A convergence rate for the ADI method for infinite-dimensional Lyapunov equations

Mark R. Opmeer *

* *Department of Mathematical Sciences, University of Bath, United Kingdom (e-mail: m.opmeer@bath.ac.uk).*

Abstract: We show that the ADI method for a class of infinite-dimensional Lyapunov equations with appropriately chosen shift parameters converges exponentially in the square root. The main assumption on the class of Lyapunov equations is that the main operator generates an analytic semigroup. Rather than directly analyzing the ADI algorithm, we instead use that the ADI error is bounded by the error made by applying quadrature to the inverse Laplace transform integral of the output map and we analyze the error made by this quadrature approximation.

Keywords: Distributed-parameter systems, Lyapunov equation, ADI method, Laplace transforms, Convergence of numerical methods.

1. INTRODUCTION

The standard method for solving the Lyapunov equation

$$A^*X + XA + C^*C = 0, \tag{1}$$

where $A \in \mathbb{C}^{N\times N}$ and $C \in \mathbb{C}^{m\times N}$ are given and $X \in \mathbb{C}^{N\times N}$ is the unknown, is the Bartels–Stewart method: Bartels and Stewart (1972) (this is e.g. implemented in the matlab function `lyap`). However, the complexity of this algorithm is $\mathcal{O}(N^3)$ and the storage requirement is $\mathcal{O}(N^2)$, which makes this method unfeasible when N is large (which is typically the case when A is the discretization of a differential operator). For this case of large N, two types of methods have emerged as most popular: rational Krylov methods and the ADI method: Benner and Saak (2013); Simoncini (2016). In this article we consider convergence of the ADI method. It is known that, for finite-dimensional systems as considered above, under some mild assumptions this algorithm convergences at an exponential rate. However, instead of considering a fixed discretization A of a differential operator, one should really consider a sequence of discretizations $(A_N)_{N=1}^{\infty}$ of growing dimension. Using the above mentioned result, for each $N \in \mathbb{N}$ one then obtains an exponential convergence rate r_N. The problem is that typically $r_N \to 0$ as $N \to \infty$ (see Section 2.3). In this article we consider the infinite-dimensional case and thereby circumvent this problem.

Perhaps surprisingly, we don't analyze the ADI algorithm directly. Instead we analyze the error made by a quadrature approximation of the inverse Laplace transform integral. In Section 2, we indicate how this leads to an error-bound for ADI and we return to this in more detail in Section 5 after we have completed the analysis of the inverse Laplace transform approximation.

In addition to the application to convergence of the ADI method, our analysis of the inverse Laplace transform approximation can also be applied to approximation of the impulse response of an infinite-dimensional system by a sum of exponentials (which can be interpreted as the impulse response of a finite-dimensional system). We elaborate on this in Section 4.

2. RELATING THE ADI METHOD TO THE INVERSE LAPLACE TRANSFORM

In this section we given an informal overview of the main idea of this article.

2.1 The ADI method

The Lyapunov equation (1) appears in control theory because for its solution we have

$$\langle Xx_0, x_0\rangle = \int_0^\infty |y(t)|^2\, dt,$$

where y satisfies

$$\dot{x}(t) = Ax(t), \qquad x(0) = x_0, \qquad y(t) = Cx(t), \tag{2}$$

(and for notational simplicity we assume $m = 1$, i.e. the single output case). Let $(\alpha_k)_{k=1}^n$ be a sequence of distinct complex numbers with positive real part and consider

$$V_n := \operatorname{span}\{t \mapsto \mathrm{e}^{-\alpha_k t} : k \in \{1, \ldots, n\}\} \subset L^2(0,\infty).$$

Let y_n denote the best approximation to y in V_n (i.e. the orthogonal projection of y onto V_n) and let X_n be such that

$$\langle X_n x_0, x_0\rangle = \int_0^\infty |y_n(t)|^2\, dt.$$

It was established in Opmeer et al. (2013) that X_n is precisely what ADI computes for the "shift-parameters" $(\alpha_k)_{k=1}^n$.

