

Bit Complexity of Jordan Normal Form and Polynomial Spectral Factorization

Papri Dey ✉

Georgia Tech, Atlanta, GA, USA

Ravi Kannan ✉

Microsoft Research, Bangalore, India

Nick Ryder ✉

OpenAI, San Francisco, CA, USA

Nikhil Srivastava ✉

UC Berkeley, CA, USA

Abstract

We study the bit complexity of two related fundamental computational problems in linear algebra and control theory. Our results are: (1) An $\tilde{O}(n^{\omega+3}a + n^4a^2 + n^\omega \log(1/\epsilon))$ time algorithm for finding an ϵ -approximation to the Jordan Normal form of an integer matrix with a -bit entries, where ω is the exponent of matrix multiplication. (2) An $\tilde{O}(n^6d^6a + n^4d^4a^2 + n^3d^3 \log(1/\epsilon))$ time algorithm for ϵ -approximately computing the spectral factorization $P(x) = Q^*(x)Q(x)$ of a given monic $n \times n$ rational matrix polynomial of degree $2d$ with rational a -bit coefficients having a -bit common denominators, which satisfies $P(x) \succeq 0$ for all real x . The first algorithm is used as a subroutine in the second one.

Despite its being of central importance, polynomial complexity bounds were not previously known for spectral factorization, and for Jordan form the best previous best running time was an unspecified polynomial in n of degree at least twelve [8]. Our algorithms are simple and judiciously combine techniques from numerical and symbolic computation, yielding significant advantages over either approach by itself.

2012 ACM Subject Classification Mathematics of computing → Numerical analysis

Keywords and phrases Symbolic algorithms, numerical algorithms, linear algebra

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.42

Related Version *Full Version*: <https://arxiv.org/abs/2109.13956>

Funding *Nikhil Srivastava*: Supported by NSF Grant CCF-2009011.

Acknowledgements We thank the anonymous referees of a previous version of this paper, whose thoughtful comments greatly improved the presentation. We thank Bill Helton, Clément Pernet, Pablo Parrilo, Mario Kummer, Rafael Oliveira, and Rainer Sinn for helpful discussions, as well as the Simons Institute for the Theory of Computing, where a large part of this work was carried out during the “Geometry of Polynomials” program.

1 Introduction

We study the bit complexity of finding approximate solutions to the following problems, where the input is assumed to be given exactly as a (complex) rational matrix or collection of matrices.



© Papri Dey, Ravi Kannan, Nick Ryder, and Nikhil Srivastava;
licensed under Creative Commons License CC-BY 4.0

14th Innovations in Theoretical Computer Science Conference (ITCS 2023).

Editor: Yael Tauman Kalai; Article No. 42; pp. 42:1–42:18



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1. **Jordan Normal Form.** Given $A \in \mathbb{C}^{n \times n}$, find a similarity $V \in \mathbb{C}^{n \times n}$ such that $A = VJV^{-1}$ where J is a direct sum of *Jordan blocks*, i.e., matrices of type

$$J_\lambda := \begin{bmatrix} \lambda & 1 & 0 & 0 & \dots \\ 0 & \lambda & 1 & 0 & \dots \\ 0 & 0 & \lambda & 1 & \dots \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \lambda \end{bmatrix}$$

for eigenvalues $\lambda \in \mathbb{C}$ of A . Here J is unique up to permutations but V is not if there are eigenspaces of dimension greater than one. The existence of the JNF is taught in undergraduate linear algebra courses and has myriad applications throughout science and mathematics.

2. **Spectral Factorization.** Given an $n \times n$ monic matrix polynomial

$$P(x) = x^{2d}I + \sum_{i \leq 2d-1} x^i P_i$$

with Hermitian coefficients $P_i \in \mathbb{C}^{n \times n}$ satisfying $P(x) \succeq 0$ for all $x \in \mathbb{R}$, find a monic matrix polynomial $Q(x) = x^d I + \sum_{i \leq d-1} x^i Q_i$ such that $P(x) = Q^*(x)Q(x)$ and $\det(Q(x))$ has all of its zeros in the closed upper half complex plane (where $Q^*(x) = x^d I + \sum_{i \leq d-1} x^i Q_i^*$). Such a $Q(x)$ is guaranteed to exist and is unique [38, 45]. This fact has been rediscovered several times and goes under many names (such as matrix Féjér-Riesz/Wiener-Hopf factorization and matrix polynomial sum of squares) in different fields. Note that the $n = 1$ case is the fact that a univariate scalar polynomial nonnegative on \mathbb{R} may be expressed as a sum of squares (which can be obtained by considering the real and imaginary parts of $Q(x)$), and the $d = 0$ case is just the Cholesky factorization if we allow the leading coefficient of $P(x)$ to be an arbitrary positive semidefinite matrix (not necessarily I).

Both of the above problems have generated a large literature and several proposed methods for solving them (see Section 1.1 for a thorough discussion). Roughly speaking, these methods range on a spectrum between symbolic (relying on algebraic reasoning, performing exact computations with rational numbers, polynomials, field extensions, etc.) and numerical (relying on analytic reasoning, semidefinite optimization, homotopy continuation, etc.). With one exception in the case of problem (1) [8], to the best of our knowledge none of these methods has been rigorously shown to yield a polynomial time algorithm.

This paper provides the first polynomial time bit complexity bounds for problem (2) and significantly improves the best known bound for (1), in the case when the input matrices have integer entries¹. The algorithms we study are simple and the algorithmic ingredients employed are not essentially new; rather, our main contribution is to synthesize ideas from both the symbolic and numerical approaches to these problems, which have in the past developed largely separately across different fields over several decades, in a way which enables good bit complexity estimates. At a technical level, the main task is to find good bounds on both the bit lengths of rational numbers and on the condition numbers of matrices appearing during the execution of the algorithms. A key theme of our proofs is that bit length bounds can be used to obtain condition number bounds and *vice versa*, and that carefully passing between the two is more effective than either one alone.

¹ Or are rational with a common denominator, see the corollaries following the main theorems.

Our two main results, advertised in the abstract, appear in Sections 2 and 3 as Theorems 9 and 18. Additional preliminaries for each result are included in its section, and further history and context for our contributions is discussed in Section 1.1. Two notable common features of our results are:

- Our algorithms have good *forward error* bounds, i.e., they compute approximations to the exact solution of the given instance (as opposed to backward error, computing exact solutions of nearby instances, which is the standard notion in scientific computing). This notion of error is appropriate for mathematical (as opposed to scientific) applications where discontinuous quantities in the input (such as the size of a Jordan block) can be meaningful, but typically comes at the cost of higher running times resulting from the use of numbers with large bit length.
- The running times of our algorithms are bounded solely in terms of the number of bits used to specify the input. This type of result is easier to use than bounds depending on difficult to compute condition numbers, especially for such ill-conditioned problems. As such, the key phenomenon enabling our results is that *instances of controlled bit length cannot be arbitrarily ill-conditioned* in an appropriate sense.

We conclude with a discussion and open problems in Section 4.

1.1 Comparison to Related Work

Jordan Normal Form

As far as we are aware, the only known polynomial bit complexity algorithm for approximately computing the JNF $A = VJV^{-1}$ of a general square rational matrix $A \in \mathbb{Q}^{n \times n}$ with a -bit entries is [8], obtaining a runtime of $O(\text{poly}(n, a))$ where the degree of the polynomial is not specified but is seen to be at least twelve².

In the symbolic computation community, the works [27, 34, 17, 15, 37, 32] gave polynomial *arithmetic* complexity³ bounds for computing the “rational Jordan form” of a matrix over any field. Roughly speaking, the rational Jordan form involves a symbolic representation of the matrix J where the eigenvalues are represented in terms of their minimal polynomials over the field. These results are not adequate for our application to spectral factorization, which requires inverting submatrices of the similarity V , an operation which becomes difficult in the symbolic representation. Nonetheless our JNF algorithm is heavily inspired by the ideas in these works, relying on the same reduction to Frobenius canonical form (expressing A as a direct sum of companion matrices) used in essentially all of them. The main difference is that we compute the eigenvalues approximately using numerical techniques [35], and are able to bound the condition number of V by controlling the minimum gap between distinct eigenvalues as a function of the bit length of the input matrix.

