Maximal entropy in the moment body

Didier Henrion[1,2]

Draft of July 2, 2025

Abstract

A moment body is a linear projection of the spectraplex, the convex set of trace-one positive semidefinite matrices. Determining whether a given point lies within a given moment body is a problem with numerous applications in quantum state estimation or polynomial optimization. This moment body membership oracle can be addressed with semidefinite programming, for which several off-the-shelf interior-point solvers are available. In this paper, inspired by techniques from quantum information theory, we argue analytically and geometrically that a much more efficient approach consists of minimizing globally a smooth strictly convex log-partition function, dual to a maximum entropy problem. We analyze the curvature properties of this function and we describe a neat geometric pre-conditioning algorithm. A detailed complexity analysis reveals a cubic dependence on the matrix size, similar to a few eigenstructure computations. Basic numerical experiments illustrate that dense (i.e. non-sparse) projections of size 1000 of a dense semidefinite matrix of size 1000-by-1000 can be routinely handled in a few seconds on a standard laptop, thereby moving the main bottleneck in large-scale semidefinite programming almost entirely to efficient gradient storage and manipulation.

1   Introduction

Semidefinite programming is a versatile framework for convex optimization. It consists of optimizing (typically linear functions) over spectrahedra (linear sections of the semidefinite cone, described by linear matrix inequalities) or spectrahedral shadows (linear projections of spectrahedra). These sets capture a large class of convex semialgebraic sets [3]. Polynomial optimization relies heavily on semidefinite optimization, and the moment-SOS hierarchy constructs a nested family of spectrahedral shadows of increasing size that provide increasingly tight approximations of convex hulls of semialgebraic sets, see e.g. [10, 21, 32] and references therein.

Semidefinite optimization problems can be solved with interior-point algorithms [24, 3]. However, as second-order methods, these algorithms do not scale well at the age of data science.

---

[1] CNRS; LAAS; Université de Toulouse, 7 avenue du colonel Roche, F-31400 Toulouse, France.

[2] Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czechia.

Most of the computational burden is concentrated on computing and storing the Hessian matrix of second-order derivatives of a logarithmic barrier function. First-order algorithms scale better, since they use only gradient information, but they are also more sensitive to problem scaling and conditioning. Conditioning of semidefinite optimization problems is understood theoretically [30], but evaluating the conditioning of a given problem is as expensive as solving the original problem. From that point of view, the versatily and generality of semidefinite programming can also be seen as a weakness: currently, there is no simple recipe that can be systematically used to cure all numerical issues, see e.g. [27] for a survey of recent attempts. There are at least three geometric pathologies that can occur in semidefinite programming: (i) a linear image of an unbounded spectrahedron need not be closed; (ii) a spectrahedron or its shadow can lack interior points; (iii) the linear map defining a spectrahedral shadow can be ill-conditioned (i.e. with singular values largely differing in magnitude). In this paper, we propose to focus on pathology (iii), namely ill-conditioning of the linear map, and our strategy is as follows. First, we restrict our attention to semidefinite feasibility problems whose spectrahedral shadows are full-dimensional and bounded. This eliminates the pathologies (i) and (ii). Second, we focus on analytic, quantitative aspects of a standard first-order optimization algorithm, in which issue (iii) appears explicitly through curvature parameters. This allows us to design a simple and cheap pre-conditioning algorithm.

Our focus is on the moment body membership oracle problem: finding a point in a linear projection of the spectraplex, defined as the compact convex set of trace-one positive semidefinite matrices, a non-polyhedral generalization of the simplex. Determining whether a given point lies within a given moment body is a problem with numerous applications in polynomial optimization or quantum information theory. This includes for example the problem of decomposing a given multivariate polynomial as a sum of squares (SOS) of other polynomials, see e.g. [21, Section 2.4] and references therein. In order to address this problem with a first-order algorithm, we use an approach inspired from quantum information theory [14, 13], namely the global minimization of a smooth and strictly convex log-partition function dual to a maximum entropy problem. Quantum state estimation aims to recover a density matrix (i.e. an element of the spectraplex, a trace-one positive semdefinite matrix) consistent with observed measurement statistics (i.e. the linear projection of the spectraplex) [2] - and this is exactly our moment body membership oracle problem. A particularly effective method for solving this problem consists of selecting, among all compatible density matrices, the one maximizing entropy. The dual of this problem leads to the minimization of a convex, smooth function called the log-partition function. We analyze its curvature properties, and based on geometric quantities appearing during the analysis, we describe a neat and simple geometric pre-conditioning algorithm. A detailed complexity analysis reveals a cubic dependence on the matrix size, similar to a few eigenstructure computations.

Basic numerical experiments illustrate that a rudimentary Matlab prototype can be competitive with SDPNAL+ [36, 31] a state-of-the-art solver for large-scale semidefinite programming. On a standard laptop we can solve in a few seconds the moment body membership for a dense (i.e. non-sparse) linear projection of size 1000 of a dense semidefinite matrix of size 1000-by-1000, at expected accuracy 10^{-8}. Note however that for these sizes, just storing the problem data requires almost 8 gigabytes. Practically speaking, this implies that, for this problem class, the bottleneck of large-scale semidefinite solvers is pushed further and almost

exclusively to the efficient storage and manipulation of gradient information.

