A

STRUCTURED MATRICES AND NEWTON'S ITERATION: UNIFIED APPROACH

by

YOUSSEF RAMI

A dissertation submitted to the graduate Faculty in Mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The city university of New York

2000

UMI Number: 9969724

UMI®

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Mathematics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 18 2000
Date

Victor Pan
Chair of Examining Committee

April 18 2000
Date

Executive Officer

MICHAEL ANSHEL Michael Anshel

ALEXEI MIASNIKOV Alexei Miasnikov

VICTOR Y. PAN

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

NEWTON'S ITERATION FOR STRUCTURED MATRICES : UNIFIED APPROACH

By

YOUSSEF RAMI

Advisor: Professor Victor Y. Pan

Recent progress in the study of structured matrices shows advantages of unifying the treatment of various classes of such matrices. We show such a unified treatment of Newton's iteration which rapidly improves an initial approximation to matrix inverse by performing two matrix multiplications per recursive step. The iteration is particularly suitable for the inversion of structured matrices because in this case matrix operations are performed much more rapidly and with $O(n)$ entries of short generators of structured matrices rather than with order of n^2 entries of the input matrix, that is, using much less memory space. A major problem is to control the length of the generator, which tends to grow quite rapidly in the iterative process. Some techniques of 1992-93 propose to control the growth in the Toeplitz-like case by reliying on the concept of approximate orthogonal displacement rank, other techniques of 1997-99 used substitution of the computed approxima-

tions for the inverse matrix in the expression for the generators of Newton's iterates. We study both variant of the iteration for the more general class of structured matrices, propose some variations of the iterative process and of the techniques for controlling the length of the generators, and investigate the convergence rate as well as computational complexity. Some novel techniques are introduced in this study, in particular for the estimation of the norms of the auxiliary operators. We also summarize and unify the representation of structutred matrices via their short generators.

Acknowledgements

I am grateful to many people who encouraged and supported me in the completion of this work. In particular, from the start Prof. Victor Pan for all his guidance, spare time and several important comments and suggestions, my committee members, Michael Anshel and Alexei Miasnikov for their time and support and my colleagues for their moral support. I would like to thank all the faculty members from the mathematics Department of the Graduate School, specifically Prof. Linda keen for her excellent lectures in the field of Kleinian Groups, which made me appreciate the true meaning of pure mathematics. Completion of my education would have been impossible without the support of my family, specifically my brother Najem and his wife Anne Marie.

Contents

1 Introduction

1.1 Toeplitz and Hankel matrices and the displacement rank approach

The modern study of structured matrices was largely motivated by the seminal paper [KKM79] and, in particular, by the basic concept of the displacement rank introduced there. The idea was to measure the Toeplitz-like (or Hankel-like) structure of a matrix M by the rank of its displacement, that is, of the image matrix of some standard linear operators of shift (displacement) applied to the matrix M.

The rank of the displacement (called the *displacement rank*) of M is at most 2 for Toeplitz (and Hankel) matrices, and an $m \times n$ matrix M is said to be of *Toeplitz* (or *Hankel*) type or, alternatively, to be *Toeplitz-like* (or *Hankel-like*) if the rank r of its displacement is small (say, bounded by a small constant independent of n). In this case, the matrix can be represented by $(m + n)r$ entries of its short displacement generators rather than by its own mn entries. This enables more efficient storage of such matrices in computer memory as well as much faster computations with them [KS99].

1.2 Extension to the study of other structured matrices

Several other important classes of structured matrices can be defined and treated similarily based on using other linear operators, in particular the scaling operators of multiplication by diagonal matrices and the operators that combine scaling and displacement.

In Table 1 we represent four basic classes of structured matrices, which themselves are highly important in numerous applications to sciences, eigeneering and communication and also have been naturally extended in the above way to cover several other popular and important classes of structured matrices.

The matrices of the four classes of Table 1:

1) are represented with a few parameters (from m to $m+n$ per an $m \times n$ matrix),

2) can be multiplied by vectors much faster than general matrices,

3) are closely related to some operations with polynomials,

4) can be naturally associated with some linear operators of displacement (shift) and/or scaling.

We refer the reader to [BP94] on property 3), we will specify properties

Table 1: **Four classes of structured matrices**

Toeplitz matrices, $T = [t_{i-j}]_{i,j=0}^{n-1}$ Hankel matrices, $H = [h_{i+j}]_{i,j=0}^{n-1}$

$$\begin{bmatrix} t_0 & t_{-1} & \cdots & t_{1-n} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_{n-1} & \cdots & t_1 & t_0 \end{bmatrix} \qquad \begin{bmatrix} h_0 & h_1 & \cdots & h_{n-1} \\ h_1 & \cdots & \cdots & h_n \\ \vdots & \cdots & \cdots & \vdots \\ h_{n-1} & h_n & \cdots & h_{2n-2} \end{bmatrix},$$

Vandermonde matrices, $V = [x_i^j]_{i,j=0}^{n-1}$ Cauchy matrices, $C = [\frac{1}{x_i - y_j}]_{i,j=1}^{n}$

$$\begin{bmatrix} 1 & x_0 & \cdots & x_0^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & \cdots & x_{n-1}^{n-1} \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{x_1 - y_1} & \cdots & \frac{1}{x_1 - y_n} \\ \vdots & & \vdots \\ \frac{1}{x_n - y_1} & \cdots & \frac{1}{x_n - y_n} \end{bmatrix}$$

1) and 2) in Table 2, and then will study property 4).

The four classes of matrices are naturally associated with various linear operators L of scaling and displacement (shift), of the Sylvester type,

$$L(M) = \nabla_{A,B}(M) = AM - MB, \tag{1.1}$$

and/or the Stein type,

$$L(M) = \Delta_{A,B}(M) = M - AMB, \tag{1.2}$$

where A and B are some fixed $m \times m$ and $n \times n$ operator matrices.

Among the common choices for A and B are the diagonal matrices $D(s) = diag(s_1, \dots, s_k)$ for a fixed vector $\mathbf{s} = (s_i)_{i=1}^k$, which are the matrices of

Table 2:

Matrix class	Number of parameters in the $m \times n$ matrix	Number of flops required for multiplication by a vector
Toeplitz and Hankel	$m+n-1$	$O((m+n)\log n)$
Vandermonde	n	$O((m+n)\log^2 n)$
Cauchy	$m+n$	$O((m+n)\log^2 n)$

operators of scaling, and the unit f-circulant matrices (for a fixed scalar f) $Z_f = Z + f\mathbf{e}_0\mathbf{e}_{n-1}^T$ and their transposes Z_f^T, which are the matrices of the shifts (displacement) operators. Here and hereafter, $\mathbf{e}_j$ is the j-th coordinate vector. $Z = (z_{i,j})_{i,j=0}^{n-1}$ is the shift matrix, $z_{i,j} = 1$ for $i = j+1$, $z_{i,j} = 0$ otherwise, $\mathbf{v}^T$ and W^T are the transposes of a vector $\mathbf{v}$ and a matrix W.

To each linear operator $L = \nabla_{A,B}$ or $L = \Delta_{A,B}$, a class of structured matrices M is associated for which the images $\nabla_{A,B}(M)$ and/or $\Delta_{A,B}(M)$ of application of this operator to a matrix M have small rank r relatively to the matrix size, or equivalently, where

$$L(M) = \nabla_{A,B}(M) = AM - MB = GH^T, \tag{1.3}$$

$$L(M) = \Delta_{A,B}(M) = M - AMB = GH^T, \tag{1.4}$$

$$G = (\mathbf{g}_1, \dots, \mathbf{g}_r), \quad H = (\mathbf{h}_1, \dots, \mathbf{h}_r), \tag{1.5}$$

r is the minimum for the given matrices M, A, B and is small relatively to the matrix size.

We will follow [KKM79], [BP94], and [P2000] and will call the matrix pair (A, B) by *operator matrices*, the matrix pair (G, H) by both *generator matrices* for the operator images $L(M)$ and *L-generator matrices* for the matrix M, and the rank of $L(M)$ by the *L-rank* of M (for the operators L of (1.1) and (1.2)).

Table 3 represents some examples of operator matrices associated naturally with the matrices M of Table 1 and with ones having similar structure. The matrices whose L-rank is small for the operator L of the respective lines of Table 3 will be called hereafter *Toeplitz-like, Hankel-like, Vandermonde-like*, and *Cauchy-like* matrices (some authors called them matrices of *Toeplitz, Hankel, Vandermonde*, and *Cauchy* types).

Hereafter, $D(\mathbf{v})$ for $\mathbf{v} = (v_i)$ denotes the $n \times n$ diagonal matrix $diag(v_0, \dots, v_{n-1})$. This includes several important classes of structured matrices such as Sylvester (resultant), Frobenius, Lowener, and Pick matrices [BP94], [OP98]. Furthermore, the block submatrices, the products and the inverses of structured matrices inherit their structure [KKM79], [P90], [BP94], [P2000], that is,

Table 3: **Operator matrices A, B of the operators $\Delta_{A,B}$ associated with four basic classes of structured matrices.**

Class of Matrices	Matrix pairs (A, B) for the operators $\Delta_{A,B}$	rank
Toeplitz	(Z, Z^T) or (Z_{-1}, Z_1^T)	2
Hankel	(Z^T, Z^T) or (Z_{-1}^T, Z_1^T)	2
Vandermonde	$(D(\mathbf{t}), Z)$ or $(D(\mathbf{t}), Z_{-1})$	1
Cauchy	$(D(\mathbf{t}), D(\mathbf{s}))$	1

have small L-rank for appropriate operators L too.

There are also several natural extension of the above matrix classes, such as polynomial Vandermonde-like matrices and Toeplitz+Hankel-like matrices.

1.3 Representation of structured matrices and some related basic operations

The cited operators L associated with structured matrices are readily invertible, that is, we have simple expressions of matrices M via the generators matrices G and H and the operator matrices A and B. Hereafter, $L(\mathbf{v})$ is the lower triangular Toeplitz matrix $\sum_{i=0}^{n-1} v_i Z^i$, $Circ(\mathbf{v})$ is the circulant matrix $\sum_{i=0}^{n-1} v_i Z_1^i$, $J = \begin{bmatrix} & & 1 \\ & \cdots & \\ 1 & & \end{bmatrix}$ the reflection matrix, and $\mathbf{t}^n = (t_i^n)$ for a vector

$\mathbf{t} = (t_i)$.

Theorem 1.1. *[KKM79], [BP94], [GO94].*

$$M = \sum_{i=1}^{\alpha} L(\mathbf{g}_i)L^T(\mathbf{h}_i), \ \text{if } \Delta_{Z,Z^T}(M) = GH^T, \tag{1.6}$$

and consequently (since $JZJ = Z^T$, $J^2 = I$),

$$MJ = \sum_{i=1}^{\alpha} L(J\mathbf{g}_i)L^T(\mathbf{h}_i)J, \ \text{if } \Delta_{Z^T,Z^T}(MJ) = GH^T. \tag{1.7}$$

Furthermore,

$$M = \sum_{i=1}^{\alpha} L(\mathbf{g}_i)D(\mathbf{t})V^{-T}(\mathbf{t})D(\mathbf{h}_i), \ \text{if } \nabla_{D^{-1}(\mathbf{t}),Z}(M) = GH^T, \tag{1.8}$$

$$M = \sum_{i=1}^{\alpha} D(\mathbf{g}_i)D^{-1}(\mathbf{e} - \mathbf{t}^n)V(\mathbf{t})Circ(\mathbf{h}_i^T), \ \text{if } \nabla_{D(\mathbf{t}),Z_1}(M) = GH^T, \tag{1.9}$$

$$M = -\sum_{i=1}^{\alpha} D(\mathbf{g}_i)D(\mathbf{s})C(\mathbf{s},\mathbf{t})D(\mathbf{h}_i), \ \text{if } \nabla_{D(\mathbf{t}),D(\mathbf{t})}(M) = GH^T. \tag{1.10}$$

Theorem 1.1 enables an immediate extension of the cost bounds of Table 2 to more general classes of structured matrices:

We need some additional definitions (cf. [BP94], [GO94]) and basic results.

Definition 1.1. *$v_\alpha = v_\alpha(\nabla_{A,B})$ denotes the arithmetic cost (in flops) of multiplication by a vector of an $n \times n$ matrix M represented by its $\nabla_{A,B}$-generator of length α. $m_\alpha = m_\alpha(\nabla_{A,B})$ denotes the arithmetic cost of multiplying a pair of $n \times n$ matrices, where the input and output matrices are represented by their generators of length α and $O(\alpha)$, respectively.*

Theorem 1.2. *We have* $v_{\alpha,n}(L) = O(\alpha n \log n)$ *for* $L = \nabla_{A,B}, L = \Delta_{A,B}$, *and for any* A, B *from the set* $\{Z_e, Z_f^T\}$ *for any pair of scalars* e *and* f, *we have* $v_{\alpha,n}(L) = O(\alpha n \log^2 n)$ *for* $L = \nabla_{A,B}$ *and* $L = \Delta_{A,B}$, *where* $A = D(\mathbf{s}), B = D(\mathbf{t})$, *or* $A = D(\mathbf{s}), B \in \{Z_f, Z_f^T\}$, *or* $A \in \{Z_f, Z_f^T\}$, $B = D(\mathbf{s})$ *for any pair of vectors* $\mathbf{s}$ *and* $\mathbf{t}$ *and any scalar* f. *Furthermore, in both cases* $m_{\alpha,n}(L) = O(\alpha v_{\alpha,n}(L))$.

