Lecture Notes in
Mathematics

Edited by A. Dold, B. Eckmann and F. Takens

1457

O. Axelsson L. Yu. Kolotilina (Eds.)

Preconditioned Conjugate Gradient Methods

Proceedings, Nijmegen 1989

Springer-Verlag

Lecture Notes in Mathematics

Edited by A. Dold, B. Eckmann and F. Takens

1457

O. Axelsson L. Yu. Kolotilina (Eds.)

Preconditioned Conjugate Gradient Methods

Proceedings of a Conference held in Nijmegen,
The Netherlands, June 19–21, 1989

Springer-Verlag
Berlin Heidelberg New York London
Paris Tokyo Hong Kong Barcelona

Editors

Owe Axelsson
University of Nijmegen
Faculty of Mathematics and Informatics
Toernooiveld 1
6525 Nijmegen, The Netherlands

Lily Yu. Kolotilina
Steklov Mathematical Institute
LOMI
Fontanka 27, Leningrad D-111, USSR

Mathematics Subject Classification (1980): 65F10, 65N20, 65N30

ISBN 3-540-53515-2 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-53515-2 Springer-Verlag New York Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is only permitted under the provisions of the German Copyright Law of September 9, 1965, in its current version, and a copyright fee must always be paid. Violations fall under the prosecution act of the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1990
Printed in Germany

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2146/3140-543210 – Printed on acid-free paper

Preface

An International Conference on Preconditioned Conjugate Gradient Methods was held at the Faculty of Mathematics and Informatics, University of Nijmegen, The Netherlands on June 19-21, 1989.

The main motivation in organizing this Conference was the wish to bring together specialists working on iterative solution methods, in particular using preconditioning methods. The conference was preceeded by a short course on preconditioned conjugate gradient methods, on June 15-17, 1989.

Although the conference was organized and announced only a few months prior to its taking place, many scientists from different countries attended the conference. The director of the faculties of sciences, dr.ir. L.H.J. Wachters, had kindly accepted the invitation to open the conference.

The topics presented at the conference contained both analysis and implementational aspects of the methods. The proceedings contains the full text of 11 invited or contributed papers.

We are grateful to the participants, and especially to the speakers who made this meeting an important scientific event. We are also most indebted to the Administration of the Faculty of Mathematics and Informatics, to the younger members of the numerical analysis group and to the secretatiat of the Institute of Mathematics whose support was an indispensable contribution to the success of this conference.

O. Axelsson and L. Kolotilina

MODIFIED INCOMPLETE FACTORIZATION STRATEGIES

R. BEAUWENS
UNIVERSITÉ LIBRE DE BRUXELLES, SERVICE DE MÉTROLOGIE NUCLÉAIRE
50, AV. F.D. ROOSEVELT, B-1050 BRUSSELS, BELGIUM

Abstract. We review here the conditioning properties of modified incomplete symmetric factorizations of Stieltjes matrices, appropriate for the PCG solution of symmetric positive definite linear systems. Emphasizing the algorithmic interpretation of the required assumptions, we analyse the theoretical support that can be given to practical factorization algorithms and the choice left open between "perturbed" and "unperturbed" policies. Recent results extending the scope of unperturbed modified factorizations are included and discussed.

1. Introduction. We consider in this work the PCG solution of a linear system

$$Ax = b \tag{1.1}$$

with the purpose of reviewing a priori upper bounds on the spectral condition number

$$\kappa(B^{-1}A) = \frac{\lambda_{max}(B^{-1}A)}{\lambda_{min}(B^{-1}A)} \tag{1.2}$$

of the preconditioned matrix $B^{-1}A$ and factorization "strategies" that are supported or suggested by these results, under the assumptions that A is a Stieltjes matrix and that B is determined from a modified incomplete factorization of A.

We notice here that the case of an arbitrary symmetric positive definite matrix A may be reduced to the Stieltjes case (or even to some more restrictive class of matrices) provided that one can determine spectral bounds for a pencil of the form $A - \lambda A_o$ where A_o is Stieltjes since, if α and β are positive numbers such that, for all $z \in C^n, z \neq 0$,

$$\alpha \leq \frac{(z, Az)}{(z, A_o z)} \leq \beta, \tag{1.3}$$

then

$$\kappa(B^{-1}A) \leq \frac{\beta}{\alpha}\kappa(B^{-1}A_o). \tag{1.4}$$

The assumption that B is determined from an incomplete factorization of A means that B is of the form

$$B = U^T P^{-1} U \tag{1.5}$$

where U is upper triangular with positive diagonal entries, $P = diag(U)$, and that the entries of $U = (u_{ij})$ are related to those of $A = (a_{ij})$ through the relations

$$u_{ij} = a_{ij} - \beta_{ij} \sum_{r<i} \frac{u_{ri}u_{rj}}{u_{rr}} \qquad i \leq j \tag{1.6}$$

where $\beta = (\beta_{ij})$ is a given (0,1) matrix. The assumption, made here, that B is determined from a *modified* incomplete factorization of A means that the determination of the diagonal entries of U is modified with respect to the previous scheme. These entries are now determined from the relations

$$B\hat{x} = A\hat{x} + \Lambda D\hat{x} \tag{1.7}$$

试读结束，需要全本PDF请购买 www.ertongbook.com

where $\hat{x}$ is a given positive vector, $D = diag(A)$ and Λ is a nonnegative diagonal matrix. The latter relations may equivalently be written

$$U\hat{x} = A\hat{x} + \Lambda D\hat{x} - (U^T - P)P^{-1}U\hat{x}. \tag{1.8}$$

In other words, modified incomplete factorizations use the relations (1.6) to determine $(U - P)$ and (1.8) to determine $U\hat{x}$, whence P.

