Linear Algebra and its Applications ••• (••••) •••–•••

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Algebraic properties of solutions to certain nonlinear matrix equations

Valeria Simoncini [a,b,*]

[a] *Dipartimento di Matematica and (AM)^2, Alma Mater Studiorum - Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy*
[b] *IMATI-CNR, Pavia, Italy*

ARTICLE INFO

Article history:
Received 3 March 2025
Received in revised form 6 July 2025
Accepted 21 September 2025
Available online xxxx
Submitted by V. Mehrmann

This paper is dedicated to Daniel Szyld, on the occasion of his 70th birthday.

MSC:
65F30

Keywords:
Linear matrix equations
Matrix iterations
Singular value decay

ABSTRACT

We are interested in algebraic properties of the solution X to the linear or mildly nonlinear symmetric matrix equation $AX + XA + U\Phi(X)U^T + BB^T = 0$ with symmetric A. We analyze monotonicity and low-rank properties of closed form solutions, whenever available, with respect to the solution of the equation $AX + XA + BB^T = 0$. We extend this analysis to approximation recurrences, for which monotonicity and singular value decay properties are discussed.

* Correspondence to: Dipartimento di Matematica and (AM)^2, Alma Mater Studiorum - Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy.
E-mail address: valeria.simoncini@unibo.it.

https://doi.org/10.1016/j.laa.2025.09.017

1. Introduction

We are interested in investigating certain structural properties of the solution X to the following class of nonlinear equations,

$$AX + XA + U\Phi(X)U^T + BB^T = 0, \tag{1.1}$$

where the $n \times n$ matrix A is assumed to be symmetric and negative or positive definite, and U, B are tall matrices with full column rank. The function Φ stands for a linear or mildly nonlinear symmetric function depending on X, and is such that the dimensions of $\Phi(X)$ are conforming with those of U. In the following we shall denote with $\mathcal{M} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ the coefficient operator, $\mathcal{M}(X) = AX + XA + U\Phi(X)U^T$, and with $\mathcal{L} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ the portion of $\mathcal{M}$ corresponding to the Lyapunov operator, $\mathcal{L}(X) = AX + XA$.

The equation in (1.1) includes the problem

$$AX + XA + \sum_{k=1}^{\ell} N_k X N_k^T + BB^T = 0 \tag{1.2}$$

as a special (linear) case. This equation arises in the stability analysis of bilinear dynamical systems, see, e.g., [6],[13], [16], [20], where usually N_k has rank much smaller than its dimension. The equation in (1.2) has recently emerged as a natural formulation in different contexts, beyond control. This is the case, for instance, in problems where the unmixing of heterogeneous variables represents a computationally attractive solution strategy, such as in stochastic problems and space-time computations; see, e.g., [39]. The numerical solution of (1.2) is significantly more challenging than the Lyapunov equation $AX + XA + BB^T = 0$, since in general obtaining a closed form solution matrix is computationally expensive. Because of this, (1.2) has recently attracted great interest in the numerical community, and different strategies have been proposed, see, e.g., [4], [14], [19], [34], [36], [11], [30] and their references.

The generalization to (1.1) further broadens the class of problems that can be treated with similar procedures, including problems arising in different engineering applications, see, e.g., [5], [12], [18], [32]. While several successful computational advances have appeared in the recent literature, structural properties of the solution to (1.1) have been less analyzed. The fact that the solution is a matrix, and that it may be symmetric, poses questions on structural properties that have been nicely addressed for the Lyapunov equation. In this paper we are thus interested in analyzing how properties such as definiteness and singular value decay of the solution matrix to the Lyapunov equation carry over to the setting in (1.1). Assessing these properties is crucial when targeting a low rank approximation to the exact solution. Our analysis exploits the closed form of the solution that can be derived for certain choices of Φ. In particular, when available, a closed form highlights the composition of the solution in terms of two distinct components, from which the singular value decay can be readily deduced.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

In case a closed form is not available or computationally unfeasible, we analyze sequences of approximate solutions, some of which are well established. For these approximate solutions, we study how definiteness and monotonicity are preserved throughout the iteration. In doing so, we also derive a new iteration for the nonlinear case $\Phi(X) = V^T(X \otimes X)V$. This iteration leverages recently introduced closed form solution techniques for quasi-linear equations of type (1.1), where $\Phi(X)$ is replaced by a real valued function in X.

The new contributions of this manuscript are:

1. We derive closed form solutions for quadratic-bilinear matrix equations under precise hypotheses on the data (section 2.3);
2. We propose a new fixed-point iteration for solving quadratic-bilinear matrix equations for low-rank quadratic term (section 3.1);
3. We prove monotonicity results of fixed-point iterations for (1.1) in Proposition 4.1;
4. We prove error minimization in the matrix-energy norm in fixed-point iterations, building upon generic vector-based results in Proposition 4.2;
5. We prove error minimization in the matrix-energy norm of the Galerkin approximation in Proposition 4.3;
6. We generalize singular value decay features to the solution of (1.1) in section 5.

An outline of the paper is as follows. In the subsections of section 2 we single out three distinct but highly connected classes of problems that fall in the framework of (1.1). In section 3.1 we summarize some known and less exercised structural properties of approximation sequences, together with a new recurrence for one of the discussed nonlinear problems, which builds upon the construction methodology of other solutions. In section 3.2 we recall certain optimality properties of recently developed projection methods. In section 4 we derive monotonicity results for the discussed recurrences, their norms and their error norm. In section 5 we discuss decay properties that can be inferred by the already derived structure. Section 6 contains our final considerations.

1.1. Notation

Given two matrices $A = (A_{i,j}) \in \mathbb{R}^{m\times n}$ and B, the Kronecker product is defined in block form as

$$A \otimes B = \begin{bmatrix} A_{1,1}B & \cdots & A_{1,n}B \\ \vdots & \ddots & \vdots \\ A_{m,1}B & \cdots & A_{m,n}B \end{bmatrix}.$$

This matrix operator satisfies

$$\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X), \tag{1.3}$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

where $\mathrm{vec}(X)$ stacks the columns of X one below the other.

We use the notation $X \succeq 0$ to say that the square matrix X is symmetric and positive semi-definite, while $X \succeq Y$ means that $X - Y \succeq 0$. We define the energy-norm associated with a symmetric and positive definite operator $\mathcal{A}$ as $\|X\|_{\mathcal{A}}^2 = \mathrm{trace}(X\mathcal{A}(X))$. Moreover, $\|X\|$ denotes the (spectral) matrix norm induced by the Euclidean vector norm, while $\|X\|_F$ denotes the Frobenius norm.

We specifically use bold face letters for matrices and vectors of leading dimension n^2, while calligraphic fonts are used for matrix operators.

2. Problems with a quasi-linear matrix equation structure

In this section we discuss algebraic properties and closed form solutions of different closely related matrix equations. The problems and techniques of section 2.1 and section 2.2 are known, and are briefly reported as background for the developments of section 2.3 and of later relations. We refer to the references cited in the sections for a more detailed account of the related literature and the main algebraic properties of the two problems.

2.1. Multiterm linear matrix equations with low-rank structure

We consider the problem (1.2), where for the sake of the presentation, we work with $\ell = 1$, so that the sum consists of a single term. Let $N \equiv N_1$ be written as $N = UV^T$ and assume that U, V are full column rank matrices, with r columns.