Assumption 1.1. *Hereafter, we will assume (motivated by Theorem 1.2) dealing with operators* L *for which*

$$v_{\alpha,n}(L) = O(\alpha n \log^d n), \quad d \leq 2,$$

$$m_{\alpha,n}(L) = O(\alpha v_{\alpha,n}(L)).$$

To take advantage of the matrix structure, we will represent structured matrices M via their short L-generators based on Theorem 1.1 and then will manipulate with L-generators rather than with the matrices themselves.

In particular for linear combinations, products, inverses and blocks of structured matrices, we may define appropriate operator matrices for the input matrices and obtain short L-generators for the output [BP94], [P2000].

Theorem 1.3. *For any linear operator* L *(in particular, for* $L = \nabla_{A,B}$ *and* $L = \Delta_{A,B}$ *for any pair matrices* A *and* B*), we have* $L(aM + bN) = aL(M) + bL(N)$.

Theorem 1.4.

$$\nabla_{A,C}(MN) = \nabla_{A,B}(M)N + M\nabla_{B,C}(N),$$

$$\Delta_{A,C}(MN) = \Delta_{A,B}(M)N + AM\nabla_{B,C}(N).$$

Theorem 1.5. *Let a pair of $n \times \alpha$ matrices G and H form a $\Delta_{A,B}$-generator of length α for a nonsingular matrix M. Write $M^{-1}G = U$, and $H^T M^{-1} = W^T$. Then $\nabla_{B,A}(-M^{-1}) = UW^T$.*

Proof. Note that $M^{-1}\nabla_{A,B}(M)M^{-1} = M^{-1}A - BM^{-1} = M^{-1}GH^TM^{-1} = \nabla_{B,A}(-M^{-1})$. □

Theorems 1.3-1.5 define the operator and generator matrices of the output in terms of the ones of the input, and we may have various options for the operators L and for L-generator matrices associated with the same input matrix. When we choose an operator L for a given class of structured matrices, we may care most about decreasing the multiplication cost bounds $v_{\alpha,n}(L)$. For numerical computation, L-generator matrices may be chosen to better numerical stability requirement, and here the *orthogonal (SVD-based) representation* seems to be most effective. That is, compute the SVD of the image $L(M)$,

$$L(M) = U\Sigma^2 V^T, \tag{1.11}$$

$$U^*U = V^*V = I_r, \quad \Sigma = diag(\sigma_1, \dots, \sigma_r),$$
$$\sigma_1 \geq \dots \geq \sigma_r \geq 0, \quad r = rankL(M), \tag{1.12}$$

U and V are $m \times n$ and $n \times n$ matrices respectively,

$$\sigma_1^2 = \sigma_1^2(M), \dots, \sigma_r^2 = \sigma_r^2(M) \tag{1.13}$$

denoting the singular values of the matrix M, and write

$$G = U\Sigma, \quad H = V\Sigma. \tag{1.14}$$

The SVD computation is quite inexpensive in this case, involving $O((m + n)r^2)$ flops (see [P93]). These recipe and techniques were first used in [P92], [P93], [P93a] and were recently revived in [PBRZ99].

1.4 Our subject

We assume an available crude initial approximation to $\nabla_{A,B}(M^{-1})$, supplied, for example, by the preconditioned conjugate gradient method, which converges to the output rather slowly, with the linear rate. The approximations can be rapidly refined by means of some well known methods of matrix computation, in particular Newton's iteration for matrix inversion:

$$X_{i+1} = 2X_i - X_iMX_i, \quad i = 0, 1, \dots \tag{1.15}$$

such that $I - X_{i+1}M = (I - X_iM)^2$, and/or its extensions [PS91]. Such methods have higher order of convergence, but involve the costly operations of matrix multiplication in each iteration step.

In the case of structured matrices, however, these methods become much more effective because matrix multiplication is inexpensive, as long as the operand matrices are represented with their short L-generators (see Theorem 1.4). The main technical problem, however, is to control the length of the L-generators in the process of the iteration. Without special treatment, the length grows quite rapidly, but several effective techniques were proposed in [P92], [P93a], [PZHD97] and [PBRZ99] to counter such a mishap in the case of Toeplitz, Toeplitz-like and Cauchy-like matrices.

Our main goal in our thesis is to extend these techniques to various other classes of structured matrices in a unified way.

1.5 Our main results

We present two modifications of Newton's iteration that generalize the idea and the techniques of [P92], [P93a] and [PBRZ99], where Newton's iteration was modified to preserve the initial displacement structure of a Toeplitz-like input matrix during the iteration. One idea, due to [P92], was to control the growth of the length of short displacement generators by periodically chopping-off their components corresponding to the smallest singular values in the SVD of these generators. Note that $r = rank\nabla_{B,A}(M^{-1}) = rank\nabla_{A,B}(M)$ by Theorem 1.5, so that we will not deviate much from X_i and M^{-1} if we approximate $\nabla_{B,A}(X_i)$ by a closest matrix $\nabla_{B,A}(Y_i)$ of rank r and then recover Y_i and substitute for X_i in (1.15). We will apply this idea to cover various classes of structured matrices in a unified way.

Our second version of unified Newton's iteration extends [PZHD97], [PBRZ99]. We will control the length of the associated generators but avoid the SVD computation. Namely, we recall Theorem 1.5, substitute X_i for M^{-1}, write $\nabla_{B,A}(Y_i) = (-X_iG)(H^TX_i) = G_yH_y^T$, compute G_y, H_y, recover Y_i, and substitute for X_i in (1.15).

We estimate the convergence rate and arithmetic computational complexity of both our algorithms in a unified way for various classes of structured matrices and then specialize the algorithms to some important specific classes.

Our estimates involve the norm of the inverse operator L^{-1}, which expresses a matrix M via $L(M)$ (cf. Theorem 1.1), and we propose some novel techniques for estimating this norm for various classes of structured matrices. Our next research topic is the study of more refined processes of iterative improvement (cf. [PS91]) specialized to structured matrices with the goal to enhance further the power of our Newton-structured matrix algorithms.

The presented algorithms require an initial approximation lying sufficiently close to the solution, otherwise the iterations may converge too slowly or diverge. Here, we assume such an approximation available but we plan to relax this assumption by extending the homotopy techniques of [P92] from the Toeplitz-like to various matrix structures.

Our results have been reported in the papers [PR2000], [PRW2000], [PRWa].

Remark 1.1. *Instead of the unification, one may use some known transformations to Toeplitz-like or Cauchy-like case to extend the cited successful algorithms to other classes of structured matrices. This is a special case of the general idea of extending successful algorithms from one class to other classes of structured matrices. The idea was proposed in [P90] together with the sample transformations among Toeplitz-like, Hankel-like, Vandermonde-like and Cauchy-like matrices. The approach turned out to be useful for the design of practical Toeplitz-like solvers, for handling singularities, for polynomial interpolation and multipoint evaluation, and in the computational complexity analysis of structured matrix operations [PLST93], [H95], [GKO95], [PZHY97], [P2000]. Our application of Newton's iteration directly to the input matrix, however, simplifies the computation by removing the transfor-*

mation stage, prone to numerical stability problems.

1.6 Organization of our thesis

We organize our thesis as follows. In the next section, we state some basic definitions and assumptions. In section 3, we briefly compare Newton's iteration for general and structured matrices. In section 4, we very briefly outline our first algorithm. In section 5, we present this algorithm in some detail. We define a unified Newton-structured matrix iteration (associated with operators of Sylvester type), where we bound the L-rank of Newton's iterates by zeroing the smallest singular values of the matrix $L(M)$. We also estimate the convergence rate and the computational complexity of the resulting algorithm. In section 6, we present our second algorithm. That is, we define a unified Newton-structured matrix iteration (associated with operators of Sylvester type), where we bound the L-rank of Newton's iterates avoiding the SVD computation. In section 7, we estimate the convergence rate and the computational complexity of this algorithm. In section 8, we extend both algorithms to the case of singular operators. In section 9, we apply our algorithms to the Stein type case. In section 10, we state some basic preliminaries for estimating the norm of the inverse operator. In sections

11-13, we estimate the norm ν^- of the inverse operator L^{-1}, which is the main ingredient of our estimates for the output errors and the convergence rate of our algorithms. We present these techniques for estimating ν^-, both in general and for the specific operators L associated with specific classes of structured matrices.

2 Some Definitions

We will need some further definitions in addition to the ones mentioned in the introduction.

Definition 2.1. $||M||$ *denotes any fixed operator norm of a matrix* M. $||M||_2, ||M||_1$, *and* $||M||_\infty$ *are the special cases denoting the 2-norm, the (column) 1-norm, and the (row)* ∞*-norm, respectively (cf. [BP94], [GL96]), that is,* $||M||_l$ *is the* l*-norm,* $l = 1, 2, \infty$. $\kappa(M) = cond_2(M) = \sigma_1^2(M)\sigma_r^2(M)$, *where* $\sigma_i^2(M)$ *is the* i*-th singular value of* M, $i = 1, \dots, r, r = rank M$.

Theorem 2.1. *(cf. [BP94], [GL96]). We have* $||M||_2 = \sigma_1^2(M)$ *for every matrix* M *and* $\kappa(M) = ||M||_2||M^{-1}||_2$ *for an* $n \times n$ *nonsingular matrix* $M = (m_{i,j})$*.(cf. (1.11)-(1.13)). Furthermore we have* $||M||_l/\sqrt{n} \leq ||M||_2 \leq |M||_l\sqrt{n}, l = 1, \infty$; $||M||_1 = ||M^T||_\infty = \max_j \sum_i |m_{i,j}|$. *Moreover, if* $M \in \mathbb{R}^{n\times n}$*, that is, if the matrix* M *has real entries, then* $||M||_2^2 \leq ||M||_1||M||_\infty$. $\kappa(M) = cond_2(M) = \sigma_1^2(M)/\sigma_r^2(M)$ *where* $\sigma_i^2(M)$ *is the* i*-th singular value of* $M, r = rank M$.

Definition 2.2. *A linear operator L of Sylvester type and/or Stein type is called* nonsingular *if it can be inverted, that is, if for any fixed matrix pair (G, H) of (1.5) there exists a unique matrix M having its L-generator (G, H) of length α, that is, satisfying equations (1.3)-(1.5). Such an operator $L = \nabla_{A,B}$ or $L = \Delta_{A,B}$ is called* regular *if it is nonsingular and if (1.3)-(1.5) imply that the matrix M can be expressed as a* unique *bilinear or trilinear matrix form with at most α terms in the entries of the L-generator matrices G and H.*

Definition 2.3. *We define the* norms *of the linear operator L and (if this operator is nonsingular) of its inverse L^{-1}:*

$$\nu = \nu_{\alpha,l}(L) = \sup_M(\|L(M)\|_l/\|M\|_l),$$

$$\nu^- = \nu_{\alpha,l}(L^{-1}) = \sup_M(\|M\|_l/\|L(M)\|_l),$$

where the supremum is over all matrices M having positive L-rank at most α. We also define the condition number *of the operator L as*

$$\chi = \chi(L) = cond(L) = \nu\nu^- = \nu_{\alpha,l}(L)\nu^-_{\alpha,l}(L).$$

Assumption 2.1. *Except for section 9, we will assume with regular operators L only.*

3 Newton's Iteration and the Unified Newton-Structured Matrix Iteration

Newton's iteration

$$X_{i+1} = X_i(2I - MX_i), \quad i = 0, 1, \dots, \tag{3.1}$$

is a well known technique for rapid improvement of a crude initial approximation, X_0, to the inverse of a matrix M. Indeed, the above matrix equations imply that

$$I - MX_{i+1} = (I - MX_i)^2$$

and, therefore,

$$||I - MX_{i+1}|| \leq ||I - MX_i||^2$$

for all i. That is, we have quadratic convergence if $||I - MX_0|| < 1$. Various modifications of this iteration allow its faster convergence, and various policies are known for efficient choices of the initial approximation X_0 [IK66], [PS91].

