

Efficient Solution of Algebraic Bernoulli Equations Using \mathcal{H} -Matrix Arithmetic

Ulrike Baur and Peter Benner

Abstract The algebraic Bernoulli equation (ABE) has several applications in control and system theory, e.g., the stabilization of linear dynamical systems and model reduction of unstable systems arising from the discretization and linearization of parabolic partial differential equations (PDEs). As standard methods for the solution of ABEs are of limited use for large-scale systems, we investigate approaches based on the matrix sign function method. This includes the solution of a linear least-squares (LLS) problem. Due to the large-scale setting we propose to solve this LLS problem via normal equations. To make the whole approach applicable in the large-scale setting, we incorporate structural information from the underlying PDE model into the approach. By using data-sparse matrix approximations, hierarchical matrix formats, and the corresponding formatted arithmetic we obtain an efficient solver having linear-polylogarithmic complexity. The proposed solver computes a low-rank representation of the ABE solution.

1 Introduction

We consider the algebraic Bernoulli equation (ABE)

$$A^T X + XA - XBB^T X = 0, \quad (1)$$

where $A \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$, and $X \in \mathbf{R}^{n \times n}$ is the matrix of unknowns. Recent methods for model order reduction of unstable dynamical systems [17] give the motiva-

Baur

Mathematics in Industry and Technology, Faculty of Mathematics, Chemnitz University of Technology e-mail: baur@mathematik.tu-chemnitz.de

Benner

Mathematics in Industry and Technology, Faculty of Mathematics, Chemnitz University of Technology e-mail: benner@mathematik.tu-chemnitz.de

tion for an efficient numerical solution of large-scale ABEs. Thereby, the general assumption

$$\Lambda(A) \cap \mathbf{C}^- \neq \emptyset, \quad \Lambda(A) \cap \mathbf{C}^+ \neq \emptyset, \quad \Lambda(A) \cap i\mathbf{R} = \emptyset \quad (2)$$

is given, using the notation $\Lambda(A)$ for the spectrum of A , $i = \sqrt{-1}$ and \mathbf{C}^- (\mathbf{C}^+) for the open left (right) half complex plane. The major part of the computational complexity of the balancing-related methods for model order reduction stems from the solution of large-scale ABEs and Lyapunov equations. In general, numerical methods for matrix equations have a complexity of $\mathcal{O}(n^3)$ (see, e.g., [7, 16]) and therefore, all these approaches are restricted to problems of moderate size. To overcome this limitation for a special class of practically relevant large-scale systems, recent approaches for the solution of Lyapunov equations [2, 3] combine iterative solvers based on the *sign function method* [15] with the hierarchical matrix format and the corresponding arithmetic. This idea is extended to the numerical solution of large-scale ABEs in the following.

This paper is organized as follows. In Section 2, we describe the sign function iteration for the solution of matrix equations and provide some basic facts of the \mathcal{H} -matrix format and the corresponding formatted arithmetic. The \mathcal{H} -matrix based sign function method is introduced in Section 3. In Section 3.1 we explain how the ABE solution is computed in low-rank factorized form by solving an LLS problem via normal equations. Symmetrizing this solution is explained in Section 3.2 and the derived method is tested on a numerical example in Section 4.

2 Theoretical Background

Necessary basics of the sign function iteration and of the data-sparse hierarchical matrix format are provided in this section.

2.1 The Matrix Sign Function

It is well known that a solution X of an ABE (which is a homogeneous algebraic Riccati equation) can be derived from the invariant subspace of the associated Hamiltonian matrix Z as

$$\underbrace{\begin{pmatrix} A & BB^T \\ 0 & -A^T \end{pmatrix}}_{=:Z} \begin{pmatrix} I_n \\ -X \end{pmatrix} = \begin{pmatrix} I_n \\ -X \end{pmatrix} (A - BB^T X),$$

see for instance [14]. Thus, if (A, B) is stabilizable and (2) holds, the unique stabilizing solution X of (1) can be computed by the Z -invariant subspace corresponding to the stable eigenvalues, i.e. $\Lambda(A - BB^T X) \subset (\Lambda(Z) \cap \mathbf{C}^-)$. Using spectral

projection, the kernel of the projector onto the anti-stable Z -invariant subspace $P_+ := (I_{2n} + \text{sign}(Z))/2$ describes the stable Z -invariant subspace. Thus, the stabilizing solution X can be derived from the LLS problem

$$\frac{1}{2}(I_{2n} + \text{sign}(Z)) \begin{pmatrix} I_n \\ -X \end{pmatrix} = 0. \quad (3)$$