Methods for computing the JNF must inherently involve a symbolic component since the Jordan structure can be changed by infinitesimal perturbations. It is worth mentioning that JNF is still not a solved problem “in practice” as trying to compute the JNF of a 50×50 matrix using standard software packages reveals.

² The related paper [1] proposed using JNF as an “uncheatable benchmark” for certifying that a device has high computational power.

³ The works [34, 17] derived bit complexity bounds for certain special cases of input matrices, but not in general.

Spectral Factorization

Polynomial spectral factorization has been rediscovered many times. The earliest references we are aware of are [38, 24, 45, 39, 9]; the reader may consult any of the excellent surveys [40, 2, 10, 25] for a detailed discussion of the history. More recently, several constructive proofs of the spectral factorization theorem have been proposed e.g. [23, 2, 12, 26, 11, 13], [3, §2] (this list is not meant to be comprehensive). While these may be considered constructive from a mathematical standpoint, bit complexity bounds are not pursued and are not readily evident from the techniques used⁴. Two particularly simple algorithms on this list are [2] (which requires exactly computing the Schur form of a certain matrix and inverting some of its submatrices) and [3, §2] (which requires solving a semidefinite program). We remark that unlike JNF, spectral factorization is actually a problem that is frequently solved in practice, with several of the papers above including numerical experiments.

The work most relevant for this paper is the important paper [18] (see also [31]), which reduces spectral factorization to computing the JNF of a block companion matrix (see Section 3 for a definition), and inverting and multiplying some matrices derived from it. Our contribution is to analyze the conditioning of this approach and combine it with our JNF algorithm, yielding concrete bit complexity bounds.

One notable advantage of our algorithm is that it works even when the input is degenerate – i.e., $P(x)$ is only positive semidefinite rather than positive definite – which frequently occurs in applications. This is in contrast to almost all of the works mentioned above, which only consider the strictly positive definite case (or even require all roots of $\det(P(x))$ to be distinct) and appeal to nonconstructive limiting arguments to handle the degenerate case.

A more stringent variant of the problem is to find a real factorization $P(x) = Q^T(x)Q(x)$ in the case when $P(x)$ is real symmetric, possibly allowing $Q(x)$ to be rectangular. The recent works [6, 22] have obtained optimal bounds on the dimensions of $Q(x)$. In this paper, we restrict our attention to the Hermitian setting.

1.2 Preliminaries

Asymptotic Notation. We will use $O^*(\cdot)$ to suppress polylogarithmic factors in the input parameters n (dimension), a (bit length of input numbers), d (degree), and b (desired bits of accuracy). Logarithmic factors are not the focus of this paper and can be safely ignored everywhere because all proofs in this paper invoke this notation at most a constant number of times (in particular, our algorithms do not contain any loops which could lead to blowups in the exponents of the logarithms).

Numbers and Arithmetic. We say that $x \in \mathbb{Z}\langle\langle a \rangle\rangle$ if x is an integer with bit length at most a and $x \in \mathbb{Z}\langle a \rangle$ if x is an integer with bit length at most $O^*(a)$. We use $\mathbb{Q}\langle a/c \rangle$ (resp. $\mathbb{Q}\langle\langle a/c \rangle\rangle$) to denote the rationals p/q with $p \in \mathbb{Z}\langle a \rangle, q \in \mathbb{Z}\langle c \rangle$ (resp. $\mathbb{Z}\langle\langle a \rangle\rangle, \mathbb{Z}\langle\langle c \rangle\rangle$), and $\mathbb{Q}_{\text{dy}}\langle a/c \rangle$ to denote the elements of $\mathbb{Q}\langle a/c \rangle$ with denominator equal to a power of two; the latter will sometimes be useful since adding rationals with dyadic denominators does not increase the bit length of the denominator. For a rational x , let $\text{round}_c(x)$ denote the nearest rational with denominator 2^c , which clearly satisfies

⁴ This is due in each case to one or more of the following operations: solving a linear system without bounding its condition number, computing the eigenvalues of a matrix or roots of a univariate polynomial or system of multivariate polynomials “exactly” (which is impossible), solving a semidefinite program without controlling the volume of its feasible region, computing a Schur or Jordan form of a matrix “exactly” (also impossible), using an iterative scheme with no rigorous proof of convergence, and assuming arithmetic is carried out in infinite precision.

$$|x - \text{round}_c(x)| \leq 2^{-c} \tag{1}$$

and can be computed in time nearly linear in the bit length of x . This notation extends to complex numbers with rational real and imaginary parts in the natural way. The bit complexity of arithmetic with rational numbers is nearly linear in the bit length (see e.g. [20]).

Matrices. We use $\mathbb{Z}^{n \times n} \langle a \rangle$ (resp. $\mathbb{Z}^{n \times n} \langle a \rangle$) to denote integer matrices with entries of bit length a (resp. $O^*(a)$), and $\mathbb{Q}^{n \times n} \langle a/c \rangle$ (similarly $\mathbb{Q}_{\text{dy}}^{n \times n} \langle a/c \rangle$, $\mathbb{C}^{n \times n} \langle a/c \rangle$, and $\mathbb{C}_{\text{dy}}^{n \times n} \langle a/c \rangle$) to denote matrices with entries in $\mathbb{Q} \langle a/c \rangle$ having a *common denominator*. For $A \in \mathbb{Z}^{n \times n}$, the notation $\langle A \rangle$ refers to the maximum bit length of an entry of A .

We record the following easy facts about inverses as well as products and sums of pairs of matrices⁵, which follow from the adjugate formula for the inverse and the assumption on common denominators⁶:

► **Fact 1** (Bit Length of Matrix Arithmetic).

1. If $A \in \mathbb{Z}^{n \times n} \langle a \rangle$ then $A^{-1} \in \mathbb{Q}^{n \times n} \langle an/an \rangle$.
2. If $A \in \mathbb{K}^{n \times n} \langle a/c \rangle$ then $A^{-1} \in \mathbb{K}^{n \times n} \langle c + an/an \rangle$, for $\mathbb{K} = \mathbb{Q}, \mathbb{Q}_{\text{dy}}, \mathbb{C}, \mathbb{C}_{\text{dy}}$.
3. If $A, B \in \mathbb{K}^{n \times n} \langle a/c \rangle$ then

$$A + B \in \mathbb{K}^{n \times n} \langle a/c \rangle \quad \text{and} \quad AB \in \mathbb{K}^{n \times n} \langle a/c \rangle,$$

for $\mathbb{K} = \mathbb{Q}, \mathbb{Q}_{\text{dy}}, \mathbb{C}, \mathbb{C}_{\text{dy}}$.

Perturbation Theory. We use $\|\cdot\|$ to denote the operator norm and $\|\cdot\|_{\max}$ to denote the entrywise ℓ_∞ norm of a matrix, noting that $\|M\|_{\max} \leq \|M\| \leq n\|M\|_{\max}$ for an $n \times n$ matrix M . We use $\kappa(M) := \|M\| \|M^{-1}\|$ to denote the condition number of an invertible matrix. We will frequently use the elementary fact:

$$\|(M + E)^{-1} - M^{-1}\| \leq \frac{\|E\| \|M^{-1}\|}{1 - \|E\| \|M^{-1}\|} \cdot \|M^{-1}\| \tag{2}$$

provided $\|E\| \|M^{-1}\| < 1$, which follows from a Neumann series argument, as well as its consequence

$$\kappa(M + E) \leq \kappa(M) \frac{1 + \|E\| \|M^{-1}\|}{1 - \|E\| \|M^{-1}\|} \tag{3}$$

whenever $\|E\| \|M^{-1}\| < 1$.

Polynomials. We use $\text{mingap}(\cdot)$ to indicate the minimum gap between *distinct* roots of a polynomial. We use $\|P(\cdot) - Q(\cdot)\|_{\max}$ to denote the coefficient-wise $\|\cdot\|_{\max}$ norm of two matrix polynomials. We extend the notations $\langle \cdot \rangle, \langle \cdot \rangle$ to polynomials by applying them to each scalar or matrix coefficient.

⁵ We do not rely on matrix arithmetic with a superconstant number of matrices in this paper, for which the bit length bounds necessarily depend on the number of matrices.