In the case of a general input matrix M, the computational cost of the iteration as well as its known modifications is high because matrix products must be computed at each step. For structured matrices, matrix multiplication has a lower computational cost of $m_\alpha(L)$ (see Theorem 1.2), because one may operate with short generators rather than with the more numerous entries of the input and output matrices of each operation. In particular, for structured matrices, M and X_0, having short $\nabla_{A,B}$- and $\nabla_{B,A}$-generators, respectively, the iteration can be performed by operating with some short ∇-generators of the matrices M, X_i, and MX_i (or X_iM), to decrease dramatically the computational time and memory space used in the above classical version of Newton's iteration. This, however, requires some special techniques for controlling the length of the $\nabla_{B,A}$-generators of X_i, which otherwise tend to be tripled at every iterative step. Similar comments apply where $\Delta_{A,B}$- and $\Delta_{B,A}$-generators are used. In the Toeplitz-like case, two approaches to such a control are described in [P92], [P93], [P93a], and [PBRZ99]. The resulting two processes are called *Newton-Toeplitz iteration* in [PBRZ99]. In [PZHD97] one of the approaches was developed for Cauchy-like matrices. We will extend the two approaches to a more general class of structured matrices in a unified way and will call the resulting two processes the *Unified Newton-Structured Matrix Iteration* for matrix inversion.

Let us outline it assuming that M and X_0 are structured matrices with $rank\nabla_{A,B}(M) = rank\nabla_{B,A}(M^{-1})) = \alpha$ (cf. Theorem 1.5), $rank\nabla_{B,A}(X_0) = \beta_0 \leq 3\alpha$ where α and β_0 are small relative to n. We will modify Newton's iteration to ensure that for every i the matrix X_i has a $\nabla_{B,A}$-generator of length at most 3α. By assumption, $rank \nabla_{B,A}(M^{-1}) = \alpha$. Hence the matrices X_i, which approximate M^{-1} closely for larger i, have a nearby matrix of $\nabla_{B,A}$-rank α. This fact suggests the following approach (cf. [P92], [P93a], [PBRZ99]) to decreasing the computational cost of the iterative scheme: shift from X_i to a nearby matrix Y_i having $\nabla_{B,A}$-rank at most α and then restart the iteration with Y_i instead of X_i. Here is a formal description of this approach for the Sylvester type operators. (The same ideas work similarily for the Stein type operators $\Delta_{A,B}$ and $\Delta_{B,A}$.)

Algorithm 3.1. *Unified Newton-Structured Matrix Iteration for the Sylvester Type Operators.*

Input: A positive integer α, a pair of $n \times n$ matrices A and B (defining a regular operator $\nabla_{A,B}$), an $n \times n$ nonsingular structured matrix M having $\nabla_{A,B}$-rank α and defined by its $\nabla_{A,B}$-generator (G, H) of length α (cf. (1.3), (1.5)), a sufficiently close initial approximation Y_0 to M^{-1} given with its $\nabla_{B,A}$-generator of length at most α, a bound on the number of iteration steps (such a bound will be specified in Corollary 5.1), and a Subroutine **R** for the transition from a $\nabla_{A,B}$-generator of length at most 3α, for an $n \times n$ matrix approximating M^{-1} to an $\nabla_{A,B}$-generator of length at most α for a nearby matrix.

Output: A $\nabla_{B,A}$–generator of length at most α for a matrix Y_{l+1} approximating M^{-1}.

Computations: Recursively compute $\nabla_{B,A}$-generators of length at most 3α and α for the matrices

$$X_{i+1} = Y_i(2I - MY_i), \quad i = 0, 1, \ldots, l, \tag{3.2}$$

and $\nabla_{B,A}$-generators of length at most α for the matrices Y_i defined by a transformation from X_i by means of the Subroutine **R**.

Proposition 3.1. *Under the assumptions of Algorithm 3.1, for any* $i = 0, 1, 2, \ldots,$ *a* $\nabla_{B,A}$*-generator of length at most* 3α *for the matrix* $X_{i+1} = 2Y_i - Y_i M Y_i$ *can be computed at the cost of performing* $O(M_\alpha) = O(\alpha V_\alpha) = O(\alpha^2 n \log^d n)$ *flops.*

Proof. Write $Y = Y_i, X = X_{i+1} = 2Y - YMY$, and observe that $\nabla_{B,A}(X) = \nabla_{B,A}(2Y - MY)$. Combine the equations

$$\nabla_{B,A}(X) = \nabla_{B,A}(Y)(2I - MY) - Y\nabla_{A,B}(M)Y - YM\nabla_{B,A}(Y),$$

$$\nabla_{B,A}(X) = G_X H_X^T, \quad \text{and} \quad \nabla_{B,A}(Y) = G_Y H_Y^T$$

to obtain that

$$G_X^T = \left(G_Y^T, (YG)^T, (YMG_Y)^T\right),$$

$$H_X^T = \left(H_Y^T(2I - MY), -H^T Y, -H_Y^T\right).$$

Because of the assumed regularity of operator $\nabla_{B,A}$, these observations imply Proposition 3.1. □

To complete the description of the Unified Newton-Structured Matrix Iteration, it remains to specify the Subroutine **R**, which controls over the length of the computed L-generators. We will do this in two ways, to be specified in sections 4 and 6.

4 A Variant of the Subroutine R for Controlling the Length of L-generators of Approximate Inverses: Zeroing the Smallest Singular Values

In this section we will rely on the following result, which will enable us to control the length of an L-generator for a matrix Y_i lying near X_i and M^{-1} (cf. [GL96], pp. 72, 230):

Theorem 4.1. *Given a matrix W of rank r, it holds that*

$$\sigma_r^2 = \min_{rank B \leq r-1} \|W - B\|_2,$$

that is, the r-th singular value of W is equal to the minimal 2-norm of the error of approximation of W by a matrix of rank less than r.

We will represent the image matrix $\nabla_{B,A}(X_i)$ via its SVD, which defines the unique orthogonal $\nabla_{B,A}$-generator for X_i, and then will se to zero its smallest singular values leaving at most α of them nonsingular and thus obtaining a shorter $\nabla_{B,A}$-generator of length at most α for a nearby matrix Y_i. Both Y_i and M lie near X_i, and we have $\|\nabla_{B,A}(X_i) - \nabla_{B,A}(Y_i)\|_2 \leq \|\nabla_{B,A}(X_i) - \nabla_{B,A}(M^{-1})\|_2$ by Theorem 4.1 because $rank\nabla_{B,A}(M^{-1}) \leq \alpha$, and thus implies that also Y_i lies near X_i. To specify and to analyze formally the transition from the matrices X_i to Y_i, we will use some further definitions and simple preliminary results.

Hereafter, we will write $\beta = \beta_i = rank\nabla_{B,A}(X_i)$. We will prove shortly that $\beta \leq 3\alpha$ (for all i) in our case. By the definition of Y_i, we have $rank\nabla_{B,A}(Y_i) = \alpha$ for all i. We also have by Theorem 1.5 that $rank\nabla_{B,A}(M^{-1}) = \alpha$. Let us write

$$e_{l,i} = \|X_i - M^{-1}\|_l, \quad l = 1, 2, \infty, \tag{4.1}$$

$$e_i = \|X_i - M^{-1}\|, \tag{4.2}$$

$$\hat{e}_i = \|Y_i - M^{-1}\|. \tag{4.3}$$

Now, we are ready to describe Variant 1 of Subroutine **R** for Algorithm 3.1.

Subalgorithm 4.1. *Variant 1 for the Control of the Generator Length: Zeroing the Smallest Singular Values.*

Input: A positive integer α, operator matrices A and B, a $\nabla_{A,B}$-generator of length at most α for a nonsingular $n \times n$ matrix M where $\alpha = rank\nabla_{A,B}(M) = rank\nabla_{B,A}(M^{-1})$, and a $\nabla_{B,A}$-generator (G_i, H_i) of length at most $\beta = \beta_i$, for a matrix X_i such that $\alpha \leq \beta, \nabla_{B,A}(X_i) = G_i H_i^T$.

Output: A $\nabla_{B,A}$-generator of length at most α for a matrix Y_i such that

$$||Y_i - M^{-1}||_2 \leq (1 + (||A||_2 + ||B||_2)\nu^-)e_{2,i} \tag{4.4}$$

for $e_{2,i}$ of (4.2) and $\nu^- = \nu_{2,\alpha}(\nabla_{A,B})$ of Definition 2.3.

Computations:

a. Compute the SVD of the displacement $\nabla_{B,A}(X_i) = U_i \Sigma_i^2 V_i^T$.

b. Set to zero the diagonal entries $\sigma_{\alpha+1}^2, \ldots, \sigma_\beta^2$ of the matrix Σ_i, thus turning Σ_i into a diagonal matrix of rank α. ($\sigma_{\alpha+1}^2, \ldots, \sigma_\beta^2$ are the $\beta - \alpha$ smallest singular values of the matrix $\nabla_{B,A}(X_i)$.)

c. Compute and output the matrices G_i^* and H_i^* obtained from the matrices $U_i\Sigma_i$ and $V_i\Sigma_i$, respectively, by deleting their last $\beta - \alpha$ columns.

Correctness of Subalgorithm 4.1 is implied by the following result, which shows that bound (4.4) holds under our assumptions on the input of Algorithm 3.1 and Subalgorithm 4.1.

Theorem 4.2. *Let the structured matrices M^{-1}, X_i, and Y_i be defined as above and let a positive scalar $e_{2,i}$ be defined by equation (4.2). Let $\nabla_{B,A}$ be a regular linear operator associated with the matrix specified for M. Then bound (4.4) is satisfied.*

Theorem 4.2 generalizes a result of [P92], [P93], and [P93a] proved for the Toeplitz-like case. To prove Theorem 4.2, we need the next two propositions.

Proposition 4.1. *Under the notation of Algorithm 3.1, we have*

$$\|\nabla_{B,A}(X_i) - \nabla_{B,A}(Y_i)\|_2 = \sigma^2_{\alpha+1}(\nabla_{B,A}(X_i)), \tag{4.5}$$

$$\|\nabla_{B,A}(X_i) - \nabla_{B,A}(M^{-1})\| \leq (\|A\| + \|B\|)e_i, \tag{4.6}$$

for e_i of (4.2).

proof Equation (4.5) follows immediately from Theorem 2.1. To prove bound (4.6), recall that

$$\nabla_{B,A}(M^{-1}) = M^{-1}B - AM^{-1}, \quad \nabla_{B,A}(X_i) = X_iB - AX_i.$$

Therefore,

$$\begin{aligned}
&\|\nabla_{B,A}(X_i) - \nabla_{B,A}(M^{-1})\| \\
= \ & \|X_iB - AX_i - M^{-1}B + AM^{-1}\| \\
= \ & \|(X_i - M^{-1})B - A(X_i - M^{-1})\| \\
\leq \ & \|X_i - M^{-1}\| \cdot \|B\| + \|A\| \cdot \|X_i - M^{-1}\| \\
\leq \ & (\|A\| + \|B\|)\|X_i - M^{-1}\| = (\|A\| + \|B\|)e_i.
\end{aligned}$$

□

Proposition 4.2. $\|\nabla_{B,A}(X_i) - \nabla_{B,A}(Y_i)\|_2 \leq (\|A\|_2 + \|B\|_2)e_{2,i}$.

proof Apply the well known estimate ([GL96], p.428) and deduce that

$$|\sigma_i^2(\nabla_{B,A}(X_i)) - \sigma_i^2(\nabla_{B,A}(M^{-1}))| \leq \|(\nabla_{B,A}(X_i)) - (\nabla_{B,A}(M^{-1}))\|_2$$

for all i, where $\sigma_i^2(W)$ is defined by (1.11)-(1.13). For all $i > \alpha$, recall that $\sigma_i^2(\nabla_{B,A}(M^{-1})) = 0$ and obtain

$$\sigma_i^2(\nabla_{B,A}(X_i)) \leq \|\nabla_{B,A}(X_i) - \nabla_{B,A}(M^{-1})\|_2.$$

Now, substitute inequality (4.6) and deduce that

$$\sigma_i^2(\nabla_{B,A}(X_i)) \leq (\|A\|_2 + \|B\|_2)e_{2,i} \text{ for } i > \alpha.$$

Combine this bound for $i = \alpha + 1$ with equation (4.5) and deduce Proposition 4.2. □

Now, we are prepared to prove Theorem 4.2.

Proof of Theorem 4.2: We have $||M^{-1}-Y_i||_2 \leq ||M^{-1}-X_i||_2+||X_i-Y_i||_2$. By first applying Definition 2.2 for $l = 2$ and $L = \nabla_{B,A}$ and then applying the linearity of the operator $\nabla_{B,A}$, we obtain that $||X_i - Y_i||_2 \leq \nu^-||\nabla_{B,A}(X_i - Y_i)||_2 \leq ||\nabla_{B,A}(X_i) - \nabla_{B,A}(Y_i)||_2$. Substitute equation (4.1) for $l = 2$, that is, $e_{2,i} = ||X_i - M^{-1}||_2$, substitute the latter bound on $||X_i - Y_i||_2$ and the one of Proposition 4.2, and obtain $||M^{-1}-Y_i||_2 \leq e_{2,i}+||\nabla_{B,A}(X_i)-\nabla_{B,A}(Y_i)||_2\nu^- \leq e_{2,i} + (||A||_2 + ||B||_2)e_{2,i}\nu^-$. □

5 Unified Newton-Structured Matrix Iteration I: Its Convergence Rate and Computational Complexity Estimates

Combining Algorithm 3.1 with Subalgorithm 4.1 applied as a Subroutine **R** defines Unified Newton-Structured Matrix Iteration I. Next, we will estimate its convergence rate and the computational complexity of its performance. Our study extends to various classes of structured matrices a similar study presented in [P92], [P93], [P93a], and [PBRZ99] for Toeplitz-like matrices.