One of the numerical methods to compute the sign of Z is based on the Newton iteration for $Z^2 = I$ [15]. To describe this method, consider a matrix $Z \in \mathbf{R}^{n \times n}$ with no eigenvalues on the imaginary axis. The *matrix sign function* for Z is defined by the real version of the Jordan canonical form

$$\text{sign}(Z) := S^{-1} \begin{bmatrix} I_\ell & 0 \\ 0 & -I_{n-\ell} \end{bmatrix} S, \quad \text{with} \quad Z = S^{-1} \begin{bmatrix} J_\ell^+ & 0 \\ 0 & J_{n-\ell}^- \end{bmatrix} S,$$

and $S \in \mathbf{R}^{n \times n}$, $\Lambda(J_\ell^+) \subset \mathbb{C}^+$, $\Lambda(J_{n-\ell}^-) \subset \mathbb{C}^-$.

To compute the matrix sign function, we use the Newton iteration applied to $(\text{sign}(Z))^2 = I_n$:

$$Z_0 \leftarrow Z, \quad Z_{j+1} \leftarrow \frac{1}{2}(Z_j + Z_j^{-1}).$$

This so called sign function iteration converges globally quadratically to the sign of Z and is well-behaved in finite-precision arithmetic. In order to solve the ABE (1) satisfying (2), the sign function is applied to the Hamiltonian Z associated with (1). By the block structure of Z , the iteration splits into two parts

$$\begin{aligned} A_0 &\leftarrow A, & A_{j+1} &\leftarrow \frac{1}{2}(A_j + A_j^{-1}), \\ B_0 &\leftarrow B, & B_{j+1} &\leftarrow \frac{1}{\sqrt{2}} \begin{bmatrix} B_j & A_j^{-1} B_j \end{bmatrix}, \quad j = 0, 1, 2, \dots, \end{aligned} \quad (4)$$

with quadratic convergence rate and

$$\text{sign}(Z) = \begin{pmatrix} A_\infty & B_\infty B_\infty^T \\ 0 & -A_\infty^T \end{pmatrix}, \quad (5)$$

using the notations

$$A_\infty := \lim_{j \rightarrow \infty} A_j, \quad B_\infty := \lim_{j \rightarrow \infty} B_j.$$

In [5, 6], this iteration scheme is used for solving Lyapunov equations and modified for the direct computation of the Cholesky (or full-rank) factors. We review the iteration scheme in Section 3 and propose further improvements using the hierarchical matrix format as briefly introduced in the next section.

2.2 \mathcal{H} -Matrix Arithmetic Introduction

In [11], the sign function method for solving algebraic Riccati equations is combined with a data-sparse matrix representation and a corresponding approximate arithmetic. This initiated the idea to use the \mathcal{H} -matrix format for computing low-rank solutions of ABEs. As our approach also makes use of this \mathcal{H} -matrix format, we will introduce some of its basic facts in the following.

The \mathcal{H} -matrix format is a data-sparse representation for a special class of matrices, which often arise in applications. Matrices that belong to this class result, for instance, from the discretization of partial differential or integral equations. Exploiting the special structure of these matrices in computational methods yields reduced computing time and memory requirements. A detailed description of the \mathcal{H} -matrix format can be found, e.g., in [9, 10, 12, 13].

The basic idea of the \mathcal{H} -matrix format is to partition a given matrix recursively into submatrices $M_{|_{r \times s}}$ that admit low-rank approximations, $\text{rank}(M_{|_{r \times s}}) \leq k$, where k denotes the block-wise rank. The corresponding submatrix is stored in factorized form as

$$M_{|_{r \times s}} = AB^T, \quad A \in \mathbb{R}^{r \times k}, B \in \mathbb{R}^{s \times k},$$

all remaining blocks correspond to submatrices which are stored in the usual dense matrix format.

The set of \mathcal{H} -matrices of block-wise rank k is denoted by $\mathcal{M}_{\mathcal{H},k}$. The storage requirements for a matrix $M \in \mathcal{M}_{\mathcal{H},k}$ are

$$\mathcal{N}_{\mathcal{M}_{\mathcal{H},k}\text{St}} = \mathcal{O}(n \log(n)k)$$

instead of $\mathcal{O}(n^2)$ for the original (full) matrix. We denote by $M_{\mathcal{H}}$ the hierarchical approximation of a matrix M . The formatted arithmetic \oplus (\ominus), \odot , $(\cdot)_{\mathcal{H}}^{-1}$ is a means to close the set of \mathcal{H} -matrices under addition, multiplication and inversion. These operations in formatted arithmetic are performed block-wise with exact addition or multiplication followed by truncating the resulting block back to rank k using a best Frobenius norm approximation. The truncation operator, denoted by \mathcal{T}_k , is realized by a truncated singular value decomposition, see, e.g., [10] for more details. For two matrices $A, B \in \mathcal{M}_{\mathcal{H},k}$ and a vector $v \in \mathbf{R}^n$ we consider the formatted arithmetic operations, which all have linear-polylogarithmic complexity:

$$\begin{aligned} v \mapsto Av &: && \mathcal{O}(n \log(n)k), \\ A \oplus B &= \mathcal{T}_{\mathcal{H},k}(A+B) &: & \mathcal{O}(n \log(n)k^2), \\ A \odot B &= \mathcal{T}_{\mathcal{H},k}(AB) &: & \mathcal{O}(n \log^2(n)k^2). \end{aligned} \quad (6)$$

In this work the \mathcal{H} -inverse $A_{\mathcal{H}}^{-1}$ of a matrix A is computed using an approximate \mathcal{H} -LU factorization $A \approx L_{\mathcal{H}}U_{\mathcal{H}}$ followed by an \mathcal{H} - forward and \mathcal{H} - backward substitution. The complexity of the \mathcal{H} -inversion is $\mathcal{O}(n \log^2(n)k^2)$.

Note that in practice, the blockwise rank is chosen adaptively for each matrix block instead of using a fixed rank k . Thus, the rank in each block $M_{|_{r \times s}}$ is deter-

mined so that the formatted operation yields an error less than or equal to a prescribed accuracy ε . We will use the \mathcal{H} -matrix structure to compute solution factors of ABEs, which reduces the complexity and the storage requirements of the underlying iteration scheme.

3 Efficient Solution of Large-Scale ABEs by use of the \mathcal{H} -matrix arithmetic

It is observed that in usual applications which stem from the discretization of some elliptic partial differential operator the B -iterates in (4) and thus the solution X of the ABE (1) have a small (numerical) rank. Thus, memory requirements are reduced by computing low-rank approximations to the factors directly. Furthermore, the hierarchical matrix format is incorporated in the sign function iteration (7) to reduce the cubic complexity and the quadratic storage requirements:

$$A_{j+1} \leftarrow \frac{1}{2}(A_j \oplus A_{\mathcal{H},j}^{-1}), \quad (7)$$

$$B_{j+1} \leftarrow \frac{1}{\sqrt{2}} \left[B_j, A_{\mathcal{H},j}^{-1} B_j \right], \quad j = 0, 1, 2, \dots \quad (8)$$

for details see [1, 3, 6]. Since the number of columns of B_j in (8) is doubled in each iteration step, it is proposed in [6] to apply a rank-revealing LQ factorization (RRLQ) [8] in order to reveal the expected low numerical rank. We denote the numerical rank determined by a threshold τ by t . Thus, after convergence, $B_\infty \in \mathbb{R}^{n \times t}$. Since the spectral norm of an \mathcal{H} -matrix can be computed without much effort, it is advised to choose

$$\|A_{j+1} - A_j\|_2 \leq \text{tol} \|A_{j+1}\|_2$$

as stopping criterion for the iteration.

3.1 Solving the LLS Problem (3)

When the sign function iteration has converged, including (5) for the sign of Z in (3), the LLS problem is equivalently given by

$$\underbrace{\begin{pmatrix} B_\infty B_\infty^T \\ I_n - A_\infty^T \end{pmatrix}}_{\tilde{A}} X = \underbrace{\begin{pmatrix} I_n + A_\infty \\ 0_n \end{pmatrix}}_{\tilde{b}}. \quad (9)$$

It admits a unique solution if $\text{rank}(\tilde{A}) = n$. Since $\text{rank}(I_n - A_\infty^T) = n - \ell$, where ℓ is the number of unstable eigenvalues of A , we must have $\text{rank}(B_\infty B_\infty^T) \geq \ell$.

To proceed the computations in low complexity we solve the problem using the normal equations

$$\tilde{A}^T \tilde{A} X = \tilde{A}^T \tilde{b},$$

exploiting that B_∞ is of low rank and A_∞ is stored as \mathcal{H} -matrix. The matrices involved are computed in the following way:

$$\begin{aligned} \underbrace{\tilde{A}^T \tilde{A}}_{\mathcal{H}\text{-matrix}} &= \underbrace{B_\infty B_\infty^T B_\infty B_\infty^T}_{\text{low rank}} + \underbrace{(I_n - A_\infty)}_{\mathcal{H}\text{-matrix}} \underbrace{(I_n - A_\infty)^T}_{\mathcal{H}\text{-matrix}} \\ \underbrace{\tilde{A}^T \tilde{b}}_{\text{low rank}} &= \underbrace{B_\infty [B_\infty^T + B_\infty^T A_\infty]}_{\text{low rank}} \end{aligned}$$

where the notation ‘‘low rank’’ indicates that the matrix is stored as product of two rectangular matrices.

