A Proposal for Toeplitz Matrix Calculations

By Gilbert Strang

In contrast to the usual (and successful) direct methods for Toeplitz systems $Ax = b$, we propose an algorithm based on the conjugate gradient method. The preconditioner is a circulant, so that all matrices have constant diagonals and all matrix-vector multiplications use the Fast Fourier Transform. We also suggest a technique for the eigenvalue problem, where current methods are less satisfactory. If the first indications are supported by further experiment, this new approach may have useful applications--including nearly Toeplitz systems, and parallel computations.

Address for correspondence: Professor Gilbert Strang
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139.

*This research was supported by the National Science Foundation (84-03222) and by the Army Research Office (DAAG29-83-K-0025).

submitted to Studies in Applied Mathematics 1986 [April], Number 2
Vol LXXIV
pg. 171

A Toeplitz matrix is one with constant diagonals. The entries on the main diagonal share a common value a_0, those on the first subdiagonal equal a_1, and in general the i,j entry is a_{i-j} --depending only on the difference $i-j$, which is fixed down each diagonal. Thus A is a "convolution matrix." It is symmetric when $a_k = a_{-k}$.

These matrices--or matrices that are nearly Toeplitz--arise practically everywhere. In time series or signal processing they depend on stationarity; in difference equations the coefficients need to be constant; a wide range of problems are invariant in time and in space. All the requirements of Fourier analysis are satisfied, except one. The only flaw is that the matrix is finite. It starts and ends, and the presence of those boundaries adds a little zest to an otherwise important but unexciting problem.

The continuous counterpart is a differential equation with constant coefficients, or a convolution equation $\int a(t-s)x(s)ds = b(t)$. On the whole line the problem is again straightforward; after Fourier transform it is $\hat{a}\hat{x} = \hat{b}$. But if the limits of integration are 0 and 1, so that this is a "finite section" of an infinite problem, transform methods cannot give such an explicit solution. They are still the key, but only in one case do they go through without difficulty. That is the periodic case, which corresponds in the discrete problem to a circulant matrix.

An n by n matrix C is a circulant if it not only has constant entries c_k down each diagonal, but also satisfies $c_k = c_{k+n}$. The matrix is Toeplitz, and furthermore the i,j entry depends only on $i-j$ modulo n. Each lower diagonal agrees with an upper diagonal, and the distinction between Toeplitz matrices and circulants is seen in

$$A = \begin{bmatrix} a_0 & a_{-1} & \cdot & \cdot & a_{1-n} \\ a_1 & a_0 & a_{-1} & \cdot & \cdot \\ \cdot & a_1 & a_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & a_{-1} \\ a_{n-1} & \cdot & \cdot & a_1 & a_0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} c_0 & c_{n-1} & \cdot & \cdot & c_1 \\ c_1 & c_0 & c_{n-1} & \cdot & \cdot \\ \cdot & c_1 & c_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & c_{n-1} \\ c_{n-1} & \cdot & \cdot & c_1 & c_0 \end{bmatrix}.$$

One case is determined by $2n-1$ entries (the first row and column), the other by n. With symmetry those numbers are nearly halved. A symmetric Toeplitz matrix has n degrees of freedom (its first column) and a symmetric circulant has $[n/2] + 1$.

Numerically both A and C are much simpler than typical full matrices. The usual choice for Toeplitz systems $Ax = b$ is a direct method based on Levinson's algorithm [1]; the operation count grows like n^2 instead of the normal n^3 in Gaussian elimination. For circulants, which are identical with discrete convolutions, the Fast Fourier Transform enters three times--to find the discrete transforms of the sequences

c and x and the inverse of their component-by-component product. This is the convolution rule; we mention in [2] that it is exactly a diagonalization $Cx = F\Lambda F^{-1}x$ by the Fourier matrix containing the eigenvectors that are common to every circulant. It requires $\frac{3}{2}n \log n$ complex multiplications, an important improvement over n^2.