⁶ Allowing distinct denominators in the entries of A, B could increase the bit lengths of AB and $A + B$ by a factor of n if the denominators are, say, relatively prime

Bit Length of Inverse and Characteristic Polynomial. We will frequently appeal to the bounds

$$\|A^{-1}\| \leq n!2^{an} \quad \text{whenever } A \in \mathbb{Z}^{n \times n}\langle a \rangle \text{ is invertible,} \quad (4)$$

which is easily seen by considering the adjugate formula for the inverse, and

$$\langle\langle \chi_A(x) \rangle\rangle \leq n!2^{an} \quad \text{for } A \in \mathbb{Z}^{n \times n}\langle a \rangle, \quad (5)$$

where $\chi_A(x) := \det(xI - A)$ denotes the characteristic polynomial.

2 Jordan Normal Form

The *companion matrix* of a scalar monic polynomial $p(x) = x^d + \sum_{i < d} p_i x^i$ is the $d \times d$ matrix:

$$C_p := \begin{bmatrix} 0 & 1 & & & & \\ 0 & 0 & 1 & & & \\ & & \vdots & & & \\ & & & & & 1 \\ -p_0 & -p_1 & -p_2 & -p_3 & \dots & -p_{d-1} \end{bmatrix}^T \quad (6)$$

It is easily seen that $\det(xI - C_p) = p(x)$. The high level idea of our algorithm is to use symbolic techniques to reduce the input matrix to a direct sum of companion matrices, and then use explicit formulas and root finding algorithms to compute the JNF of the companion matrices. We will rely on the following tools.

► **Theorem 2** (Exact Frobenius Canonical Form, [16] Theorems 2.2 & 3.2). *There is a randomized Las Vegas algorithm which given $A \in \mathbb{Z}^{n \times n}\langle a \rangle$, outputs a matrix $F \in \mathbb{Z}\langle an \rangle$ which is a direct sum of companion matrices and an invertible $U \in \mathbb{Z}\langle an^2 \rangle$ satisfying $A = UFU^{-1}$, with an expected running time of $O^*(n^5 a + n^4 a^2)$ bit operations.*

► **Theorem 3** (Approximate Polynomial Roots in the Unit Disk, [35] Corollary 2.1.2). *There is an algorithm which given bitwise access⁷ to the coefficients of a polynomial $p \in \mathbb{Q}[x]$ of degree n with all roots $z_1, \dots, z_n \in \mathbb{C}$ satisfying $|z_i| \leq 1$ and a parameter⁸ $b' \geq \log n$, computes numbers $z'_1, \dots, z'_n \in \mathbb{C}_{\text{dy}}\langle b' \rangle$ such that $|z'_i - z_i| \leq 2^{2-b'}$ for all $i \leq n$, using at most $O^*(n^2 b')$ bit operations.*

► **Theorem 4** (Minimum Gap of Integer Polynomials, [33]). *If $p \in \mathbb{Z}[x]\langle a \rangle$ is monic of degree n , then*

$$\text{mingap}(p) \geq 2^{-an - 2n \lg n}.$$

► **Corollary 5** (Approximate Roots and Multiplicities of Integer Polynomials). *There is an algorithm which given an integer polynomial $p \in \mathbb{Z}[x]\langle a \rangle$ of degree $n \geq 2$ with roots $z_1, \dots, z_n \in \mathbb{C}$ and a parameter $b' \geq an + 4n \lg n$, computes numbers $z'_1, \dots, z'_n \in \mathbb{C}_{\text{dy}}\langle (a + b')/b' \rangle$ such that $|z'_i - z_i| < 2^{-b'}$ for all $i \leq n$, using at most $O^*(n^2(b' + a))$ bit operations. Each z'_i appears a number of times exactly equal to the multiplicity of z_i in $p(x)$.*

⁷ i.e., the algorithm can query the i th bit of the binary expansion of each coefficient in constant time. This is slightly different from the access model in this paper where rational numbers are given as numerator and denominator, but it is easy to see that given a rational in $\mathbb{Q}\langle a/c \rangle$, the desired binary expansion needed to apply Theorem 3 can be produced in $O^*(a + c)$ time.

⁸ The parameter b' here corresponds to b/n in [35]

Proof. The largest root of $p(x)$ has magnitude at most the sum of the absolute values of its coefficients, which is at most $M = n2^a$. Apply Theorem 3 to the polynomial $p(Mx)$, which has roots in the unit disk, with error parameter $b' + \lg(M) = b' + \lg(n) + a + 1$ to obtain numbers $\tilde{z}_1, \dots, \tilde{z}_n \in \mathbb{C}_{\text{dy}}\langle(a+b')/(a+b')\rangle = \mathbb{C}_{\text{dy}}\langle b'/b'\rangle$ (since $b' \geq a$) with common denominator. Then for $i = 1, \dots, n$ we have $z'_i := \tilde{z}_i M \in \mathbb{C}\langle(a+b')/b'\rangle$ with common denominator and $|z'_i - z_i| \leq 2^{-b'}$. By Theorem 4 the minimum gap between distinct z_i is at least $2^{-b'+1}$ since $2n \lg n \geq 1$ so this is sufficient to correctly determine the multiplicity of each z_i and replace all z'_i corresponding to a root with the same value. \blacktriangleleft

We will also use an explicit formula for the JNF of a companion matrix as a confluent Vandermonde matrix (see e.g. [14, 5] for a discussion) in the roots of the corresponding polynomial.

► **Theorem 6** ([7]). *If $C \in \mathbb{C}^{n \times n}$ is a companion matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_k \in \mathbb{C}$ of multiplicities m_1, \dots, m_k , then $C = WJW^{-1}$ with*

$$J = \bigoplus_{j \leq k} J_{\lambda_j}$$

$$W = [W_{\lambda_1}, W_{\lambda_2}, \dots, W_{\lambda_k}]$$

where J_{λ_j} an $m_j \times m_j$ Jordan block with eigenvalue λ_j and W_{λ_j} is the $n \times m_j$ matrix:

$$W_{\lambda_j} := \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \lambda_j & 1 & 0 & \dots & 0 \\ \lambda_j^2 & 2\lambda_j & 1 & \dots & 0 \\ \lambda_j^3 & 3\lambda_j^2 & \binom{3}{2}\lambda_j & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_j^{n-1} & (n-1)\lambda_j^{n-2} & \binom{n-1}{2}\lambda_j^{n-3} & \dots & \binom{n-1}{m_j}\lambda_j^{n-m_j} \end{bmatrix}, \quad (7)$$

so that W is a confluent Vandermonde matrix.

Note that the entries of W_λ in (7) are univariate polynomials of degree n in the λ_i with coefficients in $\mathbb{Z}\langle n \rangle$. We now present the algorithm.

We begin by fully defining and analyzing Step 3 of JNF.

► **Lemma 7** (Rounded Approximate Eigenvalue Powers). *The approximate powers of the eigenvalues $\widetilde{\lambda}_{ij}^p \in \mathbb{C}_{\text{dy}}\langle b'/b'\rangle$ required in Step 3 may be computed in $O^*(n^2b')$ bit operations and satisfy*

$$|\widetilde{\lambda}_{ij}^p - \lambda_{ij}^p| \leq 2^{-b'} \cdot n2^{2n+1} \|A\|^{n^2+n} \leq 2^{-b'+an^2+5an}$$

for every i, j .

Proof. Suppose we wish to compute powers $\lambda, \lambda^2, \dots, \lambda^r$ for some nonzero eigenvalue $\lambda = \lambda_{ij}$ appearing in W . Let $\widetilde{\lambda}$ be the approximate eigenvalue produced in Step 2, satisfying $|\lambda - \widetilde{\lambda}| \leq 2^{-b'}$. We use the following inductive scheme for $p = 2, \dots, r \leq n$:

$$\widetilde{\lambda}^p := \text{round}_{b'}(\widetilde{\lambda}^{p-1} \cdot \widetilde{\lambda}).$$

First, observe that from Step 2 we have the error estimate $|\lambda - \widetilde{\lambda}| \leq 2^{-b'}$, which implies that for every $p \leq n$:

$$|\lambda^p - (\widetilde{\lambda})^p| \leq 2^{-b'} \cdot p \cdot |\lambda^{p-1}| \leq 2^{-b'} n \|A\|^n \leq 2^{-b'} \cdot n2^{2n} \|A\|^{n^2+n} \quad (9)$$

■ **Algorithm 1** Algorithm JNF.

