relatively expensive. They can be used in combination with Krylov methods, either as polynomial type preconditioners in order to damp dominanting parts of the spectrum, or to continue the iteration with the eigenvalue guesses that can be obtained from the Krylov methods (the Ritz values).

A second alternative is to use a polygon H that contains $\sigma(A)$ [147, 139]. A notable advantage of using polygons is that they may better represent the shape of an arbitrary spectrum. The polynomial is not explicitly known but it may be computed by a Remez algorithm. As in the symmetric case an alternative is to use some weighted L_2-norm instead of the infinity norm. Saylor and Smolarski used a discrete norm on the polygon. [147]. Saad [139] used an L_2-norm associated with Chebyshev weights on the edges of the polygon and expressed the best polynomial as a linear combination of Chebyshev polynomials associated with the ellipse of smallest area containing H.

Yet another attractive possibility, when polygons instead of ellipses, proposed by Fischer and Reichel [64] is to avoid the problem of best approximation altogether and interpolate the function $1/z$ with a polynomial at the Fejer points of E, i.e., the points $e^{2ji\pi/k}, j = 0, 1, \cdots, k$ that are conformally mapped from the unit circle to H. This is known to be an asymptotically optimal process. There are numerous publications related to this approach and the use of Faber polynomials; see the references in [64].

6 Krylov subspace methods: the first period

In the early 50s a number of new methods appeared that dramatically changed the landscape of iterative methods. In separate contributions Lanczos [106] and Hestenes and Stiefel [89] propose in effect different versions of what is now known as the Conjugate Gradient method. Amazingly, Hestenes and Stiefel actually discovered the same method independently[1]. The method proposed by Lanczos is, for symmetric positive definite matrices, mathematically equivalent to the Conjugate gradient method, but it was described for the general case of nonsymmetric matrices.

There is no doubt that the origin of this class of methods was deeply rooted in approximation theory and, in particular in orthogonal polynomials. The ideas behind "gradient methods" as this class of

[1] The anecdote told at the recent "Conference on preconditioning methods for sparse matrix problems in industrial applications" held in Minneapolis, by Emer. Prof. Marvin Stein, the post-doc who programmed the algorithm for M. Hestenes the first time, is that Stiefel was visiting UCLA at the occasion of a conference in 1951. Hestenes, then a faculty member at UCLA, offered to demonstrate this effective new method to Stiefel, in the evening after dinner. Stiefel was impressed by the algorithm. After seeing the deck of cards he discovered that this was the same method as the one he had developed independently in Zurich. Stiefel also had an assistant, by the name of Hochstrasser, who programmed the method.

methods was referred to, are based on some kind of global minimization. For instance, for positive definite symmetric A, the CG method minimizes the so-called A-norm: $\|x_i - x\|_A^2 \equiv (x_i - x, A(x_i - x))$, for x_i that are in the Krylov subspace $K_i(A, r_0) \equiv \{r_0, \ldots, A^{i-1}r_0\}$. For some PDE problems this norm is known as the Energy norm, which has physical relevance. Another interpretation of the gradient methods is that the residual is orthogonal to the space of previously generated residuals, or some related space. Both interpretations are useful for the formulation of methods as well as for the analysis. A very useful consequence from the Krylov subspace basis is that x_i can be expressed as a polynomial in A of degree $i-1$, acting on r_0. The minimization interpretation makes it possible to bound the error for CG by replacing the "CG-polynomial", by easier to analyse polynomials, for instance a Chebyshev polynomial. This leads to the well-known upper bound [99, 43, 40, 4]

$$\|x_i - x\|_A \le \sqrt{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^i \|x_i - x\|_A, \tag{5}$$

for symmetric positive definite matrices, in which $\kappa = \lambda_{\max}(A)/\lambda_{\min}(A)$. This upper bound describes well the convergence behavior for matrices A of which the eigenvalues are distributed rather homogeneously. For more uneven distributions one can obtain bounds by making more subtle choices for the approximating polynomials, for instance, products of suitably chosen Chebyshev polynomials [4]. These choices do not reveal the superlinear convergence that is often observed for CG and also for other Krylov subspace methods. The notion of superlinear convergence refers to the observation that the convergence of CG accelerates in the course of the process. Proofs for superlinear convergence had been given already in the early fifties [101, 87], but these did not reveal that the superlinear behavior may take place in early phases of the iteration process; they were qualitative rather than quantitative. Concus et al [40] related this convergence behavior to the Krylov subspace approximations, by stating that "the extremal eigenvalues are approximated especially well (by the Ritz values corresponding to the Krylov subspace) as CG proceeds, the iteration then behaving as if the corresponding eigenvectors are not present (thus leading to a smaller "effective" condition number in (5), which might then explain the faster convergence." In 1986, this was proven in a quantitative way [163] and it was shown that the relevant eigenvalues needed to be approximated only in a modest degree by Ritz values, for an acceleration to set in.

Surprisingly, it took some time before the ideas of the CG were generalized to other classes of matrices. Paige and Saunders [124], in 1975, showed that the underlying Lanczos method, could also be used for efficient implementations of related approaches for symmetric indefinite matrices. In MINRES, the norm of the residual $\|Ax_i - b\|_2$ is minimized, and this required a more careful approach in order to avoid breakdowns. Conjugate Gradient can be interpreted as a process in which A is projected to the Krylov subspace (in fact, the Lanczos approach), which leads to a tridiagonal matrix T of low dimension. The actual approximation x_i is determined by solving a small linear system with T, and this is done with LU without pivoting. This leads to the elegant short recurrences in CG. For indefinite matrices, the LU factorization could lead to numerical difficulties.

Paige and Saunders circumvented this by employing a QR decomposition of T, which leads again to regular short recurrences, making the MINRES method attractive because of its minimal overhead and its economy of storage. It may come as a small wonder that one can also minimize $\|x_i - x\|_2$, without knowledge of the solution x. Paige and Saunders accomplished this in their SYMMLQ, by restricting x_i to $AK_i(A, r_0)$. The advantage of SYMMLQ over MINRES appears to be in less sensitivity to ill-conditioning of A, the price one has to pay is that SYMMLQ often takes more steps. A slight disadvantage is also that although the method minimizes the norm of the error, the value of this norm is not known and the only practical information one has is the norm of the residual.

In 1976, Concus and Golub [38] and Widlund [181] came up with the idea of splitting a matrix into its symmetric and nonsymmetric parts and using the symmetric part as a preconditioner. With the proper inner product, the resulting algorithm corresponds to an iteration with a skew-Hermitian matrix - and therefore a three-term recurrence - CG-like - algorithm (called CGW) can be formulated.

The Bi-CG method, proposed in 1976 by Fletcher [66], is actually an implementation of the two-sided Lanczos process, which was suggested by Lanczos in 1952. In Bi-CG, the residual is constructed to be orthogonal to a Krylov subspace generated with A^T and some vector s_0. Initially, many numerical analysts were very skeptical of the Bi-CG, mainly because of the various breakdown situations that may occur. Also, Bi-CG did not minimize any norm of interest and the convergence behavior can be very irregular. The fact that the underlying two-sided Lanczos process works with non-orthogonal projections led to serious doubts on the usefulness of the method. A good example of this concern is in [182, page 394, 395], where the two-sided Lanczos method (viewed as a finite method for reducing a nonsymmetric matrix to tridiagonal form) is commented on: "... we may well have to pay a heavy price in terms of numerical instability..." and "... it is difficult to think of any reason why we should use Lanczos' method in preference of Householder's". Wilkinson's analysis is, of course, still valid. However in the context of solving large sparse linear systems, we have learned to make the two-sided Lanczos method and the Bi-CG into useful solution techniques thanks to a number of enhancements and a better understanding of these processes.

7 Krylov subspace methods: the second period

Of course, the success of Krylov methods for symmetric matrices has inspired the construction similar methods for unsymmetric matrices. The classical Lanczos method leads to a tridiagonal system, the projection of A with respect to the Krylov subspace. Factorization of this tridiagonal matrix as the product of a lower and an upper bidiagonal matrix, leads to the coupled two-term recurrences as in Bi-CG. As said before, Bi-CG suffered from several breakdown conditions. One breakdown, associated with the indefiniteness of the implicitly generated projected tridiagonal system, can be cured by admitting two by two blocks along the diagonal of one of the factors. This requires the combination of two successive iteration steps, which explains the name *Composite step Bi-CG* [11]. A similar idea had been used much earlier by Luenberger, in order to make the Conjugate Gradient algorithm robust for symmetric indefinite matrices [108]. The other breakdown, a more serious one, arises when the bi-orthogonalization process leads to a zero inner product of the two new vectors in the Krylov subspace and its adjoint space (that is the Krylov subspace, generated in the two-sided Lanczos process, with A^T). Likewise, a near breakdown should also be avoided since it may lead to inaccuracies in the approximated solution. This breakdown can be cured with a *look-ahead strategy*, first proposed by Parlett, Taylor, and Liu [127]. The idea is to expand the Krylov subspaces by two vectors simultaneously, and to make the new vectors block bi-orthogonal with respect to the similarly expanded adjoint space. Parlett et al considered only look-aheads of length two, but a few years later, around 1990, the idea was picked up almost simultaneously by a number of other researchers who generalized it to look-aheads of arbitrary length. The most well-known of these approaches were those published by Gutknecht and co-authors [83], Joubert [95], Parlett [126], Freund and Nachtigal [71], and Brezinski and co-authors [25, 26]. In the latter work, the look-ahead strategy was related to the theory of orthogonal polynomials and referred to as *recursive zoom* technique.

