Iterative Solution of Algebraic Riccati Equations for Damped Systems

Kirsten Morris and Carmeliza Navasca

Abstract—Algebraic Riccati equations (ARE) of large dimension arise when using approximations to design controllers for systems modelled by partial differential equations. We use a modified Newton method to solve the ARE. Since the modified Newton method leads to a right-hand side of rank equal to the number of inputs, regardless of the weights, the resulting Lyapunov equation can be more efficiently solved. A low-rank Cholesky-ADI algorithm is used to solve the Lyapunov equation resulting at each step. The algorithm is straightforward to code. Performance is illustrated with an example of a beam, with different levels of damping. Results indicate that for weakly damped problems a low rank solution to the ARE may not exist. Further analysis supports this point.

INTRODUCTION

A classical controller design objective is to find a control $u(t)$ so that the objective function

$$
\int_0^\infty \langle x(t), Q x(t) \rangle + u^*(t) R u(t) dt
$$

is minimized where $R$ is a symmetric positive definite matrix and $Q$ is symmetric positive semi-definite. It is well-known that the solution to this problem is found by solving an algebraic Riccati equation

$$
A^* P + PA - P B R^{-1} B^* P = -Q.
$$

(2)

for a feedback operator

$$
K = -R^{-1} B' P.
$$

(3)

If the system is modelled by partial differential or delay differential equations then the state $x(t)$ lies in an infinite-dimensional space. The theoretical solution to this problem for many infinite-dimensional systems parallels the theory for finite-dimensional systems [11], [12], [18], [19]. In practice, the control is calculated through approximation. The matrices $A$, $B$, and $C$ arise in a finite dimensional approximation of the infinite dimensional system. Let $n$ indicate the order of the approximation, $m$ the number of control inputs and $p$ the number of observations. Thus, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$. There have been many papers written describing conditions under which approximations lead to approximating controls that converge to the control for the original infinite-dimensional system [4], [12], [15], [18], [19]. In this paper we will assume that an approximation has been chosen so that a solution to the Riccati equation (2) exists for sufficiently large $n$ and also that the approximating feedback operators converge. Unfortunately, particularly in partial differential equation models with more than one space dimension, many infinite-dimensional control problems lead to Riccati equations of large order. A survey of currently used methods to solve large ARE can be found in [8]. Two common iterative methods are Chandrasekhar [3], [9] and Newton-Kleinman iterations [14], [17], [28]. Here we use a modification to the Newton-Kleinman method first proposed by Banks and Ito [3] as a refinement for a partial solution to the Chandrasekhar equation.

Solution of the Lyapunov equation is a key step in implementing either modified or standard Newton-Kleinman. The Lyapunov equations arising in the Newton-Kleinman method have several special features: (1) the model order $n$ is generally much larger than the number of controls $m$ or number of observations $p$ and (2) the matrices are often sparse. We use a recently developed method [20], [26] that uses these features, leading to an efficient algorithm. Use of this Lyapunov solver with a Newton-Kleinman method is described in [27], [6], [7], [8].

In the next section we describe the algorithm to solve the Lyapunov equations. We use this Lyapunov solver with both standard and modified Newton-Kleinman to solve a number of standard control examples, including one with several space variables. Our results indicate that modified Newton-Kleinman achieves considerable savings in computation time over standard Newton-Kleinman. We also found that using the solution from a lower-order approximation as an ansatz for a higher-order approximation significantly reduced the computation time.

I. DESCRIPTION OF ALGORITHM

One approach to solving large Riccati equations is the Newton-Kleinman method [17]. The Riccati equation (2) can be rewritten as

$$
(A - B K)^* P + P (A - B K) = -Q - K^* R K.
$$

(4)

We say a matrix $A_n$ is Hurwitz if $\sigma(A_n) \subset \mathbb{C}_-$. Theorem 1.1: [17] Consider a stabilizable pair $(A, B)$ with a feedback $K_0$ so that $A - B K_0$ is Hurwitz. Define $S_i = A - B K_i$, and solve the Lyapunov equation

$$
S_i^* P_i + P_i S_i = -Q - K_i^* R K_i
$$

(5)

for $P_i$ and then update the feedback as $K_{i+1} = -R^{-1} B^* P_i$. Then

$$
\lim_{i \to \infty} P_i = P
$$

with quadratic convergence.

