Differential and subdifferential properties of symplectic eigenvalues

Hemant Kumar Mishra

Indian Statistical Institute

April 19, 2021

Differential and subdifferential properties of symplectic eigenvalues

Hemant Kumar Mishra

Thesis Supervisor: Prof. Tanvi Jain

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy

April, 2021

Indian Statistical Institute
7, S.J.S. Sansanwal Marg, New Delhi, India.

To my parents and teachers

Acknowledgement

Foremost, I would like to express deepest gratitude to my advisor Prof. Tanvi Jain for her continued support, enthusiasm and patience. I am especially thankful to her for always being available for discussion and for her guidance throughout the research. She always encouraged me to become an independent researcher and guided me at every step to becoming one. Her immense knowledge, commitment to research and attention to detail brought the best out of me. I am also thankful to her for bearing with my mistakes, and for providing critical comments and corrections in the thesis. Thank you Tanvi ma'am for the wonderful opportunity of working with you.

I am profoundly indebted to the faculty of Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi Centre for providing us an inspiring research atmosphere. My special thanks to Prof. Rajendra Bhatia for exposing us to potential research problems during the courses he taught us. He shared some great insights into solving the problems and motivated us to pursue new ideas. I am thankful to Arup K. Pal, Rahul Roy, Maneesh Thakur, Shanta Laishram and Anish Sarkar for teaching us various research level courses. Thank you Gursharn Kaur for being so welcoming and helping me with tons of official work related queries. I thank Gunjan Sapra for carefully reading through the chapters of the thesis and providing useful comments.

Heartfelt thanks to all my friends at Indian Statistical Institute for supporting directly or indirectly during my research. I am especially thankful to Komal Malik. She has been a wonderful friend to me even though I gave her tough times. Our passionate discussions on world issues, social issues, movies, politics, travels, etc. have been amazing pastimes.

I am thankful to my parents beyond words for their love and support throughout my life. If it wasn't for them, none of this would have been possible. Thanks mom and dad!

This research work was done in Indian Statistical Institute, Delhi. I am thankful to the institute for the research fellowship and the facilities provided to me.

Hemant Kumar Mishra

April, 2021

New Delhi, India

Abstract

A real $2n \times 2n$ matrix M is called a symplectic matrix if $M^T JM = J$, where J is the $2n \times 2n$ matrix given by $J = \left(\begin{smallmatrix} O & I_n \\ -I_n & O \end{smallmatrix}\right)$ and I_n is the $n \times n$ identity matrix. A result on symplectic matrices, generally known as Williamson's theorem, states that for any $2n \times 2n$ positive definite matrix A there exists a symplectic matrix M such that $M^T AM = D \oplus D$ where D is an $n \times n$ positive diagonal matrix with diagonal entries $0 < d_1(A) \leq \cdots \leq d_n(A)$ called the symplectic eigenvalues of A. In this thesis, we study differentiability and analyticity properties of symplectic eigenvalues and corresponding symplectic eigenbasis. In particular, we prove that *simple* symplectic eigenvalues are infinitely differentiable and compute their first order derivative. We also prove that symplectic eigenvalues and corresponding symplectic eigenbasis for a real analytic curve of positive definite matrices can be chosen real analytically. We then derive an analogue of Lidskii's theorem for symplectic eigenvalues as an application of our analysis. We study various subdifferential properties of symplectic eigenvalues such as Fenchel subdifferentials, Clarke subdifferentials and Michel-Penot subdifferentials. We show that symplectic eigenvalues are directionally differentiable and derive the expression of their first order directional derivatives.

Notations

$\mathbb{R}^n$	The set of ordered n tuples of real numbers
$\mathbb{C}^n$	The set of ordered n tuples of complex numbers
conv$\mathcal{K}$	The closed convex set generated by any subset $\mathcal{K}$ of a Euclidean space
int$\mathcal{K}$	The interior of any subset $\mathcal{K}$ of a Euclidean space
$\mathbb{M}_{m,n}(\mathbb{R})$	The set of $m \times n$ real matrices
$\mathbb{M}_n(\mathbb{C})$	The set of $n \times n$ complex matrices
$\mathbb{H}_n(\mathbb{C})$	The set of $n \times n$ Hermitian matrices
$\mathbb{S}_n(\mathbb{R})$	The set of $n \times n$ real symmetric matrices
$\mathbb{P}_n(\mathbb{R})$	The set of $n \times n$ real positive definite matrices
$\mathrm{Diag}(x)$	The diagonal matrix whose diagonal entries are the components of x
tr$A$	The trace of any square matrix A
$\det A$	The determinant of any square matrix A
ker$A$	The kernel of any matrix A
I_n	The identity matrix of size n
J_{2n}	The $2n \times 2n$ block matrix $\left(\begin{smallmatrix} O & I_n \\ -I_n & O \end{smallmatrix}\right)$, denoted by J when n is clear from the context
$Sp(2n)$	The set of $2n \times 2n$ symplectic matrices
$\|A\|$	The operator norm of any matrix A
$\|A\|_F$	The Frobenius norm of any matrix A
$\|\|\|A\|\|\|$	A unitarily invariant norm of any matrix A
$\kappa(A)$	The condition number of any invertible matrix A
$\partial\Phi(x)$	The Fenchel subdifferential of Φ at x
$\partial^\circ\Phi(x)$	The Clarke subdifferential of Φ at x
$\partial^\diamond\Phi(x)$	The Michel-Penot subdifferential of Φ at x

Contents

Introduction

A real $2n \times 2n$ matrix M is called a *symplectic matrix* if $M^T J M = J$, where J is the $2n \times 2n$ matrix given by $J = \left(\begin{smallmatrix} O & I_n \\ -I_n & O \end{smallmatrix}\right)$, I_n is the $n \times n$ identity matrix. Symplectic matrices are ubiquitous in various fields such as symplectic geometry [21], quantum optics [67, 69], quantum information [24, 62], hamiltonian dynamics [37] and optimization problems [16, 26]. A well known class of matrices in classical linear algebra is the set of orthogonal matrices. Replacing J by the identity matrix in the definition of symplectic matrices yields orthogonal matrices. As the orthogonal matrices form a group under matrix multiplication so do the symplectic matrices. The group of $2n \times 2n$ symplectic matrices is called the symplectic group and is denoted by $Sp(2n)$. The symplectic group exhibits some similar properties as the orthogonal group, e.g., $M \in Sp(2n)$ implies $M^T \in Sp(2n)$ and $\det M = 1$. But unlike the orthogonal group, this group is not compact. See [23].

By the spectral theorem we know that any positive definite matrix can be reduced to a diagonal matrix by an orthogonal congruence. The diagonal entries of the diagonal matrix are called the eigenvalues of the given matrix. A symplectic counterpart of the spectral theorem, generally known as *Williamson's theorem*, states that for any $2n \times 2n$ positive definite matrix A there exists $M \in Sp(2n)$ such that

$$M^T A M = \begin{pmatrix} D & O \\ O & D \end{pmatrix}, \tag{1}$$

where D is an $n \times n$ positive diagonal matrix with diagonal entries $d_1(A) \leq \cdots \leq d_n(A)$. The positive numbers $d_1(A), \ldots, d_n(A)$ are uniquely determined by (1). These are the complete invariants of A under the action of the symplectic group $Sp(2n)$ and are called the *symplectic* *eigenvalues* of A. See [21, 36].

Symplectic eigenvalues appear in various applications such as classical and quantum mechanics [21], symplectic topology [37] and harmonic oscillator systems [1, 58]. Recently there has been a heightened interest in the study of symplectic eigenvalues by both physicists and mathematicians. A particular reason for this being their growing importance and applications in quantum information. Associated with an n mode quantum state is a $2n \times 2n$ positive definite matrix known as the covariance matrix of the quantum state. The Heisenberg uncertainty principle tells us that a $2n \times 2n$ positive definite matrix A is the covariance matrix of a Gaussian state if and only if $d_j(A) \geq \frac{1}{2}$ for all $j = 1, \dots, n$. See [2, 23]. The class of Gaussian states in a quantum system is being widely studied. Of interest there are various *entropy* functions associated with Gaussian states that are useful in measurement of the degree of mixedness and entanglement of the Gaussian states. These entropy functions can be expressed as smooth maps of symplectic eigenvalues of the covariance matrices. See [3, 36, 45]. So it is useful and important to have a well developed theory for symplectic eigenvalues as we have for eigenvalues.

Eigenvalue problems can be classified as quantitative and qualitative in nature. The quantitative problems include variational principles, eigenvalues of functions of matrices, majorisation inequalities and computation of eigenvalues and eigenvectors. See [9, 41]. There has been much interest in the study of relationships between the eigenvalues of Hermitian matrices A and B and those of their sum $A + B$. Suppose $\lambda^{\uparrow}(A) = \left(\lambda_1^{\uparrow}(A), \dots, \lambda_n^{\uparrow}(A)\right)$ denotes the tuple of eigenvalues of an $n \times n$ Hermitian matrix A arranged in increasing order. In 1912, H. Weyl discovered several relationships between the eigenvalues of sums of Hermitian matrices one of which states that

$$\lambda_j^{\uparrow}(A + B) \geq \lambda_j^{\uparrow}(A) + \lambda_1^{\uparrow}(B), \qquad 1 \leq j \leq n. \tag{2}$$

The maximum principle given by Ky Fan [25] in 1949 implies that for all $1 \leq k \leq n$,

$$\sum_{j=1}^{k} \lambda_j^{\uparrow}(A + B) \geq \sum_{j=1}^{k} \lambda_j^{\uparrow}(A) + \sum_{j=1}^{k} \lambda_j^{\uparrow}(B). \tag{3}$$

In 1950, V. B. Lidskii [50] proved the inequalities

$$\sum_{j=1}^{k} \lambda_{i_j}^{\uparrow}(A+B) \geq \sum_{j=1}^{k} \lambda_{i_j}^{\uparrow}(A) + \sum_{j=1}^{k} \lambda_{j}^{\uparrow}(B) \tag{4}$$

for all $k = 1, \ldots, n$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. Inequalities (2) and (3) are special cases of (4). Lidskii's inequalities played a fundamental role in the study of eigenvalues of sums of matrices and proved to be an important stimulant for the much celebrated Horn's conjecture. See, for instance, [8, 28]. These inequalities have attracted much attention and a number of different proofs for these are now available in literature. See [7, 48]. But all the proofs are generally more difficult than those for the earlier two families of inequalities (2) and (3). An example of qualitative problems is the study of continuity, differentiability, subdifferentiability and analyticity of eigenvalues and eigenvectors as functions of matrices. These problems have been studied for a long time. See the classical books by Kato [43] and Rellich [64]. See also [33, 34, 42, 44, 51, 66, 73, 74]. These have applications in optimisation [20], linear programming [49], numerical analysis [53, 74] and physics [70]. It is natural and fundamental to study the analogous quantitative and qualitative problems of symplectic eigenvalues.

A positive number d is a symplectic eigenvalue of A if and only if $\pm d$ is an eigenvalue of the Hermitian matrix $\imath A^{1/2} J A^{1/2}$ and of the matrix $\imath JA$. This connection has been used to study various quantitative properties of symplectic eigenvalues such as perturbation theorems, majorisation inequalities, variational principles and an interlacing theorem in the past few years. See [10, 11, 38]. For instance, the following minmax principles for symplectic eigenvalues were given in [37]. For all $1 \leq j \leq n$, we have

$$\frac{1}{d_j(A)} = \max_{\substack{\mathcal{M} \subset \mathbb{C}^{2n} \\ \dim \mathcal{M} = j}} \min_{\substack{x \in \mathcal{M} \\ \langle x, Ax \rangle = 1}} \langle x, iJx \rangle, \tag{5}$$

and also

$$\frac{1}{d_j(A)} = \min_{\substack{\mathcal{M} \subset \mathbb{C}^{2n} \\ \dim \mathcal{M} = 2n-j+1}} \max_{\substack{x \in \mathcal{M} \\ \langle x, Ax \rangle = 1}} \langle x, iJx \rangle. \tag{6}$$

T. Hiroshima [36] gave the following symplectic counterpart of Ky Fan's extremal principle

$$2\sum_{j=1}^{m} d_j(A) = \min_{M} \mathrm{tr} M^T A M, \tag{7}$$

for all $1 \leq m \leq n$ where the minimum is taken over $2n \times 2m$ real matrices M satisfying $M^T J_{2n} M = J_{2m}$ with $J_{2k} = \begin{pmatrix} O & I_k \\ -I_k & O \end{pmatrix}$. Using equation (7) he also derived a symplectic analogue of the inequalities (3) which states that

$$\sum_{j=1}^{k} d_j(A+B) \geq \sum_{j=1}^{k} d_j(A) + \sum_{j=1}^{k} d_j(B) \tag{8}$$

for $1 \leq k \leq n$. However, some quantitative results on symplectic eigenvalues are not easy to derive using the theory of eigenvalues. The Weyl's inequalities (2) are a direct consequence of the Courant-Fischer-Weyl minmax principle. But it is difficult to obtain a symplectic analogue of the Weyl's inequalities using the minmax principles (5) and (6) because these involve $\frac{1}{d_j(A)}$ and the minmax conditions are dependent on A. The fact that the symplectic group is not compact makes it tricky to get a nice analogue of the Courant-Fischer-Weyl minmax principle for symplectic eigenvalues. Using the connection that $\pm d_j(A)$ are the eigenvalues of $\imath A^{1/2} J A^{1/2}$, a symplectic analogue of the Weyl's inequalities for a special class of positive definite matrices appeared in [10] which states that for all $j = 1, \ldots, n$

$$d_j(A+B) \geq d_j(A) + d_1(B) \tag{9}$$

when A and B are of the form $A = \begin{bmatrix} D & O \\ O & D \end{bmatrix}$, $B = \begin{bmatrix} X & O \\ O & X^{-1} \end{bmatrix}$, where D is an $n \times n$ positive diagonal matrix and X is any $n \times n$ positive definite matrix. A generalisation of the inequalities (9) to any $2n \times 2n$ positive definite matrices A, B was derived recently in [13] using an independent theory of symplectic eigenvalues.

There have not been any explicit study of the qualitative properties of symplectic eigenvalues in the literature to the best of our knowledge. In principle, it could be possible to study qualitative properties of symplectic eigenvalues by using their connection with eigenvalues but it is not feasible in practice. The matrix $\imath JA$ is not even normal, and the matrix $\imath A^{1/2} J A^{1/2}$ has a complicated form which makes it difficult to obtain results for symplectic eigenvalues from the well-developed theory of eigenvalues of Hermitian matrices. For instance, even though the matrix square root map is differentiable [22], no closed-form expression for its derivative is known. Therefore, it is difficult to compute derivative and generalised derivative expressions of d_j when treating $d_j(A)$ as eigenvalues of $\imath A^{1/2}JA^{1/2}$. Moreover, the appearance of $A^{1/2}$ obscures even the simplest properties of symplectic eigenvalues. By the characterisation (7), it is easy to verify that d_j is a difference of concave functions. On the other hand, the map $A \mapsto \imath A^{1/2}JA^{1/2}$ is neither convex nor concave as illustrated by the following example.

Example 1. For any 2×2 positive definite matrix A, define $\phi(A) = \imath A^{1/2}JA^{1/2}$. Let $A = \mathrm{Diag}(1, 4)$. We have $\phi(I_2) = \imath J$, $\phi(A) = 2\imath J$ and $\phi\left((I_2 + A)/2\right) = (\imath\sqrt{10}/2)J$. This gives

$$\phi\left(\frac{I_2 + A}{2}\right) - \frac{1}{2}(\phi(I_2) + \phi(A)) = \frac{\imath(\sqrt{10} - 3)}{2}J$$

which is neither negative nor positive semidefinite. This implies ϕ is neither convex nor concave.

This makes it difficult to establish the fact that d_j is a difference of concave functions from the corresponding properties of eigenvalues of Hermitian matrices. Therefore, it is important and necessary to develop a theory for symplectic eigenvalues independent of eigenvalues. Very recently, some quantitative results on symplectic eigenvalues using an independent theory of symplectic eigenvalues were given in [12, 13].

In this thesis, we study various qualitative properties of symplectic eigenvalues such as differentiability, subdifferentiability and analyticity as well as some fundamental class of inequalities on symplectic eigenvalues. We also develop, in the course of the thesis, a novel theory and techniques that can be used to study symplectic eigenvalues further. The symplectic eigenvalue maps d_j are known to be continuous. See for example Theorem 7 of [11]. But the following example shows that they are not differentiable in general.

Example 2. Let B be the 4×4 matrix $B = I_2 \otimes \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. We have $d_1(I_4 + tB) = 1 - |t|$ and $d_2(I_4 + tB) = 1 + |t|$ for any $t \in (-1, 1)$. The modulus function is not differentiable at $t = 0$. So d_1 and d_2 are not differentiable at I_4.

We show that d_j is infinitely differentiable at A if $d_j(A)$ is a *simple* symplectic eigenvalue, i.e., $d_i(A) \neq d_j(A)$ for all $i \neq j$, and derive its first order derivative expression. We also study the differentiability and analyticity property of symplectic eigenvalues of positive definite matrices depending on a real parameter. In particular, we prove that symplectic eigenvalues of a real analytic curve of positive definite matrices can be chosen real analytically. Even though the symplectic eigenvalue maps are not differentiable, we can talk about *generalised derivatives* of symplectic eigenvalues. Generalised derivatives are weaker versions of derivatives. In non-smooth optimisation, one often deals with non-differentiable functions. In the absence of differentiability, generalised derivatives play important roles. See [17, 19, 32, 52]. Among some useful generalised derivatives are directional derivative, Clarke directional derivative and Michel-Penot directional derivative. Using the generalised derivatives, several notions of generalised gradients or *subdifferentials* are defined for various class of functions, e.g., convex functions, locally Lipschitz functions. These include Fenchel subdifferential, Clarke subdifferential and Michel-Penot subdifferential. Subdifferentials are useful in obtaining various optimality conditions in non-smooth optimisation, see e.g., [5, 19, 65]. Let $\sigma_m(A) = -2\sum_{j=1}^{m} d_j(A)$. By the extremal characterisation (7), it is easy to see that σ_m is a convex function. We use this fact to show that σ_m is directionally differentiable and derive the expressions for its Fenchel subdifferential and directional derivative. We also show that symplectic eigenvalues are directionally differentiable and compute the expressions for their directional derivatives. We use the fact that symplectic eigenvalues are locally Lipschitz to derive the expressions for their Clarke and Michel-Penot directional derivatives and subdifferentials. We also derive a quantitative property of symplectic eigenvalues, an analogue of the Lidskii's inequalities (4), using the analyticity of symplectic eigenvalues. We show that for all $k = 1, \dots, n$ and all $1 \leq i_1 < \cdots < i_k \leq n$,

$$\sum_{j=1}^{k} d_{i_j}(A+B) \geq \sum_{j=1}^{k} d_{i_j}(A) + \sum_{j=1}^{k} d_j(B). \tag{10}$$

As for the case of eigenvalues of Hermitian matrices, these greatly generalise the inequalities (8) and (9). We emphasise that the proofs of Lidskii's inequalities (4) for eigenvalues are non-trivial. So deriving the inequalities (10) using the connection of symplectic eigenvalues with eigenvalues would be a more difficult problem.

The thesis is organised as follows. In Chapter 1, we introduce the notion of *symplectic eigenvector pairs*, and give some preliminary results on symplectic eigenvalues and symplectic eigenvector pairs. We introduce the notion of *symplectic projection* in Section 1.3 and give an extension of the Williamson's theorem to a class of positive semidefinite matrices. In the beginning of Chapter 2, we recall some basic definitions on differentiability of functions on Banach spaces. We review the theory of differentiability and analyticity of eigenvalues of Hermitian matrices in Section 2.2. In Section 2.3, we prove that d_j is infinitely differentiable at A if $d_j(A)$ is simple, and also derive its derivative expression. We study the differentiability and analyticity properties of symplectic eigenvalues of curves of positive definite matrices in Section 2.4. We prove that the symplectic eigenvalues and symplectic eigenvector pairs for a real analytic curve of positive definite matrices can be chosen analytically. In Section 2.5, we derive the majorisation inequalities (10) and give some other applications of our analysis of symplectic eigenvalues. In Chapter 3, we recall some basic theory of Fenchel subdifferentials and directional derivatives of convex functions. We also review the theory of directional derivatives of eigenvalues of symmetric matrices. In Section 3.3, we derive the expression for the Fenchel subdifferential and the first order directional derivative of σ_m. We then show in Section 3.4 that the directional derivative of d_j exists and compute its expression. In Chapter 4, we review the theory of Clarke and Michel-Penot directional derivatives and subdifferenatials of locally Lipschitz functions on real Banach spaces. In Section 4.2, we compute the expressions for their Clarke and Michel-Penot subdifferentials and also the expressions for the Clarke and Michel-Penot directional derivatives. We use these subdifferentials to give an alternate proof of the monotonicity principle of symplectic eigenvalues.

Chapter 1

Preliminaries

In this chapter, we recall the theory of symplectic spaces and symplectic matrices. We establish some preliminary results that is fundamental to our study of symplectic eigenvalues.

We review some basic properties of real symplectic spaces in Section 1.1. In Section 1.2 we recall the definition of symplectic matrices and discuss some useful results on these matrices. We state a fundamental result on symplectic matrices known as Williamson's theorem, recall the definition of symplectic eigenvalues and introduce the notion of *symplectic eigenvector pairs*. We also derive some preliminary results on symplectic eigenvalues and symplectic eigenvector pairs. In Section 1.3, we introduce the notion of *symplectic projection* that is useful in our analysis of symplectic eigenvalues. We give an extension of Williamson's theorem to a class of semidefinite matrices at the end of the section.

Throughout the chapter, we only deal with finite dimensional real vector spaces and real matrices unless stated otherwise.

1.1 Symplectic spaces

In this section, we discuss some geometrical properties of symplectic spaces useful in our present work. These properties are in contrast to the properties of inner product spaces familiar to us. This section is based on Chapter 1 of [21].

1.1.1 Symplectic forms

Let $\mathcal{X}$ be a vector space. A symplectic form ω on $\mathcal{X}$ is a mapping $\omega : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is

(i) bilinear:

$$\omega(au + bv, w) = a\omega(u, w) + b\omega(v, w),$$

$$\omega(w, au + bv) = a\omega(w, u) + b\omega(w, v)$$

for all $u, v, w \in \mathcal{X}$ and $a, b \in \mathbb{R}$;

(ii) antisymmetric: $\omega(u, v) = -\omega(v, u)$ for all $u, v \in \mathcal{X}$;

(iii) non-degenerate: $\omega(u, v) = 0$ for all $v \in \mathcal{X}$ if and only if $u = 0$.

By virtue of bilinearity, the antisymmetric property is equivalent to the condition $\omega(u, u) = 0$ for all $u \in \mathcal{X}$. A symplectic space $(\mathcal{X}, \omega)$ is a vector space $\mathcal{X}$ equipped with a symplectic form ω.

Example 3. Let K be a $2n \times 2n$ skew-symmetric and non-singular matrix. The K induced map $(x, y) \mapsto x^T K y$ for all $x, y \in \mathbb{R}^{2n}$ is a symplectic form on $\mathbb{R}^{2n}$. The map is clearly bilinear. Its antisymmetry follows from $K^T = -K$, and non-degeneracy is a consequence of the fact that K is non-singular.

Symplectic spaces are even dimensional. Let $(\mathcal{X}, \omega)$ be a symplectic space and dim$\mathcal{X} = m$. Fix a basis $\mathcal{B} = \{u_1, \ldots, u_m\}$ of $\mathcal{X}$. Let Ω be the $m \times m$ matrix with the ijth entry given by $\omega(u_i, u_j)$ for $1 \leq i, j \leq m$. By antisymmetry of ω, the matrix Ω is skew-symmetric. Also, bilinearity and non-degeneracy of ω imply that kerΩ is trivial. So, Ω is a non-singular skew-symmetric matrix and therefore its size m must be even.

1.1.2 Symplectic orthogonality and symplectic basis

Let $(\mathcal{X}, \omega)$ be a symplectic space and $u_j, v_j \in \mathcal{X}$ for $j = 1, 2$. We say that the pairs (u_1, v_1) and (u_2, v_2) are *symplectically orthogonal* to each other if

$$\omega(u_i, v_j) = \omega(u_i, u_j) = \omega(v_i, v_j) = 0,$$

for $i \neq j$, $i, j = 1, 2$. We call a set of vectors $\{u_1, \ldots, u_m, v_1, \ldots, v_m\}$ of $\mathcal{X}$ *symplectically orthogonal* if the pairs of vectors (u_j, v_j) are mutually symplectically orthogonal to each other. We call the set $\{u_1, \ldots, u_m, v_1, \ldots, v_m\}$ *symplectically orthonormal* if it is symplectically orthogonal as well as satisfies $\omega(u_j, v_j) = 1$ for all $j = 1, \ldots, m$.

Symplectically orthonormal sets are linearly independent. Let $\{u_1, \ldots, u_m, v_1, \ldots, v_m\}$ be a symplectically orthonormal set. Let $x = \sum_{j=1}^{m}(\alpha_j u_j + \beta_j v_j)$, where α_j, β_j be real numbers for $j = 1, \ldots, m$. By the definition of symplectically orthonormal sets we have,

$$\alpha_j = \omega(x, v_j), \qquad \beta_j = \omega(u_j, x) \tag{1.1}$$

for all $j = 1, \ldots, m$. So, $x = 0$ implies $\alpha_j = \beta_j = 0$ for all $j = 1, \ldots, m$. Consequently, a symplectically orthonormal set containing dim$\mathcal{X}$ number of vectors is a basis to the symplectic space.

Definition 1.1.1. A basis of a symplectic space is said to be symplectic basis if it is a symplectically orthonormal set.

In Theorem 1.1.3 we prove that every symplectic space has a symplectic basis. Symplectic bases are analogous to orthonormal bases in inner product spaces. For a fixed symplectic basis, every element in the symplectic space has a canonical representation. Let $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ be a symplectic basis of $(\mathcal{X}, \omega)$ and x be any vector in $\mathcal{X}$. Suppose $x = \sum_{j=1}^{n}(\alpha_j u_j + \beta_j v_j)$, where α_j, β_j are real numbers for $j = 1, \ldots, n$. By relations (1.1) we have

$$x = \sum_{j=1}^{n} \left(\omega(x, v_j)u_j - \omega(x, u_j)v_j\right).$$

1.1.3 Symplectic subspaces

Let X be any subset of a symplectic space $(\mathcal{X}, \omega)$. Define a set

$$X^{\perp_s} = \{u \in \mathcal{X} : \omega(u, x) = 0 \ \ \forall x \in X\}.$$

The set $X^{\perp_s}$ is known by various names in the literature [21, 23], we call it the *symplectic complement* of X. A subspace W of $\mathcal{X}$ is called a symplectic subspace if the restriction of the symplectic form ω on W is again a symplectic form. The symplectic form ω is bilinear and antisymmetric on W by definition. This means W is a symplectic subspace if ω is non-degenerate on W. Therefore, W is a symplectic subspace if and only if $W \cap W^{\perp_s} = \{0\}$.

We know that every subspace of an inner product space decomposes the space into the direct sum between the subspace and its orthogonal complement. Symplectic complements show an analogous property.

Proposition 1.1.2. *Let $(\mathcal{X}, \omega)$ be a symplectic space and W be a subset of $\mathcal{X}$. The set $W^{\perp_s}$ is a subspace of $\mathcal{X}$. If W is a subspace of $\mathcal{X}$ then*

$$dim W + dim W^{\perp_s} = dim \mathcal{X} \text{ and } (W^{\perp_s})^{\perp_s} = W.$$

Also, the following statements are equivalent.

(i) *W is a symplectic subspace.*

(ii) *$W^{\perp_s}$ is a symplectic space.*

(iii) *$\mathcal{X} = W \oplus W^{\perp_s}$.*

Proof. It is easy to verify that $W^{\perp_s}$ is a subspace of $\mathcal{X}$. Suppose W is a subspace of $\mathcal{X}$. Let $\{w_1, \ldots, w_k\}$ be a basis of W. The symplectic complement of W is given by

$$W^{\perp_s} = \{u \in \mathcal{X} : \omega(w_j, u) = 0, 1 \leq j \leq k\}. \tag{1.2}$$

Extend the basis of W to a basis $\{w_1, \ldots, w_{2n}\}$ of $\mathcal{X}$. Define the vector space isomorphism $\Psi : \mathbb{R}^{2n} \to \mathcal{X}$ by

$$\Psi(x) = \sum_{j=1}^{2n} x_j w_j$$

for all $x = (x_1, \ldots, x_{2n}) \in \mathbb{R}^{2n}$. Let $T_j : \mathbb{R}^{2n} \to \mathbb{R}$ be the linear functional given by

$$T_j(x) = \omega(w_j, \Psi(x))$$

for all $x \in \mathbb{R}^{2n}$ and $1 \leq j \leq k$. By (1.2) we have

$$\begin{aligned} W^{\perp_s} &= \{u \in \mathcal{X} : T_j(\Psi^{-1}(u)) = 0, 1 \leq j \leq k\} \\ &= \bigcap_{j=1}^{k} \Psi\left(\ker T_j\right) \\ &= \Psi\left(\bigcap_{j=1}^{k} \ker T_j\right). \end{aligned}$$

Thus we have

$$\dim W^{\perp_s} = \dim\left(\bigcap_{j=1}^{k} \ker T_j\right). \tag{1.3}$$

It follows from the linear independence of $\{w_1, \ldots, w_k\}$ and non-degeneracy of the symplectic form that $T_1, \ldots, T_k$ are linearly independent. Thus $\bigcap_{j=1}^{k} \ker T_j$ is the set of solutions of a system of k independent linear equations in $2n$ variables. Therefore by (1.3) we have $\dim W^{\perp_s} = 2n - k = \dim\mathcal{X} - \dim W$. Thus we have proved that

$$\dim W^{\perp_s} + \dim W = \dim\mathcal{X}. \tag{1.4}$$

Similarly we have

$$\dim W^{\perp_s} + \dim(W^{\perp_s})^{\perp_s} = \dim\mathcal{X}.$$

Thus we have $\dim(W^{\perp_s})^{\perp_s} = \dim W$. By definition $W \subseteq (W^{\perp_s})^{\perp_s}$. This implies $(W^{\perp_s})^{\perp_s} = W$.

We know that W is a symplectic subspace of $\mathcal{X}$ if and only if $W \cap W^{\perp_s} = \{0\}$. But $(W^{\perp_s})^{\perp_s} = W$. Therefore the statements (i) and (ii) are equivalent. It follows from (1.4) that $\mathcal{X} = W \oplus W^{\perp_s}$ if and only if $W \cap W^{\perp_s} = \{0\}$. This proves the equivalence of the statements (i) and (iii).

□

Theorem 1.1.3. *Every non-zero symplectic space has a symplectic basis.*

Proof. Let $(\mathcal{X}, \omega)$ be any non-zero symplectic space of dimension $2n$. Let u_1 be any non-zero element of $\mathcal{X}$. By non-degeneracy of ω and a suitable scaling, we get $v_1 \in \mathcal{X}$ such that $\omega(u_1, v_1) = 1$. The set $\{u_1, v_1\}$ is a symplectically orthonormal set. Let $W_1 = \text{span}\{u_1, v_1\}^{\perp_s}$. By Proposition 1.1.2, W_1 is a symplectic subspace of $\mathcal{X}$ of dimension $2n-2$. If W_1 is a non-zero subspace, we similarly get $u_2, v_2 \in W_1$ such that $\omega(u_2, v_2) = 1$. The set $\{u_1, u_2, v_1, v_2\}$ is a symplectically orthonormal set. Repeat this process to get at the kth step a symplectically orthonormal set $\{u_1, \ldots, u_k, v_1, \ldots, v_k\}$. Let $W_k = \text{span}\{u_1, \ldots, u_k, v_1, \ldots, v_k\}^{\perp_s}$. By Proposition 1.1.2 we have $\dim W_k = 2n - 2k$. The process stops when W_k is the zero subspace of $\mathcal{X}$ in which case $k = n$. The set $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ so obtained is a symplectic basis of $(\mathcal{X}, \omega)$. □

Corollary 1.1.4. *Any symplectically orthonormal subset of a symplectic space can be extended to a symplectic basis.*

Proof. Let $(\mathcal{X}, \omega)$ be a symplectic space and let $\{u_1, \ldots, u_k, v_1, \ldots, v_k\}$ be any symplectically orthonormal set. If it is a symplectic basis then there is nothing to prove. Otherwise, let $W = \text{span}\{u_1, \ldots, u_k, v_1, \ldots, v_k\}$. The subspace W is symplectic. Therefore $W^{\perp_s}$ is symplectic and $\mathcal{X} = W \oplus W^{\perp_s}$ by Proposition 1.1.2. By Theorem 1.1.3, there exists a symplectic basis

$\{u_{k+1}, \ldots, u_n, v_{k+1}, \ldots, v_n\}$ of $W^{\perp_s}$. The set $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ is a symplectic basis of $\mathcal{X}$ extending the given symplectically orthonormal set. □

We say that two symplectic spaces $(\mathcal{X}, \omega)$ and $(\mathcal{Y}, \eta)$ are isomorphic if there exists a vector space isomorphism $\Phi : \mathcal{X} \to \mathcal{Y}$ such that $\eta(\Phi(u), \Phi(v)) = \omega(u, v)$ for all $u, v \in \mathcal{X}$. In the view of Theorem 1.1.3, we prove in the next theorem that two symplectic spaces of the same dimension are essentially the same.

Theorem 1.1.5. *Two symplectic spaces of the same dimension are isomorphic.*

Proof. Let $(\mathcal{X}, \omega)$ and $(\mathcal{Y}, \eta)$ be symplectic spaces of dimension $2n$. By Theorem 1.1.3, choose symplectic bases $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n, x_1, \ldots, x_n\}$ of $(\mathcal{X}, \omega)$ and $(\mathcal{Y}, \eta)$ respectively. Let $\Phi : \mathcal{X} \to \mathcal{Y}$ be the vector space isomorphism given by $\Phi(u_j) = w_j$ and $\Phi(v_j) = x_j$ for all $j = 1, \ldots, n$. By definition we have $\eta(\Phi(x), \Phi(y)) = \omega(x, y)$, for all $x, y \in \{u_1, \ldots, u_n, v_1, \ldots, v_n\}$. By using linearity of Φ and bilinearity of the symplectic forms we get $\eta(\Phi(u), \Phi(v)) = \omega(u, v)$ for all $u, v \in \mathcal{X}$. □

By Theorem 1.1.5 it thus suffices to study only one symplectic space of a given dimension. Let J be the $2n \times 2n$ skew-symmetric matrix defined as $\begin{pmatrix} O & I_n \\ -I_n & O \end{pmatrix}$, where I_n is the $n \times n$ identity matrix. Let $\langle \cdot, \cdot \rangle_s : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R}$ be the symplectic form on $\mathbb{R}^{2n}$ induced by J. More precisely,

$$\langle x, y \rangle_s = \langle x, Jy \rangle \qquad \text{for all } x, y \in \mathbb{R}^{2n}.$$

Here $\langle \cdot, \cdot \rangle$ is the Euclidean inner product on $\mathbb{R}^n$. Let $x, y \in \mathbb{R}^{2n}$ with components $x_1, \ldots, x_{2n}$ and $y_1, \ldots, y_{2n}$ respectively. We have

$$\langle x, y \rangle_s = \sum_{j=1}^{n} (x_j y_{n+j} - x_{n+j} y_j) = \sum_{j=1}^{n} \det \begin{pmatrix} x_j & x_{n+j} \\ y_j & y_{n+j} \end{pmatrix}. \tag{1.5}$$

The symplectic space $(\mathbb{R}^{2n}, \langle \cdot, \cdot \rangle_s)$ is called the standard symplectic space. We call the symplectic form $\langle \cdot, \cdot \rangle_s$ the symplectic inner product on $\mathbb{R}^{2n}$. The symplectic inner product is also known as the *standard symplectic form* [21]. The geometrical properties of symplectic spaces are quite different from that of inner product spaces. For example, the oriented area of a parallelogram in $\mathbb{R}^2$ made by two vectors (a, b) and (c, d) is given by $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. So by equation (1.5), the symplectic inner product of x, y is the sum of the oriented areas of the parallelograms in $\mathbb{R}^2$

made by the pair of vectors (x_j, x_{n+j}) and (y_j, y_{n+j}) for all $j = 1, \ldots, n$. The geometrical properties of inner product spaces such as length of a vector, angle between two vectors are missing in symplectic spaces. Despite these differences, we prove several results in symplectic spaces analogous to some well known results in inner product spaces. For our present work we only deal with the standard symplectic space in the latter part of the thesis.