2.2 The inverse Laplace transform

If F is the Laplace transform of f, then for a suitable contour Γ in the complex plane we have

$$f(t) = \frac{1}{2i\pi}\int_\Gamma \mathrm{e}^{st}\, F(s)\, ds. \tag{3}$$

If we apply quadrature to this contour integral with nodes $(s_k)_{k=1}^n$ and weights $(w_k)_{k=1}^n$ (both chosen to be independent of t) then we obtain the approximation to $f(t)$:

$$\frac{1}{2i\pi}\sum_{k=1}^{n} \mathrm{e}^{s_k t} F(s_k) w_k.$$

In particular, we obtain an approximation to f in the subspace

$$\mathrm{span}\{t \mapsto \mathrm{e}^{s_k t} : k \in \{1, \dots, n\}\}.$$

Applying the above with $f(t) = C\,\mathrm{e}^{At} x_0$ and therefore $F(s) = C(sI - A)^{-1} x_0$ gives an approximation to y from (2) in the space V_n with $\alpha_k = -s_k$. Since the ADI method gives the best approximation in this space, we have

$$\|y - y_n^{\mathrm{ADI}}\|_{L^2(0,\infty)} \leq \|y - y_n^{\mathrm{QUAD}}\|_{L^2(0,\infty)}.$$

Therefore, by analyzing the quadrature error, we obtain an upper-bound on the ADI error.

2.3 Issues with the ADI error equation

For the sequence $(X_n)_{n=0}^{\infty}$ produced by the ADI method we have the error equation

$$X - X_n = M_n (X - X_0) M_n^*,$$

where

$$M_n := \Pi_{k=1}^{n} \mathcal{C}(\alpha_k), \qquad \mathcal{C}(\alpha) := (A + \alpha I)(A - \overline{\alpha} I)^{-1}.$$

When A is a Hurwitz matrix, the spectral radius of $\mathcal{C}(\alpha)$ is strictly smaller than 1 for any α with positive real part. It then follows that the spectral radius of M_n is also strictly smaller than 1. From the error equation it can then be concluded that $X - X_n$ decays exponentially in n as long as the sequence $(\alpha_k)_{k=1}^{\infty}$ belongs to a compact subset of the open right half-plane. When α_k converges to the boundary of the open right half-plane, then convergence may not occur: see (Massoudi et al., 2016, Example 1.1). More importantly for us, if A is an unbounded operator on an infinite-dimensional space, then the spectral radius of $\mathcal{C}(\alpha)$ equals one (regardless of what α is). Therefore, the above argument breaks down in this case.

In fact, there are examples (Opmeer, 2015, Section 5) where the singular values of X do not decay exponentially and therefore, for these examples, no method (including ADI) can possibly obtain exponential convergence.

For a sequence of finite-dimensional approximations $(A_N)_{N=1}^{\infty}$, $(C_N)_{N=1}^{\infty}$ of such examples, the ADI error equation analysis referred to above will give an exponential decay rate r_N, but necessarily $r_N \to 0$ as $N \to \infty$, since in the limiting infinite-dimensional case there is no exponential decay.

3. APPROXIMATING THE INVERSE LAPLACE TRANSFORM

In this section we consider the approximation of the inverse Laplace transform integral (3) by quadrature. In Subsection 3.1 we specify the assumptions made on the Laplace transform F, in Subsection 3.2 we specify the contour Γ and in Subsection 3.3, we specify the quadrature nodes and weights. In subsection 3.4 we consider an integral estimate which is crucial in the error analysis. In Subsection 3.5 we consider the discretization error whereas in Subsection 3.6 we consider the truncation error. These results are combined in Subsection 3.7 to obtain an L^p error bound. Finally in Subsection 3.8 we make a comparison with the literature on quadrature methods for the inverse Laplace transform.