The key assumption in the above theorem is that an initial estimate, or ansatz, $K_0$, such that $A - B K_0$ is Hurwitz is available. For an arbitrary large Riccati equation, this condition may be difficult to satisfy. However, this condition is not restrictive for Riccati equations arising in control of infinite-dimensional systems. First, many of these systems
are stable even when uncontrolled and so the initial iterate $K_0$ may be chosen as zero. Second, if the approximation procedure is valid then convergence of the feedback gains is obtained with increasing model order. Thus, a gain obtained from a lower order approximation, perhaps using a direct method, may be used as an ansatz for a higher order approximation. This technique was used successfully in [14], [24], [28].

A modification to this method was proposed by Banks and Ito [3]. In that paper, they partially solve the Chandrasekhar equations and then use the resulting feedback $K$ as a stabilizing initial guess. Instead of the standard Newton-Kleinman iteration (5) above, Banks and Ito rewrote the Riccati equation in the form

$$(A - BK_i)^*X_i + X_i(A - BK_i) = -D_i^*RD_i,$$

$$D_i = K_i - K_{i-1}, K_{i+1} = K_i - B^*X_i.$$

Solution of a Lyapunov equation is a key step in each iteration of the Newton-Kleinman method, both standard and modified. Consider the Lyapunov equation

$$XA_o + A_o^*X = -DD^*$$  

where $A_o \in \mathbb{R}^{n \times n}$ is Hurwitz and $D \in \mathbb{R}^{n \times r}$. Factor $Q$ into $C^*C$ and let $n_Q$ indicate the rank of $C$. In the case of standard Newton-Kleinman, $r = m + n_Q$ while for modified Newton-Kleinman, $r = m$. If $A_o$ is Hurwitz, then the Lyapunov equation has a symmetric positive semidefinite solution $X$. As for the Riccati equation, direct methods such as Bartels-Stewart [5] are only appropriate for low model order and do not take advantage of sparsity in the matrices.

In [3] Smith’s method was used to solve the Lyapunov equations. For $p < 0$, define $U = (A_o - pI)(A_o + pI)^{-1}$ and $V = -2p(A_o^* + pI)^{-1}DD^*(A_o + pI)^{-1}$. The Lyapunov equation (6) can be rewritten as

$$X = U^*XU + V.$$  

In Smith’s method [30], the solution $X$ is found by using successive substitutions: $X = \lim_{i \to \infty} X_i$ where

$$X_i = U^*X_{i-1}U + V$$  

with $X_0 = 0$. Convergence of the iterations can be improved by careful choice of the parameter $p$ e.g. [29, pg. 297]. This method of successive substitution is unconditionally convergent, but has only linear convergence.

The ADI method [21], [32] improves Smith’s method by using a different parameter $p_i$ at each step. Two alternating linear systems,

$$(A_o^* + p_i I)X_{i-\frac{1}{2}} = -DD^* - X_{i-1}(A_o - p_i I)$$

$$(A_o^* + p_i I)X_{i+\frac{1}{2}} = -DD^* - X_{i-\frac{1}{2}}^*(A_o - p_i I)$$

are solved recursively starting with $X_0 = 0 \in \mathbb{R}^{n \times n}$ and parameters $p_i < 0$. If all parameters $p_i = p$ then the above equations reduce to Smith’s method. If the ADI parameters $p_i$ are chosen appropriately, then convergence is obtained in $J$ iterations where $J \ll n$. The parameter selection problem has been studied extensively [10], [21], [31]. For symmetric $A$ the optimal parameters can be easily calculated.

