Pontifícia Universidade Católica
do Rio de Janeiro

Antonio Maria Vasconcellos Mac Dowell da Costa

Recovery of tridiagonal matrices from spectral data

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Matemática, do Departamento de Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática.

Advisor: Prof. Carlos Tomei

Rio de Janeiro,
February 2024

Pontifícia Universidade Católica
do Rio de Janeiro

Antonio Maria Vasconcellos Mac Dowell da Costa

Recovery of tridiagonal matrices from spectral data

Thesis presented to the Programa de Pós–graduação em Matemática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Matemática. Approved by the Examination Committee:

Prof. Carlos Tomei
Advisor
Departamento de Matemática – PUC-Rio

Prof. Nicolau Corção Saldanha
Departamento de Matemática – PUC-Rio

Prof. Peter C. Gibson
York University

Rio de Janeiro, February the 29th, 2024

Antonio Maria Vasconcellos Mac Dowell da Costa

Majored in mathematics by Pontifícia Universidade Católica do Rio de Janeiro (Rio de Janeiro, Brasil).

Bibliographic data

Mac Dowell da Costa, Antonio Maria Vasconcellos

Recovery of tridiagonal matrices
from spectral data / Antonio Maria Vasconcellos Mac Dowell da Costa; advisor: Carlos Tomei. – 2024.

46 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, 2024.

Inclui bibliografia

1. Matemática – Teses. 2. Matrizes de Jacobi. 3. Algoritmos Espectrais Inversos. 4. Matrizes tridiagonais. 5. Variedades Isoespectrais. 6. Coordenadas Bidiagonais. I. Tomei, Carlos. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 530

In loving memory of my grandmother Mara.

Acknowledgments

I would like to specially thank my advisor and friend Carlos Tomei for all the support, patience and knowledge.

I would like to thank my parents, for all the love and education.

I also would like to thank my girlfriend Sofia for her love and support during the hard moments of this work.

I deeply thank PUC-Rio for the scholarship and all its members for making the great work environment that I have been using in the latest years.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Mac Dowell da Costa, Antonio Maria Vasconcellos; Tomei, Carlos (Advisor). **Recovery of tridiagonal matrices from spectral data**. Rio de Janeiro, 2024. 46p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Algorithms relating Jacobi matrices and spectral variables are standard objects in numerical analysis. The recent discovery of bidiagonal coordinates led to the search of an appropriate algorithm for these new variables. The new algorithm is presented and compared with previous techniques.

Keywords

Jacobi Matrices; Inverse Spectral Algorithms; Tridiagonal matrices; Isospectral Manifolds; Bidiagonal Coordinates.

Resumo

Mac Dowell da Costa, Antonio Maria Vasconcellos; Tomei, Carlos. **Recuperação de matrizes tridiagonais a partir de dados espectrais**. Rio de Janeiro, 2024. 46p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

A identificação algorítmica de matrizes de Jacobi a partir de variáveis espectrais é um tema tradicional de análise numérica. Uma nova representação, as coordenadas bidiagonais, naturalmente exigiu que fosse considerado um novo algoritmo. O algoritmo é apresentado e confrontado com as técnicas habituais.

Palavras-chave

Matrizes de Jacobi; Algoritmos Espectrais Inversos; Matrizes tridiagonais; Variedades Isoespectrais; Coordenadas Bidiagonais.

Table of contents

List of figures

Empty your mind, be formless, shapeless, like water. If you put water into a cup, it becomes the cup. You put water into a bottle and it becomes the bottle. You put it in a teapot it becomes the teapot. Now, water can flow or it can crash. Be water, my friend.

Bruce Lee, *The Pierre Berton Show, 1971.*

1
Introduction

We consider the recovery of Jacobi matrices from *spectral data* [1–3]. Our main theoretical tool is geometric — the identification of the closure of Jacobi matrices with a special convex polytope, a *permutohedron* [4–6]. We consider different types of spectral data and, from the identification above, we describe regions of instability for each type.

Traditional spectral data associated with a Jacobi matrix consists of its spectrum and first coordinates of appropriately normalized eigenvectors. This choice is the discrete counterpart of the so called *inverse variables* for the Schrödinger operator on the line with a decaying potential, and fits well with some applications in numerical integration of functions [7]. Geometric properties of such data are presented in Chapter 2 to keep the material self-contained. In Chapter 3, we describe the celebrated RKPW (Rutishauser-Kahan-Pal-Walker) algorithm, which retrieves a matrix from this spectral data.

From the geometric approach, two limitations of these standard variables become evident. The first is the fact that they break down at reduced matrices (i.e., matrices with some entry $(i, i+1)$ equal to zero). The second is a stability issue: small perturbations of the spectral data give rise to very different matrices. These difficulties probably did not receive much attention due to the fact that the RKPW algorithm always provides an answer. In particular, the algorithm provides an (incorrect) output for incompatible data: a non-simple spectrum implies that at least one of the first coordinates of eigenvectors is zero. Yet, RKPW generates a tridiagonal symmetric matrix out of such spectrum and nonzero first eigenvector coordinates.

In Chapter 4 we consider alternative spectral data, the *bidiagonal coordinates*, introduced in [3]. Jacobi matrices form an open subset of the *tridiagonal isospectral manifold* [4] and bidiagonal coordinates provide charts for the manifold. Convergence issues of QR-type eigenvalue algorithms are reduced to local theory [8, 9]. We present in Section 4.1 the basic theoretical information, in preparation for the description of an algorithm in Section 4.2 which recovers tridiagonal symmetric matrices from bidiagonal coordinates.

Finally, Chapter 5 provides examples and comparisons between both algorithms. Both algorithms have distinct weak points, geometrically well

indicated in figures 2.2 and 4.3. There is space for substantial improvement.

1.1 Notation

We denote by $\mathcal{S}$ the vector space of $n \times n$ real, symmetric matrices and by $\mathcal{J} \subset \mathcal{S}$ be the set of Jacobi matrices, more precisely, tridiagonal matrices $J \in \mathcal{S}$ for which $J_{i,i+1} = J_{i+1,i} > 0, i = 1, \ldots n-1$. For a real diagonal matrix Λ with simple spectrum, let the entries of $\lambda = (\lambda_1 = \Lambda_{11} < \ldots < \lambda_n = \Lambda_{nn}) \in \mathbb{R}^n$ be its eigenvalues. The set $\mathcal{J}_\Lambda$ consists of Jacobi matrices with spectrum λ. As usual, $\overline{\mathcal{J}_\Lambda} \subset \mathcal{S}$ is the closure of $\mathcal{J}_\Lambda$ in $\mathcal{S}$. Finally, $\mathcal{T}_\Lambda \subset \mathcal{S}$ is the set of tridiagonal symmetric matrices with spectrum λ.

We denote by $O(n)$ the *orthogonal group*, consisting of $n \times n$ matrices X, such that $X^T X = I$, the identity. The *special orthogonal group* $SO(n) \subset O(n)$ consists of orthogonal matrices with determinant one; $Up^+(n)$ is the group of *upper triangular* matrices with strictly positive diagonal entries; $Lo^1(n)$ is the group of *lower triangular* matrices with diagonal entries equal to one, and finally $\mathcal{E}(n) \subset O(n)$ is the subgroup of *signed diagonal matrices*, consisting of diagonal matrices with ± 1 at its entries. The matrix dimension – the index n – is frequently omitted.

Vector spaces are real, finite dimensional. Euclidean space $\mathbb{R}^n$ is endowed with the standard inner product, with associated $\mathcal{L}^2$-norm denoted by $\| \cdot \|$. We also denote by $e_1, e_2, \ldots, e_n$ the canonical vectors of $\mathbb{R}^n$.

2
Geometry of Jacobi matrices

The starting point of this text is the geometry of Jacobi matrices, given by different kind of coordinates, which in turn give rise to inverse algorithms related to *spectral data*.

2.1
Jacobi matrices with fixed spectrum

We begin with a well known fact from linear algebra [2, 10].

Proposition 1. *Jacobi matrices have simple spectrum. The first and last coordinates of their eigenvectors are nonzero. Also, dropping the signs does not change the spectrum.*

Thus, for a Jacobi matrix $J \in \mathcal{J}$, every eigenvalue λ has a *unique* normalized eigenvector c with (strictly) positive first coordinates.

Theorem 1. *[1, 2] A Jacobi matrix is determined by its spectrum and the first coordinates of its appropriately normalized eigenvectors. Explicitly, set*

$$\mathcal{M} = \{(\lambda, c) = (\lambda_1, \ldots, \lambda_n, c_1, \ldots, c_n) \mid \lambda_1 < \ldots < \lambda_n \text{ and } \sum_i c_i^2 = 1,\ c_i > 0\}$$

and $\psi : \mathcal{J} \to M$, $\psi(J) = (\lambda_1, \ldots, \lambda_n, c_1, \ldots, c_n) = (\lambda, c)$, where λ_i and c_i denote the i-th smallest eigenvalue and the first coordinate of the associated normalized eigenvector of J, respectively. Then ψ is a diffeomorphism.

The hardest part of the proof is the construction of the inverse map, which we sketch. Let Λ be a diagonal matrix with entries given by λ and K be the invertible matrix with columns $c, \Lambda c, \ldots, \Lambda^{n-1}c$. By a standard argument with Vandermonde determinants, K is invertible. The unique QR-factorization of an invertible matrix then gives

$$K = QR, \quad Q \in O(n),\ R \in Up^+(n)\ .$$

Define $\tilde{J} = Q^T \Lambda Q$. The required matrix J is obtained by dropping the signs of the off-diagonal entries of $\tilde{J}$.

This (inverse) algorithm recovers Jacobi matrices from (λ, c), the eigenvalues and first coordinates of normalized eigenvectors. We will refer to vectors c in the pair (λ, c) as *Moser vectors*. The algorithm is related to the *Lanczos method*, essentially a Gram-Schmidt process with appropriate simplifications.

The *inverse variables* (λ, c) come up naturally in many contexts in numerical linear algebra [1,11]. They are also discrete counterparts of the usual inverse variables associated with the point spectrum of Schrödinger operators — the Moser vector c plays the role of the familiar norming constants [12].

As we shall see in Section 2.2 (and especially Figure 2.2), the algorithm is numerically unstable for coordinates c_i's close to zero, as the restriction to a fixed spectrum

$$\psi_\Lambda : \mathcal{J}_\Lambda \to \mathcal{C} = \{c \in \mathbb{R}^n, \sum_i c_i^2 = 1, c_i > 0\}$$

extends continuously to the closure $\overline{\mathcal{J}_\Lambda}$ but is not injective.

2.2 Reduced matrices and the permutohedron

From Theorem 1, the set $\mathcal{J}_\Lambda \subset \mathcal{S}$ is diffeomorphic to the positive octant of the sphere in $\mathbb{R}^n$, which is in turn diffeomorphic to $\mathbb{R}^{n-1}$. We consider its closure $\overline{\mathcal{J}_\Lambda}$, starting with the case $n = 3$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$.

There are six diagonal matrices in $\overline{\mathcal{J}_\Lambda}$, corresponding to the six permutations of three symbols. The elements in the boundary of $\overline{\mathcal{J}_\Lambda}$ are *reduced matrices* – at least one (hence two, by symmetry) off-diagonal entries are equal to zero. For $n = 3$, the matrices in $\overline{\mathcal{J}_\Lambda}$ with entries (12) and (21) equal to zero have one of its three eigenvalues in entry (11). Once such eigenvalue is fixed, the block consisting of entries in rows and columns 2 and 3 has spectrum given by the remaining two eigenvalues, and it is easy to see that such set is diffeomorphic to a half-circle, as entry (23) is greater or equal to zero, and negative values for these entries correspond to the other half of the circle. In Figure 2.1, three boundary arcs are labeled 0+, indicating that entries (12) and (21) are zero and entries (23) and (32) are positive. Accordingly, diagonal and Jacobi matrices are respectively labeled 00 and ++.

It is not surprising then (and the general case will be described later) that the boundary $\partial\overline{\mathcal{J}_\Lambda} \subset \mathcal{S}$ consists of a hexagon with curved sides. From Theorem 1, non-reduced (Jacobi) matrices form a set diffeomorphic to $\mathbb{R}^2$, in accordance with our final claim: $\overline{\mathcal{J}_\Lambda}$ is a closed hexagon.