Curing the breakdowns in Bi-CG was important, but there were other aspects as well that motivated further research. The convergence behavior of Bi-CG is usually quite irregular, in the sense that the norms of successive residuals can behave erratically. This motivated Freund and Nachtigal [71] to propose an algorithm in which the projected overdetermined tridiagonal system is solved in a least-squares sense. Since the basis vectors for the Krylov subspace, generated by the two-sided Lanczos process, are in general not orthogonal, this approach does not lead to a minimum residual approximate solution (as with MINRES), and this inspired for the name QMR (*Quasi Minimum Residual*). The full QMR method includes a look-ahead strategy, but it became also popular without it, since the first breakdown condition is cured by the least squares solution of the tridiagonal system. For a template for this simplified QMR, see [12].

The other clear disadvantage in the basic two-sided Lanczos method was the necessity to construct two Krylov subspaces: $K^i(A; r_0)$, and $K^i(A^T; s_0)$. Of the two vector bases generated only one is exploited for the solution, the other can be regarded as an auxiliary set used to generate the inner products needed to generate the bi-orthogonal basis. Sonneveld [152] made the clever observation that the operations with A^T could be reformulated to operations with A, and these operations can be used for a further reduction of the residual norm. Whereas the previous approaches, look-ahead, composite-step, and QMR, help to cure deficiencies of the Lanczos method, they do not lead to essential faster convergence. However, Sonneveld's trick can lead to much faster convergence, for practically the same costs per iteration. The idea is based on the observation that the residual r_i can be expressed formally as $r_i = p_i(A)r_0$, with p_i a polynomial of degree i. Likewise, the shadow residual in the adjoint space can be expressed as $s_i = p_i(A^T)s_0$. The iteration coefficients for Bi-CG are computed from inner-products such as (r_i, s_i), and such an inner product can be rewritten, formally, as $(p_i(A)r_0, p_i(A^T)s_0) = (p_i^2(A)r_0, s_0)$. This observation leads to an algorithm that generates $\widetilde{x}_i \in K^{2i}(A; r_0)$, for which the corresponding residual $\widetilde{r}_i$ can be expressed as $\widetilde{r}_i = p_i^2(A)r_0$. For the situation where Bi-CG delivers a residual r_i $(= p_i(A)r_0)$ that is small compared with r_0, one may conclude that $p_i(A)$ has transformed r_0 into a small vector, and hopefully, if we apply $p_i(A)$ twice, then this leads to a double reduction. Indeed, the resulting method, CGS (Conjugate Gradients Squared), often leads to a convergence about twice as fast as Bi-CG. This algorithm is also referred to as Bi-CGS, which is actually more appropriate. The downside of the squaring of p_i is that the convergence of CGS is usually much more irregular than for Bi-CG. This can lead to serious accuracy problems in finite precision arithmetic; we will come back to this aspect later.

Soon after the discovery of the CGS method, it was recognized that the operations with A^T could also be transformed to other polynomials in A. The first idea in this direction was Bi-CGSTAB [168], in which Bi-CG was combined with minimum residual steps of degree one. This led to a convergence that is rather smooth as well as faster than Bi-CG and it gave rise to many other hybrids. Gutknecht suggested to combine $2i$ Bi-CG steps with i times a minimum residual method of degree 2. This was generalized by Sleijpen and Fokkema [151] to Bi-CGSTAB(ℓ). The same principles can also be applied to QMR, and the analogue of CGS led to TFQMR [70]. The analogue of Bi-CGSTAB is QMRSTAB, suggested by Chan et al [31]. Zhang [188] describes more general product methods based on Bi-CG. His framework includes the previously described methods, but also admits hybrid variants in which one can shift from CGS to Bi-CGSTAB at some iteration step. This principle admits further possibilities for reducing the residual in some norm. An interesting variant of CGS has been suggested by Fokkema et al [68]. Here, the polynomial p_i^2 that generates the residuals, is replaced by the product $p_i\widetilde{p}_i$, where $\widetilde{p}_i$ is the polynomial that corresponds to a 'nearby' Bi-CG process. This principle can be used to help reduce severe irregularities in the convergence, while the quadratically reduced errors in important eigenvector directions are still realized. According to the authors, this is an advantage in the context of iteratively solving for the Jacobian system in a Newton process for nonlinear systems.

A different direction is to try to minimize the norm of the residual over all vectors in the Krylov subspace, similar to the MINRES approach for symmetric A. A number of methods were proposed that achieved this goal, among them ORTHODIR [93], GCR [55], and ORTHOMIN [172], but many of these methods suffered from some numerical instability. GMRES [145] developed in 1986, was mathematically equivalent to these techniques but soon it came to be preferred because of its better numerical behavior and its lower cost, both in terms of memory and arithmetic. An advantage of GMRES is its guarantee to compute the approximate solution with minimum residual norm, but the price to be paid is that the overhead costs per iteration, that is the computations other than the matrix vector product, increase linearly with the iteration count. Also, all basis vectors for the Krylov subspace have to be stored. The obvious solution seems to be to restart after a convenient number of iterations, before the costs for computation and storage become prohibitive. This is known as restarted GMRES, or simply GMRES(m). The disadvantage of this approach is that it decreases the robustness of the method since convergence is no longer guaranteed. Moreover, by restarting the convergence behavior may become very slow and one runs the risk to miss the faster convergence that might have occurred for a larger value of m. For that reason, researchers have tried to find ways to reduce the number of iteration steps, other than by preconditioning, or even in addition to preconditioning. One idea is to try to improve the preconditioner with updates from the Krylov subspace. This has been suggested first by Eirola and Nevanlinna [54]. Their approach leads to iterative methods that are related to Broyden's method [29], which is a Newton-type method. For specific but obvious choices, one recovers a method that is equivalent to GMRES. The Broyden method can be obtained from this update-approach if we do not restrict ourselves to Krylov subspaces. See [173] for a discussion on the relation of these methods.

The idea of preconditioning is to approximate $A^{-1}p$ for vectors p generated by the iteration method.

One could do this in a different way for every iteration step, for instance, by incorporating information from previous iteration steps in the preconditioner, or by approximating $A^{-1}p$ by some iteration method again. The updated preconditioners can not be applied immediately to GMRES, since the preconditioned operator now changes from step to step, and we are not forming a regular Krylov subspace. However, we can still minimize the residual over the new subspace. The idea of variable preconditioning has been exploited in this sense, by different authors. Axelsson and Vassilevski [10] have proposed a Generalized Conjugate Gradient method with variable preconditioning, Saad [142] has proposed a scheme very similar to GMRES, called Flexible GMRES (FGMRES), and Van der Vorst and Vuik [169] have published a scheme called GMRESR. FGMRES has received more attention, possibly because it is fairly easy to implement: only the update directions in GMRES have to be preconditioned, and each update may be preconditioned differently. This means that only one line in the GMRES algorithm has to be adapted. The price to be paid is that the method is no longer robust; it may breakdown. GMRESR and the Generalized Conjugate Gradient method produce, in exact arithmetic, the same results, but GMRESR is numerically more stable and more efficient. In GMRESR the residual vectors are preconditioned and if this gives a further reduction then GMRESR does not breakdown. This gives slightly more control over the method in comparison with FGMRES. In most cases though the results are about the same.

Other methods that proved to be very useful include the LSQR method, suggested in 1982 by Paige and Saunders [125]. LSQR is a clever implementation of the Lanczos method that leads to a factorization of the tridiagonal reduced matrix for A^TA. The method is often the method of choice for overdetermined or underdetermined systems; it minimizes the norm of the residual over the Krylov subspace generated with A^TA. For square systems the method is not so effective, unless one finds a good preconditioner, since the convergence behavior of LSQR is dictated by the square of the condition number of the system involved. LSQR may also be viewed, in exact arithmetic, as CG applied to the normal equations. If one applies CG in this fashion, however, then one will see a loss in stability. Craig's method [125] is a Krylov method that also works with A^TA and in which the error $||x_i - x||_2$ is minimized, without computing the value of this norm.