The matrix $\Lambda = (\lambda_i \delta_{ij})$ is a nonnegative diagonal matrix of "small" parameters λ_i that may be given (static factorization algorithms) or that may be determined during the computation of U (dynamic factorization algorithms). The proper choice of Λ is crucial for the convergence behaviour of the associated PCG method and we shall review below the recipes that can be given for its choice on the basis of available upper bounds on $\kappa(B^{-1}A)$.

Standard definitions and notation used throughout the paper are the following. All matrices are $n \times n$ matrices and all vectors are n vectors. The order relation between real matrices and vectors is the usual componentwise order : with $A = (a_{ij})$ and $B = (b_{ij})$, then $A \leq B$ $(A < B)$ if $a_{ij} \leq b_{ij}$ $(a_{ij} < b_{ij})$ for all i, j. A is said nonnegative (positive) if $A \geq 0$ $(A > 0)$ and monotone if it is nonsingular with $A^{-1} \geq 0$. An M-matrix is a monotone matrix with nonpositive offdiagonal entries. A Stieltjes matrix is a symmetric M-matrix. If $A = (a_{ij})$, we denote by $diag(A)$ the (diagonal) matrix with entries $a_{ii}\delta_{ij}$. By e we denote the vector with all components equal to unity.

2. Spectral Bounds.

We first review the most general spectral bounds obtained to date for the pencil $A - \nu B$, trying to emphasize the algorithmic interpretation of the assumptions required by these results. The following general assumptions will be made in most theorems reported in this work and referred to as (GA). Proofs that can be found elsewhere are omitted. See [5,6,8,10] for those of the results reported in the present section.

DEFINITION 2.1. *General assumptions* (GA) *mean the following.*

$$\begin{aligned} &A \text{ is a real symmetric matrix,}\\ &D = diag(A),\\ &U \text{ is a real upper triangular matrix,}\\ &P = diag(U) \text{ is nonnegative and nonsingular,}\\ &B = U^T P^{-1} U,\\ &\hat{x} \text{ is a given positive vector.}\end{aligned}$$

We first separate upper and lower spectral bounds to display the proper relevance of additional assumptions.

THEOREM 2.2. *Adding to* (GA) *that*

$$U \text{ is an } M - matrix, \tag{2.1}$$

$$offdiag(U^T + U) \leq offdiag(A), \tag{2.2}$$

$$B\hat{x} \geq (1 - \tau)A\hat{x}, \tag{2.3}$$

$$U\hat{x} \geq (1 - \tau)P\hat{x}, \tag{2.4}$$

$$\tau < 1, \tag{2.5}$$

then

$$\lambda_{max}(B^{-1}A) \leq \frac{1}{1-\tau} \quad . \tag{2.6}$$

THEOREM 2.3. *Adding to (GA) that*

$$offdiag(A-(U^T-P)P^{-1}(U-P)) \leq offdiag(U^T+U), \tag{2.7}$$

$$B\hat{x} \leq A\hat{x}, \tag{2.8}$$

then

$$\lambda_{min}(B^{-1}A) \geq 1. \tag{2.9}$$

Gathering now these results, we get :

THEOREM 2.4. *Adding to (GA) that*

(2.10) U *is an* $M-matrix$,

(2.11) $offdiag(A-(U^T-P)P^{-1}(U-P)) \leq offdiag(U^T+U) \leq offdiag(A)$,

(2.12) $(1-\tau)A\hat{x} \leq B\hat{x} \leq A\hat{x}$,

(2.13) $U\hat{x} \geq (1-\tau)P\hat{x}$,

(2.14) $\tau < 1$,

then

$$1 \leq \lambda(B^{-1}A) \leq \frac{1}{1-\tau} \quad , \tag{2.15}$$

whence

$$\kappa(B^{-1}A) \leq \frac{1}{1-\tau} \quad . \tag{2.16}$$

The additional assumptions required by these results may be subdivided in two groups.

The first group, comprising (2.1), (2.2) and (2.7) or, in the combined version, (2.10) and (2.11), bears on the offdiagonal entries of U and A. They essentially determine the class of matrices A and B that are covered by the results. It is readily checked that, for incomplete or modified incomplete factorizations, (2.10) and (2.11) are always satisfied when A is a Stieltjes matrix. We refer to [10] for a recent generalization of these assumptions.

The second group, comprising (2.3), (2.4), (2.5) and (2.8) or, in the combined version (2.12), (2.13), (2.14), bears on the diagonal entries of U and deserves careful attention.

It should first be noticed that compatibility of (2.3) and (2.8) requires that $\hat{x}$ is chosen such as $A\hat{x} \geq 0$. This again is always possible when A is a Stieltjes matrix. The existence of $\hat{x} > 0$ with $A\hat{x} > 0$ is indeed a well-known criterion for A with $offdiag(A) \leq 0$ to be an M-matrix (cf. [18] : Theorem 1). In that case, $A\hat{x} > 0$ is further a sufficient condition for the existence of B satisfying (1.5)-(1.7). Existence criteria useful under the less stringent condition $A\hat{x} \geq 0$ will be mentioned below.