We now consider the natural extension of the map ψ_Λ for reduced matrices $J \in \partial\mathcal{J}_\Lambda$ with spectral decomposition $J = Q^T \Lambda Q$, represented in

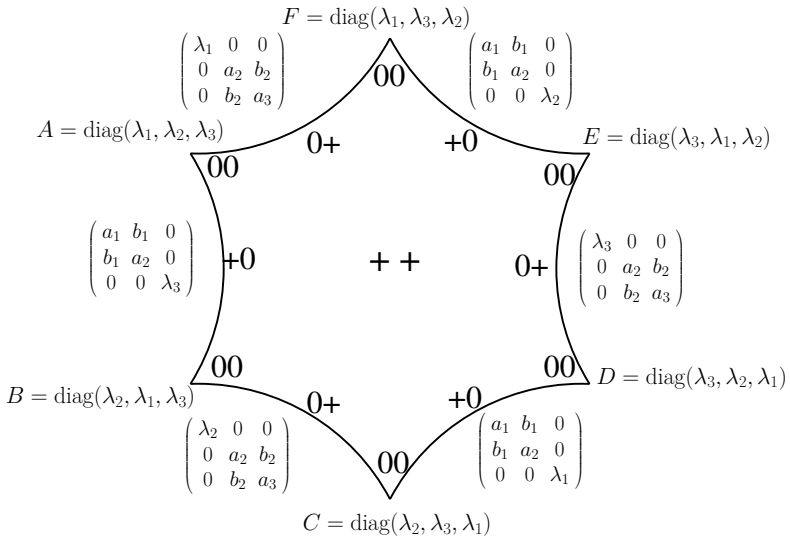

Figure 2.1: A representation of the (topological) hexagon $\overline{\mathcal{J}_\Lambda}$.

Figure 2.2. If the eigenvalue λ_i is in position (11), i.e., J is in a cell of the form 0+, we must have $c = e_i$, a canonical vector: the map extension of ψ_Λ is not injective. Matrices of the form +0 are taken injectively to arcs where some coordinate equals zero, as the reader can easily verify. Summarizing, the extension $\psi_\Lambda : \overline{\mathcal{J}_\Lambda} \to \overline{\mathcal{C}}$ is surjective, but not invertible.

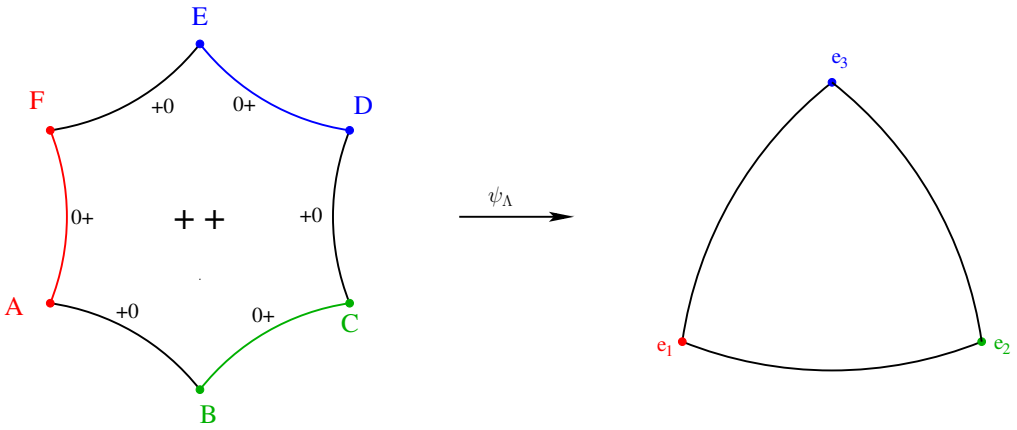

Figure 2.2: The extension of the Moser map, ψ_Λ.

More generally, $\overline{\mathcal{J}_\Lambda}$ can be identified with the *permutohedron* $\mathcal{P}_\Lambda \subset \mathbb{R}^n$, the convex hull of the $n!$ points $(\lambda_{\pi(1)}, \ldots, \lambda_{\pi(n)})$, where $\pi \in S_n$ is a permutation in n elements [13]. These points are indeed vertices of $\mathcal{P}_\Lambda$.

Theorem 2. *[4–6] There is a homeomorphism from $\overline{\mathcal{J}_\Lambda}$ to $\mathcal{P}_\Lambda$ which restricts to a diffeomorphism between interiors.*

In the original proof [4], both boundary and interior of $\overline{\mathcal{J}_\Lambda}$ are shown to be PL-homeomorphic to the boundary and interior of the sphere $\mathcal{S}^{n-1}$. The result then follows from Schoenflies theorem, a generalization of the familiar

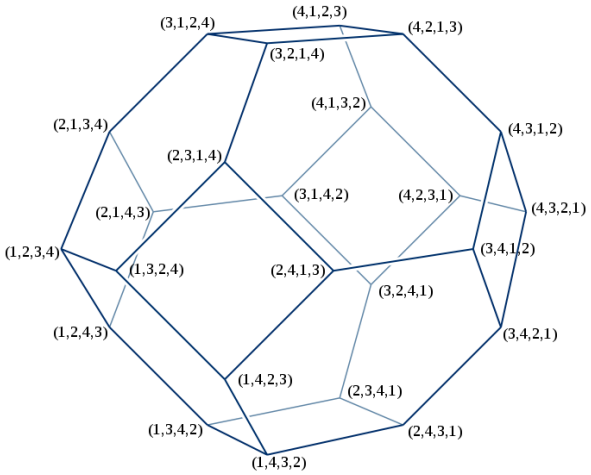

Figure 2.3: The permutohedron, $n = 4$

Jordan theorem for curves in the plane. Later, Bloch, Flaschka and Ratiu [5] obtained an explicit homeomorphism, which we now present.

For $J \in \mathcal{J}_\Lambda$, consider the spectral decomposition $J = Q^T \Lambda Q$, where Q is well defined once we prescribe that its first column equals the Moser vector of J. For Jacobi matrices, the *BFR map* is

$$BFR : \mathcal{J}_\Lambda \to V = \{v \in \mathbb{R}^n \mid \sum_i v_i = \sum_i \lambda_i\}$$

$$J \mapsto \operatorname{diag} \tilde{J} = \operatorname{diag} Q\Lambda Q^T \ .$$

This map extends to $\overline{\mathcal{J}_\Lambda}$: first coordinates may become zero, and the eigenvector normalization are known only up to sign, but this is innocuous – diagonal entries of $\tilde{J}$ are well defined. The result then follows from the (nontrivial) interpretation of the BFR map as a moment map of a Hamiltonian torus action and then using a celebrated result of Atiyah [14]. Leite and Tomei [6] obtained a simpler argument showing that the extension of the BFR map to $\overline{\mathcal{J}_\Lambda}$ indeed satisfies the properties stated in Theorem 2. In Figure 2.4, we display the BFR map, for $n = 3$. The labels A, B, C, D, E and F on the right represent the images of the diagonal matrices denoted by the same symbols on the left. As shown in [6] and indicated in the Figure, points of $\mathcal{J}_\Lambda$ near reduced matrices are mapped closer to the boundary of the hexagon $\mathcal{P}_\Lambda$.

The above theorem suggests different inverse variables for $\overline{\mathcal{J}_\Lambda}$. Inversion, however, is not explicit, requiring the inversion of polynomials of high degree. Moreover, points in the boundary of the domain are critical, and the map contracts to the boundary [6].

What about the Moser vector c for general reduced matrices $\tilde{J}$? A reduced matrix splits into blocks associated with invariant subspaces generated by

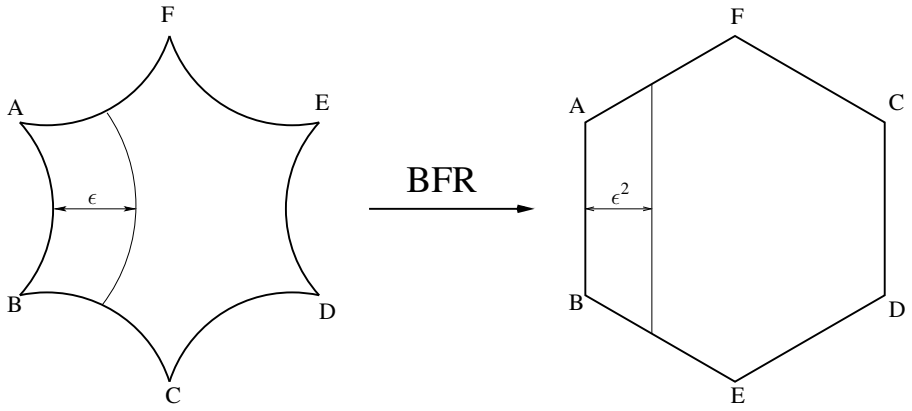

Figure 2.4: The BFR map for $n = 3$. On the left, Figure 2.1.

canonical vectors. More concretely, if, say, $\tilde{J}_{i,i+1} = \tilde{J}_{j,j+1} = 0$, $\tilde{J}$ splits in three (unreduced) blocks. For the first block, the first coordinates of the eigenvectors are nonzero, but this is not the case of the subsequent blocks — injectivity of $\tilde{J} \in \mathcal{J}_\Lambda \mapsto c \in \overline{\mathcal{C}}$ breaks down. In particular, the Moser vector does not provide a coordinate system which is injective in the neighborhood of a diagonal matrix, as shown in Figure 2.2 for $n = 3$.

The arguments above keep λ fixed — can one say something about equal eigenvalues? By Proposition 1, multiple eigenvalues imply that some entries of c must be zero. In particular, the map $\tilde{J} \in \overline{\mathcal{J}} \to \overline{\mathcal{C}}$ is neither injective or surjective.

In Chapter 4, we describe *bidiagonal coordinates* which behave well locally at any reduced matrix and for which the inverse algorithm is explicit.

3
Reconstruction from Moser variables — the RKPW algorithm

From Theorem 1, there is a bijection between Jacobi matrices and pairs $(\lambda, c) \in \mathcal{M}$. Its proof suggests a concrete inverse algorithm, but there is an alternative with better stability properties [1]. Rather surprisingly, the RKPW algorithm (for Rutishauser, Kahan, Pal and Walker) does not require the simplicity of eigenvalues nor the positivity of the Moser vectors. But things are not that simple, as we shall see – instability arises close to boundaries. We first discuss the algorithm for $(\lambda, c) \in \mathcal{M}$, then consider degenerate data.

3.1
Recovering Jacobi matrices

Recall that the entries of λ are strictly increasing. For $\lambda, c \in \mathbb{R}^n$, denote by λ_k and c_k the k-th entries of λ and c, and by λ^k and c^k the vectors of the first k entries, respectively. The RKPW algorithm is inductive: to obtain a Jacobi matrix J_k associated with (λ^k, c^k), suppose that J_{k-1} associated with (λ^{k-1}, c^{k-1}) is known.

Consider the auxiliary $(k+1) \times (k+1)$ matrix

$$S = \begin{pmatrix} 1 & \|c^{k-1}\| e_1^T & c_k \\ \|c^{k-1}\| e_1 & J_{k-1} & 0 \\ c_k & 0 & \lambda_k \end{pmatrix}. \tag{3-1}$$

Clearly S is real, symmetric, but the corner entries $(1, k+1)$ and $(k+1, 1)$ spoil tridiagonality. As we shall see, a sequence of $(k-1)$ appropriate conjugations of S by *Givens rotations*[1] chase the bulge introduced by the corner entries, yielding a matrix

$$\tilde{J} = \begin{pmatrix} 1 & \|c^k\| e_1^T \\ \|c^k\| e_1 & J_k \end{pmatrix}.$$

Repeating the process, eventually one obtains the required $J = J_n$.

There are three things to show:

[1]Let P_{ij} be the plane with (ordered) basis $e_i, e_j,\ i < j$. A Givens rotation $G_{ij}(\theta) \in SO$ is a rotation of θ on P_{ij}, which is equal to the identity on the orthogonal complement of P_{ij}.

(1) J is tridiagonal (Proposition 3).
(2) J has inverse variables (λ, c) (Proposition 2).
(3) J is a Jacobi matrix (Proposition 4).

Proposition 2. *Fact (1) implies (2).*

The proof requires essentially no detailed knowledge of the Givens rotations leading to $\tilde{J}$.