Finally, we mention that Weiss, in the early nineties, gave generalizations of SYMMLQ for unsymmetric matrices. These methods are known as GMERR methods [177, 179]. It may have an advantage to have a method in which the norm of the error is minimized, but since this is done over a different subspace, namely $A^TK_i(A^T, r_0)$, it is not clear yet when this leads to advantages over, for example, GMRES, in terms of efficiency or stability.

Convergence results for Krylov methods in the non-Hermitian case were established following essentially similar ideas as for the CG algorithm, see [144, 78] for overviews. However, this simple analysis which was given, for example, for GMRES [145] was soon viewed as insufficient. One fundamental question that was asked in the early 80s was whether a short recurrence iterative process could be found that was also optimal for non-Hermitian matrices. Indeed, it was known how to generate short-term recurrence algorithms for nonsymmetric matrices (e.g. the BiCG) but these do not verify obvious optimality properties. On the other hand the optimal processes that were known required long recurrences (e.g., GMRES). The answer to this question was given by the excellent paper by Faber and Manteuffel in 1984, and alas it was a negative one [63]: short-term solution algorithms that are also optimal can essentially be devised only for a restricted class of matrices; for all practical purposes, these matrices are either Hermitian or Skew-Hermitian. Many attempts to get a better understanding of the behavior of GMRES were made. Work by Brown [28] and later by Cullum and Greenbaum [42] established relations between certain methods (GMRES and FOM, and then BiCG and QMR). Greenbaum and Strakos [79] showed a number of interesting properties of GMRES - in particular they established that essentially any convergence behavior is possible for the same spectrum. For the BiCG and related methods, very little theory is available on convergence. An attempt on this was to introduce variable metrics, see [13] and the survey by Weiss [178].

8 Accelerators are not enough: preconditioning methods

The convergence of iterative methods depends on spectral properties of the matrix of the linear system and in order to improve these properties one often transforms the linear system by a suitable linear transformation. This process is known as preconditioning.

We do not know for sure who coined the term 'preconditioning' first – it may have been Turing (according to Golub and O'Leary [76]) or Forsythe (see below). Regardless, the idea was known quite early on. In the 1952 paper by Lanczos [105] the notion of polynomial preconditioning is clearly defined: "*The construction of the inverse matrix is equivalent to a linear transformation which transforms the given matrix into the unit matrix. The unit matrix can be conceived as the extreme case of a well-conditioned matrix whose eigenvalues are all 1. We will ask for much less if we merely demand the transformation of the original system whose dispersion is moderate.*" Lanczos then states that the goal of the procedure is to "reduce the initial skewness" of the system, not bring about the exact solution. Forsythe in his report on this paper in the mathematical reviews (review MR 16-751, 1955) does employ the term "preconditioning" explicitly. Polynomial preconditioning is also clearly mentioned in the review paper by Stiefel [156] - in 1959. Hestenes in 1956 [88], viewed the conjugate gradient method as an acceleration technique for suitable matrix splittings. His formulation of the algorithm is equivalent with preconditioned conjugate gradients.

Modern preconditioning methods started in the late 60s and early seventies. Evans [62] suggested to *precondition* a matrix by using the sparse triangular factor approximations, in combination with the conjugate gradient method. In 1972, Axelsson [3] proposed to use the SSOR method as a preconditioner. The incomplete Choleski decompositions (Meijerink and van der Vorst, 1974, 1977), became quite popular and led to the ICCG process [114]. Concus, Golub and O'Leary [40] wrote an influential paper on the usage and effect of preconditioning for the CG method.

8.1 Incomplete Factorizations

Preconditioning as we know it today refers mostly to approximate or incomplete factorizations of the coefficient matrix. Some of the early publications on such factorizations that are often cited include Buleev [30], Varga [171]. and Oliphant [119]. Later in the 1960s a few other procedures were developed specifically for matrices arising from finite difference approximations to elliptic operators, these include the work by Dupont, Kendall, and Rachford [53]. In 1977 Meijerink and Van der Vorst introduced the more general incomplete LU factorization [114]. The paper suggests that the combination of this "preconditioning" and the conjugate gradient method could lead to a robust and very fast combination. Similar ideas were in existence before. However the paper provided a much needed stimulus to the whole area of iterative methods.

The Dupont-Kendall-Rachford splitting can be viewed as an incomplete LU factorization with zero fill-in, in which the elimination errors are compensated by corrections to the diagonal of the decomposition. In 1977 this procedure was generalized by Gustafsson [82] in 1978, as a modified form of the incomplete LU factorizations: MILU.

Several development marked the years that followed. Two distinct ways of developing incomplete factorization preconditioners with improved accuracy were developed. The first approach is based on a symbolic factorization view, i.e., it only requires the nonzero structure of the matrix to determine which fill-ins to drop. A method proposed by Watts [92] for irregular sparse matrices attributes a "level of fill" recursively to each fill-in element from the levels of fill-in of its parents, in the Gaussian elimination process. Then each fill-in that is introduced and whose level exceeds a certain threshold is dropped. In practice for M-matrices, the higher the fill-in the lower the elelent. The second common approach is to modify a given direct solver by including a dropping rule, based on the numerical size of the fill-ins introduced [73, 121, 45, 44, 189, 187].

Although the relation between the size of the dropped elements and the number of iterations required to achieve convergence is far from being understood, *on the average* dropping small elements is more likely to produce a better quality preconditioner than dropping large elements. However, experience reveals that this is not always true. Another drawback of the level-of-fill approach is that it is difficult to predict the amount of fill-in that will be generated.

The number of variations that can be found on incomplete factorization preconditioner is truly astounding and we will not attempt to list them all. It suffices to say that there were variants developed for specific architectures (e.g., Twisted Factorizations), or for specific applications (e.g., element-by-element preconditioners), or to exploit specific features of the equations (e.g., block factorizations), among other classes. See [50] for an overview of these preconditioners, specially in view of their implementation for high-speed computers.

One of the interesting recurring themes in preconditioning methods is whether or not reordering the matrix prior to applying the ILU factorization can be helpful. Two early papers examined this carefully and concluded rather negatively. The first is a paper by Simon [149] who considered large nonsymmetric linear systems. For the systems he considered he concluded that standard techniques used for sparse direct solvers were not too helpful for use in preconditioners based on level-of-fill. Immediately following this was a paper by Duff and Meurant [52] which concluded, similarly, that ICCG does not in general benefit in any significant manner form reordering. These studies were limited to certain types of reorderings and certain types of preconditioners. It is now known [16] that certain reorderings, such as Reverse Cuthill McKee are beneficial in preconditioning methods, in particular with some form of dropping strategy. What seems to be also clear is that the best approaches for direct solvers (such as Nested Dissection and minimal degree ordering) are not the best for iterative solvers.

Since ILU and IC factorizations were the most popular preconditioners, at least in a sequential environment, many attempts have been made to improve them, for instance by including more fill [115], by modifying the diagonal of the ILU factorization in order to force rowsum constraints [82, 7, 6, 118, 164, 57], or by changing the ordering of the matrix [165, 166]. A set of experiments with respect to the effects of ordering is contained in [52].

Saad [140] proposed a few variants on the incomplete LU approach for the matrix A, one of which is in fact an incomplete LQ decomposition. In this approach it is not necessary to form the matrix Q explicitly, and it turns out that the lower triangular matrix L can be viewed as the factor of an incomplete Choleski factorization of the matrix $A^T A$. This can be exploited in the preconditioning step, avoiding the use of Q. The second approach was to introduce partial pivoting in ILU, which appears to have some advantages for convection-dominated problems. This approach was further improved by including a threshold technique for fill-in as is done in the ILUT algorithm, see [144, p. 287].

Another major step forward, for important classes of problems, was the introduction of block variants of incomplete factorizations [160, 39, 5], and modified variants of them [5, 109]. It was observed, by Meurant, that these block variants were more successful for discretized 2-dimensional problems than for 3-dimensional problems, unless the '2-dimensional' blocks in the latter case were solved accurately. For discussions and analysis on ordering strategies, in relation to modified block incomplete factorizations, see [110].

8.2 Parallel preconditioners

Parallel preconditioners were discussed as early as with the first appearance of vector and parallel computers. It soon became apparent that the standard ILU-based preconditioners which were just becoming quite popular, were also very sequential in nature and had either to be replaced or implemented differently. The first ideas that were promoted or discussed were based on approximating the LU-solves by means of Neuman expansions in the L and U solves [49] as well as from the start by approximating the inverse of the original matrix by the Neuman series expansion of its inverse [51]. This gave rise to a number of papers articles on "Polynomial preconditioners". The survey paper [141] gives an account of the state of the art toward the end of the 1980s and it can be seen that polynomial preconditioners figured prominently in the article. Another approach - termed "level-scheduling" or "wavefront" approach, was to unravel parallelism from the forward and backward solves. Because of sparsity many equationd can be solved at the same time in several levels during the forward and the backward solves - and a technique known in graph theory as "topological sorting" allows to determine these levels [1, 14, 18, 146, 166, 167].