1.2 The symplectic group and Williamson's theorem

1.2.1 Symplectic and orthosymplectic matrices

A $2n \times 2n$ matrix M is called a symplectic matrix if

$$M^T J M = J. \tag{1.6}$$

The matrix J is a symplectic matrix. By (1.6) we have $(\det M)^2 = 1$ which implies $\det M = \pm 1$. It turns out that the determinant of symplectic matrices is always equal to one, a non-trivial fact discussed in [23]. We denote by $Sp(2n)$ the set of $2n \times 2n$ symplectic matrices. The set $Sp(2n)$ forms a group under multiplication, it is closed under transpose, and is called the symplectic group. The symplectic group $Sp(2n)$ is precisely the set of $2n \times 2n$ matrices M that preserve the symplectic inner product on $\mathbb{R}^{2n}$, i.e.,

$$\langle Mx, My \rangle_s = \langle x, y \rangle_s \quad \text{for all } x, y \in \mathbb{R}^{2n}.$$

The symplectic group is analogous to the group of orthogonal matrices in the sense that orthogonal matrices preserve the Euclidean inner product. But unlike the set of orthogonal matrices, $Sp(2n)$ is not compact as it contains the matrices of the form $kI_n \oplus k^{-1} I_n$ where k is any positive integer. The symplectic group occurs naturally in quantum mechanics and optics [29]. In an n mode quantum continuous variable system we have self adjoint operators $x_1, \ldots, x_n$ and $p_1, \ldots, p_n$ on a suitable Hilbert space $\mathcal{H}$ called the position and momentum operators. These operators obey the Canonical Commutation Relations (CCR) given by

$$[x_j, p_k] = \imath \delta_{jk} \hbar, \qquad [x_j, x_k] = [p_j, p_k] = 0 \qquad \text{for } k, j = 1, \ldots, n$$

where $\hbar = h/2\pi$, h is the Planck's constant, $\delta_{jk} = 1$ if $j = k$, and 0 otherwise. Here $[p, q] = pq - qp$ for all operators p, q on the Hilbert space $\mathcal{H}$. Any linear *canonical transformation* $x_j \mapsto y_j$, $p_j \mapsto q_j$ such that $y_1, \ldots, y_n$ and $q_1, \ldots, q_n$ are functions of $x_1, \ldots, x_n, p_1, \ldots, p_n$ and satisfy the CCR can be specified by a symplectic matrix $S \in Sp(2n)$. See [23, 68].

Let $M \in Sp(2n)$. Suppose $u_1, \ldots u_n, v_1, \ldots, v_n$ are the $2n$ columns of M. The condition $M^T JM = J$ is equivalent to the set $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ being a symplectic basis. This gives a one to one correspondence between $Sp(2n)$ and the set of all symplectic bases of the standard symplectic space $(\mathbb{R}^{2n}, \langle \cdot, \cdot \rangle_s)$. Consider M in the following block matrix form

$$M = \begin{pmatrix} A & B \\ C & G \end{pmatrix}, \tag{1.7}$$

where A, B, C, G are $n \times n$ matrices. The condition $M^T JM = J$ is equivalent to the following conditions on the blocks:

$$A^T C = C^T A, \quad B^T G = G^T B, \quad A^T G - C^T B = I. \tag{1.8}$$

We call a matrix *orthosymplectic* if it is symplectic as well as orthogonal. Orthosymplectic matrices have a specific structure stated in the following proposition. See ([23], Sec. 4).

Proposition 1.2.1. *An element M of $Sp(2n)$ is orthogonal if and only if M is of the form*

$$M = \begin{pmatrix} U & V \\ -V & U \end{pmatrix},$$

where U, V are $n \times n$ real matrices such that $U + \imath V$ is a unitary matrix.

Proof. Consider the block decomposition of M given by equation (1.7). Suppose M is orthosymplectic. We have $MJ = JM$ which implies $A = G$ and $B = -C$. Therefore from (1.8) we get

$$A^T B = B^T A, \quad A^T A + B^T B = I, \tag{1.9}$$

where I is the $n \times n$ identity matrix. The matrix $A + \imath B$ is unitary precisely when the conditions in (1.9) are satisfied. By choosing $U = A$ and $V = B$ we get the desired form of M. Conversely, suppose U, V are $n \times n$ real matrices such that $U + \imath V$ is unitary. This implies that U and V satisfy the conditions

$$U^T V = V^T U, \quad U^T U + V^T V = I. \tag{1.10}$$

By (1.10) and (1.8) it directly follows that $\begin{pmatrix} U & V \\ -V & U \end{pmatrix}$ is orthosymplectic. $\square$

1.2.2 Doubly superstochastic matrices associated with symplectic matrices

An $n \times n$ matrix A with non-negative entries is said to be *doubly stochastic* if

$$\sum_{i=1}^{n} a_{ij} = 1 \qquad \text{for all } j = 1, \dots, n;$$

$$\sum_{j=1}^{n} a_{ij} = 1 \qquad \text{for all } i = 1, \dots, n.$$

An $n \times n$ matrix B is said to be *doubly superstochastic* if there exists an $n \times n$ doubly stochastic matrix A such that $b_{ij} \geq a_{ij}$ for all $i, j = 1, \dots, n$. Let $x = (x_1, \dots, x_n)$ be an element of $\mathbb{R}^n$. We denote by $x^{\uparrow} = (x_1^{\uparrow}, \dots, x_n^{\uparrow})$ the vector obtained by arranging the components of x in increasing order. Let $x, y \in \mathbb{R}^n$. We say x is *supermajorised* by y if

$$\sum_{j=1}^{k} x_j^{\uparrow} \geq \sum_{j=1}^{k} y_j^{\uparrow} \tag{1.11}$$

for $1 \leq k \leq n$, and is denoted by $x \prec^w y$. We say that x is *majorised* by y (or y *majorises* x) if equality holds in (1.11) for $k = n$, and is denoted by $x \prec y$. There is an intimate relationship between doubly stochastic matrices and majorisation, and between doubly superstochastic matrices and supermajorisation. A well known characterisation of doubly stochastic and doubly superstochastic matrices in the theory of majorisation is stated in the following result. See ([6], Sec. 2).

Theorem 1.2.2. *An $n \times n$ matrix B is doubly superstochastic if and only if $Bx \prec^w x$ for every vector $x \in \mathbb{R}^n$ with non-negative components. Further, B is doubly stochastic if and only if $Bx \prec x$ for every vector $x \in \mathbb{R}^n$.*

Let M be any element of $Sp(2n)$, and consider the block form of M given by the equation (1.7). Suppose $a_{ij}, b_{ij}, c_{ij}, g_{ij}$ are the ijth entries of the $n \times n$ matrices A, B, C, G respectively. Let $\widetilde{M}$ be the $n \times n$ matrix whose ijth entry is given by

$$\widetilde{m}_{ij} = \frac{1}{2}(a_{ij}^2 + b_{ij}^2 + c_{ij}^2 + g_{ij}^2). \tag{1.12}$$

The matrix $\widetilde{M}$ is doubly superstochastic as stated in the next result. See ([11], Theorem 6).

Theorem 1.2.3. *Let M be an element of $Sp(2n)$ and $\widetilde{M}$ be the $n \times n$ matrix associated with M by the rule (1.12). The matrix $\widetilde{M}$ is doubly superstochastic. Further, $\widetilde{M}$ is doubly stochastic if and only if M is orthogonal.*

1.2.3 Symplectic eigenvalues and symplectic eigenvector pairs

The following result is of fundamental importance in our present study, it is known as Williamson's theorem. There are multiple proofs of Williamson's theorem available in the literature. See [21, 72]. The proofs discussed here is based on [72]. We denote by $\mathbb{P}_n(\mathbb{R})$ the set of $n \times n$ real positive definite matrices.

Theorem 1.2.4. *For every $A \in \mathbb{P}_{2n}(\mathbb{R})$ there exists an $M \in Sp(2n)$ such that*

$$M^T A M = \begin{pmatrix} D & O \\ O & D \end{pmatrix}, \tag{1.13}$$

where D is an $n \times n$ positive diagonal matrix with diagonal entries $d_1(A) \leq \cdots \leq d_n(A)$.

Proof. Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. The matrix $A^{-1/2}JA^{-1/2}$ is a $2n \times 2n$ non-singular skew-symmetric matrix. Therefore we get a $2n \times 2n$ orthogonal matrix R and an $n \times n$ diagonal matrix D with positive diagonal entries $d_1(A) \leq \cdots \leq d_n(A)$ such that,

$$\begin{aligned} R^T A^{-1/2} J A^{-1/2} R &= \begin{pmatrix} O & D^{-1} \\ -D^{-1} & O \end{pmatrix} \\ &= \begin{pmatrix} D^{-1/2} & O \\ O & D^{-1/2} \end{pmatrix} J \begin{pmatrix} D^{-1/2} & O \\ O & D^{-1/2} \end{pmatrix}. \end{aligned}$$

This gives

$$\begin{pmatrix} D^{1/2} & O \\ O & D^{1/2} \end{pmatrix} R^T A^{-1/2} J A^{-1/2} R \begin{pmatrix} D^{1/2} & O \\ O & D^{1/2} \end{pmatrix} = J. \tag{1.14}$$

Choose $M = A^{-1/2} R \left(\begin{smallmatrix} D^{1/2} & O \\ O & D^{1/2} \end{smallmatrix}\right)$. The relation (1.14) implies that M is a symplectic matrix. One can easily verify that M and D satisfy (1.13). $\square$

The numbers $d_1(A), \ldots, d_n(A)$ are called the *symplectic eigenvalues* of A. See [36]. These numbers are the complete invariants of A under the action of the symplectic group $Sp(2n)$. By the spectral theorem we know that Hermitian matrices can be diagonalised by unitary matrices.

Williamson's theorem is analogous to the spectral theorem in the sense that positive definite matrices of even size can be diagonalised by symplectic matrices.

Symplectic eigenvalues appear in various applications such as classical and quantum mechanics [21], symplectic topology [37] and harmonic oscillator systems [1, 58]. See also [24, 45, 67]. Recently there has been a heightened interest in the study of symplectic eigenvalues by both physicists and mathematicians. A particular reason for this being their growing importance and applications in quantum information. See, for instance, [23, 62]. In quantum mechanics, an object of importance is the class of Gaussian states in a quantum system. Associated with every n mode quantum state is a $2n \times 2n$ positive definite matrix known as the covariance matrix of the state. Gaussian states are completely characterised by their covariance matrices. By Heisenberg uncertainty principle, a quantum state with the associated covariance matrix A is Gaussian if and only if

$$A - \frac{\imath}{2} J \geq 0. \tag{1.15}$$

The condition (1.15) is equivalent to $d_j(A) \geq \frac{1}{2}$ for all $j = 1, \ldots, n$ which follows from Williamson's theorem.

The class of Gaussian states in a quantum system is being widely studied. Of interest there are various *entropy* functions associated with Gaussian states that are useful in measurement of the degree of mixedness and entanglement of the Gaussian states. Let A be the covariance matrix associated with an n mode Gaussian state. The von-Neuman entropy of the Gaussian state can be expressed as

$$S(A) = \sum_{j=1}^{n} H(d_j). \tag{1.16}$$

Here $H : [\frac{1}{2}, \infty) \to \mathbb{R}$ is the smooth function given by

$$H(t) = \left(\frac{2t+1}{2}\right) \log_2 \left(\frac{2t+1}{2}\right) - \left(\frac{2t-1}{2}\right) \log_2 \left(\frac{2t-1}{2}\right) \text{ for } t > 1/2,$$

and $H(\frac{1}{2}) = 0$. Another useful entropy function is Rényi-α entropy for $\alpha > 1$ which can be expressed as

$$R_\alpha(A) = \sum_{j=1}^{n} G(d_j). \tag{1.17}$$

Here $G : [\frac{1}{2}, \infty) \to \mathbb{R}$ is the smooth function

$$G(t) = \left(\frac{1}{\alpha - 1}\right) \log_2 \left(\left(\frac{2t+1}{2}\right)^{\alpha} - \left(\frac{2t-1}{2}\right)^{\alpha} \right).$$

Both the entropies are smooth functions of the symplectic eigenvalues of the covariance matrices of Gaussian states. A detailed account of applications of symplectic eigenvalues in quantum mechanics is given in [68]. So it is useful and important to have a well developed theory for symplectic eigenvalues as we have for eigenvalues.

The following proposition is a direct consequence of Williamson's theorem.

Proposition 1.2.5. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. There exists a symplectic basis $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ of $\mathbb{R}^{2n}$ such that for each $i = 1, \ldots, n$*

$$Au_i = d_i(A)Jv_i, \; Av_i = -d_i(A)Ju_i. \tag{1.18}$$

Proof. Let M be a symplectic matrix that diagonalises A in the Williamson's theorem. Let $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ to be the set of columns of M. We know that the set $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ is a symplectic basis of $\mathbb{R}^{2n}$. We also have $M^{-T} = JMJ^T$. By (1.13) we get

$$AM = JMJ^T \begin{pmatrix} D & O \\ O & D \end{pmatrix}$$

which is equivalent to (1.18). ☐

We call a pair of vectors (u_j, v_j) satisfying the conditions (1.18) *symplectic eigenvector pair* of A corresponding to the symplectic eigenvalue $d_j(A)$, and *normalised* symplectic eigenvector pair if it also satisfies $\langle u_j, v_j \rangle_s = 1$. For any positive integer $m \leq n$, denote by $Sp(2n, 2m)$ the set of $2n \times 2m$ matrices $S = [u_1, \ldots, u_m, v_1, \ldots, v_m]$ such that $\{u_1, \ldots, u_m, v_1, \ldots, v_m\}$ is a symplectically orthonormal set. In particular, $Sp(2n, 2n) = Sp(2n)$. We denote by $Sp(2n, 2m, A)$ the subset of $Sp(2n, 2m)$ consisting of matrices $S = [u_1, \ldots, u_m, v_1, \ldots, v_m]$ such that (u_j, v_j) is a symplectic eigenvector pair of A corresponding to $d_j(A)$. We write $Sp(2n, A) = Sp(2n, 2n, A)$. A symplectic basis consisting of symplectic eigenvector pairs of A is called a *symplectic eigenbasis* corresponding to A.

Example 4. Let A be the $2n \times 2n$ matrix of the form $\left(\begin{smallmatrix} P & O \\ O & P \end{smallmatrix}\right)$ where P is an $n \times n$ positive definite matrix. Let U be an orthogonal matrix of size n such that $U^T PU$ is the diagonal matrix with diagonal entries $\alpha_1 \leq \ldots \leq \alpha_n$. The matrix $M = \left(\begin{smallmatrix} U & O \\ O & U \end{smallmatrix}\right)$ is an (ortho)symplectic matrix that diagonalises A as in (1.13). The symplectic eigenvalues of A are given by $\alpha_1, \ldots, \alpha_n$.

Example 5. Let A be of the form $\left(\begin{smallmatrix} D_1 & O \\ O & D_2 \end{smallmatrix}\right)$ where D_1, D_2 are diagonal matrices with positive diagonal entries $\alpha_1, \ldots, \alpha_n$, and $\beta_1, \ldots, \beta_n$ respectively. Let $M = A^{-1/2} J A^{1/4} J A^{1/4} J$. One can verify that M is a symplectic matrix and

$$M^T A M = \begin{pmatrix} D & O \\ O & D \end{pmatrix},$$

where D is the diagonal matrix with diagonal entries $\sqrt{\alpha_1 \beta_1}, \ldots, \sqrt{\alpha_n \beta_n}$, and hence these are the symplectic eigenvalues of A.

1.2.4 Preliminary results on symplectic eigenvalues and symplectic eigenvector pairs

The results of this subsection are based on Section 2 of [39].

Lemma 1.2.6. *Let $A \in \mathbb{P}_{2n}(\mathbb{R})$ and let d be a positive number. The following statements are equivalent.*

(i) d is a symplectic eigenvalue of A and (u, v) is a corresponding symplectic eigenvector pair.

(ii) $\pm d$ is an eigenvalue of $\imath JA$ and $u \mp \imath v$ is a corresponding eigenvector.

(iii) $\pm d$ is an eigenvalue of $\imath A^{1/2} J A^{1/2}$ and $A^{1/2} u \mp \imath A^{1/2} v$ is a corresponding eigenvector.

Proof. For any two matrices X and Y, the spectrum (accounting multiplicities) of XY is the same as that of YX. Hence d is an eigenvalue of $\imath JA$ if and only if it is so for the matrix $\imath A^{1/2} J A^{1/2}$. Further $\imath JA(u - \imath v) = d(u - \imath v)$ if and only if $\imath A^{1/2} J A^{1/2}(A^{1/2} u - \imath A^{1/2} v) = d(A^{1/2} u - \imath A^{1/2} v)$. Hence (ii) and (iii) are equivalent.

We next prove the equivalence of (i) and (ii). It is easy to see that

$$Au = dJv \text{ and } Av = -dJu$$

if and only if

$$(\imath JA)u = -d\imath v \text{ and } (\imath JA)\imath v = -du. \tag{1.19}$$

The two expressions in (1.19) can equivalently be written as

$$\imath JA(u - \imath v) = d(u - \imath v).$$

This proves the required equivalence. $\square$

Since $d_1, \ldots, d_n$ denote the symplectic eigenvalues arranged in increasing order, we usually denote any collection of symplectic eigenvalues by $\tilde{d}_1, \ldots, \tilde{d}_n$.

Proposition 1.2.7. *For A in $\mathbb{P}_{2n}(\mathbb{R})$, the set $\{\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{v}_1, \ldots, \tilde{v}_m\}$ is a symplectically orthogonal set of symplectic eigenvector pairs of A corresponding to the symplectic eigenvalues $\tilde{d}_1, \ldots, \tilde{d}_m$, respectively, if and only if $\{A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j : j = 1, \ldots, m\}$ is an orthogonal set of eigenvectors of $\imath A^{1/2} J A^{1/2}$ corresponding to the eigenvalues $\tilde{d}_1, \ldots, \tilde{d}_m$ respectively. Further, for each $j = 1, \ldots, k$*

$$\|A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j\|^2 = 2\tilde{d}_j \langle \tilde{u}_j, J\tilde{v}_j\rangle. \tag{1.20}$$

Proof. We know by Lemma 1.2.6 that $(\tilde{u}_j, \tilde{v}_j)$ is a symplectic eigenvector of A corresponding to $\tilde{d}_j$ if and only if $A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j$ is an eigenvector of $\imath A^{1/2} J A^{1/2}$ corresponding to $\tilde{d}_j$. For $j, k = 1, \ldots, m$, we have

$$\begin{aligned}
&\langle A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j, A^{1/2}\tilde{u}_k - \imath A^{1/2}\tilde{v}_k\rangle \\
&= \langle \tilde{u}_j - \imath\tilde{v}_j, A(\tilde{u}_k - \imath\tilde{v}_k)\rangle \\
&= \langle \tilde{u}_j, A\tilde{u}_k\rangle + \langle \tilde{v}_j, A\tilde{v}_k\rangle - \imath\langle \tilde{u}_j, A\tilde{v}_k\rangle + \imath\langle \tilde{v}_j, A\tilde{u}_k\rangle \\
&= \tilde{d}_k \left(\langle \tilde{u}_j, J\tilde{v}_k\rangle + \langle \tilde{u}_k, J\tilde{v}_j\rangle + \imath\langle \tilde{u}_j, J\tilde{u}_k\rangle + \imath\langle \tilde{v}_j, J\tilde{v}_k\rangle\right).
\end{aligned} \tag{1.21}$$

If $\{\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{v}_1, \ldots, \tilde{v}_m\}$ is symplectically orthogonal, then from (1.21) we see that the set $\{A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j : j = 1, \ldots, m\}$ is orthogonal. When $j = k$ in (1.21), we get (1.20). It is easy to see that $\tilde{d}_j$ is an eigenvalue of $\imath A^{1/2} J A^{1/2}$ with corresponding eigenvector $A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j$ if and only if $-\tilde{d}_j$ is its eigenvalue with corresponding eigenvector $A^{1/2}\tilde{u}_j + \imath A^{1/2}\tilde{v}_j$. So, in a way similar to as in (1.21), we can have

$$\begin{aligned}
&\langle A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j, A^{1/2}\tilde{u}_k + \imath A^{1/2}\tilde{v}_k\rangle \\
&= \tilde{d}_k \left(\langle \tilde{u}_j, J\tilde{v}_k\rangle - \langle \tilde{u}_k, J\tilde{v}_j\rangle - \imath\langle \tilde{u}_j, J\tilde{u}_k\rangle + \imath\langle \tilde{v}_j, J\tilde{v}_k\rangle\right).
\end{aligned} \tag{1.22}$$

Let $\{A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j : j = 1, \ldots, m\}$ be orthogonal. Since eigenvectors corresponding to distinct eigenvalues of a Hermitian matrix are orthogonal, $A^{1/2}\tilde{u}_j - \imath A^{1/2}\tilde{v}_j$ and $A^{1/2}\tilde{u}_k + \imath A^{1/2}\tilde{v}_k$ are orthogonal for all $j, k = 1, \ldots, m$. Hence from (1.21) and (1.22), we get the symplectic orthogonality of the set $\{\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{v}_1, \ldots, \tilde{v}_m\}$. □

Corollary 1.2.8. *Any two symplectic eigenvector pairs corresponding to two distinct symplectic eigenvalues of a positive definite matrix are symplectically orthogonal.*

Proof. It follows immediately from the above result using the fact that eigenvectors corresponding to distinct eigenvalues of a Hermitian matrix are orthogonal. □

We know that any orthogonal (orthonormal) set of eigenvectors of a Hermitian matrix can be extended to an orthogonal (orthonormal) eigenbasis. The following proposition is a symplectic analogue of this fact.

Proposition 1.2.9. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. Every symplectically orthogonal set consisting of symplectic eigenvector pairs of A can be extended to a symplectically orthogonal basis consisting of symplectic eigenvector pairs of A.*

Proof. We know that any orthogonal subset of $\mathbb{C}^{2n}$ consisting of eigenvectors of a Hermitian matrix can be extended to an orthogonal basis of $\mathbb{C}^{2n}$. So the result directly follows from Proposition 1.2.7. □

Corollary 1.2.10. *Every symplectically orthonormal set consisting of symplectic eigevector pairs of A can be extended to a symplectic eigenbasis corresponding to A.*

Proof. Let $\{u_1, \ldots, u_m, v_1, \ldots, v_m\}$ be a symplectically orthonormal subset of $\mathbb{R}^{2n}$ consisting of m symplectic eigenvector pairs of A. By Proposition 1.2.9, extend the above set to a symplectically orthogonal set

$$\{u_1, \ldots, u_m, v_1, \ldots, v_m\} \cup \{\tilde{u}_{m+1}, \ldots, \tilde{u}_n, \tilde{v}_{m+1}, \ldots, \tilde{v}_n\}$$

consisting of n symplectic eigenvector pairs of A. By Proposition 1.2.7 we know that $\langle \tilde{u}_j, J\tilde{v}_j \rangle > 0$. Let $u_j = \langle \tilde{u}_j, J\tilde{v}_j \rangle^{-1/2} \tilde{u}_j$ and $v_j = \langle \tilde{u}_j, J\tilde{v}_j \rangle^{-1/2} \tilde{v}_j$ for all $j = m+1, m+2, \ldots, n$. The set $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$ is a symplectic eigenbasis corresponding to A. □

1.3 Symplectic projection

We are familiar with orthogonal projections in classical linear algebra. One can state the spectral theorem for Hermitian matrices in terms of orthogonal projections. We introduce a symplectic analogue of orthogonal projections and state Williamson's theorem in terms of symplectic projection. This section is based on our work in Section 5 of [39].

1.3.1 Symplectic projections associated with symplectically orthonormal sets

Let $S = \{x_1, \ldots, x_k, y_1, \ldots, y_k\}$ be a symplectically orthonormal subset of $\mathbb{R}^{2n}$. We introduce a map P_S on $\mathbb{R}^{2n}$ given by

$$P_S(x) = \sum_{i=1}^{k} \left(\langle x, Jx_i \rangle Jx_i + \langle x, Jy_i \rangle Jy_i \right). \tag{1.23}$$

Suppose M is the $2n \times 2k$ matrix

$$M = \Big[Jx_1, \ldots, Jx_k, Jy_1, \ldots, Jy_k \Big]. \tag{1.24}$$

We have

$$\begin{aligned} P_S(x) &= \sum_{i=1}^{k} \left(\langle x, Jx_i \rangle Jx_i + \langle x, Jy_i \rangle Jy_i \right) \\ &= \sum_{i=1}^{k} J \left((x^T Jx_i)x_i + (x^T Jy_i)y_i \right) \\ &= \sum_{i=1}^{k} J \left(x_i x_i^T J^T x + y_i y_i^T J^T x \right) \\ &= \sum_{i=1}^{k} \left((Jx_i)(Jx_i)^T + (Jy_i)(Jy_i)^T \right) x \\ &= MM^T x. \end{aligned}$$

Therefore P_S is a positive semidefinite matrix. We call P_S the *symplectic projection* associated with the set S. One can verify that the kernel of P_S is the symplectic complement of S. If $k = n$, i.e., S is a symplectic basis of $\mathbb{R}^{2n}$, then P_S is a positive definite symplectic matrix with all its symplectic eigenvalues 1. The symplectic projections associated with two symplectically orthonormal sets spanning the same space need not be equal. This can be seen by the following example.

Example 6. Let $S = \{(1,0)^T, (0,1)^T\}$ and $T = \{(1,0)^T, (1,1)^T\}$. The sets S and T are symplectically orthonormal and span $\mathbb{R}^2$. The symplectic projection P_S is the 2×2 identity matrix whereas the symplectic projection P_T is the matrix $\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.

In the following proposition we give a necessary and sufficient condition for the equality of two symplectic projections.

Proposition 1.3.1. *Let $S = \{u_1, \ldots, u_k, v_1, \ldots, v_k\}$ and $T = \{x_1, \ldots, x_m, y_1, \ldots, y_m\}$ be two symplectically orthonormal subsets of $\mathbb{R}^{2n}$, and let P and Q be the symplectic projections associated with them. Let M and N be the $2n \times 2k$ and $2n \times 2m$ matrices given by (1.24) corresponding to the sets S and T, respectively. Then $P = Q$ if and only if $k = m$ and $M = NU$ for some $2k \times 2k$ orthosymplectic matrix U.*

Proof. If $k = m$ and $M = NU$, the equality $P = Q$ easily follows from the orthogonality of U, and the fact that $P = MM^T$ and $Q = NN^T$.

Conversely, let $P = Q$. Clearly the subspaces spanned by S and T are the same, and hence $k = m$. By equation (1.23) we have

$$Px_j = \sum_{i=1}^{k} (\alpha_{ij} Ju_i + \beta_{ij} Jv_i)$$

for all $j = 1, \ldots, k$. Here $\alpha_{ij} = \langle x_j, Ju_i \rangle$ and $\beta_{ij} = \langle x_j, Jv_i \rangle$, $1 \leq i, j \leq k$. Since $P = Q$, $Px_i = Jy_i$. This gives

$$y_j = \sum_{i=1}^{k} (\alpha_{ij} u_i + \beta_{ij} v_i) \,. \tag{1.25}$$

Also since x_j belongs to the span of the symplectically orthonormal vectors $u_1, \ldots, u_k, v_1, \ldots, v_k$,

$$x_j = \sum_{i=1}^{k} (\langle x_j, Jv_i \rangle u_i - \langle x_j, Ju_i \rangle v_i) \tag{1.26}$$

$$= \sum_{i=1}^{k} (\beta_{ij} u_i - \alpha_{ij} v_i) \,. \tag{1.27}$$

Let X and Y be the $k \times k$ matrices $X = \begin{bmatrix} \alpha_{ij} \end{bmatrix}$ and $Y = \begin{bmatrix} \beta_{ij} \end{bmatrix}$, and U be the $2k \times 2k$ matrix

$$U = \begin{bmatrix} Y & X \\ -X & Y \end{bmatrix}^T .$$

Using the fact that $x_1, \ldots, x_k, y_1, \ldots, y_k$ are symplectically orthonormal, we can see that the columns of U are orthonormal as well as symplectically orthonormal vectors in $\mathbb{R}^{2k}$. Finally, from (1.25) and (1.27) we obtain $M = NU$. $\square$

1.3.2 Williamson's theorem through symplectic projection

Let H be an $m \times m$ Hermitian matrix and $\eta_1, \ldots, \eta_k$ be the distinct eigenvalues of H. Suppose $E_1, \ldots, E_k$ are the orthogonal projections onto the eigenspaces of H corresponding to the eigenvalues $\eta_1, \ldots, \eta_k$ respectively. The spectral decomposition of H is also given by

$$H = \sum_{j=1}^{k} \eta_j E_j .$$

In the same spirit, we give an alternate statement for Williamson's theorem which states that every element of $\mathbb{P}_{2n}(\mathbb{R})$ can be written as a linear combination of symplectic projections.

Proposition 1.3.2. *For every A in $\mathbb{P}_{2n}(\mathbb{R})$ there exist distinct positive numbers $\mu_1, \ldots, \mu_m$ and symplectic projections $P_1, \ldots, P_m$ that satisfy the following conditions.*

(i) $P_jJP_k = 0$ *for all* $j \neq k$, $j, k = 1, \ldots, m$.

(ii) $\sum_{k=1}^{m} P_kJP_k = J$.

(iii) $A = \sum_{k=1}^{m} \mu_k P_k$.

The numbers $\mu_1, \ldots, \mu_m$ and the symplectic projections $P_1, \ldots, P_m$ are uniquely determined by the above conditions. Further, for every $1 \leq j \leq m$, μ_j is a symplectic eigenvalue of A and P_j is the symplectic projection associated with a symplectically orthonormal set of eigenvector pairs of A corresponding to μ_j.

Proof. Let $\mu_1, \ldots, \mu_m$ be the distinct symplectic eigenvalues of A with multiplicities $k_1, \ldots, k_m$, respectively. For every $j = 1, \ldots, m$ let $S_j = \{u_{j,1}, \ldots, u_{j,k_j}, v_{j,1}, \ldots, v_{j,k_j}\}$ be a symplectically orthonormal set of symplectic eigenvector pairs of A corresponding to μ_j. Let P_j be the symplectic projection associated with S_j. By the definition of symplectic projections and Williamson's theorem, we can see that $\mu_1, \ldots, \mu_m$ and $P_1, \ldots, P_m$ satisfy (i)-(iii).

Now, let $\eta_1, \ldots, \eta_l$ be l distinct positive numbers and $Q_1, \ldots, Q_l$ be symplectic projections that also satisfy (i)-(iii). For every $j = 1, \ldots, l$, let $T_j = \{x_{j,1}, \ldots, x_{j,r_j}, y_{j,1}, \ldots, y_{j,r_j}\}$ be a symplectically orthonormal set corresponding to Q_j. By using (i) and (iii), we can see that each η_j is a symplectic eigenvalue of A, and $(x_{j,i}, y_{j,i})$, $1 \leq i \leq r_j$, are the symplectically orthonormal symplectic eigenvector pairs corresponding to η_j. Condition (ii) implies that $\{\eta_1, \ldots, \eta_l\}$ forms the set of all distinct symplectic eigenvalues of A. By the uniqueness of symplectic eigenvalues, we have $l = m$ and $\{\mu_1, \ldots, \mu_m\} = \{\eta_1, \ldots, \eta_l\}$. We can assume that $\mu_j = \eta_j$ for all $j = 1, \ldots, m$. By (iii) we see that r_j is equal to the multiplicity of μ_j. Since symplectic eigenvector pairs corresponding to different eigenvalues are symplectically orthogonal, S_j is symplectically orthogonal to T_k for all $j \neq k$. Consequently $P_jx = 0$ for all $x \in T_k$ and for all $k \neq j$. Thus for every $(x_{j,i}, y_{j,i})$ in T_j we have

$$\mu_j Q_j x_{j,i} = \mu_j J y_{j,i} = A x_{j,i} = \mu_j P_j x_{j,i}.$$

and since $\mu_j \neq 0$, $P_j x_{j,i} = Q_j x_{j,i}$. Similarly $P_j y_{j,i} = Q_j y_{j,i}$. Since $\cup T_j$ forms a basis for $\mathbb{R}^{2n}$, we get $P_j = Q_j$ for all $j = 1, \ldots, m$. $\square$

We can see that if $d_1(B), \ldots, d_n(B)$ are the symplectic eigenvalues of B and $\{u_1, \ldots, u_n, v_1, \ldots, v_n\}$

is a corresponding symplectic eigenbasis, then

$$B = \sum_{j=1}^{n} d_j(B) P_j,$$

where P_j is the symplectic projection corresponding to $\{u_j, v_j\}$.

Corollary 1.3.3. *Let $A \in \mathbb{P}_{2n}(\mathbb{R})$, and let d be its symplectic eigenvalue with multiplicity k. Let $S = \{w_1, \dots, w_k, z_1, \dots, z_k\}$ be a symplectically orthonormal set of symplectic eigenvector pairs of A corresponding to d. Then the set $T = \{x_1, \dots, x_k, y_1, \dots, y_k\}$ is also a symplectically orthonormal set of symplectic eigenvector pairs corresponding to d if and only if there exists a $2k \times 2k$ orthosymplectic matrix U such that*

$$N = MU,$$

where M and N are $2n \times 2m$ matrices with columns $w_1, \dots, w_k, z_1, \dots, z_k$ and $x_1, \dots, x_k, y_1, \dots, y_k$, respectively.

