Extending Four Displacement Principles to Solve Matrix Equations

David Harlan Wood

Department of Computer and Information Sciences
University of Delaware, Newark, DE 19716, USA
wood@cis.udel.edu

Draft of October 22, 2004; 5:32PM

Abstract

For fixed matrices M and N, the linear transformation $A \mapsto A - MAN$ is called a displacement of the matrix A. Displacements can simplify matrix *equations*, as well as matrices themselves. Four principles to solve matrix equations are identified: 1) a variety of displacements are needed, 2) we want to recover a matrix from its displacement, 3) changing M and N is natural when A is transposed or inverted, and 4) formulas for displacements of matrix products are required. All four principles are extended in this paper. These extensions are illustrated using the class of Krylov matrices, which includes circulant, Vandermonde, Toeplitz, Hankel and other structured matrices as simple special cases.

0 Introduction

Displacement is an operation that can simplify matrices, but we want to use displacement to simplify matrix *equations*. This goal is greatly facilitated by results on finding the displacement of the product of two matrices. Pan's product formula [Pan90] illustrates one of four basic principles for applying displacements to matrix equations. These four principles are listed at the end of this introductory section. The four sections following are each devoted to extending one of these four principles. These extensions indicate that displacements are suitable for a larger class of problems than are presently treated. We do not attempt to sketch an unknown terrain; rather, we only establish four outposts there. In fact, we do not sketch even the *known* terrain, but merely indicate some familiar landmarks.

A general $m \times m$ matrix has m^2 elements. However, many matrices of interest have special structure, for example, Vandermonde and Toeplitz matrices. The m^2 elements of structured matrices are typically generated by $O(m)$ parameters. The objective of displacement methods is to reduce the number of arithmetic calculations for matrix operations on structured matrices by a factor of $O(m/m^2)$ (or more). For some structured matrices it is well known that this can be accomplished through using displacement methods [CK91] and [HR84].

To give particular examples of some of our results and to compare them with known results, consider the problem of solving a Vandermonde system of linear equations. For simplicity, let us assume that we do not encounter any technical difficulties such as singularity of principle minors of intermediate results. The problem of interest is to solve $V^t x = b$, where V is a $m \times m$ Vandermonde matrix. (Without the transpose, we would have $Vx = b$, which would amount to polynomial interpolation, for which there are well-known fast algorithms.) In a 1990 paper[Pan90], Pan first solves $V^t V z = b$ for z and then recovers x from $x = Vz$. The key idea [CKL89] is that $V^t V$ happens to be a Hankel matrix. Once the Hankel matrix is constructed, one can find z in $\mathrm{O}(m \lg^2 m)$ operations with a known algorithm [AG88]. The remaining step of computing $x = Vz$ from z is equivalent to polynomial evaluation at m points. Hence, this step can also be done in $\mathrm{O}(m \lg^2 m)$ operations. Furthermore, by considering the usual displacement of a Vandermonde matrix and a theorem of Kailath, et al. [KKM79], it is easy to see that V^{-1} is uniquely determined by a vector x that satisfies an equation of the type we are considering, $V^t x = b$.

These known results for a Vandermonde system are generalized in five ways in this paper. First, we use Krylov matrices, a very general class of matrices that includes circulant, Vandermonde, Toeplitz, Hankel and many more structured matrices as simple special cases. Second, we introduce appropriate displacements for Krylov matrices and show how to invert these displacements. Third, for a given Krylov matrix K_1, we show how to find another matrix K_2 such that the equation $K_1^t K_2 z = b$ can also be solved in $\mathrm{O}(m \lg^2 m)$ operations, even though K_1 need not be a Vandermonde matrix. Fourth, we show how the inverse of a Krylov matrix K can be constructed from such a solution vector z. Fifth, we obtain our results by exploiting a *new formula* for the displacement of the product of two matrices. Pan's product formula is unsuitable for this purpose.

There is strong interest in displacement methods for numerical computation, but there are many questions about the numerical stability of these methods in such approximate computation [Bun85]. These questions need to be clarified for displacement methods using floating point arithmetic. Our methods are, however, suitable for computations in exact arithmetic, for example in computer algebra systems. Displacement methods are advantaged when exact arithmetic is used, because then *rank is unambiguous.* After all, the standard use of displacement methods is to map structured matrices into matrices of significantly reduced rank. Symbolic computations concerning polynomials and their interpolation naturally involve structured matrices including Sylvester (resultant), Bezoutian, and Vandermonde matrices.

In this paper, a *displacement* of a matrix A will refer to the operation

$$\Delta_{M,N}(A) = A - MAN, \tag{1}$$

for some fixed matrices M and N. Unless stated otherwise, all matrices are real, rectangular matrices. All matrices are assumed to be *conformable*, that is, restricted only by the definitions of the indicated operations of addition, multiplication, inversion, etc. For example, if A is an $m \times n$ matrix, then for the displacement $\Delta_{M,N}(A)$ to be defined, conformablity requires M to be an $m \times m$ matrix and N to be an $n \times n$ matrix. We index rows and columns starting at zero, as in $m = 0, 1, 2, \ldots, m-1$. We use A^t to denote the *transpose* of A. Obvious modifications correspond to complex matrices and Hermitian transposes. The notation A^c is used for the *countertranspose* of the matrix A, which is A reflected about its main counterdiagonal. We use e_j to denote the unit vector found in the jth column of the identity matrix. Its length will be apparent from the context. The square *unit shift matrix*, having ones on the first subdiagonal and zeros elsewhere, is denoted by Z. The *circular shift matrix* Π^t is the unit shift matrix with an additional one in the upper right corner. The *spectrum* of a matrix or operator A is a list of all of its eigenvalues, and is denoted by $\sigma(A)$.

