Title

Thomas Kailath

Information Systems Laboratory
Stanford University

- The solution of problems in many fields often reduces to the solution of large sets of linear equations

$$Ax = b \in \mathcal{C}^{m \times 1}$$

- There are "standard" methods of solutions, esp. Gaussian elimination. The cost is $O(n^3)$ flops $n = m$ for simplicity.

- Is this the end of the problem?

NO:

1. $O(n^3)$ can be impossibly large in many communications, signal processing, IC design, . . . , problems.

2. The entries of $\{A, b\}$ are "flexible", and often need to be set on the basis of preliminary solutions, more experimentation, etc.

These issues can often be significantly ameliorated imposing "mathematical" structure of the problem.

From RICCATI to WIENER-HOPF to AMBARTZUMIAN-CHANDRASEKHAR to DISPLACEMENT STRUCTURE

The minimization of quadratic forms in variables that obey linear quadratic equations has been known, since at least Legendre (1786), to be reducible to the solution of a certain (nonlinear) Riccati differential equations.

In the 1960's it gained a lot of attention in systems theory from its appereance in the Kalman filter for linear least-squares prediction and filtering. Here it takes the form

$$\dot{\mathsf{P}}(t) = F(t)P(t) + P(t)F^*(t) + G(t)G^*(t) - K(t)K^*(t), P(t_0) = \Pi_0$$

where

$$K(t) = P(t)H^*(t) \in \mathcal{C}^{n\times p},$$

and where $P = p^* \in \mathcal{C}^{n\times n}$, $G \in \mathcal{C}^{n\times m}$, $H \in \mathcal{C}^{p\times n}$.

These $n(n+1)/2$ nonlinear coupled initial value differential equations can be solved by digitalization. In the sampled form,

$$P(t+\delta) = P(t) + \delta[F(t)P(t) + \ldots], \quad P(t_0) = \Pi_0,$$

which requires $O(n^3)$ flops per iteration.

This is true whether the coefficients $\{F(\cdot), G(\cdot), H(\cdot)\}$ are time-variant <u>or</u> not.

Can we exploit the time-invariance in some way? Yes.

When the coefficients are time-invariant:

$$\ddot{\mathsf{P}}(t) = F\dot{\mathsf{P}}(t) + \dot{\mathsf{P}}(t)F - \dot{\mathsf{P}}(t)HK(t) - K(t)H^*(t)\dot{\mathsf{P}}(t) = (F - K(t)H)\dot{\mathsf{P}}(t) + \dot{\mathsf{P}}(t)(F - K(t)$$

The solution of this homogeneous d.e. can be written as

$$\dot{\mathsf{P}}(t) = \Psi(t,t_0)\dot{\mathsf{P}}(0)\Psi^*(t,t_0),$$

where

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

Now suppose $\dot{\mathsf{P}}(0)$ has (low) rank, say r, and write

$$\dot{\mathsf{P}}(0) = L_0 J L_0^*, J = I_p \oplus -I_q, p + q = r.$$

Let

$$\dot{\mathsf{P}}(t) = \Psi(t, t_0) L_0 J L_0^* \Psi(t, t_0) = L(t) J L^*(t), \text{ say,}$$

where

$$\dot{\mathsf{L}}(t) = (F - K(t)H)L(t), L(t_0) = L_0.$$

Also

$$\dot{\mathsf{K}}(t) = \dot{\mathsf{P}}(t)H^* = L(t)JL^*(t)H^*, K(t_0) = \Pi_0 H^*.$$

So we have a coupled set of $n(m+p)$ nonlinear d.e.s for propagation of $L(\cdot)$. And the Riccati variable can be found, when ?, by quadrature

$$P(t) = \Pi_0 + \int_0^t L(\tau) J L^*(\tau) d\tau.$$

Whenever

$$n(m+p) \ll n(n+1)/2,$$

we have a fast algorithm for solving the matrix Riccati d.e.

We encountered the Riccati equation in the study of the Kalman filter for the linear least-squares prediction of stochastic processes. Now a linear Wiener-Hopf-type of integral equation, underlines such prediction problems (Wiener, 1942).

Actually, Wiener first encountered such equation in 1931, when Hopf pointed out to him that no solution was known for the (Wiener-Hopf) equation

$$h(t) + \int_0^\alpha h(\tau)K(t-\tau)d\tau = K(t), t \geq 0,$$

introduced into radiative transfer theory by Milne and Schwarzschild around 1920. So unexpectedly brilliant was Wiener's solution (based on spectral factorization) that the equation was henceforth called by the name Wiener-Hopf!

However Wiener's solution was computationally difficult. In 1943, V. A.

Ambartzumian showed that when $K(\cdot)$ had the form

$$K(t-s) = \int_0^1 e^{-\alpha|t-s|} w(\alpha) d\alpha, \tag{1}$$

the problem could be reduced to the solution of a certain Riccati P.D.E. in 3 variables. In 1947, S. Chandrasekhar showed that one could instead use a pair of coupled nonlinear equations for two functions X and Y of two variables each (the famous X and Y equations).

The results of Ambartzumian-Chandrasekhar were extended to Wiener-Hopf-type equations

$$h(t,s) + \int_0^t h(t,\tau) K(\tau - s) d\tau = K(t,s), 0 \le s \le t$$

where $K(\cdot)$ had "displacement" form

$$K(t-s) = \int_0^1 e^{-\alpha|t-s|} w(\alpha) d\alpha,$$

by Casti, Kalaba, Marthy (1972).

Inspired by a seminar at Stanford by Casti, I derived the $\{K(\cdot), L(\cdot)\}$

equations mentioned earlier, which are associated with Wiener-Hopf-type equations with "semi-separable" kernels,

$$K(t,s) = \begin{matrix} \sum_1^n \alpha_i(t)\beta_i(s), t \geq s \\ \sum_1^n \alpha_i(s)\beta_i(t), t \leq s. \end{matrix}$$

The time-invariance of the $\{F(\cdot), G(\cdot), H(\cdot)\}$ matrices in out original Riccati equation turned out to correspond to the property

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) K(t,s) &= \sum_1^r \Phi_i(t)\Phi_i(s), r \leq n \\ &= \text{a degenerate low rank kernel.} \end{aligned}$$

Note that $r = 0$ when $K(t,s) = f(|t-s|)$. When $K(\cdot)$ is a Wiener covariance kernel,

$$K(t,s) = \min(t,s), r = 1.$$

We pursued these studies for general Wiener-Hopf and Fredholm equations and then for linear matrix equation

$$Ax = b,$$

initially for A Toeplitz and then for matrices A such that

$$A - ZAZ^* = \text{ a low rank matrix,}$$

where Z is the lower shift matrix with ones on the first subdiagonal and zeros elsewhere.

A first interesting result was that

$$A^\# - ZA^\# Z^\# = \text{ low rank,}$$

where $A^\#$ is the ("natural") inverse of A,

$$A^\# = \tilde{I} A^{-1} \tilde{I}, \tilde{I} = \text{ the anti-identity.}$$