Classification and Approximation of Solutions to Sylvester Matrix Equation

Bogdan D. Djordjević, Nebojša Č. Dinčić

Abstract. In this paper we solve Sylvester matrix equation with infinitely-many solutions and conduct their classification. If the conditions for their existence are not met, we provide a way for their approximation by least-squares minimal-norm method.

1. Introduction and preliminaries

For given vector spaces \( V_1 \) and \( V_2 \), let \( A \in L(V_2) \), \( B \in L(V_1) \) and \( C \in L(V_1, V_2) \) be linear operators. Equations of the form

\[
AX - XB = C
\]

with solution \( X \in L(V_1, V_2) \) are called Sylvester equations, Sylvester-Rosenblum equations or algebraic Ricatti equations. Such equations have various application in vast fields of mathematics, physics, computer science and engineering (see e.g. [5], [19] and references therein). Fundamental results, established by Sylvester and Rosenblum themselves, are now-days the starting point in solving contemporary problems where these equations occur. These results are

Theorem 1.1. [23] (Sylvester matrix equation) Let \( A, B \) and \( C \) be matrices. Equation \( AX - XB = C \) has unique solution \( X \) iff \( \sigma(A) \cap \sigma(B) = \emptyset \).

Theorem 1.2. [22] (Rosenblum operator equation) Let \( A, B \) and \( C \) be bounded linear operators. Equation \( AX - XB = C \) has unique solution \( X \) if \( \sigma(A) \cap \sigma(B) = \emptyset \).

Equations with unique solutions have been extensively studied so far. There are numerous results regarding this case, some of them theoretical (e.g. Lyapunov stability criteria and spectral operators), which can be found in [2], [5] or [9], and some of them computational (matrix sign function, factorization of matrices and operators, various iterative methods etc.). It should be mentioned that matrix eq. (1) with unique solution \( X \) has been solved numerically (among others) in [4], [6], [13], [14], [18] and [21]. The case where
A, B and C are unbounded operators but solution X is unique and bounded has been studied in [16] and [20].

Solvability of eq. (1) in matrices, discarding uniqueness of solution, has been studied in [7] and partially in [18]. Main results in [7] are based on the idea that solutions X can be provided as parametric matrices, where number of parameters at hand depends on dimensions of the corresponding eigenspaces for A and B.

The case where A, B and C are unbounded, with infinitely many unbounded solutions X has been studied in [8]. This particular research paper provides insight on new solutions (called the weak solutions), which are only defined on the corresponding eigenspaces for A and B.

This research paper concerns the case when A and B are matrices whose spectra intersect, while matrix C is a rectangular matrix of appropriate dimensions. We obtain sufficient conditions for existence of infinitely-many solutions and provide a way for their classification. If the conditions for their existence are not met, we give a way of approximating particular solutions. This study relies on the eigenspace-analysis conducted in [7] and [8].

We assume $V_1$ and $V_2$ to be finite dimensional Hilbert spaces over the same scalar filed C or $\mathbb{R}$, while $A \in \mathcal{B}(V_2)$, $B \in \mathcal{B}(V_1)$ and $C \in \mathcal{B}(V_1, V_2)$ are assumed to be operators which correspond to the afore-mentioned matrices. Further, $\mathcal{N}(L)$ and $\mathcal{R}(L)$ denote null-space and range of the given operator L. Recall that every finite-dimensional subspace $W$ of a Hilbert space $V$ is closed. Consequently, there exists orthogonal projector from $V$ to $W$, which will be denoted as $P_W$.

2. Existence and classification of solutions

Throughout this paper, we assume that A and B share s common eigenvalues and denote that set by $\sigma$:

$$\{\lambda_1, \ldots, \lambda_s\} =: \sigma = \sigma(A) \cap \sigma(B).$$

For more elegant notation, we introduce $E^k_B = \mathcal{N}(B - \lambda_k I)$ and $E^k_A = \mathcal{N}(A - \lambda_k I)$ whenever $\lambda_k \in \sigma$. Different eigenvalues generate mutually orthogonal eigenvectors, so the spaces $E^k_B$ form an orthogonal sum. Put $E_B := \sum_{k=1}^s E^k_B$. It is a closed subspace of $V_1$ and there exists $E^\perp_B$ such that $V_1 = E_B \oplus E^\perp_B$. Take $B = B_E \oplus B_1$ with respect to that decomposition and denote $C_1 = CP^\perp_{E^\perp_B}$.

**Proposition 2.1.** Let $V$ be a Hilbert space and $L \in \mathcal{B}(V)$. If $W$ is $L$–invariant subspace of $V$, then $W^\perp$ is $L^*$–invariant subspace of $V$.

**Theorem 2.1.** *(Existence of solutions)* For every $k \in \{1, \ldots, s\}$, let $\lambda_k, E^k_A$ and $E^k_B$ be provided as in the previous paragraph. If

$$\mathcal{N}(C_1)^\perp = \mathcal{R}(B_1) \quad \text{and} \quad C \left( E^k_B \right) \subset \mathcal{R}(A - \lambda_k I),$$

then there exist infinitely many solutions $X$ to the equation (1).

**Proof.** For every $1 \leq k \leq s$, let $E^k_B$, $E^k_B$, $E^k_E$ and $B_1$ be provided as in the previous paragraph. Note that $\mathcal{N}(C_1)^\perp = \mathcal{R}(C_1^*)$, where $C_1^* \in \mathcal{B}(V_2, E^\perp_B)$.

**Step 1:** solutions on $E^\perp_B$. 


We first conduct analysis on $E_B^\perp$. Space $E_B$ is $BP_{E_B^\perp}$-invariant subspace of $V_1$ and Proposition 2.1 yields $E_B^\perp$ to be $(BP_{E_B})^\perp$-invariant subspace of $V_1$, so without loss of generality we can observe $B_1'$ as $B_1' : E_B^\perp \rightarrow E_B^\perp$. Since $\sigma(B_E) = \{\lambda_1, \ldots, \lambda_n\}$, it follows that

$$\sigma(B_1') \subseteq \{0\} \cup \sigma(B') \setminus \{\lambda_1, \ldots, \lambda_n\}.$$ 

**Case 1.** Assume that $\sigma(B_1') \cap \sigma(B') = \emptyset$. Then there exists unique $X_1' \in \mathcal{B}(V_2, E_B^\perp)$ such that

$$X_1'A' - B_1'X_1' = C_1',$$

that is, there exists unique $X_1 \in \mathcal{B}(E_B^\perp, V_2)$ such that

$$AX_1 - X_1B_1 = C_1$$

holds.

**Case 2.** Assume that $\sigma(A') \cap \sigma(B_1') \neq \emptyset$. It follows that $\sigma(A') \cap \sigma(B_1') = \{0\}$. But then $A'$ cannot be nilpotent. Truly, if $\sigma(A') = \{0\} = \sigma(A)$, then by assumption, $\sigma(B) \cap \sigma(A) \neq \emptyset$, therefore, $0 \in \sigma(B)$, that is, $0 \in \sigma$. If $u \in N(B_1)$, then $Bu = 0$ and $u \in E_B^\perp$, but then $Bu = B_1u = 0$, so $u \in N(B) \subseteq E_B$, therefore $u \in E_0 \cap E_B = \{0\}$. Hence contradiction, implying that $A'$ is not nilpotent, but rather has finite ascent, $\text{asc}(A') = m \geq 1$, where $N(A'(A'))$ is a proper subspace of $V_2$.

Now observe $B_1' : E_B^\perp \rightarrow E_B^\perp$, which is not invertible by assumption. Take arbitrary $Z_0' \in \mathcal{B}(N(A'), N(B_1'))$ operator. Then for every $d \in N(A')$, there exists (by (2)) unique $u \in N(B_1')$ such that

$$B_1'u = C_1'd.$$ 

Define $X_1'(Z_0')$ on $N(A')$ as $X_1'(Z_0')d := Z_0'd + u$. Since $\text{asc}(A') = m$, the following recursive formula applies.

Assume that $m = 1$. Precisely, decompose $V_2 = N(A') \oplus N(A')^\perp$ and $A' = 0 \oplus A_1'$. Then $A_1'$ is injective from $N(A')^\perp$ to $N(A')^\perp$ and $X_1'$ can be defined on $N(A')^\perp$ as restriction of $X_1'$ from Case 1.

Assume that $m > 1$. Then proceed to decompose $N(A')^\perp = N(A_1') \oplus N(A_1')^\perp$ and and define $X_1'$ on $N(A_1')$ as $X_1'(N_1')u := N_1'u + d$, where $Z_1' \in \mathcal{B}(N(A_1'), N(B_1'))$ is arbitrary operator and

$$B_1'u = C_1'd.$$ 

If $A_1'$ is injective on $N(A_1')^\perp$, i.e. if $m = 2$, then $X_1$ can be defined on $N(A_1')^\perp$ as restriction of $X_1$ from Case 1. If not, then proceed to decompose $N(A_1')^\perp = N(A_2') \oplus N(A_2')^\perp$ and so on. Eventually, one would get to iteration no. $m$, in a manner that

$$V_2 = N(A') \oplus N(A_1') \oplus N(A_2') \oplus \cdots \oplus N(A_m') \oplus N(A_m')^\perp$$

and $A_1' : N(A_m') \rightarrow N(A_m')$ is injective. Then $\sigma(B_1') \cap \sigma(A_1') = \emptyset$, ergo define $X_1'$ on $N(A_m')$ as restriction of $X_1'$ from Case 1 to $N(A_m')^\perp$. Further, for $0 \leq n \leq m$, let $Z_n' \in \mathcal{B}(N(A_n'), N(B_1'))$ be arbitrary operators. Then define $X_1'$ on $N(A_m')$ as

$$X_1'(Z_n')d := Z_n'd + u,$$

where once again $u \in N(B_1')^\perp$ is unique element such that $B_1'u = C_1'd$. Equivalently, there exists $X_1 \in \mathcal{B}(E_B^\perp, V_2)$ such that

$$AX_1 - X_1B_1 = C_1,$$

where

$$X_1 = X_1(Z_0', Z_1', \ldots, Z_m').$$

Condition $\mathcal{R}(C_1') = N(B_1')^\perp = \mathcal{R}(B_1)$ yields $X_1$ to be well defined on the entire $E_B^\perp$. 