Proof. Let P and Q be the symplectic projections associated with the symplectically orthonormal sets S and T respectively. By the uniqueness of symplectic projections in Proposition 1.3.2 we have $P = Q$. Therefore by Proposition 1.3.1 we get a $2k \times 2k$ orthosymplectic matrix U such that

$$[Jx_1, \dots, Jx_k, Jy_1, \dots, Jy_k] = [Jw_1, \dots, Jw_k, Jz_1, \dots, Jz_k]U,$$

which is the same as $N = MU$. □

1.3.3 An extension of Williamson's theorem

The proof of Proposition 1.3.2 is based mainly on the fact that there is a symplectically orthonormal set $S_j = \{u_{j,1}, \dots, u_{j,k_j}, v_{j,1}, \dots, v_{j,k_j}\}$ associated with each distinct symplectic eigenvalue μ_j of multiplicity k_j. The arguments in the proof hold even if we assume one of the symplectic eigenvalues $\mu_j = 0$. In that case A is a positive semidefinite matrix whose kernel is the symplectic subspace spanned by S_j. This observation leads to an extension of Williamson's theorem to $2n \times 2n$ positive semidefinite matrices whose kernel is a symplectic subspace of $\mathbb{R}^{2n}$. A statement of the extended result is given in ([39], Remark 2.6) without proof. We set up some notations and definitions for convenience.

Let Z be any matrix with m columns and let $\mathcal{J}$ any subset of the index set $\{1, \dots, m\}$. We denote by $Z_{\mathcal{J}}$ the submatrix of Z obtained by removing those columns of Z with indices not in

$\mathcal{J}$. Let $m_1, \ldots, m_k$ be positive integers with $m_1 + \ldots + m_k = n$. For all $j = 1, \ldots, k$ define

$$\mathcal{I}_j = \{m_{j-1} + i : 1 \leq i \leq m_j\} \cup \{n + (m_{j-1} + i) : 1 \leq i \leq m_j\}$$

with $m_0 = 0$. The sets $\mathcal{I}_1, \ldots, \mathcal{I}_k$ form a partition of $\{1, \ldots, 2n\}$. Given a matrix S with $2n$ columns, we say that

$$S = S_{\mathcal{I}_1} \diamond \ldots \diamond S_{\mathcal{I}_k} \tag{1.28}$$

is the *symplectic column partition* of S of order $(m_1, \ldots, m_k)$.

Example 7. Let $n = 6$ and $m_1 = 2, m_2 = 3, m_3 = 1$. Then we have $\mathcal{I}_1 = \{1, 2\} \cup \{7, 8\}$, $\mathcal{I}_2 = \{3, 4, 5\} \cup \{9, 10, 11\}$, $\mathcal{I}_3 = \{6\} \cup \{12\}$.

Proposition 1.3.4. *Let S be a matrix with $2n$ columns and $m_1, \ldots, m_k$ be positive integers whose sum is n. Let $S = S_{\mathcal{I}_1} \diamond \ldots \diamond S_{\mathcal{I}_k}$ be the symplectic column partition of S of order $(m_1, \ldots, m_k)$. We have*

$$TS = TS_{\mathcal{I}_1} \diamond \ldots \diamond TS_{\mathcal{I}_k}, \tag{1.29}$$

where T is a matrix of appropriate size.

Proof. We know that the jth column of TS is given by Ts_j, where s_j is the jth column of S. Therefore we have $(TS)_{\mathcal{I}_j} = TS_{\mathcal{I}_j}$ for all j, and this implies that (1.29) holds. $\square$

We recall a symplectic version of matrix direct sum introduced by Bhatia and Jain in [11]. If A_j are $m_j \times m_j$ matrices for $j = 1, \ldots, k$ then $\oplus A_j$ is their usual direct sum. It is the $n \times n$ block-matrix with $A_1, \ldots, A_k$ on its diagonal and zeros elsewhere. Let

$$A_j = \begin{pmatrix} P_j & Q_j \\ R_j & S_j \end{pmatrix}, \qquad j = 1, \ldots, k$$

be $2m_j \times 2m_j$ block-matrices with each block of size $m_j \times m_j$. The s-direct sum of A_j is defined as the $2n \times 2n$ matrix

$$\oplus_s A_j = \begin{pmatrix} \oplus P_j & \oplus Q_j \\ \oplus R_j & \oplus S_j \end{pmatrix}, \qquad j = 1, \ldots, k.$$

Theorem 1.3.5. *Let A be a $2n \times 2n$ positive semidefinite matrix. Then there exists a symplectic matrix M such that (1.13) holds for some $n \times n$ nonnegative diagonal matrix D if and only if* *the kernel of A is a symplectic subspace of $\mathbb{R}^{2n}$. If dim Ker $A = 2m$, then exactly m diagonal entries of D are zero. In this case, we call the nonnegative diagonal entries of D to be the symplectic eigenvalues of the positive semidefinite matrix A.*

Proof. The result is trivial for $m = 0, n$. So we assume $1 < m < n$. Suppose ker$A$ is a symplectic subspace of $\mathbb{R}^{2n}$ of dimension $2m$. By Proposition 1.1.2 we know that $\dim(\ker A)^{\perp_s} = 2(n-m)$. Let $\{x_1, \ldots, x_m, y_1, \ldots, y_m\}$ and $\{x_{m+1}, \ldots, x_n, y_{m+1}, \ldots, y_n\}$ be symplectic bases of ker$A$ and $(\ker A)^{\perp_s}$ respectively. Let S be the element of $Sp(2n)$ whose ith column is x_i and $(n+i)$th column is y_i for $i = 1, \ldots, n$. Let $S = \overline{S} \diamond \widetilde{S}$ be the symplectic column partition of S of order $(m, n-m)$. We have

$$\begin{aligned} S^T A S &= \overline{S}^T A \overline{S} \oplus_s \widetilde{S}^T A \widetilde{S} \\ &= O_{2m} \oplus_s \widetilde{S}^T A \widetilde{S}, \end{aligned}$$

where O_{2m} is the $2m \times 2m$ zero matrix. The columns of $\widetilde{S}$ form a symplectic basis of $(\ker A)^{\perp_s}$ and $\ker A \cap (\ker A)^{\perp_s} = \{0\}$. This implies $\widetilde{S}^T A \widetilde{S}$ is a $2(n-m) \times 2(n-m)$ positive definite matrix. By Williamson's theorem get $Q \in Sp(2(n-m))$ such that

$$Q^T \widetilde{S}^T A \widetilde{S} Q = \begin{pmatrix} \widetilde{D} & O \\ O & \widetilde{D} \end{pmatrix},$$

where $\widetilde{D}$ is the $(n-m) \times (n-m)$ diagonal matrix with diagonal entries $\alpha_1, \ldots, \alpha_{n-m}$. Let M be the $2n \times 2n$ matrix with symplectic column partition $M = \overline{S} \diamond \widetilde{S}Q$ of order $(m, n-m)$. The matrix M is symplectic and we have

$$\begin{aligned} M^T A M &= \overline{S}^T A \overline{S} \oplus_s Q^T \widetilde{S}^T A \widetilde{S} Q \\ &= O_{2m} \oplus_s \begin{pmatrix} \widetilde{D} & O \\ O & \widetilde{D} \end{pmatrix}. \end{aligned}$$

Choose D as the $n \times n$ diagonal matrix with diagonal entries $\underbrace{0, \ldots, 0}_{m \text{ times}}, \alpha_1, \ldots, \alpha_{n-m}$.

Conversely, suppose $M^T A M$ is a diagonal matrix given by (1.13), where D is a non-negative diagonal matrix with exactly m zero entries. Let $M = \overline{M} \diamond \widetilde{M}$ be the symplectic column partition of M of order $(m, m-n)$. The kernel of A is spanned by the columns of $\overline{M}$ which is a symplectic subspace of $\mathbb{R}^{2n}$ of dimension $2m$.

□

Chapter 2

Differentiability and analyticity of symplectic eigenvalues

This chapter is based on our work in the paper [39]. The main aim of the chapter is to investigate the differential properties of symplectic eigenvalues and symplectic eigenvector pair maps on $\mathbb{P}_{2n}(\mathbb{R})$, and on a curve in $\mathbb{P}_{2n}(\mathbb{R})$. We recall some preliminary definitions and results on differentiability of functions on real Banach spaces in Section 2.1. In particular, we discuss the matrix square root function and the Implicit Function Theorem. We review the theory of differentiability and analyticity of eigenvalues of Hermitian matrices in Section 2.2. The results in this section are useful later in the chapter. In Section 2.3, we prove that if $d_j(A)$ is a *simple* symplectic eigenvalue of A then d_j is smooth at A and there exists a smooth symplectic eigenvector pair map corresponding to d_j. We also compute the first order derivative expressions for these maps. In Section 2.4, we show that if $t \mapsto \mathbf{A}(t)$ is a real analytic curve in $\mathbb{P}_{2n}(\mathbb{R})$ over an open interval $\mathcal{J}$, then one can choose symplectic eigenvalues and symplectic eigenbasis real analytically over the interval $\mathcal{J}$. Further, the symplectic eigenvalue maps $t \mapsto d_j(A(t))$ are piecewise real analytic over any subinterval $[a, b]$ of $\mathcal{J}$. In Section 2.5, we give some applications of our analysis of symplectic eigenvalues. In particular, we prove a symplectic analogue of Lidskii's theorem that gives a majorisation inequality between the symplectic eigenvalues of two positive definite matrices and their sum.

We denote by $\mathbb{M}_n(\mathbb{C})$ the set of $n \times n$ complex matrices equipped with the usual inner product $\langle A, B\rangle = \operatorname{tr} A^*B$ for all $A, B \in \mathbb{M}_n(\mathbb{C})$. Let $\mathbb{H}_n(\mathbb{C})$ and $\mathbb{S}_n(\mathbb{R})$ be the real subspaces of $\mathbb{M}_n(\mathbb{C})$ consisting of Hermitian matrices and real symmetric matrices respectively.

2.1 Differentiability of functions on real Banach spaces

Let X, Y be real Banach spaces and U be an open subset of X. A function $\Psi : U \to Y$ is said to be differentiable at $a \in U$ if there exists $T \in \mathcal{B}(X,Y)$ such that

$$\lim_{\|h\| \to 0} \frac{\|\Psi(a+h) - \Psi(a) - T(h)\|}{\|h\|} = 0,$$

where $\mathcal{B}(X,Y)$ is the Banach space of bounded linear maps from X to Y and $\|\cdot\|$ is the norm in the appropriate Banach space. The linear map T is unique and it is called the derivative of Ψ at a. It is also known as the Fréchet derivative of Ψ at a and is usually denoted by $D\Psi(a)$. If Ψ is differentiable at every point of U, we say that it is differentiable on U. The map Ψ is said to be continuously differentiable at a if it is differentiable on an open neighbourhood $V \subset U$ of a and the map $D\Psi : V \to \mathcal{B}(X,Y)$ is continuous at a. The higher order derivatives are defined inductively. Suppose for some $p \geq 2$ and all $1 \leq k < p$, the k^{th} order derivative has been defined; the k^{th} order derivative of Ψ at a being an element $D^k\Psi(a) \in \mathcal{B}_k(X,Y)$ where $\mathcal{B}_k(X,Y)$ is the Banach space of bounded k-linear maps from the k-fold Cartesian product of X to Y. Here $D^1\Psi(a) = D\Psi(a)$ and $\mathcal{B}_1(X,Y) = \mathcal{B}(X,Y)$. The map Ψ is p^{th} order differentiable at a if it is $(p-1)^{\text{th}}$ order differentiable on an open neighbourhood of a and $D^{p-1}\Psi$ is differentiable at a. The p^{th} order derivative of Ψ at a is given by $D^p\Psi(a) = D(D^{p-1}\Psi)(a)$.

The map Ψ is said to be C^p or p-times continuously differentiable at a if it is p^{th} order differentiable on an open neighbourhood V of a and the map $D^p\Psi : V \to \mathcal{B}_p(X,Y)$ is continuous at a; it is said to be C^p on U if it is C^p at every point in U. If the map is C^p at a for all $p \geq 1$, it is said to be C^∞ at a; and if it is C^∞ at every point in U then it is said to be C^∞ on U. C^∞ maps are also known as infinitely differentiable or smooth maps.

Example 8 (Constant maps)**.** Let $\kappa : X \to Y$ be a constant map and a be any element of X. We have $\kappa(a+h) - \kappa(a) = 0$ for all $h \in X$. So, the constant map is differentiable at a and $D\kappa(a) = 0$.

Example 9 (Linear maps)**.** Let L be any element of $\mathcal{B}(X,Y)$ and a be any element of X. We have $L(a+h) - L(a) - L(h) = 0$ for all $h \in X$. Therefore L is differentiable at a and

$DL(a) = L$. Since DL is a constant map, we have $D^pL(a) = 0$ for all $p \geq 2$.

Example 10 (The matrix square map)**.** Let $\Phi : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_n(\mathbb{C})$ be the matrix square map and $A \in \mathbb{M}_n(\mathbb{C})$. For any $B \in \mathbb{M}_n(\mathbb{C})$ we have

$$\Phi(A+B) = (A+B)^2 = \Phi(A) + (AB+BA) + B^2.$$

Note that $B \mapsto AB+BA$ is a linear map and $\lim\limits_{B\to 0} \|B^2\|/\|B\| = 0$. Therefore Φ is differentiable at A and we have $D\Phi(A)(B) = AB+BA$.

Example 11 (The matrix inverse map)**.** Let $\Psi : GL_n(\mathbb{C}) \to \mathbb{M}_n(\mathbb{C})$ be the matrix inverse map, where $GL_n(\mathbb{C})$ is the complex general linear group. Let $A \in GL_n(\mathbb{C})$ and $B \in \mathbb{M}_n(\mathbb{C})$ such that $\|B\| < 1/\|A^{-1}\|$. We then have $A+B \in GL_n(\mathbb{C})$, and

$$\begin{aligned}\Psi(A+B) &= (A+B)^{-1}\\ &= A^{-1}(I+BA^{-1})^{-1}\\ &= A^{-1}\left(I - BA^{-1} + \sum_{k=2}^{\infty}(-1)^k(BA^{-1})^k\right).\end{aligned}$$

By using the geometric sum we get

$$\|\Psi(A+B) - \Psi(A) + A^{-1}BA^{-1}\| \leq \frac{\|A^{-1}\|}{(1-\|B\|\|A^{-1}\|)}\|B\|^2.$$

This gives

$$\lim_{\|B\|\to 0} \frac{\|\Psi(A+B) - \Psi(A) + A^{-1}BA^{-1}\|}{\|B\|} = 0.$$

Therefore, Ψ is differentiable and we have $D\Psi(A)(C) = -A^{-1}CA^{-1}$ for all $C \in \mathbb{M}_n(\mathbb{C})$.

2.1.1 The matrix square root function

The set $\mathbb{P}_{2n}(\mathbb{R})$ is open in $\mathbb{S}_{2n}(\mathbb{R})$. Let $\varrho : \mathbb{P}_{2n}(\mathbb{R}) \to \mathbb{S}_{2n}(\mathbb{R})$ be the matrix square root map and $A \in \mathbb{P}_{2n}(\mathbb{R})$. The map ϱ is infinitely differentiable on $\mathbb{P}_{2n}(\mathbb{R})$, and its first order derivative at A is given by

$$D\varrho(A)(H) = \int_0^{\infty} e^{-t\varrho(A)}He^{-t\varrho(A)}\mathrm{d}t \tag{2.1}$$

for all $H \in \mathbb{S}_{2n}(\mathbb{R})$. See Moral and Niclas ([22], Theorem 1.1). We know by Lemma 1.2.6 that the symplectic eigenvalues of A are eigenvalues of $\imath\varrho(A)J\varrho(A)$. Therefore the regularity properties of symplectic eigenvalues follow directly from the corresponding properties of eigenvalues.

But getting a closed form for the derivative expression (2.1) of ϱ is a non-trivial problem. So the computation of derivatives of symplectic eigenvalues demands a different approach to dealing with symplectic eigenvalues.

In the latter part of the chapter, we use the fact that if $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ is a real analytic curve on an open interval $\mathcal{J}$ then the composite map $\varrho \circ \mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ is also real analytic. Since we could not find an explicit proof of this in the literature, we include its proof in here for the sake of completeness. The following two results can be found in the appendix of [39].

Lemma 2.1.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces, and let $T : \mathcal{X}^k \to \mathcal{Y}$ be a bounded k-linear map. Suppose $\sum_{n=0}^{\infty} a_{jn}$ is an absolutely convergent series in $\mathcal{X}$ with sum a_j for all $j = 1, \ldots, k$. For each n, let $c_n = \sum_{j_1+\cdots+j_k=n} T(a_{1j_1}, \ldots, a_{kj_k})$. Then the series $\sum_{n=0}^{\infty} c_n$ is absolutely convergent in $\mathcal{Y}$ and has sum $T(a_1, \ldots, a_k)$.*

Proof. The absolute convergence of the series $\sum_{n=0}^{\infty} c_n$ follows from Merten's theorem for Cauchy products of series of real numbers. We shall prove that its sum is $T(a_1, \ldots, a_k)$ by induction on k. When $k = 1$, the statement directly follows from the boundedness and linearity of T. Assume that the result holds for k. Let $\sum_{n=0}^{\infty} a_{jn}$ $(1 \le j \le k)$ and $\sum_{n=0}^{\infty} b_n$ be absolutely convergent series in $\mathcal{X}$ such that $a_j = \sum_{n=0}^{\infty} a_{jn}$ and $b = \sum_{n=0}^{\infty} b_n$.

For each m, define the map $\tilde{T}_m$ from $\mathcal{X} \to \mathcal{Y}$ as

$$\tilde{T}_m(x) = \sum_{j_1+\cdots+j_k=m} T(a_{1j_1}, \ldots, a_{kj_k}, x).$$

It is easy to see that $\tilde{T}_m$ is linear and bounded with $\|\tilde{T}_m\| \le \|T\| \sum_{j_1+\cdots+j_k=m} \|a_{1j_1}\| \cdots \|a_{kj_k}\|$. Since each $\sum_{n=0}^{\infty} \|a_{jn}\|$ is convergent, by Merten's theorem for Cauchy products of series of real numbers, we see that $\sum_{m=0}^{\infty} \|\tilde{T}_m\|$ converges. Let $K = \sum_{m=0}^{\infty} \|\tilde{T}_m\|$. For each $j \ge 0$, let

$$x_j = \tilde{T}_j(b),$$

and

$$c_j = \sum_{l=0}^{j} \tilde{T}_{j-l}(b_l).$$

Clearly $c_j = \sum_{j_1+\cdots+j_k+l=j} T(a_{1j_1}, \ldots, a_{kj_k}, b_l)$. We need to show that $\sum_{j=0}^{\infty} c_j$ is convergent to

$T(a_1, \ldots, a_k, b)$. Let (X_n), (C_n) and (B_n) be the sequences of partial sums of the series $\sum_{j=0}^{\infty} x_j$, $\sum_{j=0}^{\infty} c_j$ and $\sum_{j=0}^{\infty} b_j$, respectively. By induction hypothesis, $\sum_{j=0}^{\infty} x_j$ is absolutely convergent and its sum equals $T(a_1, \ldots, a_j, b)$. Take $d_n = b - B_n$ and $E_n = \sum_{j=0}^{n} \tilde{T}_j(d_{n-j})$. We have

$$\begin{aligned}
C_n &= \sum_{j=0}^{n} \sum_{l=0}^{j} \tilde{T}_l(b_{j-l}) \\
&= \sum_{l=0}^{n} \sum_{j=l}^{n} \tilde{T}_l(b_{j-l}) \\
&= \sum_{l=0}^{n} \tilde{T}_l \left(\sum_{j=0}^{n-l} b_j \right) = \sum_{l=0}^{n} \tilde{T}_l(B_{n-l}) \\
&= \sum_{l=0}^{n} \tilde{T}_l(b) - \sum_{l=0}^{n} \tilde{T}_l(d_{n-l}) \\
&= X_n - E_n.
\end{aligned}$$

It suffices to show that $E_n \to 0$ as $n \to \infty$. Since $d_n \to 0$, we can find a positive number M such that $\|d_n\| \leq M$ for all $n \geq 0$. Given an $\epsilon > 0$, choose N in $\mathbb{N}$ such that for all $n \geq N$

$$\|d_n\| < \frac{\epsilon}{2(K+1)}$$

and

$$\sum_{j=n+1}^{\infty} \|\tilde{T}_j\| < \frac{\epsilon}{2M}.$$

Then for all $n > 2N$ we can write

$$\begin{aligned}
\|E_n\| &\leq \sum_{j=0}^{N} \|\tilde{T}_j\| \|d_{n-j}\| + \sum_{j=N+1}^{n} \|\tilde{T}_j\| \|d_{n-j}\| \\
&< \frac{\epsilon}{2(K+1)} \sum_{j=0}^{N} \|\tilde{T}_j\| + M \sum_{j=N+1}^{n} \|\tilde{T}_j\| \\
&< \frac{\epsilon}{2(K+1)} K + M \frac{\epsilon}{2M} \leq \epsilon.
\end{aligned}$$

This proves $\lim_{n\to\infty} C_n = \lim_{n\to\infty} X_n = T(a_1, \ldots, a_k, b)$. □

Proposition 2.1.2. *Let $\mathbf{A} : \mathfrak{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve on an open interval $\mathfrak{J}$ that is real analytic at $t_0 \in \mathfrak{J}$. Then the composite map $\varrho \circ \mathbf{A}$ is also real analytic at t_0.*

Proof. Without loss of generality, we can assume that $\mathfrak{J} = (-1, 1)$ and $t_0 = 0$. Since the curve $\mathbf{A}$ is real analytic at $t=0$, there exists an $r>0$ such that $\mathbf{A}(t)$ can be expressed as $\mathbf{A}(t)=\mathbf{A}(0)+\sum\limits_{j=1}^{\infty} C_j t^j$ for all $|t|<r$. Here $\sum\limits_{j=1}^{\infty} C_j t^j$ is absolutely convergent for $|t|<r$. Since each kth order derivative $D^k\varrho\circ\mathbf{A}(0)$ is k-linear and bounded, by Lemma 2.1.1 we have

$$\begin{aligned}D^k\varrho\circ\mathbf{A}(0)&(\mathbf{A}(t)-\mathbf{A}(0),\ldots,\mathbf{A}(t)-\mathbf{A}(0))\\&=\sum_{n=0}^{\infty}\sum_{j_1+\cdots+j_k=n} t^n D^k\varrho\circ\mathbf{A}(0)(C_{j_1},\ldots,C_{j_k}).\end{aligned}$$

Let $B_{k,n}$ denote the matrix $\sum\limits_{j_1+\cdots+j_k=n} D^k\varrho\circ\mathbf{A}(0)(C_{j_1},\cdots,C_{j_k})$ for $k\leq n$, the zero matrix otherwise. We have the Taylor expansion of ϱ at $\mathbf{A}(0)$ in a neighbourhood $U\subseteq\mathbb{P}_{2n}(\mathbb{R})$ of $\mathbf{A}(0)$ [22].

$$\varrho(A)=\varrho(\mathbf{A}(0))+\sum_{k=1}^{\infty}\frac{1}{k!}D^k\varrho(\mathbf{A}(0))(A-\mathbf{A}(0),\cdots,A-\mathbf{A}(0)).$$

We use the same notation ϱ to denote the square root function on positive real numbers for convenience. Suppose λ_0 is the minimum eigenvalue of $\mathbf{A}(0)$. Since $\lambda_0>0$, the square root function ϱ is real analytic at λ_0, i.e., there exists an $r_0>0$ such that the series $\sum\limits_{k=1}^{\infty}\frac{1}{k!}\varrho^{(k)}(\lambda_0)(t-\lambda_0)^k$ is absolutely and locally uniformly convergent in $(\lambda_0-r_0,\lambda_0+r_0)$. Choose δ, $0<\delta<r$ such that $\sum\limits_{j=1}^{\infty}\|C_j\|\delta^j<r_0$ and $\mathbf{A}(t)\in U$ for all $t\in(-\delta,\delta)$. Thus for all $|t|<\delta$,

$$\varrho(\mathbf{A}(t))=\varrho(\mathbf{A}(0))+\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{n=k}^{\infty}B_{k,n}t^n. \tag{2.2}$$

We show that the iterated sum $\sum\limits_{k=1}^{\infty}\frac{1}{k!}\sum\limits_{n=k}^{\infty}\|B_{k,n}\||t|^n<\infty$. Let C be the sum $\sum\limits_{j=1}^{\infty}\|C_j\|\delta^j$. For $|t|<\delta$, we have

$$\begin{aligned}\sum_{n=k}^{\infty}\|B_{k,n}\||t|^n&\leq\sum_{n=k}^{\infty}\|B_{k,n}\|\delta^n\\&\leq\sum_{n=k}^{\infty}\sum_{j_1+\cdots+j_k=n}\|D^k\varrho\circ\mathbf{A}(0)(C_{j_1},\ldots,C_{j_k})\|\delta^n\\&\leq\|D^k\varrho\circ\mathbf{A}(0)\|\sum_{n=k}^{\infty}\sum_{j_1+\cdots+j_k=n}(\|C_{j_1}\|\delta^{j_1})\cdots(\|C_{j_k}\|\delta^{j_k})\\&=\|D^k\varrho\circ\mathbf{A}(0)\|C^k.\end{aligned}$$

The last equality follows from the convergence of Cauchy product of the series $\sum\limits_{j=1}^{\infty}\|C_j\|\delta^j$. By

[14] we have

$$\|D^k \varrho \circ \mathbf{A}(0)\| = \|\varrho^{(k)}(\mathbf{A}(0))\| = |\varrho^{(k)}(\lambda_0)|.$$

For $|t| < \delta$, we have $C < r_0$ and hence

$$\sum_{k=1}^{\infty} \frac{1}{k!} \sum_{n=k}^{\infty} \|B_{k,n}\| |t|^n \leq \sum_{k=1}^{\infty} \frac{1}{k!} |\varrho^{(k)}(\lambda_0)| C^k < \infty.$$

This implies that the iterated sum on the right hand side of (2.2) is equal to the sum $\sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{k!} B_{k,n} t^n$. This shows that $\sqrt{\mathbf{A}(t)}$ can be expressed as the power series

$$\sqrt{\mathbf{A}(t)} = \sqrt{\mathbf{A}(0)} + \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} \frac{1}{k!} B_{k,n} \right) t^n \text{ for all } |t| < \delta.$$

□

2.1.2 Implicit Function Theorem

We use the Implicit function Theorem for maps on real Banach spaces in the main result of Section 2.3. Let us recall the statement of the theorem. We refer the reader to [57] for a detailed account of differential calculus in real Banach spaces.

Let X, Y, Z be real Banach spaces and $O \subset X \times Y$ be an open set containing a point (x_0, y_0). Let $f : O \to Z$ be a C^p-map for some $p \geq 1$. We denote by (x, y) any element of $X \times Y$. The derivative of f at (x_0, y_0) is a linear map $Df(x_0, y_0) : X \times Y \to Z$. The *partial derivative* of f at (x_0, y_0) with respect to the first component x is the linear map $D_1 f(x_0, y_0) : X \to Z$ given by

$$D_1 f(x_0, y_0)(x) = Df(x_0, y_0)(x, 0),$$

and with respect to the second component y is the linear map $D_2 f(x_0, y_0) : Y \to Z$ given by

$$D_2 f(x_0, y_0)(y) = Df(x_0, y_0)(0, y)$$

for all $(x, y) \in X \times Y$. The Implicit Function Theorem states that if

(i) $f(x_0, y_0) = z_0$,

(ii) $D_2 f(x_0, y_0) : Y \to Z$ is an isomorphism, then there exists an open set $U \subset X$ containing x_0 and a C^p-map $\phi : U \to Y$ such that $\phi(x_0) = y_0$ and for all $x \in U$, $(x, \phi(x)) \in O$ and $f(x, \phi(x)) = z_0$. Further, $D_2 f(x, \phi(x)) : Y \to Z$ is an isomorphism for all $x \in U$ and the derivative of ϕ at x is given by

$$D\phi(x) = -D_2 f(x, \phi(x))^{-1} \circ D_2 f(x, \phi(x)).$$

2.2 Differentiability and analyticity of eigenvalues and eigenvectors of Hermitian matrices

Given $H \in \mathbb{H}_n(\mathbb{C})$, we denote by $\lambda_1(H) \leq \ldots \leq \lambda_n(H)$ the eigenvalues of H arranged in increasing order. This defines n eigenvalue maps $\lambda_1, \ldots, \lambda_n$ which are continuous on $\mathbb{H}_n(\mathbb{C})$. In fact, by the well known Weyl's Perturbation Theorem ([9], Corollary III.2.6), the eigenvalue maps are Lipschitz continuous on $\mathbb{H}_n(\mathbb{C})$. The following simple example of Rellich [64] shows that the eigenvalue maps λ_j are not differentiable in general.

Example 12. Define a map $\mathbf{H}(t) = \begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix}$ for $t \in (-1, 1)$. This is a C^∞ curve in $\mathbb{H}_2(\mathbb{C})$. The eigenvalues of $\mathbf{H}(t)$ in increasing order are given by $\lambda_1(\mathbf{H}(t)) = -|t|, \lambda_2(\mathbf{H}(t)) = |t|$. So the composite maps $\lambda_1 \circ \mathbf{H}$ and $\lambda_2 \circ \mathbf{H}$ are not differentiable at 0. This implies that both the eigenvalue maps λ_1 and λ_2 are non-differentiable at $\mathbf{H}(0)$.

Let $H \in \mathbb{H}_n(\mathbb{C})$ and λ be an eigenvalue of H. The cardinality of the set $\{j : \lambda_j(H) = \lambda, 1 \leq j \leq n\}$ is called the *multiplicity* of the eigenvalue λ. An eigenvalue with multiplicity one is called a simple eigenvalue. The multiplicity of eigenvalues plays an important role in their differentiability. If $\lambda_j(H)$ is a simple eigenvalue H, then λ_j is infinitely differentiable at H. A proof of this result using complex function theory is given by Kato ([43], Ch.II, Theorem 5.16). Furthermore, Magnus ([51], Theorem 2) also proved the existence of infinitely differentiable eigenvector maps corresponding to simple eigenvalues of Hermitian matrices, and derived the derivative expressions for simple eigenvalues and the corresponding eigenvector maps. We state this result in the following theorem. The Implicit Function Theorem is used in proving the existence of the smooth eigenvalue and eigenvector maps.

Theorem 2.2.1. *Let $H \in \mathbb{H}_n(\mathbb{C})$ and $\lambda_j(H)$ be a simple eigenvalue for some $j = 1, \ldots, n$. Let $x_0 \in \mathbb{C}^n$ be a unit eigenvector of H corresponding to $\lambda_j(H)$. There exists a neighbourhood*

$\mathcal{U} \subseteq \mathbb{H}_n(\mathbb{C})$ of H and a map $x : \mathcal{U} \to \mathbb{C}^n$ such that for every $A \in \mathcal{U}$ and $B \in \mathbb{H}_n(\mathbb{C})$ we have

(i) *λ_j and x are C^∞ on $\mathcal{U}$ with $x(H) = x_0$;*

(ii) *the eigenvalue $\lambda_j(A)$ is simple and $x(A)$ is a corresponding unit eigenvector of A;*

(iii) *the derivative of λ_j at H is given by*

$$D\lambda_j(H)(B) = \langle x(H), Bx(H)\rangle; \tag{2.3}$$

(iv) *the derivative of x at H is given by*

$$Dx(H)(B) = (\lambda_j(H)I - H)^\dagger Bx_0,$$

where I is the $n \times n$ identity matrix and $(\lambda_j(H)I - H)^\dagger$ is the Moore-Penrose inverse of $\lambda_j(H)I - H$.

Eigenvalues and eigenvectors of one parameter family of Hermitian matrices are widely studied in the literature and several interesting results have been obtained [34, 43, 44, 64, 73]. Let $\mathfrak{J}$ be an open interval in $\mathbb{R}$ and $\mathbf{H} : \mathfrak{J} \to \mathbb{H}_n(\mathbb{C})$ be a curve. We know that the eigenvalue maps obtained by arranging eigenvalues in increasing order are continuous functions of Hermitian matrices. Therefore, the maps given by $\lambda_j(t) = \lambda_j(\mathbf{H}(t))$ are continuous on $\mathfrak{J}$ for all $j = 1, \ldots, n$. We know by Example 12 that the eigenvalue curves $\lambda_j : \mathfrak{J} \to \mathbb{R}$ are not differentiable in general even though $\mathbf{H}$ is infinitely differentiable. But a suitable arrangement of the eigenvalues given by $\tilde{\lambda}_1(t) = t$ and $\tilde{\lambda}_2(t) = -t$ gives differentiable eigenvalue maps $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$, not necessarily in order. This suggests that fixing eigenvalues in increasing or decreasing order can compromise their differentiability properties. In the following theorem, we see the existence of eigenvalue maps that inherit the differentiability of curves in $\mathbb{H}_n(\mathbb{C})$. See Kato([43], pp.111-114).

Theorem 2.2.2. *Let $\mathfrak{J}$ be an open interval and $\mathbf{H} : \mathfrak{J} \to \mathbb{H}_n(\mathbb{C})$ be a curve. If $\mathbf{H}$ is differentiable at $t_0 \in \mathfrak{J}$, then there exist n real valued functions $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ on $\mathfrak{J}$ differentiable at t_0 such that $\tilde{\lambda}_1(t), \ldots, \tilde{\lambda}_n(t)$ are the n eigenvalues of $\mathbf{H}(t)$, not necessarily in increasing or decreasing order, for all $t \in \mathfrak{J}$. Further, if $\mathbf{H}$ is C^1 on $\mathfrak{J}$, then we can choose the aforementioned eigenvalue maps to be C^1 on $\mathfrak{J}$ as well.*

By Theorem 2.2.2, we get C^1 eigenvalue maps for C^1 curves in $\mathbb{H}_n(\mathbb{C})$. Also, for the C^∞ curve in Example 12, we get C^∞ eigenvalue maps by a suitable arrangement of the eigenvalues.

So, one would expect the existence of C^k eigenvalue maps for C^k curves in $\mathbb{H}_n(\mathbb{C})$ for all $1 \leq k \leq \infty$. But this is not true. The following example illustrates that there do not exist C^2 eigenvalue maps for C^∞ curves in $\mathbb{H}_n(\mathbb{C})$.

Example 13. Let $\mathbf{H} : (-1,1) \to \mathbb{H}_2(\mathbb{C})$ be the C^∞ curve given by

$$\mathbf{H}(t) = \begin{pmatrix} \sin(1/t)e^{-1/2t} & e^{-1/t} \\ e^{-1/t} & -\sin(1/t)e^{-1/2t} \end{pmatrix} \text{ for } t > 0, \qquad \mathbf{H}(t) = 0 \text{ for } t \leq 0.$$

The characteristic equation of $\mathbf{H}(t)$ is given by $\lambda^2 - f(t)$, where

$$f(t) = \sin^2(1/t)e^{-1/t} + e^{-2/t} \text{ for } t > 0, \qquad f(t) = 0 \text{ for } t \leq 0.$$

Therefore, the eigenvalues of $\mathbf{H}(t)$ are given by the square roots of $f(t)$. But the function f does not possess a C^2 square root on the interval as shown in ([4], Sec. 2).

The differentiability properties of eigenvectors of curves of Hermitian matrices are rather pathological in nature. The following example due to Rellich [64] shows that one cannot even guarantee continuous eigenvector functions for C^∞ curves in $\mathbb{H}_n(\mathbb{C})$, much less differentiable.

