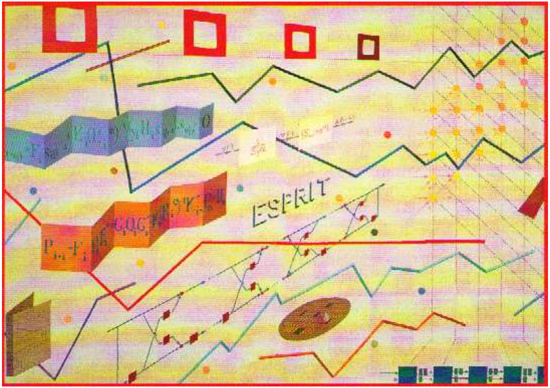
From Radiative Transfer Theory To Fast Algorithms For Cell Phones

Thomas Kailath

Stanford University

Abstract

We describe how analogies between certain problems in statistical prediction and in radiative transfer led, *inter alia*, to the development of new fast algorithms (and efficient integrated circuit implementations thereof) for a host of problems in the fields of communications, control, signal processing, linear algebra and operator theory.



Wiener-Hopf Equation

The story begins in 1931, when the astronomer Eberhard Hopf paid a visit to the summer home of the already famous mathematician Norbert Wiener of MIT.

As was his wont, Wiener enquired of his guest what the most outstanding problem was in his field. The response was that no solution method was known for an integral equation put forward by the astronomers Milne and Schwarzschild to characterize certain problems of radiative transfer in an atmosphere of infinite height:

$$\int_0^\infty w(\tau)K(t-\tau)\,d\tau = g(t)\,, \quad t \geq 0\,,$$

where $K(\cdot)$ and $g(\cdot)$ are known, and $w(\cdot)$ is to be found.

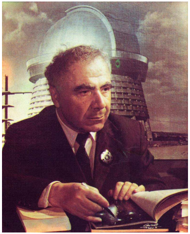
Wiener-Hopf Equation

At breakfast the next morning Wiener presented a solution! It was not quite correct, but the mistakes were easily fixed, and the paper on it (in German) was published by Wiener and Hopf in 1931.

So unexpectedly brilliant was the solution that the equation itself came to be known as the Wiener-Hopf equation and the solution method as the Wiener-Hopf technique.

Ambartsumian ...

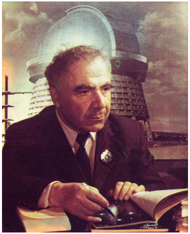
The Wiener-Hopf technique requires something called spectral factorization (an infinite-dimensional version of writing a matrix as a product of upper- and lower-triangular matrices), which was too difficult to carry out for several radiative transfer problems of interest.



Then, in 1943, the renowned Soviet/Armenian astronomer V. A. Ambartsumian showed — by using certain invariance features of the problem — that a solution could be obtained using a quadratic integral equation of Riccati type, solving it for the scattering function $S(x, y, \tau)$ of the atmosphere at surface location (x, y) and height τ.

Ambartsumian ...

The Wiener-Hopf technique requires something called spectral factorization (an infinite-dimensional version of writing a matrix as a product of upper- and lower-triangular matrices), which was too difficult to carry out for several radiative transfer problems of interest.



Then, in 1943, the renowned Soviet/Armenian astronomer V. A. Ambartsumian showed — by using certain invariance features of the problem — that a solution could be obtained using a quadratic integral equation of Riccati type, solving it for the scattering function $S(x, y, \tau)$ of the atmosphere at surface location (x, y) and height τ.

The Riccati equation that Ambartsumian derived has to be solved numerically, but the computational burden is forbidding, because the scattering function $S(x, y, \tau)$ has to be determined over a 3-dimensional grid of points.

... and Chandrasekhar

In 1947, S. Chandrasekhar in Chicago further studied the problem, and treated the case of a finite atmosphere of height T, for which the integral equation is now

$$\int_0^t w(t,\tau)K(t-\tau)\,d\tau = g(t)\ , \quad T \geq t \geq 0\ .$$

Again exploiting invariance, he showed that the scattering function S could be determined by solving a pair of so-called X and Y nonlinear differential equations, where X and Y now were functions of only 2 variables, hence to be determined over a 2-dimensional grid of points.