A heuristic procedure to calculate the parameters in the general case is in [26]. If the spectrum of $A$ is complex, we estimated the complex ADI parameters as in [10]. As the spectrum of $A$ flattens to the real axis the ADI parameters are closer to optimal.

If $A_o$ is sparse, then sparse arithmetic can be used in calculation of $X_i$. However, full calculation of the dense iterates $X_i$ is required at each step. By setting $X_0 = 0$, it can be easily shown that $X_i$ is symmetric and positive semidefinite for all $i$, and so we can write $X = ZZ^*$ where $Z$ is a Cholesky factor of $X$ [20], [26]. (A Cholesky factor does not need to be square or be lower triangular.) Substituting $Z_iZ_i^*$ for $X_i$, setting $X_0 = 0$, and defining $M_i = (A_o^* + p_i I)^{-1}$ we obtain the following iterates

$$Z_1 = \sqrt{-2p_1M_1D}$$

$$Z_i = [\sqrt{-2p_iM_iD}, M_i(A_o^* - p_i I)Z_{i-1}].$$

Note that $Z_1 \in \mathbb{R}^{n \times r}$, $Z_2 \in \mathbb{R}^{n \times 2r}$, and $Z_i \in \mathbb{R}^{n \times ir}$.

The algorithm is stopped when the Cholesky iterations converge within some tolerance. In [20] these itertes are reformulated in a more efficient form, using the observation that the order in which the ADI parameters are used is irrelevant. This leads to the algorithm shown in Table II.

This solution method results in considerable savings in computation time and memory. In calculation of the feedback $K$, the full solution $X$ never needs to be constructed. Also, the complexity of this method (CF-ADI) is also considerably less than that of standard ADI as shown in Table I. Recall that for standard Newton-Kleinman, $r$ is the sum of the rank of the state weight $n_Q$ and controls $m$. For modified Newton-Kleinman $r$ is equal to $m$. The reduction in complexity of CF-ADI method over ADI is marked for the modified Newton-Kleinman method.

### Table I

<table>
<thead>
<tr>
<th>Complexity of CF-ADI and ADI [20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse A</td>
</tr>
<tr>
<td>Full A</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>CHOLESKY-ADI METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given $A_o$ and $D$</td>
</tr>
<tr>
<td>Choose ADI parameters ${p_1, \ldots, p_J}$ with $\Re(p_i) &lt; 0$</td>
</tr>
<tr>
<td>Define $z_1 = \sqrt{-2p_1(A_o^* + p_1 I)^{-1}D}$</td>
</tr>
<tr>
<td>and $z_i = [z_i]$</td>
</tr>
<tr>
<td>For $i = 2, \ldots, J$</td>
</tr>
<tr>
<td>Define $W_i = (\sqrt{-2p_{i+1}})[I - (p_{i+1} - p_i)(A_o^* + p_i I)^{-1}]$</td>
</tr>
<tr>
<td>(1) $z_i = W_i z_{i-1}$</td>
</tr>
<tr>
<td>(2) If $|z| &gt; \text{tol}$</td>
</tr>
<tr>
<td>$Z_i = [Z_{i-1}, z_i]$</td>
</tr>
<tr>
<td>Else, stop.</td>
</tr>
</tbody>
</table>

The algorithm stops when the Cholesky iterations converge within some tolerance.
<table>
<thead>
<tr>
<th>$E$</th>
<th>$2.68 \times 10^{10}$ N/m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_b$</td>
<td>$1.64 \times 10^{-9}$ m⁴</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$1.02087$ kg/m</td>
</tr>
<tr>
<td>$C_v$</td>
<td>$2$ Ns/m²</td>
</tr>
<tr>
<td>$C_d$</td>
<td>$2.5 \times 10^8$ Ns/m²</td>
</tr>
<tr>
<td>$L$</td>
<td>$1$ m</td>
</tr>
<tr>
<td>$I_h$</td>
<td>$121.9748$ kg m²</td>
</tr>
<tr>
<td>$d$</td>
<td>$.041$ kg^{-1}</td>
</tr>
</tbody>
</table>

**Table III**

Table of physical parameters.