1.1 Outline

The paper is organized as follows. Section 2 defines the moment body and presents a few examples to illustrate its geometry in low dimension. In Section 3 we show how testing membership in a moment body of size m defined by a spectraplex of size n-by-n can be formulated as the unconstrained minimization in $\mathbb{R}^m$ of a smooth, strictly convex log-partition function, and we prove that this dual problem is equivalent (via strong duality) to a primal maximum-entropy formulation. Section 4 is devoted to a first geometric analysis of the dual objective. We derive explicit upper and lower bounds on its Hessian in terms of the Gram matrix of the linear map defining the moment body, showing that the dual is globally λ-smooth and α-strongly convex on sublevel sets, with λ and α depending only on the spectrum of the Gram matrix. In Section 5 we present a simple preconditioning algorithm: by centering and orthonormalizing the linear map, one can force the dual to become $\frac{1}{2}$-smooth and $\frac{1}{n^3}$-strongly convex. In Section 6 we exploit these curvature estimates to bound the size of the unique minimizer in terms of the input data. Section 7 gives a detailed iteration-complexity analysis of L-BFGS when applied to the preconditioned dual. Section 8 discusses how the same dual framework detects weakly feasible points (boundary membership) and certifies strict infeasibility. Section 9 briefly explains how our analysis extends to block-separable (direct-sum) moment-body problems, in which the primal density matrix splits into several independent blocks. Finally, Section 10 presents numerical experiments on random dense instances: we compare our Matlab prototype againt off-the-shelf semidefinite solvers and demonstrate that dense problems of size $n = m = 1000$ can be solved in a few seconds on a standard laptop.

1.2 Notations

$\mathbb{S}^n$ is the space of real valued symmetric matrices of size n,

$$\mathbb{S}^n_+ := \{X \in \mathbb{S}^n, X \succeq 0\}$$

is the convex closed cone of positive semidefinite elements of $\mathbb{S}^n$, called the *semidefinite cone*. Its interior

$$\operatorname{int} \mathbb{S}^n_+ := \{X \in \mathbb{S}^n, X \succ 0\}$$

is the convex open cone of positive definite elements of $\mathbb{S}^n$, and

$$\mathbb{S}^n_1 := \{X \in \mathbb{S}^n, X \succeq 0, \operatorname{tr} X = 1\}$$

is called the *spectraplex*, a generalization to non-diagonal matrices of the polyhedral simplex. It is a spectrahedron, an affine slice of the semidefinite cone, see e.g. [21, Section 7.3].

Given a matrix X, $\log X$ denotes its logarithm, $\exp X$ its exponential, and

$$\exp_1 X := \frac{\exp X}{\operatorname{tr} \exp X}$$

is the normalized or trace-one exponential.

3

2 The moment body

Let $A_i \in \mathbb{S}^n$, $i = 1, \ldots, m$ be given matrices. Define the linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $X \mapsto \mathrm{tr}(A_i X)_{i=1,\ldots,m}$ and its adjoint $\mathcal{A}^T : \mathbb{R}^m \to \mathbb{S}^n$, $\mathbf{y} \mapsto A(\mathbf{y}) := \sum_{i=1}^m y_i A_i$. The *moment body* of $\mathcal{A}$ is the set

$$\boxed{\mathscr{M} := \{\mathcal{A}(X) : X \succeq 0, \mathrm{tr} X = 1\} \subset \mathbb{R}^m}$$

or equivalently $\mathscr{M} := \mathcal{A}(\mathbb{S}^n_1)$. In words, a moment body is the linear image of a spectraplex. As a linear projection of a convex and compact set, set $\mathscr{M}$ is also convex and compact. The terminology moment body is motivated as follows. Let $\mathscr{X}$ be a topological space, and let $\phi : \mathscr{X} \to \mathscr{U}$ be a map, where $\mathscr{U} := \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u}^T \mathbf{u} = 1\}$ is the unit sphere. For example, $\phi(\mathbf{x})$ can be the result of a measurement for $\mathbf{x} \in \mathscr{X}$ in some given set of Euclidean space, with ϕ a basis for the vector space of polynomials of $\mathbf{x}$ up to some degree. We can identify each matrix A_k with the Gram matrix of a function $a_k : \mathscr{X} \to \mathbb{R}$, $\phi(\mathbf{x}) \mapsto \phi^T(\mathbf{x}) A_k \phi(\mathbf{x})$ and then write $\mathrm{tr}(A_k X) = \int_{\mathscr{X}} a_k(\mathbf{x}) d\mu(\mathbf{x})$ where $X = \int_{\mathscr{X}} \phi(\mathbf{x}) \phi(\mathbf{x})^T d\mu(\mathbf{x})$ is the moment matrix of μ, an element of $\mathscr{P}(\mathscr{X})$, the set of probability measures on $\mathscr{X}$. Equivalently, if we define ν as the image measure of μ through ϕ, $X = \int_{\mathscr{U}} \mathbf{u}\mathbf{u}^T d\nu(\mathbf{u})$ is the covariance matrix of $\nu \in \mathscr{P}(\mathscr{U})$. Both measures satisfy $\int_{\mathscr{X}} d\mu(\mathbf{x}) = \int_{\mathscr{U}} d\nu(\mathbf{u}) = \int_{\mathscr{U}} \mathbf{u}^T \mathbf{u} \, d\mu(\mathbf{u}) = \mathrm{tr} X = 1$. The moment body is therefore the set of all moments of such probability measures, i.e.

$$\mathscr{M} = \left\{ \int_{\mathscr{X}} \mathbf{a}(\mathbf{x}) d\mu(\mathbf{x}) \; : \; \mu \in \mathscr{P}(\mathscr{X}) \right\}.$$

If X is complex Hermitian, $\mathbb{S}^n_1$ is also called the set of mixed quantum states in quantum information theory [2]. Its elements are known as density operators or density matrices. Its extreme points are rank-one matrices generated by vectors of the complex unit sphere. Alternatively, we can also interpret the moment body as a generalized numerical range – see e.g. [29, 22] and references therein – defined as the convex hull of the image of the complex unit sphere through the linear map $\mathcal{A}$, i.e.