Example 14. Let $\mathbf{H} : (-1,1) \to \mathbb{H}_2(\mathbb{C})$ be given by

$$\mathbf{H}(t) = e^{-\frac{1}{t^2}} \begin{pmatrix} \cos\frac{2}{t} & \sin\frac{2}{t} \\ \sin\frac{2}{t} & -\cos\frac{2}{t} \end{pmatrix} \text{ for } t \neq 0, \qquad \mathbf{H}(0) = 0.$$

The eigenvalues of $\mathbf{H}(t)$ are given by $e^{-\frac{1}{t^2}}$ and $-e^{-\frac{1}{t^2}}$ with corresponding eigenvectors $x(t) = \left(\cos\frac{1}{t}, \sin\frac{1}{t}\right)$ and $y(t) = \left(-\sin\frac{1}{t}, \cos\frac{1}{t}\right)$ respectively for $t \neq 0$. These eigenvalues are simple and hence any other eigenvectors are scalar multiples of one of the two vectors $x(t), y(t)$. The functions $\cos\frac{1}{t}$ and $\sin\frac{1}{t}$ oscillate near zero. By elementary real analysis arguments one can show that there does not exist any eigenvector curve for $\mathbf{H}$ that is continuous and does not vanish at $t = 0$.

Alekseevsky, et al. ([4], Theorem 7.6) showed that under an additional condition, both eigenvalues and eigenvectors can be chosen smoothly for smooth curves in $\mathbb{H}_n(\mathbb{C})$. We say that two functions f and g continuous at t_0 meet with infinite order if for every $p \in \mathbb{N}$ there exists a function h_p continuous at t_0 such that $f(t) - g(t) = t^p h_p(t)$.

Theorem 2.2.3. *Let $\mathfrak{J}$ be an open interval and $\mathbf{H} : \mathfrak{J} \to \mathbb{H}_n(\mathbb{C})$ be a smooth curve such that for all $1 \leq i \neq j \leq n$ either $\lambda_i(t) = \lambda_j(t)$ for all $t \in \mathfrak{J}$ or $\lambda_i(t)$ and $\lambda_j(t)$ do not meet with infinite* *order at any point in $\mathfrak{J}$. Then all the eigenvalues and corresponding eigenbasis can be chosen smoothly in t on $\mathfrak{J}$.*

Eigenvalues and eigenvectors exhibit nicer regularity properties for real analytic curves. If $\mathbf{H}$ is a curve in $\mathbb{H}_n(\mathbb{C})$ that is real analytic at a point, then there exist eigenvalue and eigenvector maps real analytic at the given point. See Kato ([43], Ch.II, Sec.6) and Rellich ([64], Ch.1, Sec.1, Theorem 1).

Theorem 2.2.4. *Let $\mathfrak{J}$ be an open interval and $t_0 \in \mathfrak{J}$. Suppose $\mathbf{H} : \mathfrak{J} \to \mathbb{H}_n(\mathbb{C})$ is a curve real analytic at t_0. If λ is an eigenvalue of $\mathbf{H}(t_0)$ with multiplicity m, then there exists an $\epsilon > 0$ so that we can find m eigenvalue functions $\lambda_1, \ldots, \lambda_m : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{R}$ and m corresponding orthonormal eigenvector functions $x_1, \ldots, x_m : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{C}^n$ that are real analytic at t_0. Also $\lambda_i(t_0) = \lambda$ for all $i = 1, \ldots, m$.*

The following result shows that eigenvalue and eigenvector maps can be chosen analytically over an interval for real analytic curves in $\mathbb{H}_n(\mathbb{C})$. A proof of this result is given in Kato ([43], Ch.VII, Theorem 3.9).

Theorem 2.2.5. *Let $\mathfrak{J}$ be an open interval and $\mathbf{H} : \mathfrak{J} \to \mathbb{H}_n(\mathbb{C})$ be a real analytic curve on $\mathfrak{J}$. Then there exist real analytic curves $\lambda_j : \mathfrak{J} \to \mathbb{R}$ and $x_j : \mathfrak{J} \to \mathbb{C}^n$ for $j = 1, \ldots, n$ such that $\lambda_1(t), \ldots, \lambda_n(t)$ are the eigenvalues of $\mathbf{H}(t)$ with a corresponding set of orthonormal eigenvectors $x_1(t), \ldots, x_n(t)$ for all $t \in \mathfrak{J}$.*

As an application of the analysis of eigenvalues and eigenvectors of analytic curves of Hermitian matrices, Kato ([43], Ch.II, Sec.6.5) gives an analytic proof of the well known Lidskii's theorem which states that for any two matrices $A, B \in \mathbb{H}_n(\mathbb{C})$ we have

$$\lambda^{\uparrow}(A + B) - \lambda^{\uparrow}(A) \prec \lambda^{\uparrow}(B). \tag{2.4}$$

Here $\lambda^{\uparrow}(A)$ denotes the n tuple whose j^{th} component is $\lambda_j(A)$.

2.3 Differentiability of simple symplectic eigenvalues

This section is based on our work in Section 3 of [39]. Given $A \in \mathbb{P}_{2n}(\mathbb{R})$ and a symplectic eigenvalue d of A, we say that m is the multiplicity of d if the set $\{j : d_j(A) = d, 1 \leq j \leq n\}$ has exactly m elements. A symplectic eigenvalue is called simple if its multiplicity is one. Recall that $d_1(A) \leq \ldots \leq d_n(A)$ denote the symplectic eigenvalues of A in increasing order. So we have the symplectic eigenvalue maps $d_j : \mathbb{P}_{2n}(\mathbb{R}) \to \mathbb{R}$ for $j = 1, \ldots, n$.

A result on symplectic eigenvalues analogous to the Weyl's perturbation theorem states that for $A, B \in \mathbb{P}_{2n}(\mathbb{R})$ we have

$$\max_{1\leq j\leq n} |d_j(A) - d_j(B)| \leq (\kappa(A)\kappa(B))^{1/2} \|A - B\|, \tag{2.5}$$

where $\kappa(A)$ is the condition number of A. Also, one can not replace $(\kappa(A)\kappa(B))^{1/2}$ in (2.5) with any constant independent of A, B. See [38]. Therefore the symplectic eigenvalue maps d_j are locally Lipschitz (but not Lipschitz), and hence continuous on $\mathbb{P}_{2n}(\mathbb{R})$. The following example is based on Example 12. It shows that d_j are not differentiable in general.

Example 15. Let B be the 4×4 matrix $B = I_2 \otimes \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. We have $d_1(I_4 + tB) = 1 - |t|$ and $d_2(I_4 + tB) = 1 + |t|$ for any $t \in (-1, 1)$. The modulus function is not differentiable at $t = 0$. So d_1 and d_2 are not differentiable at I_4.

The multiplicity of $d_j(A)$ determines the differentiability of d_j at A. We prove in this section that simple symplectic eigenvalues are infinitely differentiable. Observe that in Example 15 that the symplectic eigenvalues of I_4 are not simple.

2.3.1 Infinite differentiability

The following proposition is an easy consequence of continuity of symplectic eigenvalues that is useful later in the chapter.

Proposition 2.3.1. *Let A be a $2n \times 2n$ real positive definite matrix, and let d be a symplectic eigenvalue of A with multiplicity m. Let $r_0 = \min\{|d - \tilde{d}| : \tilde{d}$ is a symplectic eigenvalue of $A, \tilde{d} \neq d\}$. Then for any positive number $r < r_0$, there exists an open neighbourhood U of A in $\mathbb{P}_{2n}(\mathbb{R})$ such that every P in U has exactly m symplectic eigenvalues (counted with multiplicities) in $(d - r, d + r)$.*

Proof. Let $d_i(A) < d_{i+1}(A) = \cdots = d_{i+m}(A) < d_{i+m+1}(A)$ such that $d_{i+1}(A) = \cdots = d_{i+m}(A) = d$. By our choice of r we see that

$$d_i(A) < d - r < d + r < d_{i+m+1}(A).$$

Since each d_j is continuous, we can find an open neighbourhood U of A such that for every

$P \in U$,

$$d_{i+1}(P), \ldots, d_{i+m}(P) \in (d-r, d+r),$$

$$d_i(P) < d - r \text{ and } d_{i+m+1}(P) > d + r.$$

Thus for every $P \in U$, there are exactly m symplectic eigenvalues $d_{i+1}(P), \ldots, d_{i+m}(P)$ of P that are contained in $(d-r, d+r)$. The cases $d = d_1$ and $d = d_n$ can be proved in a similar way. □

In the following lemma, we use the Implicit Function Theorem and the proof idea of Theorem 2.2.1 by Magnus [51]. Here we prove that simple symplectic eigenvalues are smooth. Moreover, we prove the existence of smooth symplectic eigenvector pair maps corresponding to simple symplectic eigenvalues. Recall that $\mathbb{P}_{2n}(\mathbb{R})$ is an open subset of the real Banach space $\mathbb{S}_{2n}(\mathbb{R})$. We view $\mathbb{C}^{2n} \times \mathbb{C}$ as a real Banach space in the proof of the following lemma.

Lemma 2.3.2. *Let A be a $2n \times 2n$ real positive definite matrix. Suppose d_0 is a simple symplectic eigenvalue of A with corresponding normalised symplectic eigenvector pair (u_0, v_0). Then there exists an open subset U of $\mathbb{P}_{2n}(\mathbb{R})$ containing A, and C^∞ maps $d : U \to \mathbb{R}$ and $u, v : U \to \mathbb{R}^{2n}$ that satisfy the following conditions.*

(i) *For every $P \in U$, $d(P)$ is a simple symplectic eigenvalue of P with the corresponding normalised symplectic eigenvector pair $(u(P), v(P))$.*

(ii) *$d(A) = d_0$, $u(A) = u_0$ and $v(A) = v_0$.*

(iii)

$$\langle u_0, Ju(P)\rangle + \langle v_0, Jv(P)\rangle = 0. \tag{2.6}$$

Proof. Since d_0 is a simple symplectic eigenvalue of A with symplectic eigenvector pair (u_0, v_0), by Lemma 1.2.6, it is a simple eigenvalue of $\imath JA$ with eigenvector $x_0 = u_0 - \imath v_0$. Also $\langle x_0, Jx_0\rangle = -2\imath\langle u_0, Jv_0\rangle = -2\imath$. Define the map $\varphi : \mathbb{P}_{2n}(\mathbb{R}) \times \mathbb{C}^{2n} \times \mathbb{C} \to \mathbb{C}^{2n} \times \mathbb{C}$ as

$$\varphi(P, x, d) = ((\imath JP - d)x, \langle x_0, Jx\rangle + 2\imath)\,.$$

Clearly, φ is a C^∞ map and $\varphi(A, x_0, d_0) = 0$. Let $D_2\varphi$ denote the partial derivative of φ with respect to (x, d). Then

$$D_2\varphi(A, x_0, d_0) = \begin{pmatrix} \imath JA - d_0 & -x_0 \\ x_0^* J & 0 \end{pmatrix}.$$

Thus $\det\, D_2\varphi(A, x_0, d_0) = -\langle x_0, J(\imath JA - d_0)^{\mathrm{adj}} x_0\rangle$. For any $m \times m$ matrix X, X^{adj} denotes the adjoint of X. This is the $m \times m$ matrix with the ijth entry $(-1)^{i+j} X(j, i)$, where $X(j, i)$

is the (j, i) minor of X. Since d_0 is a simple eigenvalue of $\imath JA$, 0 is a simple eigenvalue of $\imath JA - d_0$. So we have $(\imath JA - d_0)^{\text{adj}} x_0 = c x_0$, where c is the product of all nonzero eigenvalues of $\imath JA - d_0$. This gives

$$\langle x_0, J(\imath JA - d_0)^{\text{adj}} x_0 \rangle = c \langle x_0, J x_0 \rangle = -2\imath c \neq 0.$$

By the Implicit Function Theorem, there exists an open subset U of $\mathbb{P}_{2n}(\mathbb{R})$ containing A, and C^∞ maps $d : U \to \mathbb{C}$ and $x : U \to \mathbb{C}^{2n}$ that satisfy $\imath JPx(P) = d(P)x(P)$, $\langle x_0, Jx(P)\rangle = -2\imath$, $x(A) = x_0$ and $d(A) = d_0$. Clearly $x(P) \neq 0$, and hence $d(P)$ is an eigenvalue of $\imath JP$. All the eigenvalues of $\imath JP$ are real. Hence $d(P)$ is real. Since $d_0 > 0$, we can assume that $d(P) > 0$ for all $P \in U$. By Lemma 1.2.6, we see that $d(P)$ is a symplectic eigenvalue of P for every $P \in U$. Also since $D_2\varphi(P, x(P), d(P))$ is invertible, $(\imath JP - d(P))^{\text{adj}} \neq 0$ and this implies that $d(P)$ has multiplicity 1. Let $x(P) = \tilde{u}(P) - \imath \tilde{v}(P)$ be the Cartesian decomposition of $x(P)$. By Lemma 1.2.6 we see that $(\tilde{u}(P), \tilde{v}(P))$ is a symplectic eigenvector pair of P corresponding to $d(P)$. Also, the maps $P \mapsto \tilde{u}(P)$ and $P \mapsto \tilde{v}(P)$ are C^∞ on U, and $\tilde{u}(A) = u_0$ and $\tilde{v}(A) = v_0$. We know that $\langle u_0, Jv_0\rangle = 1$. Hence we can assume that $\langle \tilde{u}(P), J\tilde{v}(P)\rangle > 0$ for all $P \in U$. This implies that the map $P \mapsto \langle \tilde{u}(P), J\tilde{v}(P)\rangle^{-1/2}$ is C^∞ on U. Define the maps $u, v : U \to \mathbb{R}^{2n}$ as

$$u(P) = \langle \tilde{u}(P), J\tilde{v}(P)\rangle^{-1/2} \tilde{u}(P)$$

and

$$v(P) = \langle \tilde{u}(P), J\tilde{v}(P)\rangle^{-1/2} \tilde{v}(P).$$

The maps u and v are C^∞ and $(u(P), v(P))$ forms a normalised symplectic eigenvector pair of P corresponding to $d(P)$. This shows the existence of infinitely differentiable maps d, u, v on U that satisfy (i) and (ii). Moreover, since the real part of $\langle x_0, Jx(P)\rangle$ is zero,

$$\langle u_0, Ju(P)\rangle + \langle v_0, Jv(P)\rangle = 0$$

This proves (iii). $\square$

Remark 2.3.3. By Proposition 1.2.7, d_0 is a simple symplectic eigenvalue of A if and only if it is a simple eigenvalue of $\imath A^{1/2} J A^{1/2}$. We also know that the matrix square root map is infinitely differentiable on $\mathbb{P}_{2n}(\mathbb{R})$. So we can obtain (i) and (ii) of Lemma 2.3.2 from the corresponding differentiability properties of simple eigenvalues given in (i), (ii) of Theorem 2.2.1. But we give an independent proof as (2.6) is required in the computation of the derivatives of symplectic eigenvector pair in Theorem 2.3.5.

Theorem 2.3.4. *Let $A \in \mathbb{P}_{2n}(\mathbb{R})$, and suppose that $d_j(A)$ is simple. Then there exists a neighbourhood U of A in $\mathbb{P}_{2n}(\mathbb{R})$ such that for every $P \in U$, $d_j(P)$ is simple and the map $P \mapsto d_j(P)$ is smooth on U. Further, if (u_0, v_0) is a normalised symplectic eigenvector pair of A*

corresponding to $d_j(A)$, then there exist smooth maps $u_j, v_j : U \to \mathbb{R}^{2n}$ such that for every P in U, $(u_j(P), v_j(P))$ is a normalised symplectic eigenvector pair of P corresponding to $d_j(P)$, $u_j(A) = u_0$ and $v_j(A) = v_0$, and $u_j(P), v_j(P)$ satisfy (2.6).

Proof. If $d_j(A)$ is a simple symplectic eigenvalue of A, then by Lemma 2.3.2, we can find an open neighbourhood V of A in $\mathbb{P}_{2n}(\mathbb{R})$, and C^∞ maps $d : V \to \mathbb{R}$ and $u, v : V \to \mathbb{R}^{2n}$ that satisfy (i)-(iii) of Lemma 2.3.2; i.e., $d(P)$ is a simple symplectic eigenvalue of P and $(u(P), v(P))$ is a corresponding normalised symplectic eigenvector pair such that $d(A) = d_j(A)$, $u(A) = u_0, v(A) = v_0$, and $u(P), v(P)$ satisfy (2.6). Let r be a positive number with $r < \min\{d_{j+1}(A) - d_j(A), d_j(A) - d_{j-1}(A)\}$. By the continuity of the map $P \mapsto d(P)$ and Proposition 2.3.1, we can assume that for every P in V, $d(P)$ is the only symplectic eigenvalue of P contained in $(d_j(A) - r, d_j(A) + r)$. We know that the map $P \mapsto d_j(P)$ is continuous. Hence there exists an open neighbourhood W of A such that $d_j(P) \in (d_j(A) - r, d_j(A) + r)$ for every P in W. But this implies that $d(P) = d_j(P)$ for every $P \in V \cap W$. Take $U = V \cap W$. Hence the map d_j is infinitely differentiable on U with the corresponding normalised symplectic eigenvector maps u, v that satisfy the required conditions. □

2.3.2 Computation of first order derivatives

Let d be a simple symplectic eigenvalue of A and (u, v) be a normalised symplectic eigenvector pair of A corresponding to d. We know by Proposition 1.2.1 and Corollary 1.3.3 that if (x, y) is any normalised symplectic eigenvector pair of A corresponding to d then there exist real numbers a, b with $a^2 + b^2 = 1$ such that

$$[x, y] = [u, v]\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Therefore we have

$$x = au - bv \text{ and } y = bu + av.$$

Using this observation, we now compute the first order derivative expressions for simple symplectic eigenvalues and a corresponding symplectic eigenvector pair maps.

Theorem 2.3.5. *Let $A \in \mathbb{P}_{2n}(\mathbb{R})$ be such that $d_j(A)$ is simple, and let (u_j, v_j) be a normalised symplectic eigenvector pair map through $(u_j(A), v_j(A))$ obtained from Theorem 2.3.4. Let $M \in Sp(2n, A)$ be fixed. Then the derivatives of d_j, u_j and v_j at A are given as follows: for all*

$B \in \mathbb{S}_{2n}(\mathbb{R})$ we have

$$Dd_j(A)(B) = \frac{\langle u_j(A), Bu_j(A)\rangle + \langle v_j(A), Bv_j(A)\rangle}{2}, \tag{2.7}$$

$$Du_j(A)(B) = M\hat{D}M^T Bu_j(A) + M\overline{D}JM^T Bv_j(A), \tag{2.8}$$

and

$$Dv_j(A)(B) = M\hat{D}M^T Bv_j(A) - M\overline{D}JM^T Bu_j(A), \tag{2.9}$$

where $\hat{D}$ and $\overline{D}$ are the $2n \times 2n$ diagonal matrices with respective diagonal entries given by

$$\left(\hat{D}\right)_{kk} = \begin{cases} \frac{d_k(A)}{d_j^2(A)-d_k^2(A)} & k \neq j,\ 1 \leq k \leq n \\ -\frac{1}{4d_j(A)} & k = j,\ 1 \leq k \leq n \\ \left(\hat{D}\right)_{ii} & k = n+i,\ 1 \leq i \leq n, \end{cases} \tag{2.10}$$

and

$$\left(\overline{D}\right)_{kk} = \begin{cases} \frac{d_j(A)}{d_j^2(A)-d_k^2(A)} & k \neq j,\ 1 \leq k \leq n \\ \frac{1}{4d_j(A)} & k = j,\ 1 \leq k \leq n \\ \left(\overline{D}\right)_{ii} & k = n+i,\ 1 \leq i \leq n. \end{cases} \tag{2.11}$$

Proof. Since $d_j(A)$ is simple, by Theorem 2.3.4, we know that the map d_j is infinitely differentiable at A. Since (u_j, v_j) is a normalised symplectic eigenvector pair map obtained from Theorem 2.3.4, we have

$$Pu_j(P) = d_j(P)Jv_j(P), \tag{2.12}$$

$$Pv_j(P) = -d_j(P)Ju_j(P), \tag{2.13}$$

$$\langle u_j(P), Jv_j(P)\rangle = 1, \tag{2.14}$$

$$\langle u_j(A), Ju_j(P)\rangle + \langle v_j(A), Jv_j(P)\rangle = 0. \tag{2.15}$$

Let $B \in \mathbb{S}(2n)$ be arbitrary. Differentiating (2.12) and (2.13) at A we have

$$Bu_j(A) + ADu_j(A)(B) = Dd_j(A)(B)Jv_j(A) + d_j(A)JDv_j(A)(B), \tag{2.16}$$

and

$$Bv_j(A) + ADv_j(A)(B) = -Dd_j(A)(B)Ju_j(A) - d_j(A)JDu_j(A)(B). \tag{2.17}$$

Taking the inner product of (2.16) with $u_j(A)$ and using the fact that $\langle u_j(A), Jv_j(A)\rangle = 1$, we get

$$\begin{aligned}\langle u_j(A), Bu_j(A)\rangle + \langle u_j(A), ADu_j(A)(B)\rangle \\ = Dd_j(A)(B) + \langle u_j(A), d_j(A)JDv_j(A)(B)\rangle. \end{aligned} \tag{2.18}$$

Since

$$\begin{aligned}\langle u_j(A), ADu_j(A)(B)\rangle &= \langle Au_j(A), Du_j(A)(B)\rangle \\ &= d_j(A)\langle Du_j(A)(B), Jv_j(A)\rangle,\end{aligned}$$

we can write (2.18) as

$$\begin{aligned}Dd_j(A)(B) &= \langle u_j(A), Bu_j(A)\rangle + d_j(A)\langle Du_j(A)(B), Jv_j(A)\rangle \\ &\quad -d_j(A)\langle u_j(A), JDv_j(A)(B)\rangle.\end{aligned} \tag{2.19}$$

Similarly, taking the inner product of (2.17) with $v_j(A)$, we get

$$\begin{aligned}Dd_j(A)(B) &= \langle v_j(A), Bv_j(A)\rangle - d_j(A)\langle Du_j(A)(B), Jv_j(A)\rangle \\ &\quad +d_j(A)\langle u_j(A), JDv_j(A)(B)\rangle.\end{aligned} \tag{2.20}$$

Adding (2.19) and (2.20) finally gives (2.7).

We next compute the derivatives $Du_j(A)$ and $Dv_j(A)$.

Let the columns of M be $\tilde{u}_1, \ldots, \tilde{u}_n, \tilde{v}_1, \ldots, \tilde{v}_n$. We know that the columns of M form a symplectic basis of $\mathbb{R}^{2n}$. We can express $Du_j(A)(B)$ and $Dv_j(A)(B)$ uniquely as

$$Du_j(A)(B) = \sum_{k=1}^n \alpha_k \tilde{u}_k + \sum_{k=1}^n \beta_k \tilde{v}_k$$

and

$$Dv_j(A)(B) = \sum_{k=1}^n \gamma_k \tilde{u}_j + \sum_{k=1}^n \delta_k \tilde{v}_k,$$

where $\alpha_k = \langle Du_j(A)(B), J\tilde{v}_k\rangle$, $\beta_k = -\langle Du_j(A)(B), J\tilde{u}_k\rangle$, $\gamma_k = \langle Dv_j(A)(B), J\tilde{v}_k\rangle$ and $\delta_k = -\langle Dv_j(A)(B), J\tilde{u}_k\rangle$ for all $k = 1, \ldots, n$. Since $d_j(A)$ is simple, we can assume that $\tilde{u}_j = au_j(A) - bv_j(A)$ and $\tilde{v}_j = bu_j(A) + av_j(A)$ for some $a, b \in \mathbb{R}$ with $a^2 + b^2 = 1$. Thus

$$\langle \tilde{u}_k, Jv_j(A)\rangle = \langle u_j(A), J\tilde{v}_k\rangle = \delta_{kj} a \tag{2.21}$$

and

$$\langle \tilde{u}_k, Ju_j(A)\rangle = \langle \tilde{v}_k, Jv_j(A)\rangle = \delta_{kj} b \tag{2.22}$$

for all $k = 1, \ldots, n$. Here $\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jk} = 1$ otherwise. Taking inner product of (2.16) with $\tilde{u}_k$ we get

$$\begin{aligned} &\langle \tilde{u}_k, Bu_j(A)\rangle + \langle \tilde{u}_k, ADu_j(A)(B)\rangle \\ &\quad = Dd_j(A)(B)\langle \tilde{u}_k, Jv_j(A)\rangle + d_j(A)\langle \tilde{u}_k, JDv_j(A)(B)\rangle. \end{aligned}$$

Using (2.21) and the values of α_k and δ_k, this reduces to

$$d_k(A)\alpha_k - d_j(A)\delta_k = aDd_j(A)(B)\delta_{kj} - \langle \tilde{u}_k, Bu_j(A)\rangle. \tag{2.23}$$

Similarly, taking inner products of (2.16) with $\tilde{v}_k$, and of (2.17) with $\tilde{u}_k$ and $\tilde{v}_k$, and using (2.21) and (2.22), we obtain the expressions

$$d_k(A)\beta_k + d_j(A)\gamma_k = bDd_j(A)(B)\delta_{kj} - \langle \tilde{v}_k, Bu_j(A)\rangle, \tag{2.24}$$

$$d_j(A)\beta_k + d_k(A)\gamma_k = -bDd_j(A)(B)\delta_{kj} - \langle \tilde{u}_k, Bv_j(A)\rangle, \tag{2.25}$$

$$-d_j(A)\alpha_k + d_k(A)\delta_k = aDd_j(A)(B)\delta_{kj} - \langle \tilde{v}_k, Bv_j(A)\rangle. \tag{2.26}$$

Thus for each $k = 1, \ldots, n$ we have a system of four linear equations in four unknowns $\alpha_k, \beta_k, \gamma_k$ and δ_k. When $k \neq j$, this system is

$$\begin{pmatrix} d_k(A) & 0 & 0 & -d_j(A) \\ 0 & d_k(A) & d_j(A) & 0 \\ 0 & d_j(A) & d_k(A) & 0 \\ -d_j(A) & 0 & 0 & d_k(A) \end{pmatrix} \begin{pmatrix} \alpha_k \\ \beta_k \\ \gamma_k \\ \delta_k \end{pmatrix} = - \begin{pmatrix} \langle \tilde{u}_k, Bu_j(A)\rangle \\ \langle \tilde{v}_k, Bu_j(A)\rangle \\ \langle \tilde{u}_k, Bv_j(A)\rangle \\ \langle \tilde{v}_k, Bv_j(A)\rangle \end{pmatrix}.$$

Here $d_j(A) \neq d_k(A)$ therefore the coefficient matrix above is invertible. Left multiplying by the inverse we get

$$\begin{pmatrix} \alpha_k \\ \beta_k \\ \gamma_k \\ \delta_k \end{pmatrix} = (d_j^2(A) - d_k^2(A))^{-1} \begin{pmatrix} d_k(A) & 0 & 0 & d_j(A) \\ 0 & d_k(A) & -d_j(A) & 0 \\ 0 & -d_j(A) & d_k(A) & 0 \\ d_j(A) & 0 & 0 & d_k(A) \end{pmatrix} \begin{pmatrix} \langle \tilde{u}_k, Bu_j(A)\rangle \\ \langle \tilde{v}_k, Bu_j(A)\rangle \\ \langle \tilde{u}_k, Bv_j(A)\rangle \\ \langle \tilde{v}_k, Bv_j(A)\rangle \end{pmatrix}.$$

The solution is thus given by the following equations

$$\alpha_k = \frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k(A)\langle \tilde{u}_k, Bu_j(A)\rangle + d_j(A)\langle \tilde{v}_k, Bv_j(A)\rangle \right), \tag{2.27}$$

$$\beta_k = \frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k(A)\langle \tilde{v}_k, Bu_j(A)\rangle - d_j(A)\langle \tilde{u}_k, Bv_j(A)\rangle \right), \tag{2.28}$$

$$\gamma_k = \frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k(A)\langle \tilde{u}_k, Bv_j(A)\rangle - d_j(A)\langle \tilde{v}_k, Bu_j(A)\rangle \right), \tag{2.29}$$

$$\delta_k = \frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k(A)\langle \tilde{v}_k, Bv_j(A)\rangle + d_j(A)\langle \tilde{u}_k, Bu_j(A)\rangle \right). \tag{2.30}$$

Now, for $k = j$ we have the following system

$$\begin{pmatrix} d_j(A) & 0 & 0 & -d_j(A) \\ 0 & d_j(A) & d_j(A) & 0 \\ 0 & d_j(A) & d_j(A) & 0 \\ -d_j(A) & 0 & 0 & d_j(A) \end{pmatrix} \begin{pmatrix} \alpha_j \\ \beta_j \\ \gamma_j \\ \delta_j \end{pmatrix} = - \begin{pmatrix} \langle \tilde{u}_j, Bu_j(A)\rangle - aDd_j(A)(B) \\ \langle \tilde{v}_j, Bu_j(A)\rangle - bDd_j(A)(B) \\ \langle \tilde{u}_j, Bv_j(A) + bDd_j(A)(B)\rangle \\ \langle \tilde{v}_j, Bv_j(A)\rangle - aDd_j(A)(B) \end{pmatrix}.$$

Using the expression for $Dd_j(A)(B)$, the fact that B is symmetric, and the relationship between $(\tilde{u}_j, \tilde{v}_j)$ and $(u_j(A), v_j(A))$ one can see that the solution to the above system exists and is given by

$$\alpha_j - \delta_j = \frac{1}{2d_j(A)} \left(\langle \tilde{v}_j(A), Bv_j(A)\rangle - \langle \tilde{u}_j(A), Bu_j(A)\rangle \right), \tag{2.31}$$

$$\beta_j + \gamma_j = \frac{-1}{2d_j(A)} (\langle \tilde{v}_j(A), Bu_j(A)\rangle) + \frac{-1}{2d_j(A)} (\langle \tilde{u}_j(A), Bv_j(A)\rangle). \tag{2.32}$$

Differentiating (2.14) and (2.15), respectively, gives

$$\langle Du_j(A)(B), Jv_j(A)\rangle + \langle u_j(A), JDv_j(A)(B)\rangle = 0$$

and

$$\langle u_j(A), JDu_j(A)(B)\rangle + \langle v_j(A), JDv_j(A)(B)\rangle = 0.$$

These in turn imply $\alpha_j + \delta_j = 0$ and $\beta_j - \gamma_j = 0$. Thus

$$\alpha_j = -\delta_j = \frac{1}{4d_j(A)} \left(\langle \tilde{v}_j(A), Bv_j(A)\rangle - \langle \tilde{u}_j(A), Bu_j(A)\rangle \right) \tag{2.33}$$

and

$$\beta_j = \gamma_j = \frac{-1}{4d_j(A)} (\langle \tilde{v}_j(A), Bu_j(A)\rangle) + \frac{-1}{4d_j(A)} (\langle \tilde{u}_j(A), Bv_j(A)\rangle). \tag{2.34}$$

Simplifying the above expressions we get for $k \neq j$,

$$\alpha_k = \frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k^2(A)\langle J\tilde{v}_k, A^{-1}Bu_j(A)\rangle + d_j(A)\langle J\tilde{v}_k, JBv_j(A)\rangle \right),$$

$$\beta_k = -\frac{1}{d_j^2(A) - d_k^2(A)} \left(d_k^2(A)\langle J\tilde{u}_k, A^{-1}Bu_j(A)\rangle + d_j(A)\langle J\tilde{u}_k, JBv_j(A)\rangle \right),$$

$$\alpha_j = -\frac{1}{4}\langle J\tilde{v}_j, A^{-1}Bu_j(A)\rangle + \frac{1}{4d_j(A)}\langle J\tilde{v}_j, JBv_j(A)\rangle,$$

$$\beta_j = \frac{1}{4}\langle J\tilde{u}_j, A^{-1}Bu_j(A)\rangle - \frac{1}{4d_j(A)}\langle J\tilde{u}_j, JBv_j(A)\rangle.$$

Let x be the $2n$ real vector with components $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$. Then we see that x can be written as

$$x = \hat{D}\tilde{D}M^{-1}A^{-1}Bu_j(A) + \overline{D}M^{-1}JBv_j(A),$$

where $\tilde{D}$ is the $2n \times 2n$ diagonal matrix with diagonal entries the symplectic eigenvalues of A, $d_1(A), \ldots, d_n(A), d_1(A), \ldots, d_n(A)$, and $\hat{D}$ and $\overline{D}$ are the diagonal matrices given by (2.10) and (2.11), respectively. Therefore

$$\begin{aligned} Du_j(A)(B) &= M\hat{D}\tilde{D}M^{-1}A^{-1}Bu_j(A) + M\overline{D}M^{-1}JBv_j(A) \\ &= M\hat{D}M^TBu_j(A) + M\overline{D}JM^TBv_j(A). \end{aligned}$$

The last equality follows from the fact that $M^TAM = \tilde{D}$ and $MJM^T = J$. This proves (2.8). Similar computations give (2.9). □

The derivative expressions for simple eigenvalues and simple symplectic eigenvalues are similar. The role of the eigenvector in (2.3) is played by a symplectic eigenvector pair in (2.7).

Remark 2.3.6. Let $A \in \mathbb{P}_{2n}(\mathbb{R})$, and let d, u, v be maps on a neighbourhood U of A such that $d(P)$ is a symplectic eigenvalue of P and $(u(P), v(P))$ is a pair of normalised symplectic eigenvectors. If d, u, v are differentiable at A, then by following the same steps used to prove (2.7), we can compute the derivative of d at A, even if $d(A)$ is not simple, as

$$Dd(A)(B) = \frac{1}{2}\left(\langle u(A), Bu(A)\rangle + \langle v(A), Bv(A)\rangle\right). \tag{2.35}$$

As a corollary to Theorem 2.3.5 we compute the first order derivative for simple symplectic eigenvalues of differentiable curves of positive definite matrices. For a given curve $\mathbf{A} : \mathfrak{J} \to \mathbb{P}_{2n}(\mathbb{R})$ we denote the symplectic eigenvalue $d_j(\mathbf{A}(t))$ by $d_j(t)$, $1 \leq j \leq n$ and $t \in \mathfrak{J}$.

Corollary 2.3.7. *Let $\mathfrak{J}$ be an open interval and $\mathbf{A} : \mathfrak{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve that is infinitely differentiable at $t_0 \in \mathfrak{J}$. Suppose that $d_j(t_0)$ is simple. Then there exists an open interval $\mathfrak{J}_0 \subset \mathfrak{J}$ containing t_0 such that the map d_j is infinitely differentiable on $\mathfrak{J}_0$. If (u_0, v_0) is a corresponding normalised symplectic eigenvector pair of $\mathbf{A}(t_0)$, then we can find an infinitely* *differentiable normalised symplectic eigenvector pair map (u_j, v_j) on $\mathfrak{J}_0$ corresponding to d_j such that $(u_j(t_0), v_j(t_0)) = (u_0, v_0)$, and $(u_j(t), v_j(t))$ satisfies*

$$\langle u_0, Ju_j(t)\rangle + \langle v_0, Jv_j(t)\rangle = 0$$

for all $t \in \mathfrak{J}_0$. Further, for any fixed $M \in Sp(2n, \mathbf{A}(t_0))$,

$$d_j'(t) = \frac{\langle u_j(t), \mathbf{A}'(t)u_j(t)\rangle + \langle u_j(t), \mathbf{A}'(t)u_j(t)\rangle}{2} \text{ for all } t \in \mathfrak{J}_0, \tag{2.36}$$

$$u_j'(t_0) = M\hat{D}M^T\mathbf{A}'(t_0)u_0 + M\overline{D}JM^T\mathbf{A}'(t_0)v_0, \tag{2.37}$$

and

$$v_j'(t_0) = M\hat{D}M^T\mathbf{A}'(0)v_0 - M\overline{D}JM^T\mathbf{A}'(t_0)u_0, \tag{2.38}$$

where $\hat{D}$ and $\overline{D}$ are the diagonal matrices associated with $\mathbf{A}(t_0)$ given by (2.10) and (2.11), respectively.