B. D. Djordjevi´c, N.ˇC. Dinˇci´c / Filomat 33:13 (2019), 4261–4280
Step 2: solutions on \( E_B \).

We now conduct our analysis on \( E_B \). Define \( E_A = \sum_{k=1}^{s} E_A^k \) and split \( V_2 \) into orthogonal sum \( V_2 = E_A \oplus E_A^\perp \). Decompose \( A = A_E \oplus A_1 \) with respect to that sum. Then \( A_1 \) is injective on \( E_A^\perp \) and \( A_1 v = Av \), for every \( v \in E_A^\perp \). For every \( k \in \{1, \ldots, s\} \) let \( N_k \in \mathcal{B}(E_B^k, E_A^k) \) be arbitrary. For every \( u \in E_B^k \) by assumption (2), there exists unique \( d(u) \in \left( E_A^k \right)^\perp \) such that

\[
(A - \lambda_k I)d(u) = Cu.
\]

Define

\[
X^k_E : u \mapsto N_k u + d(u), \quad u \in E_B^k.
\]

Then \( X^k_E : E_B^k \to E_A^k \oplus \left( P_{E_A^k} (A_1 - \lambda_k I)^{-1} C E_B^k \right) \) defines a linear map. What is left is to check whether \( X_E := \sum_{k=1}^{s} X^k_E \) is a solution to the equation

\[
AX_E - X_E B_E = CP_{E_B}
\]

restricted to \( E_B \). However, this is directly verifiable. For any \( u \in E_B \) there exist unique \( \alpha_1, \ldots, \alpha_s \in \mathbb{C}(\mathbb{R}) \) and unique \( u_k \in E_B^k, 1 \leq k \leq s \), such that \( u = \sum \alpha_k u_k \). Then

\[
(AX_E - X_E B_E)u = A \sum_{k=1}^{s} \alpha_k X^k_E u_k - \sum_{k=1}^{s} \lambda_k \alpha_k X^k_E u_k = \sum_{k=1}^{s} (\alpha_k (A - \lambda_k I)) (N_k u_k + d(u_k)) = \sum_{k=1}^{s} \alpha_k Cu_k = Cu.
\]

It follows that

\[
X = \begin{pmatrix} X_E & 0 \\ 0 & X_1 \end{pmatrix}.
\]

is a solution to eq. (1). \( \square \)

Remark. Notice that in proof of Theorem 2.1 Case 2. only emerges if \( \sigma(A) \cap \sigma(B_1) = \{0\} \) while \( 0 \notin \sigma(B) \). Because of special circumstances under which this problem takes place, this situation will be analyzed separately from the standard problem (see Corollary 2.3).

Theorem 2.1 naturally inquires answers to the following questions:

**Question 1.** Is every solution to the equation (1) of the form (3)?

**Question 2.** Under which conditions is the solution to (1) unique?

Both of these questions have affirmative answers, which is justified by analysis of the following eigen-problem associated with given Sylvester equation:

Assume that \( 0 \in \sigma = \sigma(A) \cap \sigma(B) \) and let \( N_\lambda \in \mathcal{B}(E_B^\lambda, E_A^\lambda) \), for every \( \lambda \in \sigma \) be arbitrary. Define \( N_0 := \oplus_{\lambda \in \sigma} N_\lambda \). Find a solution \( X \) to Sylvester equation such that the following eigen-problem is uniquely solved

\[
\begin{cases}
AX - XB = C \\
Xu_\lambda := P_{(E_B^\lambda)^\perp} (A - \lambda I)^{-1} Cu_\lambda + N_\lambda u_\lambda, \quad u_\lambda \in E_B^\lambda, \quad \lambda \in \sigma \cup \{0\}.
\end{cases}
\]

**Theorem 2.2.** (Uniqueness of the solution to the eigen-problem) With respect to the previous notation, assume that \( 0 \in \sigma \).

1) If the condition (2) holds for every shared eigenvalue \( \lambda \in \sigma \), then solution \( X \) depends only on the choice of operator \( N_\lambda \), that is, for fixed \( N_\lambda \), there exists unique solution \( X \) such that (4) holds.

2) Conversely, for every solution \( X \) to (1) and for every shared eigenvalue \( \lambda \) for matrices \( A \) and \( B \), there
exists unique quotient class \((A - \lambda I)^{-1}C(N(B - \lambda I)) \oplus N(A - \lambda I)\) such that \(X\) is unique solution to the quotient eigen-problem

\[
\begin{cases}
AX - XB = C \\
X : N(B - \lambda I) \to (A - \lambda I)^{-1}C(N(B - \lambda I)) \oplus N(A - \lambda I).
\end{cases}
\]  

(5)

**Proof.** Recall notation from proof of Theorem 2.1.

1) The first statement of the theorem is proved directly. Namely, take \(V_1 = E_B \oplus E_A^1, B = B \oplus B_1, V_2 = E_A \oplus E_A^1, A = A_E \oplus A_1\) like in Theorem 2.1. Then there exists \(X = X_E \oplus X_1\), which is a solution to (1). By construction, since \(\sigma(B) \cap \sigma(A) = \emptyset\), Case 1. applies and \(X_1\) is uniquely determined in \(\mathcal{B}(E^2, V_2)\) while \(X_E^1\) is uniquely determined in the class \(\mathcal{B}(E_B/E_B^2, V_2/E_A^1)\) for every \(\lambda \in \sigma\). Varying \(\lambda\) in \(\sigma\) completes the proof.

2) Conversely, let \(X\) be a solution to the eq. (1). Let \(\lambda\) be one of the shared eigenvalues for \(A\) and \(B\) and fix \(u\) as a corresponding eigenvector for \(B\). Then \(X Bu = \lambda X u\). Hence

\[AXu - XB u = (A - \lambda I)Xu = Cu.\]

Split \(X u\) into the orthogonal sum \(X u = v_1 + v_2\), where \(v_1 \in N(A - \lambda I)\) and \(v_2 \in (N(A - \lambda I))^\perp\). Then \(v_2\) is the sought expression \(P_{N(A - \lambda I)^\perp}(A - \lambda I)^{-1}Cu\) and \(X u = v_2 + (N(A - \lambda I))\). Condition (2) follows immediately. Repeating the same procedure for every shared eigenvalue for \(A\) and \(B\) completes the proof. \(\square\)

**Corollary 2.1.** (Number of solutions) let \(\Sigma\) be the set of all \(N_\sigma\) introduced in the eigen-problem associated with given Sylvester equation (1), that is

\[\Sigma = \{ N_\sigma : N_\sigma = \oplus_{\lambda \in \sigma} N_\lambda, \quad \forall \lambda \in \sigma(A) \cap \sigma(B) = \sigma \ni \{0\} \}.\]

Let \(S\) be the set of all solutions to (1) which satisfy condition (2). Then \(|\Sigma| = |S|\).

**Proof.** For arbitrary \(N_\sigma \in \Sigma\), there exits unique \(X \in S\) such that (4) holds. Further, for arbitrary \(X \in S\) and arbitrary \(\lambda \in \sigma\) there exist quotient classes \(E_\lambda^1\) and \(E_\lambda^2\) such that (5) holds. Define \(N_\lambda : E_B^1 \to E_A^1\) to be bounded. Then \(N_\sigma = \oplus_{\lambda \in \sigma} N_\lambda\). It follows that \(N_\sigma \in \Sigma\). There is one-to-one surjective correspondence \(S \leftrightarrow \Sigma\). \(\square\)

**Remark.** Due to Corollary 2.1, solution \(X(N_\sigma) \in S\), \((N_\sigma \in \Sigma)\), can be referred to as particular solution.

**Corollary 2.2.** (Size of particular solution) With the assumptions and notation from Theorem 2.1, Theorem 2.2 and Corollary 2.1, norm of \(X(N_\sigma)\) due to

\[|X(N_\sigma)|^2 = |X_E|^2 + |X_1|^2 = |N_\sigma|^2 + \sum_{k=1}^{s} |P_{(E_k^1)}(A - \lambda_k I)^{-1}CP_{E_k^1}|^2 + |X_1|^2.\]

(6)

**Proof.** Taking the same decomposition as in Theorem 2.1, let \(X = X_E + X_1\). Since \(X_E\) annihilates \(E_B^1\) and \(X_1\) annihilates \(E_B\), it follows that

\[|X|^2 = |X_E + X_1|^2 = |X_E|^2 + |X_1|^2.\]

By the same argument, taking

\[|X_E|^2 = |N_\sigma|^2 + \sum_{k=1}^{s} |P_{(E_k^1)}(A - \lambda_k I)^{-1}CP_{E_k^1}|^2\]

completes the proof. \(\square\)

**Corollary 2.3.** (Singularities on \(E_B^1\)) Assume that \(0 \notin \sigma\) but \(0 \in \sigma(A) \cap \sigma(B_1)\) and let \(dsc\(A) = m \geq 1\). For every \(0 \leq n \leq m\), define \(Z_n \in \mathcal{B}(R(B_1)^\perp, R(A^{n+1})^\perp \cap R(A^n))\) and let \(Z = \sum_{n=0}^{m} Z_n\). If \(N(C_1)^\perp = R(B_1)\), then there are infinitely many solutions to (1) on \(E_B^1\). Those solutions depend only on choice for \(Z\), that is, if \(Z\) is fixed then there exists unique solution \(X_1(Z)\) on \(E_B^1\).