$$\mathscr{M} = \mathrm{conv}\mathcal{A}(\mathscr{U}) = \mathrm{conv}\{[\mathbf{u}^T A_i \mathbf{u}]_{i=1,\ldots,m}, \; \mathbf{u} \in \mathscr{U}\}.$$

Finally, as a linear projection of a spectrahedron, the moment body is a *spectrahedral shadow*, see e.g. [21, Section 7.3]. Note however that not all spectrahedral shadows can be modeled as moment bodies. Since $\mathbf{b}_0 := \mathcal{A}(\frac{1}{n} I_n) \in \mathscr{M}$, we can write $\mathscr{M} = \mathbf{b}_0 + \mathscr{M}_0$ and represent the translated moment body $\mathscr{M}_0 := \{\mathcal{A}_0(X) : X \succeq 0, \mathrm{tr} X = 1\}$ as the dual to the spectrahedron $\{\mathbf{y} \in \mathbb{R}^m : I_n + A_0(\mathbf{y}) \in \mathbb{S}^n_+\}$, see e.g. [29, Corollary 5.3]. Both convex bodies contain the origin. The translated linear map $\mathcal{A}_0(X) := \mathcal{A}(X - \frac{1}{n} I_n)$ is traceless, i.e. $\mathcal{A}_0(\frac{1}{n} I_n) = 0$.

Throughout the paper, we make the following natural assumption on the linear map.

Assumption 1 *Matrices $I_n, A_1, \ldots, A_m$ are linearly independent in $\mathbb{S}^n$.*

Note that Assumption 1 implies that linear map $\mathcal{A}$ is surjective. It also implies that $\mathbf{b}_0$ is an interior point of $\mathscr{M}$.
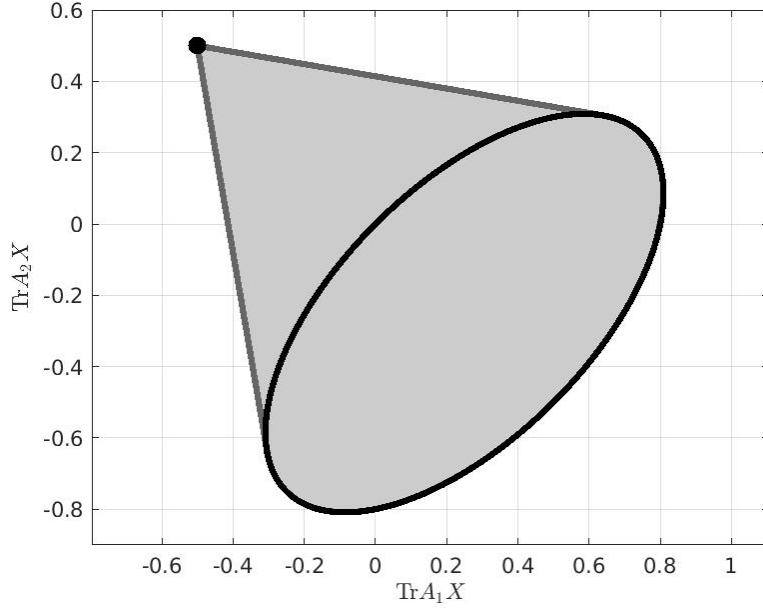
Figure 1: The moment body (light gray) of Example 1 is the convex hull of an ellipse (black, bottom right) and a point (black, top left).

Example 1 *Let $n = 3$, $m = 2$ and*

$$A_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad A_2 = \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

As explained e.g. in [9], the moment body of $\mathcal{A}$ is the convex hull of the algebraic curve dual to the curve $\{\mathbf{y} \in \mathbb{R}^2 : p(\mathbf{y}) = \det(I_3 + A_1 y_1 + A_2 y_2)\}$, i.e. the envelope of all tangent lines. The determinant factors into $p(\mathbf{y}) = \frac{1}{8}(4 + 2y_1 - 2y_2 - y_1^2 - 2y_1 y_2 - y_2^2)(2 - y_1 + y_2)$, so the dual curve is the union of the ellipse $\{\mathbf{x} \in \mathbb{R}^2 : 5x_1^2 - 6x_1 x_2 + 5x_2^2 - 4x_1 + 4x_2 = 0\}$ and the point $(-\frac{1}{2}, \frac{1}{2})$. Equivalently, in parametric form, the moment body of $\mathcal{A}$ is the convex hull of the ellipse

$$\left\{ \left(\operatorname{tr} \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}^T, \operatorname{tr} \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}^T \right), \theta \in [0, 2\pi] \right\}$$

and the point

$$\left\{ \left(\operatorname{tr} \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^T, \operatorname{tr} \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) \right\}.$$
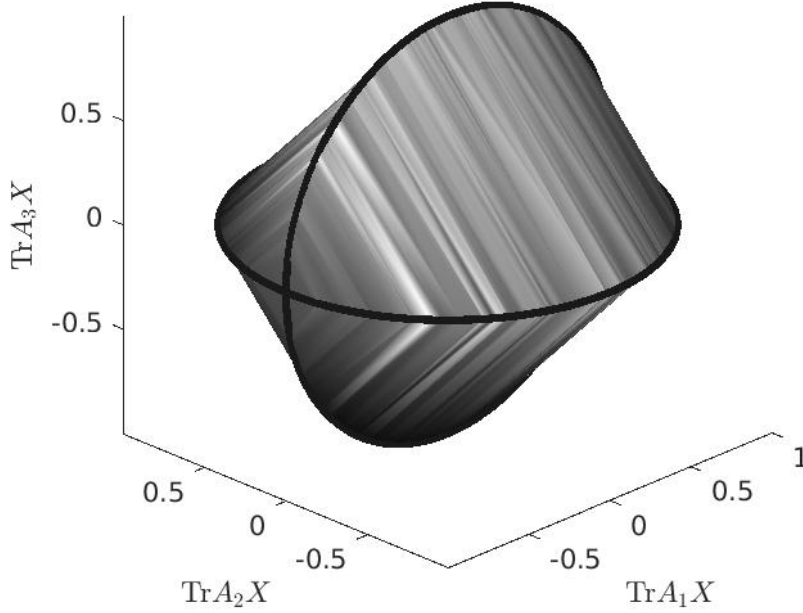
See Figure 1.