Using the Kronecker products, $\boldsymbol{A} = A \otimes I + I \otimes A$, $\boldsymbol{b} = \mathrm{vec}(BB^T)$, and we can define $\boldsymbol{U} = U \otimes U$, $\boldsymbol{V} = V \otimes V$, so that $\mathrm{rank}(\boldsymbol{U}) = r^2 = \mathrm{rank}(\boldsymbol{V})$. Note that $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{r^2}]$ with $\boldsymbol{u}_j = u_k \otimes u_i$ where $j = (k-1)r + i$. The matrix equation can be written in vectorized form as the linear system $(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}^T)\boldsymbol{x} = \boldsymbol{f}$, with $\boldsymbol{f} = -\boldsymbol{b}$ and $\boldsymbol{x} = \mathrm{vec}(X)$. If r^2 is still much smaller than n, then $\boldsymbol{x}$ can be determined by means of the Sherman-Morrison-Woodbury (SMW) formula ([15], [21], [37]) as follows

$$\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{U}\boldsymbol{V}^T)^{-1}\boldsymbol{f} = \boldsymbol{A}^{-1}\boldsymbol{f} - \boldsymbol{A}^{-1}\boldsymbol{U}(I + \boldsymbol{V}^T\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^T\boldsymbol{A}^{-1}\boldsymbol{f}.$$

The problem can be solved without using this vector formulation, but solely relying on matrix equation solves. This key property was carefully implemented in [14] to achieve peak performance, and recently sharpened in [19]; we refer to these references for a recollection of previous attempts. Let

$$\boldsymbol{w} = \boldsymbol{A}^{-1}\boldsymbol{f} \quad \Leftrightarrow \quad AW + WA = -BB^T, \tag{2.1}$$

and note that W is symmetric. Let $U = [u_1, \ldots, u_r]$, $V = [v_1, \ldots, v_r]$ and again let $\boldsymbol{u}_j = \mathrm{vec}(u_i u_k^T)$ for $k, i = 1, \ldots, r$ such that $j = (k-1)r + i$. Then

$$\boldsymbol{p}_j = \boldsymbol{A}^{-1}\boldsymbol{u}_j \quad \Leftrightarrow \quad AP_j + P_jA = u_i u_k^T, \quad \boldsymbol{p}_j = \mathrm{vec}(P_j). \tag{2.2}$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

Note that for each j_1 there exists a j_2 such that $P_{j_1} = P_{j_2}^T$, so that not all Ps need to be explicitly computed. The inner system has coefficient matrix

$$H = I + \boldsymbol{V}^T \boldsymbol{A}^{-1} \boldsymbol{U} = I + \boldsymbol{V}^T \boldsymbol{P} = I + \boldsymbol{V}^T [\boldsymbol{p}_1, \dots, \boldsymbol{p}_{s^2}], \tag{2.3}$$

which can equivalently be computed by noticing that $\boldsymbol{v}_j^T \boldsymbol{p}_t = v_i^T P_t v_k$, $j = (k-1)r + i$. The solution of the $r^2 \times r^2$ linear system

$$Hg = \boldsymbol{V}^T \boldsymbol{w} \tag{2.4}$$

requires first computing the right-hand side. This can be obtained in terms of the original data and of (2.1) as $\boldsymbol{V}^T \boldsymbol{A}^{-1} \boldsymbol{f} = [v_1^T W v_1, v_2^T W v_1, \dots, v_r^T W v_r]^T$. The final solution is then given by $X = W - \sum_{j=1}^{r^2} P_j (g)_j$, where $(g)_j$ is the jth component of the vector g. The effective implementation of this procedure requires some care; we refer to [19] and its references for computational details.

The solution X is expressed as a linear combination of solutions to Lyapunov equations having low-rank known term. The following result emphasizes this fact while summarizing the above derivation. The proof consists of a rewriting of the closed form above.

Proposition 2.1. *Let $\mathcal{L} : X \to AX + XA$. Let G be such that $g = \mathrm{vec}(G)$ where g is the solution to (2.4), with H defined in (2.3). Then the solution X to (6.1) satisfies $\mathcal{L}(X) = -BB^T - UGU^T$.*

We notice that since X, $\mathcal{L}(X)$ and BB^T are symmetric, then also G must be symmetric, so that $X = W - \sum_{j=1}^{r(r+1)/2} (P_j + P_j^T)(g)_j$, with the indexes properly ordered.

We remark that although the derivation of Proposition 2.1 is very simple, the perspective is very convenient to highlight the dependence of the solution X on $[B, U]$, irrespective of the composition of G. In section 5 this property will be used and generalized to the nonlinear case.

2.2. Quasi-linear matrix equations

A special setting is determined by the problem

$$AX + XA + f(X)C + BB^T = 0,$$

where the real valued function f can be either a linear or a nonlinear function of X; a typical linear example is $f(X) = \mathrm{trace}(GX)$ for some matrix G [32]. This problem corresponds to (1.1) with $\Phi(X) := f(X)I$ and $C = UU^T$. For linear f, it was shown in [32] that the solution X can be obtained in closed form as follows. Let M be the solution to $AX + XA + BB^T = 0$ and N be the solution to $AX + XA + C = 0$. Then, for $1 - f(N) \neq 0$,

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

$$X = M + \sigma N, \qquad \sigma = \frac{f(M)}{1 - f(N)}. \tag{2.5}$$

The result can be generalized to more terms $f_i(X)C_i$, $i = 1 \dots, s$.

Similarly to the Sherman-Morrison-Woodbury formula of section 2.1, the solution is split into the sum of solutions to distinct Lyapunov equations. This is not surprising, since for $f(X) = v^T X v$ and $C = uu^T$, the two problems are equivalent [32].

2.3. Quadratic-bilinear matrix equations

A quite significant generalization of the linear equation in the previous section includes a special quadratic term in the unknown matrix. More precisely, the problem can be written as[1]

$$AX + XA + \mathbb{H}^T(X \otimes X)\mathbb{H} + \sum_{j=1}^{\ell} N_j X N_j^T + BB^T = 0, \tag{2.6}$$

with $\mathbb{H} \in \mathbb{R}^{n^2 \times n}$.

This quadratic matrix equation arises in the stability analysis of quadratic-bilinear dynamical systems. The real symmetric and positive definite reachability matrix of the dynamical system associated with $(A, B, N_j, \mathbb{H})$, written as a Volterra time series, is also a solution to (2.6), see [7, Th.1]. Within this context, it is thus natural to seek solutions to (2.6) that are real symmetric and possibly positive semidefinite, although non-hermitian solutions may exist. The existence of a solution can be ensured using classical fixed point iteration arguments under certain hypotheses, see, e.g., [7]. Extra care needs to be taken from the control perspective, in case there are more than one positive semidefinite solution to the equation [7, Remark 1].

In the following, we show that if $\mathbb{H}$ has low rank, then the whole problem can be written as (1.1) with a simplified structure. This formulation allows us to study the properties of the solution matrix. In passing, by exploiting this formulation we also obtain new explicit formulas for the solution, under certain hypotheses on the data. To this end, since the novelty of the equation resides in the quadratic term, in the following we assume that $N_j = 0$.

The hypothesis of $\mathbb{H}$ low-rank may seem restrictive. For instance, the application studied in our reference problem [7] does not assume that $\mathbb{H}$ has low rank. Nonetheless, a simplified model where $\mathbb{H}$ is approximated by its leading directions may provide insightful solutions, to which our setting applies. In addition, other settings in control applications dwell with rank deficient $\mathbb{H}$, see, e.g., [26].