Proof. Suppose that J_{n-1} is the Jacobi matrix associated with variables (λ^{n-1}, c^{n-1}) and that from the sequential conjugations,

$$\tilde{J} = \begin{pmatrix} 1 & \|c\|e_1^T \\ \|c\|e_1 & J_n \end{pmatrix} = \tilde{G}^T S \tilde{G},$$

where $J_n = J$ is tridiagonal and

$$\tilde{G} = \begin{pmatrix} 1 & 0 \\ 0 & G \end{pmatrix} \tag{3-2}$$

with G being a product of Givens transformations. Write

$$J_n = Q_n^T \Lambda_n Q_n, \quad J_{n-1} = Q_{n-1} \Lambda_{n-1} Q_{n-1}, \quad \hat{Q}_{n-1} = \begin{pmatrix} Q_{n-1} & \\ & 1 \end{pmatrix},$$

for orthogonal matrices Q_{n-1} and Q_n with the first having strictly positive entries in its first column. We then have

$$\tilde{J} = \begin{pmatrix} 1 & \|c\|e_1^T \\ \|c\|e_1 & J_n \end{pmatrix} = \tilde{G}^T S \tilde{G} = \begin{pmatrix} 1 & 0 \\ 0 & G^T \end{pmatrix} \begin{pmatrix} 1 & \|c^{n-1}\|e_1^T & c_n \\ \|c^{n-1}\|e_1 & J_{n-1} & 0 \\ c_n & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & G \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \begin{pmatrix} \|c^{n-1}\|e_1^T & c_n \end{pmatrix} G \\ G^T \begin{pmatrix} \|c^{n-1}\|e_1 \\ c_n \end{pmatrix} & G^T \hat{Q}_{n-1}^T \Lambda_n \hat{Q}_{n-1} G \end{pmatrix}.$$

so that

$$\|c\|e_1 = G^T \begin{pmatrix} \|c^{n-1}\|e_1 \\ c_n \end{pmatrix} \quad \text{and} \quad Q_n E = \hat{Q}_{n-1} G \,,$$

for some signed diagonal matrix $E \in \mathcal{E}$, as Λ_n has simple eigenvalues. Clearly, the eigenvalues of $J = J_n$ are the entries of λ, as $\Lambda_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. We now consider the Moser vector:

$$Q_n e_1 = \hat{Q}_{n-1} G\, E\, e_1 = \pm \hat{Q}_{n-1} G\, e_1 = \pm \frac{1}{\|c\|} \begin{pmatrix} c^{n-1} \\ c_n \end{pmatrix} = \pm \frac{c}{\|c\|}. \tag{3-3}$$

Properly normalizing Q_n and the Moser vector c, the matrix J has (λ, c) as its inverse variables. □

Proposition 3. *[1] There is a sequence of Givens rotations leading to a symmetric, tridiagonal matrix $\tilde{J}$.*

Proof. We describe the inductive step from J_{k-1} to J_k. Conjugate S defined in equation 3-1 by a Givens rotation in the plane $P_{2,k+1}$ to obtain a zero in entry $(1, k+1)$ of the resulting matrix $S^{(1)} = G_{2,k+1}^T \; S \; G_{2,k+1}$. Set

$$J_{k-1} = \begin{pmatrix} a_1 & b_1 & & \\ b_1 & a_2 & \ddots & \\ & \ddots & \ddots & b_{k-2} \\ & & b_{k-2} & a_{k-1} \end{pmatrix}$$

and write $\gamma = \cos\theta, \sigma = \sin\theta$. Then

$$S^{(1)} = \begin{pmatrix} 1 & -c_k\sigma + \|c^{k-1}\|\gamma & 0 & 0 & \dots & 0 & c_k\gamma + \|c^{k-1}\|\sigma \\ -c_k\sigma + \|c^{k-1}\|\gamma & \gamma^2 a_1 + \sigma^2\lambda_k & \gamma b_1 & 0 & \dots & 0 & \gamma\sigma(a_1 - \lambda_k) \\ 0 & \gamma b_1 & a_2 & b_2 & & 0 & \sigma b_1 \\ 0 & 0 & b_2 & a_3 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & b_{k-2} & 0 \\ 0 & 0 & 0 & & b_{k-2} & a_{k-1} & 0 \\ c_k\gamma + \|c^{k-1}\|\sigma & \gamma\sigma(a_1 - \lambda_k) & \sigma b_1 & \dots & 0 & 0 & \sigma^2 a_1 + \gamma^2\lambda_k \end{pmatrix}, \tag{3-4}$$

from which we determine θ (or better, its sine and cosine) such that

$$c_k\gamma + \|c^{k-1}\|\sigma = 0 \; . \tag{3-5}$$

For the solutions $\pm\theta \in [0, 2\pi)$,

$$\gamma = \pm\frac{\|c^{k-1}\|}{\sqrt{\|c^{k-1}\|^2 + (c_k)^2}}, \quad \sigma = \mp\frac{c_k}{\sqrt{\|c^{k-1}\|^2 + (c_k)^2}} \; . \tag{3-6}$$

We choose θ so that the (12) entry of $S^{(1)}$ is positive,

$$-c_k\sigma + \|c^{k-1}\|\gamma = \|c^k\| \; .$$

Therefore, as γ and σ have opposite signs and $c_k > 0$, θ should be chosen so that its cosine is greater or equal to zero (and its sine less or equal). To summarize, after the first conjugation, we obtain

$$S^{(1)} = \begin{pmatrix} 1 & \|c^k\| & 0 & 0 & \dots & 0 & 0 \\ \|c^k\| & \gamma^2 a_1 + \sigma^2 \lambda_k & \gamma b_1 & 0 & \dots & 0 & \gamma\sigma(a_1 - \lambda_k) \\ 0 & \gamma b_1 & a_2 & b_2 & & 0 & \sigma b_1 \\ 0 & 0 & b_2 & a_3 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & b_{k-2} & 0 \\ 0 & 0 & 0 & & b_{k-2} & a_{k-1} & 0 \\ 0 & \gamma\sigma(a_1 - \lambda_k) & \sigma b_1 & \dots & 0 & 0 & \sigma^2 a_1 + \gamma^2 \lambda_k \end{pmatrix}, \qquad (3\text{-}7)$$

A second conjugation replaces bulges at entries $(2, k{+}1)$ and $(3, k{+}1)$ by bulges at $(3, k+1)$ and $(4, k+1)$. Repeat the process, bringing down the bulges in the last column (and last row), until reaching $\tilde{J}$. □

We illustrate the initial steps, for convenience. Clearly $J_1 = \lambda_1$, regardless of the value of c_1. To find J_2 out of $(\lambda^2 = (\lambda_1 < \lambda_2), c^2 = (c_1, c_2))$, apply one conjugation by a rotation $G_{2,3}$:

$$S = \begin{pmatrix} 1 & c_1 & c_2 \\ c_1 & \lambda_1 & 0 \\ c_2 & 0 & \lambda_2 \end{pmatrix} \mapsto \tilde{J} = \begin{pmatrix} 1 & \|c^2\| & 0 \\ \|c^2\| & * & * \\ 0 & * & * \end{pmatrix} = \begin{pmatrix} 1 & \|c^2\| e_1^T \\ \|c^2\| e_1 & J_2 \end{pmatrix}$$

For the step $(\lambda^2, c^2) \mapsto (\lambda^3, c^3)$, start with S having J_2 in its central block, indicated by asterisks. Apply Givens conjugations $G_{2,4}$ and $G_{3,4}$, the former mixing rows and columns 2 and 4 and the latter 3 and 4, obtaining $\tilde{J}$, with J_3 in its lower principal 3×3 block:

$$S = \begin{pmatrix} 1 & \|c^2\| & 0 & c_3 \\ \|c^2\| & * & * & 0 \\ 0 & * & * & 0 \\ c_3 & 0 & 0 & \lambda_3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & \|c^3\| & 0 & 0 \\ \|c^3\| & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix} \mapsto \tilde{J} = \begin{pmatrix} 1 & \|c^3\| & 0 & 0 \\ \|c^3\| & * & * & 0 \\ 0 & * & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

For any $x \in \mathbb{R}$, we write $\tilde{x}$ for some number very close to x, where we quantify the deviation as we proceed with the text. A matrix $\widetilde{M}$ has all its entries close to the entries of M.

We need a lemma.

Lemma 1. *Define the* $(n+1) \times (n+1)$ *tridiagonal symmetric matrix*

$$S(0) = \begin{pmatrix} 1 & \|c^{n-1}\| e_1^T & 0 \\ \|c^{n-1}\| e_1 & J_{n-1} & 0 \\ 0 & 0 & \lambda_n \end{pmatrix},$$

associated with eigenvalues $\lambda = (\lambda^{n-1}, \lambda_n)$ *and Moser vector* $c(0) = (c^{n-1}, 0)$. *Let* $S(\epsilon)$ *be the (unique) Jacobi matrix associated with* λ *and* $c(\epsilon) = (c^{n-1}, \epsilon)$ *obtained from Theorem 1. Then, for* $\epsilon = \tilde{0} > 0$, *we have* $S(\epsilon) = \widetilde{S(0)}$.

Proof. We follow the construction outlined after Theorem 1. The matrix $K(\epsilon)$ with columns $c(\epsilon), \Lambda c(\epsilon), \ldots, \Lambda^{n-1} c(\epsilon)$ is invertible, unless $\epsilon = 0$, for which its last row consists of zeros. Still, one obtains a smooth QR-factorization, where the last column of Q is simply a vector orthogonal to the previous columns, i.e., the canonical vector e_{n+1}. Thus, the RKPW algorithm is continuous as $\epsilon \to 0$: the result then follows. □

We finally prove (3): J is indeed a Jacobi matrix:

Proposition 4. *For the pair* $(\lambda, c) \in \mathcal{M}$ *as in Theorem 1, the matrix* $J = J_n$ *obtained by the RKPW algorithm is Jacobi.*

Proof. Denote the nontrivial entries of the Givens rotations by $\gamma = \cos\theta$ and $\sigma = \sin\theta$. For the transition $(\lambda^1, c^1) \mapsto (\lambda^2, c^2)$ we have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & -\sigma \\ 0 & \sigma & \gamma \end{pmatrix} \begin{pmatrix} 1 & c_1 & c_2 \\ c_1 & \lambda_1 & 0 \\ c_2 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & \sigma \\ 0 & -\sigma & \gamma \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \|c^2\| & 0 \\ \|c^2\| & \gamma^2\lambda_1 + \sigma^2\lambda_2 & \gamma\sigma(\lambda_1 - \lambda_2) \\ 0 & \gamma\sigma(\lambda_1 - \lambda_2) & \sigma^2\lambda_1 + \gamma^2\lambda_2 \end{pmatrix}$$

Since $\gamma = \frac{c_1}{\|c^2\|} > 0$, $\sigma = \frac{-c_2}{\|c^2\|} < 0$ and $\lambda_1 < \lambda_2$, the off-diagonal term $b_1 = \gamma\sigma(\lambda_1 - \lambda_2)$ is strictly positive, as entries of λ are increasing.

The inductive argument handles the transition $J_{n-1} \mapsto J_n$, where we assume that J_{n-1} is Jacobi.

Embed the matrix J_{n-1} in $S(0)$ of Lemma 1. To show that for $c_n > 0$, RKPW obtains a Jacobi matrix J_n embedded in $S(\epsilon)$ it suffices to consider small c_n, by continuity: if $c_n > 0$ is arbitrary and J_n is not Jacobi, then there exists some $\epsilon < \tilde{c}_n < c_n$ for which RKPW takes $(\lambda, c = (c_1, \ldots, \tilde{c}_n))$ to a reduced matrix $\tilde{J}$. But this is absurd, since a reduced matrix must have some entry of its Moser vector equal to zero.

Again by continuity, it suffices to prove the result for a Moser vector of the form $c = (\tilde{1}, \epsilon_2, \ldots, \epsilon_n)$, where $\epsilon_i = \tilde{0} > 0$. Taking ϵ_i small allows us to freely use expressions like 'x is of order ϵ'. From the definition 3-1,

$$S = \begin{pmatrix} 1 & \tilde{1} & 0 & 0 & \dots & 0 & \epsilon_n \\ \tilde{1} & \tilde{\lambda}_1 & b_1 & 0 & \dots & 0 & 0 \\ 0 & b_1 & \tilde{\lambda}_2 & b_2 & & 0 & 0 \\ 0 & 0 & b_2 & \tilde{\lambda}_3 & \ddots & \vdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & b_{n-2} & \vdots \\ 0 & 0 & 0 & \dots & b_{n-2} & \tilde{\lambda}_{n-1} & 0 \\ \epsilon_n & 0 & 0 & \dots & 0 & 0 & \lambda_n \end{pmatrix}, \tag{3-8}$$

where $b_i = \tilde{0} > 0$.