Figure 2: The moment body (light gray) of Example 2 is the convex hull of two orthogonal circles (black).

Example 2 *Consider the two unit circles in orthogonal planes in $\mathbb{R}^3$:*

$$C_{x_1x_2} = \{(\cos\theta,\ \sin\theta,\ 0)\colon \theta \in [0, 2\pi]\}, \quad C_{x_1x_3} = \{(\cos\phi,\ 0,\ \sin\phi)\colon \phi \in [0, 2\pi]\}.$$

Their convex hull can be modeled as a moment body as follows. Define the matrices

$$J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ A_1 = \begin{pmatrix} J & 0 \\ 0 & J \end{pmatrix}, \ A_2 = \begin{pmatrix} K & 0 \\ 0 & 0_2 \end{pmatrix}, \ A_3 = \begin{pmatrix} 0_2 & 0 \\ 0 & K \end{pmatrix},$$

i. e. $A_i \in \mathbb{S}^4$, $i = 1, 2, 3$. The corresponding moment body is the convex hull of the union of the circles $C_{x_1x_2}$ and $C_{x_1x_3}$. Indeed, $\mathcal{A}(X_1 \oplus 0_2) = \{(\mathrm{tr}X_1 J), \mathrm{tr}(X_1 K), 0) : X_1 \in \mathbb{S}_1^2\} = \mathrm{conv}\, C_{x_1x_2}$, $\mathcal{A}(0_2 \oplus X_2) = \{(\mathrm{tr}X_2 J), 0, \mathrm{tr}(X_2 K), 0) : X_2 \in \mathbb{S}_1^2\} = \mathrm{conv}\, C_{x_1x_2}$, and $\mathcal{A}(\mathbb{S}_1^4)$ consists of all convex combinations of these two sets, see Figure 2.

3 The moment body membership oracle

Given the linear map $\mathcal{A}$, the moment body membership oracle consists of determining whether a given vector $\mathbf{b} \in \mathbb{R}^m$ belongs to $\mathscr{M}$.

Let

$$f(\mathbf{y}) := \log \operatorname{tr} \exp A(\mathbf{y}) - \mathbf{b}^T \mathbf{y}$$

be the cumulant generating function or log-partition function.

Lemma 1 *Function f is smooth and convex on $\mathbb{R}^m$. Its gradient is*

$$\nabla f(\mathbf{y}) = [\operatorname{tr}(A_i X(\mathbf{y})) - b_i]_{i=1,\ldots,m} = \mathcal{A}(X(\mathbf{y})) - \mathbf{b}$$

and its Hessian is

$$\nabla^2 f(\mathbf{y}) = \left[\int_0^1 \operatorname{tr}\left(A_i\, X(\mathbf{y})^s\, A_j\, X(\mathbf{y})^{1-s} \right) ds \operatorname{tr}(A_i\, X(\mathbf{y})) \operatorname{tr}(A_j\, X(\mathbf{y})) \right]_{i=1,\ldots,m}.$$

where

$$X(\mathbf{y}) := \exp_1 A(\mathbf{y}) \in \mathbb{S}_1^n$$

is a so-called density matrix.

Proof: By standard matrix-calculus, $X \mapsto \exp X$ is smooth on $\mathbb{S}^n$, so $\log \operatorname{tr} \exp X$ is smooth as a composition. Let $t(\mathbf{y}) = \operatorname{tr} \exp A(\mathbf{y})$ so that $f(\mathbf{y}) = \log t(\mathbf{y}) - \mathbf{y}^T b$.

First derivatives. By the Duhamel formula for the derivative of a matrix exponential [38], we have

$$\frac{\partial \exp A(\mathbf{y})}{\partial y_i} = \int_0^1 \exp\left((1-s)A(\mathbf{y})\right) A_i\, \exp\left(sA(\mathbf{y})\right) ds. \tag{1}$$

Taking the trace gives

$$\frac{\partial t(\mathbf{y})}{\partial y_i} = \int_0^1 \operatorname{tr}\left(\exp((1-s)A(\mathbf{y}))\, A_i\, \exp(sA(\mathbf{y})) \right) ds = \operatorname{tr}\left(A_i\, \exp A(\mathbf{y}) \right) \tag{2}$$

by the cyclic property of the trace. Therefore

$$\frac{\partial \log t(\mathbf{y})}{\partial y_i} = \frac{1}{t(\mathbf{y})} \frac{\partial t(\mathbf{y})}{\partial y_i} = \frac{\operatorname{tr}(A_i\, \exp A(\mathbf{y}))}{\operatorname{tr}(\exp A(\mathbf{y}))} = \operatorname{tr}(A_i\, X(\mathbf{y}))$$

and finally

$$\frac{\partial f(\mathbf{y})}{\partial y_i} = \operatorname{tr}(A_i\, X(\mathbf{y})) - b_i.$$

Second derivatives. Let us differentiate the gradient

$$\frac{\partial^2 f(\mathbf{y})}{\partial y_i \partial y_j} = \frac{\partial \operatorname{tr}(A_i\, X(\mathbf{y}))}{\partial y_j} = \operatorname{tr}\left(A_i\, \frac{\partial X(\mathbf{y})}{\partial y_j} \right). \tag{3}$$