3.1 Assumptions

Let $\omega > 0$, $\delta \in (0, \pi/2)$ and define

$$\Sigma_{\delta,\omega} := \{s \in \mathbb{C} : s \neq -\omega, |\arg(s + \omega)| < \pi - \delta\},$$

which geometrically is the complement of the sector in the complex plane with vertex $-\omega$ and angle δ with the half-line $(-\infty, -\omega)$; see Figure 1. Let $\theta \geq 0$ and let $\mathcal{B}$ be a Banach space. We assume that $F : \Sigma_{\delta,\omega} \to \mathcal{B}$ is holomorphic and that there exists a $M > 0$ such that

$$\|F(s)\|_{\mathcal{B}} \leq M|s + \omega|^{\theta - 1}, \qquad \forall s \in \Sigma_{\delta,\omega}. \tag{4}$$

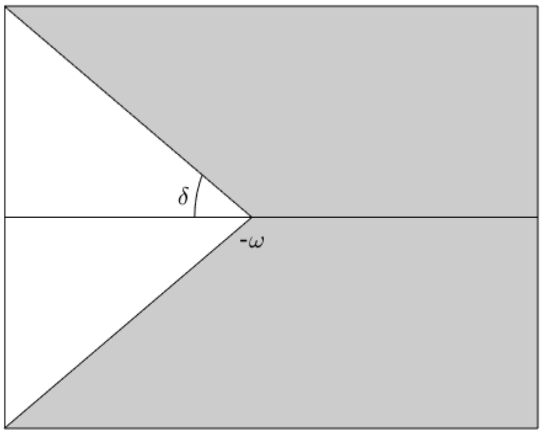

Fig. 1. The region $\Sigma_{\delta,\omega}$ in gray.

3.2 The contour

For $\gamma, \mu, \alpha \in \mathbb{R}$, consider the entire function

$$s(z) = -\gamma + \mu \sin(\alpha) + \mu \sin(iz - \alpha), \tag{5}$$

which with $z = x + iy$ can equivalently be written as

$$\begin{aligned} s(x + iy) = -\gamma + \mu \sin(\alpha) - \mu \sin(\alpha + y)\cosh(x) \\ + i\mu \cos(\alpha + y)\sinh(x). \end{aligned}$$

We assume that the parameters γ, μ, α satisfy

$$\begin{aligned} 0 < \mu < \frac{\omega}{\sin(\frac{\pi}{2} - \delta)}, \quad 0 < \alpha < \frac{\pi}{2} - \delta, \\ \mu \sin(\alpha) \leq \gamma < \omega + \mu \sin(\alpha) - \mu \sin\left(\frac{\pi}{2} - \delta\right). \end{aligned} \tag{6}$$

The function (5) then maps the strip

$$\left\{z \in \mathbb{C} : \mathrm{Im}(z) \in \left(-\alpha, \frac{\pi}{2} - \alpha - \delta\right)\right\}$$

into the sector $\Sigma_{\delta,\omega}$. Furthermore, the lines where $\mathrm{Im}(z)$ is constant are mapped onto hyperbolas with right-most point

$$x_r := -\gamma + \mu \sin(\alpha) - \mu \sin(\alpha + y),$$

and with asymptotes that make an angle

$$\frac{\pi}{2} - \alpha - y$$

with the half-line $(-\infty, x_r)$. We define for fixed $y \in \left(-\alpha, \frac{\pi}{2} - \alpha - \delta\right)$ the contour $\Gamma_y : \mathbb{R} \to \Sigma_{\delta,\omega}$ by

$$\Gamma_y(x) = -\gamma + \mu \sin(\alpha) - \mu \sin(\alpha + y) \cosh(x) + i\mu \cos(\alpha + y) \sinh(x).$$

See Figure 2 for Γ_0.

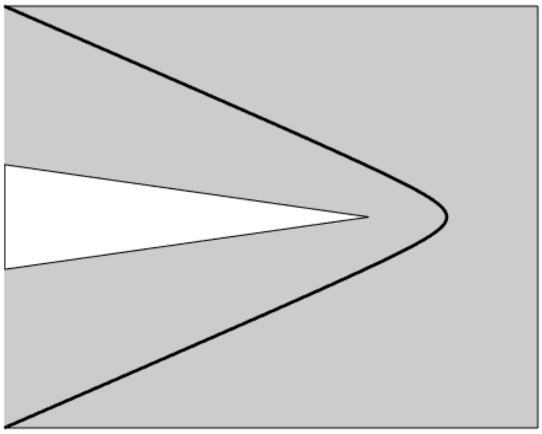

Fig. 2. The region $\Sigma_{\delta,\omega}$ in gray and the contour Γ_0 with parameters $\alpha = \frac{\pi}{4} - \frac{\delta}{2}$, $\mu = \frac{1}{2} \frac{\omega}{\sin(\frac{\pi}{2}-\delta)}$ and $\gamma = \mu \sin(\alpha) + (\omega - \mu \sin(\frac{\pi}{2} - \delta))/2$.