We start bring down the bulge along the last column. From equation 3-6, $\gamma = \tilde{1}$ and $\sigma = \tilde{0}$ (and negative). From 3-7, for $2 \leq i \leq n$, $G_{2,i+1} = \widetilde{I}$ (where I is the identity matrix) and the new bulges are given by

$$\delta_+^{(1)} = \gamma\sigma(\tilde{\lambda}_1 - \lambda_n) > 0 \quad \text{and} \quad \delta_-^{(1)} = \sigma b_1 < 0 \ .$$

Notice that $\tilde{b}_1 = \gamma b_1$ is still strictly positive.

$$S^{(1)} = \begin{pmatrix} 1 & \tilde{1} & 0 & 0 & \dots & 0 & 0 \\ \tilde{1} & \tilde{\lambda}_1 & \tilde{b}_1 & 0 & \dots & 0 & \delta_+^{(1)} \\ 0 & \tilde{b}_1 & \tilde{\lambda}_2 & b_2 & & 0 & \delta_-^{(1)} \\ 0 & 0 & b_2 & \tilde{\lambda}_3 & \ddots & \vdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & b_{n-2} & \vdots \\ 0 & 0 & 0 & \dots & b_{n-2} & \tilde{\lambda}_{n-1} & 0 \\ 0 & \delta_+^{(1)} & \delta_-^{(1)} & 0 & \dots & 0 & \tilde{\lambda}_n \end{pmatrix}$$

We show that subsequent Givens conjugations bring down the bulges, preserving their signs, until the bottom bulge disappears and the upper (positive) bulge becomes the entry b_{n-1} of J_n. We proceed by proving that intermediate conjugations — by $G_{k,n+1}$ — "pushes" the bulges $\delta_+^{(k-1)}$ and $\delta_-^{(k-1)}$ at entries $(k, n+1)$ and $(k+1, n+1)$ of the matrix $S^{(k-1)}$ to $\delta_+^{(k)}$ and $\delta_-^{(k)}$ at positions $(k+1, n+1)$ and $(k+2, n+1)$ of $S^{(k)}$, respectively.

Split $S^{(k-1)}$ (of dimension $n+1$) into nine blocks by partitioning rows and columns into three sets of sizes $k-1, 2, n-k$, as follows:

$$S^{(k-1)} = \left[\begin{array}{c|c|c} \tilde{J}_{k-2} & A & 0 \\ \hline A^T & B & C \\ \hline 0 & C^T & Z_{k+1,n} \end{array}\right],$$

where

$$\tilde{J}_{k-2} = \begin{pmatrix} 1 & \tilde{1} & & & \\ \tilde{1} & \tilde{\lambda}_1 & \tilde{b}_1 & & \\ & \tilde{b}_1 & \tilde{\lambda}_2 & \ddots & \\ & & \ddots & \ddots & \tilde{b}_{k-3} \\ & & & \tilde{b}_{k-3} & \tilde{\lambda}_{k-2} \end{pmatrix}, A = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \tilde{b}_{k-2} & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} \tilde{\lambda}_{k-1} & b_{k-1} \\ b_{k-1} & \tilde{\lambda}_k \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & \dots & 0 & \delta_+^{(k-1)} \\ b_k & 0 & \dots & 0 & \delta_-^{(k-1)} \end{pmatrix}$$

and

$$Z_{k+1,n} = \begin{pmatrix} \tilde{\lambda}_{k+1} & b_{k-1} & & & \\ b_{k-1} & \tilde{\lambda}_{k+2} & \ddots & & \\ & \ddots & \ddots & b_{n-2} & \\ & & b_{n-2} & \tilde{\lambda}_{n-1} & \\ & & & & \tilde{\lambda}_n \end{pmatrix}.$$

Split $G_{k+1,n+1}$ into blocks of the same size,

$$G_{k+1,n+1} = \left[\begin{array}{c|c|c} I & 0 & 0 \\ \hline 0 & D & E \\ \hline 0 & -E^T & \tilde{I} \end{array}\right]$$

where

$$D = \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}, \quad E = \begin{pmatrix} 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & \sigma \end{pmatrix}, \quad \tilde{I} = \begin{pmatrix} I & \\ & \gamma \end{pmatrix}.$$

Again, $\gamma = \tilde{1}$, $\sigma = \tilde{0} < 0$. Multiplying on the right (resp. left) by $G_{k+1,n+1}$ mixes columns (resp. rows) $k+1$ with $n+1$. We obtain

$$S^{(k)} = \left[\begin{array}{c|c|c} \tilde{J}_{k-2} & A & 0 \\ \hline A^T & \tilde{B} & X \\ \hline 0 & X^T & Y \end{array}\right],$$

where

$$\tilde{B} = \begin{pmatrix} \tilde{\lambda}_{k-1} & \tilde{b}_{k-1} \\ \tilde{b}_{k-1} & \tilde{\lambda}_k \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \tilde{b}_k & 0 & \dots & 0 & \delta_+^{(k)} \end{pmatrix}$$

and

$$Y = \begin{pmatrix} \tilde{\lambda}_{k+1} & b_{k-1} & & & \delta_-^{(k)} \\ b_{k-1} & \tilde{\lambda}_{k+2} & \ddots & & \\ & \ddots & \ddots & b_{n-2} & \\ & & b_{n-2} & \tilde{\lambda}_{n-1} & \\ \delta_-^{(k)} & & & & \tilde{\lambda}_n \end{pmatrix}.$$

Now, $\delta_-^{(k)} = \sigma b_k$ and $\delta_+^{(k)} = \gamma\sigma(\tilde{\lambda}_k - \tilde{\lambda}_n) + \mathcal{O}(\epsilon^2)$ have the desired signs, and $\tilde{b}_i \approx b_i > 0$ for $i = 1, \ldots, n-2$. After $n-2$ conjugations,

$$S^{(n-2)} = \begin{pmatrix} 1 & \tilde{1} & & & & \\ \tilde{1} & \tilde{\lambda}_1 & \tilde{b}_1 & & & \\ & \tilde{b}_1 & \tilde{\lambda}_2 & \ddots & & \\ & & \ddots & \ddots & \tilde{b}_{n-2} & \delta_+^{(n-2)} \\ & & & \tilde{b}_{n-2} & \tilde{\lambda}_{n-1} & \delta_-^{(n-2)} \\ & & & \delta_+^{(n-2)} & \delta_-^{(n-2)} & \tilde{\lambda}_n \end{pmatrix}.$$

A straightforward computation verifies that $b_{n-1} = \tilde{J}_{n+1,n} = S^{(n-1)}_{n+1,n} > 0$. □

Notice that the argument above uses the fact that eigenvalues are ordered increasingly. Different orderings however, would not have great impact on the result of the algorithm: the only change would be in the sign of off-diagonal entries, as the sing of $\delta_+^{(k)}$ in the proof above depends on such orderings. This is an appropriate moment to say something about *signed Jacobi matrices*.

3.1.1 Signed Jacobi matrices

A matrix J^s is a *signed Jacobi matrix* if there is a signed diagonal matrix $E \in \mathcal{E}$ such that EJ^sE is Jacobi. Said differently, dropping the signs of the off-diagonal entries of J^s obtains a Jacobi matrix. Similarly, one may think of an input vector c^s with arbitrary signs, but it makes no sense to think of it as a Moser vector: the first entries of eigenvectors of a Jacobi matrix are not well defined, and a Moser vector is the choice of positive such entries.

Still, one can rather artificially define a correspondence between vectors c^s with nonzero entries for which the last (or any other) coordinate $(c^s)_n > 0$, and signed Jacobi matrices, simply by requiring equality of the signs of $(c^s)_i$ and ${J^s}_{i+1,i}, i = 1, \ldots, n-1$.

The RKPW algorithm handles this extension in a simple fashion. Define a *signed Givens rotation* to be a matrix of the form

$$\begin{pmatrix} \gamma & -\sigma \\ -\sigma & -\gamma \end{pmatrix}.$$

Appropriate choices between standard and signed Givens rotations in the algorithm of the previous section gives rise to signed Jacobi matrices with arbitrary choices of sign of off-diagonal entries. Details are left to the reader.

3.1.2 Reduced matrices

We consider limits of Moser vectors c with non-negative entries.

Proposition 5. *Let $\Lambda = \mathrm{diag}(\lambda_1 < \ldots < \lambda_n)$ and $c \neq 0$ with non-negative entries. Let $I = \{i \mid c_i = 0\}$, with $k = |I|$. Then RKPW yields a symmetric tridiagonal matrix*

$$J_c = \begin{pmatrix} \hat{J}_{n-k} & 0 & & \\ 0 & \lambda_{i_1} & 0 & \\ & 0 & \ddots & 0 \\ & & 0 & \lambda_{i_k} \end{pmatrix},$$

where $\hat{J}_{n-k}$ is the Jacobi matrix associated with $\hat{\lambda} = \lambda \setminus \{\lambda_i, i \in I\}$ and $\hat{c} = (c_j)_{j=1}^{n-k}$, with $j \notin I$. Said differently, the eigenvalues $\lambda_i, i \in I$, are pushed to the bottom of the resulting (reduced) matrix.

Proof. Assume first that only entry c_k is zero. At the k-th step of the algorithm (i.e., at the transition $(\lambda^{k-1}, c^{k-1}) \mapsto (\lambda_k, c^k)$), the auxiliary matrix S is already tridiagonal, and therefore all the Givens rotations leading from J_{k-1} to J_k are equal to the identity, so that

$$J_k = \begin{pmatrix} J_{k-1} & 0^T \\ 0 & \lambda_k \end{pmatrix} \tag{3-9}$$

For the next step, in which $c_{k+1} \neq 0$, rotations are nontrivial and the last one, $G_{k+1,k+2}$, is a signed permutation: its $(k+1, k+2)$ block is

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

From a straightforward calculation,

$$\tilde{G}^T \begin{pmatrix} 1 & \|c^k\| e_1^T & c_{k+1} \\ \|c^k\| e_1 & J_k & 0 \\ c_{k+1} & 0^T & \lambda_{k+1} \end{pmatrix} \tilde{G} = \begin{pmatrix} 1 & \|c^{k+1}\| e_1^T & 0 \\ \|c^{k+1}\| e_1 & \hat{J}_k & 0 \\ 0 & 0^T & \lambda_k \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \|c^{k+1}\| e_1^T \\ \|c^{k+1}\| e_1 & J_{k+1} \end{pmatrix}.$$

In the subsequent steps, RKPW pushes λ_k to the bottom of the matrix. More generally, if there are other zero coordinates $c_r = 0$, the same argument, in which some Givens rotations are taken to be equal to the identity, leads to a matrix with λ_k and λ_r at its bottom, as we wanted to show. □

From the proposition above, only faces of $\mathcal{P}_\Lambda$ corresponding to isolated eigenvalues on the bottom of the reduced matrix are obtained by RKPW. Thus, no face labeled 0+ in Figure 2.1 is in the image of the inverse algorithm. Also, not every matrix close to a diagonal matrix is identifiable with some signed vector c^s. Again, the lack of continuity of the map $(\lambda, c) \mapsto T$ is clear.

3.1.3 Equiasymptotic sequences

Following [6, 15], we answer the following question: which sequences of $n \times n$ Jacobi matrices T_k converge to a given reduced matrix? The issue will be relevant when we consider stability of both RKPW and IVBI.

A sequence of vectors $c^k = (c_1^k, \ldots, c_n^k) \in \mathbb{R}^n$ admits an *equiasymptotic partition* if and only if there is an ordered partition of

$$\{1, 2, \ldots, n\} \;=\; \cup_i \, I_i$$

such that for $i_\alpha \in I_\alpha$ and $i_\beta \in I_\beta$, the quotient $c_{i_\beta}^k / c_{i_\alpha}^k$ goes to zero when $k \to \infty$ if $\alpha < \beta$ or to a nonzero real number if $\alpha = \beta$. Thus, two indices α and β in the same subset I_i label entries of c^k that have comparable asymptotic behavior. Also, the entries indexed by I_{i+1} decrease to zero faster than the ones indexed by I_i. As an example, for $c \in \mathbb{R}^5$, we represent the partition

$$\{1, 2, 3, 4, 5\} \;= \{1, 3, 4\} \cup \{2\} \cup \{5\}$$

by the sequence $[2, 1, 2, 2, 0]$, in order to suggest that entries c_1, c_3 and c_4 are asymptotically larger than entry c_2 which in turn is larger than c_5.

Clearly not every sequence c^k admits such a split, but it is easy to prove that there is always a subsequence c^{k_ℓ} admitting an equiasymptotic partition.