First develop

$$\frac{\partial X(\mathbf{y})}{\partial y_j} = \frac{1}{t(\mathbf{y})} \frac{\partial \exp A(\mathbf{y})}{\partial y_j} - \frac{\exp A(\mathbf{y})}{t(\mathbf{y})^2} \frac{\partial t(\mathbf{y})}{\partial y_j}$$

7

and use relation (2):

$$\frac{\partial t(\mathbf{y})}{\partial y_j} = \mathrm{tr}(A_j \exp A(\mathbf{y})) = t(\mathbf{y})\,\mathrm{tr}(A_j X(\mathbf{y}))$$

to obtain

$$\frac{\partial X(\mathbf{y})}{\partial y_j} = \frac{1}{t(\mathbf{y})}\,\frac{\partial \exp A(\mathbf{y})}{\partial y_j} - X(\mathbf{y})\,\mathrm{tr}(A_j X(\mathbf{y})).$$

We use again Duhamel's formula (1) to obtain

$$\frac{\partial X(\mathbf{y})}{\partial y_j} = \int_0^1 \frac{1}{t(\mathbf{y})}\exp\!\big(s\,A(\mathbf{y})\big)\,A_j\,\exp\!\big((1-s)\,A(\mathbf{y})\big)\,ds - X(\mathbf{y})\,\mathrm{tr}\!\big(A_j\,X(\mathbf{y})\big).$$

Substituting this expression into relation (3) we get

$$\frac{\partial^2 f(\mathbf{y})}{\partial y_i \partial y_j} = \int_0^1 \frac{1}{t(\mathbf{y})}\mathrm{tr}\!\Big(A_i\,\exp(sA(\mathbf{y}))\,A_j\,\exp((1-s)A(\mathbf{y}))\Big)\,ds - \mathrm{tr}(A_i\,X(\mathbf{y}))\,\mathrm{tr}(A_j\,X(\mathbf{y})).$$

Since $X(\mathbf{y})$ is symmetric and $s \in [0,1]$ it holds

$$X(\mathbf{y})^s = \frac{\exp(sA(\mathbf{y}))}{t(\mathbf{y})^s}, \quad X(\mathbf{y})^{1-s} = \frac{\exp((1-s)A(\mathbf{y}))}{t(\mathbf{y})^{1-s}}$$

and we have

$$\exp(sA(\mathbf{y}))\,A_j\,\exp((1-s)A(\mathbf{y})) = t(\mathbf{y})\,X(\mathbf{y})^s\,A_j\,X(\mathbf{y})^{1-s}.$$

Substituting this expression under the integral, we finally obtain

$$\frac{\partial^2 f(\mathbf{y})}{\partial y_i \partial y_j} = \int_0^1 \mathrm{tr}\!\Big(A_i\,X(\mathbf{y})^s\,A_j\,X(\mathbf{y})^{1-s}\Big)\,ds - \mathrm{tr}(A_i\,X(\mathbf{y}))\,\mathrm{tr}(A_j\,X(\mathbf{y})) \tag{4}$$

which is the expected expression. Note that these expressions were already studied in quantum information theory, see e.g. [34, Lem. VI], the Bogoliubov-Kubo-Mori (BKM) inner product in [1, Sect. 7.3] or [33, Prop. 6.1].

Convexity. Let us show that for any direction $\mathbf{u} \in \mathbb{R}^m$ and any vector $\mathbf{y} \in \mathbb{R}^m$ it holds

$$\mathbf{u}^T \nabla^2 f(\mathbf{y})\,\mathbf{u} \geq 0.$$

From relation (4) it holds

$$\mathbf{u}^T \nabla^2 f(\mathbf{y})\,\mathbf{u} = \int_0^1 \mathrm{tr}\!\Big(A(\mathbf{u})\,X(\mathbf{y})^s\,A(\mathbf{u})\,X(\mathbf{y})^{1-s}\Big)\,ds - \Big[\mathrm{tr}\!\big(A(\mathbf{u})\,X(\mathbf{y})\big)\Big]^2. \tag{5}$$

For each fixed $s \in [0,1]$, define the bilinear form

$$\langle Y, Z\rangle_s := \mathrm{tr}\!\Big(Y\,X^s(\mathbf{y})\,Z\,X^{1-s}(\mathbf{y})\Big).$$

This form satisfies the properties of an inner product because $X(\mathbf{y})$ is positive definite. Then by the usual Cauchy–Schwarz inequality for this inner product,

$$\langle A(\mathbf{u}), A(\mathbf{u})\rangle_s \langle I, I\rangle_s \geq \langle I, A(\mathbf{u})\rangle_s^2$$

8

since $\langle I, I \rangle_s > 0$. But

$$\langle I, A(\mathbf{u}) \rangle_s = \mathrm{tr}\Big(I\, X^s(\mathbf{y})\, A(\mathbf{u})\, X^{1-s}(\mathbf{y}) \Big) = \mathrm{tr}\Big(A(\mathbf{u})\, X(\mathbf{y}) \Big),$$

and

$$\langle I, I \rangle_s = \mathrm{tr}\Big(I\, X^s(\mathbf{y})\, I\, X^{1-s}(\mathbf{y}) \Big) = \mathrm{tr}\, X(\mathbf{y}) = 1.$$

Hence

$$\langle A(\mathbf{u}), A(\mathbf{u}) \rangle_s = \mathrm{tr}\Big(A(\mathbf{u})\, X^s(\mathbf{y})\, A(\mathbf{u})\, X^{1-s}(\mathbf{y}) \Big) \;\geq\; \Big[\mathrm{tr}\Big(A(\mathbf{u})\, X(\mathbf{y}) \Big)\Big]^2. \qquad (6)$$