3.3 The quadrature

By transforming the trapezoidal rule on the line $\mathrm{Im}(z) = 0$ to the contour Γ_0 we obtain the following quadrature nodes and weights:

$$\begin{aligned} s_k &:= -\gamma + \mu \sin(\alpha) - \mu \sin(\alpha) \cosh(kh) + i\mu \cos(\alpha) \sinh(kh), \\ w_k &:= \mu \left[-\sin(\alpha) \sinh(kh) + i \cos(\alpha) \cosh(kh) \right], \end{aligned} \tag{7}$$

where $k \in \mathbb{Z}$ and $h > 0$ is the discretization parameter. We then have the infinite sum approximation to the inverse Laplace transform

$$I_h(t) := \frac{h}{2i\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{s_k t} F(s_k) w_k,$$

and for $n \in \mathbb{N}$ the truncated sum approximation

$$I_{h,n}(t) := \frac{h}{2i\pi} \sum_{k=-n}^{n} \mathrm{e}^{s_k t} F(s_k) w_k.$$

Defining

$$I(t) := \frac{1}{2i\pi} \int_{\Gamma_0} \mathrm{e}^{st} F(s)\, ds,$$

we have the *discretization error* $\|I(t) - I_h(t)\|_{\mathcal{B}}$ and the *truncation error* $\|I_h(t) - I_{h,n}(t)\|_{\mathcal{B}}$.

3.4 An integral estimate

Let $\varepsilon \in (0, \frac{\pi}{2} - \delta)$. By standard estimates we have that there exists a $K > 0$ such that for all $t > 0$ and all $y \in (-\alpha + \varepsilon, \frac{\pi}{2} - \alpha - \delta)$

$$\int_{\Gamma_y} \| \mathrm{e}^{st} F(s) \| \, |ds| \le K(1 + t^{-\theta}) \, \mathrm{e}^{t[-\gamma + \mu \sin(\alpha)]}.$$

3.5 The discretization error

Utilizing the integral estimate from Section 3.4, it follows from standard techniques (see e.g. Trefethen and Weideman (2014)) that for the discretization error we have

$$\|I(t) - I_h(t)\|_{\mathcal{B}} \le \frac{2R(t)}{\mathrm{e}^{2\pi\beta/h} - 1},$$

where

$$\begin{aligned} R(t) &:= K(1 + t^{-\theta}) \, \mathrm{e}^{t[-\gamma + \mu \sin(\alpha)]}, \\ \beta &:= \min \left\{ \alpha + \varepsilon, \frac{\pi}{2} - \alpha - \delta \right\}. \end{aligned}$$

3.6 The truncation error

Using similar techniques as for obtaining the integral estimates in Section 3.4 we obtain the following estimate on the truncation error: for all $\varepsilon_0 > 0$ there exists a $K_0 > 0$ such that

$$\|I_h(t) - I_{h,n}(t)\|_{\mathcal{B}} \le K_0 \, \mathrm{e}^{t[-\gamma + \mu \sin(\alpha)]} (1 + t^{-\theta}) \, \mathrm{e}^{t[-\mu \sin(\alpha) + \varepsilon_0] \cosh(hn)}.$$

3.7 The L^p error

From the pointwise discretization error in Section 3.5 we obtain for $p \in [1, \infty)$ with $p\theta < 1$ that there exists a $K_1 > 0$ such that

$$\|I - I_h\|_{L^p(0,\infty;\mathcal{B})} \le \frac{K_1}{\mathrm{e}^{2\pi\beta/h} - 1}.$$