For $n = 3$, there are 13 equiasymptotic classes:

$$[0, 0, 0], [1, 0, 0], [0, 1, 0], [0, 0, 1], [1, 1, 0], [1, 0, 1], [0, 1, 1],$$

$$[0, 1, 2], [0, 2, 1], [1, 0, 2], [1, 2, 0], [2, 0, 1], [2, 1, 0] \; .$$

In Figure 3.1 below, we associate classes and cells for $n = 3$. It may be interpreted as a refinement of the information in Figure 2.2.

There is an analogous result – an identification of equiasymptotic partitions with cells of the permutohedron – for arbitrary dimensions.

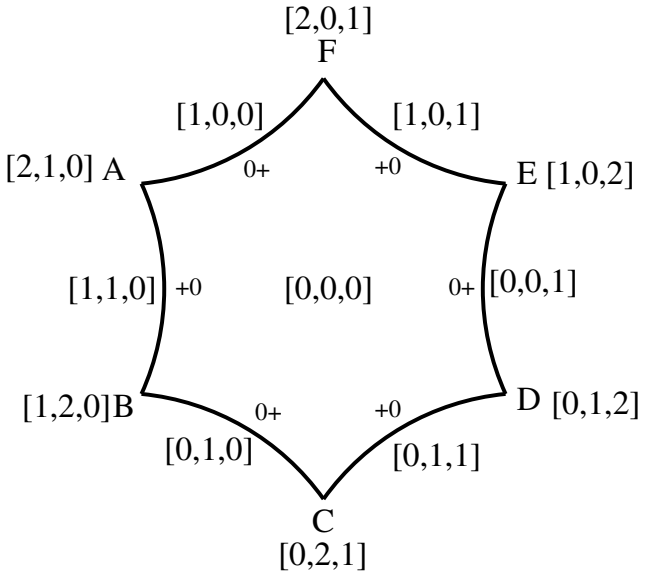

Figure 3.1: Equiasymptotic partitions and their limits.

3.1.4 Multiple spectrum

By Proposition 1, multiple eigenvalues imply that some entries of c must be zero. In particular, a point (λ, c) with non-simple λ and strictly positive Moser vector c does not belong to the image of the extension of the Moser map from $\overline{\mathcal{J}}$ to $\overline{\mathcal{M}}$.

Still, the algorithm generates an output. For $\Lambda = \mathrm{diag}(2, 2, 4, 5, 5)$ and $c = (1, 1, 1, 1, 1)$, RKPW yields

$$J = \begin{pmatrix} \frac{18}{5} & \frac{\sqrt{46}}{5} & & & \\ \frac{\sqrt{46}}{5} & \frac{381}{115} & \frac{6\sqrt{5}}{23} & & \\ & \frac{6\sqrt{5}}{23} & \frac{94}{23} & & \\ & & & 2 & \\ & & & & 5 \end{pmatrix}.$$

As seen in the example, again for a non-simple spectrum, the algorithm pushes the repeated eigenvalues to the bottom of the matrix, as in Proposition 5. The proof follows a similar argument: at some point, non-simplicity implies that $\delta_-^{(k)} = 0$. Which in turn, forces $G_{k+1,n}$ to be a signed permutation, and so on. One should be careful about meaningless outputs.

Moser vectors are frequently taken as coordinates in the study of spectral algorithms on Jacobi matrices and evolutions of physical interest, as the Toda flow [16]. It is desirable then to look for other coordinates systems with additional properties: they might include reduced matrices and be local diffeomorphisms, in particular around diagonal matrices. In the next chapter,

we introduce the *isospectral manifold* and propose a new inverse algorithm, corresponding to handling charts.

4
Reconstruction from bidiagonal coordinates — INVBI

The set $\mathcal{T}_\Lambda$ of all tridiagonal symmetric matrices with a fixed simple spectrum is actually a manifold, obtained by appropriately gluing 2^{n-1} copies of the permutohedron $\mathcal{P}_\Lambda$, each corresponding to a choice of signs on the off-diagonal entries [4]. For $n = 3$, four hexagons after identifications yield a connected sum of two tori – a bitorus, endowed with the CW-decomposition given in the figure below. Let $A = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$, $B = \mathrm{diag}(\lambda_2, \lambda_1, \lambda_3)$ and so on, so that each of the six diagonal matrices A, B, C, D, E and F correspond to its proper permutation of eigenvalues, as in Figure 2.1.

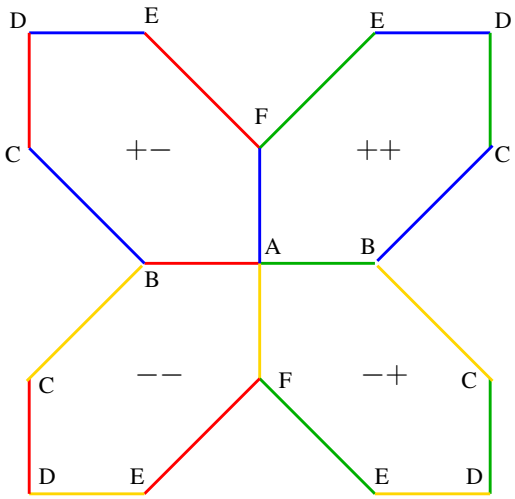

Figure 4.1: Edges with same color and same corner eigenvalue are identified.

In Section 4.1, we provide a smooth atlas for $\mathcal{T}_\Lambda$, proving that it is indeed a manifold. It turns out that the isospectral tridiagonal manifold is always compact, orientable and its universal covering is $\mathbb{R}^{n+1}$ [4].

The new inverse algorithm in Section 4.2 is a constructive inversion of these charts.

4.1
Bidiagonal charts

Again, let $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where eigenvalues are in strictly increasing order. Recall the usual QR decomposition of an invertible matrix

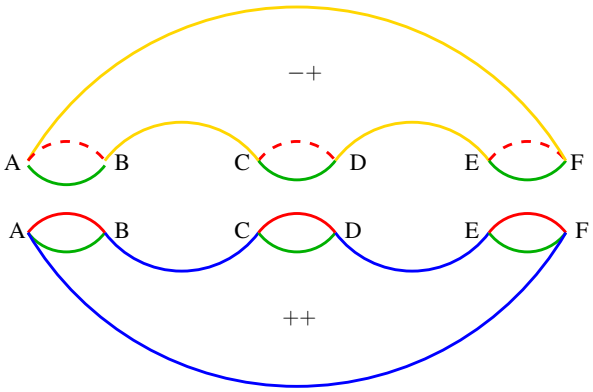

Figure 4.2: Glue the edges of the complex properly to obtain the bitorus.

$M = QR$, for which we define

$$[M]_Q = Q \in O(n) \quad , \quad [M]_R = R \in Up^+(n) \ .$$

We say a matrix M is *LU-positive* if it admits a (necessarily unique) LU decomposition $M = LU$, where

$$[M]_L = L \in Lo^1(n) \quad , \quad [M]_U = U \in Up^+(n) \ .$$

Equivalently, an invertible matrix M is LU-positive if and only if its principal upper minors are positive.

Let S_n be the group of permutations of n elements. For $\pi \in S_n$, set

$$\Lambda^\pi \ = \ \mathrm{diag}(\lambda_{\pi(1)}, \ldots, \lambda_{\pi(n)}) \ .$$

Notice that $\Lambda^\pi = P\Lambda P^T$ for a (unique) permutation matrix P. For $T \in \mathcal{T}_\Lambda$, a diagonalization $T = Q^T \Lambda Q$ yields another,

$$T \ = \ Q^T P^T P \Lambda P^T P Q \ = \ Q_\pi^T \ \Lambda^\pi \ Q_\pi \ ,$$

where $Q_\pi \in O(n)$ is not necessarily in $SO(n)$.

Now, fix a permutation $\pi \in S_n$ and define the *chart domain*

$$\mathcal{U}_\Lambda^\pi = \{T \in \mathcal{T}_\Lambda \mid T = Q^T \Lambda^\pi Q, \text{ where } Q \text{ is LU-positive}\} \ .$$

The following properties are easy to verify [3].

1. If $T = Q^T \Lambda^\pi Q$ is such that Q has nonzero principal minors, then $T \in \mathcal{U}_\Lambda^\pi$.
2. The sets $\mathcal{U}_\Lambda^\pi \subset \mathcal{T}_\Lambda$ form an open cover of $\mathcal{T}_\Lambda$.

3. If $E \in \mathcal{E}$ (i.e., E is a diagonal of signs) and $T \in \mathcal{U}_\Lambda^\pi$, then $ETE \in \mathcal{U}_\Lambda^\pi$.

4. If $T \in \mathcal{T}_\Lambda$ is unreduced, then $T \in \mathcal{U}_\Lambda^\pi$ for all $\pi \in S_n$. In particular, each $\mathcal{U}_\Lambda^\pi$ is dense in $\mathcal{T}_\Lambda$.

As an example, the interior of the polygon in Figure 4.1 is the chart $\mathcal{U}_\Lambda^\pi$ for the identity permutation.

We now introduce the charts. For $T \in \mathcal{U}_\Lambda^\pi$, set

$$T \;=\; Q_\pi^T\ \Lambda^\pi\ Q_\pi.$$

Since $Q_\pi = L_\pi\ U_\pi,\ L_\pi \in Lo^1, U_\pi \in Up^+$, we have

$$T \;=\; U_\pi^{-1} L_\pi^{-1}\ \Lambda^\pi\ L_\pi U_\pi\ .$$

Set

$$B_\pi = U_\pi T U_\pi^{-1} = L_\pi^{-1}\Lambda^\pi L_\pi. \tag{4-1}$$

Notice that B_π is *bidiagonal*, since $U_\pi T U_\pi^{-1}$ is upper Hessenberg[1] while $L_\pi^{-1}\Lambda^\pi L_\pi$ is lower triangular,

$$B_\pi = \begin{pmatrix} \lambda_{\pi(1)} & & & & \\ \beta_{\pi(1)} & \lambda_{\pi(2)} & & & \\ & \beta_{\pi(2)} & \lambda_{\pi(3)} & & \\ & & \ddots & \ddots & \\ & & & \beta_{\pi(n-1)} & \lambda_{\pi(n)} \end{pmatrix} \tag{4-2}$$

Define charts — the *bidiagonal coordinates*,

$$\phi_\pi : \mathcal{U}_\Lambda^\pi \to \mathbb{R}^{n-1}\ , \quad T \mapsto B_\pi$$

We frequently abuse notation and call bidiagonal coordinates the vector $\beta_\pi = (\beta_{\pi(1)}, \ldots, \beta_{\pi(n-1)})$. Notice the conceptual similarity between β_π and the Moser vector c (which is less sensible to reordering of eigenvalues: different orderings can change the sign of the bottom off-diagonal entry of the resulting matrix from RKPW, as shown in Proposition 4). As in Theorem 1 for Moser vectors, the inverse map is obtained from the QR-decomposition of L_π:

$$\psi_\pi : \mathbb{R}^{n-1} \to \mathcal{U}_\Lambda^\pi$$

$$(\beta_{\pi(1)}, \ldots, \beta_{\pi(n-1)}) \mapsto [L_\pi]_Q^T\ \Lambda^\pi\ [L_\pi]_Q$$

[1]A matrix M is upper Hessenberg if all entries $(i,j), i \geq j+2$, are zero.

Theorem 3. *The map $\psi_\pi : \mathbb{R}^{n-1} \to \mathcal{U}_\Lambda^\pi$ is a diffeomorphism, with inverse $\phi_\pi : \mathcal{U}_\Lambda^\pi \to \mathbb{R}^{n-1}$.*

Clearly, the maps ψ_π and $\phi_\pi = \psi_\pi^{-1}$ are smooth. Dropping the reference to the permutation π and setting differences of eigenvalues as $\delta_{ij} = \lambda_j - \lambda_i$, there is an explicit diagonalization of $B_k = L_k^{-1}\Lambda_k L_k$, obtained in [3],

$$L_k = \begin{pmatrix} 1 & & & & \\ \frac{\beta_1}{\delta_{1,2}} & 1 & & & \\ \frac{\beta_1\beta_2}{\delta_{1,3}\delta_{2,3}} & \frac{\beta_2}{\delta_{2,3}} & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \frac{\beta_1\beta_2\ldots\beta_{k-1}}{\delta_{1,k}\delta_{2,k}\ldots\delta_{k-1,k}} & \frac{\beta_2\ldots\beta_{k-1}}{\delta_{2,k}\ldots\delta_{k-1,k}} & \cdots & \frac{\beta_{k-1}}{\delta_{k-1,k}} & 1 \end{pmatrix}, \tag{4-3}$$

$$L_k^{-1} = \begin{pmatrix} 1 & & & & \\ \frac{\beta_1}{\delta_{2,1}} & 1 & & & \\ \frac{\beta_1\beta_2}{\delta_{2,1}\delta_{3,1}} & \frac{\beta_2}{\delta_{3,2}} & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \frac{\beta_1\beta_2\ldots\beta_{k-1}}{\delta_{2,1}\delta_{3,1}\ldots\delta_{k,1}} & \frac{\beta_2\ldots\beta_{k-1}}{\delta_{3,2}\ldots\delta_{k,2}} & \cdots & \frac{\beta_{k-1}}{\delta_{k,k-1}} & 1 \end{pmatrix}, \tag{4-4}$$

yielding a concrete description of the map ψ_π. However, the computational cost is overwhelming, and an alternative is suggested in the next section.