[1] We state the problem for symmetric A for consistency with respect to our setting, however the whole derivation holds in the nonsymmetric case as well.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

We first notice that for each column $\boldsymbol{h}_j$ of $\mathbb{H}$, it holds that

$$\boldsymbol{h}_i^T(X \otimes X)\boldsymbol{h}_j = \operatorname{trace}(XH_jX^TH_i^T), \qquad \text{with } H_i \text{ s.t. } \boldsymbol{h}_i = \operatorname{vec}(H_i).$$

A matrix $\mathbb{H}$ of low rank carries convenient features. Setting $\mathbb{H} = \boldsymbol{V}U^T$ with $U \in \mathbb{R}^{n\times r}$, $\boldsymbol{V} \in \mathbb{R}^{n^2\times r}$, we have

$$\mathbb{H}^T(X \otimes X)\mathbb{H} = U\boldsymbol{V}^T(X \otimes X)\boldsymbol{V}U^T,$$

which corresponds to the term $U\Phi(X)U^T$ in (1.1) with $\Phi(X) = \boldsymbol{V}^T(X \otimes X)\boldsymbol{V}$. Using these relations, and assuming first that $\mathbb{H}$ has rank equal to one (that is, $r = 1$), the following proposition provides a new closed form for the sought after solutions.

Proposition 2.2. *If $\mathbb{H} = \boldsymbol{v}u^T$ has rank equal to one, with $\boldsymbol{v} \in \mathbb{R}^{n^2}$ and $u \in \mathbb{R}^n$, then (2.6) can be written as*

$$AX + XA + \chi(X)uu^T + BB^T = 0, \tag{2.7}$$

with $\chi(X) = \operatorname{trace}(XSX^TS^T)$, $\boldsymbol{v} = \operatorname{vec}(S)$. Let M, N be solutions to $AX+XA+BB^T = 0$ and $AX + XA + uu^T = 0$, respectively. If $\alpha := \operatorname{trace}(NSNS^T) \neq 0$, then equation (2.7) admits the following two solutions,

$$X_1 = M + \chi_1 N, \qquad X_2 = M + \chi_2 N,$$

where $\chi_{1,2} = \frac{1}{2\alpha}(-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma})$, with $\beta = \operatorname{trace}(MSNS^T + NSMS^T) - 1$ and $\gamma = \operatorname{trace}(MSMS^T)$.

Before we proceed with the proof, we notice that the way we have expressed the closed form solution, yields information on the existence of *real* solutions, which are obtained only for real $\chi_{1,2}$. Moreover, for real χ both solutions are symmetric, otherwise the solutions will be complex. For real $\chi_{1,2}$, definiteness can be easily detected once the sign of $\mathcal{L}$ is known, by associating the corresponding choice of $\chi_{1,2}$.

Proof. We have

$$\mathbb{H}^T(X \otimes X)\mathbb{H} = u\boldsymbol{v}^T(X \otimes X)\boldsymbol{v}u^T = u\operatorname{trace}(XSX^TS^T)u^T = u\chi(X)u^T.$$

Hence, the original problem can be written as in (2.7). The problem is now in the form discussed in section 2.2, with χ nonlinear, and we can proceed analogously. More precisely, we first notice that

$$X = M + \chi(X)N,$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

with X symmetric. We multiply by S and by S^T, and then multiply the resulting matrices so as to obtain

$$\begin{aligned}(XS)(XS^T) &= (MS + \chi(X)NS)(MS^T + \chi(X)NS^T) \\ &= MSMS^T + \chi(X)MSNS^T + \chi(X)NSMS^T + \chi(X)^2 NSNS^T.\end{aligned}$$

Taking the trace,

$$\chi(X) = \mathrm{trace}(MSMS^T) + \chi(X)\mathrm{trace}(MSNS^T + NSMS^T) + \chi(X)^2\mathrm{trace}(NSNS^T).$$

Setting α, β and γ like in the statement, we obtain the quadratic algebraic equation $\alpha\chi^2 + \beta\chi + \gamma = 0$, whose solutions are the given $\chi_{1,2}$. □

The generalization to $\mathbb{H}$ of (small) rank greater than one gives the more general form in (1.1). We write $\mathbb{H} = \boldsymbol{V}U^T$ with $\boldsymbol{V} = [\boldsymbol{v}_1, \dots \boldsymbol{v}_r]$ and $U = [u_1, \dots, u_r]$. Then

$$\mathbb{H}^T(X \otimes X)\mathbb{H} = U\boldsymbol{V}^T(X \otimes X)\boldsymbol{V}U^T = U\Phi(X)U^T,$$

where the entries of the $r \times r$ matrix $\Phi(X)$ are given by

$$(\Phi(X))_{i,j} = \mathrm{trace}(XS_jX^TS_i^T), \quad \mathrm{vec}(S_k) = \boldsymbol{v}_k, \quad k = i, j,$$

so that Φ is nonlinear in X. Thus, we can write

$$AX + XA^T + U\Phi(X)U^T + BB^T = 0,$$

and proceed like in section 2.1. This approach leads to a feasible computational strategy as long as U has few columns.

We conclude this section by saying that disregarding the rank of $\mathbb{H}$, a general closed form solution can be described by means of an integral series, though this expression does not seem to lead to a computationally feasible strategy [7].

3. Sequences of approximations

Except when a computable closed form can be obtained, an approximation to the solution of (1.1) is derived by generating a sequence $\{X^{(k)}\}_{k\geq 0}$ approaching X as k goes to infinity. In particular, whenever the number of columns r is sizable, say a few tens, also the Sherman-Morrison-Woodbury formula for the linear case $\Phi(X) = V^TXV$ becomes infeasible, and an approximation recurrence is required.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

3.1. Fixed point iterations

A classical approach is given by a fixed point iteration. For the generic case (1.1), this can be written as

$$AX^{(k+1)} + X^{(k+1)}A = -U\Phi(X^{(k)})U^T - BB^T,$$

leading to a sequence of solves with Lyapunov equations, that is,

$$X^{(k+1)} = \mathcal{L}^{-1}(-U\Phi(X^{(k)})U^T - BB^T);$$

see, e.g., [14][36]. Computationally, the problem then becomes that of efficiently solving a sequence of Lyapunov equations, which can be a challenge for large problems, especially considering that the Lyapunov equations cannot be solved at high accuracy. This procedure gives rise to inexact iterates, for which the convergence analysis needs to be revisited, and associated with the accuracy threshold of the Lyapunov solver; we refer to [36] for this discussion. The idea is very general, so that the same approach can be used in a number of linear and nonlinear matrix equations, such as the algebraic Riccati equation [10]; see [23] for other examples related to the approximation of matrix functions. Ad-hoc iterations can be derived for specific problems, see below.

Fixed point iterations have also been proposed in [7] for solving the Quadratic-bilinear equation in section 2.3 in the large scale setting. More precisely, in [7, (41)] the following recurrence is proposed,

$$\mathcal{L}(X^{(k+1)}) = -BB^T - \mathbb{H}^T(X^{(k)} \otimes X^{(k)})\mathbb{H} - \sum_{j=1}^{\ell} U_j \Phi_j(X^{(k)}) U_j^T.$$

This general iteration is very relevant for $\mathbb{H}$ having large or column full-rank. On the other hand, for $\mathbb{H}$ having low rank, a possibly more effective fixed point iteration can be derived. To simplify the presentation, we set $\Phi_j(X^{(k)}) = 0$ for all j and focus on the quadratic term, which makes the problem special.