Input: $A \in \mathbb{Z}^{n \times n} \langle a \rangle$, desired bits of accuracy b .

Output: $\tilde{J}, \tilde{V} \in \mathbb{C}_{\text{dy}}^{n \times n} \langle an^3 + b/(an^3 + b) \rangle$.

Guarantee: $\|J - \tilde{J}\| \leq 2^{-b} \|J\|, \|V - \tilde{V}\| \leq 2^{-b} \|V\|$ for some exact JNF $A = VJV^{-1}$ and $\kappa(\tilde{V}) \leq 2^{O^*(an^3)}$.

1. Exactly compute the Frobenius Normal Form $A = UFU^{-1}$ with $F \in \mathbb{Z}^{n \times n} \langle an \rangle$ and $U \in \mathbb{Z}^{n \times n} \langle an^2 \rangle$ using Theorem 2. Let $F = \bigoplus_{i \leq \ell} C_i$ for companion matrices $C_i \in \mathbb{Z}^{n_i \times n_i} \langle an \rangle$. Let $c := \max_i \langle C_i \rangle$.
2. For $i = 1 \dots \ell$, apply Corollary 5 to the characteristic polynomial $\chi_{C_i}(x) \in \mathbb{Z}[x] \langle an \rangle$ with accuracy

$$b' := b + (n + 1) \langle U \rangle + cn^2 + an^2 + 4n^2 \lg n + 7an + 3 \lg n \quad (8)$$

to obtain approximations $\widetilde{\lambda}_{i1}, \dots, \widetilde{\lambda}_{ik_i} \in \mathbb{C}_{\text{dy}} \langle b'/b' \rangle$ to the distinct eigenvalues $\lambda_{i1}, \dots, \lambda_{ik_i}$ of C_i , with error $|\lambda_{ij} - \widetilde{\lambda}_{ij}| \leq 2^{-b'}$, as well as their multiplicities.

3. For $i = 1, \dots, \ell$, compute approximate eigenvalue powers $\widetilde{\lambda}_{i1}^p, \dots, \widetilde{\lambda}_{ik_i}^p \in \mathbb{C}_{\text{dy}} \langle b'/b' \rangle$ using Lemma 7. Let

$$\widetilde{J}_i := \bigoplus_{j \leq k_i} J_{\widetilde{\lambda}_{ij}} \in \mathbb{C}_{\text{dy}}^{n_i \times n_i} \langle b'/b' \rangle \quad \text{and} \quad \widetilde{W}_i := [W_{\widetilde{\lambda}_{i1}}, \dots, W_{\widetilde{\lambda}_{ik_i}}] \in \mathbb{C}_{\text{dy}}^{n_i \times n_i} \langle b'/b' \rangle$$

as in (7), i.e., substitute the approximate powers $\widetilde{\lambda}_{ij}^p$ into the appropriate polynomials J_λ, W_λ .

4. Output \tilde{J} and $\tilde{V} = U\widetilde{W}$.

since $|\lambda| \leq \|A\|$ and $\|A\| \geq 1$. Thus, it suffices to show that for each p :

$$|\widetilde{\lambda}^p - (\widetilde{\lambda})^p| \leq 2^{-b'} \cdot n 2^{2n} \|A\|^{n^2+n}. \quad (10)$$

Notice that $|\lambda| \geq \|A\|^{-n}$ since the product of the nonzero eigenvalues of A is given by $e_k(A) \geq 1$, for e_k the last nonzero elementary symmetric function of A , and each eigenvalue of A is at most $\|A\|$. Since

$$2^{-b'} \leq 2^{-an-2 \lg n-1} \leq \|A\|^{-n}/2, \quad (11)$$

we have $|\widetilde{\lambda}| \geq \|A\|^{-n}/2$ and thereby $|(\widetilde{\lambda})^p| \geq \|A\|^{-n^2}/2^n$ for every $p = 1, \dots, n$. It now follows by induction that:

$$|\text{round}_{b'}(\widetilde{\lambda}^{p-1} \cdot \widetilde{\lambda}) - (\widetilde{\lambda})^p| \leq 2^{-b'} p \cdot 2^n \|A\|^{n^2} |(\widetilde{\lambda})^p|$$

for every $p = 2, \dots, r$, i.e., where the inductive hypothesis is that in each step the rounding incurs a *relative* error of at most $2^{-b'} 2^n \|A\|^{n^2}$, and we observe that the relative errors simply add up since they are sufficiently smaller than one. Since we also have the upperbound $|(\widetilde{\lambda})^p| \leq 2^n \|A\|^n$, the desired inequality (10) follows. Combining this with (9) yields the advertised error bound.

The total bit complexity for one eigenvalue is n times the cost of one step of the induction, which is $O^*(nb')$. Since there are n eigenvalues, the total cost is $O^*(n^2b')$. ◀

The key condition number bounds used in proving correctness of JNF are the following, obtained via the minimum eigenvalue gap of W which is controlled using the maximum bit length of the C_i . Item (iii) is also used in the analysis of the spectral factorization algorithm in Section 3.

► **Lemma 8** (Condition Numbers from Gaps). *If $A \in \mathbb{Z}^{n \times n} \langle a \rangle$ and $A = UFU^{-1} = (UW)J(UW)^{-1} = VJV^{-1}$ for exact Frobenius and Jordan forms as above, then*

- (i) $\kappa(U) \leq n^2 \cdot n! \cdot 2^{n \langle U \rangle + \langle U \rangle} \leq 2^{O^*(an^3)}$.
- (ii) $\kappa(W) \leq 2^{O^*(an^3)}$.
- (iii) $\|V\| \leq n 2^{\langle U \rangle} \cdot 2^{n+an+n \lg n} \leq 2^{O^*(an^2)}$ and $\|V^{-1}\| \leq n \cdot (n!)^2 2^{n \langle U \rangle + cn^2 + 2n^2 \lg n} \leq 2^{O^*(an^3)}$.

Proof. For (i), note that $\|U\| \leq n 2^{\langle U \rangle}$ and since $U \in \mathbb{Z}^{n \times n} \langle an^2 \rangle$, we have

$$\|U^{-1}\| \leq n \cdot n! 2^{n \langle U \rangle} \leq 2^{O^*(an^3)}.$$

Consequently $\|U\| \|U^{-1}\| \leq 2^{O^*(an^3)}$.

The matrix W is a direct sum of W_i , which are confluent Vandermonde matrices in the eigenvalues λ_{ij} , which are roots of the $\chi_{C_i}(x)$. By Theorem 4 and $\langle \chi_{C_i} \rangle = \langle C_i \rangle$, we have

$$\delta := \min_i [\text{mingap}(\chi_{C_i})] \geq 2^{-\max_i \langle \chi_{C_i} \rangle n - 2n \lg n} = 2^{-cn - 2n \lg n} \geq 2^{-O^*(an^2)},$$

where the last inequality uses $c = O^*(an)$. Then [5, Theorem 1] implies that

$$\|W^{-1}\| \leq n! (1/\delta)^n \leq n! 2^{cn^2 + 2n^2 \lg n} \leq 2^{O^*(an^3)}.$$

On the other hand, the formula (7) reveals that $\|W\| \leq n \cdot 2^{n+an+n \lg n}$ since $|\lambda_{ij}| \leq n 2^a$. Multiplying these two bounds yields (ii).

Finally, we have $\kappa(V) \leq \kappa(U)\kappa(W)$ and $\|V^{-1}\| \leq \|W^{-1}\| \|U^{-1}\|$, establishing (iii). ◀

► **Theorem 9.** *The algorithm JNF satisfies its guarantees and runs in expected $O^*(n^{\omega+3}a + n^4a^2 + n^{\omega}b)$ bit operations.*

Proof.

Bit Length of the Output. The bit length assertions in Steps 1 and 2 are immediate from Theorem 2 and Corollary 5. The bit length of \tilde{J}, \tilde{W} in Step 3 is guaranteed by Lemma 7. The bit length of the product in Step 4 is implied by Fact 1.

Error Bounds. Step 1 is exact.