Four Principles of Displacement for Matrix Equations

With judicious choices for the matrices M and N, displacement of a matrix *equation* such as $I = AA^{-1}$ or $AX = B$ can result in a simpler equation. Four basic principles presently available for this approach are the following.

Principle 1 Displacement matrices M and N that reduce rank are sought. Effective displacements are currently known for circulant, Toeplitz, Hankel, Vandermonde, Hilbert, Cauchy and many other structured matrices. These special displacements can reduce the rank of other matrices also. These matrices are referred to as *near-circulant*, *near-Toeplitz*, etc.

Principle 2 It is necessary to *invert* displacements. The usual technique [KKM79] of inversion when the displacement is 1-to-1 exploits the inversion of rank one matrices. If the displacement is given in the form $\Delta_{M,N}(X) = \sum_i x_i y_i^t$, where the x_i and y_i are vectors, then $X = \sum_i \Delta_{M,N}^{-1}(x_i y_i^t)$, where the inverse has a special form if M and N are the unit shift matrix Z or its transpose.

Principle 3 Matrices associated with a given matrix A may call for *different* displacement operators. For example: if for a matrix A we have $\text{rank}(\Delta_{M,N}(A)) = r$ and if A^t is its transpose, then $\text{rank}(\Delta_{N^t,M^t}(A^t)) = r$ and if A is invertible, then the *reversed displacement,* interchanging the roles of M and N, yields $\text{rank}(\Delta_{N,M}(A^{-1})) = r$ [KKM79].

Principle 4 For using displacement on matrix *equations*, a formula for the displacement of a product of matrices is needed. The earliest explicit result is [Pan90], $\Delta_{M,N}(AB) = \Delta_{M,K_1}(A)B + MAK_1\Delta_{K_2,N}(B) - MA\Delta_{K_1,K_2}(I)BN$, where A, B, K_1, K_2, M, and N are arbitrary matrices (of conformable sizes).

This paper gives extensions of each of these four principles. For Principle 1, Krylov matrices are introduced to exemplify and extend the class of matrices where displacement produces low rank results. Additional displacement inversion techniques serve to extend Principle 2. Principle

3 for nonsingular matrices is generalized to singular and rectangular matrices. Additional product formulas are given that supplement those of Principle 4 for the displacement of a product of matrices. These formulas are useful for exploiting Principle 3.

The following sections, numbered 1 through 4, concern the extensions of each of the four principles. The paper ends with a section of discussion.

1 Krylov Matrices are Companions of Displacement

To extend Principle 1, Krylov matrices are introduced to exemplify and extend the class of matrices where displacement is known to be effective. This broadens the class of matrices M and N used in displacements $\Delta_{M,N}$, and develops the fact that *any* square matrix M can be used in a displacement which will reduce many Krylov matrices to rank one or even rank zero matrices. This is a significant extension because the matrices M and N used in current displacements are almost always limited to diagonal matrices, the unit shift matrix, the circular shift matrix or transposes of these. Because of this limitation remarkably few distinct displacements are in current use, yet they have been effective with circulant, resultant, Toeplitz, Hankel, Vandermonde, Hilbert, Cauchy, Leslie-Toeplitz, r-Toeplitz, Bezoutian, Lowner, and Gaussian matrices.

This section discusses displacements of Krylov matrices. Circulant, Vandermonde Toeplitz, Hankel and may more structured matrices are simple special cases of Krylov matrices. Thus, we extend and unify the classes of matrices suitable for displacement methods.

Let M be a given $m \times m$ matrix. The *Krylov matrix* with a *generator matrix* M and first column v is

$$K(v, M) = [v \; Mv \; M^2v \cdots M^{m-1}v]. \tag{2}$$

For a given matrix M, the set of all Krylov matrices $K(v, M)$ obviously form a vector space of dimension m with a basis $\{K(e_j, M) \mid j = 0, 1, 2, \ldots, m-1\}$. (Displacement methods are typically applied to $m \times m$ matrices that are determined by O(m) parameters.) Multiplying Eq 2 by M and using the Cayley-Hamilton theorem gives the fundamental equation

$$MK(v, M) = [Mv \; M^2v \cdots - \sum_{k=0}^{m-1} c_k M^k v] = K(v, M)C_M, \tag{3}$$

where C_M is the *companion* matrix of M. It is the same as the unit shift matrix Z except its last column is $[-c_0, -c_1, \ldots, -c_{m-1}]^t$. Here c_i denotes the coefficient of λ^i in the characteristic polynomial of M, namely

$$\det(\lambda I - M) = \lambda^m + c_{m-1}\lambda^{m-1} + \cdots + \lambda c_1 + c_0 \tag{4}$$

Theorem 1.1 (Characterization of Krylov matrices)

A matrix K is a Krylov matrix generated by its first column and the matrix M if and only if

$$MK = KC_M \tag{5}$$

and C_M is the companion matrix of M.