If $\mathbb{H}$ has rank equal to one, Proposition 2.2 ensures a closed form solution. If $\mathbb{H}$ is low-rank, that is $\mathbb{H} = \boldsymbol{V}U^T$ with $U = [u_1, \ldots, u_r] \in \mathbb{R}^{n\times r}$ and $\boldsymbol{V} = [v_1, \ldots, v_r] \in \mathbb{R}^{n\times r}$, then the original equation can be written as

$$AX + XA^T + \sum_{i,j=1}^{r} u_i \Phi(X)_{i,j} u_j^T + BB^T = 0, \quad \text{with} \quad \Phi(X)_{i,j} = \boldsymbol{v}_i^T (X \otimes X)\boldsymbol{v}_j. \quad (3.1)$$

We recall that $\Phi(X)_{i,j} = \mathrm{trace}(XV_jX^TV_i^T)$, where $\boldsymbol{v}_i = \mathrm{vec}(V_i)$. We can thus define a linearized fixed point iteration as follows. We replace $\Phi(X)$ with $\Psi_{i,j}(X^{(k)}, X) = \mathrm{trace}(XV_j(X^{(k)})^TV_i^T)$. For a computed $X^{(k)}$, we set $K_{i,j}^{(k)} := V_j(X^{(k)})^TV_i^T$ so that

$$\Psi_{i,j}(X^{(k)}, X) = \mathrm{trace}(XK_{i,j}^{(k)}),$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

which is now linear in X. The iteration (3.1) is transformed into the quasi-linear problem

$$\mathcal{L}(X^{(k+1)}) = -BB^T - \sum_{i,j=1}^{r} \mathrm{trace}(X^{(k+1)} K_{i,j}^{(k)}) u_i u_j^T, \tag{3.2}$$

in the unknown matrix $X^{(k+1)}$. We can obtain $X^{(k+1)}$ by a procedure similar to that recalled in section 2.2. Indeed, let $P_{\{i,j\}} = \mathcal{L}^{-1}(u_i u_j^T) \in \mathbb{R}^{n\times n}$, for $i,j = 1,\dots,r$, and let M solve $AX + XA + BB^T = 0$. Then

$$X + \sum_{i,j=1}^{r} \omega_{i,j}^{(k)} P_{\{i,j\}} = M, \qquad \omega_{i,j}^{(k)} = \mathrm{trace}(X K_{i,j}^{(k)}).$$

For each $\ell, s = 1,\dots,r$, multiply the equation by $K_{\ell,s}^{(k)}$ and take the trace, so as to obtain

$$\mathrm{trace}(X K_{\ell,s}^{(k)}) + \sum_{i,j}^{r} \omega_{i,j}^{(k)} \mathrm{trace}(P_{\{i,j\}} K_{\ell,s}^{(k)}) = \mathrm{trace}(M K_{\ell,s}^{(k)}),$$

that is,

$$\omega_{\ell,s}^{(k)} + \sum_{i,j=1,\dots,r} \omega_{i,j}^{(k)} \mathrm{trace}(P_{\{i,j\}} K_{\ell,s}^{(k)}) = \mathrm{trace}(M K_{\ell,s}^{(k)}).$$

Denote by $\tau_{i,j}^{(\ell,s)} = \mathrm{trace}(P_{\{i,j\}} K_{\ell,s}^{(k)})$ the known coefficients, and let

$$\omega^{(k)} = \mathrm{vec}\left(\mathrm{trace}(X K_{\ell,s}^{(k)})_{\ell,s=1,\dots r}\right), \quad t^{(k)} = \mathrm{vec}\left(\mathrm{trace}(M K_{\ell,s}^{(k)})_{\ell,s=1,\dots r}\right)$$

be the unknown traces and right-hand side vector of length r^2, respectively. This yields the linear system

$$(I_{r^2} + T)\omega^{(k)} = t^{(k)}, \qquad T = \begin{bmatrix} \tau_{1,1}^{1,1} & \tau_{2,1}^{1,1} & \cdots & \tau_{1,2}^{1,1} & \cdots & \tau_{r,r}^{1,1} \\ \tau_{1,1}^{2,1} & \tau_{2,1}^{2,1} & \cdots & \tau_{1,2}^{2,1} & \cdots & \tau_{r,r}^{2,1} \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \tau_{1,1}^{1,2} & \tau_{2,1}^{1,2} & \ddots & \ddots & \ddots & \tau_{r,r}^{1,2} \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \tau_{1,1}^{r,r} & \tau_{2,1}^{r,r} & \cdots & \tau_{1,2}^{r,r} & \cdots & \tau_{r,r}^{r,r} \end{bmatrix}.$$

Let $\Omega^{(k)}$ be the $r \times r$ matrix reshaping of the solution $\omega^{(k)}$. Then the next iterate $X^{(k+1)}$ is obtained as

$$X^{(k+1)} = M - \sum_{i,j=1}^{r} (\Omega^{(k)})_{i,j} P_{\{i,j\}}. \tag{3.3}$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

Table 1
Example 3.1. Performance of standard fixed point iteration and quasi-linear variant. Time is in seconds. '–' stands for divergence.

	fixed point		quasi-linear	
α	# its	CPU time	# its	CPU time
4	17	18.88	10	35.00
5	42	36.40	13	43.64
5.5	188	147.13	19	60.64
6	–	–	88	309.27

As a computational side remark, we notice that only $\Omega^{(k)}$ changes at each iteration, and that all other matrices could be computed once for all before starting the recurrence.

Example 3.1. In this example we report on our experience with the iteration (3.3), compared with the standard procedure in (3.1). To this end we consider a built-up, fully reproducible example, where we make the problem less definite by changing a problem parameter. The matrix A is given as $A = I \otimes T + T \otimes I$ of size $n \times n$ with $n = 2500$, where $T = n_h^2 \text{tridiag}(1, -2, 1)$, with $n_h^2 = n$. The nonlinear term includes $\mathbb{H} = \alpha \boldsymbol{V} U^T$ with both $U, \boldsymbol{V}$ chosen randomly with $r = 4$ columns, and α controls the definiteness of the operator $\mathcal{M}$. The known term $E = BB^T$ is a random positive definite matrix. These definitions correspond to the following matlab ([29]) commands

```
alpha=6;
nh=40; n=nh*nh; r=4;
T=toeplitz([-2,1,zeros(1,nh-2)]); I=speye(nh);
A=n*(kron(I,T)+kron(T,I));
rng(1);
E=rand(n,n);E=E*E';E=E/norm(E);
V=alpha*randn(n*n,r);
U=randn(n,r);
```

The stopping criterion used for this particular experiment was

$$\frac{\|X^{(k+1)} - X^{(k)}\|_F}{\|BB^T\|_F} < 10^{-6},$$

and a maximum of 200 iterations was considered.[2] At completion the true residual norm was also computed.