It should next be noticed that, if P is so chosen (i.e. large enough) that $U\hat{x} \geq \alpha P\hat{x}$ with $\alpha = 1 - \tau > 0$, then, by Theorem 2.2, $1/\alpha$ is an upper spectral bound for the pencil $A - \nu B$. For that purpose, it is sufficient to increase appropriate components of $B\hat{x}$ as it is readily seen from the relation

$$U\hat{x} = B\hat{x} - (U^T - P)P^{-1}U\hat{x}. \tag{2.17}$$

But $B\hat{x}$ may not be arbitrarily increased because of condition (2.8) of Theorem 2.3. The best compromise within the constraints (2.12) appears therefore to be

$$B\hat{x} = A\hat{x} \tag{2.18}$$

and, if this choice also meets the condition (2.13) with (2.14), then we get the bound (2.16). In other cases, one may try to increase $B\hat{x}$ beyond the limit $A\hat{x}$ (hereby requiring an improved version of Theorem 2.3) or philosophically live with $B\hat{x} = A\hat{x}$ (hoping that an improved version of Theorem 2.2 does apply).

Before proceeding to the consideration of these "strategies", we wish to figure out the order of magnitude of the bound $1/(1-\tau)$ with an analytical estimate which, although less accurate, is more appropriate for that purpose. It should however be appreciated that τ is readily computed during the factorization and thus, always numerically accessible.

Analytical results to be stated here and in later sections rest on the consideration of matrix graph properties and we now need a short digression into this area to recall some terminology. We refer to [11] or [14] for the general terminology on matrix graphs with the warning that all graphs considered in this work are ordered undirected graphs with node set $[1, n]$, i.e. the ordered set of the first n integers or, when subgraphs are considered, some subset of $[1, n]$. Given a graph G, we use the notation $Adj(i)$ to denote the set of neighbors of i in G; with $M \subset [1, n]$, we further set

$$Adj(M) = \bigcup_{i \in M} Adj(i).$$

DEFINITION 2.5. *An increasing path in a graph is a path $i_o, i_1, i_2, \ldots i_\ell$ such that $i_o < i_1 < i_2 < \ldots < i_\ell$.*

DEFINITION 2.6. *The maximal increasing length $\ell(M)$ of an nonempty subset M of the node set of the graph G is the length of the longest increasing path in the subgraph of G induced by M. We further set $\ell(\emptyset) = -1$.*

DEFINITION 2.7. *A node k of a graph G is called a precursor (successor) of the node i of G if $k \in Adj(i)$ with $k < i (k > i)$. The set of precursors (successors) of i is denoted by $P(i)$ $(S(i))$. If M is a subset of the node set of G, the set of precursors (successors) of the nodes of M is denoted by $P(M)$ $(S(M))$. In other words :*

$$P(M) = \bigcup_{i \in M} P(i), \qquad S(M) = \bigcup_{i \in M} S(i).$$

We further set $P(\emptyset) = S(\emptyset) = \emptyset$.

DEFINITION 2.8. *For any node i of a graph G, we define the ascent $As(i)$ of i as*

$$As(i) = \{k \mid \textit{There exists an increasing path from } k \textit{ to } i\}.$$

For any set M of nodes of G, we define the ascent $As(M)$ of M as

$$As(M) = \bigcup_{i \in M} As(i)$$

if $M \neq \emptyset$, with $As(\emptyset) = \emptyset$.

DEFINITION 2.9. *For any pair of nodes i and j of a graph G, we denote by*

$$Pc(i,j) = P(i) \cap P(j)$$

their set of common precursors. We further set

$$Pc(G) = \bigcup_{\substack{i,j=1 \\ i \neq j}}^{n} Pc(i,j)$$

and, if G is the graph of a matrix A, we also use $Pc(A)$ for $Pc(G)$.

It may be mentioned as a first application of these definitions, that the assumption (2.4) of Theorem 2.2 (whence also (2.13) of Theorem 2.4) may be weakened to

$$(2.19) \qquad (U\hat{x})_i \geq (1-\tau)(P\hat{x})_i \qquad \textit{for} \qquad i \in Pc(U) \quad ,$$

a remark which leads us to the following analytical bound.

THEOREM 2.10. *Under (GA),* (2.1), (2.2) *and*

$$(2.20) \qquad \frac{1}{k+\ell_i+1}((U^T-U)\hat{x})_i \leq (B\hat{x})_i \qquad \textit{for} \qquad i \in As(Pc(U))$$

$$(2.21) \qquad \frac{1}{k+\ell+2} \quad A\hat{x} \leq B\hat{x}$$

with $\ell_i = \ell(As(i))$, $\ell = \ell(As(Pc(U)))$ and $k \geq 0$, we have that the conditions (2.3) *and* (2.19) *are satisfied with $1-\tau = 1/(k+\ell+2)$. Therefore*

$$(2.22) \qquad \lambda_{max}(B^{-1}A) \leq k+\ell+2$$

and, if (2.7) *and* (2.8) *are also met,*

$$(2.23) \qquad \kappa(B^{-1}A) \leq k+\ell+2.$$

This result displays a relation between the ordering of U and the condition number of $B^{-1}A$ since the assumption (2.20) and the conclusion (2.22) (or (2.23)) clearly depend on the ordering of U. One may recommend, on this basis, to order U so as to satisfy (2.20) (together with $B\hat{x} = A\hat{x}$ as already discussed) with the smallest possible value of $k+\ell$.