Proof The implication in one direction was shown above. Comparing corresponding columns on each side of Eq 5 shows that if v is the first column of K, the second column of K is equal to Mv, and so on. Since C_M has the same characteristic equation as M, the last columns are also equal.$\square$

It is well known that the characteristic polynomial of C_M is the same as that of M. Also, if C_M^{-1} exists, it is the same as Z^t, except its first column is $[-c_1/c_0, -c_2/c_0, \ldots, -c_{m-1}/c_0, -1/c_0]^t$. Another useful fact about companion matrices states [LT85]

$$C_M S = SC_M^t, \tag{6}$$

where S is an upper triangular Hankel matrix with its first row equal to $[c_1, c_2, \ldots, c_{m-1}, 1]$. The necessarily nonsingular and symmetric matrix S is one of two *symmetrizers* associated with a general companion matrix.

Using Eq 5, we see that any nonsingular $m \times m$ matrix M occurs in a displacement that maps the m dimensional vector space of Krylov matrices of the form $K(v, M)$ to zero because

$$\Delta_{M, C_M^{-1}}(K(v, M)) = 0 \tag{7}$$

for any vector v. This will be referred to as a *companion displacement.* A variant of this displacement that does not require the matrix M to be nonsingular is

$$\Delta_{M, Z^t}(K(v, M)) = [v\ 0\ 0 \cdots 0], \tag{8}$$

where the right hand side is a rank one matrix. This will be referred to as a *Krylov displacement.*

Notice that if we choose any companion matrix C and choose any nonsingular matrix A, then A is the is the Krylov matrix $K(Ae_0, M)$ for some matrix M whose companion matrix is C. To see this, just solve Eq 5 for M and observe that it has the same characteristic polynomial as C_M. The challenge in applying displacement methods is to find a companion matrix C such that the generator matrix M has a simple form that makes it possible to apply and invert the displacement efficiently. Inverting displacements is addressed in the next section.

It obvious that circulant matrices are Krylov matrices generated by the circular shift matrix and that Vandermonde matrices are Krylov matrices generated by a diagonal matrix. It is less well known that Toeplitz matrices, Hankel matrices, and confluent Vandermonde matrices are also all Krylov matrices having simple generator matrices. These three examples are immediately below. Thus, the class of Krylov matrices includes many of the standard cases where displacement methods are applied.

Example 1.1.1 (Nonsingular Toeplitz matrices as Krylov matrices)

Suppose there is a vector p that satisfies

$$Tp = t, \tag{9}$$

where t is the last column of ZT. In particular, if the matrix T is nonsingular, then p must exist and be unique. Using the vector p satisfying Eq 9, we proceed in the spirit of Gover [Gov89] by introducing a companion matrix we call C with its last column given by the vector y. By comparing ZT and TZ, it is easy to see that

$$C^c T = TC. \tag{10}$$

(Recall that C^c denotes the countertranspose of C, which is the reflection of C about its main counterdiagonal.) Theorem 1.1 tells us that this equation means the Toeplitz matrix T is the Krylov matrix generated by the first column of T and the countertranspose of C,

$$T = K(Te_0, C^c). \tag{11}$$

Example 1.1.2 (Nonsingular Hankel matrices as Krylov matrices)

A similar result holds for Hankel matrices. We can see that

$$C^t H = HC, \tag{12}$$

where the last column of the companion matrix C is the solution p of

$$Hp = h, \tag{13}$$

and where h is the last column of $Z^t H$. Such a p must exist if H is nonsingular. As a consequence of Theorem 1.1 we see

$$H = K(He_0, C^t). \tag{14}$$

Example 1.1.3 (Confluent Vandermonde matrices as Krylov matrices)

The elements of each row of a Vandermonde matrix consist of successively higher powers of some variable x starting with the zeroth power. It is clear that a regular Vandermonde matrix is a Krylov matrix generated by a diagonal matrix and a first column of all ones.

For a given variable x a *confluent* Vandermonde matrix can have a group of $\mu + 1$ rows, all of which involve a variable x. The first row again consists of successively higher powers of x starting with the zeroth power in the first column. Each of the remaining μ rows is the derivative of the row above it. When such a group of rows starts at row i, the general element within the group is zero when $j < k$ and otherwise is given by

$$v(i+k, j) = \partial^k x^j = x^{j-k} j!/(j-k)!\,, \tag{15}$$

where $k = 0, 1, 2, \ldots, \mu$. Differentiation gives a recurrence relation between successive rows, but to have a Krylov matrix we need a recurrence relation between successive columns. This can be obtained as follows:

$$\begin{aligned} v(i+k, j) &= \partial^k x^j \\ &= \partial^k (x \cdot x^{j-1}) \\ &= \sum_{\ell=0}^{k} \binom{k}{\ell} \partial^\ell x \cdot \partial^{k-\ell} x^{j-1} \\ &= x\, \partial^k x^{j-1} + k\, \partial^{k-1} x^{j-1} \\ &= x\ v(i+k, j-1) + k\ v(i+k-1, j-1). \end{aligned} \tag{16}$$

This last equation is a recurrence relation giving column j as a linear function of column $j-1$.