**Proof.** Proof is the same as part 1) in Theorem 2.2. Note that \(dsc\(A) = asc\(A^*\) = m\) and \(R(A^{n+1})^\perp \cap R(A^n) = N((A^*)^{n+1}) \cap N((A^*)^n)^\perp\). Then proceed to Case 2. of proof of Theorem 2.1. \(\square\)
3. Fourier approximation minimal norm solution

As illustrated in Theorem 2.2, the system (4) has unique solution provided that all the input parameters are known and satisfy conditions (2). However, if the only input information is \( \sigma(A) \cap \sigma(B) = \emptyset \), the condition (2) is in general not easy or possible to verify. Thus approximation analysis requires more detailed approach.

The easiest assumption is that there exist eigenvalues for \( A \) and \( B \)
\[
\lambda_{k_1}, \ldots, \lambda_{k_w} \in \sigma
\]
such that
\[
C(E_B^\ell) \cap \mathcal{R}(A - \lambda \ell I) = \emptyset, \quad \ell \in \{k_1, \ldots, k_w\}
\]
Ergo any eigenvector \( u_\ell \) of \( B \) that corresponds to \( \lambda_{k_\ell} \) does not obey the condition (2) (\( \ell = k_1, k_w \)), that is, \( Cu_\ell \not\in \mathcal{R}(A - \lambda \ell I) \). There exists an orthonormed basis \( (e_k)_k \) for \( \mathcal{R}(A - \lambda \ell I) \), such that \( Cu_\ell \) can be approximated by \( \overline{C}u_\ell \in \mathcal{R}(A - \lambda \ell I) \) and this approximation is the best possible, where
\[
\overline{C}u_\ell = \sum_k \langle Cu_\ell, e_k \rangle e_k.
\]

This way, operator \( \overline{C} \) is directly defined on \( \sum_{\ell=k_1}^{k_w} E_B^\ell \equiv E_B^w \). The space \( E_B^w \) is finite-dimensional and therefore has an orthogonal complement in \( V_1 \), denoted as \( W = E_B^w \perp \). Thus the extension of \( \overline{C} \) on \( V_1 \) is admissible and we define
\[
\overline{C} := \overline{C} \oplus CP_W.
\]

Now we solve the approximate Sylvester equation \( AX - XB = \overline{C} \), and the solutions \( X \) (which exist from Theorem 2.1) are approximate solutions to the initial eq. \( AX - XB = C \). Combining Corollary 2.2, we see that the error of approximation is derived from
\[
\sup_{\|u\|=1} \| (AX - XB - C)u \| = \sup_{\|u\|=1} \| (\overline{C} - C)u \|
\]
and this approximation is the best possible, for given \( u_\ell \). However, note that \( \overline{C} \) is not uniquely determined, but still depends on the input parameters: the corresponding eigenvectors for \( B \) and the choice for bases in the spaces \( \mathcal{R}(A - \lambda \ell I) \). Hence we try to extract one particular \( C \) which is the best suited for our approximation problem:

**Problem 0.** Find those (or that one) approximations for \( C \) such that solutions have the smallest possible norm
\[
\overline{C} = \{ C : \ AX - XB = \overline{C} \Rightarrow \|X\| \text{ is the smallest possible} \}.
\]

This transfers our problem into minimum function problem, which is solvable in terms of numerical analysis.

4. Least-squares Minimum-norm solutions

When it comes to applications of matrix Sylvester equation, Frobenius norm seems to play more important role than the operator \( \sup \) –norm. Hence we continue our approximation analysis with that norm.

Frobenius norm, sometimes also called Euclidean norm, of a matrix \( A \in \mathbb{C}^{m \times n} \) is defined as
\[
\| A \|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \left( \sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A) \right)^{1/2},
\]
where \( \sigma_i(A) \) are the singular values of \( A \). Also, \( \|A\|_F = \text{tr}(AA^*)^{1/2} \) (recall that \( A^* = A^T \) is conjugate transpose). Recall that Frobenius norm is sub-multiplicative, i.e. \( \|AB\|_F \leq \|A\|_F\|B\|_F \) and unitarily invariant i.e. \( \|U_1AU_2\|_F = \|A\|_F \) for some unitary \( U_1, U_2 \). From the very definition, it also follows that for matrix \( A \) partitioned on block-matrices, \( A = [A_{ij}]_{p \times q} \) it follows that \( \|A\|_F^2 = \sum_{i=1}^p \sum_{j=1}^q \|A_{ij}\|_F^2 \). We mention that on any finite dimensional space any two norms are equivalent.

Now we state two problems about Sylvester equation \( AX - XB = C \) we are dealing with. Recall that \( A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{m \times n} \) and \( C \in \mathbb{C}^{m \times n} \) are given, and \( X \in \mathbb{C}^{m \times n} \) is an unknown matrix. Now Problem 0. can be broken down into two separate problems

**Problem 1.** Find the set \( S \) of all \( \tilde{X} \) such that \( \|AX - XB - C\|_F \) is the smallest possible, i.e.

\[
\min_{\tilde{X} \in \mathbb{C}^{m \times n}} \|AX - XB - C\|_F = \|A\tilde{X} - \tilde{X}B - C\|_F.
\]

**Problem 2.** Among all \( \tilde{X} \) find the one with the smallest Frobenius norm, i.e.

\[
\min_{\tilde{X} \in S} \|\tilde{X}\|_F = ||\tilde{X}_0\|_F.
\]

**Definition 4.1.** Matrices \( \tilde{X} \) which are solutions for Problem 1 are least-squares solutions. Matrices \( \tilde{X}_0 \) which are solutions for Problem 2 are minimal-norm least-squares solution.

Before we continue our analysis, we remark the following facts:

- if Sylvester equation is consistent for given \( C \), then set \( S \) from Problem 1 consists of all solutions of the Sylvester equation, and norm of approximation error is zero. If there is unique solution, then it solves Problem 2 as well. In the case when there are infinitely many solutions, Problem 2 gives those solutions with the smallest norm;

- for homogeneous equation (i.e. \( C = 0 \)), set \( S \) consists of all homogeneous solutions, and solution of Problem 2 is unique, namely \( X = 0 \).

It is well-known fact that eq. (1) can, by Kronecker product and vectorization operation, be transformed into

\[
(I_n \otimes A - B^T \otimes I_n) \text{vec}(X) = \text{vec}(C).
\]

The matrix \( I_n \otimes A - B^T \otimes I_n \) is often called “niveletteur” in the literature, and the least-squares minimal-norm solution is unique and is given by

\[
\text{vec}(\tilde{X}) = (I_n \otimes A - B^T \otimes I_n)^{\dagger} \text{vec}(C),
\]

where \( T^\dagger \) denotes the unique Moore-Penrose inverse of in general rectangular complex matrix \( T \). For more on the topic of the generalized inverses reader is referred to [3]. Remark that effective calculation of the Moore-Penrose inverse of niveletteur appears to be very difficult. The authors are unaware of such method, but for the group inverse there is a recent paper of Hartwig and Patricio [11].

Our aim is to reduce the problem for original Sylvester equation to the simplest Sylvester equation case, similar to the approach used in [7]. Suppose that matrices \( A \in \mathbb{C}^{m \times n} \) and \( B \in \mathbb{C}^{n \times n} \) have the following Jordan canonical forms (for some invertible matrices \( S \) and \( T \)):

\[
A = SJ_A S^{-1}, \quad B = TJ_B T^{-1}.
\]

Without loss of generality, we may assume that \( \emptyset \neq \sigma(A) \cap \sigma(B) = \{\lambda_1, ..., \lambda_s\} \), hence we excluded the unique solution case, so

\[
J_A = \text{diag}([\lambda_1; p_{11}, p_{12}, ..., p_{1k_1}], ..., [\lambda_r; p_{s1}, p_{s2}, ..., p_{sk_r}]) \in \mathbb{C}^{m \times m},
\]

\[
J_B = \text{diag}([\ell_1; q_{11}, q_{12}, ..., q_{1k_1}], ..., [\ell_r; q_{s1}, q_{s2}, ..., q_{sk_r}]) \in \mathbb{C}^{n \times n},
\]
where \( p_{ij}, \ j = 1, \ell_i, \ i = 1, s, \) and \( q_{ij}, \ j = 1, \ell_i, \ i = 1, s, \) are natural numbers, and \( k_i \) and \( \ell_i \) are geometric multiplicities of the eigenvalue \( \lambda_i, \ i = 1, s, \) of \( A \) and \( B, \) respectively. Notation \( J(\lambda_1 t_1, \ldots, t_k) \) stands for

\[
J(\lambda; t_1, \ldots, t_k) = \text{diag}(J_{h_1}(\lambda), \ldots, J_{h_k}(\lambda)) = J_{h_1}(\lambda) \oplus \cdots \oplus J_{h_k}(\lambda),
\]

where \( J_{h_i}(\lambda) \) is the Jordan block matrix of dimension \( t_i \times t_i \) with \( \lambda \) on its main diagonal.

If we put those Jordan forms in the equation, we have (we denoted \( Y = S^{-1}X T \) and \( D = S^{-1}CT)\):

\[
\|AX - XB - C\|_F^2 = \|SJ_A S^{-1}X - XTJ_B T^{-1} - C\|_F^2 = \\
= \|S(IA S^{-1}X T - (S^{-1}XT)J_B - (S^{-1}CT)T^{-1})\|_F^2 = \\
= \|S(J_A Y - Y J_B - D) T^{-1}\|_F^2 \leq \\
= \|S\|_F^2 \|T\|_F^{-1} \|T\|_F^{-1} \|J_A Y - Y J_B - D\|_F^2 = \\
= \alpha^2(S, T) \|J_A Y - J_B Y - D\|_F^2.
\]

We used the notation \( \alpha(S, T) = \|S\|_F \|T\|_F^{-1} \|T\|_F^{-1} \|. \) Remark that if \( S \) and \( T \) are unitary matrices, then equality is attained in the previous formula.