Consideration of applications to discrete PDE's shows on the other hand that this bound may be considered as an algebraic generalization of the $O(h^{-1})$ bound of the geometrical approach developed by Axelsson and his co-workers (cf. [2,13] and the references cited there).

It must finally be remarked that the conditions (2.19) and (2.14) require the strict diagonal dominance of U with respect to $Pc(U)$ and $\hat{x}$ according to the following definition.

DEFINITION 2.11. *Let $A = (a_{ij})$ be a matrix with positive diagonal entries and non-positive offdiagonal entries, let x be a positive vector and $M \subset [1,n]$. We call dominance ratio of A with respect to M and x the number*

$$t = \max_{i \in M} \left(\frac{-\sum_{j \neq i} a_{ij} x_j}{a_{ii} x_i} \right)$$

with $t = 0$ if $M = \emptyset$. A is diagonally dominant (with respect to M and x) if $t \leq 1$ and strictly diagonally dominant (with respect to M and x) if $t < 1$.

Before concluding this section,let us reconsider the existence problem of B subject to (1.5)-(1.7) when A is an M-matrix with $\hat{x} > 0$ such that $A\hat{x} \geq 0$. Notice first that Gustafsson's existence analysis (cf. [13] : Theorem 3.1) precludes the unperturbed case by requiring a strictly positive perturbation $\Lambda D\hat{x}$. The latter condition acts like strict diagonal dominance (i.e. $A\hat{x} > 0$) and dispenses with any additional requirement. Notice on the other hand that the author's criterion (cf. [5] : Corollary of Theorem 2.1) of lower semistrict diagonal dominance (i.e. $A\hat{x} \geq 0$ with $\sum_{j=1}^{j=i} a_{ij} x_j > 0$) is sufficient but not necessary. The precise necessary and sufficient additional condition has recently been obtained by Notay (cf. [17] : Theorem 3.4) and writes in the present framework, with graph notation referring to $G(U)$,

$$\forall j : S(j) = \emptyset \Rightarrow \exists i \in As(j) \textit{ with } (A\hat{x})_i + \lambda_i (D\hat{x})_i > 0 \tag{2.24}$$

and in particular, when $\Lambda = 0$,

$$\forall j : S(j) = \emptyset \Rightarrow \exists i \in As(j) \textit{ with } (A\hat{x})_i > 0. \tag{2.25}$$

It may further be added that (2.25) is true a.o. whenever the only nodes without successors are the last nodes of the connected components of $G(U)$, a condition which occurs under the assumptions considered in Section 4.

3. Factorizations algorithms.

Let us now recall that modified incomplete factorizations use the relations (1.6) to determine $(U - P)$ and (1.8) to determine $U\hat{x}$, whence P. We consider in this section that A is a Stieltjes matrix and that β is a given (0,1) matrix. It then follows from (1.6) that the assumptions (2.1), (2.2) and (2.7) of Theorems 2.2 and 2.3 are satisfied.

It remains to determine P. We consider here four examples of algorithms that may be used for that purpose following the "strategies" indicated in the previous section and we try to analyse their properties on the basis of the preceding theory.In all cases, a (small) parameter α with $0 < \alpha < 1$ and a positive vector $\hat{x}$ have to be chosen; $1/\alpha$ is the "target" upper bound on $\kappa(B^{-1}A)$; $\hat{x}$ is used to measure the degree of diagonal dominance of U and must be such that $A\hat{x} \geq 0$. We also assume here that (2.25) is satisfied. A simple modification dispensing us from that requirement will be indicated in fine.

It is understood that each line of U is computed successively beginning with u_{ij} for $i < j \leq n$ (cf. (1.6)) and ending with u_{ii} (cf. below). It is clear that, at the i^{th} stage, the formula (1.6) needs u_{ss} for $s < i$ only.

STRATEGY 1. *Compute $P = diag(U)$ by solving, at the i^{th} stage,*

$$(U\hat{x})_i = (A\hat{x})_i - ((U^T - P)P^{-1}U\hat{x})_i \tag{3.1}$$

for u_{ii}.

Strategy 2. *Denote by $P_o^{(i)}$ the diagonal matrix equal to $P = diag(U)$ with the (possible) exception of its i^{th} (diagonal) entry which we denote by u_{ii}^o and let $U_o^{(i)} = P_o^{(i)} + (U - P)$. Notice that neither $P_o^{(i)}$ nor P is entirely known at the i^{th} stage but this is irrelevant for our purpose.*

Determine first u_{ii}^o by solving

$$(U_o^{(i)}\hat{x})_i = (A\hat{x})_i - ((U-P)^T P^{-1} U\hat{x})_i \tag{3.2}$$

for u_{ii}^o and compute $\alpha_o^i = (U_o^{(i)}\hat{x})_i/(P_o^{(i)}\hat{x})_i$. If $i \notin As(Pc(U))$ or if $\alpha_o^i \geq \alpha$, set $u_{ii} = u_{ii}^o$. Otherwise, determine u_{ii} by solving

$$(U\hat{x})_i = (A\hat{x})_i + \delta_i(P_o^{(i)}\hat{x})_i - ((U-P)^T P^{-1} U\hat{x})_i \tag{3.3}$$

for u_{ii}, where δ_i is given by

$$\delta_i = \frac{\alpha^2}{1-\alpha} + \frac{\alpha}{1-\alpha}\,\frac{max\{((U^T - U)\hat{x})_i, 0\}}{(P_o^{(i)}\hat{x})_i}. \tag{3.4}$$