From the pointwise truncation error in Section 3.5 we obtain for $p \in [1, \infty)$ with $p\theta < 1$ that there exists a $K_2 > 0$ such that

$$\|I_h - I_{h,n}\|_{L^p(0,\infty;\mathcal{B})} \le \frac{K_2}{\cosh(hn)}.$$

With

$$h = \sqrt{\frac{2\pi\beta}{n}}, \tag{8}$$

we then obtain there there exists a $K_2 > 0$ and $r > 0$ such that

$$\|I - I_{h,n}\|_{L^p(0,\infty;\mathcal{B})} \le K_2 \, \mathrm{e}^{-r\sqrt{n}}.$$

3.8 Comparison with the literature

The literature on quadrature approximation of the inverse Laplace transform usually considers scalar-valued rather than Banach space valued functions, but this is a minor difference. More fundamentally, the literature seemingly only considers pointwise approximation whereas we are interested in L^2 (or more generally L^p) approximation. Our arguments do initially consider pointwise approximation (Sections 3.5 and 3.6), but for these to be usable in the L^p context, we (in contrast to the existing literature) have to explicitly take into account the blow-up rate (the term $t^{-\theta}$ in our estimates). This relates to the assumption (4), which has not been considered in the literature (but it is a standard assumption in the theory of control systems with analytic semigroups).

In some of the literature, the nodes and weights are chosen to be dependent on t and a better convergence rate is obtained. For our application to ADI this t-dependence is not allowed since the quadature approximation will then no longer be in the space V_n.

4. APPROXIMATING THE IMPULSE RESPONSE

Although, as mentioned in the introduction, we are mainly interested in providing an error bound for the ADI method, the analysis in Section 3 also applies to approximation of impulse responses. We consider such impulse response approximation in this section.

For the class of systems from (Staffans, 2005, Theorem 5.7.3) where additionally the semigroup is assumed to be exponentially stable, the transfer function satisfies (4) for an appropriate θ. Therefore, the results from Section 3 apply and the quadrature approximation of the inverse Laplace transform gives rise to a sequence of approximations to the impulse response of the original system which converges exponentially in the square root in the L^p norm for all $p \in [1, 1/\theta)$. We formalize this in the following theorem.

Theorem 1. Let A be the generator of an exponentially stable analytic semigroup T on the Banach space $\mathcal{X}$, let $B \in \mathcal{L}(\mathcal{U}, \mathcal{X}_{\alpha_B})$, $C \in \mathcal{L}(\mathcal{X}_{\alpha_C}, \mathcal{Y})$ where $\mathcal{U}$ and $\mathcal{Y}$ are Banach spaces. Define $\theta := \alpha_C - \alpha_B$. Let $p \in [1, \infty)$ be such that $p\theta < 1$. Then the impulse response $f(t) := CT(t)B$ belongs to $L^p(0, \infty; \mathcal{L}(\mathcal{U}, \mathcal{Y}))$ and there exist $\omega > 0$ and $\delta \in (0, \pi/2)$ such that the transfer function $F(s) := C(sI - A)^{-1}B$ satisfies (4). With γ, μ, α satisfying (6), $(s_k)_{k=-\infty}^{\infty}$ and $(w_k)_{k=-\infty}^{\infty}$ chosen as in (7) and h defined by (8), define the sequence $(f_n)_{n=0}^{\infty}$ with $f_n : (0, \infty) \to \mathcal{L}(\mathcal{U}, \mathcal{Y})$ by

$$f_n(t) = \frac{h}{2i\pi} \sum_{k=-n}^{n} \mathrm{e}^{s_k t}\, C(s_k I - A)^{-1} B w_k.$$

Then there exist $K, r > 0$ such that

$$\|f - f_n\|_{L^p(0,\infty;\mathcal{L}(\mathcal{U},\mathcal{Y}))} \le K\, \mathrm{e}^{-r\sqrt{n}}\, .$$

4.1 An example

Consider the PDE

$$w_t(t, \xi) = w_{\xi\xi}(t, \xi) - w(t, \xi), \qquad t > 0, \quad \xi > 0,$$

with boundary condition

$$w_\xi(t, 0) = u(t),$$

and observation

$$y(t) = -w(t, 0).$$

The transfer function is then

$$F(s) = \frac{1}{\sqrt{s+1}},$$

so that the impulse response is

$$f(t) = \frac{\mathrm{e}^{-t}}{\sqrt{t}}. \tag{9}$$

(The example was chosen so that the impulse response could be analytically determined in terms of elementary functions; this is a rare occurence.) This example fits into the framework of Theorem 1 with $\mathcal{X} = L^2(0, \infty)$, $\mathcal{U} = \mathcal{Y} = \mathbb{C}$,

$$Ax = x'' - x, \qquad D(A) = \{x \in \mathcal{X} : x'(0) = 0\},$$
$$Cx = -x(0), \qquad B^* = C.$$