Thus, a confluent Vandermonde matrix can be realized as a Krylov matrix generated by a block-diagonal matrix. The $\mu - 1 \times \mu - 1$ blocks are of the form $xI + CZ$ where the counting matrix C has elements that are $0, 1, 2, \ldots, \mu$ on the diagonal, and zero elsewhere. Corresponding to this block, the first column of the generated matrix is the unit vector is e_0. An example confluent Vandermonde matrix consisting of one block is

$$K(e_0, xI + CZ) = K(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x & 0 & 0 & 0 & 0 \\ 1 & x & 0 & 0 & 0 \\ 0 & 2 & x & 0 & 0 \\ 0 & 0 & 3 & x & 0 \\ 0 & 0 & 0 & 4 & x \end{pmatrix}) = \begin{pmatrix} 1 & x & x^2 & x^3 & x^4 \\ 0 & 1 & 2x & 3x^2 & 4x^3 \\ 0 & 0 & 2 & 6x & 12x^2 \\ 0 & 0 & 0 & 6 & 24x \\ 0 & 0 & 0 & 0 & 24 \end{pmatrix}. \tag{17}$$

2 Inversion of Displacements

Additional inversion techniques serve to extend Principle 2 concerning operators that invert displacements. It is a classical problem in matrix equations [HR84] [BGR90] to recover a matrix X from knowing $\Delta_{M,N}(X) = Y$, with given Y. This is usually called a Stein equation. Since we also want to solve this equation efficiently, we must be prepared to accept restrictions on the matrices M and N.

Displacement is a linear operator. Its spectrum is known from Stephanos' theorem [LT85],

$$\sigma(\Delta_{M,N}) = \{1 - \lambda\mu \mid \lambda \in \sigma(M) \text{ and } \mu \in \sigma(N)\}. \tag{18}$$

Theorem 2.1 (Inversion of Krylov and reversed Krylov displacements)
Any Krylov displacement Δ_{M,Z^t} or reversed Krylov displacement $\Delta_{Z^t,M}$ is linear and invertible. The same is true for $\Delta_{M,Z}$ and $\Delta_{Z,M}$.

Proof Stephanos' theorem implies each of these operators has its m^2 eigenvalues all equal to 1.

We call the next two results *local* inversion techniques because they require having some decomposition of the displacement into a sum of rank one parts. It may be helpful that these techniques do not require a *minimal* decomposition. For example, with prior knowledge of the rank, r, of the displacement and using exact arithmetic, it is highly probable that images of $2r+2$ (more than minimal) randomly generated vectors will generate a valid decomposition as a sum of $r+1$ rank one matrices.

The inverses of Krylov displacements and reversed Krylov displacements acting on rank one matrices have special forms. Since for any vector v, $K(v,Z)$ is a Toeplitz matrix and $K(v,Z^t)$ is Hankel matrix, we recognize at least one triangular Toeplitz or Hankel matrix in every term of the following inversion formulas.

Theorem 2.2 (Local inversion of Krylov and reversed Krylov displacements)
Let x_i and y_i denote vectors.

$$\text{If} \quad \Delta_{M,Z^t}(X) = \sum_i x_i y_i^t \quad \text{then} \quad X = \sum_i K(x_i, M) K(y_i, Z)^t. \tag{19}$$

$$\text{If} \quad \Delta_{Z^t,M}(X) = \sum_i x_i y_i^t \quad \text{then} \quad X = \sum_i K(x_i, Z^t) K(y_i, M^t)^t. \tag{20}$$

$$\text{If} \quad \Delta_{M,Z}(X) = \sum_i x_i y_i^t \quad \text{then} \quad X = \sum_i K(x_i, M) K(y_i, Z^t)^t. \tag{21}$$

$$\text{If} \quad \Delta_{Z,M}(X) = \sum_i x_i y_i^t \quad \text{then} \quad X = \sum_i K(x_i, Z) K(y_i, M^t)^t. \tag{22}$$

Proof We prove Eq 20, which is representative. Since $K(y_i,Z)^t$ is an upper triangular Toeplitz matrix, it commutes with Z^t. This yields

$$\Delta_{M,Z^t}(K(x_i,M)K(y_i,Z)^t) = [K(x_i,M) - MK(x_i,M)Z^t]K(y_i,Z)^t = x_i y_i^t \tag{23}$$

by using Eq 8. The result follows from linearity.□

Corollary 2.2.4 (Local inversion of Δ_{Z,Z^t}, Kailath, et al. 1979)
In the cases $M = Z$ or $M = Z^t$, the above theorem gives results due to Kailath and others [KKM79] [CK91] where the indicated Krylov matrices *all* become triangular Toeplitz and/or Hankel matrices.

Since an $m \times m$ Toeplitz or Hankel matrix can be multiplied by an arbitrary matrix with $O(m^2 \lg m)$ arithmetic operations, the results of Corollary 2.2.4 are advantageous if X is to be multiplied by a matrix.

We call the following two results *global* inversion techniques because we recover the matrix X from $\Delta_{M,N}(X) = Y$ without decomposing Y.

Theorem 2.3 (Global inversion of a Krylov displacement)
If

$$\Delta_{M,Z^t}(X) = Y, \tag{24}$$

then X *is uniquely determined by*

$$X = \sum_{k=0}^{m-1} M^k Y Z^{kt}, \tag{25}$$

where m *is the dimension of the matrix* Z*.*

Proof Substitute Y into Eq 26. The sum collapses to X since $Z^{mt} = 0$. Recall that X is unique.□

We remark that the same proof goes through for any conformable matrices, provided that the sum extends to infinity and that the product of the spectral radii of M and N is less than 1. [LT85].