As in [17], we can see that given any $T = [L]_Q^T \Lambda [L]_Q \in \mathcal{U}_\Lambda^\pi$, its Moser vectors and bidiagonal coordinates are related by

$$c_1 = \frac{1}{\sqrt{1 + \frac{\beta_1^2}{\delta_{12}^2} + \ldots + \frac{\beta_1^2\ldots\beta_{k-1}^2}{\delta_{1k}^2\ldots\delta_{k-1k}^2}}} \ , \ c_i = c_1 \frac{\beta_1 \ldots \beta_{i-1}}{\delta_{1i}\ldots\delta_{i-1,i}} \ , 2 \le i \le k \tag{4-5}$$

$$\beta_1 = \frac{\delta_{12}c_2}{c_1} \ , \ \beta_i = \frac{\delta_{1,i+1}\ldots\delta_{i,i+1}c_{i+1}}{\delta_{1i}\ldots\delta_{i-1,i}c_i} \ , 2 \le i \le k-1 \ , \tag{4-6}$$

for the Moser vector associated with T is the first column of $[L]_Q$, which is just the first column of L (given in 4-3) divided by its norm.

Figure 4.3 below indicates points of latent instability. On the left, the closure of Jacobi matrices with a fixed spectrum is represented by the hexagon $ABCDEF$. On the right, the possible values of the bidiagonal coordinates $\beta = (\beta_1, \beta_2)$ are given in a square, where two sides correspond to choosing some coordinate equal to $+\infty$. As matrices approach edges CD and DE, not in the domain of ϕ_π, their images approach a single asymptotic behavior $(+\infty, +\infty)$.

Sequences of matrices T_k admitting equiasymptotic partitions were described in terms of the Moser vector c in Section 3.1.3. A simple translation to bidiagonal coordinates, using the formulas converting c into β, yields Figure 4.4, a counterpart to Figure 3.1.

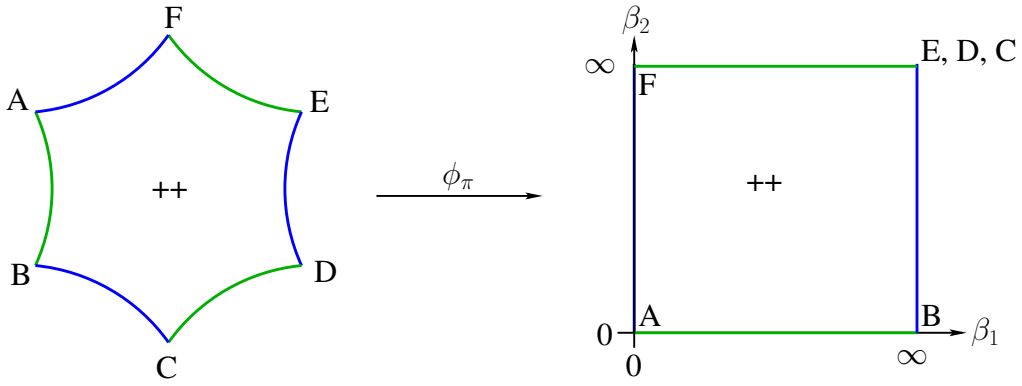

Figure 4.3: The β coordinates of $\mathcal{J}_\Lambda$.

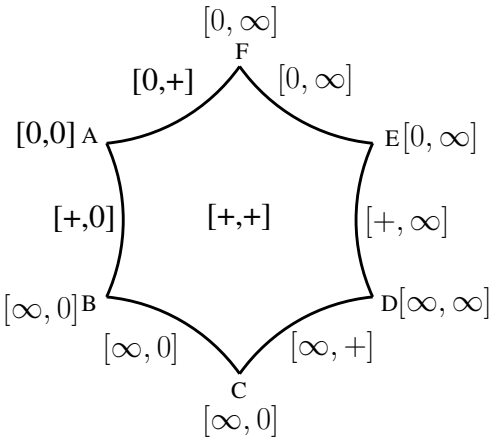

Figure 4.4: Labels $[\beta_1 \sim c_2/c_1, \beta_2 \sim c_3/c_2]$.

4.2 INVBI – a counterpart to RKPW for bidiagonal coordinates

We search for an inverse map $(\lambda_\pi, \beta_\pi) \in \mathcal{B}_\pi \mapsto T \in \mathcal{U}_\Lambda^\pi$, where

$$\mathcal{B}_\pi = \{(\lambda_\pi, \beta_\pi) = (\lambda_{\pi(1)}, \ldots, \lambda_{\pi(n)}, \beta_{\pi(1)}, \ldots, \beta_{\pi(n-1)}) \mid \lambda_{\pi(i)} \neq \lambda_{\pi(j)},\ \beta_{\pi(i)} \in \mathbb{R}\},$$

in the spirit of RKPW. We omit the reference to π: the orthogonal matrices Q below are always LU-positive.

As RKPW, the algorithm is inductive: the matrix T_k associated with $B_k = (\lambda^k, \beta^{k-1})$, is obtained from T_{k-1}, associated with $B_{k-1} = (\lambda^{k-1}, \beta^{k-2})$. The relations below hold for any $T_k \in \mathcal{U}_{\Lambda_k}$ with coordinates (λ^k, β^{k-1}):

$$T_k = Q_k^T \Lambda_k Q_k \ ,\ Q_k \in O(k), \tag{4-7}$$

$$Q_k = L_k U_k \ ,\quad L_k \in Lo^1(k) \ ,\ U_k \in Up^+(k), \tag{4-8}$$

$$B_k = L_k^{-1} \Lambda_k L_k = U_k T_k U_k^{-1}. \tag{4-9}$$

We prepare for the algorithm. Clearly, L_k is obtained from L_{k-1} by

adjoining a last row and column,

$$L_k \;=\; \begin{pmatrix} L_{k-1} & 0 \\ -\ell^T L_{k-1} & 1 \end{pmatrix} \;,$$

where the vector $\ell \in \mathbb{R}^{k-1}$ is defined in

$$L_k^{-1} \;=\; \begin{pmatrix} L_{k-1}^{-1} & 0 \\ \ell^T & 1 \end{pmatrix} \;.$$

Set

$$\hat{Q} = \begin{pmatrix} Q_{k-1} & 0 \\ 0 & 1 \end{pmatrix}, \quad \hat{U} = \begin{pmatrix} U_{k-1} & 0 \\ 0 & 1 \end{pmatrix}, \quad \hat{L} = \begin{pmatrix} L_{k-1} & 0 \\ 0 & 1 \end{pmatrix},$$

so that $L_k^{-1}\hat{L}$ is a rank one perturbation of the identity. We retrieve T_k from the auxiliary matrix

$$S = \hat{U}^{-1}\; B_k\; \hat{U} \tag{4-10}$$

$$= \begin{pmatrix} U_{k-1}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B_{k-1} & 0 \\ \beta_{k-1} e_{k-1}^T & \lambda_k \end{pmatrix} \begin{pmatrix} U_{k-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} T_{k-1} & 0 \\ \beta_{k-1} e_{k-1}^T U_{k-1} & \lambda_k \end{pmatrix} \;.$$

Combining with equation 4-9,

$$T_k \;=\; U_k^{-1}\; \hat{U}\; S\; \hat{U}^{-1} U_k \;=\; \tilde{U}^{-1} S\; \tilde{U} \tag{4-11}$$

where $\tilde{U} \;=\; \hat{U}^{-1} U_k$. We describe a procedure to obtain $\tilde{U}$ and its inverse.

Set

$$M = \hat{U}^{-1}(L_k^{-1}\hat{L})\hat{U}$$

$$= \begin{pmatrix} U_{k-1}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} L_{k-1}^{-1} & 0 \\ \ell^T & 1 \end{pmatrix} \begin{pmatrix} L_{k-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U_{k-1} & 0 \\ 0 & 1 \end{pmatrix} \;=\; \begin{pmatrix} I & 0 \\ \ell^T L_{k-1} U_{k-1} & 1 \end{pmatrix} \;,$$

also a rank one perturbation of the identity. Recall that a positive definite matrix P has a unique Cholesky decomposition $P = LL^T$, where $L \in Lo^+$.

Proposition 6. *The matrix $\tilde{U}^{-T}$ is the lower triangular factor in the Cholesky decomposition of $(MM^T)^{-1}$.*

Clearly $(MM^T)^{-1}$ is positive definite.

Proof. As $Q_k = L_k U_k$ and $\hat{Q} = \hat{L}\hat{U}$, we have $L_k U_k Q_k^T \;=\; \hat{L}\hat{U}\hat{Q}^T (= I)$. A simple algebra obtains

$$\hat{U}^{-1} U_k Q_k^T \;=\; \hat{U}^{-1} L_k^{-1} \hat{L} \hat{U} \hat{Q}^T \;,$$

and, as $\tilde{U} = \hat{U}^{-1}U_k$,

$$M \;=\; \tilde{U}Q_k^T\hat{Q}\;. \tag{4-12}$$

so that $(MM^T)^{-1} = \tilde{U}^{-T}\tilde{U}^{-1}$. Finally, as $\tilde{U} \;=\; \hat{U}^{-1}U_k$, and both $\hat{U}^{-1}$ and U_k have positive diagonal entries, so does $\tilde{U}^{-T}$. □

For $v^T = \ell^T L_{k-1}U_{k-1} = \ell^T Q_{k-1}$, the last row of M is $(v^T, 1) \in \mathbb{R}^k$. The required Cholesky decomposition only depends on v:

$$MM^T = \begin{pmatrix} I & v \\ v^T & \|v\|^2 + 1 \end{pmatrix} = \tilde{U}\tilde{U}^T\;, \tag{4-13}$$

$$(MM^T)^{-1} = \begin{pmatrix} I + vv^T & -v \\ -v^T & 1 \end{pmatrix} = \tilde{U}^{-T}\tilde{U}^{-1}\;. \tag{4-14}$$

The next proposition describes $\tilde{U}^{-1}$ in terms of v. Again, v^i and v_i denote the vector of first i entries and i-th entry of the vector $v^T = (v_1, \ldots, v_{k-1})$, respectively. For consistency, set $v_0 = v^0 = 0$. Let $n_i = \sqrt{\|v^i\|^2 + 1}$. Observe that $n_0 = 1$ and $n_{k-1} = \|v^{k-1}\|^2 + 1 = \|v\|^2 + 1$.