Theorem 2.3.8. *Following the notations of Corollary 2.3.7, the second order derivative of d_j at t_0 is given by*

$$\begin{aligned} d_j''(t_0) =&\frac{1}{2}\left(\langle u_0, \mathbf{A}''(t_0)u_0\rangle + \langle v_0, \mathbf{A}''(t_0)v_0\rangle\right) \\ &+ 2\langle \mathbf{A}'(t_0)u_0, M\overline{D}JM^T\mathbf{A}'(t_0)v_0\rangle \\ &+ \langle \mathbf{A}'(t_0)u_0, M\hat{D}M^T\mathbf{A}'(t_0)u_0\rangle + \langle \mathbf{A}'(t_0)v_0, M\hat{D}M^T\mathbf{A}'(t_0)v_0\rangle, \end{aligned} \tag{2.39}$$

where $\hat{D}$ and $\overline{D}$ are the diagonal matrices associated with $\mathbf{A}(t_0)$ given by (2.10) and (2.11), respectively.

Proof. By (2.36), we have

$$d_j'(t) = \frac{\langle u_j(t), \mathbf{A}'(t)u_j(t)\rangle + \langle v_j(t), \mathbf{A}'(t)v_j(t)\rangle}{2} \tag{2.40}$$

for every t in $\mathfrak{J}_0$. Differentiating (2.40) at $t = t_0$ and using the fact that $\mathbf{A}'(t_0)$ is real symmetric, we get

$$\begin{aligned} d_j''(t_0) \quad = \quad & \frac{1}{2}\left(\langle u_0, \mathbf{A}''(t_0)u_0\rangle + \langle v_0, \mathbf{A}''(t_0)u_0\rangle\right) \\ & + \langle u_j'(t_0), \mathbf{A}'(t_0)u_0\rangle + \langle v_j'(t_0), \mathbf{A}'(t_0)v_0\rangle. \end{aligned} \tag{2.41}$$

Using the expression (2.37) for the derivative $u_j'(t_0)$, we get

$$\begin{aligned} \langle u_j'(t_0), \mathbf{A}'(t_0)u_0\rangle =&\langle M\hat{D}M^T\mathbf{A}'(t_0)u_0, \mathbf{A}'(t_0)u_0\rangle \\ &+ \langle M\overline{D}JM^T\mathbf{A}'(t_0)v_0, \mathbf{A}'(t_0)u_0\rangle. \end{aligned} \tag{2.42}$$

Similarly using (2.38), we have

$$\begin{aligned}\langle v_j'(t_0), \mathbf{A}'(t_0)v_0\rangle =& \langle M\hat{D}M^T\mathbf{A}'(t_0)v_0, \mathbf{A}'(t_0)v_0\rangle \\ &- \langle M\overline{D}JM^T\mathbf{A}'(t_0)u_0, \mathbf{A}'(t_0)v_0\rangle.\end{aligned}$$

Since $\overline{D}J = J\overline{D}$, we have

$$\begin{aligned}\langle v_j'(t_0), \mathbf{A}'(t_0)v_0\rangle =& \langle M\hat{D}M^T\mathbf{A}'(t_0)v_0, \mathbf{A}'(t_0)v_0\rangle \\ &+ \langle M\overline{D}JM^T\mathbf{A}'(t_0)v_0, \mathbf{A}'(t_0)u_0\rangle. \end{aligned} \tag{2.43}$$

Using (2.42) and (2.43) in (2.41), we obtain (2.39). □

2.4 Symplectic eigenvalues of curves of positive definite matrices

This section is based on our work in Section 4 of [39]. In this section, we study the regularity properties of symplectic eigenvalues and symplectic eigenvector pairs for curves in $\mathbb{P}_{2n}(\mathbb{R})$. Here we use the results on eigenvalues and eigenvectors discussed in Section 2.2.

2.4.1 Differentiable curves in $\mathbb{P}_{2n}(\mathbb{R})$

The following result is a symplectic analogue of Theorem 2.2.2. We shall use $\tilde{d}_1, \ldots, \tilde{d}_n$ to represent symplectic eigenvalues in any order.

Theorem 2.4.1. *Let $\mathcal{J}$ be an open interval and $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve that is differentiable at $t_0 \in \mathcal{J}$. Then all the symplectic eigenvalues of $\mathbf{A}(t)$ can be chosen to be differentiable at t_0, i.e., we can find n functions $\tilde{d}_1, \ldots, \tilde{d}_n$ in a neighbourhood of t_0 that are differentiable at t_0 such that $\tilde{d}_1(t), \ldots, \tilde{d}_n(t)$ are the symplectic eigenvalues of $\mathbf{A}(t)$. If, in addition, $\mathbf{A}$ is C^1 on $\mathcal{J}$, then $\tilde{d}_1, \ldots, \tilde{d}_n$ can be chosen to be C^1 on $\mathcal{J}$.*

Proof. We know that the matrix square root map is C^∞ on $\mathbb{P}_{2n}(\mathbb{R})$. So the differentiability of $\mathbf{A}$ at $t_0 \in \mathcal{J}$ implies that $t \mapsto \imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ is also differentiable at t_0. By Theorem 2.2.2, we get functions $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2n} : \mathcal{J} \to \mathbb{R}$ differentiable at t_0 such that $\tilde{\lambda}_1(t), \ldots, \tilde{\lambda}_{2n}(t)$ are the eigenvalues of $\imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$. By reordering, suppose $\tilde{\lambda}_1(t_0), \ldots, \tilde{\lambda}_n(t_0)$ are the symplectic eigenvalues of $\mathbf{A}(t_0)$. Let $\mathcal{J}_0 \subset \mathcal{J}$ be an open subinterval containing t_0 such that for $t \in \mathcal{J}_0$, $\tilde{\lambda}_1(t), \ldots, \tilde{\lambda}_n(t)$ are all positive, and hence these are the symplectic eigenvalues of $\mathbf{A}(t)$. Define

$$\tilde{d}_j(t) = \begin{cases} \tilde{\lambda}_j(t) & t \in \mathcal{J}_0, \\ d_j(\mathbf{A}(t)) & t \notin \mathcal{J}_0. \end{cases}$$

Thus, $\tilde{d}_1, \ldots, \tilde{d}_n$ are the required symplectic eigenvalue maps on $\mathcal{J}$ differentiable at t_0.

If $\mathbf{A}$ is C^1 on $\mathcal{J}$ we again get by Theorem 2.2.2 eigenvalue functions $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2n} : \mathcal{J} \to \mathbb{R}$ that are C^1 on $\mathcal{J}$. We know that all the eigenvalues of $\imath\mathbf{A}(t)^{1/2} J \mathbf{A}(t)^{1/2}$ are non-zero for all $t \in \mathcal{J}$. In fact, there are equal number of positive and negative eigenvalues. Therefore we can assume, by reordering, that $\tilde{\lambda}_1(t), \ldots, \tilde{\lambda}_n(t)$ are positive for all $t \in \mathcal{J}$. Choose $\tilde{d}_j(t) = \tilde{\lambda}_j(t)$ for all $t \in \mathcal{J}$. Hence $\tilde{d}_1, \ldots, \tilde{d}_n$ are C^1 on $\mathcal{J}$ such that $\tilde{d}_1(t), \ldots, \tilde{d}_n(t)$ are the symplectic eigenvalues of $\mathbf{A}(t)$ for all $t \in \mathcal{J}$. $\square$

We see the existence of C^1 symplectic eigenvalue maps for C^1 curves in $\mathbb{P}_{2n}(\mathbb{R})$. But in general, there do not exist C^k symplectic eigenvalue maps corresponding to C^k curves in $\mathbb{P}_{2n}(\mathbb{R})$ for $k \geq 2$. Let $\mathbf{A} : (-1, 1) \to \mathbb{P}_4(\mathbb{R})$ be the C^∞ curve given by

$$\mathbf{A}(t) = I_2 \otimes (\alpha I_2 + \mathbf{H}(t)) \qquad \text{for all } t \in (-1, 1),$$

where $\mathbf{H}$ is the curve given in Example 13, and $\alpha > 0$ is such that the translated matrices $\alpha I_2 + \mathbf{H}(t)$ have positive eigenvalues for all $t \in (-1, 1)$. The symplectic eigenvalues of $\mathbf{A}(t)$ are given by $\alpha \pm \sqrt{f(t)}$, where $f(t) = \sin^2(1/t)e^{-1/t} + e^{-2/t}$ for $t > 0$ and $f(t) = 0$ for $t \leq 0$. So the existence of a C^2 symplectic eigenvalue curve corresponding to $\mathbf{A}$ would give a C^2 square root of f. But we know by Example 13 that f does not possess a C^2 square root.

Symplectic eigenvectors show the similar pathological behaviour as eigenvectors. We have a C^∞ curve in $\mathbb{P}_4(\mathbb{R})$ for which there does not exist even continuous selection of symplectic eigenvector pair maps. Let $\mathbf{B} : (-1, 1) \to \mathbb{P}_4(\mathbb{R})$ be the C^∞ curve given by

$$\mathbf{B}(t) = I_2 \otimes (I_2 + \mathbf{H}(t)) \qquad \text{for all } t \in (-1, 1)$$

where $\mathbf{H}$ is the curve given in Example 14. The symplectic eigenvalues of $\mathbf{B}(t)$ are given by $d_1(t) = 1 - e^{-1/t^2}$, $d_2(t) = 1 + e^{-1/t^2}$ for $t \neq 0$, and $d_1(0) = d_2(0) = 1$. Let $u_1(t) = e_1 \otimes [\cos\frac{1}{t}, \sin\frac{1}{t}]^T$, $v_1(t) = e_2 \otimes [\cos\frac{1}{t}, \sin\frac{1}{t}]^T$ and $u_2(t) = e_1 \otimes [\sin\frac{1}{t}, -\cos\frac{1}{t}]^T$, $v_2(t) = e_2 \otimes [\sin\frac{1}{t}, -\cos\frac{1}{t}]^T$ where $e_1 = [1, 0]^T$, $e_2 = [0, 1]^T$. One can easily verify that $(u_1(t), v_1(t))$and $(u_2(t), v_2(t))$ are normalised symplectic eigenvector pairs corresponding to $d_1(t)$ and $d_2(t)$ respectively. Suppose that there exist functions $\tilde{u}, \tilde{v} : (-1, 1) \to \mathbb{R}^4$ continuous at 0 such that $(\tilde{u}(t), \tilde{v}(t))$ is a normalised symplectic eigenvector pair of $\mathbf{B}(t)$ corresponding either to $d_1(t)$, or to $d_2(t)$. Therefore we can get a sequence $(t_j)_{j\in\mathbb{N}}$ of nonzero terms in $(-1,1)$ converging to 0 such that for all $j \in \mathbb{N}$, the pair $(\tilde{u}(t_j), \tilde{v}(t_j))$ corresponds either to $d_1(t_j)$ or to $d_2(t_j)$. Consider the case when $(\tilde{u}(t_j), \tilde{v}(t_j))$ corresponds to $d_1(t_j)$ for all j. For each j, $d_1(t_j)$ is a simple symplectic eigenvalue of $\mathbf{B}(t_j)$. This implies that the normalised symplectic eigenvector pair $(\tilde{u}(t_j), \tilde{v}(t_j))$ is of the form $\tilde{u}(t_j) = a_j u_1(t_j) - b_j v_1(t_j), \tilde{v}(t_j) = b_j u_1(t_j) + a_j v_1(t_j)$ where $a_j,\ b_j \in \mathbb{R}$ and $a_j^2 + b_j^2 = 1$. The continuity of $\tilde{u}$ and $\tilde{v}$ at $t = 0$ implies that the limits $\lim\limits_{j\to\infty} a_j \sin(1/t_j)$ and $\lim\limits_{j\to\infty} b_j \sin(1/t_j)$ exist, which in turn imply that $\lim\limits_{j\to\infty} \sin^2(1/t_j)$ exists. This is a contradiction.

We use Theorem 2.2.3 to show that under some additional conditions, symplectic eigenvalues and corresponding symplectic eigenvector pairs can be chosen smoothly for smooth curves in $\mathbb{P}_{2n}(\mathbb{R})$. Recall that two functions f and g continuous at t_0 meet with infinite order if for every $p \in \mathbb{N}$ there exists a function h_p continuous at t_0 such that $f(t) - g(t) = t^p h_p(t)$.

Theorem 2.4.2. *Let $\mathcal{J}$ be an open interval and $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a smooth curve such that for all $1 \le i \ne j \le n$ either $d_i(t) = d_j(t)$ for all $t \in \mathcal{J}$ or $d_i(t)$ and $d_j(t)$ do not meet with infinite order at any point in $\mathcal{J}$. Then all the symplectic eigenvalues and corresponding symplectic eigenbasis can be chosen smoothly in t on $\mathcal{J}$.*

Proof. By the smoothness of the matrix square root map, we know that $t \mapsto \imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ is also smooth on $\mathcal{J}$. The eigenvalues of $\imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ are $\pm d_j(t)$ for $j = 1,\ldots,n$. Therefore, no two of the ordered eigenvalues meet with infinite order at any point in $\mathcal{J}$ unless they are equal on whole $\mathcal{J}$. By Theorem 2.2.3 we get smooth maps $\tilde{\lambda}_j : \mathcal{J} \to \mathbb{R}$ and $u_j, v_j : \mathcal{J} \to \mathbb{R}^{2n}$ for all $j = 1,\ldots,2n$ such that $\tilde{\lambda}_1(t),\ldots,\tilde{\lambda}_{2n}(t)$ are the eigenvalues of $\imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ with corresponding orthonormal eigenvectors $u_1(t) - \imath v_1(t),\ldots,u_{2n}(t) - \imath v_{2n}(t)$ for all $t \in \mathcal{J}$. The eigenvalues of $\imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ are all non-zero for every $t \in \mathcal{J}$. Therefore, by continuity and reordering, we can assume that $\tilde{\lambda}_1(t),\ldots,\tilde{\lambda}_n(t)$ are the positive eigenvalues for all $t \in \mathcal{J}$. Choose $\tilde{d}_j = \tilde{\lambda}_j$, and define

$$\tilde{u}_j(t) = \sqrt{2\tilde{d}_j(t)}\mathbf{A}(t)^{-1/2}u_j(t),$$
$$\tilde{v}_j(t) = \sqrt{2\tilde{d}_j(t)}\mathbf{A}(t)^{-1/2}v_j(t)$$

for $j = 1,\ldots,n$ and $t \in \mathcal{J}$. By Proposition 1.2.7 we know that $(\tilde{u}_j(t), \tilde{v}_j(t))$ is a normalised pair of symplectic eigenvectors of $\mathbf{A}(t)$ corresponding to $\tilde{d}_j(t)$, and $\{\tilde{u}_1(t),\ldots,\tilde{u}_n(t),\tilde{v}_1(t),\ldots,\tilde{v}_n(t)\}$ is a symplectic basis of $\mathbb{R}^{2n}$ for all $t \in \mathcal{J}$. Also, the maps $\tilde{d}_j$, $\tilde{u}_j$ and $\tilde{v}_j$ are smooth on $\mathcal{J}$. This proves the existence of the required smooth maps. $\square$

2.4.2 Analytic curves in $\mathbb{P}_{2n}(\mathbb{R})$

Symplectic eigenvalues and symplectic eigenvector pairs inherit the regularity properties of real analytic curves in $\mathbb{P}_{2n}(\mathbb{R})$ as we show in the next theorem.

Theorem 2.4.3. *Let $\mathcal{J}$ be an open interval and $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve real analytic at $t_0 \in \mathcal{J}$.*

(i) If d is a symplectic eigenvalue of $\mathbf{A}(t_0)$ with multiplicity m, then for some $\epsilon > 0$, there exist m symplectic eigenvalue maps $\tilde{d}_1, \ldots, \tilde{d}_m : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{R}$, and m corresponding symplectically orthonormal symplectic eigenvector pair maps $(\tilde{u}_1, \tilde{v}_1), \ldots, (\tilde{u}_m, \tilde{v}_m) : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{R}^{2n} \times \mathbb{R}^{2n}$ that are real analytic at t_0 with each $\tilde{d}_j(t_0) = d$.

(ii) There exists an $\epsilon > 0$ such that all the n symplectic eigenvalues of $\mathbf{A}(t)$ and a corresponding symplectic eigenbasis can be chosen on $(t_0 - \epsilon, t_0 + \epsilon)$ to be real analytic at t_0.

Proof. Let $\mathbf{H} : \mathcal{J} \to \mathbb{H}_{2n}(\mathbb{C})$ be the map given by $\mathbf{H}(t) = \imath \mathbf{A}^{1/2}(t) J \mathbf{A}^{1/2}(t)$. Since $\mathbf{A}$ is real analytic at t_0, by Proposition 2.1.2, the map $\mathbf{H}$ is also real analytic at t_0. By Proposition 1.2.7, the multiplicity of the eigenvalue d of $\mathbf{H}(t_0)$ is m. Hence by Theorem 2.2.4, there exists an $\epsilon > 0$, and m functions $\tilde{d}_1, \ldots, \tilde{d}_m : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{R}$ and m functions $x_1, \ldots, x_m : (t_0 - \epsilon, t_0 + \epsilon) \to \mathbb{C}^{2n}$ that are real analytic at t_0 such that $\tilde{d}_1(t), \ldots, \tilde{d}_m(t)$ are m eigenvalues of $\mathbf{H}(t)$ and $\{x_1(t), x_2(t), \ldots, x_m(t)\}$ is a corresponding orthonormal set of eigenvectors. Also $\tilde{d}_j(t_0) = d$ for all $j = 1, \ldots, m$. Since $\mathbf{H}(t)$ is invertible for every t and $d > 0$, each $\tilde{d}_j(t) > 0$. Hence $\tilde{d}_j(t)$ is a symplectic eigenvalue of $\mathbf{A}(t)$ for every $t \in (t_0 - \epsilon, t_0 + \epsilon)$ and $j = 1, \ldots, m$. Let $x_j(t) = \overline{u}_j(t) - \imath \overline{v}_j(t)$ be the Cartesian decomposition of $x_j(t)$. For every $t \in (t_0 - \epsilon, t_0 + \epsilon)$ let $\tilde{u}_j(t) = \sqrt{2\tilde{d}_j(t)}\mathbf{A}^{-1/2}(t)\overline{u}_j(t)$ and $\tilde{v}_j(t) = \sqrt{2\tilde{d}_j(t)}\mathbf{A}^{-1/2}(t)\overline{v}_j(t)$. Since $\tilde{d}_j(t)$ and $\mathbf{A}^{-1/2}(t)$ are real analytic at t_0, $\tilde{u}_j(t)$ and $\tilde{v}_j(t)$ are real analytic at t_0. Finally by Proposition 1.2.7, $\{\tilde{u}_1(t), \ldots, \tilde{u}_m(t), \tilde{v}_1(t), \ldots, \tilde{v}_m(t)\}$ is a symplectically orthonormal set of symplectic eigenvector pairs of $\mathbf{A}(t)$ corresponding to $\tilde{d}_1(t), \ldots, \tilde{d}_m(t)$. This proves (i).

Let $\overline{d}_1 < \cdots < \overline{d}_k$ be distinct symplectic eigenvalues of $\mathbf{A}(t_0)$ with multiplicities $m_1, \ldots, m_k$, respectively. By statement (i) of the theorem, we can find an $\epsilon > 0$ and n symplectic eigenvalue functions $\tilde{d}_{1,1}(t), \ldots, \tilde{d}_{1,m_1}(t), \ldots, \tilde{d}_{k,1}(t), \ldots, \tilde{d}_{k,m_k}(t)$ of $\mathbf{A}(t)$ on $(t_0 - \epsilon, t_0 + \epsilon)$ that are real analytic at t_0. Also for each $j = 1, \ldots, k$, we can choose corresponding symplectically orthonormal symplectic eigenvector pairs $(\tilde{u}_{j,i}(t), \tilde{v}_{j,i}(t))$, $1 \le i \le m_j$, that are real analytic at t_0.Using Proposition 2.3.1, we can assume that $\epsilon > 0$ is small enough so that for all $t \in (t_0 - \epsilon, t_0 + \epsilon)$ $\tilde{d}_{r,i}(t) \neq \tilde{d}_{s,j}(t)$ for all $1 \le i \le m_r$ and $1 \le j \le m_s$, $r \neq s$. Thus by Corollary 1.2.8 the symplectic eigenvector pairs $(\tilde{u}_{j,i}(t), \tilde{v}_{j,i}(t))$, $1 \le i \le m_j$, $1 \le j \le k$, form the required symplectic eigenbasis. $\square$

We now prove that one can choose symplectic eigenvalues and a symplectic eigenbasis real analytically for real analytic curves in $\mathbb{P}_{2n}(\mathbb{R})$ over an interval. We only provide an outline of the proof since it is similar to that of Theorem 2.4.3.

Theorem 2.4.4. *Let $\mathcal{J}$ be an open interval and $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve real analytic on $\mathcal{J}$. We can choose n symplectic eigenvalue functions and a corresponding symplectic eigenbasis map such that they are real analytic on $\mathcal{J}$.*

Proof. The curve $\mathbf{A}$ being analytic on the whole interval $\mathcal{J}$ implies that the curve $t \mapsto \imath\mathbf{A}(t)^{1/2}J\mathbf{A}(t)^{1/2}$ is analytic on $\mathcal{J}$. By Theorem 2.2.5, the eigenvalues and a corresponding eigenbasis can be chosen analytically on $\mathcal{J}$. By arguing in a similar way as in the proof of Theorem 2.4.3(i), we get real analytic maps $\tilde{d}_j : \mathcal{J} \to (0, \infty)$, and $\tilde{u}_j, \tilde{v}_j : \mathcal{J} \to \mathbb{R}^{2n}$ for $j = 1, \dots, n$ such that $\tilde{d}_1(t), \dots, \tilde{d}_n(t)$ are the symplectic eigenvalues of $\mathbf{A}(t)$ with $\{\tilde{u}_1(t), \dots, \tilde{u}_n(t), \tilde{v}_1(t), \dots, \tilde{v}_n(t)\}$ corresponding symplectic basis for all $t \in \mathcal{J}$. $\square$

We know that the symplectic eigenvalue maps $d_1, \cdots, d_n$ are not differentiable on $\mathbb{P}_{2n}(\mathbb{R})$ in general. But we show that they are piecewise real analytic for real analytic curves in $\mathbb{P}_{2n}(\mathbb{R})$.

Theorem 2.4.5. *Let $\mathbf{A} : \mathcal{J} \to \mathbb{P}_{2n}(\mathbb{R})$ be a curve real analytic on the open interval $\mathcal{J}$ and let $[a, b]$ be any compact interval contained in $\mathcal{J}$. Then for each $j = 1, \dots, n$, the map $t \mapsto d_j(t) = d_j(\mathbf{A}(t))$ is piecewise real analytic on $[a, b]$. Further for each $t \in [a, b]$, we can find a symplectic eigenbasis $\{u_1(t), \dots, u_n(t), v_1(t), \dots, v_n(t)\}$ of $\mathbf{A}(t)$ corresponding to $d_1(t), \dots, d_n(t)$ such that the maps $u_1, \dots, u_n, v_1, \dots, v_n$ are also piecewise real analytic on $[a, b]$.*

Proof. By Theorem 2.4.4, we can find n symplectic eigenvalues $\tilde{d}_1(t), \dots, \tilde{d}_n(t)$ of $\mathbf{A}(t)$ and a corresponding symplectic eigenbasis $\{\tilde{u}_1(t), \dots, \tilde{u}_n(t), \tilde{v}_1(t), \dots, \tilde{v}_n(t)\}$ such that each of the maps $\tilde{d}_j, \tilde{u}_j$ and $\tilde{v}_j$ are real analytic on $\mathcal{J}$.

Define $\mathcal{I}$ to be the set of all ordered pairs (i, j), $1 \leq i \neq j \leq n$, such that $\tilde{d}_i(t) \neq \tilde{d}_j(t)$ for at least one t in $[a, b]$. Let E be the set of all points t in $[a, b]$ such that $\tilde{d}_i(t) = \tilde{d}_j(t)$ for some $(i, j) \in \mathcal{I}$. By using the real analyticity of the maps $\tilde{d}_1, \dots, \tilde{d}_n$ and the definition of the set $\mathcal{I}$, we can see that E is finite. Then for every $i = 1, \dots, n$ the multiplicity of $\tilde{d}_i(t)$ is the same for all t in $[a, b] \setminus E$. Hence $\tilde{d}_1, \dots, \tilde{d}_n$ can be reordered so that $\tilde{d}_i(t) = d_i(t)$ for all $t \in [a, b] \setminus E$. The theorem thus follows by suitably reordering the symplectic eigenvalues $\tilde{d}_1, \dots, \tilde{d}_n$ and correspondingly reordering the symplectic eigenvalue pairs $\tilde{u}_1, \dots, \tilde{u}_n, \tilde{v}_1, \dots, \tilde{v}_n$. $\square$

2.5 Symplectic analogue of Lidskii's theorem and other applications

This section is based on our work in Section 5 of [39]. Recall for elements x, y of $\mathbb{R}^n$ that x is said to be supermajorised by y, in symbols $x \prec^w y$, if for $1 \leq k \leq n$

$$\sum_{j=1}^{k} x_j^{\uparrow} \geq \sum_{j=1}^{k} y_j^{\uparrow}. \tag{2.44}$$

Here $x_1^{\uparrow} \leq \ldots \leq x_n^{\uparrow}$ are the components of x arranged in increasing order. For a positive definite matrix A, we denote by $d^{\uparrow}(A)$ the n-tuple of symplectic eigenvalues arranged in increasing order, i.e.,

$$d^{\uparrow}(A) = (d_1(A), \ldots, d_n(A)).$$

We give a symplectic analogue of the Lidskii's theorem (2.4) as an application of our analysis of symplectic eigenvalues for real analytic curves in $\mathbb{P}_{2n}(\mathbb{R})$. The proof of the result is inspired by the analytic proof of the Lidskii's theorem by Kato ([43], Ch.II, Sec.6.5).

Theorem 2.5.1. *Let A, B be two $2n \times 2n$ positive definite matrices. Then*

$$d^{\uparrow}(A+B) - d^{\uparrow}(A) \prec^w d^{\uparrow}(B). \tag{2.45}$$

Proof. Define the map $\varphi : [0,1] \to \mathbb{P}_{2n}(\mathbb{R})$ as

$$\varphi(t) = A + tB.$$

Clearly φ is real analytic with $\varphi'(t) = B$. Let $1 \leq j \leq n$, and let $d_j(t) = d_j(\varphi(t))$. By Theorem 2.4.5, d_j is piecewise real analytic. Also by the same theorem, we can find a piecewise real analytic symplectic eigenbasis $\beta(t) = \{u_1(t), \ldots, u_n(t), v_1(t), \ldots, v_n(t)\}$ of $\varphi(t)$ corresponding to $d_1(t), \ldots, d_n(t)$. For any t in $[0,1]$ at which d_j, u_j and v_j are real analytic, we have

$$d_j'(t) = \frac{1}{2}\left(\langle u_j(t), Bu_j(t)\rangle + \langle v_j(t), Bv_j(t)\rangle\right). \tag{2.46}$$

Let $\mu_1 \leq \ldots \leq \mu_n$ be the symplectic eigenvalues of B and $\beta = \{x_1, \ldots, x_n, y_1, \ldots, y_n\}$ be a corresponding symplectic eigenbasis. Let P_j be the symplectic projection corresponding to (x_j, y_j). Then $B = \sum_{j=1}^{n} \mu_j P_j$. Thus by using this expression for B and using (1.23) for P_k in

(2.46), we get

$$\begin{aligned} d_j'(t) &= \sum_{k=1}^{n} \frac{\mu_k}{2} \left(\langle u_j(t), P_k u_j(t) \rangle + \langle v_j(t), P_k v_j(t) \rangle \right) \\ &= \sum_{k=1}^{n} \frac{\mu_k}{2} \Big(\langle u_j(t), J y_k \rangle^2 + \langle u_j(t), J x_k \rangle^2 \\ &\quad + \langle v_j(t), J y_k \rangle^2 + \langle v_j(t), J x_k \rangle^2 \Big) . \end{aligned} \tag{2.47}$$

Since $\beta(t)$ and β are symplectic bases of $\mathbb{P}_{2n}(\mathbb{R})$, the matrix $M(t)$ with rsth entry

$$m_{rs}(t) = \begin{cases} \langle u_j(t), J x_k \rangle & r = j, s = k, 1 \leq j, k \leq n \\ \langle u_j(t), J y_k \rangle & r = j, s = n+k, 1 \leq j, k \leq n \\ \langle v_j(t), J x_k \rangle & r = n+j, s = k, 1 \leq j, k \leq n \\ \langle v_j(t), J y_k \rangle & r = n+j, s = n+k, 1 \leq j, k \leq n \end{cases}$$

is a symplectic matrix. Let $\widetilde{M}(t)$ be the $n \times n$ matrix with jkth entry

$$\frac{m_{jk}^2(t) + m_{j(n+k)}^2(t) + m_{(n+j)k}^2(t) + m_{(n+j)(n+k)}^2(t)}{2}.$$

Then by (2.47), we see that $d_j'(t)$ is the jth component of the vector $\widetilde{M}(t) d^{\uparrow}(B)$, i.e.,

$$d'(t) = \widetilde{M}(t) d^{\uparrow}(B). \tag{2.48}$$

where $d'(t) = (d_1'(t), \ldots, d_n'(t))^T$. Since d_j, u_j, v_j are piecewise real analytic on $[0,1]$, the maps d_j and $\widetilde{M}$ are integrable on $[0,1]$. Denote by $\overline{M}$, the $n \times n$ matrix

$$\overline{M} = \int_0^1 \widetilde{M}(t) \mathrm{d}t.$$

We know by Theorem 1.2.3, each $\widetilde{M}(t)$ is doubly superstochastic. Since the set of doubly superstochastic matrices is closed and convex, $\overline{M}$ is also doubly superstochastic. Integrating (2.48), we get

$$d^{\uparrow}(A+B) - d^{\uparrow}(A) = \overline{M} d^{\uparrow}(B).$$

We finally obtain (2.45) by Theorem 1.2.2. $\square$

Corollary 2.5.2. *For $A, B \in \mathbb{P}_{2n}(\mathbb{R})$, and for all $1 \leq i_1 < \cdots < i_k \leq n$,*

$$\sum_{j=1}^{k} d_{i_j}(A+B) \geq \sum_{j=1}^{k} d_{i_j}(A) + \sum_{j=1}^{k} d_j(B). \tag{2.49}$$

In particular,

$$d_j(A+B) \geq d_j(A) + d_1(B), \tag{2.50}$$

and

$$d_j(A+I) \geq d_j(A) + 1.$$

Here I denotes the $2n \times 2n$ identity matrix.

Proof. Let $x, y \in \mathbb{R}^m$ and $w = x^{\uparrow} - y^{\uparrow}$. Suppose $z \in \mathbb{R}^m$ is such that $w \prec^w z$. Let $1 \leq i_1 < \ldots < i_k \leq m$ be any indices. We have

$$\begin{aligned} \sum_{j=1}^{k} \left(x_{i_j}^{\uparrow} - y_{i_j}^{\uparrow} \right) &= \sum_{j=1}^{k} w_{i_j} \\ &\geq \sum_{j=1}^{k} w_j^{\uparrow} \\ &\geq \sum_{j=1}^{k} z_j^{\uparrow}. \end{aligned}$$

This gives

$$\sum_{j=1}^{k} x_{i_j}^{\uparrow} \geq \sum_{j=1}^{k} y_{i_j}^{\uparrow} + \sum_{j=1}^{k} z_j^{\uparrow}.$$

The assertion now follows directly from Theorem 2.5.1. □

When $\{i_1, \ldots, i_k\}$ is the set $\{1, \ldots, k\}$ in (2.49), we obtain

$$\sum_{j=1}^{k} d_j(A+B) \geq \sum_{j=1}^{k} d_j(A) + \sum_{j=1}^{k} d_j(B),$$

the inequalities first proved by Hiroshima. See [11, 36]. Using eigenvalue inequalities and the fact that $d_j(A)$ are eigenvalues of $\imath A^{1/2} J A^{1/2}$, the inequalities (2.50) were proved recently by R. Bhatia in [10] in the case when A and B are of the following type:

$$A = \begin{pmatrix} D & O \\ O & D \end{pmatrix}, \quad B = \begin{pmatrix} X & O \\ O & X^{-1} \end{pmatrix}$$

where D is the $n \times n$ diagonal matrix with diagonal entries $d_1(A), \ldots, d_n(A)$ and X is any $n \times n$ positive definite matrix. This also illustrates that treating $d_j(A)$ as eigenvalues of $\imath A^{1/2} J A^{1/2}$ and using eigenvalue inequalities makes it difficult to obtain general results on symplectic eigenvalues. We point out that the supermajorisation in (2.45) cannot be replaced by majorisation. Let $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ and $B = I_2$, the 2×2 identity matrix. The only symplectic eigenvalues of A, B and $A + B$ are

$$d_1(A) = \sqrt{3},\ d_1(B) = 1 \text{ and } d_1(A+B) = 2\sqrt{2}.$$

Clearly $d_1(A+B) > d_1(A) + d_1(B)$.

As a consequence of Theorem 2.5.1, we have the following local maximiser and local minimiser properties on sums of symplectic eigenvalue maps.