Eq 26 is most useful if the products $M^k Y$ can be generated with fewer than $O(m^2)$ arithmetic operations. For example, let M or Y be a permutation matrix times a diagonal matrix having mostly zeros and ones on its diagonal. The following application inverts a popular displacement by using Eq 26.

Application 2.3.5 (Global inversion of Δ_{Z,Z^t} Wood 1993)

Consider the inversion of the Δ_{Z,Z^t} displacement operator. The matrix X is recovered from $\Delta_{Z,Z^t}(X) = Y$ by using Eq 26, that is, by forming cumulative sums down the diagonals of Y which takes $O(m^2)$ additions with one processor or $O(\lg m)$ time steps with $O(m^2)$ processors [Woo93]. If X is going to be added to another matrix, this global result is preferable to that of Corollary 2.2.4.

Of course, not all displacements are 1-to-1. This can be an advantage if the desired result lies entirely in the kernel of the displacement.

Theorem 2.4 (The kernel of companion displacement with nonsingular generator)
Let C_M *be the companion matrix of a given nonsingular matrix* M*. The matrix* X *satisfies* $\Delta_{M,C_M^{-1}}(X) = 0$ *if and only if* X *is a Krylov matrix* $X = K(v, M)$*, for some vector* v*. If any column of* X *is known, then* X *is uniquely determined.*

Proof Theorem 1.1.□

In the next application, we seek our answer in the kernel of a displacement. This application is typical of results on inverses of structured matrices: the problem of finding the inverse of the matrix is reduced to the problem of *finding only a few vectors* that determine the inverse matrix.

Application 2.4.6 (The inverse of a Krylov matrix from a nonsingular generator)

A Krylov matrix K generated by a nonsingular $m \times m$ matrix M satisfies $\Delta_{M,C_M^{-1}}(K) = 0$. If K is nonsingular, it follows that $K^{-1} = C_M K^{-1} M^{-1}$. Invoking the symmetrizer matrix S in Eq 6, we have $S^{-1}K^{-1} = S^{-1}C_M K^{-1}M^{-1} = C_M^t S^{-1}K^{-1}M^{-1}$. Transposing and multiplying by M^t gives $M^t K^{-t} S^{-1} = K^{-t}S^{-1}C_M$, because S is symmetric. By Theorem 2.4, $K^{-t}S^{-1}$ is a Krylov matrix $K(y, M^t)$ and is uniquely determined by its first column y which satisfies

$$K^t y = S^{-1} e_0 = e_{m-1}, \tag{26}$$

where e_0 and e_{m-1} denote the first and last columns of the $m \times m$ identity matrix. The vector y completely determines the inverse of the matrix K,

$$K^{-1} = SK(y, M^t)^t. \tag{27}$$

This is has the form of Barnett's factorization [LT85] for a nonsingular Bezoutian matrix (which is necessarily the inverse of a Toeplitz matrix [HR84]).

3 Displacements of Associated Matrices

Roughly speaking, matrices associated with a given matrix A may call for different displacement operators. The simplest case is trivial, but we record it as a theorem for easy reference.

Theorem 3.1 (Displacement of the transpose of a matrix)
For any conformable matrices A, M, and N,

$$\operatorname{rank}(\Delta_{M,N}(A)) = \operatorname{rank}(\Delta_{N^t,M^t}(A^t)). \tag{28}$$

Proof Taking the transpose of $\Delta_{M,N}(A)$ does not change its rank.□

Here the matrix A can, of course, be rectangular.

Even with no restrictions on M and N (aside from conformablity), there is a startlingly strong theorem stating $\operatorname{rank}(\Delta_{M,N}(A)) = \operatorname{rank}(\Delta_{N,M}(A^{-1}))$, provided A is nonsingular [KKM79]. This result is also a corollary of our next theorem. We call $\Delta_{N,M}(A^{-1})$ the *reversed displacement* of A^{-1}.

Why does the reversed displacement arise in such a fundamental way? The answer does not seem to have been remarked elsewhere, but it is easy to see that the reversed displacement is merely a slightly disguised form of the adjoint of the operator $\Delta_{M,N}$. After all, for fixed matrices M and N, the displacement $\Delta_{M,N}$ can be regarded in a standard way as a linear operator on a vector space. In the present context, the adjoint of this operator enters in the guise of the reversed displacement,

$$\Delta_{N,M}(A^t) = [\operatorname{adjoint}(\Delta_{M,N})(A)]^t. \tag{29}$$

The displacement used for the transpose of a matrix in Theorem 3.1 also has a representation as an adjoint,

$$\Delta_{N^t,M^t}(A^t) = [\mathrm{adjoint}(\Delta_{M^t,N^t})(A)]^t. \tag{30}$$

We want to generalize the result $\mathrm{rank}(\Delta_{M,N}(A)) = \mathrm{rank}(\Delta_{N,M}(A^{-1}))$, without *any* restrictions on the matrix A. In particular the matrix A may be singular or even rectangular. By a *generalized inverse* of a matrix A, we mean any matrix A^+ that satisfies

$$AA^+A = A \quad \text{and} \quad A^+AA^+ = A^+. \tag{31}$$

Any matrix always has at least one generalized inverse satisfying these two conditions [LT85]. (Various specialized types of generalized inverses are defined by adding additional conditions [RM72].)