... and Chandrasekhar

In 1947, S. Chandrasekhar in Chicago further studied the problem, and treated the case of a finite atmosphere of height T, for which the integral equation is now

$$\int_0^t w(t,\tau)K(t-\tau)\,d\tau = g(t)\,, \quad T \geq t \geq 0\,.$$

Again exploiting invariance, he showed that the scattering function S could be determined by solving a pair of so-called X and Y nonlinear differential equations, where X and Y now were functions of only 2 variables, hence to be determined over a 2-dimensional grid of points.

The resulting order-of-magnitude reduction in the computational burden — from $O(N^3)$ to $O(N^2)$ — allowed important radiative transfer problems to be solved on even the clumsy hand-cranked desk calculators of the day. (An affectionate spoof of Chandrasekhar's work ends: "I wish to record my indebtedness to Miss Canna Helpit, who carried out the laborious numerical work ..."!)

Radiative Transfer — "fascinates and challenges"

The following glowing commentary is from the Preface written by the translator, I. Gaposchkin, of *A Treatise on Radiative Transfer*, V.V. Sobolev, Moscow, 1956 (English translation, 1963):

"There is no gainsaying that there always exists a scientific discipline which epitomizes the power of human intellect or a region of exact thought where man's genius shines at its best.

"For a long time such was the field of Celestial Mechanics, woven by great theoreticians whose mathematical analyses still shine with undiminishing brilliance, comparable only to that of the stars themselves.

Radiative Transfer — "fascinates and challenges"

The following glowing commentary is from the Preface written by the translator, I. Gaposchkin, of *A Treatise on Radiative Transfer*, V.V. Sobolev, Moscow, 1956 (English translation, 1963):

"There is no gainsaying that there always exists a scientific discipline which epitomizes the power of human intellect or a region of exact thought where man's genius shines at its best.

"For a long time such was the field of Celestial Mechanics, woven by great theoreticians whose mathematical analyses still shine with undiminishing brilliance, comparable only to that of the stars themselves.

"At the present time it is Radiative Transfer, linked with the Internal Constitution of Stars, which fascinates and challenges astronomers all over the world. Among the trailblazers and the leaders in this field, none are more profound, none more illustrious than two contemporary astronomers: an American, Chandrasekhar, and a Russian, Sobolev."

Wiener Filtering

Now we move to another field, the prediction and filtering of stationary random processes, i.e., those whose statistical parameters do not change with time. Though one can trace the origins back at least to the work of Legendre and Gauss ca. 1800 on least squares, the modern story started again with Wiener, now in 1941.

As part of his contributions to the scientific efforts in WW II, Wiener studied the problem of anti-aircraft fire control. This required estimating the future position of an attacking aircraft from observations of its trajectory up to the present time, which he modeled as a sample function of a stationary random process.

Wiener Filtering

Now we move to another field, the prediction and filtering of stationary random processes, i.e., those whose statistical parameters do not change with time. Though one can trace the origins back at least to the work of Legendre and Gauss ca. 1800 on least squares, the modern story started again with Wiener, now in 1941.

As part of his contributions to the scientific efforts in WW II, Wiener studied the problem of anti-aircraft fire control. This required estimating the future position of an attacking aircraft from observations of its trajectory up to the present time, which he modeled as a sample function of a stationary random process.

Wiener was excited to find that the Wiener-Hopf equation was again the key to the optimal (least squares) solution. And here, several problems of engineering interest had a structure — rational spectral densities — that made the Wiener solution and implementation feasible.

The specific filtering problem

While Wiener's approach did not turn out to be useful for the anti-aircraft problem, it ended up having a lasting impact on the problem of "filtering" signals out of noisy measurements:

Estimate the values over time t of a signal process $z(t)$, given observations $y(t)$ of the signal $z(t)$ that are corrupted by a noise process $v(t)$:

$$y(t) = z(t) + v(t) \, .$$

Wiener assumed knowledge of the mean values and correlation functions of the signal and noise processes, and that the observations began in the remote past.

The specific filtering problem

While Wiener's approach did not turn out to be useful for the anti-aircraft problem, it ended up having a lasting impact on the problem of "filtering" signals out of noisy measurements:

Estimate the values over time t of a signal process $z(t)$, given observations $y(t)$ of the signal $z(t)$ that are corrupted by a noise process $v(t)$:

$$y(t) = z(t) + v(t) \ .$$

Wiener assumed knowledge of the mean values and correlation functions of the signal and noise processes, and that the observations began in the remote past.