Corollary 2.5.3. *For all $k = 1, \ldots, n$ and $1 \leq i_1 < \cdots < i_k \leq n$, the map $A \mapsto \sum_{j=1}^{k} d_{i_j}(A)$ on $\mathbb{P}_{2n}(\mathbb{R})$ has neither a local minimiser nor a local maximiser in $\mathbb{P}_{2n}(\mathbb{R})$. In particular, for every $j = 1, \ldots, n$, the map $A \mapsto d_j(A)$ has neither a local minimiser nor a local maximiser in $\mathbb{P}_{2n}(\mathbb{R})$.*

Proof. Let I denote the $2n \times 2n$ identity matrix. Let $A \in \mathbb{P}_{2n}(\mathbb{R})$ and $\epsilon > 0$ be such that $A \pm \epsilon I \in \mathbb{P}_{2n}(\mathbb{R})$. Then replacing B by ϵI in (2.49) we get

$$\sum_{j=1}^{k} d_{i_j}(A + \epsilon I) \geq \sum_{j=1}^{k} d_{i_j}(A) + k\epsilon$$

Similarly, replacing A by $A - \epsilon I$ and B by ϵI, we get

$$\sum_{j=1}^{k} d_{i_j}(A) \geq \sum_{j=1}^{k} d_{i_j}(A - \epsilon I) + k\epsilon$$

Consequently, we get

$$\sum_{j=1}^{k} d_{i_j}(A + \epsilon I) > \sum_{j=1}^{k} d_{i_j}(A) > \sum_{j=1}^{k} d_{i_j}(A - \epsilon I)$$

□

Recall that the von Neumann entropy of a Gaussian state with the covariance matrix A is given by

$$S(A)=\sum_{i=1}^{n}\left[\left(d_i(A)+\frac{1}{2}\right)\log\left(d_i(A)+\frac{1}{2}\right)-\left(d_i(A)-\frac{1}{2}\right)\log\left(d_i(A)-\frac{1}{2}\right)\right].$$

Theorem 2.5.4. *Let $\mathcal{G}(2n)$ be the set of $2n\times 2n$ Gaussian covariance matrices and $\mathbf{A}:\mathcal{J}\to\mathcal{G}(2n)$ be a real analytic curve on an open interval $\mathcal{J}$. Then the entropy map $S(t)=S(\mathbf{A}(t))$ is monotonically increasing (decreasing) on $\mathcal{J}$ if $\mathbf{A}'(t)$ is positive (negative) semidefinite for all t in $\mathcal{J}$.*

Proof. Since $\mathbf{A}$ is real analytic on $\mathcal{J}$, by Theorem 2.4.4, we can choose the symplectic eigenvalues $\tilde{d}_1(t),\ldots,\tilde{d}_n(t)$, and a corresponding symplectic eigenbasis $\{\tilde{u}_1(t),\ldots,\tilde{u}_n(t),\tilde{v}_1(t),\ldots,\tilde{v}_n(t)\}$ of $\mathbf{A}(t)$ to be real analytic on $\mathcal{J}$. By Remark 2.3.6, we have

$$\tilde{d}_j'(t)=\frac{1}{2}\left(\langle\tilde{u}_j(t),\mathbf{A}'(t)\tilde{u}_j(t)\rangle+\langle\tilde{v}_j(t),\mathbf{A}'(t)\tilde{v}_j(t)\rangle\right). \tag{2.51}$$

If $\mathbf{A}'(t)$ is positive semidefinite, then each $\tilde{d}_j'(t)\geq 0$. Since the maps $\tilde{d}_j$ are continuous and S is a continuous map of $\tilde{d}_j$, $t\to S(t)$ is continuous on $\mathcal{J}$. The matrices $\mathbf{A}(t)$ are Gaussian covariance matrices for all t. Hence $\tilde{d}_j(t)\geq 1/2$ for all $1\leq j\leq n$ and for all $t\in\mathcal{J}$. Let F be the set $\{i:\tilde{d}_i(t)=1/2 \text{ for all } t\in\mathcal{J}\}$. If $F=\{1,\ldots,n\}$, then $S(t)=0$ for all $t\in\mathcal{J}$. So, let $F\neq\{1,\ldots,n\}$. Let $\mathcal{J}_0\subseteq\mathcal{J}$ be any open bounded interval. Clearly it suffices to show that $S(t)$ is monotonically increasing on $\mathcal{J}_0$. Consider the set $E=\{t\in\mathcal{J}_0:\tilde{d}_j(t)=1/2, 1\leq j\leq n, j\notin F\}$. By the analyticity of $\tilde{d}_j$, we know that E is finite. For all $t\in\mathcal{J}_0\setminus E$, we have

$$S'(t)=\sum_{\substack{1\leq j\leq n\\ j\notin F}}\log\left(\frac{2\tilde{d}_j(t)+1}{2\tilde{d}_j(t)-1}\right)\tilde{d}_j'(t),$$

By (2.51), $S'(t)\geq 0$ if $\mathbf{A}'(t)\geq 0$ for all $t\in\mathcal{J}_0\setminus E$. This together with the continuity of $S(t)$ proves the theorem. □

In 2015, R. Bhatia and T. Jain [11] discovered a perturbation theorem for symplectic eigenvalues. Given any unitarily invariant norm $|||\cdot|||$ and $A,B\in\mathbb{P}_{2n}(\mathbb{R})$, we have

$$|||\mathrm{Diag}(\widehat{d}(A))-\mathrm{Diag}(\widehat{d}(B))|||\leq\left(\|A\|^{1/2}+\|B\|^{1/2}\right)|||\,|A-B|^{1/2}|||, \tag{2.52}$$

where $\|\cdot\|$ is the operator norm and $\widehat{d}(A)$ the $2n$ vector with components $d_1(A),\ldots,d_n(A),d_1(A),\ldots,d_n(A)$. Another perturbation theorem on symplectic eigenvalues appeared in 2017 by M. Idel et al. [38] which states that

$$|||\mathrm{Diag}(\widehat{d}(A)) - \mathrm{Diag}(\widehat{d}(B))||| \leq (\kappa(A)\kappa(B))^{1/2}\, |||A - B|||. \tag{2.53}$$

In deriving the results (2.52) and (2.53), the authors use several matrix inequalities for Hermitian matrices and the fact that $d_j(A)$ are eigenvalues of $\imath A^{1/2} J A^{1/2}$. We give another perturbation theorem for symplectic eigenvalues using a completely different method.

Theorem 2.5.5. *Let $A, B \in \mathbb{P}_{2n}(\mathbb{R})$. Then*

$$\max_{1\leq j\leq n} |d_j(A) - d_j(B)| \leq K(A,B)\|A - B\|, \tag{2.54}$$

where $K(A,B) = \int\limits_0^1 \kappa(A + t(B-A))dt$.

Proof. Define $\varphi : [0,1] \to \mathbb{P}_{2n}(\mathbb{R})$ as

$$\varphi(t) = A + t(B - A).$$

As in the proof of Theorem 2.5.1, we see that $d_j(t) = d_j(\varphi(t))$ is piecewise real analytic on $[0,1]$, and we can choose a corresponding piecewise real analytic symplectic eigenbasis $\beta(t) = \{u_1(t), \ldots, u_n(t), v_1(t), \ldots, v_n(t)\}$. Then for t where d_j, u_j, v_j are real analytic, we have

$$d_j'(t) = \frac{1}{2}\left(\langle u_j(t), (B-A)u_j(t)\rangle + \langle v_j(t), (B-A)v_j(t)\rangle\right).$$

Integrating the above equation, we get

$$\begin{aligned}
&|d_j(B) - d_j(A)| \\
&= \left|\int_0^1 d_j'(t)\mathrm{d}t\right| \\
&\leq \frac{1}{2}\int_0^1 |\langle u_j(t), (B-A)u_j(t)\rangle + \langle v_j(t), (B-A)v_j(t)\rangle|\mathrm{d}t \\
&\leq \frac{1}{2}\int_0^1 \left(\|u_j(t)\|^2 + \|v_j(t)\|^2\right)\mathrm{d}t\, \|A - B\|.
\end{aligned} \tag{2.55}$$

Since $(u_j(t), v_j(t))$ is a normalised symplectic eigenvector pair of $\varphi(t)$ corresponding to $d_j(t)$,

$$\|u_j(t)\|^2 + \|v_j(t)\|^2 \quad \leq \quad \|\varphi(t)^{-1}\| \left(\|\varphi(t)^{1/2}u_j(t)\|^2 + \|\varphi(t)^{1/2}v_j(t)\|^2\right)$$

$$= \quad \|\varphi(t)^{-1}\| 2d_j(t) \leq 2\kappa(\varphi(t)).$$

Thus (2.55) gives (2.54). □

Even though our perturbation bound in (2.54) does not improve the recently obtained perturbation bound in [38], it does better in some cases. This is illustrated by the following example.

Example 16. Let $A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = I_2$, the 2×2 identity matrix. For any $t \in [0, 1]$ we have $\kappa(A + t(B - A)) = \frac{2}{1+t}$ which gives $K(A, B) = 2 \ln 2 \approx 1.39$. Also, $(\kappa(A)\kappa(B))^{1/2} = \sqrt{2} \approx 1.41$. Thus we have $K(A, B) < (\kappa(A)\kappa(B))^{1/2}$.

Chapter 3

First order directional derivatives of symplectic eigenvalues

We know that the symplectic eigenvalue maps d_j are smooth at $A \in \mathbb{P}_{2n}(\mathbb{R})$ whenever $d_j(A)$ is a simple symplectic eigenvalue of A. But d_j are not differentiable in general as we see in Example 15 that d_1 and d_2 fail to be differentiable at $\mathbf{A}(0) = I_4$. We study some weaker differentiability properties of the symplectic eigenvalue maps d_j. The main thrust of the chapter is to develop tools using the theory of symplectic eigenvalues and convex analysis to show that the symplectic eigenvalue maps are *directionally* differentiable, and derive the expressions for the directional derivatives.

In Section 3.1 we summarize definitions and some interesting theory of *Fenchel subdifferentials* for convex functions. We review the theory of Fenchel subdifferentials and directional differentiability of eigenvalues of real symmetric matrices in Section 3.2. We begin Section 3.3 by establishing convexity of the maps $\sigma_m = -2(d_1 + \ldots + d_m)$ and derive a simple expression for the Fenchel subdifferentials of these maps in Theorem 3.3.3. In Section 3.4, we prove that σ_m are directionally differentiable and derive the expression for their directional derivatives. We then prove that d_j are directionally differentiable and compute the expressions for their directional derivatives in Theorem 3.4.3.

3.1 Fenchel subdifferentials of convex functions

The theory of derivatives plays a significant role in studying minimisers and maximisers of differentiable functions. Several optimality conditions of differentiable functions are expressed in terms of their derivatives. Let O be an open subset of the Euclidean space $\mathbb{R}^n$ and $a \in O$. A map $f : O \to \mathbb{R}$ is said to be directionally differentiable at a in any direction $d \in \mathbb{R}^n$ if the limit

$$f'(a;d) = \lim_{t \downarrow 0} \frac{f(a+td) - f(a)}{t} \tag{3.1}$$

exists in $\mathbb{R}$. We say that f is directionally differentiable at a if the limit (3.1) exists in $\mathbb{R}$ for all $d \in \mathbb{R}^n$. The map $f'(a;\cdot) : \mathbb{R}^n \to \mathbb{R}$ is called the directional derivative of f at a. If $f'(a;\cdot)$ is a linear map then f is said to be Gâteaux differentiable at a, and the vector $\nabla f(a)$ given by $\langle \nabla f(a), d \rangle = f'(a;d)$ for all $d \in \mathbb{R}^n$ is called the gradient of f at a. The first order necessary condition for local minimisers states that if f is Gâteaux differentiable at $a \in O$ and a is a local minimiser of f then $\nabla f(a) = 0$.

In the theory of optimization, there often arise functions that are not Gâteaux differentiable. The tools of derivatives can not be used to study such functions. Apart from the directional derivatives, several other weaker notions of derivatives such as directional derivatives, *Dini directional derivatives*, *Clarke directional derivatives* and *Michel-Penot directional derivatives* have been studied in the literature, and they are known as *generelised derivatives*. These generalised derivatives give rise to *Fenchel subdifferentials*, *Dini subdifferentials*, *Clarke subdifferentials* and *Michel-Penot subdifferentials*. See [17, 19, 52]. The generalised derivatives are useful in obtaining optimality conditions for functions that are not necessarily differentiable. The utility of each generalised derivative depends on the class of functions one aims to study. An important class of functions that frequently arises in optimization problems is convex functions. But convex functions are not Gâteaux differentiable in general. This is illustrated by the following simple example.

Example 17. Let $f : \mathbb{R} \to \mathbb{R}$ be the modulus map $f(x) = |x|$ for all $x \in \mathbb{R}$. The function is directionally differentiable at $x = 0$ and its directional derivative is given by $f'(0;\alpha) = |\alpha|$ for all $\alpha \in \mathbb{R}$. But $f'(0;\alpha)$ is not linear in α, therefore f is not Gâteaux differentiable at $x = 0$.

Directional derivatives are also useful in obtaining optimality conditions. If a is a local minimizer of f and the function is directionally differentiable at a then $f'(a;d) \geq 0$ for all $d \in \mathbb{R}^n$. So it is useful to develop tools to study directionally differentiable functions. We see in Theorem 3.1.2 that convex functions are directionally differentiable. The Fenchel subdifferentials of convex functions have been widely studied in the literature. See [17, 63, 65, 75]. We review the theory of Fenchel subdifferentials for convex functions in this section. This section is based on Chapter 3 of [17].

3.1.1 Directional differentiability of convex functions

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a map. Recall that f is said to be convex if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

for all $\alpha \in (0,1)$ and $x, y \in \mathbb{R}^n$. Here we use the convention $0 \cdot \infty = 0$.

Lemma 3.1.1. *Let $h : \mathbb{R} \to (-\infty, \infty]$ be a convex function such that $h(0) = 0$. Then, the map $g : \mathbb{R}\backslash\{0\} \to (-\infty, \infty]$ defined by $g(t) = \frac{h(t)}{t}$ is non-decreasing.*

Proof. Let $0 < s \leq t$ be arbitrary numbers. We have

$$\begin{aligned} h(s) &= h\left(\frac{st}{t}\right) \\ &= h\left(\left(\frac{s}{t}\right) t + \left(1 - \frac{s}{t}\right) 0\right) \\ &\leq \frac{s}{t} h(t) + \left(1 - \frac{s}{t}\right) h(0) \\ &= \frac{s}{t} h(t). \end{aligned}$$

This implies $g(s) \leq g(t)$. Similarly, we get $g(-t) \leq g(-s)$. Also,

$$\begin{aligned} h(s) + h(-s) &= 2\left(\frac{1}{2} h(s) + \frac{1}{2} h(-s)\right) \\ &\geq 2h\left(\frac{s}{2} + \frac{-s}{2}\right) \\ &= 2h(0) \\ &= 0. \end{aligned}$$

This gives $h(s) \geq -h(-s)$, which implies $g(s) \geq g(-s)$ for all $s > 0$. Therefore, we have

$g(-t) \leq g(-s) \leq g(s) \leq g(t)$ for all $0 < s \leq t$. This proves the monotonicity of g on $\mathbb{R} - \{0\}$. $\square$

The function f is said to be sublinear if it is positively homogenous and satisfies $f(x+y) \leq f(x)+f(y)$ for all $x, y \in \mathbb{R}^n$. Here positively homogenous means $f(\lambda x) = \lambda f(x)$ for all $\lambda > 0$ and $x \in \mathbb{R}^n$. Equivalently, f is sublinear if it is convex and positively homogeneous. We denote by int K the interior of any subset K of the Euclidean space $\mathbb{R}^n$.

Theorem 3.1.2. *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and $a \in$ int $\Phi^{-1}(\mathbb{R})$. Then Φ is directionally differentiable at a and $\Phi'(a; \cdot)$ is sublinear.*

Proof. Let $d \in \mathbb{R}^n$ be arbitrary. Define $h(s) = \Phi(a+sd) - \Phi(a)$ for all $t \in \mathbb{R}$. Clearly, h is a convex function on $\mathbb{R}$ with $h(0) = 0$. By Lemma 3.1.1, the map $s \mapsto h(s)/s$ is non decreasing on $\mathbb{R}\backslash\{0\}$. Choose $t > 0$ small enough to ensure that $a \pm td \in$ int $\Phi^{-1}(\mathbb{R})$. Thus we have

$$-\infty < \frac{h(-t)}{-t} \leq \frac{h(s)}{s} \leq \frac{h(t)}{t} < \infty \tag{3.2}$$

for all $0 < s \leq t$. By (3.2), the map $s \mapsto h(s)/s$ is bounded below on $(0, t]$ by a real number. Therefore, the limit

$$\lim_{s \downarrow 0} \frac{h(s)}{s} = \lim_{s \downarrow 0} \frac{\Phi(a+sd) - \Phi(a)}{s}$$

exists in $\mathbb{R}$. So, Φ is directionally differentiable at a.

For any $\lambda > 0$, we have

$$\begin{aligned}
\Phi'(a; \lambda d) &= \lim_{t \downarrow 0} \frac{\Phi(a+t\lambda d) - \Phi(a)}{t} \\
&= \lambda \left(\lim_{t \downarrow 0} \frac{\Phi(a+t\lambda d) - \Phi(a)}{t\lambda} \right) \\
&= \lambda \Phi'(a; d).
\end{aligned}$$

Also, for any $e \in \mathbb{R}^n$,

$$\begin{aligned}
\Phi'(a; d+e) &= \lim_{t \downarrow 0} \frac{\Phi(a+t(d+e)) - \Phi(a)}{t} \\
&= \lim_{t \downarrow 0} \frac{\Phi(a+2t(d/2+e/2)) - \Phi(a)}{t} \\
&\leq \lim_{t \downarrow 0} \frac{\frac{1}{2}\left(\Phi(a+2td) + \Phi(a+2te)\right) - \Phi(a)}{t} \\
&= \lim_{t \downarrow 0} \left(\frac{\Phi(a+2td) - \Phi(a)}{2t} + \frac{\Phi(a+2te) - \Phi(a)}{2t} \right) \\
&= \Phi'(a; d) + \Phi'(a; e).
\end{aligned}$$

Therefore, $\Phi'(a;\cdot)$ is sublinear. $\square$

3.1.2 Fenchel subdifferentials of convex functions and Max formula

The theory of Fenchel subdifferentials of convex functions is used throughout the chapter. We summarise the relevant results from convex analysis useful for our present work. In particular, the *Max formula* plays an important role later in the chapter. We include a detailed proof of the result to make the exposition self contained.

Definition 3.1.3. Let $h : \mathbb{R}^n \to (-\infty, \infty]$ be a map. The Fenchel subdifferential of h at $a \in \mathbb{R}^n$ is the closed convex set

$$\partial h(a) = \{\phi \in \mathbb{R}^n : \langle \phi, x - a\rangle \leq h(x) - h(a), \quad \forall x \in \mathbb{R}^n\}.$$

We illustrate the computation of the Fenchel subdifferential of some simple functions in the following examples.

Example 18. Let h be the modulus map defined by $h(x) = |x|$ for all $x \in \mathbb{R}$. We have

$$\begin{aligned}\partial h(0) &= \{\alpha \in \mathbb{R} : \alpha \cdot (\beta - 0) \leq h(\beta) - h(0) \ \ \forall \beta \in \mathbb{R}\}\\ &= \{\alpha \in \mathbb{R} : \alpha\beta \leq |\beta| \ \ \forall \beta \in \mathbb{R}\}\\ &= [-1, 1].\end{aligned}$$

Similarly, the Fenchel subdifferentials at other points are given by

$$\partial h(a) = \begin{cases}\{1\} & \text{if } a > 0,\\ \{-1\} & \text{if } a < 0.\end{cases}$$

Example 19. Let $\|\cdot\|_\infty$ and $\|\cdot\|_1$ be the norms on $\mathbb{R}^n$ given by

$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|, \qquad \|x\|_1 = \sum_{i=1}^{n} |x_i|$$

for all $x \in \mathbb{R}^n$. Let $y \in \partial\|0\|_\infty$. By definition we have $\langle y, x\rangle \leq \|x\|_\infty$ for all $x \in \mathbb{R}^n$. This implies $\sum_{i=1}^n y_i x_i \leq \|x\|_\infty$ for all $x \in \mathbb{R}^n$. Choosing x with ith component ± 1 implies $\|y\|_1 \leq 1$. Conversely, if $y \in \mathbb{R}^n$ with $\|y\|_1 \leq 1$ then $\langle y, x\rangle \leq \|y\|_1\|x\|_\infty \leq \|x\|_\infty$. This implies $y \in \|0\|_\infty$. Thus we have

$$\partial\|0\|_\infty = \{y \in \mathbb{R}^n : \|y\|_1 \leq 1\}.$$

One can similarly verify that

$$\partial\|0\|_1 = \{y \in \mathbb{R}^n : \|y\|_\infty \leq 1\}.$$

Proposition 3.1.4. *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and $a \in$ int $\Phi^{-1}(\mathbb{R})$. The Fenchel subdifferential of Φ at a is given by*

$$\partial\Phi(a) = \{\phi \in \mathbb{R}^n : \langle \phi, d\rangle \leq \Phi'(a; d) \ \ \forall d \in \mathbb{R}^n\}.$$

Proof. Suppose $\phi \in \partial\Phi(a)$. By definition of Fenchel subdifferential we have $\langle \phi, td\rangle \leq \Phi(a + td) - \Phi(a)$ for every $d \in \mathbb{R}^n$ and $t > 0$. Dividing the expression by t and then taking the limit $t \downarrow 0$, we get $\langle \phi, d\rangle \leq \Phi'(a; d)$ for all $d \in \mathbb{R}^n$. For the other side inclusion, suppose $\phi \in \mathbb{R}^n$ satisfies $\langle \phi, d\rangle \leq \Phi'(a; d)$ for all $d \in \mathbb{R}^n$. By Lemma 3.1.1 we have for all $0 < t < 1$ and $x \in \mathbb{R}^n$,

$$\frac{\Phi(a + t(x - a)) - \Phi(a)}{t} \leq \Phi(x) - \Phi(a).$$

Taking the limit $t \downarrow 0$ we get

$$\Phi'(a; x - a) \leq \Phi(x) - \Phi(a).$$

This implies $\langle \phi, x - a\rangle \leq \Phi'(a; x - a) \leq \Phi(x) - \Phi(a)$. □

For a sublinear function $f : \mathbb{R}^n \to (-\infty, \infty]$, let lin f be the set

$$\text{lin } f = \{x \in \mathbb{R}^n : f(-x) = -f(x)\}.$$

Proposition 3.1.5. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a sublinear function. The set lin f is the largest subspace of $\mathbb{R}^n$ on which f is linear.*

Proof. By definition, lin f contains all the subsapces of $\mathbb{R}^n$ on which f is linear. Thus, we only need to prove that lin f is a vector subspace and f is linear on it. By sublinearity, we know that $f(0) = 0$. This implies $0 \in$ lin f and hence lin f is non-empty. Let $x, y \in$ lin f and $\alpha \in \mathbb{R}$. We will show that $\alpha x + y \in$ lin f. We avoid triviality by assuming that $\alpha \neq 0$. Observe that $f\left(\frac{\alpha}{|\alpha|}(-x)\right) = -f\left(\frac{\alpha}{|\alpha|}x\right)$. Thus we get,

$$\begin{aligned} f(-\alpha x - y) &\leq f(-\alpha x) + f(-y) \\ &= f\left(|\alpha|\frac{\alpha}{|\alpha|}(-x)\right) - f(y) \\ &= |\alpha| f\left(\frac{\alpha}{|\alpha|}(-x)\right) - f(y) \end{aligned}$$

$$
\begin{aligned}
&= -|\alpha| f\left(\frac{\alpha}{|\alpha|}x\right) - f(y) \\
&= -f\left(|\alpha|\frac{\alpha}{|\alpha|}x\right) - f(y) \\
&= -\left(f(\alpha x) + f(y)\right) \\
&\leq -f(\alpha x + y).
\end{aligned}
$$

Also, we have

$$f(-\alpha x - y) + f(\alpha x + y) \geq f(-\alpha x - y + \alpha x + y) = f(0) = 0,$$

which implies $f(-\alpha x - y) \geq -f(\alpha x + y)$. Therefore, we have $f(-\alpha x - y) = -f(\alpha x + y)$, and hence $\alpha x + y \in \operatorname{lin} f$. This proves that lin f is a vector subspace of $\mathbb{R}^n$. Further,

$$
\begin{aligned}
f(x + y) &= -f(-x - y) \\
&\geq -f(-x) - f(-y) \\
&= f(x) + f(y),
\end{aligned}
$$

and this implies $f(x + y) = f(x) + f(y)$. By definition we have $f(\alpha x) = \alpha f(x)$ provided $\alpha \geq 0$. Otherwise,

$$f(\alpha x) = -f(-\alpha x) = -(-\alpha) f(x) = \alpha f(x).$$

Therefore f is linear on lin f. □

The following lemma plays a key role in proving the Max formula.

Lemma 3.1.6. *Suppose that $p : \mathbb{R}^n \to (-\infty, \infty]$ is a sublinear map and $a \in p^{-1}(\mathbb{R})$. Then the function $q(\cdot) = p'(a; \cdot)$ satisfies the following conditions*

(i) $q(\lambda a) = \lambda q(a)$ *for all* $\lambda \in \mathbb{R}$,

(ii) $q \leq p$, *and*

(iii) *lin* $q \supset$ *lin* $p +$ *span*$\{a\}$.

Proof. For arbitrary $\lambda \in \mathbb{R}$, choose $t > 0$ small so that $1 + t\lambda > 0$. This implies $p(a + t\lambda a) - p(a) = (1+t\lambda)p(a) - p(a) = t\lambda p(a)$ for small $t > 0$. Thus we have $p(a+t\lambda a) - p(a) = t\lambda p(a)$ for small $t > 0$. Dividing both the sides by t and taking the limit $t \downarrow 0$ gives $q(\lambda a) = \lambda p(a)$. Similarly, $q \leq p$ directly follows from the fact that $p(a + td) \leq p(a) + tp(d)$ for all $d \in \mathbb{R}$ and $t > 0$.

Let $x \in \operatorname{lin} p$. From the second part, we have

$$q(-x) \leq p(-x) = -p(x) \leq -q(x). \tag{3.3}$$

By Theorem 3.1.2, we know that q sublinear and $q(0) = 0$. This gives

$$0 = q(0) = q(-x + x) \leq q(-x) + q(x).$$

Therefore we also have $q(-x) \geq -q(x)$. So, by (3.3) we have $q(-x) = -q(x)$. Thus we have $\text{lin } p \subset \text{lin } q$. From the first part, we have $\text{span}\{a\} \subset \text{lin } q$. We know by Proposition 3.1.5 that $\text{lin } p, \text{lin } q$ are vector subspaces. We conclude that $\text{lin } p + \text{span}\{a\} \subset \text{lin } q$. □

Theorem 3.1.7. ***(Max formula)*** *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and $a \in$ int $\Phi^{-1}(\mathbb{R})$. Then, $\partial\Phi(a)$ is non-empty and we have*

$$\Phi'(a; d) = \max\{\langle \phi, d\rangle : \phi \in \partial\Phi(a)\}$$

for all $d \in \mathbb{R}^n$.

Proof. In the view of Proposition 3.1.4 it suffices to show that for any $d \in \mathbb{R}^n$, there exists $\phi \in \partial\Phi(a)$ such that $\langle \phi, d\rangle = \Phi'(a; d)$. Let d be a fixed unit vector of $\mathbb{R}^n$. Choose an orthonormal basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$ such that $e_1 = d$. Define the function $p_0(\cdot) = \Phi'(a; \cdot)$. By Theorem 3.1.2 the function p_0 is sublinear. Define the functions $p_k(\cdot) = p_{k-1}(e_k; \cdot)$ for $k = 1, \ldots, n$. By part (iii) of Lemma 3.1.6, we know that p_n is a linear functional on $\mathbb{R}^n$. So, there exists $\phi \in \mathbb{R}^n$ such that $p_n(\cdot) = \langle \phi, \cdot\rangle$. By part (ii) of Lemma 3.1.6, we have $p_n(x) \leq p_0(x)$, which implies $\langle \phi, x\rangle \leq \Phi'(a; x)$ for all $x \in \mathbb{R}^n$. Proposition 3.1.4 thus implies $\phi \in \partial\Phi(a)$.

By Lemma 3.1.6 and sublinearity, we get the following set of inequalities

$$p_n(d) \leq p_1(d) = p_0'(d; d) = -p_0'(d; -d) = -p_1(-d) \leq -p_n(-d) = p_n(d).$$

The above inequalities are thus equalities, and we get

$$\langle \phi, d\rangle = p_0'(d; d) = p_0(d) = \Phi'(a; d). \tag{3.4}$$

The equality in (3.4) holds for all non-zero elements d if we take $e_1 = d/\|d\|$. If d is zero, then we could repeat the same process by taking the standard basis of $\mathbb{R}^n$ and the conclusion would trivially hold. □

In Example 18, the function h is convex and is Gâteaux differentiable at non-zero points. Its derivatives at non-zero points are given by

$$h'(a) = \begin{cases} 1 & \text{if } a > 0, \\ -1 & \text{if } a < 0. \end{cases}$$

Therefore we have $\partial h(a) = \{h'(a)\}$ for all $a \neq 0$. Thus, the Fenchel subdifferential of h is singleton exactly at those points where the function is Gâteaux differentiable. In the following corollary we see that this is the necessary and sufficient condition for Gâteaux differentiability of convex functions.

Corollary 3.1.8. *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and $a \in$ int $\Phi^{-1}(\mathbb{R})$. Φ is Gâteaux differentiable at a if and only if $\partial\Phi(a)$ is a singleton set. In this case, $\partial\Phi(a) = \{\nabla\Phi(a)\}$.*

Proof. Suppose $\partial\Phi(a)$ is a singleton set given by $\{y\}$. By the Max formula, we have $\Phi'(a; d) = \langle y, d\rangle$. So by definition Φ is Gâteaux differentiable at a and $y = \nabla\Phi(a)$. Conversely, suppose Φ is Gâteaux differentiable at a. By Lemma 3.1.1 we have for all $0 < t < 1$ and $x \in \mathbb{R}^n$,

$$\frac{\Phi(a + t(x - a)) - \Phi(a)}{t} \leq \Phi(x) - \Phi(a).$$

Taking the limit $t \downarrow 0$ we get

$$\langle \nabla\Phi(a), x - a\rangle \leq \Phi(x) - \Phi(a).$$

This implies $\nabla\Phi(a) \in \partial\Phi(a)$. Let $\phi \in \partial\Phi(a)$ be arbitrary. By the Max formula, we get

$$\langle \phi, d\rangle \leq \Phi'(a; d) = \langle \nabla\Phi(a), d\rangle$$

for all $d \in \mathbb{R}^n$. This implies $\phi = \nabla\Phi(a)$. Thus we have $\partial\Phi(a) = \{\nabla\Phi(a)\}$.

□

3.2 Directional derivatives of eigenvalues of symmetric matrices

The theory of eigenvalues serves as a guiding tool in the present study of symplectic eigenvalues. So, we review the theory of Fenchel subdifferentials for eigenvalues of symmetric matrices and summarise the relevant results in this section. Eigenvalues arise in various applications and optimisation problems. Among others are *graph partitioning* problems that deal with the minimisation problem of the sums of m largest eigenvalues [20], optimisation problems in structural systems where sensitivity of eigenvalues plays an important role [31, 46, 60], medical imaging area where semismoothness of eigenvalues is applied [49]. Due to the growing importance in the areas of nonsmooth optimisation problems several generalised differential properties of eigenvalues have been studied in the last few decades [33, 34, 73]. This section is based on the work by Hiriart-Urruty and Ye [34].

For a given subset $\mathcal{K}$ of $\mathbb{S}_n(\mathbb{R})$, we denote by conv$\mathcal{K}$ the closed convex set generated by $\mathcal{K}$. Let $m \leq n$ be any positive integer. Denote by $\mathbb{M}_{n,m}(\mathbb{R})$ the set of $n \times m$ real matrices. Let $\Lambda_m : \mathbb{S}_n(\mathbb{R}) \to \mathbb{R}$ be the map given by $\Lambda_m(A) = \lambda_1^{\downarrow}(A) + \ldots + \lambda_m^{\downarrow}(A)$ for all $A \in \mathbb{S}_n(\mathbb{R})$. Here we denote by $\lambda_1^{\downarrow}(A) \geq \ldots \geq \lambda_n^{\downarrow}(A)$ the eigenvalues of A arranged in decreasing order. By Ky Fan's maximum principle ([25], Theorem 1) we have

$$\Lambda_m(A) = \max \{\text{tr } X^T AX : X \in \mathbb{M}_{n,m}(\mathbb{R}), X^T X = I_m\}. \tag{3.5}$$

Using the characterisation (3.5) one can verify that Λ_m are sublinear functions. In particular, Λ_m are convex functions. The expression for the Fenchel subdifferential of Λ_m is derived in two steps. In the first step, it is proved that the Fenchel subdifferential expression for Λ_m at A is given by

$$\partial\Lambda_m(A) = \text{conv}\{XX^T : X \in \mathbb{M}_{n,m}(\mathbb{R}), \text{tr } X^T AX = \Lambda_m(A), X^T X = I_m\}. \tag{3.6}$$

A key role is played by Ky Fan's maximum principle in deriving (3.6). In the second step, the convex set on the right side of (3.6) is further simplified by using the multiplicities of the eigenvalues of A and properties of matrices. A more transparent expression for $\partial\Lambda_m(A)$ is presented in the next theorem.

Let $\alpha_m, \beta_m, \gamma_m$ be non-negative integers where $\gamma_m = \alpha_m + \beta_m$ is the multiplicity of $\lambda_m(A)$ and $\alpha_m \geq 1$. Further,

$$\lambda_{m-\alpha_m}^{\downarrow}(A) > \lambda_{m-\alpha_m+1}^{\downarrow}(A) = \ldots = \lambda_{m+\beta_m}^{\downarrow}(A) > \lambda_{m+\beta_m+1}^{\downarrow}(A).$$

In particular, $\alpha_1 = 1, \beta_1 = \gamma_1 - 1$ and $\alpha_n = \gamma_n, \beta_n = 0$. Here we assume $\lambda_0^{\downarrow}(A) = \infty$ and $\lambda_{n+1}^{\downarrow}(A) = -\infty$. Let $\nabla_m(A)$ be the subset of $\mathbb{M}_{n,m}(\mathbb{R})$ consisting of matrices Z given in the block matrix form

$$Z = \begin{pmatrix} I_k & 0 \\ 0 & W \\ 0 & 0 \end{pmatrix}, \tag{3.7}$$

where $k = m - \alpha_m$ and $W \in \mathbb{M}_{\gamma_m,\alpha_m}(\mathbb{R})$ such that $W^T W = I_{\alpha_m}$. Let U be an orthogonal matrix such that $U^T AU$ is the diagonal matrix with diagonal entries $\lambda_1^{\downarrow}(A), \ldots, \lambda_n^{\downarrow}(A)$.

Theorem 3.2.1. *The Fenchel subdifferential of Λ_m at A is given by*

$$\partial\Lambda_m(A) = conv\{UZZ^T U^T : Z \in \nabla_m(A)\}. \tag{3.8}$$

The Fenchel subdifferential expression (3.8) is characterised by the matrices of the set $\nabla_m(A)$. Using the Max formula one can now derive the expression for the directional derivative of Λ_m at A. Suppose $u_1, \ldots, u_n$ are the columns of U. Let $U_1 = [u_1, \ldots, u_{m-\alpha_m}]$ and $U_2 = [u_{m-\alpha_m+1}, \ldots, u_{m+\beta_m}]$.

Theorem 3.2.2. *The first order directional derivative of Λ_m at A is given by*

$$\Lambda_m'(A; H) = tr\, U_1^T HU_1 + \sum_{k=1}^{\alpha_m} \mu_k(U_2^T HU_2) \tag{3.9}$$

for all $H \in \mathbb{S}_n(\mathbb{R})$. Here $\mu_k(U_2^T HU_2)$ denotes the kth largest eigenvalue of the matrix $U_2^T HU_2$.

The eigenvalue maps $\lambda_1^{\downarrow}, \ldots, \lambda_n^{\downarrow}$ can be written as a difference of the directionally differentiable maps $\Lambda_1, \ldots, \Lambda_n$. This implies that the eigenvalue maps are also directionally differentiable. One can directly get $\lambda_m^{\downarrow\prime}(A; \cdot)$ using the expressions for $\Lambda_m'(A; \cdot)$.

Theorem 3.2.3. *The first order directional derivative of $\lambda_m^{\downarrow}$ at A is given by*

$$\lambda_m^{\downarrow\prime}(A; H) = \mu_{\alpha_m}(U_2^T HU_2) \tag{3.10}$$

for all $H \in \mathbb{S}_n(\mathbb{R})$.