However, these two methods were soon viewed as having a limited potential. Level-scheduling has limited parallelism and the first and last (smallest) levels were small enough to cause bottlenecks. A number of strategies could be used to improve the situation however. Polynomial preconditioners faced more serious challenges. Their performance relative to existing alternatives was not too convincing, especially for small number of processors. In addition, it is difficult to find a good polynomial in the case of indefinite matrices. Current interest in these techniques has all but vanished. This is a case where good mathematically based methods are not enough to overcome an inherent limitation of a given approach.

Red-black ordering is an obvious approach to improve parallel properties for well-structured problems, but experimental results were disappointing [52] so it was avoided. If carefully done though, they can lead to significant gains in efficiency. Elman and Golub [58] suggested such an approach, in which Red-Black ordering was combined with a reduced system technique. The idea is simply to eliminate the red points, and construct an ILU for the reduced system of black points. Recently, DeLong and Ortega [48] and Saad [143] suggested carrying out a few steps of red-black ordered SOR as a preconditioner for GMRES and Bi-CGSTAB. The key to success in these cases seems to be a combined effect of fast convergence of SOR for red-black ordering, and the ability of the Krylov subspace to remove stagnations in convergence behavior associated with a few isolated eigenvalues of the preconditioned matrix.

Another stream of ideas for deriving parallel preconditioning methods came from domain decomposition-type methods. Such methods were in existence in the Partial Differential Equations (PDE) literature already in a different form, see, e.g., the survey paper [33]. Though Domain decomposition methods were motivated by parallel computing it appeared that the approach could be used with success also for the construction of sequential preconditioners. Domain decomposition has been used for problems that arise from discretization of a PDE over a given domain. The idea is to split the given domain into subdomains, and to solve the discretized PDE's over each subdomain separately. The main problem is to find proper boundary conditions along the interior boundaries of the subdomains. Domain decomposition is used in an iterative fashion and usually the interior boundary conditions are based upon information on the approximate solution of neighboring subdomains that is available from a previous iteration step.

It was shown by Chan and Goovaerts [32] that domain decomposition can actually lead to improved convergence rates, provided the number of domains is not too large. A splitting of the matrix with overlapping subblocks along the diagonal, which can be viewed as a splitting of the domain, if the matrix is associated with a discretized PDE and has been ordered properly, was suggested by Radicati and Robert [130]. They suggested to construct incomplete factorizations for the sub-blocks. These sub-blocks are then applied to corresponding parts of the vectors involved, and some averaging was applied on the overlapping parts. A more sophisticated domain-oriented splitting was suggested in [176], for SSOR and MILU decompositions, with a special treatment for unknowns associated with interfaces between the subdomains.

The isolation of sub-blocks was done by Tang [44] in such a way that the sub-blocks corresponded to subdomains with proper internal boundary conditions. In this case it is necessary to modify the sub-blocks of the original matrix such that the subblocks could be interpreted as the discretizations for subdomains with Dirichlet and Neumann boundary conditions in order to force some smoothness of the approximated solution across boundaries. In [158] this was further improved by requiring also continuity of cross-derivatives of the approximate solution across boundaries. The local fine-tuning of the resulting interpolation formulae for the discretizations was carried out by local Fourier-analysis. It was shown that this approach could lead to impressive reductions in numbers of iterations for convection dominated problems.

Note that Domain Decomposition methods for general sparse linear systems became successful at the same time as the machines for which they were designed (distributed memory, MIMD computers) were gaining importance. Currently, most of the parallel iterative solvers packages utilize essentially DD type preconditioners.

For an overview of parallel preconditioners, and guidelines for their efficient implementation, see [50].

8.3 Multilevel Preconditioners

Methods based on multilevel techniques, such as multigrid, have been popular for solving certain types of PDEs [84]. They are often designed specifically for problems arising from PDEs with regular meshes. Algebraic Multilevel solution methods were developed as an attempt to extend the scope of these methods [136]. Clearly, nothing can prevent the use of these techniques as preconditioners for Krylov solvers. Since the multigrid method is viewed as optimal, its users have often avoided to use an accelerator such as GMRES or BICGSTAB to accelerate it. A recent paper by Oosterlee and Washio [120] did, however, indicate that such a combination could be beneficial and lead to a much more robust solver than a multigrid solver alone.

Recently, a class of preconditioners that tended to close the gap between multilevel methods and Preconditioned Krylov methods drew much attention. It was discovered that a multigrid-inspired ordering can be very effective for discretized diffusion-convection equations, leading in some cases to almost grid-independent speeds of convergence [161, 162], see also [44]. These preconditioners can be viewed also from the angle of ILU factorization combined with a reordering as in the ILUM strategy, see [144, p. 371]. This type of approach can be fairly robust and scale well with problem size, unlike other ILU preconditioners.

In earlier related work, Axelsson and Vassilevski developed a method which was later referred to as AMLI [8, 9] that is based on a set of nested finite element grids. The equations associated with the finer mesh are reduced (approximately) and the process is repeated to a number of levels until the coarsest mesh is reached.

It is interesting to note that currently, this general approach offers an excellent potential for providing a global method that can encompass most of the successful approaches for solving linear systems. By restricting the number of levels to one and performing the factorization accurately, one obtains a direct solver. A standard ILU solver can also be obtained by dropping fill-in.

8.4 Sparse Approximate Inverses

Many researchers and practitioners became aware of an important and damaging phenomenon in ILU techniques. An ILU factorization can be an accurate approximation to the original matrix but it can yield a very ill conditioned factorization [164]. This phenomenon of instability of the LU factors was analyzed in particular by Elman [56]. This weakness of ILU factorizations, coupled with their sequential nature, spurred researchers to consider radical alternatives. The approximate inverse methods which were first proposed in the late 1970s [15] were in this category. It is only with the advent of massive parallel processing that such methods were considered as serious contenders of the now standard ILU methods [81]. A flury of publications followed this work and the work by Kolotlina and Yeremin [102, 103]. To cite just a few, [41, 91] define strategies for determining the best pattern for the inverse, [36, 35, 17] define alternative schemes. While at the beginning, these preconditioning methods were received with much skepticism, it is fair to say that substantial progress has been made and a number of recent papers reported that approximate inverse schemes can often be competitive with ILU factorization methods - even in a sequential environment.

One idea for constructing an approximate inverse is to find a sparse matrix M such that $||AM - I||$ is small for some convenient norm. Kolotilina and Yeremin [103] presented an algorithm in which the inverse was delivered in factored form, which has the advantage that singularity of M is avoided. In [41] an algortithm is presented which uses the 1-norm for the minimization. We also mention Chow and Saad [36], who use GMRES for the minimization of $||AM - I||_F$. Drop-tolerance strategies are applied to limit the amount of fill-in allowed. The approach can also be used to correct explicitly some given implicit approximation, such as a given ILU decomposition.

An elegant approach was suggested by Grote and Huckle [80]. They also attempt to minimize the F-norm, which is equivalent to the Euclidean norm for the errors in the columns m_i of M:

$$||AM - I||_F^2 = \sum_{i=1}^{n} ||Am_i - e_i||_2^2.$$

Based on this observation they derive an algorithm that produces the sparsity pattern for the most error-reducing elements of M. This is done in steps, starting with a diagonal approximation, each steps adds more non-zero entries to M, and the procedure is stopped when the norms are small enough or when memory requirements are violated.

9 Multigrid methods

As was mentioned above, among the earliest preconditioning methods were the simple relaxation schemes since these have, historically, been quite popular techniques. Thus Krylov subspace methods were viewed as methods for accelerating such techniques. Another powerful way of accelerating relaxation techniques is to use multigrid - or multilevel methods. Although we have given little emphasis to these methods in this survey, they are nevertheless important methods which can give rise to very efficient solvers, actually of optimal complexity in some cases. The main observation of multigrid techniques is based on a Fourier analysis of the residual (or error) vector of a sequence of iterates that are generated by a scheme such as Jacobi or Gauss-Seidel. This means that these residual vectors are analyzed in the eigen-basis associated with the iteration matrix M - assuming that M has a complete set of eigenvectors. In the case of Jacobi, the observation is that the components associated with the largest eigenvalues (in the original matrix) will decrease rapidly. However, those associated with the smallest eigenvalues will converge much more slowly. As a result after a few steps, the "high-frequency" components may have converged while the the "low-frequency" components may have made very little progress in comparison. To correct this situation, researchers developed methods that used several grids. The simplest idea is to use two meshes one fine and one that is coarser, where the fine mesh can be viewed as the result of refining the coarse one. The iteration initially takes place on the fine mesh. After a few steps, the residual is *projected* onto the coarse mesh, by some form of restriction. Let A_{2h} be the matrix for the problem on the coarse mesh and r_{2h} this projected residual. The system $A_{2h}\delta = r_{2h}$ is then solved on the coarse mesh by means of a few steps of relaxation. This is called a correction step. The vector δ is then extrapolated into the finer mesh and the result is added as a correction to the iterate on the fine mesh.