Because of:

\[
J_A Y - Y J_B - D = [J(\lambda_i p_{i1}, \ldots, p_{i, k_i})][Y_{ij}] - [Y_{ij}][J(\lambda_j q_{j1}, \ldots, q_{j, \ell_j})] - [D_{ij}] = \\
= [J(\lambda_i p_{i1}, \ldots, p_{i, k_i}) Y_{ij} - Y_{ij} J(\lambda_j q_{j1}, \ldots, q_{j, \ell_j}) - D_{ij}]_{k_i \times \ell_j} = \\
= [J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) - D_{ij}]_{k_i \times \ell_j} \sum_{u=1}^k \sum_{v=1}^s \sum_{i=1}^t \\
\|
\]

we have

\[
\|J_A Y - J_B Y - D\|_F^2 = \sum_{i,j=1}^s \|J(\lambda_i p_{i1}, \ldots, p_{i, k_i}) Y_{ij} - Y_{ij} J(\lambda_j q_{j1}, \ldots, q_{j, \ell_j}) - D_{ij}\|_F^2 = \\
= \sum_{i,j=1}^s \sum_{u=1}^k \sum_{v=1}^s \sum_{i=1}^t \|J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) - D_{ij}\|_F^2.
\]

Now, we distinguish two cases:

- if \( \lambda_i \neq \lambda_j, \) then by Theorem 1.1 the equation \( J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) = D_{ij}\) has unique solution \( Y_{ij}^{(i)}, \)
so \( \|J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) - D_{ij}\|_F^2 = 0. \)

- if \( \lambda_i = \lambda_j, \) then the equation \( J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) = D_{ij}\), after translation for \( \lambda_i, \) reduces to \( J_{p_{i1}}(0) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(0) = D_{ij}. \) From Theorem 1.1 we already know that there may be either infinitely many solutions, or no solutions at all.

Therefore,

\[
\|AX - XB - C\|_F^2 \leq \alpha^2(S, T) \|J_A Y - J_B Y - D\|_F^2 = \\
= \alpha^2(S, T) \sum_{i,j=1}^s \sum_{u=1}^k \sum_{v=1}^s \sum_{i=1}^t \|J_{p_{i1}}(\lambda_i) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(\lambda_j) - D_{ij}\|_F^2 = \\
= \alpha^2(S, T) \sum_{i,j=1}^s \sum_{u=1}^k \sum_{v=1}^s \sum_{i=1}^t \|J_{p_{i1}}(0) Y_{ij}^{(i)} - Y_{ij}^{(i)} J_{q_{j1}}(0) - D_{ij}\|_F^2.
\]
Now we can take minimum over all $X$ (equivalently, over all $Y$, since they are similar matrices):

$$\min_X \|AX - XB - C\|_F^2 \leq \alpha^2(S,T) \min_Y \|AY - YB - D\|_F^2 =$$

$$= \alpha^2(S,T) \sum_{i=1}^k \sum_{m=1}^\ell \sum_{v=1}^\eta \min_{\nu} \|f_{\nu,v}(0)Y_{\nu,v}^{(i)}(0) - Y_{\nu,v}^{(i)}(0) - D_{\nu,v}^{(i)}\|_F^2.$$ 

Therefore, in order to solve the Problem 1, we need to investigate the following simpler versions of original problems:

**Problem 1’.** Find the set $S$ of all least-squares solutions $\hat{X}$, i.e.

$$\min_{\hat{X} \in \mathbb{C}^{m \times n}} \|f_m(0)X - Xf_n(0) - C\|_F = \|f_m(0)\hat{X} - \hat{X}f_n(0) - C\|_F.$$

**Problem 2’.** Among all $\hat{X} \in S$ find the one, $\hat{X}_0$, with the smallest Frobenius norm, i.e.

$$\min_{\hat{X} \in S} \|\hat{X}\|_F = \|\hat{X}_0\|_F.$$ 

We will prove that such $\hat{X}$ is unique, and give a method for its explicit finding.

### 5. Least-squares Solutions for the Simplest Case

Let us denote $p = \min\{m, n\}$ for $m, n \in \mathbb{N}$. For given matrix $A \in \mathbb{C}^{m \times n}$, the set

$$d_k(A) := \{a_{ij} : j - i = k\}, \quad k = -m + 1, n - 1,$$

will be called $k$–th small diagonal. For $k = 0$ we have "the" diagonal, i.e. the set $\{a_{ii} : i = 1, p\}$. When we refer to some small diagonal, we assume that its elements are ordered accordingly to increase of the index $i$. For example, if we are dealing with 0–th small diagonal, we assume that the set $d_0(A) = \{a_{11}, a_{22}, ..., a_{pp}\}$ is ordered. We will denote by $\sigma_{m+k}$ sum of all elements along the $k$–th small diagonal $d_k$:

$$\sigma_{m+k}(A) = \sum_{a_{ij} \in d_k} a_{ij}, \quad k = -m + 1, n - 1.$$ 

**Theorem 5.1.** Sylvester equation $f_m(0)X - Xf_n(0) = C$ has a least-squares solution $\hat{X}$ given by

$$\hat{X} = X_h + X_p + X_c,$$

where $X_h$ denotes the solution of appropriate homogeneous equation:

$$X_h = \begin{cases} \begin{pmatrix} p_{m-1}(f_m(0)) & 0_{(m-n) \times n} \\ 0_{n \times (m-n)} & q_{m-1}(f_n(0)) \end{pmatrix}, & m \geq n, \\ \begin{pmatrix} 0_{mx(n-m)} & q_{m-1}(f_n(0)) \end{pmatrix}, & m \leq n, \end{cases}$$

$X_p$ is an expression given by

$$X_p = \begin{cases} \sum_{k=0}^{n-1} \left(f_m(0)^T\right)^{k+1} C f_n(0)^k, & m \geq n, \\ -\sum_{k=0}^{m-1} f_m(0)^k C \left(f_n(0)^T\right)^{k+1}, & m \leq n, \end{cases}$$

$X_c$ an expression given by

$$X_c = \begin{cases} \sum_{k=0}^{n-1} \left(f_m(0)^T\right)^{k+1} C f_n(0)^k, & m \geq n, \\ -\sum_{k=0}^{m-1} f_m(0)^k C \left(f_n(0)^T\right)^{k+1}, & m \leq n, \end{cases}$$
where we denoted
\[ X_c = \begin{cases} 
\begin{bmatrix} 0_{(m-n) \times n} \\ W \\ -W^T \ 0_{m \times (n-m)} \end{bmatrix}, & m \geq n, \\
\end{cases} \]
\[ W = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\
\sigma_{p/p} & 0 & \ldots & 0 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\sigma_3/3 & 2\sigma_4/4 & \ldots & 0 & 0 \\
\sigma_2/2 & 2\sigma_3/3 & (p-1)\sigma_{p/p} & 0 & 0 \\
\end{bmatrix}_{p \times p}. \]

The magnitude of the deviation is:
\[ \Delta(\hat{X}; C) = \min_X \Delta(X; C) = \sum_{k=1}^{n} \frac{\sigma_k^2}{k}. \]

where \( \sigma_k \) is a sum of elements over the \((m+k)\)-th small diagonal from the matrix \( C \).

**Proof.** In expanded form, matrix expression \( R \equiv R(X) = J_m(0)X - XJ_n(0) - C = [r_{ij}] \) is:
\[
\begin{bmatrix}
    x_{21} - c_{11} & x_{22} - x_{11} - c_{12} & \ldots & x_{2n} - x_{1,n-1} - c_{1n} \\
    x_{31} - c_{21} & x_{32} - x_{21} - c_{22} & \ldots & x_{3n} - x_{2,n-1} - c_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    -c_{m1} & x_{m2} - x_{m-1,1} - c_{m-1,2} & \ldots & x_{mn} - x_{m-1,n-1} - c_{m-1,n} \\
    -x_{m1} & -c_{m2} & \ldots & -x_{mn} - c_{mn} \\
\end{bmatrix}.
\]

As we can see, any fixed small diagonal contains some unknowns and parameters, and those unknowns and parameters cannot be found in any other small diagonal. Therefore, we will make summation over all small diagonals, and then of all elements on each odd small diagonals. Since all those summations of the elements can have one of possible 3 forms, described by functions \( M_1, M_2 \) and \( M_4 \) for \( m \geq n \) (or \( M_1, M_3 \) and \( M_4 \) for \( m \leq n \)) from Proposition 8.1 (see Appendix), it is customary to separate into three sums. So we have:

\[
||R(X)||^2_F = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}^2 = \sum_{k=m+1}^{n-p} \sum_{r_i} r_{ij}^2 =
\]
\[
= r_{m1}^2 + \sum_{k=m+2}^{n-p} \sum_{r_i} r_{ij}^2 + \sum_{k=m+1}^{n-p} \sum_{r_i} r_{ij}^2 + \sum_{k=m+p+1}^{n} \sum_{r_i} r_{ij}^2 =
\]
\[
= c_{m1}^2 + \sum_{k=m+2}^{n-p} M_1(x_{ij} \in d_k; a_{ij} \in d_k) + \sum_{k=m+p+1}^{n-p} M_2(x_{ij} \in d_k; a_{ij} \in d_k) +
\]
\[
+ \sum_{k=m+p+1}^{n-p} M_4(x_{ij} \in d_k; a_{ij} \in d_k)
\]

