Advances in Engineering Software 000 (2015) 1–10

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Parallel alternating iterative algorithms with and without overlapping on multicore architectures

Héctor Migallón [a,*], Violeta Migallón [b], José Penadés [b]

[a] Department of Physics and Computer Architectures, University Miguel Hernández, E-03202 Elche, Alicante, Spain
[b] Department of Computer Science and Artificial Intelligence, University of Alicante, E-03071, Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Block two-stage methods
Alternating iterations
Overlapping
Parallel computing
Shared memory
Distributed memory
Laplace's equation
Markov chains

ABSTRACT

We consider the problem of solving large sparse linear systems where the coefficient matrix is possibly singular but the equations are consistent. Block two-stage methods in which the inner iterations are performed using alternating methods are studied. These methods are ideal for parallel processing and provide a very general setting to study parallel block methods including overlapping. Convergence properties of these methods are established when the matrix in question is either M-matrix or symmetric matrix. Different parallel versions of these methods and implementation strategies, with and without overlapping blocks, are explored. The reported experiments show the behavior and effectiveness of the designed parallel algorithms by exploiting the benefits of shared memory inside the nodes of current SMP supercomputers.

1. Introduction

Consider the problem of solving a linear system

$$Ax = b, \tag{1}$$

where A is an $n \times n$ matrix such that b is in $\mathcal{R}(A)$, the range of A.

Given a splitting $A = M - N$ (M nonsingular), a classical iterative method produces the following iteration scheme

$$Mx^{(l+1)} = Nx^{(l)} + b, \quad l = 0, 1, \ldots. \tag{2}$$

On the other hand, when the linear systems (2) are not solved exactly, but rather their solutions approximated by iterative methods, we are in the presence of a two-stage method (see e.g. [1,2]). That is, consider the splitting $M = F - G$ and perform, at each outer step l, $q(l)$ inner iterations of the iterative procedure induced by this splitting. Thus, the two-stage method can be written as follows

$$x^{(l+1)} = (F^{-1}G)^{q(l)}x^{(l)} + \sum_{j=0}^{q(l)-1} (F^{-1}G)^j F^{-1}(Nx^{(l)} + b), \quad l = 0, 1, \ldots. \tag{3}$$

Without loss of generality, let us assume that the matrix A has the form

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ A_{21} & A_{22} & \cdots & A_{2r} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{bmatrix}, \tag{4}$$

with the diagonal blocks A_{ii} being square of order n_i, $1 \le i \le r$, $\sum_{i=1}^{r} n_i = n$. Let $A = M - N$ be a splitting of A such that M is a block diagonal matrix $M = \mathrm{Diag}\{M_1, \ldots, M_i, \ldots, M_r\}$, and let us consider the splittings $M_i = B_i - C_i$, $M_i = F_i - G_i$, $1 \le i \le r$. Let $M = P_i - Q_i = R_i - S_i$ be splittings of the matrix M such that

$$P_i = \mathrm{Diag}\{I, \ldots, B_i, \ldots, I\}, \; R_i = \mathrm{Diag}\{I, \ldots, F_i, \ldots, I\}. \tag{5}$$

Moreover, let the $n \times n$ diagonal matrices E_i have ones in the entries corresponding to the diagonal block M_i and zero otherwise. In order to approximate the linear systems (2) we perform, at each outer iteration l, $q(i, l)$ inner iterations of the following alternating iterative scheme:

$$z_i^{(k+\frac{1}{2})} = P_i^{-1}Q_i z_i^{(k)} + P_i^{-1}(Nx^{(l)} + b),$$
$$z_i^{(k+1)} = R_i^{-1}S_i z_i^{(k+\frac{1}{2})} + R_i^{-1}(Nx^{(l)} + b), \quad k = 0, 1, \ldots, q(i, l) - 1,$$

with $z^{(0)} = x^{(l)}$, or equivalently

$$z_i^{(k+1)} = R_i^{-1}S_i P_i^{-1}Q_i z^{(k)} + R_i^{-1}(S_i P_i^{-1} + I)(Nx^{(l)} + b), \quad k = 0, 1, \ldots, q(i, l) - 1.$$