An early paper describing essentially such an idea can be traced back to 1935, when Southwell [153] discusses a "group-relaxation" scheme for a two-level setting. It is clear that we do not have to stop at two levels of meshes. Much later Fedorenko [67] described the first true multigrid technique - which employs more than two grids. The idea laid dormant for some time until Achi Brandt published a series of articles, the first of which in 1972 [22]. The paper [23] provided the needed analysis to boost this class of techniques. Many variants of multigrid methods have been developed and the literature is perhaps richer than that of Krylov subspace methods. The excellent "frontiers in applied mathematics" [113] volume published in 1987 listed already 607 references. A number of excellent books have been written on Multigrid. For a quick tutorial see [27]. More complete texts include Hackbusch [84], and Wesseling [180]. The volume [113] contains an excellent collection of articles on multigrid and algebraic multigrid.

It is often asked what is the best method to use: Preconditioned Krylov subspace methods or a Multigrid approach? Users of iterative techniques are really split in two camps: those who use exclusively multigrid methods and those who use exclusively (Preconditioned) Krylov subspace methods. Combination of the two methods have been advocated however, see Section 8.3, but this is an exception rather than a rule. When multigrid techniques work, they can be extremely efficient - far more so than Preconditioned Krylov methods. However, their efficiency relies essentially on the inter-level restriction, and prolongation operators, the choice of which will vary from one application to the next. Such efficiencies can be achieved for regular meshes and for smooth elliptic PDEs. Standard multigrid methods cannot be applied without the existence of an underlying mesh - hence its major limitation.

This led to the development of Algebraic MultiGrid (AMG) initiated by by Ruge and Stuben [136]. AMG was defined for algebraic systems - in the same manner as general sparse linear systems solvers - by defining restriction and prolongation operators algebraically. The overall success of AMG, which is derived based on an underlying PDE problem, has been somewhat limited.

10 Outlook

It is rather difficult to predict what the future will bring in the area of iterative methods. However, it is almost certain that their useage will increase substantially in the applications areas. This is partly due to the impact of parallel architectures. Direct methods are more complex to implement in parallel than are iterative methods. Also it is clear that problem sizes are increasing to the point of making direct solvers exceedingly expensive - both in terms of memory and arithmetic costs. One ray of hope for those problems that are hard to solve by iterative techniques, is to combine techniques from direct and iterative solution technologies. As the communities from direct and iterative solvers are getting to learn each others tricks, the distinction between the two methodologies is getting to be blurred and this results in better, mor robust, methods. Indeed, if memory is the only difficulties with direct solvers, it may be possible to find preconditioners that are far more accurate than current ones - but which use moderate amounts of memory.

References

[1] E. C. Anderson and Y. Saad. Solving sparse triangular systems on parallel computers. *International Journal of High Speed Computing*, 1:73–96, 1989.

[2] W. E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.

[3] O. Axelsson. A generalized SSOR method. *BIT*, 12:443–467, 1972.

[4] O. Axelsson. Solution of linear systems of equations: iterative methods. In V. A. Barker, editor, *Sparse Matrix Techniques*, pages 1–51, Berlin, 1977. Copenhagen 1976, Springer Verlag.

[5] O. Axelsson. A general incomplete block-factorization method. *Linear Alg. Appl.*, 74:179–190, 1986.

[6] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, New York, 1994.

[7] O. Axelsson and G. Lindskog. On the eigenvalue distribution of a class of preconditioning methods. *Numer. Math.*, 48:479–498, 1986.

[8] O. Axelsson and P. Vassilevski. Algebraic multilevel preconditioning methods. I. *Numer. Math.*, 56:157–177, 1989.

[9] O. Axelsson and P. Vassilevski. Algebraic multilevel preconditioning methods. II. *SIAM J. Numer. Anal.*, 27(6):1569–1590, December 1990.

[10] O. Axelsson and P. S. Vassilevski. A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. *SIAM J. Matrix Anal. Applic.*, 12(4):625–644, 1991.

[11] R. E. Bank and T. F. Chan. An analysis of the composite step biconjugate gradient method. *Numer. Math.*, 66:295–319, 1993.

[12] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the solution of linear systems: building blocks for iterative methods*. SIAM, Philadelphia, 1994.

[13] T. Barth and T. Manteuffel. Variable metric conjugate gradient methods. In *Advances in Numerical Methods for Large Sparse Sets of Linear Equations, Number 10, Matrix Analysis and Parallel Computing, PCG 94*, pages 165–188, Keio University, Yokohama, Japan, 1994.

[14] D. Baxter, J. Saltz, M. H. Schultz, S. C. Eisenstat, and K. Crowley. An experimental study of methods for parallel preconditioned Krylov methods. In *Proceedings of the 1988 Hypercube Multiprocessors Conference*, pages 1698–1711, Pasadena, CA, Jan. 1988.

[15] M. W. Benson and P. O. Frederickson. Iterative solution of large sparse linear systems arising in certain multidimensional approximation problems. *Utilitas Math.*, 22:127–140, 1982.

[16] M. Benzi, J. Marín, and M. Tůma. A two-level parallel preconditioner based on sparse approximate inverses. In D. R. Kincaid and A. C. Elster, editors, *Iterative Methods in Scientific Computing, II*, pages xx–xx+10. IMACS, 1999.

[17] M. Benzi, C. D. Meyer, and M. Tuma. A sparse approximate inverse preconditioner for the conjugate gradient method. *SIAM Journal on Scientific Computing*, 17:1135–1149, 1996.

[18] H. Berryman, J. Saltz, W. Gropp, and R. Mirchandaney. Krylov methods preconditioned with incompletely factored matrices on the CM-2. *Journal of Parallel and Distributed Computing*, 8:186–190, 1990.

[19] G. Birkhoff. Solving elliptic problems: 1930-1980. In M. H. Schultz, editor, *Elliptic Problem Solver*, pages 17–38, New York, 1981. Academic Press.

[20] A. Bjork and T. Elfving. Accelerated projection methods for computing pseudo-inverse solutions of systems of linear equations. *BIT*, 19:145–163, 1979.

[21] R. Bramley and A. Sameh. Row projection methods for large nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13:168–193, 1992.

[22] A. Brandt. Multi-level adaptive technique (MLAT) for fast numerical solutions to boundary problems. In H. Cabannes and R. Temmam, editors, *Proc. 3rd Int. Conf. Numer. Methods in Fluid Mechanics, Paris, 1972*, pages 92–89, Berlin, 1973. Springer Verlag.

[23] A. Brandt. Multi-level adaptive solutions to boundary value problems. *Mathematics of Computations*, 31:333–390, 1977.

[24] C. Brezinski. *Projection methods for systems of equations.* North-Holland, Amsterdam, 1997.

[25] C. Brezinski, M. Redivo-Zaglia, and H. Sadok. Avoiding breakdown and near-breakdown in Lanczos type algorithms. *Numerical Algorithms*, 1:261–284, 1991.

[26] C. Brezinski, M. Redivo-Zaglia, and H. Sadok. A breakdown-free lanczos' type algorithm for solving linear systems. *Numer. Math.*, 63:29–38, 1992.

[27] W. L. Briggs. *A multigrid tutorial.* SIAM, Philadelphia, PA, 1987.

[28] P. N. Brown. A theoretical comparison of the arnoldi and GMRES algorithms. *SIAM Journal on Scientific and Statistical Computing*, 12:58–78, 1991.

[29] G. C. Broyden. A new method of solving nonlinear simultaneous equations. *Comput. J.*, 12:94–99, 1969.

[30] N. I. Buleev. A numerical method for the solution of two-dimensional and three-dimensional equations of diffusion. *Math. Sb*, 51:227–238, 1960. (in Russian).

[31] T. F. Chan, E. Gallopoulos, V. Simoncini, T. Szeto, and C. H. Tong. A quasi-minimal residual variant of the Bi-CGSTAB algorithm for nonsymmetric systems. *SIAM J. Sci. Comput.*, 15:338–347, 1994.

[32] T. F. Chan and D. Goovaerts. A note on the efficiency of domain decomposed incomplete factorizations. *SIAM J. Sci. Statist. Comput.*, 11:794–803, 1990.