Theorem 3.2 (Reversed displacement of a generalized inverse)
For any conformable matrices A, M, and N, if A^+ is any generalized inverse of the matrix A, then

$$\mathrm{rank}(A^+A\Delta_{N,M}(A^+)AA^+) = \mathrm{rank}(AA^+\Delta_{M,N}(A)A^+A). \tag{32}$$

Proof Both of the partitioned matrices

$$\begin{pmatrix} I & 0 \\ -A^+ANA & I \end{pmatrix} \begin{pmatrix} A & AA^+MAA^+ \\ A^+ANA^+A & A^+ \end{pmatrix} \begin{pmatrix} I & -A^+MAA^+ \\ 0 & I \end{pmatrix}$$

and

$$\begin{pmatrix} I & -AA^+MA \\ 0 & I \end{pmatrix} \begin{pmatrix} A & AA^+MAA^+ \\ A^+ANA^+A & A^+ \end{pmatrix} \begin{pmatrix} I & 0 \\ -ANA^+A & I \end{pmatrix}$$

have the same rank because they are equivalent. In addition, both are block diagonal, giving

$$\mathrm{rank}(A) + \mathrm{rank}(A^+A\Delta_{N,M}(A^+)AA^+) = \mathrm{rank}(AA^+\Delta_{M,N}(A)A^+A) + \mathrm{rank}(A^+).$$

Applying rank inequalities to Eq 32 shows that $\mathrm{rank}(A) = \mathrm{rank}(A^+)$, and so the theorem is proved.□

It is known that invertibility of A suffices [KKM79] to independently give the following result.

Corollary 3.2.7 (Reversed displacement of an inverse, Kailath, et al. 1979)
For any invertible matrix A and any conformable matrices M and N,

$$\mathrm{rank}(\Delta_{M,N}(A)) = \mathrm{rank}(\Delta_{N,M}(A^{-1})). \tag{33}$$

Proof If A is invertible $A^+ = A^{-1}$ satisfies Eq 32. Substitute this into Eq 33.

One might hope for relations between the ranks of the displacements $\Delta_{M,N}(A)$ and $\Delta_{N,M}(A^+)$, without using reversed displacements $\Delta_{M,N}(A)$ or $\Delta_{N,M}(A^+)$. However, if the matrix A is not square, the matrices M and N can not be conformable to *both* $\Delta_{M,N}(A)$ and $\Delta_{M,N}(A^+)$. Furthermore, even if A is square and invertible, $\text{rank}(\Delta_{M,N}(A)) = \text{rank}(\Delta_{M,N}(A^{-1}))$ may be either true or false.

Nevertheless, Theorem 3.2 has a simple interpretation. The two combinations A^+A and AA^+ that clutter Eq 33 are *projection matrices.* The projection A^+A maps into the domain of A, while the projection AA^+ maps into the range of A. (Recall that P is a projection if and only if $P^2 = P$. Eq 32 immediately yields $(A^+A)(A^+A) = A^+A$ and $(AA^+)(AA^+) = AA^+$.)

4 Displacements of Products

In order to work with matrix equations, we need to be able to compute the displacement of a product of two matrices. Surprisingly, formulas for this are comparatively recent [Pan90] [Woo93] [NW95].

Theorem 4.1 (Displacement of a matrix product, Pan 1990)
For arbitrary matrices A, B, K_1, K_2, M, and N (of conformable sizes), the displacement of the product AB is given by

$$\Delta_{M,N}(AB) = \Delta_{M,K_1}(A)B + MAK_1\Delta_{K_2,N}(B) - MA\Delta_{K_1,K_2}(I)BN. \tag{34}$$

Notice that the the left hand side of this equation is independent of the matrices K_1 and K_2, which are arbitrary except for their sizes.

Theorem 4.2 (Displacement of a matrix product, Nguyen and Wood 1991)
For arbitrary but conformable matrices A, B, K, K, M,N_1, and N_2, the displacement of the product AB is given by

$$\Delta_{M,N_1}(AB) = \Delta_{M,K}(A)B - MA\Delta_{K,N_2}(B)N_1 + MAKB\Delta_{N_2,N_1}(I). \tag{35}$$

Notice that the the left hand side of this equation is independent of the matrices K and N_2, which are arbitrary except for their sizes.

For arbitrary but conformable matrices A, B, K, K, M_1 M_2, and N, the displacement of the product AB is given by

$$\Delta_{M_1,N}(AB) = -M_1\Delta_{M_2,K}(A)BN + A\Delta_{K,N}(B) + \Delta_{M_1,M_2}(I)AKBN. \tag{36}$$

Notice that the the left hand side of this equation is independent of the matrices K and M_2, which are arbitrary except for their sizes.

Proof Expand both sides of the equations.□

Actually,the formulas found in [NW95] are more general than those in Theorem 4.2.

The two formulas in Theorem 4.2 are related to each other, but both are quite different from the formula in Theorem 4.1. The results of Theorem 4.2 are more compatible with the reversed displacement needed to exploit Corollary 3.2.7, as the following example illustrates.

Example 4.2.8 (Displacement of A^{-1} from applying product formulas to $I = AA^{-1}$)

Given a matrix A, we want to find a displacement of A^{-1} using the equation $I = AA^{-1}$. We solve one of the product formulas given above for the displacement of A^{-1}. If we are aware of a displacement $\Delta_{M,N}(A)$ that has low rank, then Corollary 3.2.7 motivates us to use $\Delta_{N,M}(A^{-1})$ since it is guaranteed to have the same low rank.