The optimal causal estimate $\widehat{z}(t)$ is produced by the convolution $w(t) * y(t)$, where the impulse response $w(t)$ of the Wiener filter is the solution of the Wiener-Hopf equation. Under the assumption of "white" noise $v(\cdot)$ that is uncorrelated with the signal $z(t)$, this equation takes the form

$$w(t) + \int_0^\infty w(\tau) R_{zz}(t-\tau)\, d\tau = R_{zy}(t) \ , \quad t \geq 0 \ .$$

Kalman Filtering

Wiener's technique could not be satisfactorily extended to vector-valued and/or non-stationary processes, despite almost a decade of effort.

In 1960, Rudolf Kalman found the key to the extension by replacing knowledge of the correlation functions by knowledge of a dynamical state-space model for the signal $z(t)$:

$$\frac{dx(t)}{dt} = F(t)x(t) + u(t) \, , \qquad \text{e.g., } x(t) = \begin{bmatrix} \text{position} \\ \text{velocity} \end{bmatrix}$$

$$y(t) = \underbrace{H(t)x(t)}_{z(t)} + v(t) \, , \quad t > 0 \, . \qquad \text{e.g., } z(t) = \text{position}$$

The Kalman filter determines the optimal state estimate $\widehat{x}(t)$ — which often has an independent interest — using the past values of the observations $y(\cdot)$, and can then also obtain the estimate $\widehat{z}(t)$ as $H\widehat{x}(t)$.

The Riccati Equation again

It turns out that the critical step in specifying the Kalman filter is the solution of a nonlinear Riccati-type ordinary differential equation:

$$\frac{dP(t)}{dt} = F(t)P(t) + P(t)F'(t) + Q(t) - P(t)H'(t)H(t)P(t)\ , \quad t > 0.$$

Here $F(t)$, $H(t)$ and $Q(t)$ are known coefficient matrices, possibly varying with time, and $'$ denotes matrix transposition; $P(0)$ is known.

The Riccati Equation again

It turns out that the critical step in specifying the Kalman filter is the solution of a nonlinear Riccati-type ordinary differential equation:

$$\frac{dP(t)}{dt} = F(t)P(t) + P(t)F'(t) + Q(t) - P(t)H'(t)H(t)P(t)\ , \quad t > 0.$$

Here $F(t)$, $H(t)$ and $Q(t)$ are known coefficient matrices, possibly varying with time, and $'$ denotes matrix transposition; $P(0)$ is known.

This initial-value differential equation can be solved by standard numerical integration methods. The number of flops to implement an incremental time-step of the numerical integration is $O(n^3)$, where n is the number of components in the state vector x.

We remark that this computational burden is independent of whether $\{F(t), H(t), Q(t)\}$ in the Riccati equation are time-dependent or not.

Ambartsumian and Chandrasekhar *Redux*

Because I was familiar with the Wiener filter from my first MIT course as a graduate student in 1957, I knew well that the Kalman filter solution, though it was couched differently, ultimately also had to solve some type of Wiener-Hopf equation (specifically of the type that arises in the finite-atmosphere problem).

Through a lucky encounter, I became aware in 1972 of the 1940's work on Wiener-Hopf equations by Ambartsumian and Chandrasekhar.

I was happily surprised to see that Ambartsumian had already introduced a Riccati-type equation in 1943.

Ambartsumian and Chandrasekhar *Redux*

Because I was familiar with the Wiener filter from my first MIT course as a graduate student in 1957, I knew well that the Kalman filter solution, though it was couched differently, ultimately also had to solve some type of Wiener-Hopf equation (specifically of the type that arises in the finite-atmosphere problem).

Through a lucky encounter, I became aware in 1972 of the 1940's work on Wiener-Hopf equations by Ambartsumian and Chandrasekhar.

I was happily surprised to see that Ambartsumian had already introduced a Riccati-type equation in 1943.

And I saw also, as mentioned earlier, that Chandrasekhar had reduced the computational burden of solving Ambartsumian's Riccati equation by better exploiting some properties of the underlying radiative transfer.

So it was natural to ask whether analogs of such properties could be used to reduce the computational burden of the Riccati differential equation in the Kalman filter.

