1 Sylvester equations

The Sylvester equation (or the special case of the Lyapunov equation) is a matrix equation of the form

\[ A X + X B = C \]

where \( A \in \mathbb{R}^{m \times m}, B \in \mathbb{R}^{n \times n}, \), are known, and \( X \in \mathbb{R}^{m \times n} \) is to be determined. The Sylvester equation has important applications in control theory, and also plays a prominent role in the theory of several classes of structured matrices.

On the surface of it, this is a simple system: the expressions \( AX \) and \( XB \) are just linear in the elements of \( X \), after all. Indeed, we can rewrite the system as

\[ (I \otimes A + B^T \otimes I) \text{vec}(X) = \text{vec}(C), \]

where

\[ \text{vec}([x_1 \ldots x_n]) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \]

is a vector of length \( mn \) composed by listing the elements of \( X \) in column-major order, and the Kronecker product is defined by

\[ F \otimes G = \begin{bmatrix} f_{11}G & f_{12}G & \cdots \\ f_{21}G & f_{22}G & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \]

Alas, solving this matrix equation by Gaussian elimination would cost \( O((mn)^3) \). Can we do better?

The Bartels-Stewart algorithm is a clever approach to the problem that takes only \( O(\max(m, n)^3) \) time. The key is to compute the Schur factorizations

\[ A = U_A T_A U_A^* \quad B = U_B T_B U_B^* \]

from which we obtain

\[ T_A \tilde{X} + \tilde{X} T_B = \tilde{C} \]
where $\tilde{X} = U^*_AXUB$ and $\tilde{C} = U^*_ACUB$. Column $j$ of this system of equations can be written as

$$(T_A + t_{B,kk}I)\tilde{x}_k = \tilde{c}_k - \sum_{j=1}^{k-1} \tilde{x}_jt_{B,jk};$$

therefore, we can solve each column of $\tilde{x}$ in turn by a back-substitution procedure that involves a triangular linear solve. We only run into trouble if one of these systems is singular (or nearly so), corresponding to the case where $A$ and $-B$ (nearly) have an eigenvalue in common.

### 1.1 Riccati equations

The Sylvester equation is a linear matrix equation whose solution is accelerated via an intermediate eigendecomposition. The algebraic Riccati equation is a quadratic matrix equation that also can be expressed via an eigenvalue problem. The Riccati equation occurs in optimal control problems, as well as some other places; for the continuous-time optimal control problem, we would usually write

$$A^TX +XA -XBR^{-1}B^TX +Q = 0$$

where $R$ and $Q$ are spd matrices representing cost functions, $A$ and $B$ are general square matrices, and we seek a symmetric solution matrix $X$.

The key to thinking of the Riccati equation via eigenvalues is to write the left hand side of the equation as a pure quadratic:

$$\begin{bmatrix} I & Q \\ X & A^T \end{bmatrix} \begin{bmatrix} A & -BR^{-1}B^T \\ I & X \end{bmatrix} = 0.$$

We can also characterize this by the relation

$$\begin{bmatrix} Q \\ A^T \\ -BR^{-1}B^T \end{bmatrix} \begin{bmatrix} I \\ X \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ I \\ -I \end{bmatrix} \begin{bmatrix} I \\ X \\ 0 \end{bmatrix},$$

or, equivalently

$$Z = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix}, \quad Z \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix}L.$$
That is, we want a specific basis of an invariant subspace of a Hamiltonian matrix, i.e. a matrix $Z$ such that

$$JZ \text{ symmetric, } J \equiv \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$ 

Hamiltonian eigenvalue problems show up in a surprising variety of places in addition to optimal control. The theory of eigenvalue problems for Hamiltonian and skew-Hamiltonian matrices is reasonably well developed, and the eigenvalues have a special symmetry to them. There is now good software for these classes of problems that exploits the structure — though not in LAPACK. The right place to look for these solvers is in the SLICOT package.