Since the growth-bound of A equals -1, we can choose ω to be any number in $(0, 1)$. Since A is self-adjoint, we can choose any $\delta \in (0, \pi/2)$. By standard interpolation and trace theory for Sobolev spaces we have $\mathcal{X}_{1/4} = H^{1/2}(0, \infty)$, $C \in \mathcal{L}(\mathcal{X}_{1/4}, \mathcal{Y})$ and $B \in \mathcal{L}(\mathcal{U}, \mathcal{X}_{-1/4})$. We can therefore choose $\theta = 1/2$. This can of course also be seen directly from the explicit expression for the transfer function.

In Figure 3 we've plotted the error (with $p = 1$) for various values of n.

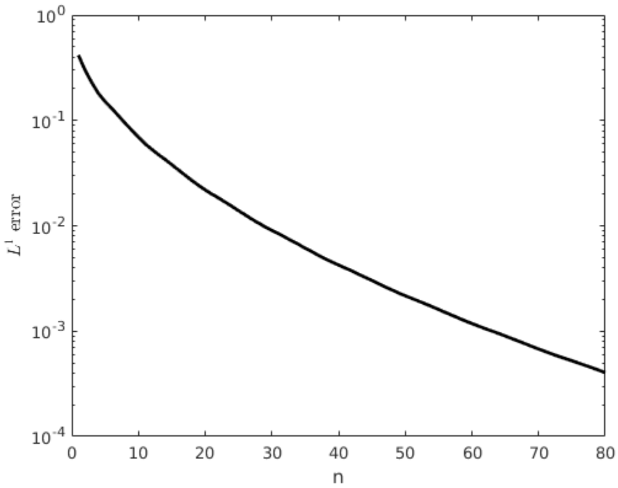

Fig. 3. Approximation error for the function (9). The dimension of the approximation is $2n + 1$.

5. THE CONNECTION WITH THE ADI METHOD

An issue in relating the quadrature of the inverse Laplace transform to the ADI method is that in ADI the shift parameters are an expanding sequence whereas in quadrature, the quadrature nodes used may completely change with a change in h and n. This can however be reconciled as follows.

When h is chosen according to (8), quadrupling n means halving h. In particular, for any $n \in \mathbb{N}$ we have $(s_k)_{k=-n}^{n} \subset (s_k)_{k=-4n}^{4n}$ (see Figures 4 and 5). In ADI we pick the shift parameters accordingly: we ensure that when n is a power of 4

$$\{\alpha_j : j \in \{1, \ldots, 2n+1\}\} = \{-s_k : k \in \{-n, \ldots, n\}\}. \tag{10}$$

Combined with the norm inequality from Section 2, this gives that, along a subsequence, the ADI method converges exponentially in the square root. Using monotonicity of the ADI method, we then obtain convergence of the whole sequence exponentially in the square root. We formalize this in the following theorem.

Theorem 2. Let A be the generator of an exponentially stable analytic semigroup T on the Hilbert space $\mathcal{X}$, let $C \in \mathcal{L}(\mathcal{X}_\theta, \mathcal{Y})$ where $\mathcal{Y}$ is a Hilbert space and $\theta < 1/2$. Then the Lyapunov equation (1) has a unique solution X. Furthermore, there exist $\omega > 0$ and $\delta \in (0, \pi/2)$ such that $F(s) := C(sI - A)^{-1}$ satisfies (4). With γ, μ, α satisfying (6), $(s_k)_{k=0}^{\infty}$ chosen as in (7) and h defined by (8), define the sequence $(X_n)_{n=1}^{\infty}$ as the ADI approximation to X where the shift parameters $(\alpha_k)_{k=1}^{n}$ satisfy (10). Then there exist $K, r > 0$ such that

$$\|X - X_n\|_{\mathcal{L}(\mathcal{X})} \le K\, \mathrm{e}^{-r\sqrt{n}}\, .$$

The assumptions of Theorem 2 on A and C are for example satisfied by the boundary observed convection–diffusion equation from (Massoudi et al., 2016, Section 8.1).