Integrating over $s \in [0,1]$ gives

$$\int_0^1 \mathrm{tr}\Big(A(\mathbf{u}) X^s(\mathbf{y}) A(\mathbf{u}) X^{1-s}(\mathbf{y}) \Big)\, ds \;\geq\; \int_0^1 \Big[\mathrm{tr}(A(\mathbf{u})\, X(\mathbf{y}))\Big]^2 ds = \Big[\mathrm{tr}(A(\mathbf{u})\, X(\mathbf{y}))\Big]^2.$$

Therefore

$$\mathbf{u}^T \nabla^2 f(\mathbf{y})\, \mathbf{u} \;=\; \int_0^1 \mathrm{tr}(A(\mathbf{u}) X^s(\mathbf{y}) A(\mathbf{u}) X^{1-s}(\mathbf{y}))\, ds \;-\; \Big[\mathrm{tr}(A(\mathbf{u})\, X(\mathbf{y}))\Big]^2 \;\geq\; 0,$$

and so $\nabla^2 f(\mathbf{y})$ is positive semidefinite. $\qquad\square$

Lemma 2 *Function f is coercive (i.e. $\lim_{\|y\| \to \infty} f(y) = +\infty$) if and only if $\mathbf{b} \in \mathrm{int}\, \mathcal{M}$.*

Proof: If $\mathbf{b} \notin \mathrm{int}\, \mathcal{M}$, by the separating hyperplane theorem there exists a vector $\mathbf{u} \in \mathcal{U}$ so that $\lambda_{\max}\big(A(\mathbf{u}) \big) < \mathbf{u}^T \mathbf{b}$. Along the ray $\mathbf{y} = t\, \mathbf{u}$,

$$f(t\, \mathbf{u}) = \log \mathrm{tr} \exp t A(\mathbf{u}) - t\, \mathbf{b}^T \mathbf{u} \;\sim\; t\, \lambda_{\max}(A(\mathbf{u})) - t\, \mathbf{b}^T \mathbf{u} \;\longrightarrow\; -\infty \text{ when } t \to +\infty,$$

so f cannot be coercive.

Conversely, from the inequality $\sup_{\mathbf{y} \in \mathcal{M}} \mathbf{u}^T \mathbf{y} = \lambda_{\max}(A(\mathbf{u}))$, if $\mathbf{b} \in \mathrm{int}\, \mathcal{M}$ then for every vector $\mathbf{u} \in \mathcal{U}$, $\lambda_{\max}(A(\mathbf{u})) > \mathbf{b}^T \mathbf{u}$. Hence along the ray $\mathbf{y} = t\mathbf{u}$,

$$f(t\, \mathbf{u}) \;\geq\; t\, \lambda_{\max}(A(\mathbf{u})) - t\, \mathbf{b}^T \mathbf{u} \;\longrightarrow\; +\infty \text{ when } t \to +\infty.$$

$\qquad\square$

Lemma 3 *If $\mathbf{b} \in \mathrm{int}\, \mathcal{M}$, function f has a unique global minimizer*

$$f^* = \arg \min_{\mathbf{y} \in \mathbb{R}^m} f(\mathbf{y}).$$

Proof:

Recall expression (5) from the proof of Lemma 1, let

$$V(\mathbf{y}, \mathbf{u}) := \int_0^1 \mathrm{tr}(A(\mathbf{u})\, X^s(\mathbf{y})\, A(\mathbf{u})\, X^{1-s}(\mathbf{y}))\, ds - \Big[\mathrm{tr}\Big(A(\mathbf{u})\, X(\mathbf{y}) \Big)\Big]^2$$

9

and let us show that $V(\mathbf{y}, \mathbf{u})$ is zero if and only if $A(\mathbf{u})$ is a multiple of the identity matrix.

Assume $A(\mathbf{u}) = cI$ for some scalar $c \in \mathbb{R}$. We substitute this into the expression for $V(\mathbf{y}, \mathbf{u})$. The first term becomes:

$$\int_0^1 \operatorname{tr}\Big((cI)\, X^s(\mathbf{y})(cI)X^{1-s}(\mathbf{y})c\Big)\, ds = \int_0^1 \operatorname{tr}\Big(c^2 X^s(\mathbf{y})X^{1-s}(\mathbf{y})\Big)\, ds = \int_0^1 c^2 \operatorname{tr} X(\mathbf{y})\, ds = c^2.$$

The second term becomes:

$$\Big[\operatorname{tr}\big((cI)X(\mathbf{y})\big)\Big]^2 = \Big[c \operatorname{tr}(X(\mathbf{y}))\Big]^2 = c^2.$$

Therefore, $V(\mathbf{y}, \mathbf{u}) = 0$.

The converse statement relies on the Cauchy-Schwarz inequality (6) obtained from the inner product $I, \langle A(\mathbf{u})\rangle_s$ defined in the proof of Lemma 1. This inequality shows that the integrand in the expression for $V(\mathbf{y}, \mathbf{u})$ is non-negative for all $s \in [0, 1]$. If $V(\mathbf{y}, \mathbf{u}) = 0$ it holds

$$\int_0^1 \Big(\operatorname{tr}(A(\mathbf{u})X^s(\mathbf{y})A(\mathbf{u})X^{1-s}(\mathbf{y}))\Big)\, ds = \Big[\operatorname{tr}\big(A(\mathbf{u})\, X(\mathbf{y})\big)\Big]^2.$$

Since the integrand is a continuous and non-negative function of s, its integral can only be zero if the integrand is identically zero for all $s \in [0, 1]$. Therefore:

$$\operatorname{tr}(A(\mathbf{u})X^s(\mathbf{y})A(\mathbf{u})X^{1-s}(\mathbf{y})) = \Big[\operatorname{tr}(A(\mathbf{u})X(\mathbf{y}))\Big]^2 \quad \text{for all } s \in [0, 1].$$

This means the Cauchy-Schwarz inequality (6) must hold with equality for all s. Equality in the Cauchy-Schwarz inequality holds if and only if the two matrices $A(\mathbf{u})$ and I are linearly dependent, meaning $A(\mathbf{u})$ must be a multiple of the identity matrix.