On one hand, applying Pan's formula, Eq 35, with $B = A^{-1}$ gives

$$\Delta_{M,M}(AA^{-1}) = \Delta_{M,N}(A)A^{-1} + MAN\Delta_{N,M}(A^{-1}) - MA\Delta_{N,N}(I)A^{-1}N. \tag{37}$$

We have no choice but to cope with $\Delta_{M,M}(I) = I - M^2$ and $\Delta_{N,N}(I) = I - N^2$. But these may not have low rank because M and N are already constrained by requiring $\Delta_{M,N}(A)$ to have low rank.

On the other hand, by way of contrast, applying Theorem 4.2 gives

$$\Delta_{M,N_1}(AA^{-1}) = \Delta_{M,N}(A)A^{-1} - MA\Delta_{N,M}(A^{-1})N_1 + MANA^{-1}\Delta_{M,N_1}(I) \tag{38}$$

and

$$\Delta_{M_1,M}(AA^{-1}) = -M_1\Delta_{M,N}(A)A^{-1}M + A\Delta_{N,M}(A^{-1}) + \Delta_{M_1,M}(I)ANA^{-1}M. \tag{39}$$

In the first of these equations, we can solve for $\Delta_{N,M}(A^{-1})$ with the freedom to choose N_1 for our convenience. For example, we could minimize the rank of $\Delta_{M,N_1}(I) = I - MN_1$. In the second equation, M_1 is similarly at our disposal for minimizing the ranks of $\Delta_{M_1,M}(I) = I - M_1M$.

Specializing A to a Toeplitz matrix in any one of the three above equations can give a Gohberg-Semencul formula for the inverse of a Toeplitz matrix [HR84] as in [Woo93].

Example 4.2.9 (Krylov inverse from a product formula, Nguyen and Wood 1991)

Let an $m \times m$ matrix M be given. Let K be a nonsingular Krylov matrix $K(v, M)$ for some vector v. Let us seek the inverse of this matrix. We first find the displacement of K^{-1} and then invert the displacement to obtain K^{-1} itself [NW95]. The displacement $\Delta_{M,Z^t}(K)$ is known to be of rank one by Eq 8 which means $\Delta_{Z^t,M}(K^{-1})$ is of rank one by Corollary 3.2.7. Applying Eq 37 to the product $K^{-1}K$ gives

$$\Delta_{Z^tZ}(I) \quad = \quad \Delta_{Z^tM}(K^{-1})K + Z^tK^{-1}MK\Delta_{Z^tZ}(I), \tag{40}$$

because $\Delta_{M,Z^t}(K)Z = 0$ by Eq 8. Solving for the displacement of K^{-1} yields

$$\Delta_{Z^t M}(K^{-1}) = \Delta_{Z^t Z}(I)K^{-1} - Z^t K^{-1} M K \Delta_{Z^t Z}(I) K^{-1}. \tag{41}$$

Now, $\Delta_{Z^t Z}(I) = e_{m-1}e_{m-1}^t$, where e_{m-1} is the last column of the $m \times m$ identity matrix. Using Eq 5, we see that $Z^t K^{-1} M K \Delta_{Z^t Z}(I) = Z^t C_M e_{m-1} e_{m-1}^t = Z^t c\, e_{m-1}^t$, where c is the last column of the companion matrix of M. If we define y by the equation

$$K^t y = e_{m-1}, \tag{42}$$

we now have $\Delta_{Z^t M}(K^{-1}) = (e_{m-1} - Z^t c) y^t$. Using Eq 21, we obtain

$$K^{-1} = K(e_{m-1} - Z^t c, Z^t) K(y, M^t)^t = SK(y, M^t)^t, \tag{43}$$

where S is the symmetrizer matrix introduced in Eq 6. This result is identical to the one found in Application 2.4.6. In short, the problem of finding the inverse of a Krylov matrix was reduced to solving Eq 43.

Notice that the theorem of Pan (Theorem 4.1) is not suitable for the above example because it introduces $I - Z^{2t}$ which is not of low rank.

Example 4.2.10 (Solution of a Krylov system of equations, Nguyen and Wood 1991)

We can solve [NW95] any system of linear equations involving a nonsingular Krylov matrix or its transpose by generalizing a device used for Vandermonde matrices [CKL89].

Let K_1 denote the nonsingular matrix $K(v, M)$ for some vector v and some matrix M. To solve the equation

$$K_1 x = b, \tag{44}$$

we rewrite it in the form

$$K_2^t K_1 x = K_2^t b, \tag{45}$$

where K_2 is a Krylov matrix $K(u, M^t)$. Here the vector u is chosen to make $K(u, M^t)$ nonsingular (such a u exists because M^t is of full rank because M^t is.)