Let us write

$$p_i = \|I - Y_i M\|_2, \quad i = 0, 1, 2, \ldots$$

We have

$$I - X_{i+1}M = (I - Y_i M)^2, \quad \|I - X_{i+1}M\|_2 \le p_i^2,$$

and therefore, $\|M^{-1} - X_{i+1}\|_2 \le p_i^2 \|M^{-1}\|_2$. By Theorem 4.2, we have

$$\|M^{-1} - Y_{i+1}\|_2 \le (1 + (\|A\|_2 + \|B\|_2)\nu^-)\|M^{-1} - X_{i+1}\|_2.$$

Consequently, we have

$$\|M^{-1} - Y_{i+1}\|_2 \le (1 + (\|A\|_2 + \|B\|_2)\nu^-)p_i^2\|M^{-1}\|_2,$$

and therefore,

$$\begin{aligned} p_{i+1} = \|I - Y_{i+1}M\|_2 &\le \|M^{-1} - Y_{i+1}\|_2\|M\|_2 \\ &\le (1 + (\|A\|_2 + \|B\|_2)\nu^-)p_i^2\|M^{-1}\|_2\|M\|_2 \\ &\le (1 + (\|A\|_2 + \|B\|_2)\nu^-)p_i^2\kappa(M), \end{aligned}$$

where $\kappa(M) = cond_2(M) = \|M^{-1}\|_2\|M\|_2$ (cf. Theorem 2.1).

Now, suppose that for a positive $\theta, \theta \le 1$, we have

$$(1 + (\|A\|_2 + \|B\|_2)\nu^-)p_i^{1-\theta}\kappa(M) \le 1 \text{ for } i = 0, 1, 2, \ldots. \tag{5.1}$$

This bound implies a convergence rate of $1+\theta$, that is,

$$p_{i+1} \leq p_i^{1+\theta} \text{ for all } i. \tag{5.2}$$

Next, observe that if $p_0 \leq 1$, then $p_i \leq 1$ and satisfies (5.1). Then (5.2) implies that $p_{i+1} \leq p_i$. Therefore, it is sufficient to assume (5.1) with p_i replaced by p_0. The next theorem summarizes our analysis.

Theorem 5.1. *(Convergence rate). Under Definition 2.3 and the stated assumptions on the input of Unified Newton-Structured Matrix Iteration I, (that is, Algorithm 3.1 combined with Subalgorithm 4.1) let the matrices X_i and M be given with their $\nabla_{B,A}$- and $\nabla_{A,B}$-generators of length β_0 and α_0, respectively. Furthermore, let*

$$(1+(\|A\|_2+\|B\|_2)\nu^-)p_0^{1-\theta}\kappa(M) \leq 1 \textit{ for } i=0,1,2,\ldots \tag{5.3}$$

for $p_0 = \|I - X_0 M\|_2, \nu^- = \nu_{2,\alpha}(\nabla_{A,B})$ of Definition 2.3, and some fixed positive θ. Then for all positive i, we have $\mathrm{rank}\nabla_{B,A}(Y_i) \leq \alpha$, $\|Y_i - M^{-1}\|_2 \leq p_0^{(1+\theta)^i}\|M^{-1}\|_2$.

Estimating the computational cost of performing Algorithm 3.1, we will rely on Definition 2.3, Theorem 5.1 and the bound $O(n\alpha^2)$ on the cost of computing the SVD of the product of a pair of $n \times \alpha$ by $\alpha \times n$ matrices [P93], [GL96]. (The latter bound should actually include the term $\alpha \log\log(1/\delta) \log \alpha$, where δ is the output approximation error bound for the

SVD; we will ignore this term assuming realistically that $\log\log(1/\delta)\log\alpha = O(n\alpha)$.)

Theorem 5.2. *Under Definition 2.3 and the assumption (5.3) on the input of Unified Newton-Structured Matrix Iteration I (that is, of Algorithm 3.1 and Subalgorithm 4.1), the matrices $X_1, Y_1, X_2, Y_2, \dots, X_i, Y_i$ can be computed by performing $O((m_\alpha + \alpha^2 n)i) = O(i\alpha^2 n \log^d n)$ flops for any fixed natural i and for $m_\alpha = O(\alpha^2 n \log^d n)$ and d of Assumption 1.1.*

We can see that unless the matrix M is very ill-conditioned, inequality (5.3) is a mild assumption on the residual norm p_0. This assumption is sufficient to ensure a rapid improvement of the initial approximation of M by X_0 at a low computational cost of performing i steps of Algorithm 3.1 by using $O(im_\alpha) = O(i\alpha^2 n \log^d n)$ flops.

Corollary 5.1. *Under the assumptions of Theorem 5.1, the residual norm bound $p_l = \|I - Y_l M\|_2 \le \epsilon\kappa(M)$ is ensured in*

$$l = \lceil \log_{1+\theta}(\log\epsilon / \log p_0) \rceil$$

steps of Unified Newton-Structured Matrix Iteration I, (that is of Algorithm 3.1 combined with Subalgorithm 4.1). These steps require $O(lm_\alpha) = O(l\alpha^2 n \log^d n)$ flops for m_α and d of Assumption 1.1.

Remark 5.1. *One may develop a* dynamic version *of Newton-Structured Matrix Iteration I, where the level of the cut-off in zeroing the smallest singular value may change with i, to optimize the overall flops count (this was* *poposed in [P92] and [BM,a] in the Toeplitz-like case). In particular, the recipe of [BM,a] is to zero all the singular values of $\nabla_{B,A}(X_i)$ that are less than a fixed tolerance ϵ. Because the residual norms $p_i \leq p_0$ are assumed to be small for all i, this recipe should normally lead to exactly the same matrix Y_i as the one output by our Subalgorithm 4.1. The only exception is the matrices X_i for which there exist approximation matrices B satisfying simultaneously the two following inequalities: $rank \nabla_{B,A}(B) < \alpha$, $\|\nabla_{B,A}(B - X_i)\|_2\| \leq \epsilon$.*

6 Variant II of Subroutine R for Controlling the Generator Length for the Inverse: Substitution of an Approximate Inverse into an Expression for $\nabla_{B,A}(M^{-1})$

Let us describe an SVD-free method of controlling the length of an L-generators of approximate inverses. By Theorem 1.5, $\nabla_{B,A}(M^{-1}) = -M^{-1}GH^TM^{-1}$. We substitute X_i for M^{-1} on the right-hand side and define a short $\nabla_{B,A}$-generator for Y_i as follows:

$$\nabla_{B,A}(Y_i) = U_i W_i^T, \tag{6.1}$$

where $U_i = -X_i G \in \mathbb{C}^{n \times \alpha}, W_i^T = H^T X_i \in \mathbb{C}^{\alpha \times n}$. This leads us to the following variant of Subroutine **R** of Algorithm 3.1.

Subalgorithm 6.1. *Variant II for the Control of the Generator Length for Approximate Inverses.*

Input: A positive integer α, a pair of $n \times n$ operator matrices A and B defining a regular operator $\nabla_{B,A}$, a $\nabla_{A,B}$-generator for a nonsingular $n \times n$ matrix M of length at most α where $\alpha = rank\nabla_{A,B}(M) = rank\nabla_{B,A}(M^{-1})$, and a $\nabla_{B,A}$-generator G_{i+1}, H_{i+1} of length at most 3α for a matrix X_{i+1} of equation (3.2).

Output: A $\nabla_{B,A}$-generator (U_{i+1}, W_{i+1}) of length at most α for a matrix Y_{i+1} such that

$$\hat{e}_{i+1} = ||Y_{i+1} - M^{-1}|| \leq C_i e_i \tag{6.2}$$

for $\hat{e}_i$ of (4.3), e_i of (4.2), $C_i = \nu^- ||GH^T||(e_i + 2||M^{-1}||)$, and a constant $\nu^- = \nu_\alpha(\nabla_{B,A}^{-1})$ of Definition 2.3.

Computations: Compute and output the matrix products $U_{i+1} = -X_{i+1}G$, $W_{i+1}^T = H^T X_{i+1}$ of (6.2).

Under Assumption 1.1 about regularity of the operator $\nabla_{B,A}$, the matrix pair (U_{i+1}, W_{i+1}) is a $\nabla_{B,A}$-generator of length at most α for a matrix Y_{i+1}, which is a unique solution to the equation (cf. (6.1))

$$\nabla_{B,A}(-Y_{i+1}) = U_{i+1} W_{i+1}^T,$$

and the computation of the $n \times \alpha$ matrices U_{i+1}, W_{i+1} is reduced to multiplication of the matrix X_{i+1} by the $n \times 2\alpha$ matrix $(-G, H)$, at the cost$O(\alpha m_{\alpha,n} = O(\alpha^2 n \log^d n)$ for $m_{\alpha,n}$ and d of Assumption 1.1.

To prove correctness of the subalgorithm, that is, to prove bound 5.3), we need some auxiliary results. Recall the matrix equations $U = M^{-1}G$ and $W^T = H^T M^{-1}$ of Theorem 1.5 and deduce that

$$U_i = X_i G = (X_i - M^{-1})G + M^{-1}G,$$

$$W_i^T = H^T X_i = H^T(X_i - M^{-1}) + H^T M^{-1}.$$

Now, write $E_i = U_i W_i^T - U W^T$ and obtain the following matrix equation

$$E_i = (X_i - M^{-1})GH^T(X_i - M^{-1}) + M^{-1}GH^T(X_i - M^{-1}) + (X_i - M^{-1})GH^T M^{-1}.$$

Proposition 6.1. *Let matrices U_i, W_i^T, and E_i be defined as above. Then*

$$||E_i|| = ||U_i W_i^T - U W^T|| \le ||GH^T|| e_i(e_i + 2||M^{-1}||),$$

for $e_i = ||X_i - M^{-1}||$ of (4.2).

pf The proposition follows from the above expression for E_i. □

Proposition 6.2. *For a nonsingular matrix M, X_{i+1} defined by equation (3.2), and Y_i of Algorithm 6.1, for $i = 0, 1, \ldots$, we have $e_{i+1} \le ||M|| \hat{e}_i^2$ for $e_{i+1} = ||X_{i+1} - M^{-1}||$ and $\hat{e}_i = ||Y_i - M^{-1}||$ (of (4.2), (4.3)).*

proof By (3.2), we have $I - MX_{i+1} = (I - MY_i)^2, i = 0, 1, \ldots$. It follows that $e_{i+1} = \|X_{i+1} - M^{-1}\| = \|M^{-1}(I - MX_{i+1})\| = \|M^{-1}(I - MY_i)^2\| = \|(M^{-1} - Y_i)M(M^{-1} - Y_i)\| \leq \|M\|\hat{e}_i^2$. □

Theorem 6.1. *For $i = 1, 2, 3, \ldots$, we have $\hat{e}_i \leq C_i e_i$ and $e_{i+1} \leq (C_i e_i)^2\|M\|$ for e_i of (4.3), $\hat{e}_i$ of (4.4), $C_i = \nu^- \|GH^T\|(e_i + 2\|M^{-1}\|)$, and a constant ν^- of Definition 2.3.*

Proof. $\hat{e}_i = \|Y_i - M^{-1}\| \leq \nu^- \|\nabla_{B,A}(Y_i - M^{-1})\|$ (see Definition 2.3). Since the operator $\nabla_{B,A}$ is linear, we have $\hat{e}_i \leq \nu^- \|\nabla_{B,A}(Y_i) - \nabla_{B,A}(M^{-1})\| \leq \nu^- \|U_i W_i^T - UW^T\| \leq \nu^- \|E_i\|$. At this point, apply Proposition 6.1 and obtain that $\hat{e}_i \leq C_i e_i$. Now, since $e_{i+1} \leq \|M\|\hat{e}_i^2$ by Proposition 6.2, we deduce that $e_{i+1} \leq (C_i e_i)^2 \|M\|$. □

7 Unified Newton-Structured Matrix Iteration II: Its Convergence Rate and Computational Complexity Estimates

Combining Algorithm 3.1 with Subalgorithm 6.1 (applied as Subroutine **R**) defines Unified Newton-Structured Matrix Iteration II. Our next goal is to estimate its convergence rate and its computational complexity.