Under Assumption 1, the matrices $A_1, \ldots, A_m$ cannot span the identity matrix, and hence $A(\mathbf{u}) = cI$ implies that $c = 0$, and thus $A(\mathbf{u}) = \mathbf{0}$. The linear independence of the matrices then forces $\mathbf{u} = \mathbf{0}$.

Therefore, $\mathbf{u}^T \nabla^2 f(\mathbf{y})\, \mathbf{u} > 0$ for all $\mathbf{u} \neq \mathbf{0}$, proving that f is strictly convex. A strictly convex function has at most one minimizer. Since f is also coercive when $\mathbf{b} \in \operatorname{int} \mathcal{M}$ (by Lemma 2), it is guaranteed to have a unique global minimizer. $\qquad\square$

The above results suggest that minimizing f solves the moment body membership oracle. Indeed, f has a unique global minimizer $\mathbf{y}^*$ at which the gradient of f vanishes, i.e. $\mathcal{A}(X(\mathbf{y}^*)) = \mathbf{b}$. Therefore the inclusion $\mathbf{b} \in \mathcal{M}$ is certified by the matrix $X(\mathbf{y}^*) = \exp_1 A(\mathbf{y}^*) \in \mathbb{S}_1^n$.

Theorem 1 *The convex unconstrained minimization problem*

$$\boxed{\min_{\mathbf{y} \in \mathbb{R}^m} f(\mathbf{y})}$$

is dual to the convex problem of maximizing the entropy in the pre-image of the moment body

$$\boxed{\max_{X \in \mathbb{S}_1^n} \operatorname{tr}(X - X \log X) \quad \text{s.t.} \quad \mathcal{A}(X) = \mathbf{b}}$$

At the optimum $(X^*, \mathbf{y}^*)$ it holds

$$\boxed{X^* = \exp_1 A(\mathbf{y}^*).}$$

Proof: Introduce multipliers $\mathbf{y} \in \mathbb{R}^m$, $z \in \mathbb{R}$, and $Z \in \mathbb{S}_+^n$ for these constraints, respectively. The Lagrangian is

$$\mathcal{L}(X, \mathbf{y}, z, Z) = \operatorname{tr}(X - X \log X) - \mathbf{y}^T(\mathcal{A}(X) - \mathbf{b}) - z(\operatorname{tr}X - 1) - \operatorname{tr}(XZ).$$

Using the matrix-derivative identity $\partial_X \operatorname{tr}(X - X \log X) = -\log X$, setting $\nabla_X \mathcal{L} = 0$ gives

$$\log X - A(\mathbf{y}) - zI - Z = 0.$$

At the optimum $X^* \succ 0$, so $Z^* = 0$. Thus

$$\log X^* - A(\mathbf{y}^*) - z^* I = 0 \implies X^* = \exp\left(A(\mathbf{y}^*) + z^* I\right) = \exp z^* \exp A(\mathbf{y}^*).$$

The constraint $\operatorname{tr}X^* = 1$ enforces $z^* = -\log \operatorname{tr} \exp A(\mathbf{y}^*)$ and hence

$$X^* = \frac{\exp A(\mathbf{y}^*)}{\operatorname{tr} \exp A(\mathbf{y}^*)} = \exp_1 A(\mathbf{y}^*).$$

From dual optimality, it holds $\mathcal{A}(X^*) = \mathbf{b}$, i.e. $X^* \in \mathcal{A}^{-1}(\mathbf{b})$. $\qquad\square$

The following result is well-known in quantum information theory, see e.g. [14, Theorem 2], where it is attributed to [34].

Lemma 4 *The map $\mathbf{y} \mapsto \mathcal{A}(\exp_1 A(\mathbf{y}))$ is a smooth diffeomorphism between $\mathbb{R}^m$ and $\operatorname{int} \mathcal{M}$.*

Proof: Define $g(\mathbf{y}) = \log \operatorname{tr} \exp A(\mathbf{y}) = f(\mathbf{y}) + \mathbf{b}^T \mathbf{y}$ and $\Phi(\mathbf{y}) = \nabla g(\mathbf{y}) = \mathcal{A}(X(\mathbf{y}))$ with $X(\mathbf{y}) = \exp_1 A(\mathbf{y})$. We claim that $\Phi : \mathbb{R}^m \longrightarrow \operatorname{int} \mathcal{M}$ is a smooth diffeomorphism from the whole space onto the interior of the moment body.

(i) Smoothness. Since f is smooth, g is smooth, and its gradient Φ is smooth.

(ii) Injectivity. g is strictly convex, so Φ is injective.

(iii) Local invertibility. $\nabla^2 g(\mathbf{y})$ is positive-definite for all $\mathbf{y}$, hence $\nabla \Phi(\mathbf{y}) = \nabla^2 g(\mathbf{y})$ is invertible everywhere. By the inverse-function theorem, Φ is a diffeomorphism.

(iv) Image equals the interior. For any $\mathbf{y}$, $\Phi(\mathbf{y}) = \mathcal{A}(X(\mathbf{y}))$ where $X(\mathbf{y}) \succ 0$, so $\Phi(\mathbf{y})$ lies in the interior of $\mathcal{M}$. Conversely, given any interior point $\mathbf{x}$ of the moment body, strict convexity of g and the Legendre-transform duality imply there is a unique $\mathbf{y}$ solving $\nabla g(\mathbf{y}) = \mathbf{x}$.

