Using fast matrix multiplication to solve structured linear systems

Alin Bostan[1], Claude-Pierre Jeannerod[2] and Éric Schost[3]

[1] Algorithms Project, INRIA Rocquencourt
Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France
Alin.Bostan@inria.fr
[2] LIP - ENS Lyon and Arenaire Project, INRIA Rhônes-Alpes
46 allée d'Italie, 69364 Lyon Cedex 07, France
claude-pierre.jeannerod@ens-lyon.fr
[3] Computer Science Department, the University of Western Ontario
London, Ontario N6A 5B7, Canada
eschost@uwo.ca

Abstract. Structured linear algebra techniques enable one to deal at once with various types of matrices, with features such as Toeplitz-, Hankel-, Vandermonde- or Cauchy-likeness. Following Kailath, Kung and Morf (1979), the usual way of measuring to what extent a matrix possesses one such structure is through its displacement rank, that is, the rank of its image through a suitable displacement operator. Then, for the families of matrices given above, the results of Bitmead-Anderson, Morf, Kaltofen, Gohberg-Olshevsky, Pan (among others) provide algorithm of complexity $O(\alpha^2 N)$, up to logarithmic factors, where N is the matrix size and α its displacement rank. We show that for Toeplitz-like or Vandermonde-like matrices, this cost can be reduced to $O(\alpha^{\omega-1} N)$, where ω is an exponent for matrix multiplication. We present consequences for Hermite-Padé approximation and bivariate interpolation.

Keywords. Structured matrices, fast matrix multiplication, Toeplitz-like matrix, Vandermonde-like matrix, Hermite-Padé approximation, interpolation.

1 Introduction

Structured linear algebra techniques are a versatile set of tools; they enable one to deal at once with various types of matrices, with features such as Toeplitz-, Hankel-, Vandermonde- or Cauchy-likeness.

Following [1], the usual way of measuring to what extent a matrix M possesses one such structure is through its *displacement rank*, that is, the rank of its image through a suitable *displacement operator*. Let N be in $\mathbb{N}$ and let k be our base field. Given two matrices A and B in $k^{N\times N}$, the displacement operator $\Delta_{A,B} : k^{N\times N} \to k^{N\times N}$ is defined as $\Delta_{A,B}(M) = M - AMB$. The displacement rank of M is the rank of $\Delta_{A,B}(M)$; the matrices G, H are *generators* for M if

Dagstuhl Seminar Proceedings 06271
Challenges in Symbolic Computation Software
http://drops.dagstuhl.de/opus/volltexte/2006/778

$\Delta_{A,B} = GH^t$. Usual choices for A and B are the lower-shift matrix $Z \in k^{N\times N}$ given by

$$Z = \begin{bmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}$$

and the diagonal matrices $D(\mathbf{s})$, for some $\mathbf{s} \in k^N$. They enable us to consider the following operators, associated with the four types of structure seen above:

- the Toeplitz-like structure, with $A = Z$ and $B = Z^t$;
- the Hankel-like structure, with $A = Z$ and $B = Z$;
- the Vandermonde-like structure, with $A = D(\mathbf{s})$ and $B = Z^t$, for some $\mathbf{s}$ in k^N;
- the Cauchy-like structure, with $A = D(\mathbf{s})$ and $B = D(\mathbf{t})$, for some $\mathbf{s}$ and $\mathbf{t}$ in k^N.

Bitmead and Anderson [2] and Morf [3] gave fast algorithms for solving Toeplitz-like systems $Mx = b$, under strong non-degeneracy assumptions, obtaining algorithms of complexity $O\tilde{~}(N)$ when $\alpha = O(1)$. Here, the soft-Oh notation $O\tilde{~}(\,.\,)$ denotes the omission of polylogarithmic factors. These first algorithms were subsequently generalized in several directions. Kaltofen [4] showed how to lift the previous non-degeneracy assumptions, using randomization and an extension of Morf's and Bitmead and Anderson's inequalities on the displacement rank of submatrices. These techniques were also extended to the other structures presented above; see [5] for a detailed account. Overall, these algorithms featured a complexity of $O\tilde{~}(N)$ when $\alpha = O(1)$, which becomes $O\tilde{~}(\alpha^2 N)$ when α varies.