Lemma 7 implies that the matrices \tilde{J}_i, \tilde{W}_i in Step 3 satisfy

$$\|J_i - \tilde{J}_i\|_{\max} \leq 2^{-b'}, \quad \|W_i - \tilde{W}_i\|_{\max} \leq 2^{-b' + n + an^2 + 5an} \quad (12)$$

This additive bound is preserved under taking direct sums. To obtain the multiplicative bound, we observe that $\|W\| \geq 1$ by (7); if there is a Jordan block of size at least two then $\|J\| \geq 1$ also, otherwise since A is integral we have

$$\prod_{\text{nonzero } \lambda_{ij}} \lambda_{ij}^{\text{mult}(\lambda_{ij})} = e_k(A) = e_k(J) \geq 1$$

for the last nonzero elementary symmetric function e_k of A , so one of the eigenvalues λ_{ij} must be at least $\|A\|^{-n}$ and we have crudely $\|J\| \geq \|A\|^{-n} \geq 2^{-2an}$. In either case, we conclude after passing to the operator norm that

$$\|J - \tilde{J}\| \leq 2^{-b' + 2an} \|J\| \leq 2^{-b} \|J\| \quad \text{and} \quad \|W - \tilde{W}\| \leq 2^{-b' + an^2 + 6an} \|W\|.$$

42:10 JNF and Spectral Factorization

To obtain the final error bound on \tilde{V} in Step 4, we observe that

$$\|V - \tilde{V}\| = \|UW - U\tilde{W}\| \leq \|U\|n2^{-b'+n+an^2+5an} \leq 2^{-b'+an^2+6an+\langle U \rangle+2\lg n} \quad (13)$$

since $\|U\| \leq 2^{\langle U \rangle+1\lg n}$. Since $\|V\| \geq \|W\|/\|U^{-1}\| \geq 2^{-n\langle U \rangle-n\lg n-\lg n}$, we obtain the conclusion

$$\|V - \tilde{V}\| \leq 2^{-b'+an^2+6an+\langle U \rangle+3\lg n+n\langle U \rangle+n\lg n}\|V\| \leq 2^{-b}\|V\| \quad (14)$$

by our choice of b' , as desired.

Condition of \tilde{V} . The bound (13) together with Lemma 8(iii) implies

$$\|V - \tilde{V}\|\|V^{-1}\| \leq 2^{-b'+an^2+6an+\langle U \rangle+2\lg n} \cdot n \cdot (n!)^2 2^{n\langle U \rangle+cn^2+2n^2\lg n} \leq 1/2$$

by the choice of b' in Step 2. It follows from (3) that

$$\kappa(\tilde{V}) \leq \kappa(V) \frac{1+1/2}{1-1/2} \leq 3\kappa(V) \leq 2^{O^*(an^3)}, \quad (15)$$

as desired.

Complexity. Step 1 takes $O^*(n^5a + n^4a^2)$ bit operations by Theorem 2.

Step 2 takes $O^*(n^2(an^3 + b))$ bit operations by Theorem 3.

Step 3 takes $O^*(an^5 + bn^2)$ bit operations by Lemma 7.

The matrix multiplication in Step 4 $O^*(n^\omega(b' + an^2))$ time.

The total running time is therefore $O^*(n^\omega b' + n^4a^2) = O^*(n^{\omega+3}a + n^4a^2 + n^\omega b)$, as advertised. \blacktriangleleft

► **Corollary 10** (JNF of Rational Matrices with Common Denominator). *The algorithm JNF can easily be used to compute the JNF of A/q for integer A and q : if \tilde{J}, \tilde{V} is an approximate JNF of A with b bits of accuracy, then $\tilde{J}/q, \tilde{V}$ is an approximate JNF of A/q with $b - \lg(q)$ bits of accuracy. This fact will be useful in the spectral factorization algorithm in the following section.*

► **Remark 11.** The proof of Theorem 9 yields an explicit estimate on $\kappa(V)$ in terms of the bit lengths $a, c, \langle U \rangle$ which may be better than the worst case bound of $2^{O^*(an^3)}$ for specific instances.

3 Spectral Factorization

We briefly review some aspects of the theory of matrix polynomials (the reader may consult [19] for a comprehensive introduction). Given a monic matrix polynomial $L(x) = x^d I + \sum_{i \leq d-1} x^i L_i$ with $L_i \in \mathbb{C}^{n \times n}$, its adjoint is $L^*(x) = x^d I + \sum_{i \leq d-1} x^i L_i^*$, and its latent roots are the $\lambda \in \mathbb{C}$ such that $L(\lambda)$ is singular. The *block companion matrix*⁹ of L is the $dn \times dn$ matrix:

⁹ This is a “row” companion matrix as opposed to the “column” companion matrices in Section 2. This is customary in the theory of matrix polynomials.

$$C_L := \begin{bmatrix} 0 & 1 & & & & \\ 0 & 0 & 1 & & & \\ & & \vdots & & & \\ & & & & 1 & \\ -L_0 & -L_1 & -L_2 & -L_3 & \dots & -L_{d-1} \end{bmatrix} \quad (16)$$

The following important theorem states the existence of spectral factorizations of positive definite monic matrix polynomials, and gives a way of computing them using the block companion matrix.

► **Theorem 12** (Theorems 5.1, 5.4 of [18]). *Suppose $P(x) = P^*(x) = x^{2d}I + \sum_{i \leq 2d-1} x^i P_i \in \mathbb{C}^{n \times n}[x]$ monic of degree $2d$ satisfies $P(x) \succeq 0$ for all $x \in \mathbb{R}$. Then:*

1. *There is a unique monic $Q(x) \in \mathbb{C}^{n \times n}[x]$ of degree d such that $P(x) = Q^*(x)Q(x)$ and Q has all of its latent roots in the closed upper half plane.*
2. *The complex eigenvalues of C_P occur in conjugate pairs, and each Jordan block in the JNF of C_P corresponding to a real eigenvalue has even size.*
3. *Let $C_P = VJV^{-1}$ be a Jordan Form of the block companion matrix of P with block decomposition*

$$J =: \begin{bmatrix} J_+ & & \\ & J_0 & \\ & & J_- \end{bmatrix}, \quad V =: \begin{bmatrix} V_+ & V_0 & V_- \\ Z_+ & Z_0 & Z_- \end{bmatrix}$$

for J_{\pm} corresponding to eigenvalues in the open upper/lower half plane and J_0 corresponding to the real eigenvalues and V_{\pm}, V_0 having dn rows. Then

$$C_Q = V_{\geq 0} J_{\geq 0} V_{\geq 0}^{-1}, \quad (17)$$

where

$$V_{\geq 0} = [V_+, V_0^{(1/2)}] \in \mathbb{C}^{dn \times dn} \quad \text{and} \quad J_{\geq 0} = J_+ \oplus J_0^{(1/2)} \in \mathbb{C}^{dn \times dn}. \quad (18)$$

Here, for each Jordan block of size $2s$ in J_0 , $J_0^{(1/2)}$ contains a Jordan block of size s with the same eigenvalue, and $V_0^{(1/2)}$ contains as columns the first s of the corresponding $2s$ columns of V_0 .

The formula (17) gives a one line algorithm for computing Q given access to the exact Jordan form of P . The key issue is that in order to use an approximate Jordan form $\tilde{V}\tilde{J}\tilde{V}^{-1}$ in the formula, we must have a good bound on the condition number of $V_{\geq 0}$ in order to control the error incurred during inversion. Note that while Lemma 8 guarantees a bound on $\kappa(V)$, this does not in general imply a bound on its submatrices; indeed, it is known that there can be square submatrices of V which are singular. The main technical contribution of this section is to prove a bound on $\kappa(V_{\geq 0})$ in terms of $\kappa(V)$ by exploiting the special structure of V which arises from the structure of C_P . This is encapsulated in the following fact, which may be found in any reference on matrix polynomials (e.g., [19, §1]).

► **Fact 13.** *If $C_P = VJV^{-1}$ is the Jordan normal form of an $n \times n$ complex matrix polynomial P of degree d , then there is a matrix $X \in \mathbb{C}^{n \times 2dn}$ such that:*

$$V = \begin{bmatrix} X \\ XJ \\ XJ^2 \\ \vdots \\ XJ^{2d-1} \end{bmatrix}. \quad (19)$$

42:12 JNF and Spectral Factorization

We show that the least singular value of a column submatrix of any matrix of type (19) may be related to the least singular values of certain block submatrices.