2 \hspace{1em} Polynomial eigenvalue problems

A nonlinear eigenvalue problem is an equation of the form

$$T(\lambda)v = 0$$

where $T : \mathbb{C} \to \mathbb{C}^{n \times n}$ is a matrix-valued function. The most common nonlinear eigenvalue problems are polynomial eigenvalue problems in which $T$ is a polynomial; and most common among the polynomial eigenvalue problems are the quadratic eigenvalue problems

$$(\lambda^2 M + \lambda D + K)u = 0.$$ 

As the notation might suggest, one of the natural sources of quadratic eigenvalue problems is in the analysis of damped unforced vibrations in mechanical (or other physical) systems. In this context, $M$, $D$, and $K$ are the mass, damping, and stiffness matrices, and the eigenvalue problem arises from the search for special solutions to the equation

$$M \ddot{x} + D \dot{x} + K x = 0$$

where $x(t) = u \exp(\lambda t)$. We note that the mass matrix is often symmetric and positive definite; in this case, we can apply a change of variables to convert to a problem in which the leading term involves an identity matrix. We will assume this case for the remainder of our discussion.

When studying the solution of higher-order differential equations, a standard trick is to put the system into first-order form by introducing auxiliary
variables for derivatives. For example, we would put our model second-order
unforced vibration equation into first-order form by introducing the variable
\( v = \dot{x} \); then (assuming \( M = I \)), we have

\[
\begin{bmatrix}
\dot{v} \\
\dot{x}
\end{bmatrix} =
\begin{bmatrix}
-D & -K \\
I & 0
\end{bmatrix}
\begin{bmatrix}
v \\
x
\end{bmatrix}.
\]

Similarly, we can convert the quadratic eigenvalue problem into a standard
linear eigenvalue problem by introducing \( w = \lambda u \); then

\[
\lambda
\begin{bmatrix}
w \\
u
\end{bmatrix} =
\begin{bmatrix}
-D & -K \\
I & 0
\end{bmatrix}
\begin{bmatrix}
w \\
u
\end{bmatrix}.
\]

This process of converting a quadratic (or higher-order polynomial) eigen-
value problem into a linear eigenvalue problem in a higher-dimensional space
is called linearization (a somewhat unfortunate term, but the standard choice).
There are many ways to define the auxiliary variables, and hence many ways
to linearize a polynomial eigenvalue problem; the version we have described
is the companion linearization. Different linearizations are appropriate to
polynomial eigenvalue problems with different structure.

More generally, a “genuinely” nonlinear eigenvalue involves a matrix \( T(\lambda) \)
that depends on the spectral parameter \( \lambda \) as a more general non-rational
function. Typically, we restrict our attention to functions that are complex-
analytic in some domain of interest; these arise naturally in many applica-
tions, particularly in problems involving delay, radiation, and similar effects.
One thread in my own research has been to extend some of the theory we have
for the standard eigenvalue problem — results like Gershgorin and Bauer-
Fike — to this more general nonlinear case.

3 Pseudospectra

We conclude our discussion of eigenvalue-related ideas by revisiting the per-
turbation theory for the nonsymmetric eigenvalue problem from a somewhat
different perspective. In the symmetric case, if \( A - \hat{\lambda}I \) is nearly singular (i.e.
\( (A - \hat{\lambda}I)\dot{x} = r \) where \( \|r\| \ll \|A\|\|\dot{x}\| \)), then \( \hat{\lambda} \) is close to one of the eigenval-
ues of \( A \). But in the nonsymmetric case, \( A - \hat{\lambda}I \) may become quite close to
singular even though \( \hat{\lambda} \) is quite far from any eigenvalues of \( A \). The approxi-
mate null vector of \( A - \hat{\lambda}I \) is sometimes called a quasi-mode, and dynamical
systems defined via such a matrix $A$ are often characterized by long-lived transient dynamics that are well-described in terms of such quasi-modes.

In order to describe quasi-modes and long-lived transients, we need a systematic way of thinking about “almost eigenvalues.” This leads us to the idea of the $\epsilon$-pseudospectrum:

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \|(A - zI)^{-1}\| \geq \epsilon^{-1}\}.$$ 

This is equivalent to

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \exists E \text{ s.t. } \|E\| < \epsilon \text{ and } (A + E - zI) \text{ singular}\},$$

or, when the norm involved is the operator 2-norm,

$$\Lambda_\epsilon(A) = \{x \in \mathbb{C} : \sigma_{\min}(A - zI) < \epsilon\}.$$ 

There is a great deal of beautiful theory involving pseudospectra; as a guide to the area, I highly recommend *Spectra and Pseudospectra* by Mark Embree and Nick Trefethen.