In this work, we are interested in "intermediate" situations, where the displacement rank may be more than constant. Then, the previous complexity results are satisfactory (quasi-linear) with respect to N, but not to α. We improve on this by reintroducing fast dense linear algebra into operations involving the generators of the given matrix. Hence we denote by ω a feasible exponent for linear algebra, that is, a real number such that $n\times n$ matrices over k can be multiplied in $O(n^\omega)$ operations in k. By the results of Coppersmith and Winograd [6], one can take $\omega \le 2.38$.

Theorem 1 *Let $M \in k^{N\times N}$ that is either Toeplitz-, Hankel-, or Vandermonde-like. Given generators G, H for M in $k^{N\times\alpha}$, and a vector b, one can find a random solution to the system $Mx = b$, or certify that none exists, in randomized time $O\tilde{~}(\alpha^{\omega-1}N)$.*

Note that when $\alpha \simeq N$ (i.e. when the matrix is loosely, or almost not structured), our complexity result is in $O\tilde{~}(N^\omega)$, that is, it matches that obtained using classical dense linear algebra.

In the Toeplitz case, we use Kaltofen's extension of Morf's and Bitmead and Anderson's algorithm. After suitable regularizations (described in [4]), the core of the algorithm in the Toeplitz case consists in a divide-and-conquer process, that mimics Strassen's reduction of matrix inversion to matrix multiplication, always working with the generators of the matrices involved. As it turns out, the cost $C(N, \alpha)$ of this algorithm satisfies the following recurrence:

$$C(N, \alpha) = 2C(\frac{N}{2}, \alpha) + O(\mathsf{K}(N, \alpha)) + O\tilde{~}(\alpha^{\omega-1}N),$$

where $\mathsf{K}(N, \alpha)$ denotes the cost of multiplying an $N \times N$ Toeplitz-like matrix of displacement rank α by α vectors.

The key tool to perform this operation is the so-called *Gohberg-Semencul* formula, which enables one to write a Toeplitz-like matrix M of displacement rank α as a sum of α terms of the form $L_i U_i$, where L_i is a lower triangular Toeplitz matrix and U_i an upper triangular Toeplitz matrix which can be read off the generators of M. Using fast polynomial multiplication [7], a single matrix-vector product $v \mapsto L_i U_i v$ can be done in time $O\tilde{~}(N)$, so that the matrix-vector product $v \mapsto \sum_{i \le \alpha} L_i U_i v$ can be done in time $O\tilde{~}(\alpha N)$. Considering α vectors $v_1, \dots, v_\alpha$, we obtain the previous estimate $\mathsf{K}(N, \alpha) \in O\tilde{~}(\alpha^2 N)$. Our contribution for the Toeplitz-like case is then summarized in the following proposition.

Proposition 2 *The inequality* $\mathsf{K}(N, \alpha) \in O\tilde{~}(\alpha^{\omega-1}N)$ *holds.*

To prove this proposition, one first rewrites the operation $v_j \mapsto \sum_{i \le \alpha} L_i U_i v_j$ in terms of polynomial multiplication. After a few simplifications, we are left to compute expressions of the form

$$\sum_{i \le \alpha} A_i (B_i V_j \bmod x^N)$$

for $j \le \alpha$, where A_i, B_i and V_i are polynomials in $k[x]$ of degree at most N deduced from L_i, U_i and v_j. Were it not for the innder reduction modulo x^N, performing this computation would be immediate. We use a recursive approach reminiscent of Mulder's *short product* [8], which enables us to get rid of these truncations, replacing them by suitable polynomial matrix multiplications, yielding our claim.

The Vandermonde case is reduced to the Toeplitz case using a slight extension of the reduction presented in [9], and relies on similar polynomial matrix multiplication techniques.

2 Applications

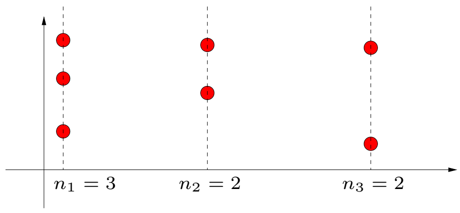
A first application is Padé-Hermite approximation: given m powers series $f_1, \dots, f_m$ in $k[[x]]$ and integers $d_1, \dots, d_m$, we want to find polynomials $p_1, \dots, p_m$ in $k[x]$ such that $\deg p_i \le d_i$ for all i, and $p_1 f_1 + \cdots + p_m f_m = O(x^\sigma)$, with $\sigma = \sum_{i=1}^m (p_i + 1) - 1$. This is a linear problem, with a Toeplitz-like structure; we deduce the following corollary, improving Beckermann and Labahn's $O\tilde{~}(m^\omega \sigma)$ deterministic result [10, 11]. Recently, following [11], Storjohann announced a deterministic $O\tilde{~}(m^{\omega-1}\sigma)$ result in [12].