* Corresponding author. Tel.: +34966658390; fax: +34966658814.
E-mail address: hmigallon@umh.es (H. Migallón).

http://dx.doi.org/10.1016/j.advengsoft.2015.10.012
0965-9978/

Thus, for $l = 0, 1, \ldots$, the alternating two-stage method can be written as follows,

$$x^{(l+1)} = \sum_{i=1}^{r} E_i z_i^{q(i,l)} = \sum_{i=1}^{r} E_i [(R_i^{-1} S_i P_i^{-1} Q_i)^{q(i,l)} x^{(l)} + \sum_{j=0}^{q(i,l)-1} (R_i^{-1} S_i P_i^{-1} Q_i)^j R_i^{-1} (S_i P_i^{-1} + I)(N x^{(l)} + b)]. \quad (6)$$

Note that the global iteration matrix of the alternating two-stage iterative method (6) can be written as $T^{(l)} = \sum_{i=1}^{r} E_i T_i^{(l)}$, with

$$T_i^{(l)} = (R_i^{-1} S_i P_i^{-1} Q_i)^{q(i,l)} + \sum_{j=0}^{q(i,l)-1} (R_i^{-1} S_i P_i^{-1} Q_i)^j R_i^{-1} (S_i P_i^{-1} + I) N, \quad (7)$$

or equivalently as

$$T^{(l)} = \sum_{i=1}^{r} E_i [(R_i^{-1} S_i P_i^{-1} Q_i)^{q(i,l)} + (I - (R_i^{-1} S_i P_i^{-1} Q_i)^{q(i,l)}) M^{-1} N]. \quad (8)$$

With the above notation, the iterative scheme (6) describes an alternating two-stage Block-Jacobi type method but note that this method is much more general if, for example other matrices M, P_i, R_i and/or E_i are chosen. Particularly if $P_i = P$, $R_i = R$ for all $i = 1, \ldots, r$, this iteration scheme includes the alternating method described in [3] but this general formulation allows us to include overlapping setting the weighting diagonal nonnegative matrices E_i such that they add up to the identity. From a theoretical point of view, under certain hypotheses, the presence of overlap can reduce the convergence rate of the iterative solvers in the nonsingular case. Therefore, if the extra work required by the use of overlap is offset by a reduction in the number of iterations, probably the computation time will be reduced; see [4] and [5]. The experiments performed in [4] have been executed in only one processor using Matlab for the code implementation. However, to run the experiments of [5], a parallel block iterative code was implemented with the Block Jacobi method as the outer iteration and the point Gauss-Seidel method as the inner iteration. The test matrix was generated from the discretization of the Laplace's equation using the standard five-point stencil and the experiments were performed on a parallel computer using 16 processors. As compared to the non-overlapping implementation, the parallel implementation with overlapping blocks achieved a time reduction about 5%, when the involved parameters in both algorithms were chosen near to the optimal values.

Despite the fact that the behavior of the convergence rate of the block-based iterative solvers with overlap is an open question, specially in the singular case, some numerical results given in [6] show that overlap can also improve the asymptotic convergence factor and the sequential execution time of iterative methods for singular systems, and specifically for ergodic Markov chains.

The use of quite general weighting matrices in (6) allows us the study of truly parallel methods (with or without overlap), i.e., methods in which each processor computes an approximation to the solution of a problem which is much smaller than the original problem. Recently convergence of (6) has been analyzed in the context of solving nonsingular linear systems obtaining similar convergence results to those obtained in [3]; see [7] and [8]. In this paper we give convergence results of these methods considering the general formulation for consistent linear systems. Concretely, in Section 3, we give convergence results of these methods when M-matrices or symmetric matrices are considered. The numerical experiments performed in Section 4 explore the behavior of these parallel algorithms for the solution of singular and nonsingular systems. Previously, in Section 2, we present some definitions and preliminaries that are used later in the paper. The conclusions are given in Section 5. This paper is based upon Migallón et al. [9], but the current paper includes the following additional research: new convergence results for symmetric positive semidefinite matrices are given and new parallel versions of these methods and implementation strategies, with and without overlapping blocks, are explored.