It was natural for the FFT ideas to be extended back to the Toeplitz problem. It is not a pure convolution, but a multiplication Ax can be quickly achieved. (A is completed to a circulant C^* of order $2n-1$, x is completed to x^* by $n-1$ zero components, and from the fast convolution C^*x^* the last $n-1$ components are dropped.) The inversion of a Toeplitz matrix, or the solution of $Ax = b$, is much more delicate. Kailath, Morf, Kung, and others have nevertheless found several methods of complexity $n \log n \log \log n$ [3-6]. The mathematics is beautiful. It uses subtle algebraic properties of transforms to produce an exact solution. The algorithms are somewhat outside the usual range of numerical linear algebra--Bunch has discussed difficulties with stability in SIAM's Journal for Scientific and Statistical Computing--but they throw new light on central problems of constructive classical analysis.

Our present suggestion is very much within the conventional framework. It is a semi-direct method--by which we mean a method that ultimately gives exact answers but should not be

— Multiplication by A
— Linear system Cz = r

carried that far. It is the standard preconditioned conjugate gradient method, in which the preconditioner is a circulant matrix C. Each cycle of the algorithm [7] includes a multiplication by A and a linear system Cz = r. The first cycle therefore uses six fast transforms, but afterwards the transforms of the sequences a and c are fixed and only four transforms enter each later cycle. We report below on some encouraging, but extremely preliminary, numerical experiments using MATLAB. The creation and testing of an efficient code will take much longer, but the basic ingredients are widely available and we hope it may be useful to propose the idea so early.

GET IT!

Our suggestion for the eigenvalue problem is even more premature. Rayleigh quotient iteration is known to have cubic convergence for symmetric problems [8]. Its drawback is the cost of solving $(A-r_kI)x_k = b_k$ with a new matrix at each step--shifted by the Rayleigh quotient value r_k taken from the approximate eigenvector at the previous step. An iterative method seems appropriate for those shifted systems, instead of repeated LU factorizations. Since a good initialization is known we intend to try the method proposed above for linear systems. The advantage of Rayleigh iterations over QR and others is that the matrices $A-r_kI$ remain Toeplitz.†

†Furthermore the sequences a and c are shifted by $(r_k,0,\cdots,0)$, and we stay at four transforms per cycle.

It seems natural to contemplate doing many of these operations in parallel, and to look at corresponding ideas for singular value decompositions. Iterative methods have also an important flexibility, that if A is only close to Toeplitz then the main idea continues to apply. Changes in boundary conditions, or variations instead of invariance in space or time, make the direct Levinson-Durbin-Yule-Walker algorithms much more difficult.

I am extremely grateful for conversations which encouraged me to hope that these proposals are reasonable. George Cybenko recalled the work of Charles Rino, who kindly collected several papers written about 1970 on the use of circulants in ordinary iterations; a splitting into C and A-C appears in [9]. This idea is revived in a preprint of Bitmead and Allen, provided by Tom Kailath, in which C is a multiple of the identity. Alberto Grunbaum emphasized the need for a new look at the eigenvalue problem (and Cybenko-Van Loan have found a Yule-Walker-Newton iteration for the minimal eigenvalue). My conversations with Gene Golub and Beresford Parlett were of immense value, as they always are. Finally, none of the experiments that follow would have been possible without the quick and generous help of Nick Trefethen.

Report on Preliminary Experiments

The first question is which circulant to choose as preconditioner. In practical applications the main diagonal and its neighbors are often strongly dominant. Therefore we copied those diagonals of A into C, and then brought them around to complete the circulant. Notice an important consequence: The entries $a_1 = c_1$ next to the main diagonal appear again in the extreme corners of C. They are comparatively large, so that the residual $A-C$ interferes with the convergence of an iteration based on ordinary splitting. Perhaps for this reason the idea was not much used. In conjugate gradients, however, it is not the distance of the iteration matrix from the identity (reflected in its norm or its condition number) that finally determines its quality. Instead it is the distribution of the eigenvalues. We will see that the large corner entries in C do indicate the largest and smallest eigenvalues of $C^{-1}A$, but the other eigenvalues (at least in our happiest examples) are clustered very near to one.