II. CONTROL OF EULER-BERNOULLI BEAM

In this section, we consider a Euler-Bernoulli beam clamped at one end ($r = 0$) and free to vibrate at the other end ($r = 1$). Let $w(r, t)$ denote the deflection of the beam from its rigid body motion at time $t$ and position $r$. The deflection is controlled by applying a torque $u(t)$ at the clamped end ($r = 0$). We assume that the hub inertia $I_h$ is much larger than the beam inertia $I_b$ so that $I_b \theta \approx u(t)$. The partial differential equation model with Kelvin-Voigt and viscous damping is

$$\rho w_{tt}(r, t) + C_v w_t(r, t) + \partial^2 w_{ttt}(r, t) = \frac{\rho}{I_b} u(t),$$

with boundary conditions

$$w(0, t) = 0 \quad w_r(1, t) = 0 \quad E I_b w_{rr}(1, t) + C_d I_b w_{rrr}(0, t) = 0 \quad \frac{\partial}{\partial r} [E I_b w_{rr}(r, t) + C_d I_b w_{rrr}(r, t)]_{r=1} = 0.$$

The values of the physical parameters in Table II are as in [2].

Let

$$H = \{ w \in H^2(0, 1) : w(0) = 0, w_t(0) = 0 \}$$

be the closed linear subspace of the Sobolev space $H^2(0, 1)$ and define the state-space to be $X = H \times L^2(0, 1)$ with state $z(t) = (w(\cdot, t), \frac{d}{dt} w(\cdot, t))$. A state-space formulation of the above partial differential equation problem is

$$\frac{d}{dt} x(t) = Ax(t) + Bu(t),$$

where

$$A = \begin{bmatrix} 0 & I \\ -E I_b \frac{d^2}{dr^2} & -C_v I_b \frac{d^4}{dr^4} & -C_d \rho \frac{d^4}{dr^4} \end{bmatrix}$$

and

$$B = \begin{bmatrix} 0 \\ I_h \end{bmatrix}$$

with domain

$$\text{dom} (A) = \{ (\phi, \psi) \in X : \psi \in H, M(L) = \frac{d}{dr} M(L) = 0 \}$$

where $M = E I_b \frac{d^2}{dr^2} \phi + C_d I_b \frac{d^2}{dr^2} \psi \in H^2(0, 1)$. We use $R = 1$ and define $C$ by the tip position:

$$w(1, t) = C[w(x, t) \cdot \dot{w}(x, t)].$$

Then $Q = C^* C$. We also solved the control problem with $Q = I$.

Let $H^{2N} \subset H$ be a sequence of finite-dimensional subspaces spanned by the standard cubic B-splines with a uniform partition of $[0, 1]$ into $N$ subintervals. This yields an approximation in $H^{2N} \times H^{2N}$ [16, e.g.] of dimension $n = 4N$. This approximation method yields a sequence of solutions to the algebraic Riccati equation that converge strongly to the solution to the infinite-dimensional Riccati equation corresponding to the original partial differential equation description [4], [22].

III. LOW RANK APPROXIMATIONS TO LYAPUNOV FUNCTION

Tables IV, V show the effect of varying $C_d$ on the number of iterations required for convergence. Larger values of $C_d$ (i.e. smaller values of $\alpha$) leads to a decreasing number of iterations. Small values of $C_d$ lead to a large number of required iterations in each solution of a Lyapunov equation.

Recall that the CF-ADI algorithm used here starts with a rank 1 initial estimate of the Cholesky factor and the rank of the solution is increased at each step. The efficiency of the Cholesky-ADI method relies on the existence of a low-rank approximation of the solution $X$ to the Lyapunov equation. This is true of many other iterative algorithms to solve large Lyapunov equations.