Since all summands are nonnegative, the minimization works independently for each of the small diagonal in some of three sums by using Proposition 8.1. We denote by “hat” the elements on which the minimum
is attained. Therefore:

\[
\min_X \| R(X) \|_F^2 = c_{m1}^2 + \sum_{k=-m+2}^{-m+p} \min M_1(x_{ij} \in d_k; a_{ij} \in d_k) + \sum_{k=-m+p+1}^{n-p} \min M_{2V3}(x_{ij} \in d_k; a_{ij} \in d_k) + \\
+ \sum_{k=n-p+1}^{n-1} \min M_4(x_{ij} \in d_k; a_{ij} \in d_k) = \\
= c_{m1}^2 + \sum_{k=-m+2}^{-m+p} M_1(\bar{x}_{ij} \in d_k; a_{ij} \in d_k) + \sum_{k=-m+p+1}^{n-p} M_{2V3}(\bar{x}_{ij} \in d_k; a_{ij} \in d_k) + \\
+ \sum_{k=n-p+1}^{n-1} M_4(\bar{x}_{ij} \in d_k; a_{ij} \in d_k) = \\
= c_{m1}^2 + \sum_{k=-m+2}^{-m+p} M_1(\bar{x}_{ij} \in d_k; a_{ij} \in d_k) + 0 + 0 = \\
= \sum_{k=1}^{p-1} \sigma_k^2.
\]

During the minimization process of the functions \( M_1, M_2 \) (or \( M_3 \)) and \( M_4 \), we have obtained the following:

- the unique \( \bar{x}_{ij} \) lying on the diagonals \( d_k, k = -m+1, -m+p-1 \);
- the unique \( \bar{x}_{ij} \) lying on the diagonals \( d_k, k = -m+p, n-p-1 \);
- the elements on each of diagonals \( d_k, k = n-p, n-1 \), depend on one real parameter, and we assume this element is from the first row in the case \( m \geq n \), and in the last column if \( m \leq n \).

If we rearrange the matrix \( \bar{X} \) whose elements are known, we obtain precisely (7). Such rearrangement looks rather cumbersome in general case, and some insight can be brought after looking at Examples 6.1 and 6.2.

Let us prove that any \( \bar{X} = X_0 + X_p + X_c \) given by (7) is indeed a least-square solution:

\[
J_m(0)X - XJ_n(0) - C = (J_m(0)X_0 - X_0J_m(0)) + (J_m(0)X_p - X_pJ_m(0)) - C + J_m(0)X_c - X_cJ_m(0) = \\
= (J_m(0)X_p - X_pJ_m(0) - C) + J_m(0)X_c - X_cJ_m(0)
\]

It is not hard to check that \( J_m(0)X_p - X_pJ_m(0) - C \) is a matrix whose all entries are zero, except the \( m \)-th row (case \( m \geq n \)) or the first column (case \( m \leq n \)):

\[
J_m(0)X_p - X_pJ_m(0) - C = \begin{bmatrix}
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 \\
-\sigma_1 & -\sigma_2 & \ldots & -\sigma_n
\end{bmatrix} \text{ or } \begin{bmatrix}
-\sigma_n & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-\sigma_n & 0 & \ldots & 0 \\
-\sigma_n & 0 & \ldots & 0
\end{bmatrix}.
\]

On the similar way it can be shown that for \( m \geq n \)

\[
J_m(0)X_c - X_cJ_m(0) = \\
\begin{bmatrix}
0 & 0 & \ldots & 0 & 0 \\
-\sigma_n/n & 0 & \ldots & \ldots & \ldots \\
-\sigma_{n-1}/(n-1) & -\sigma_n/n & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
-\sigma_3/3 & -\sigma_4/4 & \ldots & \ldots & -\sigma_n/n \\
-\sigma_2/2 & -\sigma_3/3 & -\sigma_4/4 & \ldots & -\sigma_n/n \\
0 & \sigma_2/2 & 2\sigma_3/3 & (n-2)\sigma_{n-1}/(n-1) & (n-1)\sigma_n/n
\end{bmatrix},
\]
(and just transposed matrix if \(m \leq n\)), so we conclude that

\[
J_m(0)X - XJ_n(0) - C = \begin{bmatrix}
\sigma_n/n & 0 & \cdots & 0 \\
\sigma_{n-1}/(n-1) & \sigma_n/n & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
\sigma_2/2 & \sigma_3/3 & \cdots & \sigma_n/n \\
\sigma_1 & \sigma_2/2 & \cdots & \sigma_{n-1}/(n-1) & \sigma_n/n
\end{bmatrix} - \begin{bmatrix}
0_{(m-n)\times n} \\
0_{n\times (n-1)} \\
0_{(n-1)\times n}
\end{bmatrix}
\]

if \(m \geq n\), and transpose of this matrix if \(m \leq n\). The Frobenius norm of this matrix is

\[
||J_m(0)X - XJ_n(0) - C||_F^2 = \sum_{k=1}^{n} k^2 \left( \frac{\sigma_k}{k} \right)^2 = \sum_{k=1}^{n} \frac{\sigma_k^2}{k},
\]

so we have the proof. \(\Box\)

Corollary 5.1. Sylvester equation \(J_m(0)X - XJ_n(0) = C\) is consistent if and only if \(\sigma_k = 0, k = 1, p\), i.e. iff \(X_c = 0\), and its general solution is given by (7).

Remark that this result agrees with Theorems 2.3 and 2.5 from [7].

Corollary 5.2. The least-squares solution for the equation \(J_n(0)X - XJ_n(0) = I_n\) is

\[
\hat{X} = p_{n-1}(J_n(0)),
\]

for any complex polynomial \(p_{n-1}\), and \(||J_n(0)\hat{X} - XJ_n(0) - I_n||_F^2 = n\).

6. Minimal-norm Least-squares Solution for the Simplest Case

In the previous section, we observed that there is a whole class of the least-squares solutions, depending on some parameters, whether or not the equation is consistent. Now we prove that those free parameters can be chosen such that the solution is with minimal Frobenius norm. If we look at the structure of matrices \(X_h\) and \(X_c\) from Theorem 5.1, we see that they do not have any non-zero entry on the same position (for \(m \geq n\) matrix \(X_h\) is upper triangular, while \(X_c\) is strictly lower triangular; the case \(m \leq n\) is just transposed situation), therefore only matrix \(X_p\) may have the influence on the minimization. So far, we concluded that

\[
\min_{\hat{X} \in S} ||\hat{X}||_F^2 = \min_{X_h} ||X_h + X_p + X_c||_F^2 = ||X_h||_F^2 + ||X_p||_F^2 + ||X_c||_F^2.
\]

Remark that \(||X_c||_F\) is a constant, which is zero iff the equation is consistent and strictly positive otherwise.

Theorem 6.1. There is a unique least-squares minimal-norm solution for the Sylvester equation \(J_m(0)X - XJ_n(0) = C\), given by

\[
\hat{X}_0 = \hat{X}_h + \hat{X}_p + \hat{X}_c,
\]

where

\[
\hat{X}_h = \begin{cases}
\begin{bmatrix}
p^*_{n-1}(J_n(0)) \\
0_{(m-n)\times n}
\end{bmatrix}, & m \geq n, \\
\begin{bmatrix}
0_{mx(m-n)} \\
q^*_{m-1}(J_m(0))
\end{bmatrix}, & m \leq n,
\end{cases}
\]

and \(p^*_{n-1}\) and \(q^*_{m-1}\) are uniquely determined polynomials.
Proof. According to the Theorem 5.1, such $\tilde{X}_0$ is a least-squares solution. We must show that it has minimal norm among all least-squares solutions.

Let us consider the matrix $T = X_h + X_p = [t_{ij}]_{m \times n}$. The only entries that can be minimized include $x_1, \ldots, x_n$ (when $m \geq n$) or $x_{1m}, \ldots, x_{nm}$ (when $m \leq n$), and they are precisely along the small diagonals from $n - p$ to $n - 1$:

$$
\|T\|_F^2 = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-p} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-p} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-p} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}}
$$

$$
|\tilde{X}_0| = \min_{X \in S} |X_0| = \min_{X \in S} |X_h + X_p|^2 + |X|^2 = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}} + \sum_{k=n-p}^{n-1} \sum_{i,j=1}^{t_{ij}} = \sum_{k=-m+1}^{n-1} \sum_{i,j=1}^{t_{ij}}
$$

Therefore, for such $x_{i}', i = 1, n$ (when $m \geq n$), or $x_{im}'$, $i = 1, m$ (when $m \leq n$), which is given by Proposition 8.2, we obtained the unique least-squares minimum-norm solution. Since it is already known that homogeneous solution is block-polynomial matrix, we have the result. \( \square \)

Corollary 6.1. There is unique minimal-norm solution for consistent Sylvester equation, and it is given by $\overline{X}_0 = \overline{X}_h + X_p$, with notations as in the previous Theorem.

In order to clarify and illustrate the constructions given in previous two theorems about least-squares and minimal-norm solutions, we present the following two in-detailed examples. The first is dealing with case when $m > n$, and the second one with the case $m \leq n$.

Example 6.1. Let us find the minimum-norm least-squares solution of the equation $J_4(0)X - XJ_3(0) = C$, $C \in \mathbb{R}^{4 \times 3}$, i.e. we want to find all $X \in \mathbb{R}^{4 \times 3}$ such that $\|\Delta(X)\|_F^2 = \|J_4(0)X - XJ_3(0) - C\|_F^2$ is minimal, and among them such $X$ which is minimal-Frobenius-norm matrix.