[33] T. F. Chan and D. Resasco. A framework for the analysis of domain decomposition preconditioners. In *Proc. First Internat. Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 217–230, Philadelphia, PA, 1988. Society for Industrial and Applied Mathematics.

[34] D. Chazan and M. Miranker. Chaotic relaxation. *Linear Algebra and its Applications*, 2:199–222, 1969.

[35] E. Chow and Y. Saad. Approximate inverse techniques for block-partitioned matrices. *SIAM Journal on Scientific Computing*, 18:1657–1675, 1997.

[36] E. Chow and Y. Saad. Approximate inverse preconditioners via sparse-sparse iterations. *SIAM Journal on Scientific Computing*, 19:995–1023, 1998.

[37] G. Cimmino. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *Ric. Sci. Progr. tecn. econom. naz.*, 9:326–333, 1939.

[38] P. Concus and G. H. Golub. A generalized Conjugate Gradient method for nonsymmetric systems of linear equations. Technical Report STAN-CS-76-535, Stanford University, Stanford, CA, 1976.

[39] P. Concus, G. H. Golub, and G. Meurant. Block preconditioning for the conjugate gradient method. *SIAM J. Sci. Statist. Comput.*, 6:220–252, 1985.

[40] P. Concus, G. H. Golub, and D. P. O'Leary. A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations. In James R. Bunch and Donald J. Rose, editors, *Sparse Matrix Computations*, pages 309–332, New York, 1976. Academic Press.

[41] J. D. F. Cosgrove, J. C. Díaz, and A. Griewank. Approximate inverse preconditioning for sparse linear systems. *Int. J. Comp. Math.*, 44:91–110, 1992.

[42] J. Cullum and A. Greenbaum. Relations between Galerkin and norm-minimizing iterative methods for solving linear systems. *SIAM J. Matrix Anal. Applic.*, 17:223–247, 1996.

[43] J. W. Daniel. The conjugate gradient method for linear and nonlinear operator equations. *SIAM J. Numer. Anal.*, 4:10–26, 1967.

[44] E. F. D'Azevedo, F. A. Forsyth, and W. P. Tang. Ordering methods for preconditioned conjugate gradient methods applied to unstructured grid problems. *SIAM Journal on Matrix Analysis and Applications*, 13:944–961, 1992.

[45] E. F. D'Azevedo, F. A. Forsyth, and W. P. Tang. Towards a cost effective ILU preconditioner with high level fill. *BIT*, 31:442–463, 1992.

[46] C. de Boor and J. R. Rice. Extremal polynomials with applications to Richardson iteration for indefinite systems. Technical Report 2107, M. R. C., 1980.

[47] R. Dedekind. Gauss in seiner Vorlesung über die Methode der kleisten Quadrate. *Gesammelte Math. Werke*, 2:293–306, 1901.

[48] M. A. DeLong and J. M. Ortega. SOR as a preconditioner. Technical Report CS-94-43, Department of Computer Science, University of Virginia, Charlottesville, VA, 1994.

[49] H. A. Van der Vorst. A vectorizable version of some ICCG methods. *SIAM J. Stat. and Sci. Comput.*, 3:350–356, 1982.

[50] J. J. Dongarra, I. S. Duff, D. C. Sorensen, and H. A. van der Vorst. *Numerical Linear Algebra for High-Performance Computers*. SIAM, Philadelphia, PA, 1998.

[51] P. F. Dubois, A. Greenbaum, and G. H. Rodrigue. Approximating the inverse of a matrix for use on iterative algorithms on vectors processors. *Computing*, 22:257–268, 1979.

[52] I. S. Duff and G. A. Meurant. The effect of reordering on preconditioned conjugate gradients. *BIT*, 29:635–657, 1989.

[53] T. Dupont, R. Kendall, and H. Rachford. An approximate factorization procedure for solving self-adjoint elliptic difference equations. *SIAM J. Numer. Anal.*, 5:559–573, 1968.

[54] T. Eirola and O. Nevanlinna. Accelerating with rank-one updates. *Linear Alg. Appl.*, 121:511–520, 1989.

[55] S. C. Eisenstat, H. C. Elman, and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis*, 20:345–357, 1983.

[56] H. C. Elman. A stability analysis of incomplete LU factorizations. *Mathematics of Computations*, 47:191–217, 1986.

[57] H. C. Elman. Relaxed and stabilized incomplete factorizations for non-self-adjoint linear systems. *BIT*, 29:890–915, 1989.

[58] H. C. Elman and G. H. Golub. Line iterative methods for cyclically reduced discrete convection-diffusion problems. *SIAM J. Sci. Statist. Comput.*, 13:339–363, 1992.

[59] H. C. Elman, Y. Saad, and P. Saylor. A hybrid Chebyshev Krylov subspace algorithm for solving non-symmetric systems of linear equations. *SIAM Journal on Scientific and Statistical Computing*, 7:840–855, 1986.

[60] M. Engeli. Overrelaxation and related methods. In *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, pages 79–107, Basel-Stuttgart, 1959. Institute of Applied Mathematics, Zurich, Birkhauser Verlag.

[61] M. Engeli, T. Ginsburg, H. Rutishauser, and E. Stiefel. *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*. Birkhäuser, Basel/Stuttgart, 1959.

[62] D. J. Evans. The use of pre-conditioning in iterative methods for solving linear equations with symmetric positive definite matrices. *J. Inst. Maths.Applics.*, 4:295–314, 1968.

[63] V. Faber and T. Manteuffel. Necessary and sufficient conditions for the existence of a conjugate gradient method. *SIAM Journal on Numerical Analysis*, 21:352–361, 1984.

[64] B. Fischer and L. Reichel. A stable Richardson iteration method for complex linear systems. *Numerische Mathematik*, 54:225–241, 1988.

[65] D. A. Flanders and G. Shortley. Numerical determination of fundamental modes. *J. Appl. Phy.*, 21:1328–1322, 1950.

[66] R. Fletcher. Conjugate gradient methods for indefinite systems. In G. A. Watson, editor, *Proceedings of the Dundee Biennal Conference on Numerical Analysis 1974*, pages 73–89, New York, 1975. Springer Verlag.

[67] R. P. Foderenko. On the speed of convergence of an iteration process. *USSR Comput. Math. and Math. Phys.*, 4:227–235, 1964.

[68] D. R. Fokkema, G. L. G. Sleijpen, and H. A. van der Vorst. Generalized conjugate gradient squared. *J. Comp. and Appl. Math.*, 71:125–146, 1994.

[69] S. Frankel. Convergence rates of iterative treatments of parrial differential equations. *MTAC*, pages 65–75, 1950.

[70] R. W. Freund. A Transpose-Free Quasi-Minimal Residual algorithm for non-Hermitian linear systems. *SIAM Journal on Scientific Computing*, 14(2):470–482, 1993.

[71] R. W. Freund and N. M. Nachtigal. An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices, part 2. Technical Report 90.46, RIACS, NASA Ames Research Center, 1990.

[72] R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik*, 60:315–339, 1991.

[73] K. Gallivan, A. Sameh, and Z. Zlatev. A parallel hybrid sparse linear system solver. *Computing Systems in Engineering*, 1(2-4):183–195, June 1990.

[74] N. Gastinel. *Analyse Numerique Lineaire*. Hermann, Paris, 1966.

[75] M. K. Gavurin. Application of the best approximation polynomial to improving the convergence of iterative methods. *Uspehi Mat. Nauk.*, 5(3):156–160, 1950. In Russian.

[76] G. H. Golub and D. P. O'Leary. Some history of the conjugate gradient and Lanczos algorithms: 1948-1976. *SIAM review*, 31:50–102, 1989.

[77] G. H. Golub and R. S. Varga. Chebyshev semi iterative methods successive overrelaxation iterative methods and second order Richardson iterative methods. *Numerische Mathematik*, 3:147–168, 1961.

[78] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelpha, PA, 1997.

[79] A. Greenbaum and Z. Stakos. Matrices that generate the same Krylov varieties. In G. Golub, M. Luskin, and A. Greenbaum, editors, *Recent Advances in Iterative Methods, IMA Volumes in Mathematics and Its Applications*, volume 60, pages 95–119, New York, 1994. Springer Verlag.

[80] M. Grote and T. Huckle. Effective parallel preconditioning with sparse approximate inverses. In D. H. Bailey, editor, *Proceedings of the Seventh SIAM Conference on Parallel Processing for Scientific Computing*, pages 466–471, Philadelphia, 1995. SIAM.

[81] M. Grote and H. D. Simon. Parallel preconditioning and approximate inverses on the connection machine. In R. F. Sincovec, D. E. Keyes, L. R. Petzold, and D. A. Reed, editors, *Parallel Processing for Scientific Computing – vol. 2*, pages 519–523. SIAM, 1992.