3.3 Fenchel subdifferentials of the negative sums of symplectic eigenvalues

In 2006, T. Hiroshima proved a symplectic analogue of Ky Fan's maximum principle. Given any $P \in \mathbb{P}_{2n}(\mathbb{R})$ and positive integer $m \leq n$ we have

$$\sum_{j=1}^{m} d_j(P) = \min\,\{\text{tr}\ S^T AS : S \in Sp(2n, 2m)\}. \tag{3.11}$$

See ([36], Sec. V, Lemma 1). This characterisation implies that the map $P \mapsto \sum_{j=1}^{m} d_j(P)$ is a concave function on $\mathbb{P}_{2n}(\mathbb{R})$. Define the map $\sigma_m : \mathbb{S}_{2n}(\mathbb{R}) \to (-\infty, \infty]$ by

$$\sigma_m(P) = \begin{cases} -2\sum_{j=1}^{m} d_j(P) & \text{if } P \in \mathbb{P}_{2n}(\mathbb{R}), \\ \infty & \text{otherwise.} \end{cases}$$

The set $\mathbb{P}_{2n}(\mathbb{R})$ is open and convex in $\mathbb{S}_{2n}(\mathbb{R})$. Therefore σ_m are convex functions. We use the theory of symplectic matrices and symplectic eigenvalues developed in the previous chapters to derive the Fenchel subdifferentials of σ_m. This section is based on our work in Section 3 of [56].

We note a useful property of the space $\mathbb{S}_n(\mathbb{R})$ in the following lemma. A more general result for locally convex topological vector spaces is given in Zălinescu ([75], Theorem 1.1.5).

Lemma 3.3.1. *Let $\mathcal{C}$ and $\mathcal{K}$ be non-empty subsets of $\mathbb{S}_n(\mathbb{R})$. If $\mathcal{C}$ is closed, $\mathcal{K}$ is compact and $\mathcal{C} \cap \mathcal{K} = \emptyset$, then there exist $C \in \mathbb{S}_n(\mathbb{R})$ and $\delta > 0$ such that*

$$\mathit{tr}\, CX + \delta \leq \mathit{tr}\, CY$$

for all $X \in \mathcal{C}$ and $Y \in \mathcal{K}$.

Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. We derive the expression for $\partial\sigma_m(A)$ in two steps similar to eigenvalues discussed in Section 3.2. We first compute a preliminary expression for $\partial\sigma_m(A)$ in the following proposition.

Proposition 3.3.2. *The Fenchel subdifferential of σ_m at A is given by*

$$\partial\sigma_m(A) = \mathit{conv}\left\{-SS^T : S \in Sp(2n, 2m, A)\right\}. \tag{3.12}$$

Proof. Let $\mathcal{Q} = \text{conv}\left\{-SS^T : S \in Sp(2n, 2m, A)\right\}$. For any $S \in Sp(2n, 2m, A)$ and $B \in \mathbb{S}_{2n}(\mathbb{R})$ we have

$$\begin{aligned}\langle -SS^T, B - A\rangle &= -\text{tr } SS^T B + \text{tr } SS^T A \\ &= -\text{tr } S^T BS + \text{tr } S^T AS \\ &= -\text{tr } S^T BS - \sigma_m(A) \\ &\leq \sigma_m(B) - \sigma_m(A).\end{aligned}$$

The last equality follows from the fact that $A \in Sp(2n, 2m, A)$ and the last inequality follows by the definition by σ_m. This implies that $-SS^T \in \partial\sigma_m(A)$. We know that $\partial\sigma_m(A)$ is a closed convex set. Thus we have $\mathcal{Q} \subseteq \partial\sigma_m(A)$.

For the other side inclusion, we assume $\partial\sigma_m(A)\backslash\mathcal{Q} \neq \emptyset$ and derive a contradiction. Let $B \in \partial\sigma_m(A)\backslash\mathcal{Q}$. By Lemma 3.3.1 we get $C \in \mathbb{S}_{2n}(\mathbb{R})$ and $\delta > 0$ such that for all $S \in Sp(2n, 2m, A)$ we have

$$\langle B, C\rangle \geq \langle -SS^T, C\rangle + \delta. \tag{3.13}$$

Let (a, b) be an open interval containing 0 such that $A(t) = A + tC$ is in $\mathbb{P}_{2n}(\mathbb{R})$ for all $t \in (a, b)$. By Theorem 2.4.5 we get an $\varepsilon > 0$ and continuous maps d_j, u_j, v_j on $[0, \varepsilon) \subset (a, b)$ for $j = 1, \ldots, n$ such that $d_j(t) = d_j(A(t))$ and $\{u_1(t), \ldots, u_n(t), v_1(t), \ldots, v_n(t)\}$ is a symplectic basis of $\mathbb{R}^{2n}$ consisting of symplectic eigenvector pairs of $A(t)$ for all $t \in [0, \varepsilon)$. Therefore the matrix

$$S(t) = [u_1(t), \ldots, u_n(t), v_1(t), \ldots, v_n(t)]$$

is an element of $Sp(2n, A(t))$ for all $t \in [0, \varepsilon)$. For any t in $(0, \varepsilon)$ we have

$$\begin{aligned}\langle -S(t)S(t)^T, C\rangle &= -\text{tr}S(t)^T CS(t) \\ &= \frac{-\text{tr } S(t)^T(A + tC)S(t) + \text{tr } S(t)^T AS(t)}{t} \\ &= \frac{-\text{tr } S(t)^T A(t)S(t) + \text{tr } S(t)^T AS(t)}{t} \\ &= \frac{\sigma_m(A(t)) + \text{tr } S(t)^T AS(t)}{t} \\ &\geq \frac{\sigma_m(A(t)) - \sigma_m(A)}{t} \\ &\geq \langle C, B\rangle.\end{aligned}$$

The second last inequality follows because $S(t)$ is an element of $Sp(2n, 2m)$ for all $t \in (0, \varepsilon)$, and the last inequality follows from the fact that $B \in \partial\sigma_m(A)$. By continuity we get

$$\langle -S(0)S(0)^T, C\rangle \geq \langle B, C\rangle.$$

But $S(0) \in Sp(2n, 2m, A)$ and hence we get a contradiction by (3.13). Therefore our assumption $\partial\sigma_m(A)\backslash\mathcal{Q} \neq \emptyset$ is false. This completes the proof. $\square$

We now simplify the convex set in the right side of the equation (3.12) and derive a simpler expression for $\partial\sigma_m(A)$. Let i_m, j_m, r_m be the non-negative integers given as follows. Let $r_m = i_m + j_m$ be the multiplicity of $d_m(A)$ and $i_m \geq 1$. Further,

$$d_{m-i_m}(A) < d_{m-i_m+1}(A) = \ldots = d_{m+j_m}(A) < d_{m+j_m+1}(A).$$

In particular, $i_1 = 1, j_1 = r_1 - 1$ and $i_n = r_n, j_n = 0$. Define $\Delta_m(A)$ to be the set of $2n \times 2m$ real matrices of the form

$$\left(\begin{array}{cc|cc} I & O & O & O \\ O & U & O & V \\ O & O & O & O \\ \hline O & O & I & O \\ O & -V & O & U \\ O & O & O & O \end{array}\right)$$

where I is the $(m - i_m) \times (m - i_m)$ identity matrix, and U, V are $r_m \times i_m$ real matrices such that the columns of $U + \iota V$ are orthonormal. Recall that if S is a matrix with $2n$ columns and $m_1, \ldots, m_k$ are positive integers with $m_1 + \ldots + m_k = n$ then the symplectic column partition of S of order $(m_1, \ldots, m_k)$ is given by (1.28).

Theorem 3.3.3. *Let $A \in \mathbb{P}_{2n}(\mathbb{R})$ and $M \in Sp(2n, A)$ be fixed. The Fenchel subdifferential of σ_m at A is given by*

$$\partial\sigma_m(A) = conv\left\{-MHH^TM^T : H \in \Delta_m(A)\right\}. \tag{3.14}$$

Proof. We first show that

$$\partial\sigma_m(A) \subseteq \operatorname{conv}\left\{-MHH^TM^T : H \in \Delta_m(A)\right\}.$$

By Proposition 3.3.2 it suffices to show that for every $S \in Sp(2n, 2m, A)$ there exists some $H \in \Delta_m(A)$ such that $SS^T = MHH^TM^T$. Let I denote the $2n \times 2n$ identity matrix and $I = \overline{I} \diamond \widetilde{I} \diamond \widehat{I}$ be the symplectic column partition of I of order $(m - i_m, r_m, n - m - j_m)$. Let $\overline{M} = M\overline{I}$, $\widetilde{M} = M\widetilde{I}$ and $\widehat{M} = M\widehat{I}$. The columns of $\widetilde{M}$ consist of symplectic eigenvector pairs of A corresponding to the symplectic eigenvalue $d_m(A)$. Let $S \in Sp(2n, 2m, A)$ be arbitrary and $S = \overline{S} \diamond \widetilde{S}_1$ be the symplectic column partition of S of order $(m - i_m, i_m)$. Extend S to a matrix $S \diamond \widetilde{S}_2$ in $Sp(2n, 2(m + j_m), A)$ by Corollary 1.2.10. The columns of $\overline{S}$ consist of symplectic eigenvector pairs of A corresponding to $d_1(A), \ldots, d_{m-i_m}(A)$, and the columns of $\widetilde{S}_1 \diamond \widetilde{S}_2$ consist of symplectic eigenvector pairs of A corresponding to $d_m(A)$. By Corollary 1.3.3 we can find orthosymplectic matrices Q and R of orders $2(m-i_m) \times 2(m-i_m)$ and $2r_m \times 2r_m$ respectively such that $\overline{S} = \overline{M}Q$ and $\widetilde{S}_1 \diamond \widetilde{S}_2 = \widetilde{M}R$. Let $R = \overline{R} \diamond \widetilde{R}$ be the symplectic column partition of R of order (i_m, j_m). By Proposition 1.3.4 we have $\widetilde{S}_1 \diamond \widetilde{S}_2 = \widetilde{M}\overline{R} \diamond \widetilde{M}\widetilde{R}$. This implies $\widetilde{S}_1 = \widetilde{M}\overline{R}$. Therefore

$$S = \overline{S} \diamond \widetilde{S}_1 = \overline{M}Q \diamond \widetilde{M}\overline{R}.$$

So we have

$$S = M(\overline{I}Q \diamond \widetilde{I}\overline{R}).$$

By Proposition 1.2.1 there exist $r_m \times r_m$ real matrices X, Y such that $X + \iota Y$ is unitary and

$$R = \begin{pmatrix} X & Y \\ -Y & X \end{pmatrix}.$$

Let U, V be the $r_m \times i_m$ matrices consisting of the first i_m columns of X, Y respectively. Therefore

$$\overline{R} = \begin{pmatrix} U & V \\ -V & U \end{pmatrix}. \tag{3.15}$$

We have

$$\begin{aligned} SS^T &= M(\overline{I}Q \diamond \widetilde{I}\overline{R})(\overline{I}Q \diamond \widetilde{I}\overline{R})^T M^T \\ &= M\left((\overline{I}Q)(\overline{I}Q)^T + (\widetilde{I}\overline{R})(\widetilde{I}\overline{R})^T\right) M^T \\ &= M\left(\overline{I}QQ^T\overline{I}^T + (\widetilde{I}\overline{R})(\widetilde{I}\overline{R})^T\right) M^T \\ &= M\left(\overline{I}\,\overline{I}^T + (\widetilde{I}\overline{R})(\widetilde{I}\overline{R})^T\right) M^T \\ &= M(\overline{I} \diamond \widetilde{I}\overline{R})(\overline{I} \diamond \widetilde{I}\overline{R})^T M^T. \end{aligned}$$

The second and the last equalities follow from Proposition 1.3.4. The fourth equality follows from the fact that Q is an orthogonal matrix. Let $H = \overline{I} \diamond \widetilde{I}\overline{R}$. By the definition of $\Delta_m(A)$ and (3.15) we have $H \in \Delta_m(A)$. Therefore $SS^T = MHH^TM^T$, where $H \in \Delta_m(A)$.

We now prove the reverse inclusion. By definition, observe that any $H \in \Delta_m(A)$ is of the form

$$H = \overline{I} \diamond \widetilde{I}\begin{pmatrix} U & V \\ -V & U \end{pmatrix}.$$

By Proposition 1.3.4 we thus have

$$MH = \overline{M} \diamond \widetilde{M} \begin{pmatrix} U & V \\ -V & U \end{pmatrix}.$$

We know that the columns of $\overline{M}$ correspond to the symplectic eigenvalues $d_1(A), \ldots, d_{m-i_m}(A)$. By using the fact that the columns of $\widetilde{M}$ correspond to the symplectic eigenvalue $d_m(A)$ we get

$$\begin{aligned} \begin{pmatrix} U & V \\ -V & U \end{pmatrix}^T \widetilde{M}^T A \widetilde{M} \begin{pmatrix} U & V \\ -V & U \end{pmatrix} &= d_m(A) \begin{pmatrix} U & V \\ -V & U \end{pmatrix}^T \begin{pmatrix} U & V \\ -V & U \end{pmatrix} \\ &= d_m(A) I_{2i_m}. \end{aligned}$$

Here we used the fact that the columns of $\left(\begin{smallmatrix} U & V \\ -V & U \end{smallmatrix}\right)$ are orthonormal. The above relation implies that the columns of $\widetilde{M}\left(\begin{smallmatrix} U & V \\ -V & U \end{smallmatrix}\right)$ also correspond to the symplectic eigenvalue $d_m(A)$. Therefore we have $MH \in Sp(2n, 2m, A)$ for all $H \in \Delta_m(A)$, and hence

$$\partial \sigma_m(A) \supseteq \text{conv}\left\{-MHH^T M^T : H \in \Delta_m(A)\right\}.$$

This completes the proof. □

It is interesting to see the similarities between the expressions for Fenchel subdifferentials of σ_m and Λ_m. In particular, $\Delta_m(A)$ in the expression (3.14) for $\partial \sigma_m(A)$ plays similar role as $\nabla_m(A)$ in the expression (3.8) for $\partial \Lambda_m(A)$.

3.4 First order directional derivatives of symplectic eigenvalues

This section is based on our work in Section 4 of [56]. We know that convex functions are directionally differentiable and σ_m are convex functions. Thus σ_m are directionally differentiable. We use the expression (3.14) in computing the directional derivatives of σ_m.

We set up some notations for our convenience. Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$ and M be an element of $Sp(2n, A)$. Let I denote the $2n \times 2n$ identity matrix and $I = \overline{I} \diamond \widetilde{I} \diamond \widehat{I}$ be the symplectic column partition of I of order $(m - i_m, r_m, n - m - j_m)$. Let $\overline{M} = M\overline{I}$, $\widetilde{M} = M\widetilde{I}$ and $\widehat{M} = M\widehat{I}$. Let B be any element of $\mathbb{S}_{2n}(\mathbb{R})$. Define $\overline{B} = -\overline{M}^T B \overline{M}$ and $\widetilde{B} = -\widetilde{M}^T B \widetilde{M}$. Note that $\widetilde{B}$ is a $2r_m \times 2r_m$ real symmetric matrix. Consider the block matrix form of $\widetilde{B}$ given by

$$\widetilde{B} = \begin{pmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{12}^T & \tilde{B}_{22} \end{pmatrix},$$

where each block has order $r_m \times r_m$. Denote by $\widetilde{\widetilde{B}}$ the Hermitian matrix $\tilde{B}_{11}+\tilde{B}_{22}+\iota(\tilde{B}_{12}-\tilde{B}_{12}^T)$.

Theorem 3.4.1. *The directional derivative of σ_m at A is given by*

$$\sigma_m'(A;B) = tr\,\overline{B} + \sum_{j=1}^{i_m} \lambda_j^{\downarrow}(\widetilde{\widetilde{B}}) \tag{3.16}$$

for all $B \in \mathbb{S}_{2n}(\mathbb{R})$. Here $\lambda_j^{\downarrow}(\widetilde{\widetilde{B}})$ denotes the jth largest eigenvalue of $\widetilde{\widetilde{B}}$.

Proof. By the Max formula we have

$$\sigma_m'(A;B) = \max\,\{\langle C, B\rangle : C \in \partial\sigma_m(A)\}$$

for all $B \in \mathbb{S}_{2n}(\mathbb{R})$. It suffices to take the maximum in the above expression over a subset that generates $\partial\sigma_m(A)$. By Theorem 3.3.3 we have

$$\sigma_m'(A;B) = \max\,\{\langle -MHH^TM^T, B\rangle : H \in \Delta_m(A)\}. \tag{3.17}$$

Every element of $\Delta_m(A)$ is of the form $\overline{I} \diamond \widetilde{I}\overline{R}$ where $\overline{R}$ is given by (3.15). Let $H = \overline{I} \diamond \widetilde{I}\overline{R}$ be an arbitrary element of $\Delta_m(A)$. This gives

$$\begin{aligned} MHH^TM^T &= (M(\overline{I} \diamond \widetilde{I}\overline{R}))(M(\overline{I} \diamond \widetilde{I}\overline{R}))^T \\ &= (M\overline{I} \diamond M\widetilde{I}\overline{R})(M\overline{I} \diamond M\widetilde{I}\overline{R})^T \\ &= (\overline{M} \diamond \widetilde{M}\overline{R})(\overline{M} \diamond \widetilde{M}\overline{R})^T \\ &= \overline{M}\,\overline{M}^T + \widetilde{M}\overline{R}\,\overline{R}^T\widetilde{M}^T. \end{aligned} \tag{3.18}$$

The second and the last equalities follow from Proposition 1.3.4. This implies

$$\begin{aligned} \langle -MHH^TM^T, B\rangle &= \text{tr}\,(-MHH^TM^TB) \\ &= \text{tr}\,(-\overline{M}\,\overline{M}^TB) + \text{tr}\,(-\widetilde{M}\overline{R}\,\overline{R}^T\widetilde{M}^TB) \\ &= \text{tr}\,(-\overline{M}\,\overline{M}^TB) + \text{tr}\,(-\overline{R}^T\widetilde{M}^TB\widetilde{M}\overline{R}) \\ &= \text{tr}\,(-\overline{M}^TB\overline{M}) + \text{tr}\,(\overline{R}^T\widetilde{B}\overline{R}) \\ &= \text{tr}\,\overline{B} + \text{tr}\,(U^T\tilde{B}_{11}U + V^T\tilde{B}_{22}V - 2U^T\tilde{B}_{12}V) \\ &\quad + \text{tr}\,(V^T\tilde{B}_{11}V + U^T\tilde{B}_{22}U + 2U^T\tilde{B}_{12}^TV) \\ &= \text{tr}\,\overline{B} + \text{tr}\,(U+\iota V)^*(\tilde{B}_{11} + \tilde{B}_{22} + \iota(\tilde{B}_{12} - \tilde{B}_{12}^T))(U+\iota V) \end{aligned}$$

$$= \text{tr}\, \overline{B} + \text{tr}\, (U + \iota V)^* \widetilde{\widetilde{B}}(U + \iota V). \tag{3.19}$$

Therefore by (3.17) and (3.19) we get

$$\sigma_m'(A; B) = \text{tr}\, \overline{B} + \max_{U+\iota V} \text{tr}\, (U + \iota V)^* \widetilde{\widetilde{B}}(U + \iota V),$$

where the maximum is taken over $r_m \times i_m$ unitary matrices $U + \iota V$. By Ky Fan's extremal characterisation ([25], Theorem 1) we have

$$\max_{U+\iota V} \text{tr}\, (U + \iota V)^* \widetilde{\widetilde{B}}(U + \iota V) = \sum_{j=1}^{i_m} \lambda_j^{\downarrow}(\widetilde{\widetilde{B}}).$$

This completes the proof. □

We see that the directional derivative expression (3.16) of σ_m has two terms, a linear term and a sublinear term which is analogous to the directional derivative expression (3.9) of Λ_m for eigenvalues.

Corollary 3.4.2. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$ and M be an element of $Sp(2n, A)$. If $d_m(A) < d_{m+1}(A)$ then σ_m is Gâteaux differentiable at A with the gradient*

$$\nabla \sigma_m(A) = -(\overline{M} \diamond \widetilde{M})(\overline{M} \diamond \widetilde{M})^T.$$

Here we assume $d_{m+1}(A) = \infty$ for $m = n$.

Proof. If $d_m(A) < d_{m+1}(A)$ then $j_m = 0$ and $i_m = r_m$. Therefore $\overline{R}$ is a $2r_m \times 2r_m$ orthosymplectic matrix in the proof of Theorem 3.4.1. By (3.18) we have

$$MHH^TM^T = \overline{M}\overline{M}^T + \widetilde{M}\widetilde{M}^T$$

for all $H \in \Delta_m(A)$. By Proposition 1.3.4 we have

$$-\overline{M}\overline{M}^T - \widetilde{M}\widetilde{M}^T = -(\overline{M} \diamond \widetilde{M})(\overline{M} \diamond \widetilde{M})^T.$$

Therefore we have by Theorem 3.3.3

$$\partial \sigma_m(A) = \{-(\overline{M} \diamond \widetilde{M})(\overline{M} \diamond \widetilde{M})^T\}.$$

By Corollary 3.1.8 we conclude that σ_m is Gâteaux differentiable with

$$\nabla \sigma_m(A) = -(\overline{M} \diamond \widetilde{M})(\overline{M} \diamond \widetilde{M})^T.$$

□

The symplectic eigenvalue maps $d_1, \ldots, d_n$ can be written as the difference of two directionally differentiable functions. If σ_0 is the zero map on $\mathbb{S}_{2n}(\mathbb{R})$, we have

$$d_m = \frac{1}{2}\left(\sigma_{m-1} - \sigma_m\right) \qquad \text{for } 1 \leq m \leq n.$$

Therefore, the symplectic eigenvalue maps are also directionally differentiable. By definition we get,

$$d_m'(A;\cdot) = \frac{1}{2}\left(\sigma_{m-1}'(A;\cdot) - \sigma_m'(A;\cdot)\right).$$

Theorem 3.4.3. *The directional derivative of d_m at A is given by*

$$d_m'(A;B) = -\frac{1}{2}\lambda_{i_m}^{\downarrow}(\widetilde{\widetilde{B}}) \tag{3.20}$$

for all $B \in \mathbb{S}_{2n}(\mathbb{R})$.

Proof. By definition we have $i_m \geq 1$. We deal with the following two possible cases separately.
Case: $i_m \geq 2$
This is the case when $d_m(A) = d_{m-1}(A)$. This implies

$$i_{m-1} = i_m - 1, j_{m-1} = j_m + 1, r_{m-1} = r_m.$$

Therefore we have $m - i_m = (m-1) - i_{m-1}$. From Theorem 3.4.1 we get,

$$\begin{aligned} d_m'(A;B) &= \frac{1}{2}\sigma_{m-1}'(A;B) - \frac{1}{2}\sigma_m'(A;B) \\ &= \frac{1}{2}(\operatorname{tr} \overline{B} + \sum_{j=1}^{i_m-1} \lambda_j^{\downarrow}(\widetilde{\widetilde{B}})) - \frac{1}{2}(\operatorname{tr} \overline{B} + \sum_{j=1}^{i_m} \lambda_j^{\downarrow}(\widetilde{\widetilde{B}})) \\ &= -\frac{1}{2}\lambda_{i_m}^{\downarrow}(\widetilde{\widetilde{B}}). \end{aligned}$$

Case: $i_m = 1$
In this case we have $d_{m-1}(A) < d_m(A)$. By Corollary 3.4.2 the map σ_{m-1} is Gâteaux differentiable at A and we have

$$\nabla \sigma_{m-1}(A) = -SS^T,$$

where S is the submatrix consisting of columns with indices $1, \ldots, (m-1) + j_{m-1}$ of M. But here we have $j_{m-1} = 0$ which means that $(m-1) + j_{m-1} = m - i_m$. In other words, we have

$S = \overline{M}$. This gives

$$\begin{aligned}\sigma'_{m-1}(A;B) &= \nabla\sigma_{m-1}(A)(B)\\ &= \langle -\overline{M}\overline{M}^T, B\rangle\\ &= \text{tr}\,(-\overline{M}\overline{M}^T B)\\ &= \text{tr}\,(-\overline{M}^T B\overline{M})\\ &= \text{tr}\,\overline{B}.\end{aligned}$$

By Theorem 3.4.1 we have

$$\sigma'_m(A;B) = \sigma'_{m-1}(A;B) + \lambda_1^{\downarrow}(\widetilde{\widetilde{B}}).$$

Therefore we get

$$2d'_m(A;B) = -\lambda_1^{\downarrow}(\widetilde{\widetilde{B}})$$

which is the same as (3.20) for $i_m = 1$. □

Suppose $d_m(A)$ is a simple symplectic eigenvalue of A. If u, v are the mth and $(n+m)$th columns of M then we have $\widetilde{M} = [u, v]$. For any $B \in \mathbb{S}_n(\mathbb{R})$, we have

$$\widetilde{B} = -\widetilde{M}^T B\widetilde{M} = -\begin{pmatrix}\langle u, Bu\rangle & \langle u, Bv\rangle\\ \langle v, Bu\rangle & \langle v, Bv\rangle\end{pmatrix}.$$

Hence $\widetilde{\widetilde{B}} = -\langle u, Bu\rangle - \langle v, Bv\rangle$ is a scalar. By (3.20) we thus have

$$d'_m(A;B) = \frac{\langle u, Bu\rangle + \langle v, Bv\rangle}{2}.$$

This verifies the fact that the directional derivative of d_m at A reduces to its derivative (2.3.5) whenever $d_m(A)$ is a simple symplectic eigenvalue.

Chapter 4

Clarke and Michel-Penot subdifferentials of symplectic eigenvalues

The theory of Fenchel subdifferentials for the class of convex functions has lead to new insights and inspire new methods in nonsmooth analysis and optimization. A much successful attempt in generalising the theory of subdifferentials to a larger class of functions has been for *locally Lipschitz* functions and the ideas of F. H. Clarke have played a pioneering role. In 1983, F. H. Clarke [18] introduced a notion of *generalised gradient* for locally Lipschitz functions known as *Clarke subdifferential*. Let X be a real Banach space and ϕ be a locally Lipschitz real valued function on X. The Clarke subdifferential of ϕ at x, denoted by $\partial^\circ\phi(x)$, is a nonempty, convex, weak* compact set. Clarke subdifferential has found applications in various fields such as optimal controls and mathematical programming [19]. It is also used in the geometry of Banach spaces in studying the differentiability of distance functions [27], and in establishing optimality conditions for weak efficient, global efficient and efficient solutions in vector variational inequalities [30]. In 1992, Michel and Penot [54] introduced another notion of generalised gradient called *Michel-Penot subdifferential*. We denote by $\partial^\diamond\phi(x)$ the Michel-Penot subdifferential of ϕ at x which is also a nonempty, convex, weak* compact set and $\partial^\diamond\phi(x) \subseteq \partial^\circ\phi(x)$. Michel-Penot subdifferential has applications in optimization problems such as the steepest descent direction and stochastic programming [15], composite nonsmooth programming [40], in global convergence of Newton's method for nonsmooth equations [61], optimality conditions in nonsmooth optimal control [71].

The main feature of Michel-Penot subdifferential lies in the fact that it is smaller than Clarke subdifferential and hence easier to compute. In particular, if ϕ is Gâteaux differentiable at x then $\partial^\diamond\phi(x)$ is the singleton set containing the Gâteaux derivative of ϕ at x but $\partial^\circ\phi(x)$ may contain more elements. An advantage of Clarke subdifferential over Michel-Penot subdifferential is that the map $x \mapsto \partial^\circ\phi(x)$ is *upper semicontinuous* whereas $x \mapsto \partial^\diamond\phi(x)$ is not. Michel-Penot subdifferential can not be used to extend the Lagrange multiplier rule for nonsmooth mathematical programming involving equality constraints due to the lack of upper semicontinuity which is needed to handle the equality constraints. See [59]. Therefore it is fruitful to study both the subdifferentials in order to develop a richer theory for nonsmooth functions. The aim of this chapter is to discuss the Clarke and Michel-Penot subdifferentials of the symplectic eigenvalue maps.

The work of Hiriart-Urruty and Lewis [33] on Clarke and Michel-Penot subdifferentials for eigenvalues of symmetric matrices was a motivation for our work. Let $B \in \mathbb{S}_n(\mathbb{R})$ and $m \leq n$ be a positive integer. Denote by $E_m(B)$ the eigenspace of B corresponding to the mth largest eigenvalue $\lambda_m^\downarrow(B)$. We have

$$\partial^\circ\lambda_m^\downarrow(B) = \partial^\diamond\lambda_m^\downarrow(B) = \text{conv}\{xx^T : x \in E_m(B), \|x\| = 1\}.$$

See ([33], Theorem 5.1). In particular, the Clarke and Michel-Penot subdifferentials of $\lambda_m^\downarrow$ coincide at B and are independent of the choice of m corresponding to the equal eigenvalues of B. We prove an analogous result for symplectic eigenvalues in Theorem 4.2.3.

The chapter is organised as follows. We summarise definitions and some basic theory of Clarke and Michel-Penot subdifferentials for locally Lipschitz functions in Section 4.1. In Section 4.2, we derive expressions for the Clarke and Michel-Penot subdifferentials of the symplectic eigenvalue maps and show that both the subdifferentials coincide at every element of $\mathbb{P}_{2n}(\mathbb{R})$. As an application of Clarke and Michel-Penot subdifferentials, we give a new proof of the well known monotonicity property of symplectic eigenvalues.

4.1 Clarke and Michel-Penot subdifferentials of locally Lipschitz functions

Let X be a real Banach space and O be an open subset of X. A function $f : O \to \mathbb{R}$ is said to be locally Lipschitz at $x \in O$ if there exist $K > 0$ and $r > 0$ such that

$$|f(y) - f(z)| \leq K\|y - z\|$$

for all $y, z \in O$ satisfying $\|y - x\| < r$, $\|z - x\| < r$. If f is locally Lipschitz at every element of O it is said to be locally Lipschitz on O.

Let $f : O \to \mathbb{R}$ be locally Lipschitz on O and x be any element of O. The Clarke directional derivative of f at x in the direction $d \in X$ is defined by

$$f^\circ(x; d) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + td) - f(y)}{t}. \tag{4.1}$$

The quotient in (4.1) is bounded by $K\|d\|$ for y near x and small t which implies that the quantity $f^\circ(x; d)$ is finite. The Clarke subdifferential of f at x is defined as

$$\partial^\circ f(x) = \{y \in X^* : \langle y, h\rangle \leq f^\circ(x; h) \;\; \forall h \in X\},$$

where X^* is the dual of X and $\langle y, h\rangle = y(h)$. The Michel-Penot directional derivative of f at x in the direction d is defined by

$$f^\diamond(x; d) = \sup_{y \in \mathbb{R}^n} \limsup_{t \downarrow 0} \frac{f(x + ty + td) - f(x + ty)}{t}.$$

The locally Lipschitz property of f ensures that quantity $f^\diamond(x; d)$ is finite. The Michel-Penot subdifferential of f at x is defined by

$$\partial^\diamond f(x) = \{y \in X^* : \langle y, h\rangle \leq f^\diamond(x; h) \;\; \forall h \in X\}. \tag{4.2}$$

The following result gives a relationship between the Clarke and Michel-Penot subdifferentials. See ([17], Corollary 6.1.2).

Theorem 4.1.1. *Let $\phi : O \to \mathbb{R}$ be a function locally Lipschitz on O. Given any $x \in O$ we have*

$$\partial^{\diamond}\phi(x) \subseteq \partial^{\circ}\phi(x). \tag{4.3}$$

Furthermore, Clarke and Michel-Penot directional derivatives exhibit analogues of the Max formula given by

$$\phi^{\circ}(x; d) = \max\{\langle y, d\rangle : y \in \partial^{\circ}\phi(x)\}, \tag{4.4}$$

$$\phi^{\diamond}(x; d) = \max\{\langle y, d\rangle : y \in \partial^{\diamond}\phi(x)\} \tag{4.5}$$

for all $d \in X$.

The Clarke and Michel-Penot subdifferentials also exhibit nice properties for scalar multiples of locally Lipschitz functions. The proof of the following theorem can be found in [19, 55].

Theorem 4.1.2. *Let $\phi : O \to \mathbb{R}$ be locally Lipschitz on O. Given any $x \in O$ and any real number r,*

$$\partial^{\diamond}(r\phi)(x) = r\partial^{\diamond}\phi(x), \tag{4.6}$$

$$\partial^{\circ}(r\phi)(x) = r\partial^{\circ}\phi(x). \tag{4.7}$$

It can be easily verified for convex functions that both Clarke and Michel-Penot directional derivatives are the same as the usual directional derivative. So Clarke, Michel-Penot and Fenchel subdifferentials coincide for convex functions. See ([19], Proposition 2.3.6).

The following example illustrates the limitation of Fenchel subdifferential theory. It fails to give any useful information about the directional differentiability of functions as simple as the negative of the modulus function.

Example 20. Let $g(x) = -|x|$ for all $x \in \mathbb{R}$. A simple calculation shows that $\alpha \in \partial g(0)$ must satisfy $\alpha\beta \leq -|\beta|$ for all $\beta \in \mathbb{R}$. But this implies $\alpha \leq -1$ and $\alpha \geq 1$ which is absurd. Thus we have $\partial g(0) = \emptyset$. One can easily compute the Clarke and Michel-Penot directional derivatives of this function and get

$$g^{\circ}(0; \alpha) = g^{\diamond}(0; \alpha) = |\alpha|$$

for all $\alpha \in \mathbb{R}$. This gives $\partial^{\circ} g(0) = \partial^{\diamond} g(0) = [-1, 1]$.

The Clarke and Michel-Penot subdifferentials turn out to be equal in Example 20 but they may be greatly different. This is illustrated by the following example of [30].

Example 21. Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} x^2 |\cos\left(\frac{\pi}{x}\right)| & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Here $f^\diamond(0;\alpha) = 0$ and $f^\circ(0;\alpha) = \pi|\alpha|$ for all $\alpha \in \mathbb{R}$. This gives $\partial^\diamond f(0) = \{0\}$ and $\partial^\circ f(0) = [-\pi, \pi]$.

The following simple properties of Clarke and Michel-Penot directional derivatives are useful later in the chapter.

Proposition 4.1.3. *Let $\phi : O \to \mathbb{R}$ be locally Lipschitz on O. Let $x \in O$ be fixed and $d \in X$ be arbitrary. If ϕ is directionally differentiable at x then*

$$\phi^\diamond(x;d) = \sup_{y \in X}\{\phi'(x;d+y) - \phi'(x;d)\}. \tag{4.8}$$

Furthermore, if ϕ is directionally differentiable in an open neighbourhood of x then

$$\phi^\circ(x;d) = \limsup_{y \to x} \phi'(y;d). \tag{4.9}$$

Proof. For arbitrary $y \in X$ and $t > 0$ we have

$$\frac{\phi(x+ty+td) - \phi(x+ty)}{t} = \frac{\phi(x+ty+td) - \phi(x)}{t} - \frac{\phi(x+ty) - \phi(x)}{t}.$$

Taking $\limsup$ on both the sides as $t \downarrow 0$ and using the fact that the directional derivative of ϕ exists at x, we get

$$\limsup_{t \downarrow 0} \frac{\phi(x+ty+td) - \phi(x+ty)}{t} = \phi'(x;y+d) - \phi'(x;d).$$

By taking supremum over $y \in X$ we get (4.8).