Let us first restate Theorem 6.1 in a more constructive way – by replacing the values $e_0, \|M^{-1}\|$ and C_i by more readily available values. We will write

$$r_0 = \|I - MY_0\|, \tag{7.1}$$

$$e_0^* = r_0\|Y_0\|/(1 - r_0) \tag{7.2}$$

and will assume realistically that

$$r_0 \le 1, \tag{7.3}$$

$$e_i \le \|M^{-1}\|, \tag{7.4}$$

for e_i of (4.2) and for all i.

Proposition 7.1. *Assuming (7.1)-(7.4), we have*

$$\hat{e}_0 \le e_0^*, \tag{7.5}$$

$$\|M^{-1}\| \le \|Y_0\|/(1 - r_0), \tag{7.6}$$

$$C_i \le C = 3\nu^-\|GH^T\| \cdot \|Y_0\|/(1 - r_0), \forall i. \tag{7.7}$$

Proof. We have $|M^{-1}\| - \|Y_0\| \le \|M^{-1} - Y_0\| \le \|M^{-1}\|/r_0$, assuming (7.3), and now (7.6) follows. Furthermore, $\hat{e}_0 = \|Y_0 - M^{-1}\| \le r_0 M^{-1}\|$. Substitute (7.6) and obtain (7.5). Substitute (7.3) into the expression of Theorem 6.1 for C_i, then substitute (7.6), and obtain (7.7). □

By combining Theorem 6.1 and Proposition 7.1, we obtain

Theorem 7.1. *Assume (7.1)-(7.4). Then we have* $\hat{e}_i \leq Ce_i, e_{i+1} \leq (Ce_i)^2\|Y_0\|/(1-r_0)$ *for* C *of (7.7)* e_i *of (4.2),* $\hat{e}_i$ *of (4.3) and* $i = 1, 2, \ldots$.

Corollary 7.1. *Assume (7.1)-(7.4) and let* $e_1 \leq 1$ *and* $(e_1)^{1-\theta}\|M\|C^2 \leq 1$ *for* $0 < \theta \leq 1$. *Then* $e_{i+1} \leq (e_i)^{1+\theta} \leq \ldots \leq (e_1)^{(1+\theta)^i}$, *for* $i = 1, 2, \ldots$.

By applying Proposition 6.2 and then bound (7.5) we obtain $e_1 \leq \hat{e}_0^2\|M\| \leq (e_0^*)^2\|M\|$. Substitute the latter bound into Corollary 7.1 and obtain

Corollary 7.2. *Assume relations (7.1)-(7.4) and the bounds* $(e_0^*)^2\|M\| \leq 1, \theta \leq 1, (e_0^*)^{1-\theta}\|M\|^{1-\theta/2}C \leq 1$ *for* C *of (7.7). Then we have* $e_{i+1} \leq ((e_0^*)^2\|M\|)^{(1+\theta)^i}$ *and* $\hat{e}_{i+1} \leq Ce_{i+1}, i = 0, 1, \ldots$.

Corollary 7.3. *Write*

$$i^* = \lceil \log(\log \epsilon^* / \log((e_0^*)^2\|M\|)) / \log(1+\theta) \rceil,$$

$$\hat{i} = \lceil \log(\log \hat{\epsilon} / \log((e_0^*)^2\|M\|)) / \log(1+\theta) \rceil.$$

Then, under the assumptions of Corollary 7.2, $i+1 \geq i^*+1$ *recursive steps of (6.1) and (6.2) are sufficient in order to ensure that* $e_{i+1} = \|X_{i+1} - M^{-1}\| \leq e^*$. *Furthermore,* $i+1 \geq \hat{i}+1$ *such recursive steps are sufficient in order to ensure that* $\hat{e}_{i+1} = \|Y_{i+1} - M^{-1}\| \leq \epsilon C$. *Each recursive step of (6.1) and (6.2) can be performed by using* $O(\alpha m_{\alpha,n}) = O(\alpha^2 n \log^d n)$ *flops for* $m_{\alpha,n}$ *and* $d \leq 2$ *of Assumption 1.1.*

8 Extension to Singular Operators

In some cases structured matrices are studied by using singular operators L. (Most celebrated are singular operators associated with Toeplitz-like matrices, in particular, $\nabla_{Z,Z}$, ∇_{Z^T,Z^T} (cf. [BP93], [BP94]), ∇_{Z_f,Z_f} and $\nabla_{Z_f^T,Z_f^T}$ (cf. [GO94]), with Hankel-like, and Hankel+Toeplitz-like matrices (cf. [BP93], [BP94]).) In such cases, we may apply our Newton-Structured Matrix Iteration in two ways.

1. Associate a regular linear operator to the same class of matrices and recover the respective new L-generators from the old ones.

This is sometimes a very simple task. For instance, given a singular operator ∇_{Z_f,Z_f} and a ∇_{Z_f,Z_f}-generator for a matrix M, we observe that the operator ∇_{Z_e,Z_f} is regular if and only if $e \neq f$ and immediately obtain the ∇_{Z_f,Z_e}-generator for the same matrix M as foolows:

$$\nabla_{Z_e,Z_f}(M) = Z_f M - M Z_e = \nabla_{Z_f,Z_f}(M) + M(Z_f - Z_e),$$

where $M(Z_f - Z_e) = (f - e)M\mathbf{e}_0$, $M\mathbf{e}_0$ is the first column of M.

2. For some singular operators L almost all entries of the matrix M can be recovered from $L(M)$. Then modify the iteration process to recover also the remaining entries of M.

For example, for the operator $\nabla_{Z_f^T, Z_f^T}$ we have the expression

$$M = Z_{f,lc}(Me_{n-1}) + \frac{e}{e-f}\sum_{i=1}^{\alpha} Z_f(Z_f \mathbf{g}_i) Z_{1/e}^T(\mathbf{h}_i)$$

provided that $\nabla_{Z_f^T, Z_f^T}(M) = GH^T$ for G, H of (1.5), $\mathbf{e}_{n-1}$ is the n-th coordinate vector so that $M\mathbf{e}_{n-1}$ is the last column of an $n \times n$ matrix M; $Z_f(\mathbf{v})$ is the f-circulant matrix with the first column $\mathbf{v}$, and $Z_{f,lc}(\mathbf{v})$ is the f-circulant matrix with the last column $\mathbf{v}$ (cf. [GO94], [PBRZ99]). Thus, we have a simple bilinear expresion for an $n \times n$ matrix via its $\nabla_{Z_f^T, Z_f^T}$-generator and its last column.

We now easily extend Algorithm 3.1 (cf. [PBRZ99])provided that together with a short $\nabla_{Z_f^T, Z_f^T}$-generator of the matrices X_{i+1}, we also compute the last columns of X_{i+1} as follows:

$$X_{i+1}\mathbf{e}_{n-1} = Y_i(2I + AY_i)\mathbf{e}_{n-1}, \quad i = 0, 1, \ldots .$$

For both approaches 1 and 2, the estimates for the convergence rate and the computational cost are easily extended from the case of the regular operators such as ∇_{Z_f, Z_e} [PBRZ99]. Moreover, the techniques and analysis of such an extension to the singular case apply to all other irregular bilnear operators $\nabla_{A,B}$ (so far encountered) for the study of Toeplitz-like, Hankel-like, and Toeplitz+Hankel-like matrices.

9 Extension to the Case of the Stein Type Operators

We may extend our algorithms by replacing Sylvester type operators $\nabla_{A,B}$ by Stein type operators $\Delta_{A,B}$. First, the formula for the recovery of a matrix W from its image $\Delta_{A,B}(W)$ changes, and all the algorithms change respectively. Second, minor changes appear in the computation of the $\Delta_{A,B}$-generators of the matrices $X_{i+1} = 2Y_i - Y_i M Y_i$, because of the changes of the expressions for the matrix products and inverses. Let us specify these changes.

Assume that the matrices M and A are nonsingular and write $\Delta_{A,B}(M) = GH^T$. Then we have the following expression for the inverses:

$$\Delta_{B,A}(M^{-1}) = M^{-1} - BM^{-1}A = M^{-1}A^{-1}\Delta_{A,B}(M)M^{-1}A = G_-H_-^T$$

where $G_- = M^{-1}A^{-1}G$ and $H_-^T = H^TM^{-1}A$. Similarily, if M and B are nonsingular, we have

$$\Delta_{B,A}(M^{-1}) = M^{-1} - BM^{-1}A = TM^{-1}\Delta_{A,B}(M)T^{-1}M^{-1} = G_+H_+^T$$

where $G_+ = BM^{-1}G$ and $H_+^T = H^TB^{-1}M^{-1}$. In both cases, the length of the $\Delta_{A,B}$-generator G, H for M equals the length of the respective $\Delta_{B,A}$-generator for M^{-1}.

Likewise, for the product YMY we deduce the following expression without nonsingularity assumptions:

$$YMY - BYMYA = (Y - BYA)MY + BYAM(Y - BYA) - BY(M - AMB)YA.$$

This expression furnishes us with $\Delta_{B,A}$-generators (of Stein type) of length at most 3α for Y_iMY_i and, consequently, for $X_{i+1} = 2Y_i - Y_iMY_i$, provided that M and Y_i are given with their $\Delta_{A,B}$- and $\Delta_{B,A}$-generators of length α, respectively.

The resulting changes of our algorithms will be further specified in the next two subsections. For some more elaborate techniques that enable extension of our algorithms to some operators $\Delta_{A,B}$ where both matrices A and B are singular.

9.1 Specific changes for Subalgorithm 4.1

We change the requirements to the output of Subalgorithm 4.1 and its computation as follows:

New Output: A $\Delta_{B,A}$-generator of a length at most α for a matrix Y_i satisfying the bound

$$||Y_i - M^{-1}||_2 \leq (1 + (1 + ||A||_2||B||_2)\nu^-)e_{2,i} \tag{9.1}$$

for $e_{2,i}$ of (4.1) and ν^- of Definition 2.3. The latter change is motivated by the following argument extending the proof of Theorem 5.1

$$\begin{aligned}
& \|X_i - BX_iA - M^{-1} + BM^{-1}A\|_2 \\
= \; & \|(X_i - M^{-1}) - B(X_i - M^{-1})A\|_2 \\
\leq \; & \|(X_i - M^{-1})\|_2 + \|B\|_2\|(X_i - M^{-1})\|_2\|A\|_2 \\
\leq \; & (1 + \|A\|_2\|B\|_2)\|(X_i - M^{-1})\|_2 \\
\leq \; & (1 + \|A\|_2\|B\|_2)e_i.
\end{aligned}$$

The mild assumption (5.3) for $i = 0$, which ensures rapid convergence of Algorithm 3.1 with the rate of $1+\theta$, turns into the following one in the Stein type case:

$$(1 + (1 + \|A\|_2\|B\|_2)\nu^-)\kappa(M)p_0^{1-\theta} \leq 1. \tag{9.2}$$

for $\nu^- = \nu_{2,\alpha}(\Delta_{A,B})$ of Definition 2.3.

9.2 Specific changes for Subalgorithm 6.1

We change the requirement regarding the output of Subalgorithm 6.1 and its computations as follows:

New Output: A positive integer α, two operator matrices A and B, A being nonsingular, a $\Delta_{A,B}$-generator of length at most α for a nonsingular matrix M. A $\Delta_{B,A}$-generator of a length at most α for a matrix Y_i satisfying the bound

$$\hat{e}_{i+1} = \|Y_{i+1} - M^{-1}\| \leq C_i e_i \tag{9.3}$$

for e_i of (4.2), $hate_i$ of (4.3), and

$$C_i = \nu^{-}\|GH^T\| \cdot \|A\| \cdot \|A^{-1}\|(e_i + 2\|M^{-1}\|). \tag{9.4}$$

New Computations: Compute and output $U_{i+1} = -X_{i+1}A^{-1}G, W_{i+1}^T = H^T X_{i+1} A$. The latter changes are motivated by the following argument extending Proposition 6.1: the expression for $E_i = U_i W_i^T - UW^T = X_i A^{-1} GH^T X_i A - M^{-1}A^{-1}GH^T M^{-1}A$ changes as follows:

$$E_i = (X_i - M^{-1})A^{-1}GH^T(X_i - M^{-1})A - UH^T(X_i - M^{-1})A$$
$$+ (X_i - M^{-1})A^{-1}GW_i^T,$$
$$\|E_i\| \leq \|GH^T\| \cdot \|A\| \cdot \|A^{-1}\|e_i(e_i + 2\|M^{-1}\|)$$
$$= \|GH^T\|\chi(A)e_i(e_i + 2\|M^{-1}\|).$$

Corollary 7.2, which ensures the rapid convergence of Algorithm 3.1 combined with Subalgorithm 6.1, turns into the following result in the Stein type case:

Corollary 9.1. *Let* $r_0 = \|I - MY_0\| \leq 1, e_0^* \sqrt{\|M\|} \leq 1, \theta \leq 1$ *for* e_0^* *of (7.2) and let* $(e_0^*)^{1-\theta}\|M\|^{1-\theta/2}C \leq 1$. *Then* $e_{i+1} \leq ((e_0^*)^2\|M\|)^{(1+\theta)^i}$ *and* $\hat{e}_{i+1} \leq$

Ce_{i+1} for $i = 0, 1, 2, \ldots$. Here C is a new constant derived from the constant C_i of equation (9.4) by assuming that $C = 3\nu^- \|GH^T\| \|\chi(A)\| \|Y_0\|/(1-r_0) \geq C_i$ for C_i of equation (9.4) and provided that $e_i \leq \|M^{-1}\|$ for all i and that $r_0 \leq 1$.