Table 1 reports the performance of both iterations for $\alpha \in [4, 6]$. For small α, the standard iteration (3.1) is faster in terms of CPU time, whereas the new approach is able to get good information on the solution in fewer iterations. For the second value

[2] For more robust software-oriented criteria, the one above should be accompanied by typical criteria for nonlinear problems, ensuring true convergence, and not stagnation.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

of α, the gap in the number of iterations increases, although the CPU time of the new iteration remains larger. After that, the standard fixed point iteration starts showing convergence difficulties, until a blow up for $\alpha = 6$, while good properties of (3.1) can still be observed. Although not reported here, we notice that up to $\alpha = 5.5$ convergence of the iteration in (3.3) is monotonic, in terms of $\|X^{(k+1)} - X^{(k)}\|$ whereas for $\alpha = 6$ oscillations occur. Larger values of α than those considered here will eventually lead to lack of convergence also for the new recurrence.

We consider these results very promising. However, assessing the overall quality of the new method requires a deeper investigation, which is beyond the scope of this work. ⋄

The previous examples all generate matrix sequences $\{X^{(k)}\}_{k\geq 0}$ such that

$$AX^{(k+1)} + X^{(k+1)}A = -BB^T - \sum_{i,j} u_i w_{i,j}^{(k)} u_j^T$$

for some $w_{i,j}^{(k)}$, so that

$$X^{(k+1)} = -\mathcal{L}^{-1}(BB^T) - \sum_{i,j} w_{i,j}^{(k)} \mathcal{L}^{-1}(u_i u_j^T). \tag{3.4}$$

We stress once again that pairs of solutions of $\mathcal{L}^{-1}(u_i u_j^T)$ are the transpose of each other, and that overall the sum is symmetric; see the similar discussion in section 2.1.

The solution matrix is a linear combination of solutions to distinct Lyapunov equations, where the first term is the solution one would obtain with $\Phi \equiv 0$. Hence, inspecting the magnitude of the coefficients $w_{i,j}^{(k)}$ enables one to appreciate the influence on the solution of the extra term in (1.1) compared with the plain Lyapunov problem.

3.2. Projection methods

An alternative to fixed point iterations is given by projection-type methods for the original equation (1.1), which have been mostly explored in the case of linear matrix equations, that is for $\Phi(X) = V^T XV$. Given an approximation space of dimension s_k and a matrix W_k with orthonormal columns with the same range, an approximate solution is obtained as $X^{(k)} = W_k Y^{(k)} W_k^T$, where $Y^{(k)} \in \mathbb{R}^{s_k \times s_k}$ is determined, e.g., by imposing an orthogonality (Galerkin) condition to the residual $R^{(k)} = \mathcal{M}(X^{(k)}) + BB^T$ [39]. Such condition can be formalized as requiring that $W_k^T R^{(k)} W_k = 0$. It can be shown, see [30] and its references, that for linear and symmetric positive definite $\mathcal{M}$ in (1.1), the approximate solution $X^{(k)}$ obtained with the Galerkin condition minimizes the error in the energy-norm, that is

$$X^{(k)} = \arg \min_{\substack{Z = W_k \tau W_k^T \\ \tau \in \mathbb{R}^{s_k \times s_k}}} \|X - Z\|_{\mathcal{M}}. \tag{3.5}$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

This result will be used in the next section. In [30] a two-term-type recurrence was also derived for this approach, so that the approximation can be written as

$$X^{(k+1)} = X^{(k)} + P_k \alpha^{(k)} P_k^T.$$

This solution update generalizes the well known two-term recurrence of the Conjugate Gradient method for vector linear systems. In spite of the similarity with respect to the previous recurrences, here P_k has low rank, which grows at each iteration, and the dimensions of the *matrix* α_k grow accordingly. As opposed to (stationary) fixed point iterations, this matrix update is nonstationary, since α_k and P_k are chosen at each iteration so as to satisfy certain optimality properties; see [30] for more details.

For nonlinear problems such as (2.7) in the large scale regime, projection-type methods have also been developed, in the context of model order reduction, see, e.g., [18], [5]. A major challenge for these approaches on the quadratic-bilinear problem is the selection of the approximation space, which needs to include information both on A and N_j, but also account in some way for the quadratic term; we refer to [5], [12] for a special selection that enjoys promising interpolation properties. There are two major difficulties in using projection methods: first, lack of theoretical ground for a chosen subspace selection as equation solver; second, lack of explicit or direct methods for the reduced problem after projection, which has the same quadratic structure and similar though less dramatic memory limitations due to the Kronecker term. The described situation is different from what occurs for other quadratic problems such as the Riccati equations, for which the approximation space can be built from the linear terms, and the reduced problem can be reliably solved by means of explicit decomposition methods; see, e.g., [9], [25],[22], [8], [38] and references therein. Here, the reduced problem still possesses all difficulties of the original problems except the reduced size. Hence, devising reliable small scale solvers is paramount to be able to address the large scale case with projection methods.

4. Monotonicity results

In this section we are concerned with monotonicity properties of the solution matrix or of its approximation iterates. These may be expressed in terms of definiteness, or in terms of some norm. We mainly dwell with Φ linear, hence the equation's solution is assumed to be unique, that is, the equation in Kronecker form is solvable [39].

If A is symmetric and negative definite, we have that the Lyapunov solution $X_{\mathcal{L}} = \mathcal{L}^{-1}(-BB^T)$ is positive semi-definite, and that the solution X to (1.1) satisfies

$$X \succeq X_{\mathcal{L}} \quad \text{if and only if} \quad \Phi(X) \succeq 0.$$

Positive definiteness can be checked in all instances where we have derived a symmetric closed form of the solution and it readily depends on the involved scalars (see, e.g., (2.5)), whereas the property $X \succeq X_{\mathcal{L}}$ needs to be derived case by case.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

For the sequences $\{X^{(k)}\}_{k\geq 0}$ derived in the previous sections we next prove a new, corresponding result.

Proposition 4.1. *Let the problem $AX + XA + U\Phi(X)U^T + BB^T = 0$ be given. In the iteration (3.4), assume that Φ is linear and such that $\Phi(Y) \succeq 0$ for all $Y \succeq 0$. Assume that $X^{(0)} = 0$. Then,*

(i) *If $\mathcal{L}$ is negative definite, then for any $k \geq 0$, it holds that $X^{(k+1)} \succeq X^{(k)} \succeq 0$;*
(ii) *If $\mathcal{L}$ is positive definite, the semidefiniteness alternates between successive iterates.*

Proof. We notice that $\mathcal{L}(X^{(k+1)}) = -BB^T - U\Phi(X^{(k)})U^T$, so that

$$X^{(k+1)} - X^{(k)} = -\mathcal{L}^{-1}(U\Phi(X^{(k)} - X^{(k-1)})U^T).$$

Hence, the definiteness sign of $X^{(k+1)} - X^{(k)}$ depends on that of $\mathcal{L}$ and on the sign of the approximate solutions difference at the previous step k. i) Taking $X^{(0)} = 0$ ensures $X^{(1)} \succeq 0$, from which all subsequent terms satisfy $X^{(k+1)} - X^{(k)} \succeq 0$. ii) Taking $X^{(0)} = 0$ ensures $0 \succeq X^{(1)}$ and from there on, the signs alternate. □