Proposition 7. *The lower triangular matrix $W = \tilde{U}^{-T}$ in the Cholesky decomposition $(MM^T)^{-1} = WW^T$ is*

$$W = \begin{pmatrix} n_1 & & & & & & \\ \frac{v_1v_2}{n_1} & \frac{n_2}{n_1} & & & & & \\ \frac{v_1v_3}{n_1} & \frac{v_2v_3}{n_1n_2} & \frac{n_3}{n_2} & & & & \\ \vdots & \vdots & \ddots & \ddots & & & \\ \frac{v_1v_{k-1}}{n_1} & \frac{v_2v_{k-1}}{n_1n_2} & \frac{v_3v_{k-1}}{n_2n_3} & \cdots & \frac{v_{k-2}v_{k-1}}{n_{k-2}n_{k-3}} & \frac{n_{k-1}}{n_{k-2}} & \\ \frac{-v_1}{n_1} & \frac{-v_2}{n_1n_2} & \frac{-v_3}{n_2n_3} & \cdots & \frac{-v_{k-2}}{n_{k-2}n_{k-3}} & \frac{-v_{k-1}}{n_{k-1}n_{k-2}} & \frac{1}{n_{k-1}} \end{pmatrix}\;.$$

More explicitly,

$$W_{ij} = (\tilde{U}^{-T})_{ij} = \begin{cases} \frac{n_i}{n_{i-1}}\;, i = j \neq k \\ \frac{v_jv_i}{n_in_{i-1}}\;, j < i < k \\ \frac{-v_i}{n_in_{i-1}}\;, j < i = k \\ \frac{1}{n_{k-1}}\;, i = j = k \end{cases}\;.$$

Proof. From equation 4-14,

$$(MM^T)^{-1} \;=\; \begin{pmatrix} 1 + v_1^2 & v_1v_2 & \ldots & v_1v_{k-1} & -v_1 \\ v_1v_2 & 1 + v_2^2 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & v_{k-2}v_{k-1} & -v_{k-2} \\ v_1v_{k-1} & \ldots & v_{k-2}v_{k-1} & 1 + v_{k-1} & -v_{k-1} \\ -v_1 & \ldots & -v_{k-2} & -v_{k-1} & 1 \end{pmatrix}\;.$$

Let $w^{(i)}$ be the i-th row of W. Thus $W = \tilde{U}^{-T}$ is equivalent to

(1) $\langle w^{(i)}, w^{(j)} \rangle = v_i v_j, \ \forall i \neq j \in \{1, \ldots, k-1\}$
(2) $\langle w^{(i)}, w^{(i)} \rangle = 1 + (v_i)^2, \ \forall i \in \{1, \ldots, k-1\}$
(3) $\langle w^{(i)}, w^{(k)} \rangle = -v_i, \ \forall i \in \{1, \ldots, k-1\}$
(4) $\langle w^{(k)}, w^{(k)} \rangle = 1.$

Suppose $i < j$. In the inner product $\langle w^{(i)}, w^{(j)} \rangle$, only the i first entries are possibly nonzero and we may collect $v_i v_j$:

$$\langle w^{(i)}, w^{(j)} \rangle = v_i v_j \left(\frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} + \frac{v_3^2}{n_2^2 n_3^2} + \ldots + \frac{v_{i-1}^2}{n_{i-1}^2 n_{i-2}^2} + \frac{1}{n_{i-1}^2} \right) \tag{4-15}$$

We prove by induction on i that K, the sum of the terms between parentheses above, equals one. Let F be the sum of the first $i-1$ terms of K. Start with $i = 3$: $\langle w^{(3)}, w^{(j)} \rangle = v_3 v_j$. Indeed, the sum has only the same three terms for all $j > 3$,

$$K_3 = \frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} + \frac{1}{n_2^2} = F_3 + \frac{1}{n_2^2} = \left(\frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} \right) + \frac{n_1^2}{n_1^2 n_2^2}$$

$$= \left(\frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} \right) + \frac{v_1^2 + 1}{n_1^2 n_2^2}$$

$$= \frac{v_1^2 n_2^2 + (v_2^2 + v_1^2 + 1)}{n_1^2 n_2^2} = \frac{v_1^2 n_2^2 + n_2^2}{n_1^2 n_2^2} = 1 \ .$$

We mimic the steps above. As $n_i^2 = v_1^2 + v_2^2 + \ldots + v_i^2 + 1$, we have

$$n_i^2 + v_{i+1}^2 = n_{i+1}^2 \ . \tag{4-16}$$

Now, suppose by induction that

$$K_{i-1} = \frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} + \frac{v_3^2}{n_2^2 n_3^2} + \ldots + \frac{v_{i-2}^2}{n_{i-2}^2 n_{i-3}^2} + \frac{1}{n_{i-2}^2} = F_{i-1} + \frac{1}{n_{i-2}^2} = 1 \ .$$

and then we must prove that $K_i = 1$, where

$$K_i = \frac{v_1^2}{n_1^2} + \frac{v_2^2}{n_1^2 n_2^2} + \frac{v_3^2}{n_2^2 n_3^2} + \ldots + \frac{v_{i-1}^2}{n_{i-1}^2 n_{i-2}^2} + \frac{1}{n_{i-1}^2} = F_{i-1} + \frac{v_{i-1}^2}{n_{i-1}^2 n_{i-2}^2} + \frac{1}{n_{i-1}^2} \ .$$

From equation 4-16 and the inductive hypothesis,

$$K_i = F_{i-1} + \frac{v_{i-1}^2 + n_{i-2}^2}{n_{i-1}^2 n_{i-2}^2} = F_{i-1} + \frac{1}{n_{i-2}^2} = 1 \ ,$$

which implies (1). For (3), set $v_j = -1$, and (4) is straightforward. For (2),

$$\langle w^{(i)}, w^{(i)} \rangle = \frac{v_1^2 v_i^2}{n_1^2} + \ldots + \frac{n_i^2}{n_{i-1}^2}$$

$$= v_i^2 \left(\frac{v^1}{n_1^2} + \ldots + \frac{v_{i-1}^2}{n_{i-2}^2 n_{i-1}^2} \right) + \frac{v_1^2 + \ldots + v_{i-1}^2 + v_i^2 + 1}{n_{i-1}^2}$$

$$= v_i^2 \left(\frac{v^1}{n_1^2} + \ldots + \frac{v_{i-1}^2}{n_{i-2}^2 n_{i-1}^2} + \frac{1}{n_{i-1}^2} \right) + \frac{v_1^2 + \ldots + v_{i-1}^2 + 1}{n_{i-1}^2}$$

$$= \; v_i^2 K_i + \frac{n_{i-1}^2}{n_{i-1}^2} \; = \; v_i^2 + 1$$

□

Actually, we need much less then the full matrices $\tilde{U}$ and $\tilde{U}^{-1}$. According to equation 4-11, since T_k is symmetric and tridiagonal, we only need diagonal and super-diagonal entries of the $k \times k$ matrices $\tilde{U}$ and $\tilde{U}^{-1}$, with indices (ii) and $(i, i+1)$. Such entries of $\tilde{U}$ are easily expressed in terms of the analogous entries of $\tilde{U}^{-1}$, as shown in the example below for $k = 3$:

$$\tilde{U}^{-1} = \begin{pmatrix} a & b & * \\ & d & e \\ & & f \end{pmatrix} \;\Rightarrow\; \tilde{U} = \begin{pmatrix} 1/a & -b/ad & * \\ & 1/d & -e/df \\ & & 1/f \end{pmatrix} \tag{4-17}$$

Thus, the algorithm only needs entries

$$(\tilde{U}^{-1})_{ii} = \begin{cases} \frac{n_i}{n_{i-1}} = \sqrt{\frac{\|v^i\|^2+1}{\|v^{i-1}\|^2+1}} \;, & i = 1, \ldots, k-1 \\ \frac{1}{n_{k-1}} = \frac{1}{\sqrt{\|v\|^2+1}} \;, & i = k \end{cases} \tag{4-18}$$

$$(\tilde{U}^{-1})_{i,i+1} = \begin{cases} \frac{v_i v_{i+1}}{n_i n_{i-1}} = \frac{v_i v_{i+1}}{\sqrt{(\|v^i\|^2+1)(\|v^{i-1}\|^2+1)}} \;, & i \;= 1, \ldots, k-2 \\ \frac{-v_{k-1}}{n_{k-1} n_{k-2}} = \frac{-v_{k-1}}{\sqrt{(\|v\|^2+1)(\|v^{k-2}\|^2+1)}} \;, & i = k-1 \;. \end{cases} \tag{4-19}$$

We now obtain v.

Proposition 8. *The vector v satisfies the linear system*

$$(T_{k-1} - \lambda_k I)v = u_{k-1}\beta_{k-1}e_{k-1} \;, \tag{4-20}$$

where

$$u_{k-1} = \|e_{k-1}^T L_{k-1}^{-1}\| = \sqrt{\frac{\beta_1^2 \ldots \beta_{k-2}^2}{\delta_{12}^2 \ldots \delta_{1,k-1}^2} + \ldots + \frac{\beta_{k-2}^2}{\delta_{k-2,k-1}^2} + 1} \;. \tag{4-21}$$

Proof. From equation 4-9, we have $L_k^{-1}\Lambda_k = B_k L_k^{-1}$. Equating bottom rows,

$$(\ell^T, 1)\Lambda_k = (\beta_{k-1} e_{k-1}^T L_{k-1}^{-1}, 0) + (\lambda_k \; \ell^T, \lambda_k) \;,$$

which implies that

$$\ell^T = \beta_{k-1} e_{k-1}^T L_{k-1}^{-1} (\Lambda_{k-1} - \lambda_k)^{-1} \tag{4-22}$$

so that, as $v^T = \ell^T L_{k-1} U_{k-1} = \ell^T Q_{k-1}$,

$$v^T = \beta_{k-1} e_{k-1}^T U_{k-1} Q_{k-1}^T (\Lambda_{k-1} - \lambda_k)^{-1} Q_{k-1}$$

$$= u_{k-1} \beta_{k-1} e_{k-1}^T (T_{k-1} - \lambda_k)^{-1} \ ,$$

where $u_{k-1} = (U_{k-1})_{k-1,k-1}$ is the bottom entry of U_{k-1}. From equation 4-8, $U_{k-1} U_{k-1}^T = L_{k-1}^{-1} L_{k-1}^{-T}$ and, equating the bottom right entries,

$$u_{k-1} = \| e_{k-1}^T L_{k-1}^{-1} \|.$$

The expression for u_{k-1} in terms of bidiagonal coordinates follows from the formula for L_{k-1}^{-1}, equation 4-4. □

Applying the above propositions to equation 4-11, we are led to the outcome of INVBI, the inverse algorithm associated with bidiagonal coordinates.

Theorem 4. *The nontrivial entries of T_k are*

$$(T_k)_{ii} \ = \ a_i \ = \ \begin{cases} \tilde{U}_{11} \begin{pmatrix} (\tilde{U}^{-1})_{11} & (\tilde{U}^{-1})_{12} \end{pmatrix} \begin{pmatrix} S_{11} \\ S_{12} \end{pmatrix}, \textit{ for } i = 1 \\ \begin{pmatrix} (\tilde{U}^{-1})_{ii} & (\tilde{U}^{-1})_{i,i+1} \end{pmatrix} \begin{pmatrix} S_{i-1,i} \tilde{U}_{i-1,i} + S_{ii} \tilde{U}_{ii} \\ S_{i+1,i} \tilde{U}_{ii} \end{pmatrix}, \ \forall i \in \{2, \ldots, k-1\} \\ (\tilde{U}^{-1})_{kk} \begin{pmatrix} S_{k,k-1} & S_{kk} \end{pmatrix} \begin{pmatrix} \tilde{U}_{k-1,k} \\ \tilde{U}_{kk} \end{pmatrix}, \textit{ for } i = k. \end{cases} \tag{4-23}$$

$$(T_k)_{i,i+1} = (T_k)_{i+1,i} = b_i = (\tilde{U}^{-1})_{i+1,i+1} S_{i+1,i} \tilde{U}_{ii}, \ \forall i \in \{1, \ldots, k-1\}. \tag{4-24}$$

In summary, the inductive step of INVBI proceeds as follows.

(1) Compute u_{k-1} as in formula 4-21.
(2) Build the auxiliary matrix S.
(3) Solve the system $(T_{k-1} - \lambda_k) v = u_{k-1} \beta_{k-1} e_{k-1}$.
(4) Obtain relevant entries of $\tilde{U}^{-1}$ and $\tilde{U}$ from equations 4-18 and 4-19.
(5) Compute the entries a_i and b_i of T_k, according to 4-23 and 4-24.