The above changes in Corollary 7.2 imply the respective changes in Corollary 7.3.

10 Preliminaries for Estimating the Norm of the Inverse Operator

To complete our analysis presented in the previous sections, we must estimate the norm ν^- of the inverses of the operators $\nabla_{A,B}$ or $\Delta_{A,B}$ associated with each of our two variants of Newton-Structured Matrix Iteration (cf. Definition 2.3).

In section 11 we will estimate ν^- for the four major classes of structured matrices, Toeplitz-like, Hankel-like, Vandermonde-like and Cauchy-like matrices (depending on the choice of the basic bilinear or trilinear representation of such matrices). Technically, we will follow the line of the Appendix of [P92], in particular we will rely on the SVD-representation of the image GH^T of the operator applied to a matrix M.

In section 12, we will present two distinct types of techniques for the estimation of ν^- not involving the SVD-representation of the image GH^T of the operators applied to a matrix M. The first technique, will exploit some specific properties of the displacement operators and will enable us to estimate the norm ν^- for the operators associated with Toeplitz-like, Hankel-like, Vandermonde-like, and Chebyshev-Vandermonde-like matrices. We will present these techniques in the next two subsections.

In section 13, we will present an alternative technique, which can be used for more general classes of operators. We will specify it in a unified way for the operators associated with the important classes of the Toeplitz-plus-Hankel-like, and Cauchy-like matrices.

In the present section, we will state some definitions and simple auxiliary results.

Definition 10.1. *For a positive integer k, we will say that a matrix S is* nilpotent of order k *if $S^k = 0$ and*

f-idempotent of order k *if $S^k = aI$.*

Example 10.1. $Z^n = 0, Z_f^n = fI$.

Theorem 10.1. *For an operator $\nabla_{A,B}(M) = AM - MB$ of Sylvester type, if A is a nonsingular matrix, then $||A||\nu^-(\nabla_{A,B}) \leq \nu^-(\Delta_{A^{-1},B})$, and if B is a nonsingular matrix, then $||B||\nu^-(\nabla_{A,B}) \leq \nu^-(\Delta_{A,B^{-1}})$.*

Definition 10.2. *Write* $S = [\sqrt{\frac{2}{n+1}}\sin(\frac{kj\pi}{n+1})]$, *S being the (normalized) matrix of the Discrete Sine Transform-I.*

Definition 10.3. *Write* $Q = [\sqrt{\frac{2}{n}} q_j \cos(\frac{(2k-1)(j-1)\pi}{2n})]$, $q_1 = \frac{1}{\sqrt{2}}, q_2 = \ldots = q_n = 1)$, *Q is the (normalized) matrix of the Discrete Cosine Transform-II,* $Q^T Q = I$. $D_Q = 2diag\left(1, \cos(\frac{\pi}{n}), \ldots, 2\cos(\frac{(n-1)\pi}{n})\right)$.

Definition 10.4. *Let the following structured matrix be defined by integers* γ *and* δ

$$Y_{\gamma\delta} = \begin{bmatrix} \gamma & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & \delta \end{bmatrix},$$

The next basic results are simple and well known (see, eg, [P92], [KO96]).

Fact 10.1. *We have* $||Circ(\mathbf{v})||_1 = ||\mathbf{v}||_1, ||L(\mathbf{v})||_l = ||\mathbf{v}||_1$, *for any vector* $\mathbf{v}, l = 1, \infty$; *furthermore,* $||D(\mathbf{v})|| \leq ||\mathbf{v}||$.

Fact 10.2. *For an orthogonal L-generator* (G, H) *of a matrix (cf. (1.3)-(1.9)), we have* $||\mathbf{g}_i||_2 = ||\mathbf{h}_i||_2 = \sigma_i(GH^T), i = 1, \ldots, \alpha, ||GH^T||_2 = \sigma_1^2(GH^T)$.

Fact 10.3. $||S||_1 = \max_k \sum_{j=1}^n \sqrt{\frac{2}{n+1}}|\sin(\frac{kj\pi}{n+1})| \leq \sum_{j=1}^n \sqrt{\frac{2}{n+1}} \leq \sqrt{2n}$.

Fact 10.4. $S^2 = I, Z + Z^T = SD_SS$, *and therefore,* $S(Z + Z^T)S = D_S$, *where* $D_S = diag\left(2\cos(\frac{\pi}{n+1}), \ldots, 2\cos(\frac{n\pi}{n+1})\right)$.

Fact 10.5. *The matrix* Y_{11} *of Definition 10.4 can be diagonalized by the Discrete Cosine Transform matrix Q, that is,* $Y_{11} = QD_QQ^T$.

Fact 10.6. $||Q||_1 = \max_k \sum_{j=1}^{n} \sqrt{\frac{2}{n}} |q_j \cos(\frac{(2k-1)(j-1)\pi}{2n})| \leq n\sqrt{\frac{2}{n}} \leq \sqrt{2n}$, $||D_Q||_1 = 2$.

11 Estimating the Norm Based on the Truncation of the SVD of the Image Matrix $L(M)$

In this section, we will estimate the scalar factor ν^- for the operators associated with Toeplitz-like, Hankel-like, Cauchy-like, and Vandermonde-like matrices.

Theorem 11.1. *Let* $\mathbf{s}$ *and* $\mathbf{t}$ *be a pair of vectors of dimension n filled with $2n$ distinct coordinates, s_i and t_j, none of the t_j being zero. Let $\nabla = \nabla_{A,B}$ and $\Delta = \Delta_{A,B}$ be nonsingular operators of (1.3), (1.4). Then we have the following bounds on the l-norm $\nu_{\alpha,l}, l = 1, 2, \infty, 1 \leq \alpha \leq n$, of the inverse operators ∇^{-1} and Δ^{-1} over the $n \times n$ complex matrices:*

$$\nu_{l,\alpha}(\Delta^{-1}_{Z,Z^T}) = \nu_{l,\alpha}(\Delta^{-1}_{Z^T,Z^T}) \leq \alpha n^{1.5}, \tag{11.1}$$

$$\nu_{l,\alpha}(\nabla^{-1}_{D^{-1}(\mathbf{t}),Z}) \leq \alpha\sqrt{n}||D(\mathbf{t})V^T(\mathbf{t})||_l, \tag{11.2}$$

$$\nu_{l,\alpha}(\nabla^{-1}_{D(\mathbf{t}),Z_1}) \leq \alpha\sqrt{n}||D^{-1}(\mathbf{e} - \mathbf{t}^n)V(\mathbf{t})||_l, \tag{11.3}$$

$$\nu_{l,\alpha}(\nabla^{-1}_{D(\mathbf{t}),D(\mathbf{s})}) \leq \alpha\sqrt{n}||D(\mathbf{s})C(\mathbf{s},\mathbf{t})||_l, \tag{11.4}$$

for $l = 1, 2, \infty$. For $l = 2$ over the real matrices all these upper bounds are *decreased by a factor $\sqrt{n}$.*

Proof. The bounds of Theorem 11.1 are obtained based on the bilinear or trilinear representations for each matrix M of Δ-rank (respectively, ∇-rank) at most α such that $\Delta(M) = GH^T$ (respectively, $\nabla(M) = GH^T$) for the matrices G and H of (1.5). In particular, we deduce each of bounds (11.1)-(11.4) based on the equations of Theorem 1.1 for the operators Δ or ∇.

We first deduce from (1.6) that $||M|| \leq \sum_{i=1}^{\alpha} ||L(\mathbf{g}_i)L^T(\mathbf{h}_i)||$. By applying Facts 10.1, 10.2 and Theorem 2.1, we obtain that $||L(\mathbf{g}_i)||_1 = ||\mathbf{g}_i||_1 \leq \sigma_i\sqrt{n}$, $||L^T(\mathbf{h}_i)||_1 = ||L(\mathbf{h}_i)||_\infty \leq ||\mathbf{h}_i||_1 \leq \sigma_i\sqrt{n}$, $||L(\mathbf{g}_i)L^T(\mathbf{h}_i)||_l \leq \sigma_i^2 n$ for $l = 1, \infty$ and for all i. Therefore, $||M||_l \leq n\sum_{i=1}^{\alpha}\sigma_i^2 \leq n\alpha\sigma_1^2 = n\alpha||GH^T||_2$ for $l = 1$ and $l = \infty$.

Using Theorem 2.1, we reconcile the l-norm and the 2-norm on both sides of the latter inequality and arrive at the bound (11.1) for the complex matrices $M \in \mathbb{C}^{n\times n}$ and the operator $\Delta_{\{Z,Z^T\}}$. The equation of (11.1) follows by comparison of (1.6) and (1.7). For real $M \in \mathbb{R}^{n\times n}$, we combine the latter bounds on $||M||_l$ for $l = 1, \infty$ with the bound $||M||_2^2 \leq ||M||_1||M||_\infty$ of Theorem 2.1 and obtain an improvement of the bound of (12.1) by a factor $\sqrt{n}$. All other estimates of Theorem 11.1, that is, (11.2)-(11.4), are derived similarly, based on the representations (1.8)-(1.10), respectively. (We leave the details to the reader.) □

Remark 11.1. *The operators Δ and ∇ are associated with Toeplitz-like and Hankel-like matrices (for Δ of (11.1)), Vandermonde-like matrices (for ∇ of (11.2) and (11.3)), and Cauchy-like matrices (for ∇ of (11.4)). We will omit the straightforward extensions of Theorem 1.1 to various other operators.*

12 Estimating the Norm ν^- Based on the Decomposition of a Matrix M via the Powers of Operator Matrices and the Image $L(M)$ and on the f-Idempotent Properties of Displacement Operators

12.1 Estimating the operator norm ν^- in the case of nilpotent operator matrices A and/or B

In this section, we estimate the scalar factor ν^- for the operators associated with Toeplitz-like, Hankel-like, Vandermonde-like, and Chebyshev-Vandermonde-like matrices. We explicitly estimate ν^- for the Stein type operators L, but we may extend the estimate immediately to the case of the operators (1.1) of Sylvester type provided that one of the operator matrices

A and B is nonsingular (see section 11).

In particular the matrix equation $\nabla_{A,B}(M) = A\Delta_{A^{-1},B}(M)$ implies that $\nu^-_{1,\alpha}(\Delta_{A^{-1},B}) \geq \|A\|_1\nu^-_{1,\alpha}(\nabla_{A,B})$ and similarly $\nabla_{A,B}(M)A = \Delta_{A,B^{-1}}(M)B$ implies that $\nu^-_{1,\alpha}(\Delta_{A,B^{-1}}) \geq \|B\|_1\nu^-_{1,\alpha}(\nabla_{A,B})$.

Theorem 12.1. *[GO92], [W93]. Let $\Delta = \Delta_{A,B}$ be a regular operator of (1.4) of Stein type. Let A and/or B be any nilpotent matrix. Then we have*

$$M = \Delta(M) + A\Delta(M)B + A^2\Delta(M)B^2 + \ldots + A^{k-1}\Delta(M)B^{k-1}, \tag{12.1}$$

and consequently,

$$\nu^-_{1,\alpha} \leq 1 + \|A\|_1\|B\|_1 + \|A^2\|_1\|B^2\|_1 + \ldots + \|A^{k-1}\|_1\|B^{k-1}\|_1. \tag{12.2}$$

Proof. Recursively pre-multiply the equation

$$\Delta_{A,B}(M) = M - AMB \tag{12.3}$$

of (12.3) by A and post-multiply it by B to obtain

$$A^i\Delta_{A,B}(M)B^i = A^iMB^i - A^{i+1}MB^{i+1} \tag{12.4}$$

for all i. Sum over i from 0 to k to obtain that

$$\Delta(M) + A\Delta(M)B + A^2\Delta(M)B^2 + \ldots + A^{k-1}\Delta(M)B^{k-1} = M - A^kMB^k. \tag{12.5}$$

Substitute $A^k = 0$ or $B^k = 0$ and obtain equation (12.1) and the bound of (12.2). $\square$

Next, we will specialize Theorem 12.1 to some specific classes of structured matrices.