For given symmetric matrices Z_1, Z_2 such that $Z_1 \succeq Z_2$ it holds that $\lambda_i(Z_1) \geq \lambda_i(Z_2)$ where $\lambda_i(\cdot)$ are the increasingly ordered eigenvalues of the argument matrix [24, Corollary 7.7.4]. Hence, the result of Proposition 4.1(i) implies the following norm inequalities, under the same hypotheses

$$\|X^{(k+1)}\| \geq \|X^{(k)}\|, \qquad \|X^{(k+1)}\|_F \geq \|X^{(k)}\|_F.$$

In the linear setting, by just reformulating classical results, we can explicitly write a strictly decreasing bound for the error. We consider the linear case $AX + XA + UV^TXVU^T + BB^T = 0$, which in Kronecker form corresponds to

$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = 0 \quad \text{with} \quad \boldsymbol{A} = \boldsymbol{L} + \boldsymbol{N},$$

where $\boldsymbol{L} = A \otimes I + I \otimes A$ and $\boldsymbol{N} = (U \otimes U)(V \otimes V)^T$. Let $\boldsymbol{x}^*$ be the exact solution to the linear system above. Given the recurrence $\boldsymbol{L}\boldsymbol{x}^{(k+1)} = -\boldsymbol{b} - \boldsymbol{N}\boldsymbol{x}^{(k)}$, the following result provides a bound for the error $\boldsymbol{e}^{(k)} = \boldsymbol{x}^* - \boldsymbol{x}^{(k)}$ for a stationary iteration. We include the proof for completeness, though it may be derived similarly to other proofs in the literature[3] [33, Corollary 4.1].

Proposition 4.2. *Let $\boldsymbol{A} = \boldsymbol{L} + \boldsymbol{N}$ with both $\boldsymbol{A}$ and $\boldsymbol{L}$ symmetric and positive definite. If $\rho(\boldsymbol{L}^{-1}\boldsymbol{A}) < 2$, then*

[3] We thank Silvia Noschese for pointing us to [33] for this type of results.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

$$\|\boldsymbol{e}^{(k+1)}\|_{\boldsymbol{A}} \le \rho(I - \boldsymbol{L}^{-1}\boldsymbol{A})\|\boldsymbol{e}^{(k)}\|_{\boldsymbol{A}}.$$

Proof. Let $\boldsymbol{B} = \boldsymbol{L}^{-1}\boldsymbol{N} = I - \boldsymbol{L}^{-1}\boldsymbol{A}$. We first observe that $\widehat{\boldsymbol{B}} := \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{-\frac{1}{2}} = I - \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{L}^{-1}\boldsymbol{A}^{\frac{1}{2}}$ is symmetric and that thanks to the similarity transformations, $\rho(\widehat{\boldsymbol{B}}) = \rho(\boldsymbol{B})$. Hence, using $\|\boldsymbol{e}^{(k+1)}\|_{\boldsymbol{A}} = \|\boldsymbol{B}\boldsymbol{e}^{(k)}\|_{\boldsymbol{A}}$, we have

$$\begin{aligned}\|\boldsymbol{e}^{(k+1)}\|_{\boldsymbol{A}} &= \|\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{e}^{(k)}\| = \|\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{e}^{(k+1)}\| \\ &\le \|\widehat{\boldsymbol{B}}\|\,\|\boldsymbol{e}^{(k+1)}\|_{\boldsymbol{A}} = \rho(\boldsymbol{B})\|e^{(k)}\|_{\boldsymbol{A}}. \quad \square\end{aligned}$$

In our setting, $\boldsymbol{L}$ corresponds to $\mathcal{L}$, and the result above reads

$$\|E^{(k+1)}\|_{\mathcal{M}} \le \rho(I - \mathcal{L}^{-1}(\mathcal{M}(\cdot)))\|E^{(k)}\|_{\mathcal{M}},$$

where $\mathcal{M} : X \mapsto AX + XA + U\Phi(X)U^T$ is symmetric and positive definite.

Projection methods satisfy similar inequalities, though stemming from quite different properties. More precisely, the minimization property in (3.5) ensures monotonicity of the approximate solution obtained via a projection method. The following new result makes this statement more precise, under the only hypothesis that the linear operator $\mathcal{M}$ be symmetric and positive definite.

Proposition 4.3. *Let $\mathcal{M}$ be symmetric and positive definite. Then the Galerkin approximate solution $X^{(k)} = W_k Y^{(k)} W_k^T$ to (1.1) converges monotonically increasing towards X in the $\mathcal{M}$-norm, that is*

$$\|X^{(k+1)}\|_{\mathcal{M}} \ge \|X^{(k)}\|_{\mathcal{M}}.$$

Proof. Let $E^{(k)} = X - X^{(k)}$ and $R^{(k)} = \mathcal{M}(X^{(k)}) + BB^T$. Observe that $R^{(k)} = -\mathcal{M}(E^{(k)})$ and that $\mathrm{trace}(\mathcal{M}(Y)X) = \mathrm{trace}(Y\mathcal{M}(X))$ for symmetric matrices X, Y of conforming dimensions. From (3.5) we have

$$\|E^{(k+1)}\|_{\mathcal{M}} \le \|E^{(k)}\|_{\mathcal{M}}. \tag{4.1}$$

We have $\|E^{(k)}\|^2_{\mathcal{M}} = \mathrm{trace}(E^{(k)}\mathcal{M}(E^{(k)})) = -\mathrm{trace}(E^{(k)}R^{(k)})$. The Galerkin condition on the residual ensures that $\mathrm{trace}(X^{(k)}R^{(k)}) = 0$, since $\mathrm{trace}(X^{(k)}R^{(k)}) = \mathrm{trace}(Y^{(k)}W_k^T R^{(k)} W_k)$ and $W_k^T R^{(k)} W_k = 0$. Then,

$$\begin{aligned}\mathrm{trace}(E^{(k)}R^{(k)}) &= \mathrm{trace}(XR^{(k)}) = -\mathrm{trace}(\mathcal{M}(X)E^{(k)}) \\ &= -\mathrm{trace}(\mathcal{M}(X)X) + \mathrm{trace}(\mathcal{M}(X)X^{(k)}).\end{aligned}$$

We recall that $\mathrm{trace}(\mathcal{M}(X)X^{(k)}) = \mathrm{trace}(-BB^T X^{(k)}) = \mathrm{trace}(\mathcal{M}(X^{(k)})X^{(k)})$, where once again we used that $\mathrm{trace}(X^{(k)}R^{(k)}) = 0$. Hence,

$$\|E^{(k)}\|^2_{\mathcal{M}} = -\mathrm{trace}(E^{(k)}R^{(k)}) = \mathrm{trace}(\mathcal{M}(X)X) - \mathrm{trace}(\mathcal{M}(X^{(k)})X^{(k)}).$$

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

The result follows from substituting into the inequality (4.1). □

This property has applications in different directions, for instance in approximating the $\mathcal{H}$-energy norm of a linear dynamical system, see, e.g., [1].

5. Decay properties of the solution's singular values

For BB^T of very low rank p, it is known that the singular values of the solution to the Lyapunov equation $AX + XA + BB^T = 0$ exhibit an exponentially decaying pattern, which depends on the spectral properties of A and to a lesser extent on the right-hand side matrix B. In addition to being very helpful when questing for low-rank approximate solutions, the topic is fascinating and has attracted a lot of attention, especially in the case of non-normal A, see, e.g., [31], [3], [40], [2], [27], [35], [17].

Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ be the singular values of the solution matrix X, and recall that (Eckart-Young-Mirsky theorem)

$$\frac{\sigma_k}{\sigma_1} = \min_{\mathrm{rank}(Y)=k} \frac{\|X - Y\|}{\|X\|}.$$

As a consequence of this property, for any rank-k approximation $X^{(k)}$ to X, any bound for the error $\|X - X^{(k)}\|$ can be used as an upper estimate for the decay of the kth singular value. This has been a common strategy in the literature cited above. A very realistic estimate stemming from rational function approximations of X was derived in [28],

$$\lambda_{pk+1}(X) \leq \frac{16\|B\|_2^2}{\lambda_{\min}(A)} \exp\left(-\frac{k\pi^2}{\log(8\kappa(A))}\right), \quad 1 \leq pk < n, \tag{5.1}$$

where $\lambda_{\min}(A), \lambda_{\max}(A)$ are the extreme eigenvalues of the symmetric and positive definite matrix A, and $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ is the condition number of A; a similar estimate explicitly using elliptic functions was earlier established in [35, sec.2.1.2].

The question thus arises whether the decaying properties carry over to the general form (1.1). The answer is in the affirmative, despite the extra term $U\Phi(X)U^T$. In fact, the main role is played by the rank of $[B, U]$. This can be readily seen by writing (1.1) as $AX + XA = -BB^T - U\Phi(X)U^T$, with

$$BB^T + U\Phi(X)U^T = [B, U] \begin{bmatrix} I & \\ & \Phi(X) \end{bmatrix} [B, U]^T. \tag{5.2}$$

Hence, independently of the action of Φ on X, the solution X has a worst possible decay completely described by A, $[B, U]$, and the rank of $[B, U]$. The decay can be faster in case $\Phi(X)$ is rank deficient. More precisely, to generalize (5.1) to the new setting, we use (5.2) to obtain

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

$$\frac{\lambda_{pk+1}(X)}{1+\|\Phi(X)\|} \le \frac{16\|[B,U]\|_2^2}{\lambda_{\min}(A)} \exp\left(-\frac{k\pi^2}{\log(8\kappa(A))}\right), \quad 1 \le pk < n, \tag{5.3}$$

where p is the rank of $[B,U]$.

We refer to [41] for bounds that explicitly take into account the factor $[B,U]$, though their relevance can be better appreciated for A non-symmetric. Interestingly, the formalization via $[B,U]$ was used in the proof of Theorem 4.1 in [4] to derive a closed form approximation to (1.1) as a sum of exponentials.

In the previous sections we gave a solution in closed form for some of the considered problems. This makes the decay analysis more precise, while avoiding explicit reference to $\Phi(X)$. Indeed, in all cases, we have

$$X = M + \sum_i \gamma_i P_i,$$

where M and all P_i are solutions to Lyapunov equations, hence they inherit the corresponding decay pattern. As an example, consider the solution $X_1 = M + \chi_1 N$ as derived in Proposition 2.2. A completely analogous result can be obtained with the procedure described in section 2.2, see also the example below. The matrices M and N are solutions to Lyapunov equations with the same coefficient matrix, that is $M = \mathcal{L}^{-1}(-BB^T)$ and $N = \mathcal{L}^{-1}(-uu^T)$. Then by linearity,

$$X_1 = \mathcal{L}^{-1}(-BB^T - \chi_1 uu^T),$$

so that

$$\lambda_{pk+1}(X_1) \le \frac{16\|Y\|_2}{\lambda_{\min}(A)} \exp\left(-\frac{k\pi^2}{\log(8\kappa(A))}\right), \quad 1 \le pk < n, \tag{5.4}$$

where $Y = [B,U]\mathrm{blkdiag}(I, \chi_1 I)[B,U]^T$, and p is its rank.

Example 5.1. We consider the problem $AX + XA + uv^T Xvu^T + BB^T = 0$ with A diagonal with diagonal entries logarithmically distributed in the interval $[10^{-10}, 1]$ and dimension $n = 1000$, u a vector with normally distributed random components (Matlab seed `rng=1`), $v = e_1$, and $B = \mathbf{1}$. In Fig. 1 we report the distribution of the first few tens singular values of the solutions $Y_1 = -\mathcal{L}^{-1}(BB^T)$, $Y_2 = -\mathcal{L}^{-1}(uu^T)$, together with that of $X = Y_1 + \sigma Y_2$, obtained with the procedure described in section 2.2. Moreover, the normalized estimate in (5.4) is also shown. The estimate is quite descriptive of the true decay for $p = 2$, since $[B,u]$ is full rank. The decay of the two solutions Y_1, Y_2 is very similar, and since $\sigma = \mathcal{O}(1)$, Y_1 and Y_2 equally contribute to the spectral decay of X. ⋄

6. Generalizations and outlook

Many of the results we have stated can be generalized to the nonlinear non-symmetric case

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

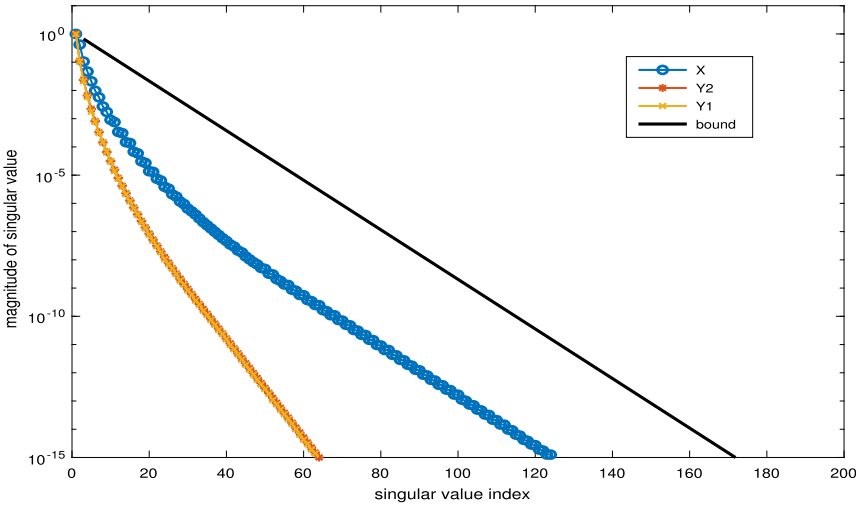

Fig. 1. Example 5.1. Singular values of the solutions Y_1, Y_2 and $X = Y_1 + \sigma Y_2$, and the estimate in (5.1).

$$A_1 X + X A_2 + \sum_{i=1}^{\ell} U_i \Phi_i(X) U_i^T + B_1 B_2^T = 0, \tag{6.1}$$

as long as the associated coefficient operator is symmetric and positive definite. This setting may have broader applications, such as in the discretization of time- and parameter-dependent partial differential equations and in the numerical treatment of problems dealing with uncertainty in the data; see, e.g., [39].

The quadratic-bilinear matrix equation is a new algebraic problem largely unexplored. We have proposed a new iterative method for its solution, which under low-rank assumption of the matrix $\mathbb{H}$, seems to give promising computational results with respect to the only (very recent) attempt in the literature. We believe there is plenty of room for improvements, both in the theoretical understanding of the algebraic equation, and in the development of new effective methods.

We have shown that writing the solution to (1.1) as a sum of terms, allows us to devise a simple indicator of the effect on the solution of the extra possibly nonlinear addend in the equation, that is not associated with the Lyapunov equation. Whenever explicitly available, easily characterized positive definiteness and monotonicity properties of the approximate solutions can help calibrate the approximation phase, thus monitoring computational costs and dynamic rank truncations.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Valeria Simoncini reports travel was provided by Ministry of Education and Merit (20227PCCKZ – CUP J53D23003620006). If there are other authors, they declare that they have no known competing financial inter-

ests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Peter Benner for a discussion on the problem analyzed in [7].