► **Lemma 14** (Condition of Submatrices of Companion JNF). *Given any $Y \in \mathbb{C}^{n \times D}$ and $K \in \mathbb{C}^{D \times D}$ with $\|K\| \geq 1$, define for $k = 1, 2, \dots$ the $nk \times D$ matrices:*

$$W_k := \begin{bmatrix} Y \\ YK \\ YK^2 \\ \vdots \\ YK^{k-1} \end{bmatrix}.$$

Then

$$\sigma_D(W_D) \geq \frac{\sigma_D(W_k)}{\sqrt{k}(4\|K\|)^{D(k-D+1)}}$$

for every $k \geq D$.

Proof. Suppose $x \in \mathbb{C}^D$ is a unit vector satisfying $\|W_D x\| = \sigma_D(W_D) =: \sigma$. We will show that

$$\|W_k x\| \leq \sigma \cdot \sqrt{k}(2D^{1/D}\|K\|)^{D(k-D+1)}, \quad (20)$$

which yields the Lemma by using $D^{1/D} \leq 2$. Let q be the characteristic polynomial of K . By the Cayley-Hamilton theorem, we have

$$q(K) = K^D + \sum_{0 \leq i \leq D-1} c_i K^i = 0,$$

for some complex coefficients c_i crudely bounded as

$$\max_{i \leq D-1} |c_i| \leq 2^D \|K\|^D := \alpha,$$

by considering their expansion as elementary symmetric functions in the eigenvalues of K . Using this expression, we obtain the identity:

$$YK^j x = YK^{j-D} K^D x = - \sum_{0 \leq i \leq D-1} c_i YK^{j-D} K^i x,$$

for every $j \geq D$. By the triangle inequality, this yields:

$$\|YK^j x\| \leq \alpha D \cdot \max_{i < j} \|YK^i x\|,$$

which applied recursively gives:

$$\|YK^j x\| \leq (\alpha D)^{j-D+1} \cdot \max_{i < D} \|YK^i x\| \leq (\alpha D)^{j-D+1} \sigma.$$

Summing over all $j \leq k$, we have:

$$\|W_k x\|^2 \leq \sigma^2 + \sum_{j=D}^k (\alpha D)^{2(j-D+1)} \sigma^2 \leq k(\alpha D)^{2(k-D+1)}.$$

Taking a square root establishes (20) and finishes the proof. ◀

► **Remark 15.** Lemma 14 is a quantitative version of the main claim of [18, §2.3] showing that $V_{\geq 0}$ is invertible whenever V is invertible, which is central to the theory of matrix polynomials. The proof above is an arguably simpler proof of this fact, and may be of independent interest. The original proof of [18] relies on a delicate analysis of a certain indefinite quadratic form.

Finally, we are able to bound $\kappa(V_{\geq 0})$.

► **Lemma 16.** *In the setting of Theorem 12,*

$$\|V_{\geq 0}^{-1}\| \leq \|V^{-1}\| \cdot \sqrt{2dn}(4 + 4\|C_P\|)^{dn(dn+1)}$$

and

$$\kappa(V_{\geq 0}) \leq \kappa(V) \cdot \sqrt{2dn}(4 + 4\|C_P\|)^{dn(dn+1)}.$$

Proof. Letting $C_P = VJV^{-1}$, the similarity V has the form (19) for some $X \in \mathbb{C}^{n \times 2dn}$. Let $X_{\geq 0}$ be the $n \times dn$ submatrix of X with columns corresponding to the columns in $V_{\geq 0}$. Apply Lemma 14 with $D = dn, k = 2dn, K = J_{\geq 0}, Y = X_{\geq 0}$, noting that $\|K\| \leq 1 + \|C_P\|$ since all of the diagonal entries of $J_{\geq 0}$ are eigenvalues of C_P and bounded by its norm. This yields:

$$\sigma_{dn}(V_{\geq 0}) \geq \frac{\sigma_{dn} \left(\begin{bmatrix} V_{\geq 0} \\ Z_{\geq 0} \end{bmatrix} \right)}{\sqrt{2dn}(4 + 4\|C_P\|)^{dn(dn+1)}},$$

where $Z_{\geq 0}$ has the obvious meaning, yielding the first claim. But $\begin{bmatrix} V_{\geq 0} \\ Z_{\geq 0} \end{bmatrix}$ is a column submatrix of V so

$$\sigma_{dn} \left(\begin{bmatrix} V_{\geq 0} \\ Z_{\geq 0} \end{bmatrix} \right) \geq \sigma_{dn}(V) \geq \sigma_{2dn}(V).$$

Combining this with $\sigma_1(V_{\geq 0}) \leq \sigma_1(V)$, we obtain the second claim. ◀

We now present the algorithm SF which approximately computes the $Q(\cdot)$ guaranteed by Theorem 12 using an approximate Jordan normal form computation and exact inversion. We rely on the following tool from symbolic computation.

► **Theorem 17** (Fast Exact Inversion, [43]). *There is a randomized algorithm which given an invertible matrix $A \in \mathbb{Z}^{n \times n} \langle a \rangle$ exactly computes its inverse $A^{-1} \in \mathbb{Q}^{n \times n} \langle an/an \rangle$ in time $O^*(n^3a + n^3 \log \kappa(A))$.*

► **Theorem 18.** *The algorithm SF satisfies its guarantees and runs in $O^*((dn)^6a + (dn)^4a^4 + (dn)^3b)$ bit operations.*

Proof. Item (2) of Theorem 12 shows that $P(x) \not\equiv 0$ if there is an odd size Jordan block with real eigenvalue.

Assuming this is not the case, that theorem shows that the exact spectral factor Q is given by the last row of $C_Q = V_{\geq 0}J_{\geq 0}V_{\geq 0}^{-1}$. We now prove that the quantity $\widetilde{V}_{\geq 0}\widetilde{J}_{\geq 0}\widetilde{J}_{\geq 0}^{-1}$ computed by SF is close to C_Q . This is a consequence of the following estimates. Given Lemma 16, the arguments are essentially identical to those in the proof of Lemma 8 and Theorem 9 (the key point being that the inverse of a well-conditioned matrix is stable under small enough perturbations).

■ **Algorithm 2** Algorithm SF:

Input: Coefficients $P_0, \dots, P_{2d-1} \in \mathbb{Q}^{n \times n} \langle a/a \rangle$ (with a common denominator) of a monic matrix polynomial $P(x)$, desired bits of accuracy $b \in \mathbb{N}$.

Output: $\widetilde{Q}_0, \dots, \widetilde{Q}_{d-1} \in \mathbb{C}_{\text{dy}}^{n \times n} \langle a(dn)^3 + b \rangle$ or a certificate that $P(x) \not\equiv 0$ for some $x \in \mathbb{R}$.

Guarantee: If $P(x) \succeq 0$ then $\|\widetilde{Q}(\cdot) - Q(\cdot)\|_{\max} \leq 2^{-b} \|Q(\cdot)\|_{\max}$ for $P(x) = Q^*(x)Q(x)$.

1. Compute an approximate Jordan Normal Form $(\widetilde{V}, \widetilde{J}) = \text{JNF}(C_P, b'')$ of C_P using Corollary 10, with $\widetilde{J}, \widetilde{V} \in \mathbb{C}_{\text{dy}}^{n \times n} \langle b''/b'' \rangle$ where b'' is chosen to be the least integer such that

$$f_1(b'') + f_2(b'') \leq 2^{-b} \|P(\cdot)\|_{\max},$$

where f_1, f_2 are defined in (24),(30). Determine its eigenvalues on¹⁰, below, and above the real line. If any Jordan block corresponding to a real eigenvalue has odd size, output “ $P(x) \not\equiv 0$ ”.

2. Let

$$\widetilde{J} =: \begin{bmatrix} \widetilde{J}_+ & & \\ & \widetilde{J}_0 & \\ & & \widetilde{J}_- \end{bmatrix}, \quad V =: \begin{bmatrix} \widetilde{V}_+ & \widetilde{V}_0 & \widetilde{V}_- \\ * & * & * \end{bmatrix}$$

be a block decomposition such that \widetilde{J}_+ corresponds to eigenvalues of C_P in the open upper half plane and \widetilde{J}_0 corresponds to real eigenvalues of C_P .