To solve the related equation

$$K_1^t y = d, \tag{46}$$

for y we use the same matrix K_2 to rewrite this related problem in the form

$$y = K_2 z, \quad \text{where} \quad K_1^t K_2 z = d. \tag{47}$$

In either case, we have to solve an equation involving the product of the transpose of a Krylov matrix on the left with a Krylov matrix on the right. One Krylov matrix is generated by the matrix M and the other is generated by M^t. Consider the product $K_1^t K_2$ as representative. For this matrix, the special displacement $\Delta_{Z,Z}(K_1^t K_2)$ is always of low rank, independent of M, as can be shown by using Theorem 4.2. First, notice that Theorem 3.1 together with Eq 8 implies that $\Delta_{Z,M^t}(K_1^t)$ has rank one. Second, because of Eq 8, the displacement $\Delta_{M^t,Z^t}(K_2)$ is of rank one. Third, $\Delta_{Z^t,Z}(I)$ is also of rank one. These three facts imply that $\Delta_{Z,Z}(K_1^t K_2)$ has rank at most three because Eq 36 shows that

$$\Delta_{Z,Z}(K_1^t K_2) = \Delta_{Z,M^t}(K_1^t)K_2 - ZK_1^t \Delta_{M^t,Z^t}(K_2)Z + ZK_1^t M^t K_2 \Delta_{Z^t,Z}(I). \tag{48}$$

The low rank of this displacement is important because a system of m linear equations with small $\Delta_{Z,Z}$ rank can be solved in $O(m^2)$ arithmetic operations [HR84]. Thus $K_1^t K_2$ can be recovered from Eq 49 by using Eq 22. With additional hypothesis, the operations count can be reduced to $O(m \lg^2 m)$ [AG88], as was done for Vandermonde systems in [Pan90].

Pan's Theorem 4.1 is not suited to the above example unless $I - MM^t$ happens to be of low rank.

5 Discussion

The class of Krylov matrices is a pleasant generalization that includes many of the usual structured matrices (circulant, Vandermonde, Toeplitz, and Hankel matrices) as simple special cases.

Product formulas appear to be a fruitful means of dealing with matrix *equations*. These formulas appeared fairly recently [Pan90] [Woo93] [NW95], but Pan points out that a special case of his formula was implicit in earlier results [CKLA87]. In fact, by using a different approach without explicit product formulas, one can sometimes embed matrices in a larger matrix so that a Schur complement would deal with the desired product [CK91].

With a variety of product formulas to choose from, we need to identify what displacements are natural to use in a given equation. When equations involve the matrices A, A^t, A^{-1}, or A^+, we have the results of Section 3. These are far too limited, and much more work is needed to answer questions like, "If $\Delta_{N,M}(A)$ is of low rank, what are the natural displacements for the matrices Q and Λ in the equation $AQ = Q\Lambda$, where Λ is diagonal?" A particular case of this question is addressed in [Woo93] using Pan's product formula. Also, displacements need to be explored that in some sense simplify the structure of matrices rather than lowering their rank.

The four principles cited in this paper, and the extensions given, seem likely to find a number of interesting applications.

Acknowledgment

I have quoted from joint work with Quyen L. Nguyen, an undergraduate Science and Engineering Scholar in the Department of Computer and Information Science at the University of Delaware. I am pleased to acknowledge how helpful our discussions have been. I want to formally thank both him and the Honors Program of the University of Delaware that sponsored our interaction.

References

[AG88] Gregory S. Ammar and William B. Gragg. Superfast solution of real positive definite Toeplitz systems. *SIAM Journal on Matrix Analysis and Applications*, 9:61–76, 1988.

[BGR90] Joseph A. Ball, Israel Gohberg, and Leiba Rodman. *Interpolation of Rational Matrix Functions*. Birkhäuser Verlag, Boston, 1990.

[Bun85] James R. Bunch. Stability of methods for solving Toeplitz systems of equations. *SIAM Journal on Scientific and Statistical Computing*, 6:349–364, 1985.

[CK91] J. Chun and T. Kailath. Displacement structure for Hankel, Vandermonde, and related (derived) matrices. *Linear Algebra and Its Applications*, 151:199–227, 1991.

[CKL89] J. F. Canny, E. Kaltofen, and Y. Lakshman. Solving systems of non-linear polynomial equations faster. In *Proceedings of the ACM-SIGSAM International Symposium on Symbolic and Algebraic Computations*, pages 34–42, New York, 1989.

[CKLA87] J. Chun, T. Kailath, and H. Lev-Ari. Fast parallel algorithms for QR and triangular factorization. *SIAM Journal on Scientific and Statistical Computing*, 8:899–913, 1987.

[Gov89] M. J. C. Gover. The determination of companion matrices characterizing Toeplitz and r–Toeplitz matrices. *Linear Algebra and Its Applications*, 117:81–92, 1989.

[HR84] Georg Heinig and Karla Rost. *Algebraic Methods for Toeplitz-like Matrices and Operators*. Birkhäuser Verlag, Boston, 1984.

[KKM79] T. Kailath, S.Y. Kung, and M. Morf. Displacement ranks of a matrix. *Bulletin of the American Mathematical Society*, 1:769–773, 1979.

[LT85] Peter Lancaster and Miron Tismenetsky. *The Theory of Matrices, 2nd Edition*. Academic Press, New York, 1985.

[NW95] Quyen L. Nguyen and David H. Wood. Displacement of matrix products. In Géard Cohen, Marc Giusti, and Teo Mora, editors, *Applied Algebra and Error-Correcting Codes II, 11th International Symposium, AAECC-11*, volume 948 of *Lecture Notes in Computer Science*, pages 383–392, Berlin, 1995. Springer-Verlag.

[Pan90] Victor Pan. On computations with dense structured matrices. *Mathematics of Computation*, 55:179–190, 1990.

[RM72] C. R. Rao and S. K. Mitra. *Generalized Inverse of of Matrices and its Applications*. John Wiley and Sons, Inc., New York, 1972.

[Woo93] David H. Wood. Product rules for the displacement of near-Toeplitz matrices. *Linear Algebra and Its Applications*, 188:641–663, 1993.