The author is a member of the INdAM Research Group GNCS. Moreover, her work was partially supported by the European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice PRIN 2022 - DD N. 104 of 2/2/2022, entitled "Low-rank Structures and Numerical Methods in Matrix and Tensor Computations and their Application", code 20227PCCKZ – CUP J53D23003620006.

Data availability

No data was used for the research described in the article.

References

[1] A.C. Antoulas, Approximation of Large-Scale Dynamical Systems, Advances in Design and Control, SIAM, Philadelphia, 2005.
[2] A.C. Antoulas, D.C. Sorensen, Y. Zhou, On the decay rate of Hankel singular values and related issues, Syst. Control Lett. 46 (2002) 323–342.
[3] J. Baker, M. Embree, J. Sabino, Fast singular value decay for Lyapunov solutions with nonnormal coefficients, SIAM J. Matrix Anal. Appl. 36 (2015) 656–668.
[4] P. Benner, T. Breiten, Low rank methods for a class of generalized Lyapunov equations and related issues, Numer. Math. 124 (2013) 441–470.
[5] P. Benner, T. Breiten, Two-sided projection methods for nonlinear model order reduction, SIAM J. Sci. Comput. 37 (2015) B239–B260.
[6] P. Benner, T. Damm, Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems, SIAM J. Control Optim. 49 (2011) 686–711.
[7] P. Benner, P. Goyal, Balanced truncation for quadratic-bilinear control systems, Adv. Comput. Math. 50 (2024) 88.
[8] P. Benner, J. Saak, Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey, GAMM-Mitt. 36 (2013) 32–52.
[9] D. Bini, B. Iannazzo, B. Meini, Numerical Solution of Algebraic Riccati Equations, SIAM, Philadelphia, 2012.
[10] D.A. Bini, B. Iannazzo, B. Meini, Numerical Solution of Algebraic Riccati Equations, Society for Industrial and Applied Mathematics, 2011.
[11] I. Bioli, D. Kressner, L. Robol, Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations, Tech. Rep. 2408.16416, 2024.
[12] X. Cao, J. Maubach, S. Weiland, W. Schilders, A novel Krylov method for model order reduction of quadratic bilinear systems, in: IEEE Conference on Decision and Control, Miami Beach, FL, USA, Dec. 17–19, 2018, pp. 3217–3222.
[13] R. Choudhary, K. Ahuja, Stability analysis of bilinear iterative rational Krylov algorithm, Linear Algebra Appl. 538 (2018) 56–88.
[14] T. Damm, Direct methods and ADI-preconditioned Krylov subspace methods for generalized Lyapunov equations, Numer. Linear Algebra Appl. 15 (2008) 853–871, Special issue on Matrix equations.
[15] G. Golub, C.F. Van Loan, Matrix Computations, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.
[16] W. Gray, J. Mesko, Energy functions and algebraic Gramians for bilinear systems, IFAC Proc. Vol. 31 (17) (1998) 101–106.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017

[17] L. Grubisić, D. Kressner, On the eigenvalue decay of solutions to operator Lyapunov equations, Syst. Control Lett. 73 (2014) 42–47.
[18] C. Gu, QLMOR: a projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 30 (2011) 1307–1320.
[19] Y. Hao, V. Simoncini, The Sherman-Morrison-Woodbury formula for generalized linear matrix equations and applications, Numer. Linear Algebra Appl. 28 (2021) e2384.
[20] C. Hartmann, A. Zueva, B. Schaefer-Bung, Balanced model reduction of bilinear systems with applications to positive systems, Tech. Rep., Institut fuer Mathematik, Freie Universitaet, Berlin, 2010.
[21] H.V. Henderson, S.R. Searle, On deriving the inverse of a sum of matrices, SIAM Rev. 23 (1981) 53–60.
[22] M. Heyouni, K. Jbilou, An extended block Krylov method for large-scale continuous-time algebraic Riccati equations, Electron. Trans. Numer. Anal. 33 (2008–2009) 53–62.
[23] N.J. Higham, Functions of Matrices - Theory and Computation, SIAM, Philadelphia, USA, 2008.
[24] R.A. Horn, C.R. Johnson, Matrix Analysis, II ed., Cambridge University Press, Cambridge, 2013.
[25] K. Jbilou, Block Krylov subspace methods for large algebraic Riccati equations, Numer. Algorithms 34 (2003) 339–353.
[26] A.E. Kafshgarkolaei, M.S. Hemati, Local stability and stabilization of quadratic-bilinear systems using Petersen's lemma, arXiv:2503.21040, 2025.
[27] N. Komaroff, Simultaneous eigenvalue lower bounds for the Lyapunov matrix equation, IEEE Trans. Autom. Control 33 (1988) 126–128.
[28] D. Kressner, C. Tobler, Krylov subspace methods for linear systems with tensor product structure, SIAM J. Matrix Anal. Appl. 31 (2010) 1688–1714.
[29] The MathWorks, Inc., MATLAB 7, r2020b ed., 2020.
[30] D. Palitta, M. Iannacito, V. Simoncini, A subspace-conjugate gradient method for linear matrix equations, Tech. Rep. 2501.02938, January 2025, Available from http://arxiv.org/abs/2501.02938.
[31] T. Penzl, Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case, Syst. Control Lett. 40 (2000) 139–144.
[32] M. Porcelli, V. Simoncini, Numerical solution of a class of quasi-linear matrix equations, Linear Algebra Appl. 664 (2023) 349–368.
[33] A. Quarteroni, R. Sacco, F. Saleri, Matematica Numerica, Springer, 2008.
[34] E. Ringh, G. Mele, J. Karlsson, E. Jarlebring, Sylvester-based preconditioning for the waveguide eigenvalue problem, Linear Algebra Appl. 542 (2018) 441–463.
[35] J. Sabino, Solution of Large-Scale Lyapunov Equations via the Block Modified Smith Method, PhD thesis, Rice University, 2006.
[36] S.D. Shank, V. Simoncini, D.B. Szyld, Efficient low-rank solutions of generalized Lyapunov equations, Numer. Math. 134 (2016) 327–342.
[37] J. Sherman, W.J. Morrison, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, Ann. Math. Stat. 21 (1950) 124–127.
[38] V. Simoncini, Analysis of the rational Krylov subspace projection method for large-scale algebraic Riccati equations, SIAM J. Matrix Anal. Appl. 37 (2016) 1655–1674.
[39] V. Simoncini, Computational methods for linear matrix equations, SIAM Rev. 58 (2016) 377–441.
[40] D. Sorensen, Y. Zhou, Bounds on eigenvalue decay rates and sensitivity of solutions to Lyapunov equations, Tech. Rep. 02-07, Rice University, Houston, Texas, 2002.
[41] N. Truhar, K. Veselić, Bounds on the trace of a solution to the Lyapunov equation with a general stable matrix, Syst. Control Lett. 56 (2007) 493–503.

Please cite this article in press as: V. Simoncini, Algebraic properties of solutions to certain nonlinear matrix equations, Linear Algebra Appl. (2025), https://doi.org/10.1016/j.laa.2025.09.017