Large-scale methods for Lyapunov equations

We give a hint of the methods used for large-scale equations.

We focus on Lyapunov equations, $AX + XA^* + BB^* = 0$.
(Then we can solve CAREs using Newton's method, for instance.)

Assumptions: A large and sparse with $\Lambda(A) \subset LHP$. $B \in \mathbb{R}^{n \times m}$, with $m \ll n$.

Actually, we may suppose $B = b \in \mathbb{R}^n$ without loss of generality: a rank-m matrix is the sum of m rank-1 matrices, and the equation is linear.

Assume A symmetric, normal or 'almost normal'. The algorithms often work for generic A, but the analysis works better for normal matrices.

ADI (alternating-direction implicit) iteration

Let $p_k < 0 \in \mathbb{R}$, so that $A + p_k I$ is never singular (and $A - p_k I$ may be singular if p_k is an eigenvalue of A).

Rewrite the Lyapunov equation as

$$(A + p_k I)X + X(A^* - p_k I) + bb^* = 0.$$

and formulate as a fixed-point equation (starting from $X_0 = 0$)

$$X_{k+1} = (A + p_k I)^{-1}(-bb^* - X_k(A^* - p_k I)).$$

Ugly, breaks symmetry: to restore it, make two steps with same p_k:

$$\begin{aligned} X_{k+\frac{1}{2}} &= (A + p_k I)^{-1}(-bb^* - X_k(A^* - p_k I)), \\ X_{k+1} &= (-bb^* - (A - \overline{p_k} I)X_{k+\frac{1}{2}})(A^* + \overline{p_k} I)^{-1} \\ &= -2\,\mathrm{Re}(p_k)(A + p_k I)^{-1} bb^* (A + p_k I)^{-*} + f_{p_k}(A) X_k f_{p_k}(A)^*, \end{aligned}$$

with $f_{p_k}(x) = \frac{x - \overline{p_k}}{x + p_k} = 1 - 2\,\mathrm{Re}(p_k)\frac{1}{x + p_k}$.

Low-rank ADI

$$X_{k+1} = -2\,\mathrm{Re}(p_k)(A + p_k I)^{-1} bb^* (A + p_k I)^{-*} + f_{p_k}(A) X_k f_{p_k}(A)^*$$

Can be rewritten in terms of the 'low-rank factor' of $X_k = Z_k Z_k^*$:

$$Z_{k+1} = \begin{bmatrix} \sqrt{-2\,\mathrm{Re}(p_k)}(A + p_k I)^{-1} b & f_{p_k}(A) Z_k \end{bmatrix}.$$

or in a more efficient form with only one inversion at each step:

$$Z_k = \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix}, \quad v_{k+1} = \alpha_k v_k + \beta_k (A + p_k I)^{-1} v_k.$$

(easier to see by looking at the first steps).

ADI: convergence

Convergence depends on the choices of p_k. Intuitively: good if $A - p_k I$ is small and $A + p_k I$ is large. Suggests taking p_k as (some of) the eigenvalues of A.

More formally:

$$X_{k+1} - X_* = f_{p_k}(A)(X_k - X_*)f_{p_k}(A)^* = \cdots = g(A)(X_0 - X_*)g(A)^*,$$

where $g(x) = \prod_{i=0}^{k} \frac{x-\overline{p_i}}{x+p_i}$.

If $A = V\Lambda V^{-1}$, then

$$\|g(A)\| \leq \kappa(V) \max_{\lambda\in\Lambda(A)} \prod_{i=0}^{k} \frac{|\lambda - \overline{p_i}|}{|\lambda + p_i|}.$$

How to choose p_k's that make this small? Easy if A has few / clustered eigenvalues.

ADI convergence

$$\eta_k = \min_{p_0,\dots,p_k} \max_{\lambda \in \Lambda(A)} \prod_{i=0}^{k-1} \frac{|\lambda - \overline{p_i}|}{|\lambda + p_i|}.$$

In general, tricky approximation theory problem. Typical approach: find an enclosing region for the eigenvalues of A (for instance, if $A = A^*$, all eigenvalues are in $[\lambda_{\min}, \lambda_{\max}]$); look for a polynomial that is 'small' on this region and 'large' on $[-\lambda_{\max}, -\lambda_{\min}]$.

In many cases, η_k decays with k — exponentially, or anyway fast.

Consequence The solution X has decaying singular values / low numerical rank.

(Good thing, because otherwise the problem would be hopeless: X is full.)

Residual computation

For $X_k = Z_k Z_k^*$, with $Z_k \in \mathbb{R}^{n\times k}$, we have

$$AZ_k Z_k^* + Z_k Z_k^* A^* + BB^* = \begin{bmatrix} Z_k & AZ_k & B \end{bmatrix} \begin{bmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} Z_k & AZ_k & B \end{bmatrix}^* .$$

Using QR or SVD of the tall thin $\begin{bmatrix} Z_k & AZ_k & B \end{bmatrix}$, we can compute residual norms in $O(nk^2)$.

Rational Arnoldi

The computed Z_k has columns of the form $r(A)b$, where $r(x) = q(x)/p(x)$, with denominator $p(x) = (x - p_0)(x - p_1)\dots(x - p_{k-1})$.

Definition

The rational Krylov subspace with poles $p_0, p_1, \dots, p_{k-1}$ is

$$R_k(A, b) = \{p(A)^{-1}q(A)b \colon \deg q < k\} = p(A)^{-1}K_k(A, b).$$

An orthogonal basis can be computed with a variant of Arnoldi: at each step, multiply the last vector v_k by $(A - p_k I)^{-1}$ and orthogonalize against the previous ones.

Plan First compute this subspace, then solve the projected equation.

Solving projected equations

Given an orthonormal basis U_k of $R_k(A, b)$:

1. Set $X_k = U_k Y_k U_k^*$;
2. Assume 'orthogonal residual': $U_k^*(AX_k + X_k A^* + BB^*)U_k = 0$.

Produces a projected Lyapunov equation

$$(U_k^* A U_k)Y + Y(U_k^* A U_k)^* + U_k^* BB^* U_k = 0.$$

Difficulty 1 Even if A stable, $U_k^* A U_k$ is not necessarily so.
Difficulty 2 (main one, common to ADI): good pole selection.
Pole selection can be critical for convergence. No good general strategies. Usually one tries some extremal eigenvalues of A and A^{-1} as p_k.