We can arrange the small diagonals of matrix $C$ as follows (below the matrix there are the indices for small diagonals):

$$
\begin{bmatrix}
   c_{21} & c_{11} & c_{12} & c_{13} \\
   c_{31} & c_{32} & c_{22} & c_{23} \\
   c_{41} & c_{42} & c_{43} & c_{33} \\
   -3 & -2 & -1 & 0 & 1 & 2
\end{bmatrix},
$$

and let us denote by $a_k$, $k = 1, 3$, sum over the $(4 + k)$–th small diagonal ($k = \overline{3, 2}$) of the matrix $C$, i.e.

$$
a_1 = c_{41}, \ a_2 = c_{31} + c_{42}, \ a_3 = c_{21} + c_{32} + c_{43}.
$$

We can arrange the small diagonal for the matrix $J_4(0)X - XJ_3(0) - C$ as well:

$$
\begin{bmatrix}
   x_{31} - c_{31} & x_{32} - c_{32} & x_{33} - c_{33} \\
   x_{41} - c_{41} & x_{42} - c_{42} & x_{43} - c_{43} \\
   -c_{41} - x_{41} - c_{42} & -x_{42} - c_{43} & x_{43} - x_{32} - c_{33}
\end{bmatrix}, \begin{bmatrix}
   x_{21} - c_{21} & x_{22} - c_{22} & x_{23} - c_{23} \\
   x_{11} - c_{11} & x_{12} - c_{12} & x_{13} - c_{13}
\end{bmatrix}, \begin{bmatrix}
   x_{21} - c_{21} & x_{22} - c_{22} & x_{23} - c_{23} \\
   x_{11} - c_{11} & x_{12} - c_{12} & x_{13} - c_{13}
\end{bmatrix}, \begin{bmatrix}
   x_{21} - c_{21} & x_{22} - c_{22} & x_{23} - c_{23} \\
   x_{11} - c_{11} & x_{12} - c_{12} & x_{13} - c_{13}
\end{bmatrix}
$$
Vertical bars partition the small diagonals according to the entries they have. By using the properties of the Frobenius norm, it is clear that \( ||A(0)X - B(0) - C ||_F^2 \) can be obtained as a sum of squared entries over each of the columns. For left submatrix, those sums are precisely the function \( M_1 \) applied to all its columns except the first one (we enlist unknown variables \( x_{ij} \) and parameters \( c_{ij} \) from top to bottom); for central submatrix we have function \( M_2 \), while for right submatrix function \( M_4 \) is applied for each of its columns. Hence, we have

\[
H(x_{11}, ..., x_{43}) = (-c_{41})^2 + ((x_{41} - c_{31})^2 + (-x_{41} - c_{42})^2) + \\
+ ((x_{31} - c_{21})^2 + (x_{42} - x_{31} - c_{32})^2 + (-x_{42} - c_{43})^2) + \\
+ ((x_{21} - c_{11})^2 + (x_{32} - x_{21} - c_{22})^2 + (x_{43} - x_{32} - c_{33})^2) + \\
+ ((x_{22} - x_{11} - c_{12})^2 + (x_{33} - x_{22} - c_{23})^2) + (x_{23} - x_{12} - c_{13})^2 = \\
= c_{41}^2 + M_1(x_{41}; c_{31}, c_{42}) + M_2(x_{31}, x_{42}; c_{21}, c_{32}, c_{43}) + \\
+ M_3(x_{21}, x_{32}, x_{43}; c_{11}, c_{22}, c_{33}) + \\
+ M_4(x_{11}, x_{22}, x_{33}; c_{12}, c_{23}) + M_4(x_{12}, x_{23}; c_{13}).
\]

By the Theorem 8.1, we have unique \( \tilde{x}_{41}, \tilde{x}_{31}, \tilde{x}_{32}, \tilde{x}_{41}, \tilde{x}_{42}, \tilde{x}_{43} \), given by:

\[
\tilde{x}_{41} = \frac{1}{2}(c_{31} - c_{42}), \\
\tilde{x}_{31} = \frac{1}{3}(2c_{21} - c_{32} - c_{43}), \quad \tilde{x}_{42} = \frac{1}{3}(c_{21} + c_{32} - 2c_{43}), \\
\tilde{x}_{21} = c_{11}, \quad \tilde{x}_{32} = c_{11} + c_{22}, \quad \tilde{x}_{43} = c_{11} + c_{22} + c_{33},
\]

such that \( M_1(\tilde{x}_{41}; c_{31}, c_{42}) = \sigma_1^2/2 \) and \( M_1(\tilde{x}_{31}, \tilde{x}_{42}; c_{21}, c_{32}, c_{43}) = \sigma_2^2/3 \), while \( \tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{13}, \tilde{x}_{22}, \tilde{x}_{33}, \tilde{x}_{33} \) are depending of parameters denoted by \( x_{11}, x_{12}, x_{13} \):

\[
\tilde{x}_{11} = x_{11}, \quad \tilde{x}_{22} = x_{11} + c_{12}, \quad \tilde{x}_{33} = x_{11} + c_{12} + c_{23}, \\
\tilde{x}_{12} = x_{12}, \quad \tilde{x}_{23} = x_{12} + c_{13}, \\
\tilde{x}_{13} = x_{13},
\]

for them minima of all \( M_2 \) and \( M_4 \) are zeros. Therefore, the minimum of the function \( H \) is:

\[
H(\tilde{x}_{11}, ..., \tilde{x}_{43}) = \sigma_1^2 + \frac{\sigma_2^2}{2} + \frac{\sigma_3^2}{3} = \sum_{k=1}^{3} \frac{\sigma_k^2}{k}.
\]

It is clear that \( H(\tilde{x}_{11}, ..., \tilde{x}_{43}) = 0 \iff \sigma_k = 0, \quad k = 1, 3, \) which is precisely the consistency condition (Theorem 2.3 from [7]). If we decompose such \( \tilde{X} \) as:

\[
\tilde{X} = \begin{bmatrix}
    x_{11} & x_{12} & x_{13} \\
    c_{11} & x_{11} + c_{12} & x_{12} + c_{13} \\
    \frac{1}{2}(2c_{21} - c_{32} - c_{43}) & c_{11} + c_{22} & x_{11} + c_{12} + c_{23} \\
    \frac{1}{2}(c_{31} - c_{42}) & \frac{1}{2}(c_{31} + c_{32} - 2c_{43}) & c_{11} + c_{22} + c_{33}
\end{bmatrix} = \\
= x_b + x_p + x_c,
\]

Filomat 33:13 (2019), 4261–4280
where $X_b$, $X_p$ and $X_c$ are as in the Theorem 7. Also, remark that:

\[ J_4(0)\tilde{X} - XJ_5(0) - C = \begin{bmatrix} 0 & 0 & 0 \\ -\frac{\sigma_2}{3} & 0 & 0 \\ -\frac{\sigma_3}{3} & -\frac{\sigma_2}{3} & 0 \\ -\sigma_1 & -\frac{\sigma_2}{3} & -\frac{\sigma_3}{3} \end{bmatrix} = -\frac{0_{1x3}}{p_2(J_3(0))}, \]

where $p_2(t) = \sigma_3/3 + \sigma_2/2 \cdot t + \sigma_1 \cdot t^2$.

In order to obtain minimum-norm solution among all least-squares solutions, let us arrange small diagonals of $\tilde{X}$ as follows (vertical bars separate those small diagonals which take parts in minimization from the others):

\[
\begin{bmatrix}
    c_{21} - \frac{\sigma_3}{3} & c_{11} & x_{11} & x_{12} & x_{13} \\
    c_{31} - \frac{\sigma_2}{2} & c_{21} + c_{22} - \frac{2\sigma_3}{3} & c_{11} + c_{22} + c_{33} & x_{11} + c_{12} + c_{23} & x_{11} + c_{12} + c_{23} \\
    c_{31} & c_{32} & c_{33} & 0 & 0 \\
    c_{31} & c_{32} & c_{33} & 0 & 0
\end{bmatrix}.
\]

There are three parameters, $x_{11}, x_{12}$ and $x_{13}$, which can be used for the minimization of sums over the columns of right submatrix. In order to obtain such solution, we should minimize the following function (by using the Theorem 8.2) $h : \mathbb{R}^3 \to \mathbb{R}$:

\[
h(x_{11}, x_{12}, x_{13}) = x_{11}^2 + (x_{11} + c_{12})^2 + (x_{11} + c_{12} + c_{23})^2 + x_{12}^2 + (x_{12} + c_{13})^2 + x_{13}^2 = \\
F(x_{11}; c_{12}, c_{12} + c_{23}) + F(x_{12}; c_{13}) + x_{13}^2.
\]

According to Theorem 8.2, the minimum is obtained for

\[
x^*_{11} = -\frac{2c_{12} + c_{23}}{3}, \quad x^*_{12} = -\frac{c_{13}}{2}, \quad x^*_{13} = 0,
\]

and it is $h(x^*_{11}, x^*_{12}, x^*_{13}) = (3c_{13}^2 + 4(c_{12}^2 + c_{12}c_{23} + c_{23}^2))/6$.

Therefore, the least-squares minimal-norm solution is:

\[
\tilde{X}_0 = \begin{bmatrix}
    -2 & -1 & 0 & 1 & 2 & 3 \\
    c_{11} & c_{12} & c_{13} & c_{14} \\
    c_{21} & c_{22} & c_{23} & c_{24} \\
    c_{31} & c_{32} & c_{33} & c_{34}
\end{bmatrix}.
\]

Note that the solution depends on all entries of the matrix $C$ except $c_{41}$.

**Example 6.2.** Let us find the minimum-norm least-squares solution of the equation $J_3(0)X - XJ_4(0) = C$, $C \in \mathbb{R}^{3x4}$, i.e. we want to find all $X \in \mathbb{R}^{3x4}$ such that $\|\Delta(X)\|^2 = \|J_3(0)X - XJ_4(0) - C\|^2$ is minimal, and among them such $X$ which is minimal-Frobenius-norm matrix.