[82] I. Gustafsson. A class of first order factorization methods. *BIT*, 18:142–156, 1978.

[83] M. H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms, Part I. *SIAM J. Matrix Anal. Applic.*, 13:594–639, 1992.

[84] W. Hackbusch. *Multi-Grid methods and applications*. Springer Verlag, New York, 1985.

[85] W. Hackbusch. *Iterative solution of large linear systems of equations*. Springer Verlag, New York, 1994.

[86] A. L. Hageman and D. M. Young. *Applied Iterative Methods*. Academic Press, New York, 1981.

[87] R. M. Hayes. Iterative methods of solving linear systems on hilbert space. *National Bureau of Standards, Appl. Math. Ser.*, 39:71–103, 1954.

[88] M. R. Hestenes. The Conjugate-Gradient method for solving linear systems. In *Sympos. Appl. Math.*, Numerical Analysis, Volume VI, pages 83–102, New York, 1956. McGrwa-Hill.

[89] M. R. Hestenes and E. L. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, Section B, 49:409–436, 1952.

[90] A. S. Householder. *Theory of Matrices in Numerical Analysis*. Blaisdell Pub. Co., Johnson, CO, 1964.

[91] T. Huckle and M. Grote. A new approach to parallel preconditioning with sparse approximate inverses. Technical Report SCCM-94-03, Stanford University, Scientific Computing and Computational Mathematics Program, Stanford, California, 1994.

[92] J. W. Watts III. A conjugate gradient truncated direct method for the iterative solution of the reservoir simulation pressure equation. *Society of Petroleum Engineers Journal*, 21:345–353, 1981.

[93] K. C. Jea and D. M. Young. Generalized conjugate-gradient acceleration of nonsymmetrizable iterative methods. *Linear Alg. Appl.*, 34:159–194, 1980.

[94] O. G. Johnson, C. A. Micchelli, and G. Paul. Polynomial preconditionings for conjugate gradient calculations. *SIAM Journal on Numerical Analysis*, 20:362–376, 1983.

[95] W. D. Joubert. *Generalized conjugate gradient and Lanczos methods for the solution of nonsymmetric systems of linear equations.* PhD thesis, University of Texas at Austin, Center for Numerical Analysis, Austin, TX, 1990.

[96] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bulletin international de l'Academie polonaise des Sciences et lettres*, pages 355–357, 1939.

[97] W. Kahan. *Gauss-Seidel methods of solving large systems of linear equations.* PhD thesis, University of Toronto, 1958.

[98] C. Kamath and A. Sameh. A projection method for solving nonsymmetric linear systems on multiprocessors. Technical Report 611, CSRD, university of Illinois, Urbana, IL, 1986.

[99] S. Kaniel. Estimates for some computational techniques in linear algebra. *Math. Comp.*, 20:369–378, 1966.

[100] L. Kantorovitch. On an efficient method for solving optimization problems for quadratic functionals. *DAN SSSR*, 48(7):7–8, 1945. In Russian.

[101] W. Karush. Convergence of a method of solving linear problems. *Proc. Amer. Math. Soc.*, 3:839–851, 1952.

[102] L. Yu Kolotilina and A. Yu Yeremin. On a family of two-level preconditionings of the incomplete block factorization type. *Soviet Journal of Numerical Analysis and Mathematical Modeling*, 1:293–320, 1986.

[103] L. Yu. Kolotilina and A. Yu. Yeremin. Factorized sparse approximate inverse preconditionings I.Theory. *SIAM Journal on Matrix Analysis and Applications*, 14:45–58, 1993.

[104] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45:255–282, 1950.

[105] C. Lanczos. Chebyshev polynomials in the solution of large-scale linear systems. In *Proceedings of the ACM*, pages 124–133, 1952.

[106] C. Lanczos. Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49:33–53, 1952.

[107] H. Liebmann. Die angenäherte harmonischer Funktionen und konformer Abbildungen (nach Ideen von Boltzmann und Jacobi). *S. B. Math. Nat. Kl. Bayerischen Akad. Wiss. München*, pages 385–416, 1918.

[108] D. G. Luenberger. Hyperbolic pairs in the method of conjugate gradients. *SIAM J. Appl. Math.*, 17:1263–1267, 1979.

[109] M. M. Magolu. Modified block-approximate factorization strategies. *Numer. Math.*, 61:91–110, 1992.

[110] M. M. Magolu. Ordering strategies for modified block incomplete factorizations. *SIAM J. Sci. Comput.*, 16:378–399, 1995.

[111] T. A. Manteuffel. The Tchebychev iteration for nonsymmetric linear systems. *Numerische Mathematik*, 28:307–327, 1977.

[112] T. A. Manteuffel. Adaptive procedure for estimation of parameter for the nonsymmetric Tchebychev iteration. *Numerische Mathematik*, 28:187–208, 1978.

[113] S. F. Mc Cormick, editor. *Multigrid methods*, Philadelphia, PA, 1987. SIAM.

[114] J. A. Meijerink and H. A. van der Vorst. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. *Mathematics of Computations*, 31(137):148–162, 1977.

[115] J. A. Meijerink and H. A. van der Vorst. Guidelines for the usage of incomplete decompositions in solving sets of linear equations as they occur in practical problems. *J.of Comp. Physics*, 44:134–155, 1981.

[116] J. C. Miellou. Algorithmes de relaxation chaotiques a retard. *RAIRO*, R-1:55–82, 1975.

[117] P. A. Nekrasov. Determining unknowns using the least squares method when the number of unknowns is large. *Mat. Sb.*, 12:189–204, 1884. (in Russian).

[118] Y. Notay. DRIC: a dynamic version of the RIC method. *Num. Lin. Alg. Appl.*, 1:511–532, 1994.

[119] T. A. Oliphant. An extrapolation process for solving linear systems. *Quart. Appl. Math.*, 20:257–267, 1962.

[120] C. W. Oosterlee and T. Washio. An evaluation of parallel multigrid as a solver and a preconditioner for singularly perturbed problems. *SIAM Journal on Scientific and Statistical Computing*, 19:87–110, 1991.

[121] O. Osterby and Z. Zlatev. *Direct Methods for Sparse Matrices*. Springer Verlag, New York, 1983.

[122] A. M. Ostrowski. Uber die Determinanten mit überwiegender Hauptdiagonale. *Comment. Math. Helv.*, 10:69–96, 1937.

[123] A. M. Ostrowski. On the linear iteration procedures for symmetric matrices. *Rend. Mat. e Appl.*, 14:140–163, 1954.

[124] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.

[125] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.*, 8:43–71, 1982.

[126] B. N. Parlett. Reduction to tridiagonal form and minimal realizations. *SIAM J. Matrix Anal. Applic.*, pages 567–593, 1992.

[127] B. N. Parlett, D. R. Taylor, and Z. S. Liu. A look-ahead Lanczos algorithm for nonsymmetric matrices. *Mathematics of Computation*, 44:105–124, 1985.

[128] D. Peaceman and H. Rachford. The numerical solution of elliptic and parabolic differential equations. *Journal of SIAM*, 3:28–41, 1955.

[129] C. Pommerell. *Solution of large unsymmetric systems of linear equations*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 1992.

[130] G. Radicati di Brozolo and Y. Robert. Parallel conjugate gradient-like algorithms for solving sparse non-symmetric systems on a vector multiprocessor. *Parallel Computing*, 11:223–239, 1989.

[131] E. Reich. On the convergence of the classical iterative method of solving linear simultaneous equations. *Ann. Math. Statist.*, 20:448–451, 1949.

[132] J. K. Reid. On the method of conjugate gradients for the solution of large sparse systems of linear equations. In J. K. Reid, editor, *Large Sparse Sets of Linear Equations*, pages 231–254. Academic Press, 1971.

[133] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations with an application to the stresses to a masonry dam. *Philos. Trans. Roy. Soc. London ser. A*, 210:307–357, 1910.

[134] F. Robert. Contraction en norme vectorielle: convergence d'iterations chaotiques. *Linear Algebra and its Applications*, 13:19–35, 1976.

[135] F. Robert, M. Charnay, and F. Musy. Iterations chaotiques serie-parallele pour les equations non lineaires de point fixe. *Aplikace Mathematiky*, 20:1–38, 1975.

[136] A. Ruge and K. Stüben. Algebraic multigrid. In S. McCormick, editor, *Multigrid Methods*, volume 3 of *Frontiers in Applied Mathematics*, chapter 4. SIAM, 1987.

[137] H. Rutishauser. Theory of gradient methods. In *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, pages 24–49, Basel-Stuttgart, 1959. Institute of Applied Mathematics, Zurich, Birkhauser Verlag.