**Theorem 3.1:** For any symmetric, positive semi-definite matrix $X$ of rank $n$ let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be its eigenvalues. For any rank $k < n$ matrix $X_k$,

$$\|X - X_k\|_2 \leq \frac{\mu_{k+1}}{\mu_k} \|X\|_2.$$

**Proof:** See, for example, [13, Thm. 2.5.3].

Thus, the relative size of the largest and smallest eigenvalues determines the existence of a low rank approximation $X_k$ that is close to $X$, regardless of how this approximation is obtained.

The ratio $\mu_{k+1}/\mu_1$ is plotted for several values of $C_d$ in Figure 1. For larger values of $C_d$ the solution $X$ is closer to a low rank matrix than it is for smaller values of $C_d$.

A bound on the rank of the solution to a Lyapunov equation where $A$ is symmetric is given in [25]. A tighter bound on the error in a low-rank approximation has been obtained [1] in terms of the eigenvalues and eigenvectors of $A$. This bound is applicable to all diagonalizable matrices $A$. The bound for the case where the right-hand-side $D$ has rank one is as follows.

**Theorem 3.2:** [1, Thm 3.1] Let $V$ be the matrix of right eigenvectors of $A$, and denote its condition number by $\kappa(X)$. Denote the eigenvalues of $A$ by $\lambda_1, \ldots, \lambda_n$. There is a rank $k < n$ approximation to $X$ satisfying

$$\|X - X_k\|_2 \leq (n - k)^2 \delta_{k+1} (\kappa(V)\|D\|_2)^2$$

\[ (8) \]
where
\[
\delta_{k+1} = \frac{-1}{2 \text{Real}(\lambda_{k+1})} \prod_{j=1}^{k} \left| \frac{\lambda_{k+1} - \lambda_j}{\lambda_{k+1}^* + \lambda_j} \right|^2.
\]
The eigenvalues \( \lambda_i \) are ordered so that \( \delta_i \) are decreasing.

In order to calculate this bound, all the eigenvalues and eigenvectors of \( A \) are needed.

If the condition number \( \kappa(V) \) is not too large, for instance if \( A \) is normal, then \( \delta_{k+1}/\delta_1 \) gives a relative error estimate of the error \( \mu_{k+1}/\mu_1 \) in the approximation \( X_k \). This estimate is observed in [1] to estimate the actual decay rate of the eigenvalues of \( X \) quite well, even in cases where the decay is very slow.

Consider a parabolic partial differential equation, with \( A \)-matrix of the approximation is symmetric and negative definite. Then all the eigenvalues are real. A routine calculation yields that
\[
\delta_{k+1} \approx \frac{\lambda_1}{\lambda_k}
\]
and so the rate of growth of the eigenvalues determines the accuracy of the low rank approximation. The accuracy of the low-rank approximant can be quite different if \( A \) is non-symmetric. The solution \( X \) to the Lyapunov equation could be, for instance, the identity matrix [1], [25] in which case the eigenvalues of \( X \) do not decay.

It is observed through several numerical examples in [1] that it is not just complex eigenvalues that cause the decay rate to slow down, but the dominance of the imaginary parts of the eigenvalues as compared to the real parts, together with absence of clustering in the spectrum of \( A \). This effect was observed in the beam equation. Figure 2 shows the change in the spectrum of \( A \) as \( C_d \) is varied. Essentially, increasing \( C_d \) increases the angle that the spectrum makes with the imaginary axis. Note that the eigenvalues do not cluster. As the damping is decreased, the dominance of the imaginary parts of the eigenvalues of \( A \) increases and the decay rate of the eigenvalues of \( X \) slows. The Cholesky-ADI calculations for the beam with \( C_d = 0 \) did not terminate on a solution \( X_k \) with \( k < n \).