Theorem 12.2. *Let* $\mathbf{s}$ *and* $\mathbf{t}$ *be a pair of vectors of dimension* n *filled with* $2n$ *distinct coordinates* s_i *and* t_j, *none of the* t_j *being zero. Let* $\Delta = \Delta_{A,B}$ *be a regular operator of(1.2). Let us write* $t_- = \min_j |t_j|, t_+ = \max_j |t_j|, Z^* = 2\sum_{i=1}^{n/2}(-1)^{i-1}Z^{2i-1}$. *Then we have the following bounds on the 1-norm* $\nu_{\alpha,1}, 1 \le \alpha \le n$.

$$\nu_{1,\alpha}(\Delta^{-1}_{Z,Z^T}) = \nu_{1,\alpha}(\Delta^{-1}_{Z^T,Z^T}) \le n, \tag{12.6}$$

$$\nu_{l,\alpha}(\Delta^{-1}_{D^{-1}(\mathbf{t}),Z_1}) \le \begin{cases} \frac{1-(\frac{1}{t_-})^n}{1-\frac{1}{t_-}} & \textit{if } \frac{1}{t_-} \ne 1 \\ n & \textit{if } \frac{1}{t_-} = 1 \end{cases} \tag{12.7}$$

$$\nu_{1,\alpha}(\Delta^{-1}_{D(\mathbf{t}),Z}) \le \begin{cases} \frac{1-t_+^n}{1-t_+} & \textit{if } t_+ \ne 1 \\ n & \textit{if } t_+ = 1 \end{cases} \tag{12.8}$$

$$\nu_{1,\alpha}(\Delta^{-1}_{D(\mathbf{t}),Z^*}) \le \begin{cases} 1 + (\frac{n}{t_+})(\frac{1-(\frac{2}{t_+})^{n-1}}{1-(\frac{2}{t_+})}) & \textit{if } t_+ \ne 2 \\ 1 + \frac{n(n-1)}{2} & \textit{if } t_+ = 2 \end{cases} \tag{12.9}$$

Proof. The bounds of Theorem 12.2 are obtained based on the bound (12.2) applied for $k = n$ and the operators Δ of (12.6)-(12.9). (12.6) is immediate because $\|Z_1\|_1 = \|Z_1^T\|_1 = \ldots = \|Z_1^{n-1}\|_1 = \|(Z_1^T)^{n-1}\|_1 = 1$. (12.7) immediately follows from (12.2) because $\|D^{-1}(\mathbf{t})\|_1 = \frac{1}{t_-}, \|Z_1\|_1 = 1$, and therefore, we have $\nu^-_{1,\alpha} \le 1 + \frac{1}{t_-} + \ldots + \frac{1}{t_-^{n-1}}$. The proof of (12.8) is similar to the proof of (12.7). Finally, let us prove (12.9). Recall that $Z^{\bullet} = 2\sum_{i=1}^{n/2}(-1)^{i-1}Z^{2i-1}$. Consequently, $\|Z^{\bullet}\|_1 = n$,

$$\|(Z^{\bullet})^{n-1}\|_1 = 2^{n-1}\|(\sum_{i=1}^{n/2}(-1)^{i-1}Z^{2i-1})^{n-1}\|_1 \leq 2^{n-1}\frac{n}{2} = 2^{n-2}n,$$

$$\nu_{1,\alpha}^{-} \leq 1 + \frac{1}{t_+}n + \ldots + \frac{1}{t_+^{n-1}}2^{n-2}n,$$

$$\nu_{1,\alpha}^{-} \leq 1 + \frac{n}{t_+}\left(1 + (\frac{n}{t_+})^2 + \ldots + \frac{1}{t_+^{n-2}}2^{n-2}n\right).$$

□

Remark 12.1. *The operators Δ are associated with Toeplitz-like and Hankel-like matrices for Δ of (12.6), Vandermonde-like matrices for Δ of (12.7 and (12.8)), and Chebyshev-Vandermonde-like matrices for Δ of (12.9).*

12.2 Estimating the operator norms ν^- in the case of f-idempotent operator matrices A and/or B

In this section, we extend our first technique based on the extension of Theorem 12.1 (This will enable us to extend our estimates for the norm ν^- to cover the Toeplitz-like, Hankel-like, Vandermonde-like and Cauchy-like matrices.)

Theorem 12.3. *Let $\Delta = \Delta_{A,B}$ be a regular operator of (1.2), where $S^k = \alpha I$, for some positive integer k. Then we have*

$$M = (\Delta(M) + A\Delta(M)B + \ldots + A^{k-1}\Delta(M)B^{k-1})(I - \alpha B^k)^{-1}, \quad (12.10)$$

and consequently,

$$\nu_{1,\alpha}^{-} \leq (1 + \|A\|\ \|B\| + \ldots + \|A^{k-1}\|\ \|B^{k-1}\|)\|(I - \alpha B^{k})^{-1}\|. \tag{12.11}$$

Likewise, if $M = (I-\alpha S^k)$, $\nu_{1,\alpha}^{-} \leq \nu_{1,\alpha}^{-} \leq (1+\|A\|\ \|B\|+\ldots+\|A^{k-1}\|\ \|B^{k-1}\|)\|(I-\alpha A)^{-1}\|$.

Proof. Similar to the proof of Theorem 12.1. □

Theorem 12.4. *Let* **s** *and* **t** *be a pair of vectors of dimension* n *filled with* $2n$ *distinct coordinates* s_i *and* t_j, *none of the* t_j *being zero. Let* $\Delta = \Delta_{A,B}$ *be a regular operator of (1.4). Let the matrix* T *represent either* $D^{-1}(\mathbf{t})$ *or* $D(\mathbf{t})$. *Let us write (cf. Theorem 12.2)* $t_- = \min_j |t_j|, t_+ = \max_j |t_j|, (1+t^n)_- = \min_j |1+t_j^n|, (1+t^{-n})_- = \min_j |1+\frac{1}{t_j^n}|, Z_{-1}^{*} = 2\sum_{i=1}^{n/2}(-1)^{i-1}Z_{-1}^{2i-1}$. *(Note that* $\frac{1}{(1+t^n)_-} = \|(I+T^n)^{-1}\|_1$ *if* $B = D(\mathbf{t})$.*) Then we have* $\frac{1}{(1+t^{-n})_-} = \|(I+T^n)^{-1}\|_1$ *if* $B = D^{-1}(\mathbf{t})$,

$$\nu_{1,\alpha}^{-}(\Delta_{Z_{-1},Z_1^T}^{-1}) = \nu_{1,\alpha}^{-}(\Delta_{Z^T-1,Z_1^T}^{-1}) \leq \frac{n}{2}, \tag{12.12}$$

$$\nu_{l,\alpha}^{-}(\Delta_{D^{-1}(\mathbf{t}),Z_{-1}}^{-1}) \leq \begin{cases} \frac{1}{(1+t^{-n})_-} \cdot \frac{1-(\frac{1}{t_-})^n}{1-\frac{1}{t_-}} & \text{if } \frac{1}{t_-} \neq 1 \\ \frac{1}{(1+t^{-n})_-} \cdot n & \text{if } \frac{1}{t_-} = 1 \end{cases} \tag{12.13}$$

$$\nu_{1,\alpha}^{-}(\Delta_{D(\mathbf{t}),Z_{-1}}^{-1}) \leq \begin{cases} \frac{1}{(1+t^{n})_-} \cdot \frac{1-t_+^n}{1-t_+} & \text{if } t_+ \neq 1 \\ \frac{1}{(1+t^{n})_-} \cdot n & \text{if } t_+ = 1 \end{cases} \tag{12.14}$$

$$\nu_{1,\alpha}^{-}(\Delta_{D(\mathbf{t}),Z_{-1}^{*}}^{-1}) \leq \begin{cases} \frac{1}{(1+t^{n})_-}\left(1 + \frac{n}{t_+} \cdot \frac{1-(\frac{2}{t_+})^{n-1}}{1-(\frac{2}{t_+})}\right) & \text{if } t_+ \neq 2 \\ \frac{1}{(1+t^{n})_-}\left(1 + \frac{n(n-1)}{2}\right) & \text{if } t_+ = 2 \end{cases} \tag{12.15}$$

Proof. Similar to the proof of Theorem 12.2. □

Remark 12.2. *The operator Δ is associated with Toeplitz-like and Hankel-like matrices (for Δ of (12.12)), Vandermonde-like matrices (for Δ of (12.13) and (12.14)), and Chebyshev-Vandermonde-like matrices (for Δ of (12.15)).*

13 Estimating the Operator Norms ν^- Based on the Eigendecomposition of an Operator Matrix

In this section we will generalize our first technique and present applications to Cauchy-like and Toeplitz+Hankel-like matrices.

Theorem 13.1. *let $\Delta = \Delta_{A,B}$ be a regular operator of (1.4) with $n \times n$ operator matrices A and B. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the matrix S. Let us write $A_{\lambda_i} = A - \lambda_i I, B_{\lambda_i} = I - \lambda_i B$. Assume that $\frac{1}{\lambda_i}$ is not an eigenvalue of the matrix B. Then we have*

$$\begin{aligned} M = \Delta(M)B_{\lambda_1}^{-1} + A_{\lambda_1}\Delta(M)BB_{\lambda_1}^{-1}B_{\lambda_2}^{-1} + \ldots \\ + A_{\lambda_1}\cdots A_{\lambda_{n-1}}\Delta(M)BB_{\lambda_n}^{-1}\cdots BB_{\lambda_1}^{-1} \end{aligned} \tag{13.1}$$

and consequently,

$$\begin{aligned} \nu_{1,\alpha} \le \|B_{\lambda_1}^{-1}\|_1 + \|A_{\lambda_1}\|_1\|B\|_1\|B_{\lambda_1}^{-1}\|_1\|B_{\lambda_2}^{-1}\|_1 + \ldots \\ + \|A_{\lambda_1}\|_1\cdots\|A_{\lambda_{n-1}}\|_1\|B^{n-1}\|_1\|B_{\lambda_1}^{-1}\|_1\cdots\|B_{\lambda_n}^{-1}\|_1 \end{aligned} \tag{13.2}$$

Proof. Let λ be any eigenvalue of the matrix A. We have

$$\begin{aligned}\Delta(M) &= M - \lambda MB + \lambda MB - AMB \\ &= M(I - \lambda B) - (A - \lambda I)MB \\ &= MB_\lambda - A_\lambda MB,\end{aligned}$$

and obtain

$$\Delta(M)B_\lambda^{-1} = M - A_\lambda MBB_\lambda^{-1}. \tag{13.3}$$

In particular, substitute $\lambda = \lambda_1$ into (13.3) and obtain

$$\Delta(M)B_{\lambda_1}^{-1} = M - A_{\lambda_1} MBB_{\lambda_1}^{-1}. \tag{13.4}$$

Then substitute $\lambda = \lambda_2$ and obtain

$$\Delta(M)B_{\lambda_2}^{-1} = M - A_{\lambda_2} MBB_{\lambda_2}^{-1}. \tag{13.5}$$

Pre-multiply (13.5) by A_{λ_1}, post-multiply by $BB_{\lambda_1}^{-1}$ and obtain

$$A_{\lambda_1}\Delta(M)B_{\lambda_2}^{-1}BB_{\lambda_1}^{-1} = A_{\lambda_1}MBB_{\lambda_1}^{-1} - A_{\lambda_1}A_{\lambda_2}MBB_{\lambda_2}^{-1}BB_{\lambda_1}^{-1}. \tag{13.6}$$

Add (13.4) to (13.6) and obtain

$$\Delta(M)B_{\lambda_1}^{-1} + A_{\lambda_1}\Delta(M)B_{\lambda_2}^{-1}BB_{\lambda_1}^{-1} = M - A_{\lambda_1}A_{\lambda_2}MBB_{\lambda_2}^{-1}BB_{\lambda_1}^{-1}. \tag{13.7}$$

Substitute $\lambda = \lambda_3$ into equation (13.3). Multiply the resulting equation by the first term of (13.7)on the left, and by the second term of (13.7) on the right, then add (13.7) to the resulting equation. Repeat this process recursively and in a total of n steps obtain the following equation

$$\Delta(M)B_{\lambda_1}^{-1} + A_{\lambda_1}\Delta(M)B_{\lambda_2}^{-1}BB_{\lambda_1}^{-1} + \ldots + A_{\lambda_1}\cdots A_{\lambda_{n-1}}\Delta(M)BB_{\lambda_n}^{-1}\cdots BB_{\lambda_1}^{-1}$$
$$= M - A_{\lambda_1}A_{\lambda_2}\cdots A_{\lambda_n}MBB_{\lambda_n}^{-1}\cdots BB_{\lambda_1}^{-1}. \tag{13.8}$$

This implies (13.1) since $A_{\lambda_1}\ldots A_{\lambda_n} = 0$. □

13.1 Estimating the operator norms ν^- for Cauchy-like and Toeplitz+Hankel-like matrices using the eigendecomposition technique

To apply Theorem 13.1 to the case of Toeplitz+Hankel-like matrices we need some definitions and auxiliary results. Recall the expression for S and Q in Definitions 10.2 and 10.3, respectively.