3. Output the negative of the last row of

$$\widetilde{C}_Q := \widetilde{V}_{\geq 0} \widetilde{J}_{\geq 0} \widetilde{V}_{\geq 0}^{-1}, \quad (21)$$

where

$$\widetilde{V}_{\geq 0} := [\widetilde{V}_+, \widetilde{V}_0^{(1/2)}] \in \mathbb{C}^{dn \times dn} \quad \text{and} \quad \widetilde{J}_{\geq 0} := \widetilde{J}_+ \oplus \widetilde{J}_0^{(1/2)} \in \mathbb{C}^{dn \times dn} \quad (22)$$

and $(\cdot)^{(1/2)}$ is defined as in Theorem 12. The approximate inverse $\widetilde{V}_{\geq 0}^{-1}$ is computed by exactly computing $(\widetilde{V}_{\geq 0})^{-1}$ using Theorem 17 and letting $\widetilde{V}_{\geq 0}^{-1} = \text{round}_{b''}((\widetilde{V}_{\geq 0})^{-1})$.

Let B be the maximum of $n2^a$ (which is an upperbound on $\|C_P\|$) and the two explicit upper bounds on $\|V\|$ and $\|V^{-1}\|$ in Lemma 8(iii) (noting that the bit size $\langle U \rangle$ can be read off from the matrix U produced during the execution of JNF, so B is easily computable). It follows by Lemma 16 and the guarantees of JNF that:

$$\|V\|, \|J\|, \|\widetilde{V}_{\geq 0}\|, \|\widetilde{J}_{\geq 0}\| \leq B, \quad \|\widetilde{V}_{\geq 0}^{-1}\|, \|\widetilde{J}_{\geq 0}^{-1}\| \leq 2B, \quad \|V_{\geq 0}^{-1}\| \leq 2^{2(a+\lg n)d^2 n^2} B =: B'. \quad (23)$$

Applying the triangle inequality thrice, we decompose the output error of SF as:

$$\begin{aligned} \|\widetilde{V}_{\geq 0} \widetilde{J}_{\geq 0} \widetilde{V}_{\geq 0}^{-1} - V_{\geq 0} J_{\geq 0} V_{\geq 0}^{-1}\| &\leq \|\widetilde{V}_{\geq 0} \widetilde{J}_{\geq 0} \widetilde{V}_{\geq 0}^{-1} - \widetilde{V}_{\geq 0} \widetilde{J}_{\geq 0} V_{\geq 0}^{-1}\| \\ &\quad + \|\widetilde{V}_{\geq 0} \widetilde{J}_{\geq 0} V_{\geq 0}^{-1} - \widetilde{V}_{\geq 0} J_{\geq 0} V_{\geq 0}^{-1}\| \\ &\quad + \|\widetilde{V}_{\geq 0} J_{\geq 0} V_{\geq 0}^{-1} - V_{\geq 0} J_{\geq 0} V_{\geq 0}^{-1}\| \\ &\leq 4B^2 \|\widetilde{V}_{\geq 0}^{-1} - V_{\geq 0}^{-1}\| \\ &\quad + B' \cdot 2B^2 \|\widetilde{J}_{\geq 0} - J_{\geq 0}\| \\ &\quad + B' \cdot B^2 \|\widetilde{V}_{\geq 0} - V_{\geq 0}\|. \end{aligned}$$

The sum of the last two terms is bounded by

$$2^{2(a+\lg n)d^2n^2} \cdot 2B^2(2^{-b''} \|J\| + 2^{-b''} \|V\|) \leq 2^{-b''+2(a+\lg n)d^2n^2} \cdot 2B^3 =: f_1(b''). \quad (24)$$

For the first term, observe that whenever

$$2^{-b''} \leq B'/2 \quad (25)$$

we have

$$4B^2 \|\widetilde{V}_{\geq 0}^{-1} - V_{\geq 0}^{-1}\| = 4B^2 \|\text{round}_{b''}((\widetilde{V}_{\geq 0})^{-1}) - V_{\geq 0}^{-1}\| \quad (26)$$

$$\leq 4B^2 (\|\text{round}_{b''}((\widetilde{V}_{\geq 0})^{-1}) - (\widetilde{V}_{\geq 0})^{-1}\| + \|(\widetilde{V}_{\geq 0})^{-1} - V_{\geq 0}^{-1}\|) \quad (27)$$

$$\leq 4B^2 (n2^{-b''} + \frac{2^{-b''} \|V_{\geq 0}^{-1}\|}{1 - 2^{-b''} \|V_{\geq 0}^{-1}\|} \|V_{\geq 0}^{-1}\|) \quad (28)$$

$$\leq 4B^2 (n2^{-b''} + \frac{2^{-b''} B'}{(1/2)} B') \quad (29)$$

$$\leq 4B^2 (2n2^{-b''} (B')^2) =: f_2(b''). \quad (30)$$

By our choice of b'' in Line 1, the advertised error bound for the output follows. Note that $B = 2^{O^*(a(dn)^3)}$ in the worst case.

Complexity. The running time of JNF in Step 1 is $O^*((dn)^{\omega+3}a + (dn)^4a^2 + (dn)^\omega b'')$. Step 2 does not involve any computation. The time taken to exactly invert $\widetilde{V}_{\geq 0}$ in Step 3 using Theorem 17 (after pulling out the common dyadic denominator to obtain an integer matrix) is $O^*((dn)^3 \cdot (b'' + a(dn)^3 + b))$ by the estimate $\kappa(\widetilde{V}_{\geq 0}) = 2^{O^*(a(dn)^3)}$ which follows from (3) and the bound on the first term above. The time taken to round down the entries of $\widetilde{V}_{\geq 0}^{-1}$ to b'' bits is $O^*((dn)^2 b'')$. The time taken to multiply together the three matrices is $O^*((dn)^\omega b'')$. Thus, the total number of bit operations is dominated by

$$O^*((dn)^{\omega+3}a + (dn)^4a^2 + (dn)^3b'') = O^*((dn)^6a + (dn)^4a^2 + (dn)^3b),$$

as advertised. \blacktriangleleft

► Corollary 19 (Spectral Factorization of Non-Monic Polynomials). *Suppose $P(x) = VV^*x^{2d} + \sum_{i=0}^{2d-1} x^i P_i$ is a positive semidefinite Hermitian matrix polynomial with $V, P_i \in \mathbb{Q}^{n \times n} \langle a/a \rangle$ with a common denominator and V invertible. Then an approximate spectral factorization of $P(x)$ accurate to $b - O^*(a)$ bits (as in Theorem 18) can be computed in expected $O^*((dn)^6an + (dn)^4(an)^2 + (dn)^3b)$ bit operations.*

Proof. The rescaled polynomial $\tilde{P}(x) := x^{2d}I + \sum_{i=0}^{2d-1} x^i V^{-1} P_i V^{-*}$ is also positive semi-definite, has coefficients in $\mathbb{Q}^{n \times n} \langle an/an \rangle$, and is monic. Applying Theorem 18 yields an approximate spectral factor \tilde{Q} , for which $Q(x) = V\tilde{Q}(x)V^*$ is an approximate spectral factor of $P(x)$ with at most a loss of $O^*(a)$ bits of accuracy. \blacktriangleleft

4 Discussion and Future Work

For historical context, proving bit complexity bounds on fundamental linear algebra computations (such as inversion, polynomial matrix inversion, Hermite/Smith/Frobenius normal forms [30, 29, 44, 41, 42, 21, 46, 28]) has been a vibrant topic in theoretical computer science and symbolic computation since the 70's, with near-optimal arithmetic and bit complexity bounds being obtained for several of these problems within the last decade (e.g. [43]).

However, this program did not reach the same level of completion for problems of a spectral nature, such as the ones studied in this paper. While the polynomial time bounds obtained in this paper are modest, we hope they will stimulate further work on these fundamental problems, as well as the important special case of efficiently diagonalizing a diagonalizable matrix in the forward error model, which remains unresolved (in that we don't know the correct exponent of n ; the recent work [4] obtains nearly matrix multiplication time for the backward error formulation of the problem).