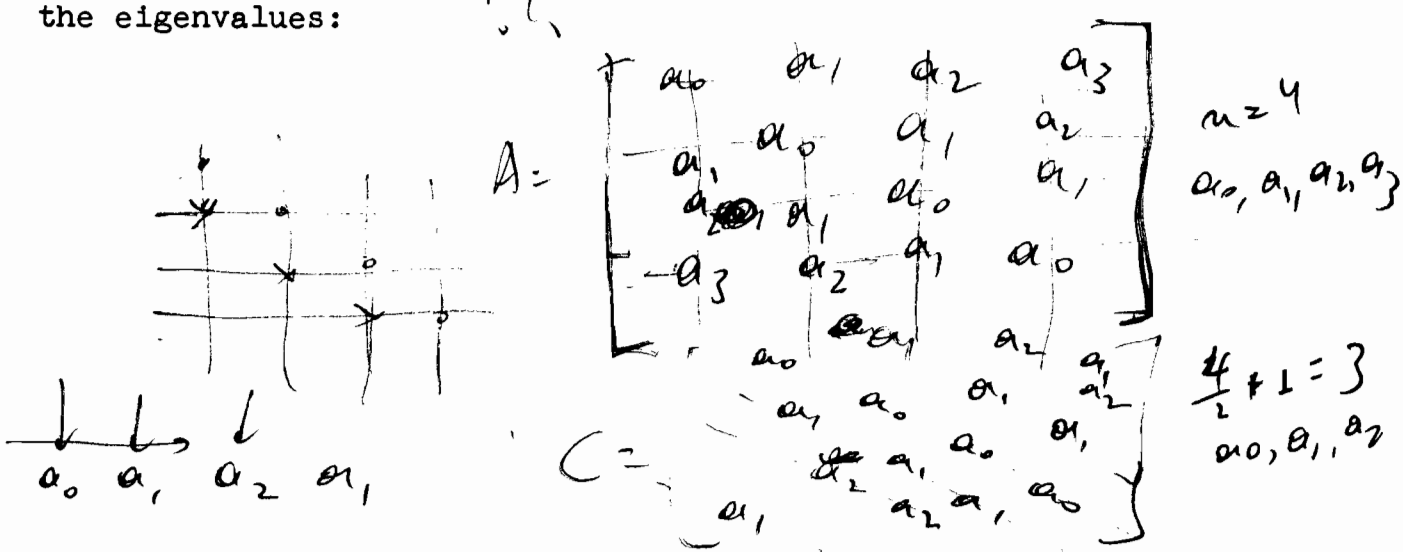
If the Toeplitz matrix A is a finite section of an underlying infinite matrix, another choice of c_k is the sum $\Sigma\, a_{k+jn}$ over diagonals with index k modulo n. This assures that C is positive definite when the infinite matrix is. However it uses information that lies far beyond the finite section A, and in our chief example below (where $a_k = 1/(1+k)$)

$$Cx_{n+1} = (C-A)x_n + b$$

the sum would not converge. In any case the positive definiteness of C is tested when the first cycle takes the discrete transform of $c_0, c_1, \cdots, c_{n-1}$. That transform gives the eigenvalues in $C = F\Lambda F^{-1}$, up to a constant factor n, and triggers a change in C when the eigenvalues are not all positive.

We mention here a common case in differential equations, when A is banded and the sum along a typical row is zero. It is only because the diagonals are cut off in the corners that A is invertible; the circulant C will be singular! The constant vector $x = (1,1,\cdots,1)$ will satisfy $Cx = 0$, as it does for the tridiagonal second-difference matrix--when inserting -1's in the two corners makes the problem periodic. In such a case, with $\Sigma\, a_k = 0$, we use the Fast Sine Transform in place of the FFT in constructing the preconditioner. It is no longer a pure circulant, but it is equally fast (and nonsingular).

We report here on symmetric positive definite experiments. The eigenvalues of $C^{-1}A$ govern the convergence of preconditioned conjugate gradients. Therefore the first test took $a_k = 1/(1+k)$ in a Toeplitz matrix A of order $n = 12$, copied $c_k = a_k$ on the central 13 diagonals of a circulant C, and computed the eigenvalues:



eigenvalues

C	A	$C^{-1}A$
0.376	0.390	0.707
0.413	0.401	0.957
0.413	0.421	0.958
0.443	~~0.451~~ NG	0.973
0.443	0.494	0.974
0.590	0.556	1.000
0.590	0.642	1.000
0.776	0.769	1.026
0.776	0.959	1.028
1.568	1.282	1.041
1.568	1.868	1.047
4.043	3.765	1.880