Strategy 3. *Proceed in the same way as in the former case, but define δ_i by*

$$\delta_i = \frac{\alpha^2}{1-\alpha} + \frac{\alpha}{1-\alpha}\,\frac{((U^T - U)\hat{x})_i}{(P_o^{(i)}\hat{x})_i}. \tag{3.5}$$

Strategy 4. *Determine u_{ii}^o, $P_o^{(i)}$, $U_o^{(i)}$ and α_o^i as in the former case. If $i \notin Pc(U)$ or if $\alpha_o^i \geq \alpha$, set $u_{ii} = u_{ii}^o$. Otherwise, determine u_{ii} by solving* (3.3) *for u_{ii} with*

$$\delta_i = \frac{\alpha - \alpha_o^i}{1-\alpha}. \tag{3.6}$$

The first strategy was called "unmodified" factorization in [2] (with $\hat{x} = e$). We prefer to call it "unperturbed" to avoid confusion with the case where P is also determined by the relation (1.6) and because it amounts to set $\Lambda = 0$ in the relation (1.7). The second strategy is an obvious adaptation to our framework of the Axelsson-Barker scheme (defined by Eqs (7.18) in [2]) and the third one is a simplified version of the same algorithm. The fourth strategy is new although supported by the same philosophy as we shall see below. Notice that in all cases, the diagonal matrix $\Lambda = (\lambda_i \delta_{ij})$ of Eq. (1.7) may be determined from

$$\lambda_i = \frac{u_{ii}^o}{a_{ii}}\,\delta_i \tag{3.7}$$

provided that δ_i is initialized to zero at the beginning of the i^{th} stage.

The first remark to be done when discussing these strategies is that, would it happen that $\alpha_o^i \geq \alpha$ at all stages, then all four strategies reduce to the first one. It does happen in particular when the assumptions of Theorem 2.10 are satisfied, with $\alpha = \frac{1}{k+\ell+2}$ and this shows that the indications provided by the latter result should be used before starting any of these procedures to :

(1) try to find an ordering of U such that $k+\ell$ be not too large and that the condition (2.20) is approximately satisfied at most if not all nodes of $As(Pc(U))$;

试读结束，需要全本PDF请购买 www.ertongbook.com

(2) choose α once the ordering has been chosen.

Our next remark is to emphasize the opposite points of view which support the first strategy on the one hand and the other three on the other hand : the first strategy preserves an exact lower bound for the pencil $A - \nu B$ since 1 is then the lowest eigenvalue of $A - \nu B$ (with $\hat{x}$ as associated eigenvector) while, as it will now be shown, the other strategies imply that $1/\alpha$ is an upper bound on the spectrum of $A - \nu B$.

THEOREM 3.1. *Adding to* (GA) *that*

$$A \textit{ is a Stieltjes matrix with } A\hat{x} \geq 0 \ , \tag{3.8}$$

$$U - P \textit{ satisfies } (1.6) \textit{ with } 0 \leq \beta_{ij} \leq 1 \textit{for } i < j \leq n \ , \tag{3.9}$$

$$P \textit{ is determined from strategy } n^o i \textit{ with } i = 2 \textit{ or } 3 \textit{ or } 4 \ , \tag{3.10}$$

$$0 < \alpha < 1 \ , \tag{3.11}$$

then (2.3) *and* (2.19) *are satisfied with* $1 - \tau = \alpha$. *Therefore*

$$\lambda_{max}(B^{-1}A) \leq \frac{1}{\alpha} \quad . \tag{3.12}$$

Proof. It is clearly sufficient to show that the i^{th} stage leads to u_{ii} such that $(U\hat{x})_i \geq \alpha(P\hat{x})_i$, whenever $i \notin Pc(U)$.

In the case of the fourth strategy, this follows from

$$\begin{aligned}(U\hat{x})_i - \alpha(P\hat{x})_i &= (U_o^{(i)}\hat{x})_i + \delta_i(P_o^{(i)}\hat{x})_i - \alpha((P_o^{(i)}\hat{x})_i + \delta_i(P_o^{(i)}\hat{x})_i) \\ &= (\alpha_o^i - \alpha + \delta_i(1-\alpha))(P_o^{(i)}\hat{x})_i = 0\end{aligned}$$

when $\alpha_o^i < \alpha$ while there is nothing to prove when $\alpha_o^i \geq \alpha$.

In the case of the third strategy, we prove by induction that $(U\hat{x})_i \geq \alpha(P\hat{x})_i$ whenever $i \in As(Pc(U))$ (whence also in particular for $i \in Pc(U)$). If $\alpha_o^i < \alpha$, we have

$$\begin{aligned}(U\hat{x})_i - \alpha(P\hat{x})_i &= (U_o^{(i)}\hat{x})_i + ((1-\alpha)\delta_i - \alpha)(P_o^{(i)}\hat{x})_i \\ &= (U_o^{(i)}\hat{x})_i + \alpha((U^T - U)\hat{x})_i - \alpha(1-\alpha)(P_o^{(i)}\hat{x})_i \\ &= (U_o^{(i)}\hat{x})_i + \alpha(U^T - P)\hat{x})_i \\ &\qquad - \alpha((U - P)\hat{x})_i - \alpha(1-\alpha)(P_o^{(i)}\hat{x})_i \\ &= (A\hat{x})_i - ((U^T - P)P^{-1}(U - \alpha P)\hat{x})_i \\ &\qquad + \alpha(1 - \alpha_o^i)(P_o^{(i)}\hat{x})_i - \alpha(1-\alpha)(P_o^{(i)}\hat{x})_i \\ &= (A\hat{x})_i - ((U^T - P)^{-1}(U - \alpha P)\hat{x}_i + \alpha(\alpha - \alpha_o^i)(P_o^{(i)}\hat{x})_i \quad ,\end{aligned}$$

but the first term of the RHS of the latter equation is nonnegative and the last one positive by assumption while the second one is nonnegative by the induction hypothesis, entailing

$$(U\hat{x})_i - \alpha(P\hat{x})_i > 0$$

while again, there is nothing to prove if $\alpha_o^i \geq \alpha$.