Fast Estimation Algorithms

It turns out that the simplest analog of the special properties exploited by Chandrasekhar was to assume time-invariance of the coefficient matrices $\{F(t), H(t), Q(t)\}$ and a special choice of $P(0)$.

Then, unlike with the Riccati equation, the computational effort could be reduced from $O(n^3)$ to $O(n^2)$ flops per iteration, by using an analog of Chandrasekhar's X and Y equations. This can be a very significant reduction when n is large (when $n = 1000$, for example, n^2 is evidently negligible compared to n^3).

Fast Estimation Algorithms

It turns out that the simplest analog of the special properties exploited by Chandrasekhar was to assume time-invariance of the coefficient matrices $\{F(t), H(t), Q(t)\}$ and a special choice of $P(0)$.

Then, unlike with the Riccati equation, the computational effort could be reduced from $O(n^3)$ to $O(n^2)$ flops per iteration, by using an analog of Chandrasekhar's X and Y equations. This can be a very significant reduction when n is large (when $n = 1000$, for example, n^2 is evidently negligible compared to n^3).

I actually had to use a different argument than Chandrasekhar's, based on an identity I derived that turned out to have been found in a simpler context by George Stokes, studying the passage of light through a pile of plates.

Moreover, this different derivation allowed me to extend the results to the case where the coefficients $\{F(t), H(t), Q(t)\}$ changed with time in a certain way that was actually encountered in many problems. (One such problem is that of Lambert scattering in radiative transfer.)

Derivation of Chandrasekhar equations

When the coefficient matrices $\{F(t), H(t), Q(t)\}$ are time-invariant, differentiation of the Riccati differential equation yields

$$\begin{aligned}\frac{d^2P(t)}{dt^2} &= F\dot{P}(t) + \dot{P}(t)F' - \dot{P}(t)H'HP(t) - P(t)H'H\dot{P}(t) \\ &= (F - K(t)H)\dot{P}(t) + \dot{P}(t)(F - K(t)H)' ,\end{aligned}$$

where $\dot{P}(t) = dP(t)/dt$ and we have defined $K(t) = P(t)H'$.

The solution of this homogeneous differential equation can be written as

$$\dot{P}(t) = \Psi(t, t_0)\dot{P}(0)\Psi'(t, t_0) ,$$

where

$$\frac{d\Psi(t, t_0)}{dt} = (F - K(t)H)\,\Psi(t, t_0) , \quad \Psi(t_0, t_0) = I .$$

Chandrasekhar equations

Now suppose $\dot{P}(0)$ has (low) rank r, so

$$\dot{P}(0) = L_0 J L_0' , \quad J = I_\alpha \oplus -I_\beta , \quad \alpha + \beta = r .$$

Since $\dot{P}(t) = \Psi(t, t_0)\dot{P}(0)\Psi'(t, t_0)$, we get

$$\dot{P}(t) = L(t)JL'(t) , \quad \text{where} \quad L(t) = \Psi(t, t_0)L_0 .$$

Hence

$$\frac{dL(t)}{dt} = (F - K(t)H)L(t) , \quad L(t_0) = L_0 .$$

Also

$$\frac{dK(t)}{dt} = \dot{P}(t)H' = L(t)JL'(t)H' , \quad K(t_0) = P(0)H' .$$

So we have a coupled set of $n(r + p)$ nonlinear differential equations for propagation of $L(t)$ and $K(t)$, where p is the number of measured outputs in the Kalman filter.

Further comments on Chandrasekhar equations

- The Riccati variable can be found by quadrature:

$$P(t) = P(0) + \int_0^t L(\tau) J L'(\tau)\, d\tau \ .$$

- Whenever $n(r+p) \ll n(n+1)/2$ the Chandrasekhar equations provide a fast algorithm for solving the matrix Riccati differential equation, $O(n^2)$ instead of $O(n^3)$. Typically $p \ll n$, so what is needed is $r \ll n$.

Further comments on Chandrasekhar equations

- The Riccati variable can be found by quadrature:

$$P(t) = P(0) + \int_0^t L(\tau) J L'(\tau)\, d\tau \ .$$

- Whenever $n(r+p) \ll n(n+1)/2$ the Chandrasekhar equations provide a fast algorithm for solving the matrix Riccati differential equation, $O(n^2)$ instead of $O(n^3)$. Typically $p \ll n$, so what is needed is $r \ll n$.
- A special case is when F is stable and $P(0)$ obeys the stationarity condition $FP(0) + P(0)F' + Q = 0$. Then

$$\dot{P}(0) = FP(0) + P(0)F' + Q - P(0)H'HP(0) = -P(0)H'HP(0) \ ,$$

  so $r = p \ll n$.
- Another special case: $\mathrm{rank}(Q) \ll n$ and $P(0) = 0$, so $r = \mathrm{rank}(Q) \ll n$.