Using the \mathcal{H} -Cholesky-decomposition of $\tilde{A}^T \tilde{A} = CC^T$ and \mathcal{H} -based forward and backward substitutions, we compute the stabilizing solution X of (1) as low rank matrix, i.e.

$$X = X_1 X_2^T, \quad X_1 = C^{-1} B_\infty, \quad X_2 = C^{-1} (A_\infty^T B_\infty + B_\infty) = C^{-1} A_\infty^T B_\infty + X_1. \quad (10)$$

3.2 Symmetrizing the low-rank presentation of ABE solutions obtained by normal equations

The stabilizing solution X of an ABE is known to be symmetric [4]. This property is not reflected in the representation (10) which is not a problem for certain applications as model order reduction of unstable systems. But in case that symmetry of X is required we give a procedure that achieves this task.

Let $B_\infty \in \mathbb{R}^{n \times t}$. From [4] we know that $\text{rank}(X) = \ell$, thus $t \geq \ell$. As $X = X^T \geq 0$ we have $X_1 X_2^T = X_2 X_1^T$ and $X_1 X_2^T$ is the positive semidefinite square root of X^2 ,

$$X^2 = (X_1 X_2^T)^2 = X_1 X_2^T X_2 X_1^T. \quad (11)$$

Now let $X_1 = Q_1 R_1$ be a thin QR decomposition with $Q_1 \in \mathbb{R}^{n \times t}$, $R_1 \in \mathbb{R}^{t \times t}$ upper triangular and compute a singular value decomposition

$$X_2 R_1^T = U \Sigma V^T. \quad (12)$$

We then get from (11)

$$\begin{aligned} X^2 &= Q_1 R_1 X_2^T X_2 R_1^T Q_1^T = Q_1 (U \Sigma V^T)^T (U \Sigma V^T) Q_1^T \\ &= (Q_1 V \Sigma^T) (Q_1 V \Sigma^T)^T = (Q_1 V \hat{\Sigma}^2 V^T Q_1^T)^2, \end{aligned}$$

where $\hat{\Sigma} = \Sigma(1:t, :) \in \mathbb{R}^{t \times t}$.

Hence, $Q_1 V \hat{\Sigma} V^T Q_1^T$ is the positive semidefinite square root of X^2 and $Q_1 V \hat{\Sigma}^{\frac{1}{2}} V^T Q_1^T$ is the positive semidefinite square root of X . Due to uniqueness of semidefinite square roots,

$$X = X_1 X_2^T = Q_1 V \hat{\Sigma} V^T Q_1^T.$$

A rank- t factor of X is thus given by

$$Y = Q_1 V \hat{\Sigma}^{\frac{1}{2}}$$

as $X = YY^T$.

Remark 1. Note that the accumulation of $U \in \mathbb{R}^{n \times t}$ in (12) is not necessary which reduces the cost of the SVD computation from $14nt^2 + 8t^3$ to $4nt^2 + 8t^3$ flops.

4 Numerical Results

In this section we examine the accuracy and complexity of the data-sparse approach for the numerical solution of ABEs. As exemplary system we consider the following reaction-diffusion equation

$$\frac{\partial \mathbf{x}}{\partial t}(t, \xi) = \Delta \mathbf{x}(t, \xi) + c \mathbf{x}(t, \xi) + b(\xi)u(t), \quad \xi \in (0, 1)^2, t \in (0, \infty),$$

which is discretized in space by finite elements, leading to the LTI system

$$\dot{x}(t) = \underbrace{(\tilde{A} + cI_n)}_{:=A} x(t) + Bu(t). \quad (13)$$

For the problem sizes $n = 4096$ and $n = 16,384$, we choose the parameter c such that one eigenvalue of the coefficient matrix A has positive real part: $\lambda \approx 0.25$. We compute the relative residual

$$\frac{\|A^T X + XA - XBB^T X\|_F}{2(\|A\|_F \|X\|_F) + \|X\|_F^2 \|BB^T\|_F}$$

of the ABE (1) obtained by applying (7), (8) and (10) to the unstable system (13). We vary the parameters τ for the numerical rank decision in the RRLQ factorization and ε , the approximation error in the adaptive rank choice of the \mathcal{H} -matrix arithmetic. The numerical rank of B_∞ , the computational time and the accuracy are depicted in Table 1. We observe in Table 1 high accuracy in the solution factors computed with the algorithm in \mathcal{H} -matrix arithmetic by low numerical ranks of B_∞ and in low execution time. The results of the parameter variation show the expected behavior, we have increasing accuracy as ε gets smaller and the relative residual is observed to remain bounded from above for increasing problem size.