The same argument holds a fortiori in the case of the second strategy. ∎

It is now clear that the last three strategies satisfy all the assumptions of Theorem 2.4 with the possible exception of the right inequality (2.12). Defining $\Lambda = (\lambda_i \delta_{ij})$ through (3.7) we actually have

$$B\hat{x} = A\hat{x} + \Lambda D\hat{x}$$

and the question raises to evaluate the influence of the perturbation $\Lambda D\hat{x}$ on the lower spectral bound of $A - \nu B$. One has the following result.

THEOREM 3.2. *Under (GA), (2.7) and*

$$B\hat{x} \leq A\hat{x} + \Lambda D\hat{x} \tag{3.13}$$

$$\Lambda = (\lambda_i\, \delta_{ij}) \geq 0 \tag{3.14}$$

$$\xi = \max_{z \neq 0} \frac{(z, \Lambda D z)}{(z, Az)} \tag{3.15}$$

one has

$$\lambda_{min}(B^{-1}A) \geq \frac{1}{1+\xi} \quad . \tag{3.16}$$

Proof. We have by (2.7) that

$$offdiag(A + \Lambda D - B) = offdiag(A - B) \leq 0$$

and, by (3.13) with (3.14) that

$$(A + \Lambda D - B)\hat{x} \geq 0.$$

It follows that $A + \Lambda D - B$ is nonnegative definite, whence by (3.15),

$$(z, Bz) \leq (z, (A + \Lambda D)z) \leq (z, Az)(1 + \xi)$$

which implies (3.16). ■

This result reduces the question raised above to the estimation of ξ. Clearly

$$\xi \leq \frac{\lambda}{\lambda_1(D^{-1}A)} \quad , \tag{3.17}$$

where

$$\lambda = \max_{1 \leq i \leq n} (\lambda_i) \tag{3.18}$$

and

$$\lambda_1(D^{-1}A) = \min_{z \neq 0} \frac{(z, Az)}{(z, Dz)} \quad . \tag{3.19}$$

But this bound on ξ is often too large; more precisely, with

$$<\lambda> = \frac{(\hat{x}, \Lambda D\hat{x})}{(\hat{x}, D\hat{x})} \quad , \tag{3.20}$$

it is found that (3.17) is accurate when $\lambda \simeq <\lambda>$ but an order of magnitude too large when $\lambda \gg <\lambda>$ (see [6] and [8] for a more detailed account).

Unfortunately, no closed form formula has been rigorously obtained to deal with the latter case. Instead, appropriate procedures have been set up for obtaining upper bounds on ξ in specific situations. Since these procedures are somewhat involved, we must however refer the reader to the literature (cf. [2,8,12,15]) for more detailed information.

Nevertheless, the following heuristic formula can be proposed

$$\xi \simeq \frac{<\lambda>}{\lambda_1(D^{-1}A)} \tag{3.21}$$

inasmuch as one may admit that the fundamental eigenvector z_1 of th pencil $A - \lambda D$ is a valid trial vector for the Raleigh quotient estimate of ξ, i.e. that

$$\xi \simeq \frac{(z_1, \Lambda D z_1)}{(z_1, A z_1)} = \frac{(z_1, \Lambda D z_1)}{(z_1, D z_1)} \frac{1}{\lambda_1(D^{-1}A)} \tag{3.22}$$

and further that

$$\frac{(z_1, \Lambda D z_1)}{(z_1, D z_1)} \simeq < \lambda > . \tag{3.23}$$

Despite the lack of rigorous justification of the formula (3.21), it has the essential merit of representing the most concise summary of the general conclusions that can be drawn from the rigorous analysis of a variety of examples to be found a.o. in [2,8,12,15], reproducing the exact formula when $\lambda = < \lambda >$, exhibiting the right order of magnitude when $\lambda \gg < \lambda >$ and displaying correctly the influence of $\lambda_1(D^{-1}A)$.

The main practical conclusion to be drawn from the latter remark is that the strategies N^o 2, 3 and 4 are likely to provide a good spectral condition number $\kappa(B^{-1}A)$ whenever $< \lambda > \simeq \lambda_1(D^{-1}A)$, i.e. whenever $\lambda_1(D^{-1}A)$ is not too small. In other cases (quasi singular problems) one may attempt to limit the perturbations λ_i so that $< \lambda > \simeq \lambda_1(D^{-1}A)$, i.e. essentially shift to the first strategy. As noticed in Section 2, such an attitude relies on the hope that an improved version of Theorem 2.2 applies. The results to be reported in the next section show that such an hope needs not be unrealistic.