These observations are supported by results in the theory of control of partial differential equations. If \( C_d > 0 \) the semigroup for the original partial differential equation is parabolic and the solution to the Riccati equation converges uniformly in operator norm [18, chap.4]. However, if \( C_d = 0 \), the partial differential equation is hyperbolic and only strong convergence of the solution is obtained [19]. Thus, one might expect a greater number of iterations in the Lyapunov loop to be required as \( C_d \) is decreased. Consider the bound (8) for the beam with \( C_d = 0 \), and an eigenfunction basis for the approximation so that the eigenvalues of \( A \) are the exact eigenvalues. Defining \( c \), \( c = C_v/\rho \), \( c^2 = E I \), /\( \omega \) and indicating the roots of \( 1 + \cos(\omega) \cos(b(\omega)) \) by \( a_k \), the eigenvalues are
\[
\lambda_k = -c \rho / 2 \pm i \omega_k
\]
where
\[
\omega_k = \sqrt{c^2 - c^2 / 4}
\]

**TABLE IV**

<table>
<thead>
<tr>
<th>( C_w )</th>
<th>( C_d )</th>
<th>( \alpha )</th>
<th>N.-K. It's</th>
<th>Lyapunov It's</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1E04</td>
<td>1.5699</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3E05</td>
<td>1.5661</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>63.14</td>
</tr>
<tr>
<td>4E05</td>
<td>1.5654</td>
<td>3</td>
<td>1620:1620:1620</td>
<td>16.21</td>
<td></td>
</tr>
<tr>
<td>4E07</td>
<td>1.5370</td>
<td>3</td>
<td>1316:1316:1316</td>
<td>42.91</td>
<td></td>
</tr>
<tr>
<td>1E08</td>
<td>1.4852</td>
<td>3</td>
<td>744:744:744</td>
<td>18.01</td>
<td></td>
</tr>
<tr>
<td>5E08</td>
<td>1.3102</td>
<td>3</td>
<td>301:301:301</td>
<td>5.32</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE V**

<table>
<thead>
<tr>
<th>( C_w )</th>
<th>( C_d )</th>
<th>( \alpha )</th>
<th>N.-K. It's</th>
<th>Lyapunov It's</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1E04</td>
<td>1.5699</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3E05</td>
<td>1.5661</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>24.83</td>
</tr>
<tr>
<td>4E05</td>
<td>1.5654</td>
<td>2</td>
<td>1620:1</td>
<td>16.79</td>
<td></td>
</tr>
<tr>
<td>4E07</td>
<td>1.5370</td>
<td>2</td>
<td>1316:1</td>
<td>7.49</td>
<td></td>
</tr>
<tr>
<td>5E08</td>
<td>1.4852</td>
<td>2</td>
<td>301:1</td>
<td>2.32</td>
<td></td>
</tr>
</tbody>
</table>

[23] A simple calculation yields that for all \( k \),
\[
\delta_k / \delta_1 \approx \prod_{j=1}^{k-1} \frac{1}{1 + \left( \frac{w_k - w_j}{\omega} \right)^2} \approx 1.
\]

This indicates that the eigenvalues of the Lyapunov solution do not decay.

**IV. CONCLUSIONS**

The results indicate a problem with solving the algebraic Riccati equation for systems with light damping where the eigenvalues are not decaying rapidly. Although better choice of the ADI parameters might help convergence, the fact that the low rank approximations to the solution converge very slowly when the damping is small limits the achievable improvement. This may have consequences for control of coupled acoustic-structure problems where the spectra are closer to those of hyperbolic systems than those of parabolic systems.

![Fig. 1. Eigenvalue ratio for solution to beam ARE for varying \( C_d \)](chart.png)
Fig. 2. Beam: Spectrum of A for varying $C_d$

Fig. 3. Beam: Spectrum of uncontrolled and controlled systems

REFERENCES