Fast Algorithms for Linear Equations

Seeking extensions of these ideas to different areas, we were led all the way to a much studied problem, namely that of solving the system of linear equations

$$Ax = b$$

where A is a known $n \times n$ matrix, b is a known n-vector, and x is an n-vector of unknowns.

There are many libraries of computational algorithms for this problem. However, for general matrices A the computational effort is $O(n^3)$, which is prohibitive for the large n found in many interesting problems. How do we get around this?

Fast Algorithms for Linear Equations

Seeking extensions of these ideas to different areas, we were led all the way to a much studied problem, namely that of solving the system of linear equations

$$Ax = b$$

where A is a known $n \times n$ matrix, b is a known n-vector, and x is an n-vector of unknowns.

There are many libraries of computational algorithms for this problem. However, for general matrices A the computational effort is $O(n^3)$, which is prohibitive for the large n found in many interesting problems. How do we get around this?

The answer is that we often make simplifying modeling assumptions on the underlying physical situations, idealizations such as homogeneity, isotropy, time-invariance, infinite duration, finite bandwidth, etc.. These lead to special structures for the matrix A, which can allow the computational burden to be reduced to $O(n^2)$ or $O(nlogn)$ or even $O(n)$.

Structured Matrices: Explicit

$$Toeplitz = [x_{i-j}] = \begin{bmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & a \end{bmatrix} \quad ; \quad Hankel = [x_{i+j}] = \begin{bmatrix} d & c & b & a \\ c & b & a & e \\ b & a & e & f \\ a & e & f & g \end{bmatrix} \quad ;$$

$$Vandermonde = [x_j^{i-1}] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & b & c & d \\ a^2 & b^2 & c^2 & d^2 \\ a^3 & b^3 & c^3 & d^3 \end{bmatrix} \quad ; \quad Cauchy = \left[\frac{1}{x_i - y_j}\right]$$

- Banded, Pick, Schur-Cohn, Routh-Hurwitz, Bezoutian, ...
- Controllability, observability, impulse response matrices for state-space descriptions, ...

Structured Matrices: Implicit

If A, B, C, D are all (explicitly) structured matrices, we would like the same to be true for various composites of these matrices that are often encountered in applications, e.g., the following that arise in least squares problems:

$$AB \; ; \qquad (A'A)^{-1}A \; ; \qquad A - B(C)^{-1}D$$

The composites unfortunately do *not* generally inherit the explicit structure of their constituent matrices.

Nevertheless, it is not unreasonable to expect that the composites will still share some sort of common structure with their constituent matrices.

For many classes of matrices, this common structure turns out to be what we have named Displacement Structure, which can be exploited to design fast algorithms.

Displacement Rank

Define left and right displacement operators L and R. Then

$$\text{Displacement rank}(A) = \text{Rank}(A - LAR) \; .$$

Example: For a Toeplitz matrix A, choose

$$L = R' = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ (the canonical \textit{shift} operator) ,}$$

then

$$L \begin{bmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & da \end{bmatrix} R = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & a & b & c \\ 0 & e & a & b \\ 0 & f & e & a \end{bmatrix}$$

and

$$\text{Rank}(A - LAR) = \text{Rank} \begin{bmatrix} a & b & c & d \\ e & 0 & 0 & 0 \\ f & 0 & 0 & 0 \\ g & 0 & 0 & 0 \end{bmatrix} \leq 2 \; .$$

Displacement Rank of Composite Matrices

Example: If A, B and

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

are all Toeplitz (constant along diagonals), and assuming the matrix products below are conformable, we have

- Displacement rank$(A) \leq 2$
- Displacement rank$(A^{-1}) \leq 2$
- Displacement rank$(AB) \leq 4$
- Displacement rank$(C_{11} - C_{12}C_{22}^{-1}C_{21}) \leq 2$

Displacement Rank of Composite Matrices

Example: If A, B and

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

are all Toeplitz (constant along diagonals), and assuming the matrix products below are conformable, we have

- Displacement rank$(A) \leq 2$
- Displacement rank$(A^{-1}) \leq 2$
- Displacement rank$(AB) \leq 4$
- Displacement rank$(C_{11} - C_{12}C_{22}^{-1}C_{21}) \leq 2$

Significance: If a matrix has displacement rank α, then standard $O(n^3)$ algorithms for a variety of associated matrix problems — solving linear equations, inversion, triangular and orthogonal factorization, Schur complementation, ... — can be replaced by $O(n^2\alpha)$ algorithms.