It is now appropriate to reconsider the assumption (2.25). It follows from the (2.24) version of Notay's result that a simple way to take care of any violation of (2.25) at some node j is to introduce a corresponding positive perturbation $\lambda_j(D\hat{x})_j$. Since $S(j) = \emptyset$, this perturbation does not affect later coefficients nor does it affect the upper spectral bound discussed above. Since on the other hand, its size is arbitrary, it can be chosen small enough to have a negligible influence on the lower spectral bound. With this slight modification of our factorization algorithms, the only requirement to be put on $\hat{x}$ is that $A\hat{x} \geq 0$.

Before leaving this section, let us finally notice that the essential limitation of Theorem 2.2 arises from the requirement (2.5) which (together with (2.19)) means that U must be strictly diagonally dominant (with respect to $Pc(U)$ and $\hat{x}$). In this respect, it is of interest to mention the following result by Axelsson [1] (cf. also [3]) which applies to the generalized SSOR method.

THEOREM 3.3. *Adding to* (GA) *that :*

$$\begin{aligned} &A \textit{ is symmetric positive definite,} \\ &offdiag(U^T + U) = offdiag(A), \\ &D\hat{x} \leq 2\ \tau_a\ P\hat{x}, \\ &\tau_a < 1, \end{aligned}$$

then

$$\lambda_{max}(B^{-1}A) \leq \frac{1}{2} \frac{1}{1 - \tau_a} .$$

Exchanging Theorem 2.2 for the latter result, a new family of strategies can be developed (for the generalized SSOR factorization of Stieltjes matrices) in much the same way as done above, but where the perturbations have now to take care of the condition $\tau_a < 1$ (i.e. of the strict positive definiteness of $2P - D$) rather than $\tau < 1$. Unfortunately this new requirement is almost as restrictive as the preceding one and leads us to similar difficulties.

4. The unperturbed strategy. We now report recent results on upper spectral bounds which do neither require the strict diagonal dominance of U nor the strict positive definiteness of $2P - D$. These improved versions of Theorem 2.2 have been obtained in [9] and [16] under the additional assumption that the approximate triangular factors are "S/P consistently ordered". We therefore begin with a few definitions to clarify this notion.

We first rephrase Young's definition of consistent ordering [19,20] through the notion of level structure [11].

DEFINITION 4.1. *A level structure of a graph G is a partitioning of its node set, $\mathcal{L} = (L_k)_{k=0,1,\dots,\ell}$, such that*

$$\begin{array}{lll} (a) & Adj(L_o) \subset L_o \cup L_1 \quad ; & \\ (b) & Adj(L_k) \subset L_{k-1} \cup L_k \cup L_{k+1} & for \quad 0 < k < \ell \ ; \\ (c) & Adj(L_\ell) \subset L_{\ell-1} \cup L_\ell \quad ; & \end{array}$$

ℓ is called the length of $\mathcal{L}$.

DEFINITION 4.2. *A graph G is consistently ordered if there exists a level structure $\mathcal{L} = (L_k)_{k=0,1,\dots,\ell}$ of G such that*

$$i \in L_k \Rightarrow P(i) \subset L_{k-1} \quad for \quad 0 < k \leq \ell$$

and

$$i \in L_k \Rightarrow S(i) \subset L_{k+1} \quad for \quad 0 \leq k < \ell.$$

A matrix is said to be consistently ordered when its graph is consistently ordered.

A level structure satisfying the preceding properties will be called associated with the consistent ordering of G. As observed in [7], if G is consistently ordered and connected, it has a unique level structure associated with its consistent ordering.

Next, complementing Definition 2.9 with the notation $Sc(i,j)$ to denote the common successors of any pair of nodes i and j in a graph G :

$$Sc(i,j) = S(i) \cap S(j),$$

we introduce the notion of S/P consistent ordering.

DEFINITION 4.3. *We say that a graph G is S/P consistently ordered if, for any pair of nodes i, j with $i \neq j$,*

$$Pc(i,j) \neq \emptyset \Rightarrow Sc(i,j) \neq \emptyset \quad .$$

A matrix is said to be S/P consistently ordered when its graph is S/P consistently ordered.

We recall from [7] that an S/P consistently ordered graph is consistently ordered and that if it is connected, its (unique) associated level structure satisfies :

$$\begin{array}{ll} (a) & \textit{its last level is } L_\ell = \{n\} \quad ; \\ (b) & \textit{its } k^{th} \textit{ level is } L_k = P(L_{k+1}) \textit{ for } 0 \leq k < \ell. \end{array}$$

To deal with S/P consistently ordered graphs and matrices, it is appropriate to further complement Definition 2.9 with the following notation : for any couple of nodes i and j in a graph G, we set

$$Sp(i,j) = \begin{cases} Sc(i,j) & \text{if} \quad Pc(i,j) \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}$$

and we define

$$Sp(G) = \bigcup_{\substack{i,j=1 \\ i \neq j}}^{n} Sp(i,j).$$

If G is the graph of the matrix A, we also write $Sp(A)$ for $Sp(G)$.

To introduce our next definition, let $U = (u_{ij})$ be an upper triangular matrix and consider products of the form $u_{ri}u_{rj}$ with $r < i < j$; such expressions are nonzero only if $r \in Pc(U)$; now if U is S/P consistently ordered, to nonzero expressions of this type, we can associate nonzero products of the form $u_{is}u_{js}$ with $i < j < s$ since the existence of $r \in Pc(i,j)$ entails that of $s \in Sc(i,j)$; further, if U is an M-matrix, we can introduce positive normalization constants u_{ijrs} such that $u_{is}u_{js}/u_{ijrs} = u_{ri}u_{rj}/u_{rr}$; more generally, to any pair of nodes i,j with $i \neq j$ such that $Pc(i,j) \neq 0$, we can associate some family of positive parameters u_{ijs}, $s \in Sc(i,j)$ such that

$$\sum_{s \in Sc(i,j)} \frac{u_{is}u_{js}}{u_{ijs}} = \sum_{r \in Pc(i,j)} \frac{u_{ri}u_{rj}}{u_{rr}}.$$

These remarks show that any upper triangular S/P consistently ordered M-matrix has S/P images according to the following definitions.