[138] Y. Saad. Practical use of polynomial preconditionings for the conjugate gradient method. *SIAM Journal on Scientific and Statistical Computing*, 6:865–881, 1985.

[139] Y. Saad. Least squares polynomials in the complex plane and their use for solving sparse nonsymmetric linear systems. *SIAM Journal on Numerical Analysis*, 24:155–169, 1987.

[140] Y. Saad. Preconditioning techniques for indefinite and nonsymmetric linear systems. *Journal of Computational and Applied Mathematics*, 24:89–105, 1988.

[141] Y. Saad. Krylov subspace methods on supercomputers. *SIAM Journal on Scientific and Statistical Computing*, 10:1200–1232, 1989.

[142] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM Journal on Scientific and Statistical Computing*, 14:461–469, 1993.

[143] Y. Saad. Highly parallel preconditioners for general sparse matrices. In G. Golub, M. Luskin, and A. Greenbaum, editors, *Recent Advances in Iterative Methods, IMA Volumes in Mathematics and Its Applications*, volume 60, pages 165–199, New York, 1994. Springer Verlag.

[144] Y. Saad. *Iterative Methods for Sparse Linear Systems.* PWS publishing, New York, 1996.

[145] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7:856–869, 1986.

[146] J. Saltz, , R. Mirchandaney, and K. Crowley. Run-time paralellization and scheduling of loops. *IEEE Transactions on Computers*, 40:603–612, 1991.

[147] P. E. Saylor and D. C. Smolarski. Computing the roots of complex orthogonal kernel polynomials. *SIAM Journal on Scientific and Statistical Computing*, 9:1–13, 1988.

[148] G. H. Shortley. Use of Tschebyscheff-polynomial operators in the solution of boundary value problems. *J. Appl. Phy.*, 24:392–396, 1953.

[149] H. D. Simon. Incomplete LU preconditioners for conjugate gradient type iterative methods. In *Proceedings of the SPE 1985 reservoir simulation symposium*, pages 302–306, Dallas, TX, 1988. Society of Petroleum Engineers of AIME. Paper number 13533.

[150] Horst D. Simon. Direct sparse matrix methods. In James C. Almond and David M. Young, editors, *Modern Numerical Algorithms for Supercomputers*, pages 325–444, Austin, 1989. The University of Texas at Austin, Center for High Performance Computing.

[151] G. L. G. Sleijpen and D. R. Fokkema. BICGSTAB(ℓ) for linear equations involving unsymmetric matrices with complex spectrum. *ETNA*, 1:11–32, 1993.

[152] P. Sonneveld. CGS: a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 10:36–52, 1989.

[153] R. V. Soutwell. Stress calclation in frameworks by the method of systematic relaxation of constraints. *Proc. Roy. Soc. London, Ser. A*, 151:56–95, 1935.

[154] W. Spakman and G. Nolet. Imaging algorithms, accuracy and resolution in delay time tomography. In N. J. Vlaar, N. Nolet, M. Wortel, and S. Cloetingh, editors, *Mathematical Geophysics: A survey of recent developments in seismology and geodynamics*, pages 155–188. Reidel, Dordrecht, 1987.

[155] P. Stein and R. L. Rosenberg. On the solutiion of linear simultaneous equations by iteration. *J. London Math. Soc.*, 23:111–118, 1948.

[156] E. L. Stiefel. Kernel polynomials in linear algebra and their applications. *U. S. National Bureau of Standards, Applied Mathematics Series*, 49:1–24, 1958.

[157] H. L. Stone. Iterative solution of implicit approximations of multidimensional partial differential equations. *SIAM Journal on Numerical Analysis*, 5:530–558, 1968.

[158] K. H. Tan. *Local coupling in domain decomposition.* PhD thesis, Utrecht University, Utrecht, the Netherlands, 1995.

[159] K. Tanabe. Projection method for solving a singular system of linear equations and its applications. *Numer. Math.*, 17:203–214, 1971.

[160] R. R. Underwood. An approximate factorization procedure based on the block Cholesky factorization and its use with the conjugate gradient method. Technical Report Tech Report, General Electric, San Jose, CA, 1976.

[161] A. van der Ploeg. *Preconditioning for sparse matrices with applications.* PhD thesis, Department of Mathematics, University of Groningen, 1995.

[162] A. van der Ploeg, E. F. F. Botta, and F. W. Wubs. Nested grids ILU-decomposition (NGILU). *J. Comp. and Appl. Math.*, 66:515–526, 1996.

[163] A. van der Sluis and H. A. van der Vorst. The rate of convergence of conjugate gradients. *Numerische Mathematik*, 48:543–560, 1986.

[164] H. A. van der Vorst. Iterative solution methods for certain sparse linear systems with a non-symmetric matrix arising from PDE-problems. *J. Comp. Phys.*, 44:1–19, 1981.

[165] H. A. van der Vorst. Large tridiagonal and block tridiagonal linear systems on vector and parallel computers. *Parallel Computing*, 5:45–54, 1987.

[166] H. A. van der Vorst. High performance preconditioning. *SIAM J. Sci. Statist. Comput.*, 10:1174–1185, 1989.

[167] H. A. van der Vorst. ICCG and related methods for 3-D problems on vector computers. In D. Truhlar, editor, *Workshop on Practical Iterative Methods for Large Scale Computations, Minneapolis MN. , Oct. 23-25 1988*. Computer Physics Communication, vol. 53, 1989.

[168] H. A. van der Vorst. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13:631–644, 1992.

[169] H. A. van der Vorst and C. Vuik. GMRESR: A family of nested GMRES methods. *Num. Lin. Alg. Appl.*, 1:369–386, 1994.

[170] R. S. Varga. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1962.

[171] R.S. Varga. Factorization and normalized iterative methods. In R.E. Langer, editor, *Boundary problems in differential equations*, pages 121–142. Univ. of Wisconsin Press, Madison, 1960.

[172] P. K. W. Vinsome. ORTHOMIN: an iterative method for solving sparse sets of simultaneous linear equations. In *Proc.Fourth Symposium on Reservoir Simulation*, pages 149–159. Society of Petroleum Engineers of AIME, 1976.

[173] C. Vuik and H. A. van der Vorst. A comparison of some GMRES-like methods. *Linear Alg. Appl.*, 160:131–162, 1992.

[174] E. L. Wachspress. CURE: A generalized two-space-dimension multigroup coding for the IBM-704. Technical Report Report KAPL-1724, Knolls Atomic Power Laboratory, Schenectady, New York, 1957.

[175] E. L. Wachspress. *Iterative Solution of Elliptic Systems and Applications to the Neutron Equations of Reactor Physics*. Prentice Hall, Englewood Cliffs, NJ, 1966.

[176] T. Washio and K. Hayami. Parallel block preconditioning based on SSOR and MILU. *Num. Lin. Alg. Appl.*, 1:533–553, 1994.

[177] R. Weiss. Error-minimizing Krylov subspace methods. *SIAM J. Sci. Comput.*, 15:511–527, 1994.

[178] R. Weiss. A theoretical overview of Krylov subspace methods. In W. Schönauer and R. Weiss, editors, *Special Issue on Iterative Methods for Linear Systems*, pages 33–56. Applied Numerical Methods, 1995.

[179] R. Weiss. *Parameter-Free Iterative Linear Solvers*. Akademie Verlag, Berlin, 1996.

[180] P. Wesseling. *An Introduction to Multigrid Methods*. John Wiley and Sons, Chichester, 1992.

[181] O. Widlund. A Lanczos method for a class of nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, 15:801–812, 1978.

[182] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.

[183] J. H. Wilkinson and C. Reinsch. *Handbook for automatic computation, Vol. II, Linear Algebra*. Springer Verlag, New York, 1971.

[184] D. M. Young. *Iterative Methods for Solving Partial Differential Equations of Elliptic Type*. PhD thesis, Harvadd University, 1950.

[185] D. M. Young. On Richardson's method for solving linear systems with positive definite matrices. *J. Math. Phys.*, 32:243–255, 1954.

[186] D. M. Young. *Iterative Solution of Large Linear Systems*. Academic Press, New York, 1971.

[187] D. P. Young, R. G. Melvin, F. T. Johnson, J. E. Bussoletti, L. B. Wigton, and S. S. Samant. Application of sparse matrix solvers as effective preconditioners. *SIAM Journal on Scientific and Statistical Computing*, 10:1186–1199, 1989.

[188] Shao-Liang Zhang. GPBi-CG: Generalized product-type methods based on Bi-CG for solving nonsymmetric linear systems. *SIAM J. Sci. Comput.*, 18:537–551, 1997.

[189] Z. Zlatev. Use of iterative refinement in the solution of sparse linear systems. *SIAM Journal on Numerical Analysis*, 19:381–399, 1982.