Corollary 3 *One can compute $p_1, \ldots, p_m$, randomly sampled in the solution set, in randomized time $O\tilde{}(m^{\omega-1}\sigma)$.*

The second application is bivariate polynomial interpolation. Without loss of generality, assume that the sample points are ordered as

$$\begin{array}{ccc} P_{1,1} = (x_1, y_{1,1}) & \ldots & P_{1,n_1} = (x_1, y_{1,n_1}) \\ & \ldots & \\ P_{s,1} = (x_s, y_{s,1}) & \ldots & P_{s,n_s} = (x_s, y_{s,n_s}), \end{array}$$

with $n_1 \geq \cdots \geq n_s$, and let $N = n_1 + \cdots + n_s$ be the total number of points. An example with $s = 3$ and $N = 7$ is as follows:



Given values $v_{i,j}$, $i \leq s$, $j < n_i$, there exists a unique polynomial F of the form

$$F = \sum\nolimits_{i \leq s,\, j < n_i} f_{i,j} x^i y^j$$

such that $F(P_{i,j}) = v_{i,j}$ for all i, j; this follows for instance from Lazard's theorem [13]. This is a linear problem, with Vandermonde-like structure; we deduce the following corollary of Theorem 1.

Corollary 4 *Given the values $v_{i,j}$, the coefficients $f_{i,j}$ can be computed in randomized time $O\tilde{}(n_1^{\omega-1}N)$.*

In particular, suppose that $n_1 = s, n_1 = s, \ldots, n_s = 1$, so that we are interpolating on the simplex of monomials of total degree less than s; here, $N = s(s+1)/2$. Then, our algorithm has subquadratic complexity $O\tilde{}(N^{\frac{\omega+1}{2}}) \subset O\tilde{}(N^{1.69})$ (see [14] for results on bivariate evaluation).

References

1. Kailath, T., Kung, S.Y., Morf, M.: Displacement ranks of matrices and linear equations. J. Math. Anal. Appl. **68** (1979) 395–407
2. Bitmead, R.R., Anderson, B.D.O.: Asymptotically fast solution of Toeplitz and related systems of linear equations. Linear Algebra Appl. **34** (1980) 103–116
3. Morf, M.: Doubling algorithms for Toeplitz and related equations. In: IEEE Conference on Acoustics, Speech, and Signal Processing. (1980) 954–959
4. Kaltofen, E.: Asymptotically fast solution of Toeplitz-like singular linear systems. In: ISSAC'94, ACM (1994) 297–304
5. Pan, V.Y.: Structured Matrices and Polynomials. Birkhäuser Boston Inc. (2001)
6. Coppersmith, D., Winograd, S.: Matrix multiplication via arithmetic progressions. J. Symb. Comput. **9** (1990) 251–280

7. Cantor, D.G., Kaltofen, E.: On fast multiplication of polynomials over arbitrary algebras. Acta Informatica **28** (1991) 693–701
8. Mulders, T.: On short multiplications and divisions. AAECC **11** (2000) 69–88
9. Gohberg, I., Olshevsky, V.: Complexity of multiplication with vectors for structured matrices. Linear Algebra Appl. **202** (1994) 163–192
10. Beckermann, B., Labahn, G.: A uniform approach for the fast computation of matrix-type Padé approximants. SIAM J. Matrix Anal. Appl. **15** (1994) 804–823
11. Giorgi, P., Jeannerod, C.P., Villard, G.: On the complexity of polynomial matrix computations. In: ISSAC'03, ACM (2003) 135–142
12. Storjohann, A.: Notes on computing minimal approximant bases. Technical report, Symbolic Computation Group, University of Waterloo (2006)
13. Lazard, D.: Ideal bases and primary decomposition: the case of two variables. J. Symb. Comput. **1** (1985) 261–270
14. Nüsken, M., Ziegler, M.: Fast multipoint evaluation of bivariate polynomials. In: ESA 2004. Number 3222 in LNCS, Springer (2004) 544–555