We can arrange the small diagonals of matrix $C$ as follows (below the matrix there are the indices for small diagonals):

\[
\begin{bmatrix}
    c_{11} & c_{12} & c_{13} & c_{14} \\
    c_{21} & c_{22} & c_{23} & c_{24} \\
    c_{31} & c_{32} & c_{33} & c_{34}
\end{bmatrix}.
\]

and let us denote by $\sigma_k$, $k = 1,3$, sum over the $(3 + k)$–th small diagonal ($k = -2,3$) of the matrix $C$, i.e.

\[
\sigma_1 = c_{31}, \quad \sigma_2 = c_{21} + c_{32}, \quad \sigma_3 = c_{11} + c_{22} + c_{33}.
\]
We can arrange the small diagonal for the matrix $J_3(0)X - XJ_4(0) - C$ as well:

\[
\begin{bmatrix}
    x_{31} - c_{21} & x_{32} - x_{21} - c_{22} & x_{33} - x_{22} - c_{23} & x_{34} - x_{23} - c_{24} \\
    x_{31} - c_{21} & -c_{31} & -x_{32} - c_{32} & -x_{33} - c_{34} \\
    x_{21} - c_{11} & x_{22} - x_{11} - c_{12} & x_{23} - x_{12} - c_{13} & x_{24} - x_{13} - c_{14} \\
    -c_{21} & x_{31} - c_{11} & x_{32} - x_{12} - c_{13} & x_{34} - x_{13} - c_{14}
\end{bmatrix}
\]

Vertical bars partition the small diagonals according to the entries they have. By using the properties of the Frobenius norm, it is clear that $\|J_3(0)X - XJ_4(0) - C\|_F^2$ can be obtained as a sum of squared entries over each of the columns. For left submatrix, those sums are precisely the function $M_1$ applied to all its columns except the first one (we enlist unknown variables $x_{ij}$ and parameters $c_{ij}$ from top to the bottom); for central submatrix we have function $M_2$, while for right submatrix function $M_4$ is applied for each of its columns. Hence, we have

\[
H(x_{11}, ..., x_{34}) = (-c_{41})^2 + (x_{31} - c_{21})^2 + (x_{31} - c_{23})^2 + (x_{32} - x_{21} - c_{22})^2 + (x_{32} - x_{23} - c_{24})^2 + (x_{33} - x_{23} - c_{23})^2 + (x_{34} - x_{24} - c_{24})^2 + (x_{33} - c_{33})^2 + (x_{34} - c_{34})^2 =
\]

By Theorem 8.1, we have unique $\bar{x}_{11}, \bar{x}_{21}, \bar{x}_{22}, \bar{x}_{31}, \bar{x}_{32}, \bar{x}_{33}$, given by:

\[
\begin{align*}
\bar{x}_{31} & = \frac{1}{2}(c_{21} - c_{32}), \\
\bar{x}_{21} & = \frac{1}{3}(2c_{11} - c_{22} - c_{33}), \\
\bar{x}_{32} & = \frac{1}{3}(c_{11} + c_{22} - 2c_{33}), \\
\bar{x}_{11} & = -(c_{12} + c_{23} + c_{34}), \\
\bar{x}_{22} & = -(c_{23} + c_{34}), \\
\bar{x}_{33} & = -c_{34},
\end{align*}
\]

such that $M_1(\bar{x}_{31}; c_{21}, c_{32}) = \sigma_1^2/2$ and $M_1(\bar{x}_{21}; \bar{x}_{32}; c_{11}, c_{22}, c_{33}) = \sigma_2^2/3$, while $\bar{x}_{12}, \bar{x}_{13}, \bar{x}_{14}, \bar{x}_{23}, \bar{x}_{24}, \bar{x}_{34}$ are depending of parameters denoted by $x_{14}, x_{24}, x_{34}$ (this is important difference from the previous example!):

\[
\begin{align*}
\bar{x}_{34} & = x_{34}, \\
\bar{x}_{23} & = x_{34} - c_{24}, \\
\bar{x}_{12} & = x_{34} - c_{13} - c_{24}, \\
\bar{x}_{24} & = x_{24}, \\
\bar{x}_{13} & = x_{24} - c_{14}, \\
\bar{x}_{14} & = x_{14},
\end{align*}
\]

for them minima of all $M_3$ and $M_4$ are zeros. Therefore, the minimum of the function $H$ is:

\[
H(\bar{x}_{11}, ..., \bar{x}_{34}) = \sigma_1^2 + \frac{\sigma_2^2}{2} + \frac{\sigma_3^2}{3} = \sum_{k=1}^{3} \frac{\sigma_k^2}{k}.
\]

It is clear that $H(\bar{x}_{11}, ..., \bar{x}_{34}) = 0 \Leftrightarrow \sigma_k = 0$, $k = 1, 3$, which is precisely the consistency condition (Theorem 2.5 from [7]). If we decompose such $\bar{X}$ as:

\[
\bar{X} = \begin{bmatrix}
    (-(c_{12} + c_{23} + c_{34}) & x_{34} - c_{13} - c_{24} & x_{24} - c_{14} & x_{14} \\
    \frac{1}{3}(2c_{11} - c_{22} - c_{33}) & -(c_{23} + c_{34}) & x_{34} - c_{24} & x_{24} \\
    \frac{1}{2}(c_{11} + c_{22} - 2c_{33}) & \frac{1}{2}(c_{11} + c_{22} - 2c_{33}) & -(c_{23} + 3c_{34}) & x_{34}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    0 & x_{34} & x_{24} & x_{14} \\
    0 & 0 & x_{34} & x_{24} \\
    0 & 0 & 0 & x_{34}
\end{bmatrix}
\]

\[
= X_b + X_p + X_c.
\]
where \( X_0, X_p \) and \( X_c \) are as in the Theorem 7. Also, remark that:

\[
    f_3(0)\tilde{X} - \tilde{X}f_4(0) - C = \begin{bmatrix}
    \frac{-\sigma_3}{2} & 0 & 0 & 0 \\
    \frac{-\sigma_2}{2} & \frac{-\sigma_3}{2} & 0 & 0 \\
    \sigma_1 & \frac{-\sigma_2}{2} & \frac{-\sigma_3}{2} & 0 \\
    -\sigma_1 & \frac{-\sigma_2}{2} & \frac{-\sigma_3}{2} & 0
    \end{bmatrix} = -\left[ p_2(f_3(0))^T \mid 0_{3 \times 1} \right],
\]

where \( p_2(t) = \sigma_3/3 + \sigma_2/2 \cdot t + \sigma_1 \cdot t^2 \).

In order to obtain minimum-norm solution among all least-squares solutions, let us arrange small diagonals of \( X \) as follows (vertical bars separate those small diagonals which take parts in minimization from the others):

\[
    \begin{bmatrix}
        \sigma_3/3 - c_{12} & (c_{12} + c_{23} + c_{34})/3 & x_{34} - c_{13} - c_{24} & x_{24} - c_{14} & x_{14} \\
        (c_{23} + c_{34})/3 & -c_{34} & x_{34} - c_{24} & x_{24} & x_{14} \\
        x_{34} & x_{34} & x_{34} & x_{34} & x_{34}
    \end{bmatrix}
\]

There are three parameters, \( x_{14}, x_{24}, x_{34} \), which can be used for the minimization of sums over the columns of right submatrix. In order to obtain such solution, we should minimize the following function (by using the Theorem 8.2) \( h: \mathbb{R}^3 \rightarrow \mathbb{R} \):

\[
    h(x_{14}, x_{24}, x_{34}) = x_{34}^2 + (x_{34} - c_{24})^2 + (x_{34} - c_{24} - c_{13})^2 + x_{24}^2 + (x_{24} - c_{14})^2 + x_{14}^2 = G(x_{34}; -c_{24}, -c_{13}) + G(x_{24}; -c_{14}) + x_{14}^2.
\]

According to Theorem 8.2, the minimum is obtained for

\[
    x_{34}^* = \frac{2c_{24} + c_{13}}{3}, \quad x_{24}^* = \frac{c_{14}}{2}, \quad x_{14}^* = 0,
\]

and it is \( h(x_{14}^*, x_{24}^*, x_{34}^*) = (3c_{14}^2 + 4(c_{24}^2 + c_{24}c_{13} + c_{14}^2))/6 \).

Therefore, the least-squares minimal-norm solution is:

\[
    \tilde{X}_0 = \begin{bmatrix}
        -(c_{12} + c_{23} + c_{34}) & -(2c_{24} + c_{13})/3 & -c_{14}/2 & 0 \\
        (2c_{11} - c_{22} - c_{13})/3 & -(c_{13} - c_{24})/3 & c_{14}/2 & 0 \\
        (c_{21} - c_{32})/2 & (c_{11} + c_{22} - 2c_{33})/3 & -c_{34} & (c_{13} + 2c_{24})/3 \\
    \end{bmatrix}.
\]

Note that the solution depends on all entries of the matrix \( C \) except \( c_{31} \).

7. Return to the Main Case

We have thoroughly examined the simplest case, so we can return to our original problem.

\[
    \min_X \|AX - XB - C\|_F^2 \leq \alpha^2(S, T) \min_Y \|J_A Y - Y f_B - D\|_F^2 =
\]

\[
    = \alpha^2(S, T) \sum_{i=1}^{k} \sum_{u=1}^{k} \sum_{v=1}^{k} \min_{Y} \|f_u(0)Y_{uv}^{(i)} - Y_{uv}^{(0)} f_v(0) - D_{uv}^{(0)}\|_F^2 =
\]

\[
    = \alpha^2(S, T) \sum_{i=1}^{\bar{t}} \sum_{u=1}^{k} \sum_{v=1}^{k} \sum_{k=1}^{\bar{s}} \frac{\sigma_k(D_{uv}^{(i)})}{\bar{s}}.
\]

If the right hand side is zero, i.e.

\[
    \sigma_k(D_{uv}^{(i)}) = 0, \quad k = 1, \min[u, \bar{v}], \quad i = 1, \bar{s},
\]
then the equation is consistent. This result is independent of the choice of matrices $S$ and $T$ (except the request that proper Jordan blocks should be on appropriate positions).