$S = \left[\sqrt{\frac{2}{n+1}}\sin(\frac{kj\pi}{n+1})\right]$, $Q = \left[\sqrt{\frac{2}{n}}q_j\cos(\frac{(2k-1)(j-1)\pi}{2n})\right]$,

$q_1 = \frac{1}{\sqrt{2}}, q_2 = \ldots = q_n = 1$. By using Facts 10.5 and 10.6, we deduce $\|Y_{11} - \lambda_i I\|_1 = \|QD_QQ^T - \lambda_i I\|_1 \leq \|Q\|_1\|D_Q\|_1\|Q^T\|_1 + |\lambda_i| \leq 2n + |\lambda_i|$. Now, let $\alpha_i = 2n + |\lambda_i|$ and let $\alpha = \max_i \alpha_i$. By the definition, $((Z+Z^T)^{-1})_{\lambda_i} = I - \lambda_i(Z+Z^T)^{-1}$. Recall that $S^2 = I$ and obtain that $((Z+Z^T)^{-1})_{\lambda_i} =$

$S(I-\lambda_i S(Z+Z^T)^{-1}S)S = S(I-\lambda_i D_S^{-1})S$. Therefore, $\|((Z+Z^T)^{-1})_{\lambda_i}^{-1}\|_1 = \|F(I-\lambda_i D_S^{-1})^{-1}S\|_1 \le 2n/\psi_i$ where $\psi_i = \min_j |1-2\lambda_i/\cos(\frac{j\pi}{n+1})|$. Let us write $\psi = \min_i \psi_i$ and $\alpha = \max_i \alpha_i$.

Now, we are ready to state our next theorem.

Theorem 13.2. *Let* $\mathbf{s}$ *and* $\mathbf{t}$ *be a pair of vectors of dimension* n *filled with* $2n$ *distinct coordinates* s_i *and* t_j, *none of the* t_j *being zero. Let* $\Delta = \Delta_{A,B}$ *be a regular operator of (1.2). Let us write* $t_- = \min_j |t_j|, s_+ = \max_j |s_j|, \phi_i = \min_j |1 - s_i t_j|, \phi = \min_i \phi_i$. *Then we have*

$$\nu_{1,\alpha}(\Delta^{-1}_{D(\mathbf{s}),D^{-1}(\mathbf{t})}) \le \begin{cases} \frac{1}{\phi} \cdot \frac{1-(\frac{2s_+t_-}{\phi})^n}{1-\frac{2s_+t_-}{\phi}} & \text{if } \frac{2s_+t_-}{\phi} \ne 1 \\ \frac{n}{\phi} & \text{if } \frac{2s_+t_-}{\phi} = 1 \end{cases} \tag{13.9}$$

$$\nu_{1,\alpha}(\Delta^{-1}_{(Z+Z^T)^{-1},Y_{11}}) \le \begin{cases} \frac{2n}{\psi} \cdot \frac{1-(\frac{2\alpha 2n}{\psi})^n}{1-\frac{2\alpha 2n}{\psi}} & \text{if } \frac{2\alpha 2n}{\psi} \ne 1 \\ \frac{n}{2\alpha} & \text{if } \frac{2\alpha 2n}{\psi} = 1 \end{cases} \tag{13.10}$$

Proof. Recall that $s_+ = \max_j |s_j|, A_{\lambda_i} = A - \lambda_i I$ and deduce that $\|A_{\lambda_i}\|_1 = \|A - \lambda_i I\|_1 \le \|A\|_1 + |\lambda_i|$. We have $A = D(\mathbf{s})$ in (13.9). Therefore, $\lambda_i = s_i, |\lambda_i| \le s_+$, and $\|A_{\lambda_i}\|_1 \le 2s_+$ for all i. Similarily, for $B = D^{-1}(\mathbf{t})$ of (13.9), we obtain $\|B_{\lambda_l}\|_1 \le t_-/\phi_i$ Substitute both norm bounds into (13.2) and obtain for $\nu_{1,\alpha}$ of (13.9) that

$$\nu_{1,\alpha} \le \frac{1}{\phi_1} + 2s_+t_-\frac{1}{\phi_1\phi_2} + \ldots + (2s_+t_-)^{n-1}\frac{1}{\phi_1\cdots\phi_n}.$$

Since we have $\phi = \min_i \phi_i$, it follows that

$$\nu_{1,\alpha} \le \frac{1}{\phi} + 2s_+t_-\frac{1}{\phi^2} + \ldots + (2s_+t_-)^{n-1}\frac{1}{\phi^n}$$

and we obtain (13.9). To derive (13.10), first observe that $\|(Z+Z^T)^n\|_1 \leq 2^n$. By (13.2) we have

$$\nu_{1,\alpha} \leq \frac{2n}{\psi_1} + 2\alpha_1 \frac{(2n)^2}{\psi_1 \psi_2} + \ldots + (2^{n-1})\alpha_1 \cdots \alpha_{n-1} \frac{(2n)^n}{\psi_1 \cdots \psi_n},$$

since we have $\psi = \min \psi_i$ and $\alpha = \max \alpha_i$. Therefore,

$$\nu_{1,\alpha} \leq \frac{2n}{\psi} + 2\alpha \frac{(2n)^2}{\psi^2} + \ldots + (2^{n-1})\alpha^{n-1} \frac{(2n)^n}{\psi^n}$$

$$\nu_{1,\alpha} \leq \frac{2n}{\psi}(1 + 2\alpha \frac{2n}{\psi} + \ldots + (2^{n-1})\alpha^{n-1} \frac{(2n)^{n-1}}{\psi^{n-1}}),$$

and we arrive at (13.10). □

Remark 13.1. *The operator Δ is associated with Cauchy-like matrices (for Δ of (13.9)) and Toeplitz+Hankel-like matrices (for Δ of (13.10)) (cf. [GO94], [GKO95]).*

References

[BA80] R. R. Bitmead, B. D. O. Anderson, Asymptotically Fast Solution of Toeplitz and Related Systems of Linear Equations, *Linear Algebra Appl.*, **34**, 103-116, 1980.

[BM,a] D. A. Bini, B. Meini, Approximate Displacement Rank and Applications, preprint.

[BP93] D. Bini, V. Y. Pan, Improved Parallel Computations with Toeplitz-like and Hankel-like Matrices, *Linear Algebra Appl.*, **188/189**, 3-29, 1993.

[BP94] D. Bini, V. Y. Pan, *Polynomial and Matrix Computations, Vol.1: Fundamental Algorithms*, Birkhäuser, Boston, 1994.

[CKL-A87] J. Chun, T. Kailath. H. Lev-Ari, Fast Parallel Algorithm for QR-factorization of Structured Matrices, *SIAM J. Sci. Stat. Comput.*, **8**, **6**, 899-913, 1987.

[GKO95] I. Gohberg, T. Kailath, V. Olshevsky, Fast Gaussian Elimination with Partial Pivoting for Matrices with Displacement Structure, *Math. of Computation*, **64**, 1557-1576, 1995.

[GL96] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, Maryland, 1996 (third edition).

[GO92] I. Gohberg, V. Olshevsky, Circulants, Displacements and Decompositions of Matrices, *Integral Equations and Operator Theory*, **15, 5**, 730-743, 1992.

[GO94] I. Gohberg, V. Olshevsky, Complexity of Multiplication with Vectors for Structured Matrices, *Linear Algebra Appl.*, **202**, 163-192, 1994.

[GO96] I. Gohberg, V. Olshevsky, Fast Inversion of Vandermonde and Vandermonde-like Matrices, 205-221, *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath* (A. Paulraj, V. Roychowdhury, and C. Shaper eds.), Kluwer Academic Publisher, 1996.

[GR87] L. Greengard, V. Rokhlin, A Fast Algorithm for Particle Simulation, *J. Comput. Physics*, **73, 2**, 325–348, 1987.

[H95] G. Heinig, Inversion of Generalized Cauchy Matrices and the Other Classes of Structured Matrices, *Linear Algebra for Sig-*

nal Processing, IMA Volume in Math. and Its Applications, **69**, 95-114, Springer, 1995.

[H96] G. Heinig, Transformation Techniques for Toeplitlz and Related Matrices, *Calcolo*, **33**, 1996.

[IK66] E. ISAACSON AND H. B. KELLER, Analysis of Numerical Methods, Wiley, NY, 1966

[KKM79] T. Kailath, S. Y. Kung, M. Morf, Displacement Ranks of Matrices and Linear Equations, *J. Math. Anal. Appl.*, **68, 2**, 395-407, 1979.

[KO96] T. Kailath, V. Olshevsky, Displacement Structure Approach to Discrete Transform Based Preconditioners of G. Strang Type and of T. Chan Type, *Calcolo*, **33**, 191-208, 1996.

[KS95] T. Kailath, A. Sayed, Displacement Structure: Theory and Applications, *SIAM Review*, **37**, 297-386, 1995.

[KS99] T. Kailath, A. Sayed (editors), *Fast Reliable Algorithms for Matrices with Structure*, SIAM Publications, Philadelphia, 1999.

[M80] M. Morf, Doubling Algorithms for Toeplitz and Related Equations, *Proc. IEEE Internat. Conf. on ASSP*, 954-959,1980.

[OP98] V. Olshevsky, V. Y. Pan, A Unified Superfast Algorithm for Boundary Rational Tangential Interpolation Problem and for Inversion and Factorization of Dense Structured Matrices, *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, 192-201, IEEE Computer Society Press, 1998.

[OP99] V. Olshevsky, V. Y. Pan, Polynomial and Rational Interpolation and Multipoint Evaluation (with Structured Matrices), *Proc. 26th Intern. Colloquium on Automata, Languages and Programming (ICALP'99), Springer's LNCS*, **1644**, 585-594, Berlin, 1999.

[OS99] V. Olshevsky, M. A. Shokrollahi, A Displacement Approach to Efficient Decoding of Algebraic-Geometric Codes, *Proc. 31st Ann. Symp. on Theory of Computing*, 235-244, ACM Press, New York, 1999.

[P90] V. Y. Pan, On Computations with Dense Structured Matrices, *Math. of Computation*, **55 (191)**, 179-190. 1990.

[P92] V. Y. Pan, Parallel Solution of Toeplitz-like Linear Systems, *J. of Complexity*, **8**, 1-21, 1992.

[P93] V.Y. Pan, Decreasing the Displacement Rank of a Matrix, *SIAM J. Matrix Anal. Appl.*, **14, 1**, 118-121, 1993.

[P93a] V. Y. Pan, Concurrent Iterative Algorithm for Toepliz-like Linear Systems, *IEEE Trans. on Parallel and Distributed Systems*, **4, 5**, 592-600, 1993.

[P2000] V. Y. Pan, Nearly Optimal Computations with Structured Matrices, *Proc. 11th Ann. ACM-SIAM Symp. on Discrete Algorithms, (SODA'2000)*, ACM Press, New York, and SIAM Publications, Philadephia, 2000.

[PACLS98] V. Y. Pan, M. Abu Tabanjeh, Z. Q. Chen, E. Landowne, A. Sadikou, New Transformations of Cauchy Matrices and Trummer's Problem, *Computers & Math. (with Applics.)*, **35, 12**, 1-5, 1998.

[PBRZ99] V. Y. Pan, S. Branham, R. Rosholt, A. Zheng, Newton's Iteration for Structured Matrices and Linear Systems of Equations, *SIAM volume on Fast Reliable Algorithms for Matrices with Structure*, SIAM Publications, Philadelphia, 1999.

[PLST93] V.Y. Pan, E. Landowne, A. Sadikou, O. Tiga, A New Approach to Fast Polynomial Interpolation and Multipoint Evaluation, *Computers & Math. (with Application)*, **25**, **9**, 25-30, 1993.

[PR2000] V. Y. Pan, Y. Rami, "Newton's iteration for the Inversion of Structured Matrices" (with Y. Rami), invited to "Structured Matrices: Recent Developments in Theory and Computation:", edited by D.Bini, E. Tyrtyshnikov and P. Yalamov, Nova Science Publishers, USA.

[PRW2000] V. Y. Pan, Y. Rami, X. Wang, "Newton's iteration for the Inversion of Structured Matrices" (with Y. Rami and X. Wang)", Proc.14th Intern. Symposium on Math. Theory of Network and Systems(MTNS'2000), June 2000.

[PRWa] V. Y. Pan, Y. Rami, X. Wang, To appear

[PS91] V. Y. Pan, R. Schreiber, An Improved Newton Iteration for the Generalized Inverse of a Matrix, with Applications, *SIAMJ. on Scientific and Statistical Computing*, **12**, **5**, 1109-1131, 1991.

[PZHY97] V. Y. Pan, A. Zheng, X. Huang, Y. Yu, Fast Multipoint Polynomial Evaluation and Interpolation via Computation with Structured Matrices, *Annals of Numerical Math.*, **4,** 483-510, 1997.

[PZHD97] V. Y. Pan, A. Zheng, X. Huang, O. Dias, Newton's Iteration for Inversion of Cauchy-like and Other Structured Matrices, *J. of Complexity,* **13**, 108-124, 1997.

[W93] D. H. Wood, Product Rules for the Displacement of Nearly-Toeplitz Matrices, *Linear Algebra Appl.*, **188,189**, 641-663, 1993.