Suppose bidiagonal data (λ, β) is given and we want to recover $T = T_k$. We compute the first steps of the algorithm explicitly, for the reader's convenience. Clearly, $T_1 = \lambda_1$ and $u_1 = 1$, so the auxiliary matrix is

$$S = \begin{pmatrix} \lambda_1 & 0 \\ \beta_1 & \lambda_2 \end{pmatrix} \ .$$

Solve the 1×1 linear system $(\lambda_1 - \lambda_2)v = \beta_1$, obtaining $v = \frac{\beta_1}{\delta_{21}}$. We have

$$\tilde{U}^{-1} = \begin{pmatrix} \frac{\sqrt{\beta_1^2+\delta_{21}^2}}{|\delta_{21}|} & \frac{-\beta_1}{\sqrt{\beta_1^2+\delta_{21}^2}} \\ 0 & \frac{|\delta_{21}|}{\sqrt{\beta_1^2+\delta_{21}^2}} \end{pmatrix}, \quad \tilde{U} = \begin{pmatrix} \frac{|\delta_{21}|}{\sqrt{\beta_1^2+\delta_{21}^2}} & \frac{\beta_1}{\sqrt{\beta_1^2+\delta_{21}^2}} \\ 0 & \frac{\sqrt{\beta_1^2+\delta_{21}^2}}{|\delta_{21}|} \end{pmatrix},$$

$$T_2 = \tilde{U}^{-1} S \tilde{U} = \begin{pmatrix} \lambda_1 - \frac{\beta_1^2\delta_{21}}{\beta_1^2+\delta_{21}^2} & \frac{\beta_1\delta_{21}^2}{\beta_1^2+\delta_{21}^2} \\ \frac{\beta_1\delta_{21}^2}{\beta_1^2+\delta_{21}^2} & \lambda_2 + \frac{\beta_1^2\delta_{21}}{\beta_1^2+\delta_{21}^2} \end{pmatrix}.$$

To find T_3, we must solve $(T_2 - \lambda_3)v = \beta_2 u_2 e_2$, i.e.,

$$\begin{pmatrix} \lambda_1 - \lambda_3 - \frac{\beta_1^2\delta_{21}}{\beta_1^2+\delta_{21}^2} & \frac{\beta_1\delta_{21}^2}{\beta_1^2+\delta_{21}^2} \\ \frac{\beta_1\delta_{21}^2}{\beta_1^2+\delta_{21}^2} & \lambda_2 - \lambda_3 + \frac{\beta_1^2\delta_{21}}{\beta_1^2+\delta_{21}^2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_2\sqrt{\frac{\beta_1^2}{\delta_{21}^2}+1} \end{pmatrix}.$$

As k increases, longer combinations of β's by δ_{ij} appear on the algorithm. Relative sizes may induce concerns on the stability of the algorithm. The formulas (and Figure 4.3) suggest that data closer to a diagonal matrix associated with λ are more stable.

4.2.1 Computing u_{k-1}

According to equation 4-21, the computation of u_{k-1} requires the knowledge of β_i's and differences of eigenvalues $\delta_{ij} = \lambda_j - \lambda_i$,

$$u_{k-1} = \sqrt{\frac{\beta_1^2 \dots \beta_{k-2}^2}{\delta_{12}^2 \dots \delta_{1,k-1}^2} + \dots + \frac{\beta_{k-2}^2}{\delta_{k-2,k-1}^2} + 1}\,.$$

From formula 4-4, u_{k-1} is the norm of the last row of L_{k-1}^{-1}. This leads to a recursive definition in terms of the last row of L_{k-2}^{-1}. Let

$$\hat{\ell} = \begin{pmatrix} \frac{\beta_1 \dots \beta_{k-3}}{\delta_{12}\dots\delta_{1,k-2}} & \frac{\beta_2 \dots \beta_{k-3}}{\delta_{23}\dots\delta_{2,k-2}} & \dots & \frac{\beta_{k-3}}{\delta_{k-3,k-2}} & 1 & 1 \end{pmatrix}^T = (e_{k-2}^T L_{k-2}^{-1}, 1) \in \mathbb{R}^{k-1}\,.$$

The entries of $(\ell, 1)$ — the last row of L_k^{-1} — are obtained by multiplying the entries of $\hat{\ell}$ by the k numbers

$$\frac{\beta_{k-2}}{\delta_{1,k-1}}, \quad \dots, \quad \frac{\beta_{k-2}}{\delta_{k-2,k-1}}, 1\,,$$

then we obtain u_{k-1} by taking the norm of the resulting vector.

4.2.2 Solving the tridiagonal system

Let $\alpha = \beta_{k-1}u_{k-1}$, already computed. We consider the $(k-1)\times(k-1)$ tridiagonal system $(T_{k-1} - \lambda_k)v = \alpha e_{k-1}$, where

$$T_{k-1} = \begin{pmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & a_3 & \ddots & \\ & & \ddots & \ddots & b_{k-2} \\ & & & b_{k-2} & a_{k-1} \end{pmatrix}, \tag{4-25}$$

For purposes of inversion, we may suppose $\alpha = 1$. Gaussian elimination — essentially the LU decomposition of $(T_{k-1}-\lambda_k)$ — is a possibility among many. Recall that, in this case, L is lower bidiagonal with ones along the diagonal, and U is upper bidiagonal [11]. If the eigenvalues are in strictly increasing or decreasing order, $(T_{k-1} - \lambda_k)$ has invertible principal minors, by interlacing. More precisely, λ_k is not in the smallest interval I containing the remaining eigenvalues, and by interlacing the minors have spectrum contained in I.

A possible source of loss of precision is the subtraction $T_{k-1}-\lambda_k$. this may be circumvented by an appropriate shift strategy, such that only differences δ_{ij} are used in the algorithm up to the very end, when some eigenvalue must be added to the outcome.

This system is endowed with a rare property: we know the spectrum of the associated matrix. Unfortunately, we were not able to obtain an alternative algorithm from this fact.

5
Comparing the algorithms, numerical examples

We initially test both RKPW and INVBI with the canonical example: the discretization of the second order derivative on functions in an interval satisfying Dirichlet conditions. Without real loss, we perform minor modifications on this matrix. Concretely, the $n \times n$ matrix

$$M = \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & 0 & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & 0 \end{pmatrix}$$

has eigenvalues and associated normalized eigenvectors

$$\lambda_k = 2\cos\left(\frac{k\pi}{n+1}\right) \ , \quad u_k = \frac{\sqrt{2}}{\sqrt{n+1}} \begin{pmatrix} \sin(\frac{k\pi}{n+1}) \\ \sin(\frac{2k\pi}{n+1}) \\ \vdots \\ \sin(\frac{kn\pi}{n+1}) \end{pmatrix} \ .$$

The Moser vector of M is

$$c^T = \frac{\sqrt{2}}{\sqrt{n+1}} \begin{pmatrix} \sin(\frac{\pi}{n+1}) & \sin(\frac{2\pi}{n+1}) & \dots & \sin(\frac{n\pi}{n+1}) \end{pmatrix} \ .$$

The bidiagonal coordinates β are obtained from c using formula 4-6.

We run RKPW and INVBI for dimensions $n = 10, 50, 100$ and 500. The eigenvalues were presented in increasing order, and INVBI used the LU decomposition to solve the linear system. We compare the time of execution in seconds and three measures of error with respect to M: the largest deviation among diagonal entries, ϵ_d; the largest deviation among off-diagonal entries, ϵ_{off}; the sum of all deviations of diagonal and subdiagonal entries, ϵ_t. The results are recorded in table below.

	n	Time	ϵ_d	ϵ_{off}	ϵ_t
RKPW	10	0.008	2.05391×10^{-15}	8.88178×10^{-16}	1.12895×10^{-14}
	50	0.725	2.39808×10^{-14}	6.21724×10^{-15}	2.43282×10^{-13}
	100	3.030	4.06341×10^{-14}	2.68674×10^{-14}	1.04737×10^{-12}
	500	71.602	2.61457×10^{-13}	1.32338×10^{-13}	2.35667×10^{-11}
	1000	286.953	5.34017×10^{-13}	2.54907×10^{-13}	9.04898×10^{-11}
INVBI	10	0.007	1.27675×10^{-15}	6.66133×10^{-16}	7.96585×10^{-15}
	50	0.375	5.74258×10^{-15}	3.10862×10^{-15}	9.44603×10^{-14}
	100	1.393	1.03929×10^{-14}	4.10782×10^{-15}	2.87122×10^{-13}
	500	35.295	2.91766×10^{-13}	5.93969×10^{-14}	4.03024×10^{-12}
	1000	145.206	1.12206×10^{-13}	8.17124×10^{-14}	9.91484×10^{-12}

We now give an example of how slightly different Moser vectors can generate very different matrices out of RKPW. For $\Lambda = \mathrm{diag}(1, 2, 4)$, the Moser vectors $c = (0, 1, 0)$ and $\tilde{c} = (0, 1.00001, 0.00001)$ we obtain

$$T = \begin{pmatrix} 2 & & \\ & 1 & \\ & & 4 \end{pmatrix}, \quad \tilde{T} = \begin{pmatrix} 2.0000000002 & 1.99998 \times 10^{-5} & 0 \\ 1.99998 \times 10^{-5} & 3.9999999998 & 0 \\ 0 & 0 & 1.0 \end{pmatrix}.$$

This result is predicted by the analysis of stable sequences in [15].

Finally, we show the instability of the INVBI algorithm for different asymptotic classes $[\beta_1, \beta_2]$, illustrated in Figures 4.3 and 4.4. For $\Lambda = \mathrm{diag}(1, 2, 4)$, we consider three cases: $(10^4, 10^{-5}) \sim [\infty, 0]$, $(10^4, 10^{-1}) \sim [\infty, +]$ and $(10^4, 10^4) \sim [\infty, \infty]$, with outputs

$$\begin{pmatrix} 1.99999999000556 & 0.00010005553913244 & 0 \\ 0.00010005553913244 & 1.00332964378459 & 0.0998890127057464 \\ 0 & 0.0998890127057464 & 3.99667036620985 \end{pmatrix}$$

$$\begin{pmatrix} 2.00055539127465 & 0.0333242266451282 & 0 \\ 0.0333242266451282 & 3.99941758646916 & 0.00900116874515525 \\ 0 & 0.00900116874515525 & 1.00002702225619 \end{pmatrix}$$

$$\begin{pmatrix} 3.99999928000025 & 0.00119999958150015 & 0 \\ 0.00119999958150015 & 2.00000069749975 & 0.000150000023624997 \\ 0 & 0.000150000023624997 & 1.0000000225 \end{pmatrix}.$$

Switching the order of the β coordinates yields approximations of the remaining three diagonal matrices.

5.0.0.1 Python code

The algorithms in this work were implemented using Python, and a notebook with the code and all experiments mentioned above is available at `https://colab.research.google.com/drive/13jcKEnB_Au20QtlBws5HZDJsV3K-m5ep?usp=sharing`.

Bibliography

[1] W. B. Gragg and W. J. Harrod, "The numerically stable reconstruction of jacobi matrices from spectral data," *Numerische Mathematik*, vol. 44, pp. 317–335, 1984.

[2] J. Moser, "Finitely many mass points on the line under the influence of an exponential potential," *Dynamic systems theory and applications*, pp. 467–497, 1975.

[3] R. S. Leite, N. C. Saldanha, and C. Tomei, "An atlas for tridiagonal isospectral manifolds," *Linear Algebra and its Applications*, vol. 429, pp. 387–402, 2018.

[4] C. Tomei, "The topology of isospectral manifolds of tridiagonal matrices," *Duke Mathematical Journal*, vol. 51, pp. 981–996, 1984.

[5] A. Bloch, H. Flaschka, and T. Ratiu, "A convexity theorem for isospectral manifolds of jacobi matrices in a compact lie algebra," *Duke Mathematical Journal*, vol. 61, pp. 41–65, 1990.

[6] R. S. Leite and C. Tomei, "Parametrization by polytopes of intersections of orbits by conjugation," *Linear Algebra and its Applications*, vol. 361, pp. 223–243, 2003.

[7] G. H. Golub and J. H. Welsch, "Calculation of gauss quadrature rules," *Math. Comput.*, vol. 23, pp. 221–230, 1969.

[8] R. S. Leite, N. C. Saldanha, and C. Tomei, "Dynamics of the symmetric eigenvalue problem with shift strategies," *International Mathematics Research Notices*, vol. 2013, pp. 4382–4412, 2012.

[9] R. S. Leite, N. C. Saldanha, and C. Tomei, "The asymptotics of wilkinson's shift: Loss of cubic convergence," *Foundations of Computational Mathematics*, vol. 10, pp. 15–36, 2010.

[10] B. Parlett, *The Symmetric Eigenvalue Problem*. SIAM, 1980.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[12] P. Deift and E. Trubowitz, "Inverse scattering on the line," *Communications on Pure and Applied Mathematics*, vol. 32, pp. 121–251, 1979.

[13] A. W. Marshall, I. Olkin, and A. B. C., *Inequalities: Theory of Majorization and its Applications*. Springer, 1979.

[14] M. Atiyah, "Convexity and commuting hamiltonians," *Bull. London Math. Soc.*, vol. 14, pp. 1–15, 1982.

[15] P. C. Gibson, "Spectral distributions and isospectral sets of tridiagonal matrices," *arXiv:math/0207041v1 [math.SP]*, 2002.

[16] C. Tomei, "The toda lattice, old and new," *arXiv:1508.03229v1 [math.DS]*, 2015.

[17] R. S. Leite, N. C. Saldanha, and C. Tomei, "Reconstruction of tridiagonal matrices from spectral data," *arXiv:math/0508099v1 [math. NA]*, 2005.