If the right hand side is not zero, at least we have an upper bound (which need not be the best possible!) for the error.

Remark that if $A = S J_A S^{-1}$, then $A = (y S) J_A (y S)^{-1}$ for any $y \neq 0$, so one can further analyze quantity $a(S, T)$.

8. Appendix: Some Auxiliary Results on the Minimum of the Functions

Suppose that $C \in \mathbb{C}^{m \times n}$ is a complex matrix. It can be written on the unique way as $C = C' + i C''$, where $C', C'' \in \mathbb{R}^{m \times n}$. Note that $||C||_F^2 = ||C'||_F^2 + ||C''||_F^2$. Because of:

$$\Delta(X; C) : = ||J_m(0)X - X J_n(0) - C||_F^2 =$$
$$= ||J_m(0)(X' + iX'') - (X' + iX'') J_n(0) - (C' + iC'')||_F^2 =$$
$$= ||J_m(0)X' - X' J_n(0) - C' + i(J_m(0)X'' - X'' J_n(0) - C'')||_F^2 =$$
$$= ||J_m(0)X' - X' J_n(0) - C'||_F^2 + ||J_m(0)X'' - X'' J_n(0) - C''||_F^2 =$$
$$= \Delta(X'; C') + \Delta(X''; C''),$$

it is enough to consider minimization of appropriate real function.

The following two results can be easily proven by elementary calculations, but since they are the backbone for Theorems 5.1–6.1, we formulate them as propositions.

**Proposition 8.1.** 1) The function $M_1 : \mathbb{R}^n \to \mathbb{R}$ depending on $n + 1$ real parameters $a_1, \ldots, a_{n+1}$,

$$M_1(x_1, \ldots, x_n; a_1, \ldots, a_n, a_{n+1}) = (x_1 - a_1)^2 + \sum_{k=1}^{n+1} (x_{k+1} - x_k - a_{k+1})^2 + (x_n + a_{n+1})^2$$

attains its minimum for the uniquely determined arguments

$$\bar{x}_k = \frac{1}{n+1} \left( n+1-k \right) \sum_{i=1}^{k} a_i - k \sum_{i=k+1}^{n+1} a_i = \sum_{i=1}^{k} a_i - \frac{k}{n+1} \sum_{i=1}^{n+1} a_i, \quad k = 1, n,$$

and this minimal value is

$$M_1(\bar{x}_1, \ldots, \bar{x}_n; a_1, \ldots, a_{n+1}) = \frac{1}{n+1} \left( \sum_{i=1}^{n+1} a_i \right)^2.$$

2) The function $M_2 : \mathbb{R}^n \to \mathbb{R}$, depending on $n$ real parameters $a_1, \ldots, a_n$,

$$M_2(x_1, \ldots, x_n; a_1, \ldots, a_n) = (x_1 - a_1)^2 + \sum_{k=1}^{n} (x_{k+1} - x_k - a_{k+1})^2$$

attains its minimum 0 for the uniquely determined arguments

$$\bar{x}_k = \sum_{i=1}^{k} a_i, \quad k = 1, n.$$

3) The function $M_3 : \mathbb{R}^n \to \mathbb{R}$, depending on $n$ real parameters $a_1, \ldots, a_n$,

$$M_3(x_1, \ldots, x_n; a_1, \ldots, a_n) = \sum_{k=1}^{n} (x_{k+1} - x_k - a_k)^2 + (x_n + a_n)^2$$

...
attains its minimum 0 for the uniquely determined arguments

\[ \hat{x}_k = - \sum_{i=k}^{n} a_i, \quad k = \frac{1}{2} n. \]

4) The function \( M_4 : \mathbb{R}^n \rightarrow \mathbb{R} \), depending on \( n - 1 \) real parameters \( a_1, ..., a_{n-1} \),

\[ M_4(x_1, ..., x_n; a_1, ..., a_{n-1}) = \sum_{k=1}^{n-1} (x_{k+1} - x_k - a_k)^2 \]

attains its minimum 0 for the one-parameter set of the arguments

\[ \hat{x}_k = x_1 + \sum_{i=k}^{n} a_i - 1, \quad k = \frac{1}{2} n, \]

or

\[ \hat{x}_k = x_n - \sum_{i=k}^{n-1} a_i, \quad k = \frac{3}{2} n - 1. \]

**Proposition 8.2.** 1) The function \( F : \mathbb{R} \rightarrow \mathbb{R} \), depending on \( n \) real parameters \( a_1, ..., a_n \):

\[ F(x; a_1, ..., a_n) = x^2 + (x + a_1)^2 + ... + (x + a_n)^2, \]

attains its minimum at

\[ x^* = - \frac{a_1 + ... + a_n}{n + 1}, \]

and this minimal value is

\[ F(x^*; a_1, ..., a_n) = \sum_{k=1}^{n} a_k^2 - \frac{1}{n + 1} \left( \sum_{k=1}^{n} a_k \right)^2 = \frac{1}{n + 1} \left( n \sum_{j=1}^{n} a_j^2 - 2 \sum_{j<k} a_j a_k \right). \]

2) The function \( G(x; a_1, ..., a_n) = F(x; a_1, a_1 + a_2, ..., a_1 + a_2 + ... + a_n) \), depending on \( n \) real parameters \( a_1, ..., a_n \),

attains its minimum for

\[ x^* = - \sum_{j=1}^{n} \left( \frac{n + 1 - j}{n + 1} \right) a_j, \]

and

\[ G(x^*; a_1, ..., a_n) = \sum_{k=1}^{n} \left( \sum_{j=1}^{k} a_j \right)^2 - \frac{1}{n + 1} \left( \sum_{k=1}^{n} \sum_{j=1}^{k} a_j \right)^2. \]

**References**

[1] S. Albeverio, A. K. Motovilov and A. A. Shkalikov, *Bounds on variation of spectral subspaces under \( J \)-self-adjoint perturbations*, Int. Equ. Oper. Theory 64 (2009), 455–486.

[2] W. Arendt, F. Rabiger and A. Sourour, *Spectral properties of the operator equation \( AX + XB = Y \)*, Quart. J. Math. Oxford (2) 45 (1994), 133–149.

[3] A. Ben-Israel and T. N. E. Greville, *Generalized inverses, theory and applications*, 2nd edition, Springer, 2003.

[4] P. Benner, *Factorized solutions of Sylvester equations with applications in control*, in Proc. of the 16th International Symposium on Mathematical Theory of Network and Systems (MTNS 2004), 2004.

[5] R. Bhatia and P. Rosenthal, *How and why to solve the operator equation \( AX - XB = Y \)*, Bull. London Math. Soc. 29 (1997), 1–21.

[6] R. Byers, *Solving the algebraic Ricatti equation with the matrix sign function*, Linear Algebra Appl. 85 (1987), 267–279.

[7] N. Č. Dinčić, *Solving the Sylvester equation \( AX - XB = C \) when \( \sigma(A) \cap \sigma(B) \neq \emptyset \)*, Electron. J. Linear Algebra 35 (2019), 1–23.
[8] B. D. Djordjević and N. Ć. Dinčić, Solving the operator equation $AX - XB = C$ with closed $A$ and $B$, Integral Equation and Operator Theory (2018) 90:51
[9] M. P. Drazin, On a result of J. J. Sylvester, Linear Algebra Appl. 505 (2016), 361–366.
[10] F. R. Gantmacher, The theory of matrices, vol 1, Chelsea Pub. Co., 1959.
[11] R. E. Hartwig and P. Patricio, Group inverse of the nivellateur, Linear and Multilinear Algebra 66 (12) (2018), 2409–2420
[12] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, 1985.
[13] Q. Hu, D. Cheng, The polynomial solution to the Sylvester matrix equation, Applied Mathematics Letters 19 (9) (2006), 859–864.
[14] D. Y. Hu, L. Reichel, Krylov-subspace methods for the Sylvester equation, Linear Algebra Appl. 172 (1992), 283–313.
[15] X. Jin, Y. Wei, Numerical linear algebra and its applications, SCIENCE PRESS USA Inc. Beijing, 2005.
[16] N. T. Lan, On the operator equation $AX - XB = C$ with unbounded operators $A$, $B$ and $C$, Abstract and Applied Analysis 6 (6) (2001), 317–328.
[17] G. Lumer, M. Rosenblum, Linear operator equations, Proc. Amer. Math. Soc. 10 (1959), 32–41.
[18] E-C. Ma, A finite series solution of the matrix equation $AX - XB = C$, SIAM J. Appl. Math. 14 (3) (1966), 490–495.
[19] S. Mecheri, Why we solve the operator equation $AX - XB = C$, preprint, (www.researchgate.net/publication/240626459_Why_we_solve_the_operator_equation_AX XB_C)
[20] V. Q. Phong, The operator equation $AX - XB = C$ with unbounded operators $A$ and $B$ and related abstract Cauchy problems, Math. Z. 208 (1991), 567–588.
[21] J. D. Roberts, Linear model reduction and solution of the algebraic Ricatti equation by use of the sign function, Internat. J. Control 32 (1980), 677–687.
[22] M. Rosenblum, On the operator equation $BX -XA = Q$, Duke Math. J. 23 (1956), 263–270
[23] J. J. Sylvester, Sur l'équation en matrices $px = qx$, C. R. Acad. Sci. Paris, 99 (1884) 67–71 and 115–116.