For the second part, we assume without loss that ϕ is directionally differentiable on O. Let $\epsilon > 0$ be arbitrary. By the definition of $\limsup$, for any $\delta > 0$ and $p \in \mathbb{N}$ we get $y_0 \in O$ with $\|x - y_0\| < \delta$ and $t_p \in (0, \frac{1}{p})$ such that

$$\phi^\circ(x;d) - \epsilon < \frac{\phi(y_0 + t_p d) - \phi(y_0)}{t_p}.$$

Since $\phi'(y_0;d)$ exists, by taking the limit $p \to \infty$ we get

$$\phi^\circ(x;d) - \epsilon \leq \phi'(y_0;d). \tag{4.10}$$

Again, by definition of $\limsup$ we get $\delta_0 > 0$ and $t_0 > 0$ such that for all $y \in O$ with $\|x-y\| < \delta_0$ and $t \in (0, t_0)$ we have

$$\frac{\phi(y+td)-\phi(y)}{t} < \phi^\circ(x;d)+\epsilon.$$

Taking the limit $t \to 0$ in the above inequality we get

$$\phi'(y;d) \leq \phi^\circ(x;d)+\epsilon. \tag{4.11}$$

The conditions (4.10) and (4.11) imply (4.9).

□

4.2 Clarke and Michel-Penot subdifferentials of symplectic eigenvalues

We know by (7) that the symplectic eigenvalue maps are locally Lipschitz on $\mathbb{P}_{2n}(\mathbb{R})$. We use the theory of locally Lipschitz functions discussed in Section 4.1 to study the Clarke and Michel-Penot subdifferentials of symplectic eigenvalues. The results of this section are based on our work in Section 4 of [56].

Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. We denote by $S_m(A)$ the set of normalised symplectic eigenvector pairs (u,v) of A corresponding to the symplectic eigenvalue $d_m(A)$. Let $\widehat{m}$ be the index of the smallest symplectic eigenvalue of A equal to $d_m(A)$. More precisely, $d_j(A) = d_m(A)$ implies $j \geq \widehat{m}$. Let $M \in Sp(2n, A)$ be fixed and $M = \overline{M} \diamond \widetilde{M} \diamond \widehat{M}$ be the symplectic column partition of M of order $(m-i_m, r_m, n-m-j_m)$. For any $B \in \mathbb{S}_{2n}(\mathbb{R})$ recall that $\widetilde{\widetilde{B}}$ is the $r_m \times r_m$ Hermitian matrix given as follows. Let $\widetilde{B} = -\widetilde{M}^T B \widetilde{M}$ in the block matrix form be given by

$$\widetilde{B} = \begin{pmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{12}^T & \tilde{B}_{22} \end{pmatrix},$$

where each block is of order $r_m \times r_m$. The matrix $\widetilde{\widetilde{B}}$ is given by

$$\widetilde{\widetilde{B}} = \tilde{B}_{11} + \tilde{B}_{22} + \iota(\tilde{B}_{12} - \tilde{B}_{12}^T).$$

The following result plays a key role in deriving the expression for the Michel-Penot subdifferential of symplectic eigenvalues. In the proof of the result, we use the fact that if f is a sublinear map on X then we have

$$f(d) = \sup\{\langle x, d\rangle : x \in \partial f(0)\} \tag{4.12}$$

for all $d \in X$. See ([35], p.168, Remark 1.2.3). To keep notations simple, we use $\langle\cdot,\cdot\rangle$ to denote any inner product. Their meanings become clear from the context.

Proposition 4.2.1. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. The function $-d'_{\widehat{m}}(A;\cdot)$ is sublinear and its Fenchel subdifferential at zero is given by*

$$\partial(-d'_{\widehat{m}}(A;\cdot))(0) = conv\{-\frac{1}{2}(xx^T + yy^T) : (x,y) \in S_m(A)\}.$$

Proof. By definition we have $i_{\widehat{m}} = 1$. Therefore by Theorem 3.4.3 we have

$$-d'_{\widehat{m}}(A;B) = \frac{1}{2}\lambda_1^{\downarrow}(\widetilde{\widetilde{B}})$$

for all $B \in \mathbb{S}_{2n}(\mathbb{R})$. The map $B \mapsto \widetilde{\widetilde{B}}$ is linear and the largest eigenvalue map $\lambda_1^{\downarrow}$ is sublinear. Therefore $-d'_{\widehat{m}}(A;\cdot)$ is a sublinear map. By the property (4.12) of sublinear maps, it suffices to show that

$$-d'_{\widehat{m}}(A;B) = \max\{-\frac{1}{2}\langle xx^T + yy^T, B\rangle : (x,y) \in S_m(A)\} \tag{4.13}$$

for all $B \in \mathbb{S}_{2n}(\mathbb{R})$. Let $(x,y) \in S_m(A)$ be arbitrary. By Corollary 1.2.10 extend $[x,y]$ to S in $Sp(2n, 2r_m)$ with columns consisting of symplectic eigenvector pairs of A corresponding to $d_m(A)$. We get a $2r_m \times 2r_m$ orthosymplectic matrix Q by Corollary 1.3.3 such that $S = \widetilde{M}Q$. We know that Q is of the form

$$\begin{pmatrix} U & V \\ -V & U \end{pmatrix},$$

where U, V are $r_m \times r_m$ real matrices such that $U + \iota V$ is unitary. Let u be the first column of U and v be the first column of V. This implies

$$[x,y] = \widetilde{M}\begin{pmatrix} u & v \\ -v & u \end{pmatrix}. \tag{4.14}$$

Conversely, if $u + \iota v$ is a unit vector in $\mathbb{C}^{r_m}$ and $x, y \in \mathbb{R}^{2n}$ satisfy the above relation (4.14), then $(x,y) \in S_m(A)$. Therefore (4.14) gives a one to one correspondence $(x,y) \mapsto u + \iota v$ between $S_m(A)$ and the set of unit vectors in $\mathbb{C}^{r_m}$. We consider $\mathbb{C}^{r_m}$ equipped with the usual inner product $\langle z, w\rangle = z^* w$ for all $z, w \in \mathbb{C}^{r_m}$. We have

$$\begin{aligned}
-\frac{1}{2}\langle xx^T + yy^T, B\rangle &= -\frac{1}{2}\langle [x,y][x,y]^T, B\rangle \\
&= -\frac{1}{2}\text{tr}[x,y]^T B[x,y] \\
&= -\frac{1}{2}\text{tr}\begin{pmatrix} u & v \\ -v & u\end{pmatrix}^T \widetilde{M}^T B \widetilde{M} \begin{pmatrix} u & v \\ -v & u\end{pmatrix} \\
&= \frac{1}{2}\text{tr}\begin{pmatrix} u & v \\ -v & u\end{pmatrix}^T \widetilde{B} \begin{pmatrix} u & v \\ -v & u\end{pmatrix} \\
&= \frac{1}{2}(u+\iota v)^* \widetilde{\widetilde{B}}(u+\iota v) \\
&= \frac{1}{2}\langle u+\iota v, \widetilde{\widetilde{B}}(u+\iota v)\rangle.
\end{aligned}$$

Therefore we get

$$\begin{aligned}
-d'_{\widehat{m}}(A;B) &= \frac{1}{2}\lambda_1^{\downarrow}(\widetilde{\widetilde{B}}) \\
&= \frac{1}{2}\max\{\langle u+\iota v, \widetilde{\widetilde{B}}(u+\iota v)\rangle : \|u+\iota v\| = 1\} \\
&= \max\{-\frac{1}{2}\langle xx^T + yy^T, B\rangle : (x,y) \in S_m(A)\}.
\end{aligned}$$

The last equality follows from the above observation that (4.14) is a one to one correspondence between $S_m(A)$ and the set of unit vectors in $\mathbb{C}^{r_m}$. □

We give the expression for $\partial^\diamond d_m(A)$ in the following theorem and show that it is independent of the choice of m corresponding to equal symplectic eigenvalues of A.

Theorem 4.2.2. *The Michel-Penot subdifferentials of d_m coincide at A for all the choices of m corresponding to the equal symplectic eigenvalues of A, and are given by*

$$\partial^\diamond d_m(A) = -\partial(-d'_{\widehat{m}}(A;\cdot))(0).$$

Proof. By the property (4.6) we have $-\partial^\diamond d_m(A) = \partial^\diamond(-d_m)(A)$. So, it is equivalent to showing that

$$\partial^\diamond(-d_m)(A) = \partial(-d'_{\widehat{m}}(A;\cdot))(0).$$

We know that $-d'_{\widehat{m}}(A;\cdot)$ is convex and takes value zero at zero. By Proposition 3.1.6 of [17] we have

$$\partial(-d'_{\widehat{m}}(A;\cdot))(0) = \text{conv}\{B \in \mathbb{S}_{2n}(\mathbb{R}) : \langle B, H\rangle \le -d'_{\widehat{m}}(A;H) \ \ \forall H \in \mathbb{S}_{2n}(\mathbb{R})\}.$$

By the definition of Michel-Penot subdifferential it theorefore suffices to show that $(-d_m)^{\diamond}(A;B) = -d'_{\widehat{m}}(A;B)$ for all B in $\mathbb{S}_{2n}(\mathbb{R})$. We know that $-d_m$ is directionally differentiable at A. Therefore by the property (4.8) it is equivalent to showing

$$\sup_{H\in\mathbb{S}_{2n}(\mathbb{R})}\{-d'_m(A;B+H)+d'_m(A;H)\} = -d'_{\widehat{m}}(A;B). \tag{4.15}$$

Let B,H be elements of $\mathbb{S}_{2n}(\mathbb{R})$. It is easy to see that

$$\widetilde{\widetilde{B+H}} = \widetilde{\widetilde{B}} + \widetilde{\widetilde{H}}.$$

By Theorem 3.4.3 we get

$$\begin{aligned}-d'_m(A;B+H)+d'_m(A;H) &= \frac{1}{2}\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{B+H}}) - \frac{1}{2}\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{H}})\\ &= \frac{1}{2}\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{B}}+\widetilde{\widetilde{H}}) - \frac{1}{2}\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{H}}).\end{aligned}$$

It can be verified that $\{\widetilde{\widetilde{H}} : H \in \mathbb{S}_{2n}(\mathbb{R})\} = \mathbb{H}_{r_m}(\mathbb{C})$. Also, by Theorem 3.4.3 we have $-d'_{\widehat{m}}(A;B) = \frac{1}{2}\lambda^{\downarrow}_1(\widetilde{\widetilde{B}})$. By (4.15) we thus need to show that

$$\sup_{C\in\mathbb{H}_{r_m}(\mathbb{C})}\{\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{B}}+C) - \lambda^{\downarrow}_{i_m}(C)\} = \lambda^{\downarrow}_1(\widetilde{\widetilde{B}}). \tag{4.16}$$

The inequality

$$\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{B}}+C) - \lambda^{\downarrow}_{i_m}(C) \leq \lambda^{\downarrow}_1(\widetilde{\widetilde{B}}) \tag{4.17}$$

is one of the Weyl's inequalities. See ([9], Corollary III.2.2). To see the equality in (4.17) is attained, choose a unitary matrix U such that $U^T\widetilde{\widetilde{B}}U$ is the diagonal matrix $\mathrm{Diag}(\lambda^{\downarrow}_1(\widetilde{\widetilde{B}}),\ldots,\lambda^{\downarrow}_{r_m}(\widetilde{\widetilde{B}}))$. Let $\alpha > \lambda^{\downarrow}_1(\widetilde{\widetilde{B}}) - \lambda^{\downarrow}_{r_m}(\widetilde{\widetilde{B}})$ and C be the Hermitian matrix given by

$$C = U^T\mathrm{Diag}(0,\ldots,0,\underbrace{\alpha,\ldots,\alpha}_{(i_m-1)\text{ times}})U.$$

This gives

$$\widetilde{\widetilde{B}}+C = U^T\mathrm{Diag}(\lambda^{\downarrow}_1(\widetilde{\widetilde{B}}),\ldots\lambda^{\downarrow}_{r_m-i_m+1}(\widetilde{\widetilde{B}}),\lambda^{\downarrow}_{r_m-i_m+2}(\widetilde{\widetilde{B}})+\alpha,\ldots,\lambda^{\downarrow}_{r_m}(\widetilde{\widetilde{B}})+\alpha)U.$$

We then have $\lambda^{\downarrow}_{i_m}(C) = 0$ and $\lambda^{\downarrow}_{i_m}(\widetilde{\widetilde{B}}+C) = \lambda^{\downarrow}_1(\widetilde{\widetilde{B}})$. This proves the equality (4.16). $\square$

We now give the main result of this section. The highlight of the result is the coincidence of the Clarke and Michel-Penot subdifferentials of symplectic eigenvalues. In particular, both the subdifferentials are independent of the choice of index corresponding to equal symplectic eigenvalues.

Theorem 4.2.3. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. The Clarke and Michel-Penot subdifferentials of d_m are equal at A and they are given by*

$$\partial^{\circ} d_m(A) = \partial^{\diamond} d_m(A) = conv\{\frac{1}{2}(xx^T + yy^T) : (x, y) \in S_m(A)\}.$$

In particular, the subdifferentials are independent of the choice of m corresponding to equal symplectic eigenvalues of A.

Proof. By Proposition 4.2.1 and Theorem 4.2.2 we have

$$\partial^{\diamond} d_m(A) = \text{conv}\{\frac{1}{2}(xx^T + yy^T) : (x, y) \in S_m(A)\}.$$

By the relation (4.3) we have $\partial^{\diamond} d_m(A) \subseteq \partial^{\circ} d_m(A)$. Therefore it only remains to prove that $\partial^{\circ} d_m(A) \subseteq \partial^{\diamond} d_m(A)$. Equivalently, using the properties (4.6), (4.7), we prove that $\partial^{\circ}(-d_m)(A) \subseteq \partial^{\diamond}(-d_m)(A)$.

Let B in $\mathbb{S}_{2n}(\mathbb{R})$ be fixed. By the property (4.9) we get a sequence $A_{(p)} \in \mathbb{P}_{2n}(\mathbb{R})$ for $p \in \mathbb{N}$ such that $\lim_{p\to\infty} A_{(p)} = A$ and

$$(-d_m)^{\circ}(A; B) = -\lim_{p\to\infty} d_m'(A_{(p)}; B). \tag{4.18}$$

Let $\mathcal{I}_p = \{i : d_i(A_{(p)}) = d_m(A_{(p)})\}$ for every $p \in \mathbb{N}$. There are only finitely many choices for $\mathcal{I}_p$ for each p. Therefore we can get a subsequence of $(A_{(p)})_{p\in\mathbb{N}}$ such that $\mathcal{I}_p$ is independent of p. Let us denote the subsequence by the same sequence $(A_{(p)})_{p\in\mathbb{N}}$ for convenience and let $\mathcal{I}$ denote the common index set $\mathcal{I}_p$. Let $M_{(p)}$ be an element of $Sp(2n, A_{(p)})$ for all $p \in \mathbb{N}$. If (u, v) is a normalized symplectic eigenvector pair of $A_{(p)}$ corresponding to a symplectic eigenvalue d, we get

$$\begin{aligned}
\|u\|^2 + \|v\|^2 &\leq \|A_{(p)}^{-1}\|(\|(A_{(p)}^{1/2}u\|^2 + \|A_{(p)}^{1/2}v\|^2) \\
&= \|A_{(p)}^{-1}\| \cdot \|A_{(p)}^{1/2}u - \iota A_{(p)}^{1/2}v\|^2 \\
&= 2d\langle u, Jv\rangle \|A_{(p)}^{-1}\| \\
&= 2d\|A_{(p)}^{-1}\| \\
&\leq 2\|A_{(p)}\| \cdot \|A_{(p)}^{-1}\| \\
&= 2\kappa(A_{(p)}),
\end{aligned}$$

where $\|A_{(p)}\|$ and $\|A_{(p)}^{-1}\|$ represent the operator norms of $A_{(p)}$ and $A_{(p)}^{-1}$, and $\kappa(A_{(p)})$ is the condition number of $A_{(p)}$. The second equality follows from Proposition 1.2.7, and the second inequality follows from the fact that $d \leq \|A_{(p)}\|$. Therefore we have

$$\|M_{(p)}\|_F^2 \leq 2n\kappa(A_{(p)}), \tag{4.19}$$

where $\|M_{(p)}\|_F$ represents the Frobenius norm of $M_{(p)}$ for all $p \in \mathbb{N}$. We know that κ is a continuous function and the sequence $(A_{(p)})_{p\in\mathbb{N}}$ is convergent. Therefore the sequence $(\kappa(A_{(p)}))_{p\in\mathbb{N}}$ is also convergent, and hence bounded. By (4.19) the sequence $(M_{(p)})_{p\in\mathbb{N}}$ of $2n \times 2n$ real matrices is bounded as well. By taking a subsequence we can assume that $(M_{(p)})_{p\in\mathbb{N}}$ converges to some $2n \times 2n$ real matrix M. The set $Sp(2n)$ is closed, therefore $M \in Sp(2n)$. By continuity of the symplectic eigenvalue maps we also have $M \in Sp(2n, A)$.

Let $m_1 = \min \mathcal{I}$ and $m_2 = \max \mathcal{I}$ and $M_{(p)} = \overline{M}_{(p)} \diamond \widetilde{M}_{(p)} \diamond \widehat{M}_{(p)}$ be the symplectic column partition of $M_{(p)}$ of order $(m_1 - 1, m_2 - m_1 + 1, n - m_2)$. Let

$$\widetilde{B}_{(p)} = -\widetilde{M}_{(p)}^T B \widetilde{M}_{(p)},$$

$$\widetilde{M}_{(0)} = \lim_{p\to\infty} \widetilde{M}_{(p)}$$

and

$$\widetilde{B}_{(0)} = \lim_{p\to\infty} \widetilde{B}_{(p)} = -\widetilde{M}_{(0)}^T B \widetilde{M}_{(0)}. \tag{4.20}$$

Consider the block matrix form of $\widetilde{B}_{(p)}$ given by

$$\widetilde{B}_{(p)} = \begin{pmatrix} (\widetilde{B}_{(p)})_{11} & (\widetilde{B}_{(p)})_{12} \\ (\widetilde{B}_{(p)})_{12}^T & (\widetilde{B}_{(p)})_{22} \end{pmatrix},$$

where each block has size $m_2 - m_1 + 1$. Let

$$\widetilde{\widetilde{B}}_{(p)} = (\widetilde{B}_{(p)})_{11} + (\widetilde{B}_{(p)})_{22} + \iota((\widetilde{B}_{(p)})_{12} - (\widetilde{B}_{(p)})_{12}^T) \tag{4.21}$$

be the Hermitian matrix associated with $\widetilde{B}_{(p)}$ and define

$$\widetilde{\widetilde{B}}_{(0)} = \lim_{p\to\infty} \widetilde{\widetilde{B}}_{(p)}.$$

Recall that $M = \overline{M} \diamond \widetilde{M} \diamond \widehat{M}$ is the symplectic column partition of M of order $(m - i_m, r_m, n - m - j_m)$. The matrix $\widetilde{M}_{(0)}$ is the submatrix of M consisting of the ith and $(n+i)$th columns of M for all $i \in \mathcal{I}$. By continuity of the symplectic eigenvalues we have $\mathcal{I} \subseteq \{m - i_m + 1, m - i_m + 2, \ldots, m + j_m\}$. Therefore $\widetilde{M}_{(0)}$ is also a submatrix of $\widetilde{M}$. It thus follows by relation (4.20) that each block of $\widetilde{B}_{(0)}$ is obtained by removing ith row and ith column of $\widetilde{B}$ for all i not in $\mathcal{I}$. Therefore $\widetilde{\widetilde{B}}_{(0)}$ is a compression of $\widetilde{\widetilde{B}}$. By Cauchy interlacing principle ([9], Corollary III.1.5) we have

$$\lambda_1^{\downarrow}(\widetilde{\widetilde{B}}_{(p)}) \leq \lambda_1^{\downarrow}(\widetilde{\widetilde{B}}).$$

Using equation (4.18) we get

$$\begin{aligned}
(-d_m)^{\circ}(A;B) &= -\lim_{p\to\infty} d'_m(A_{(p)};B) \\
&\leq \frac{1}{2}\lim_{p\to\infty} \lambda_1^{\downarrow}(\widetilde{\widetilde{B}}_{(p)}) \\
&= \frac{1}{2}\lambda_1^{\downarrow}(\lim_{p\to\infty} \widetilde{\widetilde{B}}_{(p)}) \\
&= \frac{1}{2}\lambda_1^{\downarrow}(\widetilde{\widetilde{B}}_{(0)}) \\
&\leq \frac{1}{2}\lambda_1^{\downarrow}(\widetilde{\widetilde{B}}) \\
&= -d'_{\widehat{m}}(A;B).
\end{aligned}$$

Thus we have proved that $(-d_m)^{\circ}(A;B) \leq -d'_{\widehat{m}}(A;B)$ for all B in $\mathbb{S}_{2n}(\mathbb{R})$. By definition this implies $\partial^{\circ}(-d_m)(A) \subseteq \partial(-d'_{\widehat{m}}(A;\cdot))(0)$. By Theorem 4.2.2 we have $\partial^{\diamond}(-d_m)(A) = \partial(-d'_{\widehat{m}}(A;\cdot))(0)$ which implies that $\partial^{\circ}(-d_m)(A) \subseteq \partial^{\diamond}(-d_m)(A)$. □

Corollary 4.2.4. *Let A be an element of $\mathbb{P}_{2n}(\mathbb{R})$. We have*

$$d_m^{\diamond}(A;H) = d_m^{\circ}(A;H) = -d'_{\hat{m}}(A;-H).$$

for all $H \in \mathbb{S}_{2n}(\mathbb{R})$.

Proof. By Theorem 4.2.3 and the formulae (4.4), (4.5) we get $d_m^{\diamond}(A;\cdot) = d_m^{\circ}(A;\cdot)$. Also, for any $H \in \mathbb{S}_{2n}(\mathbb{R})$

$$d_m^{\diamond}(A;H) = \max\{\frac{1}{2}\langle xx^T + yy^T, H\rangle : (x,y) \in S_m(A)\}.$$

By (4.13) we have $d_m^{\diamond}(A;H) = -d'_{\hat{m}}(A;-H)$. □

As an application of Clarke and Michel-Penot subdifferentials of d_m we give a new proof of the monotonicity principle of symplectic eigenvalues. We recall a result on Clarke subdifferentials of locally Lipschitz functions known as *Lebourg mean value theorem*. Let U be an open subset of a real Banach space X and $\phi : U \to \mathbb{R}$ be a locally Lipschitz function. Let $x, y \in X$ such that $tx + (1-t)y \in U$ for all $t \in [0,1]$. Then there exists $s \in (0,1)$ and $z \in \partial^{\circ}\phi(sx + (1-s)y)$ such that $\phi(x) - \phi(y) = \langle z, x-y\rangle$. See ([47], Theorem 1.7).

Corollary 4.2.5. *For every A, B in $\mathbb{P}_{2n}(\mathbb{R})$, we have $d_j(A) \leq d_j(B)$ for all $j = 1, 2, \ldots, n$, whenever $A \leq B$.*

Proof. By Lebourg mean value theorem, let $P = sA + (1 - s)B$ for some $s \in (0, 1)$ and $C \in \partial^\circ d_m(P)$ such that

$$d_m(A) - d_m(B) = \langle C, A - B \rangle.$$

By Theorem 4.2.3 we have

$$\partial^\circ d_m(P) = \text{conv}\{\frac{1}{2}(xx^T + yy^T) : (x, y) \in S_m(P)\}.$$

Therefore we have

$$d_m(A) - d_m(B) \in \text{conv}\{\frac{1}{2}\langle xx^T + yy^T, A - B \rangle : (x, y) \in S_m(P)\}. \tag{4.22}$$

Thus, $A \leq B$ implies

$$\text{conv}\{\frac{1}{2}\langle xx^T + yy^T, A - B \rangle : (x, y) \in S_m(P)\} \subseteq (-\infty, 0].$$

By (4.22) we conclude $d_m(A) \leq d_m(B)$.

$\square$

Bibliography

[1] Houssam Abdul-Rahman. Entanglement of a class of non-Gaussian states in disordered harmonic oscillator systems. *Journal of Mathematical Physics*, 59(3):031904, 2018.

[2] Gerardo Adesso, Sammy Ragy, and Antony R Lee. Continuous variable quantum information: Gaussian states and beyond. *Open Systems & Information Dynamics*, 21(01n02):1440001, 2014.

[3] Gerardo Adesso, Alessio Serafini, and Fabrizio Illuminati. Extremal entanglement and mixedness in continuous variable systems. *Physical Review A*, 70(2):022318, 2004.

[4] Dmitri Alekseevsky, Andreas Kriegl, Peter W Michor, and Mark Losik. Choosing roots of polynomials smoothly. *Israel Journal of Mathematics*, 105(1):203–233, 1998.

[5] Tijani Amahroq, Jean-Paul Penot, and Aïcha Syam. On the subdifferentiability of the difference of two functions and local minimization. *Set-Valued Analysis*, 16(4):413–427, 2008.

[6] T. Ando. Majorization, doubly stochastic matrices, and comparison of eigenvalues. *Linear Algebra and its Applications*, 118:163 – 248, 1989.

[7] Rajendra Bhatia. *Perturbation bounds for matrix eigenvalues*, volume 53. Siam, 1987.

[8] Rajendra Bhatia. Linear algebra to quantum cohomology: the story of Alfred Horn's inequalities. *The American Mathematical Monthly*, 108(4):289–318, 2001.

[9] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[10] Rajendra Bhatia. Some inequalities for eigenvalues and symplectic eigenvalues of positive definite matrices. *International Journal of Mathematics*, 30(11):1950055, 2019.

[11] Rajendra Bhatia and Tanvi Jain. On symplectic eigenvalues of positive definite matrices. *Journal of Mathematical Physics*, 56(11):112201, 2015.

[12] Rajendra Bhatia and Tanvi Jain. A Schur-Horn theorem for symplectic eigenvalues. *Linear Algebra and its Applications*, 599:133–139, 2020.

[13] Rajendra Bhatia and Tanvi Jain. Variational principles for symplectic eigenvalues. *Canadian Mathematical Bulletin*, pages 1–7, 2020.

[14] Rajendra Bhatia, Dinesh Singh, and Kalyan B Sinha. Differentiation of operator functions and perturbation bounds. *Communications in mathematical physics*, 191(3):603–611, 1998.

[15] John R Birge and Liqun Qi. Semiregularity and generalized subdifferentials with applications to optimization. *Mathematics of Operations Research*, 18(4):982–1005, 1993.

[16] Petre Birtea, Ioan Caşu, and Dan Comănescu. Optimization on the real symplectic group. *Monatshefte für Mathematik*, 191(3):465–485, 2020.

[17] Jonathan Borwein and Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[18] Frank H Clarke. Nonsmooth analysis and optimization. In *Proceedings of the international congress of mathematicians*, volume 5, pages 847–853. Citeseer, 1983.

[19] Frank H Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

[20] Jane Cullum, Wilm E Donath, and Philip Wolfe. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. In *Nondifferentiable optimization*, pages 35–55. Springer, 1975.

[21] Maurice A De Gosson. *Symplectic geometry and quantum mechanics*, volume 166. Springer Science & Business Media, 2006.

[22] Pierre Del Moral and Angele Niclas. A Taylor expansion of the square root matrix function. *Journal of Mathematical Analysis and Applications*, 465(1):259–266, 2018.

[23] Biswadeb Dutta, N Mukunda, R Simon, et al. The real symplectic groups in quantum mechanics and optics. *Pramana*, 45(6):471–497, 1995.

[24] Jens Eisert, Tomáš Tyc, Terry Rudolph, and Barry C Sanders. Gaussian quantum marginal problem. *Communications in mathematical physics*, 280(1):263–280, 2008.

[25] Ky Fan. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences of the United States of America*, 35(11):652–655, 1949.

[26] Simone Fiori. A Riemannian steepest descent approach over the inhomogeneous symplectic group: Application to the averaging of linear optical systems. *Applied Mathematics and Computation*, 283:251–264, 2016.

[27] Simon Fitzpatrick. Metric projections and the differentiability of distance functions. *Bulletin of the Australian Mathematical Society*, 22(2):291–312, 1980.

[28] William Fulton. Eigenvalues of sums of Hermitian matrices. *Séminaire Bourbaki*, 40:255–269, 1998.

[29] Victor Guillemin and Shlomo Sternberg. *Symplectic techniques in physics*. Cambridge university press, 1990.

[30] Dinh Dieu Hang et al. On optimality conditions for vector variational inequalities. *Journal of Mathematical Analysis and Applications*, 412(2):792–804, 2014.

[31] Edward J Haug and Bernard Rousselet. Design sensitivity analysis in structural mechanics. II. Eigenvalue variations. *Journal of Structural Mechanics*, 8(2):161–186, 1980.

[32] J-B Hiriart-Urruty. Generalized differentiability/duality and optimization for problems dealing with differences of convex functions. In *Convexity and duality in optimization*, pages 37–70. Springer, 1985.

[33] J-B Hiriart-Urruty and AS Lewis. The Clarke and Michel-Penot subdifferentials of the eigenvalues of a symmetric matrix. *Computational Optimization and Applications*, 13(1-3):13–23, 1999.

[34] J-B Hiriart-Urruty and D Ye. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70(1):45–72, 1995.

[35] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.

[36] Tohya Hiroshima. Additivity and multiplicativity properties of some Gaussian channels for Gaussian inputs. *Physical Review A*, 73(1):012330, 2006.

[37] Helmut Hofer and Eduard Zehnder. *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser, 2012.

[38] Martin Idel, Sebastián Soto Gaona, and Michael M Wolf. Perturbation bounds for Williamson's symplectic normal form. *Linear Algebra and its Applications*, 525:45–58, 2017.

[39] Tanvi Jain and Hemant Kumar Mishra. Derivatives of symplectic eigenvalues and a Lidskii type theorem. *Canadian Journal of Mathematics*, pages 1–29, 2020.

[40] Vaithilingam Jeyakumar. Composite nonsmooth programming with Gateaux differentiability. *SIAM Journal on Optimization*, 1(1):30–41, 1991.

[41] Charles R Johnson and Roger A Horn. *Matrix analysis*. Cambridge university press Cambridge, 1985.

[42] Jer-Nan Juang, Peiman Ghaemmaghami, and Kyong Been Lim. Eigenvalue and eigenvector derivatives of a nondefective matrix. *Journal of Guidance, Control, and Dynamics*, 12(4):480–486, 1989.

[43] Tosio Kato. *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media, 2013.

[44] Jerry L Kazdan. Matrices $A(t)$ depending on a parameter t. *unpublished note*, 1995.

[45] Robert Koenig. The conditional entropy power inequality for Gaussian quantum states. *Journal of Mathematical Physics*, 56(2):022201, 2015.

[46] Vadim Komkov, Kyung K Choi, and Edward J Haug. *Design sensitivity analysis of structural systems*, volume 177. Academic press, 1986.

[47] Gérard Lebourg. Generic differentiability of Lipschitzian functions. *Transactions of the American Mathematical Society*, 256:125–144, 1979.

[48] Chi-Kwong Li and Roy Mathias. The Lidskii-Mirsky-Wielandt theorem–additive and multiplicative versions. *Numerische Mathematik*, 81(3):377–413, 1999.

[49] Guoyin Li, Liqun Qi, and Gaohang Yu. Semismoothness of the maximum eigenvalue function of a symmetric tensor and its application. *Linear Algebra and its Applications*, 438(2):813–833, 2013.

[50] Victor B Lidskii. The proper values of the sum and product of symmetric matrices. *Doklady Akademii Nauk SSSR*, 75:769–772, 1950.

[51] Jan R Magnus. On differentiating eigenvalues and eigenvectors. *Econometric Theory*, 1(2):179–191, 1985.

[52] Şahlar F Meherrem and Refet Polat. Weak subdifferential in nonsmooth analysis and optimization. *Journal of Applied Mathematics*, 2011, 2011.

[53] Emre Mengi, E Alper Yildirim, and Mustafa Kilic. Numerical optimization of eigenvalues of Hermitian matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 35(2):699–724, 2014.

[54] Philippe Michel, Jean Paul Penot, and Haim Brezis. A generalized derivative for calm and stable functions. *Differential and Integral Equations*, 5(2):433–454, 1992.

[55] Philippe Michel, Jean Paul Penot, and Haim Brezis. A generalized derivative for calm and stable functions. *Differential and Integral Equations*, 5(2):433–454, 1992.

[56] Hemant Kumar Mishra. First order sensitivity analysis of symplectic eigenvalues. *Linear Algebra and its Applications*, 604:324–345, 2020.

[57] Kalyan Mukherjea. *Differential calculus in normed linear spaces*, volume 26. Hindustan Book Agency, 2007.

[58] Bruno Nachtergaele, Robert Sims, and Günter Stolz. Quantum harmonic oscillator systems with disorder. *Journal of Statistical Physics*, 149(6):969–1012, 2012.

[59] Triloki Nath and SR Singh. Michel-Penot subdifferential and Lagrange multiplier rule. *WSEAS Transactions on Mathematics*, 10(4):139–148, 2011.

[60] Michael L Overton. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications*, 9(2):256–268, 1988.

[61] Jong-Shi Pang and Liqun Qi. Nonsmooth equations: motivation and algorithms. *SIAM Journal on optimization*, 3(3):443–465, 1993.

[62] KR Parthasarathy. Symplectic dilations, Gaussian states and Gaussian channels. *Indian Journal of Pure and Applied Mathematics*, 46(4):419–439, 2015.

[63] Jean-Paul Penot. The directional subdifferential of the difference of two convex functions. *Journal of Global Optimization*, 49(3):505–519, 2011.

[64] Franz Rellich and Joan Berkowitz. *Perturbation theory of eigenvalue problems*. CRC Press, 1969.

[65] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[66] Lynn C Rogers. Derivatives of eigenvalues and eigenvectors. *AIAA journal*, 8(5):943–944, 1970.

[67] Dominik Šafránek and Ivette Fuentes. Optimal probe states for the estimation of Gaussian unitary channels. *Physical Review A*, 94(6):062313, 2016.

[68] Alessio Serafini. *Quantum continuous variables: a primer of theoretical methods*. CRC press, 2017.

[69] Alessio Serafini, Fabrizio Illuminati, and Silvio De Siena. Symplectic invariants, entropic measures and correlations of Gaussian states. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 37(2):L21, 2003.

[70] AP Seyranian, ON Kirillov, and AA Mailybaev. Coupling of eigenvalues of complex matrices at diabolic and exceptional points. *Journal of Physics A: Mathematical and General*, 38(8):1723, 2005.

[71] Ilya Shvartsman. Necessary optimality conditions in discrete nonsmooth optimal control. *Journal of Optimization Theory and Applications*, 153(3):578–586, 2012.

[72] R Simon, S Chaturvedi, and V Srinivasan. Congruences and canonical forms for a positive matrix: Application to the Schweinler–Wigner extremum principle. *Journal of Mathematical Physics*, 40(7):3632–3642, 1999.

[73] Mounir Torki. Second-order directional derivatives of all eigenvalues of a symmetric matrix. *Nonlinear analysis*, 46(8):1133–1150, 2001.

[74] Yongjia Xu and Yongzeng Lai. Derivatives of functions of eigenvalues and eigenvectors for symmetric matrices. *Journal of Mathematical Analysis and Applications*, 444(1):251–274, 2016.

[75] Constantin Zalinescu. *Convex analysis in general vector spaces*. World scientific, 2002.