The last column shows ten eigenvalues that are within 5% of 1, and two that are not. Those two can be roughly attributed to the change in the corner entries from 1/12 in A to 1/2 in C. The extreme eigenvalues of $C^{-1}A$ are the minimum and maximum of the Rayleigh quotient $r(x) = x^TAx/x^TCx$, and we try the two test vectors $x = (1,0,\cdots,0,\pm 1)$:

$$r = \frac{1 + \frac{1}{12}}{1 + \frac{1}{2}} = .72 \quad \text{and} \quad r = \frac{1 - \frac{1}{12}}{1 - \frac{1}{2}} = 1.83 .$$

In other tests, the clustering of eigenvalues near 1 was even more strongly pronounced for $a_k = 1/(1+k)^2$ and $a_k = 2^{-k}$ --in which the central diagonals are more dominant. For an oscillating sequence like $a_k = (\cos k)/(1+k)$, with $n = 21$, the eigenvalues of $C^{-1}A$ were $.64,.74,.93,\cdots,1.08,1.45,1.96$.

Now we report on the application of the conjugate gradient method to $Ax = b$ with a random b. The symmetric matrix A was of order $n = 21$, with diagonals $a_k = 1/(1+k)$. We computed the Euclidean norm of the residual $b-Ax_j$ after each cycle, and compared the algorithm without preconditioning ($C = I$) to the preconditioned form. Again the central band of C was copied directly from A, a choice we will investigate further. The step by step residuals were

unconditioned	preconditioned
.9847377	.2566312
.6670332	.0890435
.3181377	.0056921
.1216105	.0002231
.0442341	.0000058
.0137891	.0000001
.0051327	.000000002
.0014339	.00000000002
.0003798	
.0000708	
.0000194	
.0000032	

For the reader's convenience we reproduce from [7] and [2] the steps of the conjugate gradient method preconditioned by a matrix C:

how do you solve this

Solve $Cz_{j-1} = r_{j-1}$

$\beta_j = z_{j-1}^T r_{j-1} / z_{j-2}^T r_{j-2}$ (except $\beta_1 = 0$)

$d_j = z_{j-1} + \beta_j d_{j-1}$ (except $d_1 = z_0$)

$\alpha_j = z_{j-1}^T r_{j-1} / d_j^T A d_j$

$x_j = x_{j-1} + \alpha_j d_j$

$r_j = r_{j-1} - \alpha_j A d_j$

$x_0 = 0$

$r_0 = b - Ax_0 = b$

$|b - Ax_j|$ small enough

This has the effect of working with the symmetric matrix $C^{-1/2}AC^{-1/2}$, which is near the identity, without finding the square root of C. It converges in n steps of exact arithmetic. We started from $x_0 = 0$ and the corresponding residual $r_0 = b - Ax_0 = b$, but a quicker start is often possible. The code separated the first cycle from the others, and equivalent forms of the main conjugate gradient cycle are frequently used.

Finally we mention that all but the extreme eigenvalues of $C^{-1}A$ were again in the interval between .95 and 1.05. Asymptotic estimates with increasing n will require more analysis, and a good code (for the eigenvalue problem too) demands much more work. The possibility exists of a demonstrably fast algorithm--but with computational complexity dependent on the given problem, as expected for iterative in contrast to direct methods.

References

1. N. Levinson, The Wiener RMS (Root-Mean Square) error criterion in filter design and prediction, J. Math. and Phys. 25:261-278 (1947).

2. G. Strang, Introduction to Applied Mathematics, Wellesley-Cambridge Press, Wellesley, Massachusetts, 1986.

3. T. Kailath, S. Y. Kung, and M. Morf, Displacement ranks of matrices and linear equations, J. Math. Anal. Appls. 68:395-407 (1979).

4. T. Kailath, A. Vieira, and M. Morf, Inverses of Toeplitz operators, innovations, and orthogonal polynomials, SIAM Review 20:106-119 (1978).

5. R. P. Brent, F. G. Gustavson, and D. Y. Yun, Fast solution of Toeplitz systems and computation of Padé approximants, Lin. Alg. Appls. 34:103-116 (1980).

6. A. K. Jain, Fast inversion of banded Toeplitz matrices by circular decompositions, IEEE J. ASSP 26:121-126 (1978).

7. G. H. Golub and C. F. Van Loan, Matrix Computations, Johns Hopkins Press, Baltimore, 1983.

8. B. N. Parlett, The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, 1980.

9. C. Rino, The inversion of covariance matrices by finite Fourier transformations, IEEE Trans. Inform. Theory 16:230-232 (1970).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(Received January 27, 1986)