n	ε	τ	# it.	rank(B_∞, τ)	time [sec]	rel. residual
4096	1.e-04	1.e-04	26	36	261	7.1e-06
	1.e-06	1.e-04	22	14	391	1.8e-07
	1.e-08	1.e-04	19	14	635	3.8e-08
	1.e-06	1.e-06	22	31	395	1.8e-07
	1.e-08	1.e-06	19	21	636	3.7e-08
	1.e-08	1.e-08	19	39	639	3.7e-08
16,384	1.e-04	1.e-04	27	34	2376	2.3e-05
	1.e-06	1.e-04	26	15	4235	5.2e-07
	1.e-08	1.e-04	22	14	7136	6.1e-09
	1.e-06	1.e-06	26	42	4273	5.2e-07
	1.e-08	1.e-06	22	23	7150	6.0e-09
	1.e-08	1.e-08	22	42	7183	5.9e-09

Table 1 Accuracy and rank $\text{rank}(B_\infty, \tau)$ of the computed ABE solution for different problem sizes and parameter combinations.

References

1. Baur, U.: Control-Oriented Model Reduction for Parabolic Systems. Dissertation, Inst. f. Mathematik, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, FRG (2008)
2. Baur, U., Benner, P.: Factorized solution of the Lyapunov equation by using the hierarchical matrix arithmetic. *Proc. Appl. Math. Mech.* **4**(1), 658–659 (2004)
3. Baur, U., Benner, P.: Factorized solution of Lyapunov equations based on hierarchical matrix arithmetic. *Computing* **78**(3), 211–234 (2006)
4. Benner, P.: Computing Low-Rank Solutions of Algebraic Bernoulli Equations. Tech. rep., Fakultät für Mathematik, TU Chemnitz, 09107 Chemnitz, FRG (2007)
5. Benner, P., Claver, J., Quintana-Ortí, E.: Efficient solution of coupled Lyapunov equations via matrix sign function iteration. In: A.D. et al. (ed.) *Proc. 3rd Portuguese Conf. on Automatic Control CONTROL’98*, Coimbra, pp. 205–210 (1998)
6. Benner, P., Quintana-Ortí, E.: Solving stable generalized Lyapunov equations with the matrix sign function. *Numer. Algorithms* **20**(1), 75–100 (1999)
7. Datta, B.: *Numerical Methods for Linear Control Systems*. Elsevier Academic Press (2004)
8. Golub, G., Van Loan, C.: *Matrix Computations*, third edn. Johns Hopkins University Press, Baltimore (1996)
9. Grasedyck, L.: Theorie und Anwendungen Hierarchischer Matrizen. Dissertation, University of Kiel, Kiel, Germany (2001). In German, http://e-diss.uni-kiel.de/diss_454
10. Grasedyck, L., Hackbusch, W.: Construction and arithmetics of \mathcal{H} -matrices. *Computing* **70**(4), 295–334 (2003)
11. Grasedyck, L., Hackbusch, W., Khoromskij, B.N.: Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices. *Computing* **70**(2), 121–165 (2003)
12. Hackbusch, W.: A sparse matrix arithmetic based on \mathcal{H} -matrices. I. Introduction to \mathcal{H} -matrices. *Computing* **62**(2), 89–108 (1999)
13. Hackbusch, W., Khoromskij, B.N.: A sparse \mathcal{H} -matrix arithmetic. II. Application to multi-dimensional problems. *Computing* **64**(1), 21–47 (2000)
14. Lancaster, P., Rodman, L.: *The Algebraic Riccati Equation*. Oxford University Press, Oxford (1995)
15. Roberts, J.D.: Linear model reduction and solution of the algebraic Riccati equation by use of the sign function. *Internat. J. Control* **32**, 677–687 (1980)
16. Sima, V.: *Algorithms for Linear-Quadratic Optimization*, *Pure and Applied Mathematics*, vol. 200. Marcel Dekker, Inc., New York, NY (1996)
17. Zhou, K., Salomon, G., Wu, E.: Balanced realization and model reduction for unstable systems. *Int. J. Robust Nonlinear Control* **9**(3), 183–198 (1999)