Displacement Structure Theory

There are more aspects to displacement structure than displacement rank (e.g., displacement inertia and generators). Over the last 35 years, the initial ideas have been developed into an extensive theory of Displacement Structure.

Along the way several quite unexpected mathematical results were encountered and used, and some new ones developed. For example, a key reference is a 1917 paper by the famous mathematician, Issai Schur, on what would seem to be a purely mathematical topic: characterizing "power series that are bounded in the unit disc". Displacement Structure theory heavily builds on our generalizations of an algorithm found in that paper, which we have called Generalized Schur Algorithms.

While it can happen that fast algorithms lose some measure of numerical stability due to the accumulation of round-off errors, numerically stable variants of these algorithms can often be found.

Applications

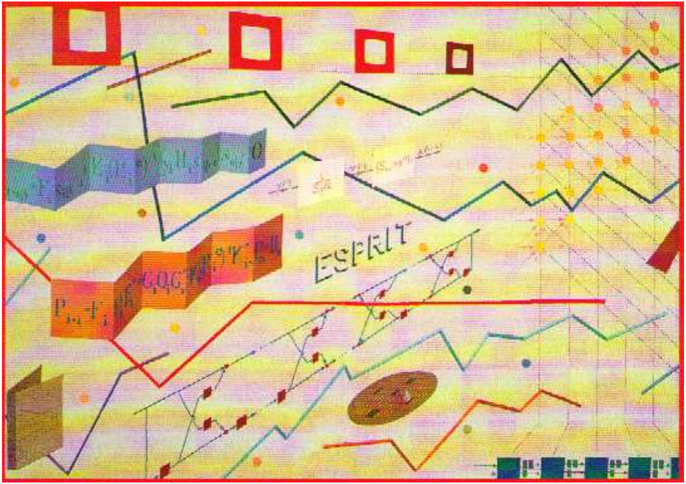
Displacement Structure theory has led to fast algorithms for solving a wide variety of problems in several different fields, including prediction and filtering, control, communications, signal processing, algebraic coding theory, queueing theory, linear algebra, matrix theory, integral equations, operator theory, inverse scattering, interpolation theory and others.

One specific communications application is to cell phones, where fast algorithms are very important in improving the speed of response and in extending battery life.

Some References

1. T. Kailath, S.Y. Kung, M. Morf, *Displacement ranks of a matrix*, Bull. Amer. Math. Soc., vol. 1, pp. 769-773, 1979.
2. T. Kailath, L. Ljung and M. Morf, *Generalised Krein-Levinson equations for efficient calculation of Fredholm resolvents of nondisplacement kernels*, in Topics in Functional Analysis, I. Gohberg and M. Kac, eds., pp. 168-184, Acad. Press, 1978.
3. T. Kailath and A H. Sayed, *Displacement Structure: theory and applications*, SIAM Review, pp. 297-386, 1995.
4. T. Kailath and A. H. Sayed, eds., *Fast Reliable Algorithm for Matrices with Structure*, SIAM Press, 1999.
5. V. Olshevsky, ed., *Structured Matrices in Mathematics, Computer Science, and Engineering*, American Mathematical Society, 2001.
6. D. A. Bini, G. Latouche and B. Meini, *Numerical Methods for Structured Markov Chains*, Oxford Press, 2005.

Thank you!



Original textile piece by Anne McKenzie Nickolson, 1992

Cell Phones

"It might interest you to know that part of my work here at Nokia involves the implementation of a pseudo-inverse of a Toeplitz structured matrix in hardware. We are using a version of displacement structure based algorithms with proprietary improvements for fast, parallel realization of the same. It has been a very interesting learning experience to consolidate the conflicting demands of precision, stability, complexity and real-time constraints into a working receiver structure."

Anand.Kannan@nokia.com