DEFINITION 4.4. *Let $U = (u_{ij})$ be an upper triangular matrix with positive diagonal entries. Then a lower triangular matrix $L = (\ell_{ij})$ is called an S/P image of U if*

(a) the offdiagonal entries of L satisfy

$$\ell_{ij} = \begin{cases} u_{ji} & \textit{if} \quad i \in Sp(U) \textit{ and } j \in S(Pc(U)), \\ 0 & \textit{otherwise,} \end{cases}$$

(b) the diagonal entries of L satisfy

$$\ell_{ss} = \begin{cases} min\{u_{ijs} | i,j \in P(s), \quad i \neq j, \quad Pc(i,j) \neq \emptyset\} & \textit{if } s \in Sp(U), \\ 1 & \textit{otherwise,} \end{cases}$$

where u_{ijs} is a family of positive parameters defined for all $s \in Sp(U)$, $i,j \in P(s)$, $i \neq j$ and $Pc(i,j) \neq \emptyset$, and which satisfy the relations

$$\sum_{s \in Sc(i,j)} \frac{u_{is}u_{js}}{u_{ijs}} = \sum_{r \in Pc(i,j)} \frac{u_{ri}u_{rj}}{u_{rr}}$$

for all couples of indices i,j with $i \neq j$ and $Pc(i,j) \neq \emptyset$.

DEFINITION 4.5. *Let $U = (u_{ij})$ be an upper triangular S/P consistently ordered M-matrix and let the diagonal matrix $K = (\kappa_i \delta_{ij})$ be such that $0 < \kappa_i \leq 1$ for all $i \leq i \leq n$. Any S/P image of KU is called a reduced S/P image of U, of reduction matrix K and of reduction ratio $\kappa = \min_{1 \leq i \leq n}(\kappa_i)$.*

With these notions in mind, we can now state the following improved version of Theorem 2.2, borrowed from [9].

THEOREM 4.6. *Under (GA), (2.1), (2.2), (2.19) and*

$$U \text{ is } S/P \text{ consistently ordered,} \tag{4.1}$$

$$L \text{ is a reduced } S/P \text{ image of } U, \text{ of reduction ratio } \eta, \tag{4.2}$$

$$B\hat{x} \geq (1 - \tau + \eta \frac{\tau}{\ell + 2})A\hat{x}, \tag{4.3}$$

$$(L\hat{x})_i \geq (1 - \tau)(Q\hat{x})_i \quad for \quad i \in Sp(U), \tag{4.4}$$

$$\tau \leq 1, \tag{4.5}$$

with $Q = diag(L)$ and $\ell = \ell(Pc(U))$, we have that

$$\lambda_{max}(B^{-1}A) \leq \frac{1}{1 - \tau + \eta \frac{\tau}{\ell + 2}} \leq \frac{\ell + 2}{\eta} \tag{4.6}$$

and therefore, if (2.7) and (2.8) are also met,

$$\kappa(B^{-1}A) \leq \frac{1}{1 - \tau + \eta \frac{\tau}{\ell + 2}} . \tag{4.7}$$

The condition (4.5) together with (2.19) means that U is diagonally dominant with respect to $Pc(U)$ and $\hat{x}$ but strict diagonal dominance is no more required. We refer to [9] for numerical results illustrating the power of Theorem 4.6, noticing here that, when U is S/P consistently ordered, it justifies the recourse to the first strategy suggested in Section 3, at least when the other ones are not satisfactory.

It remains of course to consider the implications of the other assumptions of this new bound to be sure that it truly generalizes Theorem 2.2.

It should first be noticed in this respect that (4.2), (4.4) and (4.5) require the existence of a reduced S/P image of U which must be diagonally dominant with respect to $Sp(U)$ and $\hat{x}$. That this requirement is always satisfied has recently been shown by Notay who proved in [16] that any S/P consistently ordered upper triangular M-matrix has reduced S/P images of any a priori prescribed dominance ratio t (with respect to $Sp(U)$ and any given $\hat{x} > 0$).

Because of Notay's theorem, we may introduce the following definition.

DEFINITION 4.7. *Let U be an upper triangular S/P consistently ordered M-matrix, $\hat{x}$ a positive vector and t a positive number. We call maximal reduction ratio of U with respect to $\hat{x}$ and t and we denote by $\eta_t(U, \hat{x})$ the maximal value of η such that U has a reduced S/P image of reduction ratio η and whose dominance ratio with respect to $Sp(U)$ and $\hat{x}$ does not exceed t.*

It follows from these considerations that the assumptions (4.2) and (4.4) may simply be exchanged for

$$\eta \leq \eta_\tau(U, \hat{x}). \tag{4.8}$$

On the other hand, when considering modified incomplete factorizations of a Stieltjes matrix A with $A\hat{x} \geq 0$ and $B\hat{x} = A\hat{x}$, the condition (4.3) is actually implied

试读结束，需要全本PDF请购买 www.ertongbook.com