Some concrete questions left open by this work are:

1. Improve the running time for computing the JNF of a general matrix. The best known running time for computing the eigenvalues of a matrix is roughly $O(n^{\omega+1}a)$ [36], so this seems like a reasonable goal to shoot for. The current bottleneck is the bound of $2^{O^*(an^3)}$ on the condition number of the similarity V , which could conceivably be improved to $2^{O^*(an^2)}$.
2. Improve the running time for computing the JNF of the block companion matrix of a matrix polynomial by exploiting its special structure, particularly (19). This would yield faster algorithms for spectral factorization.

References

- 1 Sigal Ar and Jin-Yi Cai. Reliable benchmarks using numerical instability. In *SODA*, pages 34–43, 1994.
- 2 Erin M Aylward, Sleiman M Itani, and Pablo A Parrilo. Explicit sos decompositions of univariate polynomial matrices and the kalman-yakubovich-popov lemma. In *2007 46th IEEE Conference on Decision and Control*, pages 5660–5665. IEEE, 2007.
- 3 Mihály Bakonyi and Hugo J Woerdeman. *Matrix completions, moments, and sums of Hermitian squares*, volume 37. Princeton University Press, 2011.
- 4 Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 529–540. IEEE, 2020.
- 5 Dmitry Batenkov. On the norm of inverses of confluent vandermonde matrices. *arXiv preprint*, 2012. [arXiv:1212.0172](https://arxiv.org/abs/1212.0172).
- 6 Grigoriy Blekherman, Daniel Plaumann, Rainer Sinn, and Cynthia Vinzant. Low-rank sum-of-squares representations on varieties of minimal degree. *International Mathematics Research Notices*, 2019(1):33–54, 2019.
- 7 Louis Brand. The companion matrix and its properties. *The American Mathematical Monthly*, 71(6):629–634, 1964.
- 8 Jin-yi Cai. Computing jordan normal forms exactly for commuting matrices in polynomial time. *International Journal of Foundations of Computer Science*, 5(03n04):293–302, 1994.
- 9 Man-Duen Choi, Tsit Yuen Lam, and Bruce Reznick. Sums of squares of real polynomials. In *Proceedings of Symposia in Pure mathematics*, volume 58, pages 103–126. American Mathematical Society, 1995.
- 10 Michael A Dritschel and James Rovnyak. The operator fejér-riesz theorem. In *A glimpse at Hilbert space operators*, pages 223–254. Springer, 2010.
- 11 Lasha Ephremidze. An elementary proof of the polynomial matrix spectral factorization theorem. *Proceedings. Section A, Mathematics-The Royal Society of Edinburgh*, 144(4):747, 2014.
- 12 Lasha Ephremidze, Gigla Janashia, and Edem Lagvilava. A simple proof of the matrix-valued fejér-riesz theorem. *Journal of Fourier Analysis and Applications*, 15(1):124–127, 2009.

- 13 Lasha Ephremidze, Faisal Saied, and Ilya Matvey Spitkovsky. On the algorithmization of janashia-lagvilava matrix spectral factorization method. *IEEE Transactions on Information Theory*, 64(2):728–737, 2017.
- 14 Walter Gautschi. On inverses of vandermonde and confluent vandermonde matrices. *Numerische Mathematik*, 4(1):117–123, 1962.
- 15 Mark Giesbrecht. Nearly optimal algorithms for canonical matrix forms. *SIAM Journal on Computing*, 24(5):948–969, 1995.
- 16 Mark Giesbrecht and Arne Storjohann. Computing rational forms of integer matrices. *Journal of Symbolic Computation*, 34(3):157–172, 2002.
- 17 Isabelle Gil. Computation of the jordan canonical form of a square matrix (using the axiom programming language). In *Papers from the international symposium on Symbolic and algebraic computation*, pages 138–145, 1992.
- 18 Israel Gohberg, Peter Lancaster, and Leiba Rodman. Spectral analysis of selfadjoint matrix polynomials. *Annals of Mathematics*, 112(1):33–71, 1980.
- 19 Israel Gohberg, Peter Lancaster, and Leiba Rodman. *Matrix polynomials*. Springer, 2005.
- 20 Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.
- 21 Somit Gupta and Arne Storjohann. Computing hermite forms of polynomial matrices. In *Proceedings of the 36th international symposium on Symbolic and algebraic computation*, pages 155–162, 2011.
- 22 Christoph Hanselka and Rainer Sinn. Positive semidefinite univariate matrix polynomials. *Mathematische Zeitschrift*, 292(1-2):83–101, 2019.
- 23 Douglas P Hardin, Thomas A Hogan, and Qiyu Sun. The matrix-valued riesz lemma and local orthonormal bases in shift-invariant spaces. *Advances in Computational Mathematics*, 20(4):367–384, 2004.
- 24 Henry Helson and David Lowdenslager. Prediction theory and fourier series in several variables. *Acta mathematica*, 99(1):165–202, 1958.
- 25 G Janashia, E Lagvilava, and L Ephremidze. Matrix spectral factorization and wavelets. *Journal of Mathematical Sciences*, 195(4):445–454, 2013.
- 26 Gigla Janashia, Edem Lagvilava, and Lasha Ephremidze. A new method of matrix spectral factorization. *IEEE Transactions on information theory*, 57(4):2318–2326, 2011.
- 27 Erich Kaltofen, M Krishnamoorthy, and B David Saunders. Fast parallel algorithms for similarity of matrices. In *Proceedings of the fifth ACM symposium on Symbolic and algebraic computation*, pages 65–70, 1986.
- 28 Erich L Kaltofen and Arne Storjohann. The complexity of computational problems in exact linear algebra, 2015.
- 29 Ravindran Kannan. Solving systems of linear equations over polynomials. *Theoretical Computer Science*, 39:69–88, 1985.
- 30 Ravindran Kannan and Achim Bachem. Polynomial algorithms for computing the smith and hermite normal forms of an integer matrix. *siam Journal on Computing*, 8(4):499–507, 1979.
- 31 Heinz Langer. Factorization of operator pencils. *Acta Sci. Math.(Szeged)*, 38(1–2):83–96, 1976.
- 32 TY Li, Zhinan Zhang, and Tianjun Wang. Determining the structure of the jordan normal form of a matrix by symbolic computation. *Linear algebra and its applications*, 252(1-3):221–259, 1997.
- 33 Kurt Mahler et al. An inequality for the discriminant of a polynomial. *Michigan Mathematical Journal*, 11(3):257–262, 1964.
- 34 Patrick Ozello. Calcul exact des formes de jordan et de frobenius d’une matrice, 1987.
- 35 Victor Y Pan. Univariate polynomials: nearly optimal algorithms for numerical factorization and root-finding. *Journal of Symbolic Computation*, 33(5):701–733, 2002.
- 36 Victor Y Pan and Zhao Q Chen. The complexity of the matrix eigenproblem. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 507–516, 1999.

- 37 Jean-Louis Roch and Gilles Villard. Fast parallel computation of the jordan normal form of matrices. *Parallel processing letters*, 6(02):203–212, 1996.
- 38 Murray Rosenblatt. A multi-dimensional prediction problem. *Arkiv för matematik*, 3(5):407–424, 1958.
- 39 Marvin Rosenblum and James Rovnyak. The factorization problem for nonnegative operator valued functions. *Bulletin of the American Mathematical Society*, 77(3):287–318, 1971.
- 40 Ali H Sayed and Thomas Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.
- 41 Arne Storjohann. An $\mathcal{O}(n^3)$ algorithm for the frobenius normal form. In *Proceedings of the 1998 international symposium on Symbolic and algebraic computation*, pages 101–105, 1998.
- 42 Arne Storjohann. Deterministic computation of the frobenius form. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 368–377. IEEE, 2001.
- 43 Arne Storjohann. On the complexity of inverting integer and polynomial matrices. *computational complexity*, 24(4):777–821, 2015.
- 44 Arne Storjohann and George Labahn. A fast las vegas algorithm for computing the smith normal form of a polynomial matrix. *Linear Algebra and its Applications*, 253(1-3):155–173, 1997.
- 45 Vladimir Andreevich Yakubovich. Factorization of symmetric matrix polynomials. In *Soviet Math. Dokl.*, volume 11(5), pages 1261–1264, 1970.
- 46 Wei Zhou, George Labahn, and Arne Storjohann. A deterministic algorithm for inverting a polynomial matrix. *Journal of Complexity*, 31(2):162–173, 2015.