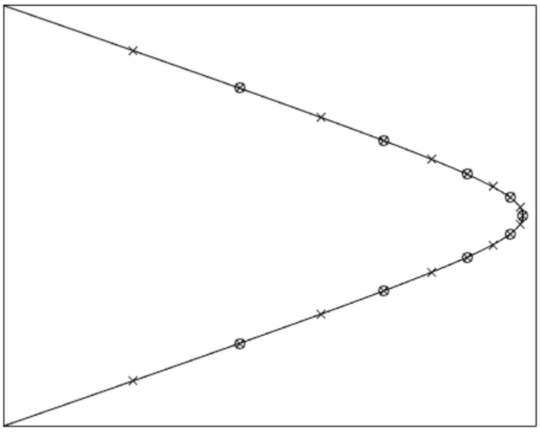

Fig. 4. Quadrature nodes with circles corresponding to 9 nodes ($n = 4$) and crosses corresponding to 33 nodes ($n = 16$): in this case the nodes are nested.

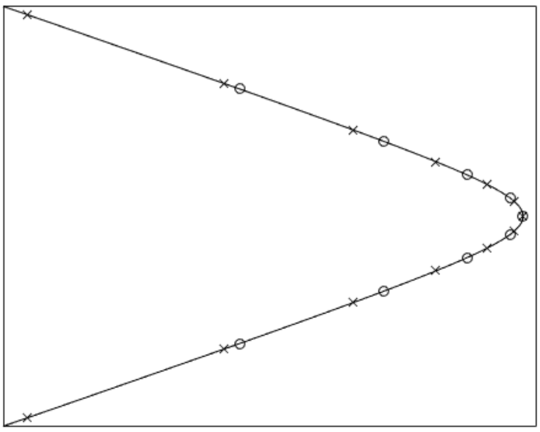

Fig. 5. Quadrature nodes with circles corresponding to 9 nodes ($n = 4$) and crosses corresponding to 13 nodes ($n = 6$): in this case the nodes are not nested.

REFERENCES

Bartels, R.H. and Stewart, G.W. (1972). Solution of the matrix equation $AX + XB = C$. *Comm. ACM*, 15(9), 820–826. doi:10.1145/361573.361582. URL `https://doi.org/10.1145/361573.361582`.

Benner, P. and Saak, J. (2013). Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey. *GAMM-Mitt.*, 36(1), 32–52. doi:10.1002/gamm.201310003. URL `https://doi.org/10.1002/gamm.201310003`.

Massoudi, A., Opmeer, M.R., and Reis, T. (2016). Analysis of an iteration method for the algebraic Riccati equation. *SIAM J. Matrix Anal. Appl.*, 37(2), 624–648. doi:10.1137/140985792. URL `https://doi.org/10.1137/140985792`.

Opmeer, M.R. (2015). Decay of singular values of the Gramians of infinite-dimensional systems. In *Proc. ECC*, 1183–1188. doi:10.1109/ECC.2015.7330700. URL `https://doi.org/10.1109/ECC.2015.7330700`.

Opmeer, M.R., Reis, T., and Wollner, W. (2013). Finite-rank ADI iteration for operator Lyapunov equations. *SIAM J. Control Optim.*, 51(5), 4084–4117. doi:10.1137/120885310. URL `https://doi.org/10.1137/120885310`.

Simoncini, V. (2016). Computational methods for linear matrix equations. *SIAM Rev.*, 58(3), 377–441. doi:10.1137/130912839. URL `https://doi.org/10.1137/130912839`.

Staffans, O. (2005). *Well-posed linear systems*, volume 103 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511543197. URL `https://doi.org/10.1017/CBO9780511543197`.

Trefethen, L.N. and Weideman, J.A.C. (2014). The exponentially convergent trapezoidal rule. *SIAM Rev.*, 56(3), 385–458. doi:10.1137/130932132. URL `https://doi.org/10.1137/130932132`.