2. Notation and preliminaries

In this section we summarize some definitions and theoretical results used later in the paper. Concretely, main results about the existence and uniqueness of splittings for stationary iterative methods are presented, the theoretical concepts of convergent and semiconvergent matrix are introduced along with the most important results that will be used in Section 3 to study the convergence of the alternating two-stage method when the coefficient matrix is both a singular M-matrix or a symmetric positive semidefinite matrix.

A general matrix A is called an M-matrix if A can be expressed as $A = sI - B$, with $B \geq O$, $s > 0$, and $\rho(B) \leq s$. The M-matrix A is singular when $s = \rho(B)$ and nonsingular when $s > \rho(B)$. Let $Z^{n \times n}$ denote the set of all real $n \times n$ matrices which have all non-positive off-diagonal entries. A splitting $A = M - N$ is called regular if $M^{-1} \geq O$ and $N \geq O$, and weak regular if $M^{-1} \geq O$ and $M^{-1}N \geq O$.

Lemma 1 ([2]). *Given a nonsingular matrix A and a matrix T such that $(I - T)^{-1}$ exists, there is a unique pair of matrices P, Q such that P is nonsingular, $T = P^{-1}Q$ and $A = P - Q$. The matrices are $P = A(I - T)^{-1}$ and $Q = P - A$.*

In the context of Lemma 1, it is said that the unique splitting $A = P - Q$ is induced by the iteration matrix T. We point out that when the matrix A is singular, the induced splitting is not unique; see e.g., [10].

Theorem 1 ([10]). *Let A be a nonsingular matrix such that $A^{-1} \geq O$. Let $A = M - N = P - Q$ be weak regular splittings. Consider the matrix $T = P^{-1}QM^{-1}N$, then $\rho(T) < 1$. Furthermore there is a unique pair of matrices B, C, such that $A = B - C$ is a weak regular splitting and $T = B^{-1}C$.*

Let $T \in \Re^{n \times n}$, by $\sigma(T)$ we denote the spectrum of the matrix T. We define $\gamma(T) = \max\{|\lambda| \ : \ \lambda \in \sigma(T), \ \lambda \neq 1\}$. We say that two subspaces S_1 and S_2 on $\Re^n$ are complementary if $S_1 \oplus S_2 = \Re^n$, i.e., if $S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \Re^n$. The *index* of a square matrix T, denoted by $ind T$, is the smallest nonnegative integer k such that $\mathcal{R}(T^{k+1}) = \mathcal{R}(T^k)$. By $ind_1 T$ we denote the *index* associated with the value one, i.e., $ind_1 T = ind(I - T)$. Note that when $\rho(T) = 1$, $ind_1 T \leq 1$ if and only if $ind_1 T = 1$. We say that a matrix $T \in \Re^{n \times n}$, is convergent if $\lim_{k \to \infty} T^k = O$. It is well known that a matrix T is convergent if and only if $\rho(T) < 1$. By $\mathcal{N}(T)$ we denote the null space of T. We say that T is semiconvergent if $\lim_{k \to \infty} T^k$ exists, although it need not be the zero matrix. If, on the other hand, $\rho(T) = 1$, two different conditions need to be satisfied to guarantee semiconvergence, as the following result shows.

Theorem 2 ([11]). *Let $T \in \Re^{n \times n}$, with $\rho(T) = 1$. The matrix T is semiconvergent if and only if the following two statements hold.*

(a) $1 \in \sigma(T)$ *and* $\gamma(T) < 1$,
(b) $\mathcal{N}(I - T) \oplus \mathcal{R}(I - T) = \Re^n$.

Condition (b) is equivalent to the existence of the group inverse $(I - T)^{\#}$, and it is also equivalent to having $ind_1 T = 1$; see, e.g., [12].

Definition 1 ([12]). Let $A \in \Re^{n \times n}$, and consider the following matrix equations.

(1) $AXA = A$,
(2) $XAX = X$, and
(3) $AX = XA$.

A {1, 2}-inverse of A is a matrix X which satisfies conditions (1) and (2). If, in addition, X satisfies condition (3), X is said to be a group inverse of A.