(v) Properness i.e. surjectivity onto the interior. Strict convexity plus coercivity of g ensure $\|\Phi(\mathbf{y})\| \to \infty$ as $\|\mathbf{y}\| \to \infty$, forcing the range of Φ to be open, closed, and nonempty in the interior of $\mathcal{M}$, hence equal to it.

$\qquad\square$

Example 3 *For the matrices of Example 1, the graph of function $f(y)$ is represented on Figure 3, together with a regular grid $\mathcal{Y}$ (black lines underneath). The moment body $\mathcal{M}$ is represented on Figure 4, together with the image of the grid through the gradient map $\mathcal{A}(\exp_1 \mathcal{A}^T(\mathcal{Y}))$.*
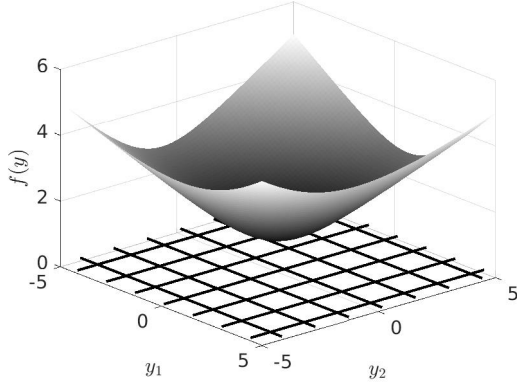
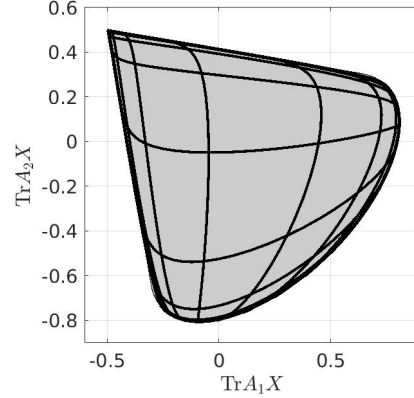Figure 3: Dual function graph (gray) and regular grid underneath (black).

Figure 4: Moment body (gray) and grid image through the gradient map (black).

4 Geometric analysis

In this section, let us make the following

Assumption 2 $\mathbf{b} \in \operatorname{int} \mathcal{M}$.

This implies that f has a unique minimizer, see Lemmas 1 and 2. Since f is smooth and convex, its minimization can be achieved with standard optimization algorithms. The performance of these algorithms depends on the geometry of f, and especially its curvature. We say that f is α-strongly convex and λ-smooth whenever

$$0 \prec \alpha I_m \preceq \nabla^2 f(\mathbf{y}) \preceq \lambda I_m \quad \forall \mathbf{y} \in \mathbb{R}^m.$$

The constant $\alpha > 0$ is called the strong convexity modulus, and the constant $\lambda > 0$ is called the smoothness constant (or Lipschitz constant of ∇f). The condition number

$$\kappa = \frac{\lambda}{\alpha}$$

governs the convergence rates of standard first- and second-order methods. With fixed step-size $1/\lambda$, gradient descent $\mathbf{y}_{k+1} = \mathbf{y}_k - \alpha \nabla f(\mathbf{y}_k)$ satisfies the linear rate $f(\mathbf{y}_k) - f(\mathbf{y}^*) \leq (1 - \frac{1}{\kappa})^k (f(\mathbf{y}_0) - f(\mathbf{y}^*))$, so that reaching ε-accuracy to the minimum $\mathbf{y}^*$ requires $O(\kappa \log \varepsilon^{-1})$ iterations, see e.g. [5, Sec. 9.3.1]. Nesterov's accelerated scheme achieves the optimal first-order complexity $O(\sqrt{\kappa} \log \frac{1}{\varepsilon})$ by combining momentum with gradient steps, see [23, Ch. 2, Thm. 2.2.2]. Near the optimum, Newton's method $\mathbf{y}_{k+1} = \mathbf{y}_k - [\nabla^2 f(\mathbf{y}_k)]^{-1} \nabla f(\mathbf{y}_k)$ converges quadratically, but its region of attraction and the quality of each step depend on κ. Ill-conditioned Hessians can force small steps or necessitate line-search/globalization strategies, whose complexity again scales with κ, see [25, Sec. 3.5]. Interior-point methods exhibit polynomial-time complexity bounds that depend on the barrier Hessian's conditioning κ (see [35, Chap. 5]), and quasi-Newton updates (e.g. BFGS) achieve superlinear convergence only when κ is moderate. Consider running L-BFGS with memory parameter m_{hist} and a standard Wolfe line-search starting from y_0. Then the L-BFGS iterates $(\mathbf{y}_k)$ satisfy

$f(\mathbf{y}_k) - f(\mathbf{y}^*) \leq \rho^k (f(\mathbf{y}_0) - f(\mathbf{y}^*))$, where the rate $\rho = 1 - \frac{c}{\kappa \, m_{\mathrm{hist}}}$ for some constant c, showing the impact of the conditioning κ. Equivalently, $\|\mathbf{y}_k - \mathbf{y}^*\|_2 \leq C \, \rho^{k/2} \|\mathbf{y}_0 - \mathbf{y}^*\|_2$, for some constant C, see e.g. [25, Chapter 8].

To solve the moment body membership problem for moderate size problems ($n, m \approx 1000$), we propose to use L-BFGS, a standard quasi-Newton algorithm constructing an approximation of the Hessian using a limited number of evaluations of the gradient. It can be interpreted as a discretization of a variable-metric generalization of the Newton flow where the true inverse Hessian is replaced by a time-varying symmetric positive-definite matrix.

In the context of semidefinite optimization, the idea of formulating and solving with BFGS a dual smooth problem was already explored in [18] for the semidefinite least-squares problem, consisting of projecting a given symmetric matrix onto a given spectrahedral shadow. It was later on used to solve polynomial SOS problems [11].

In this section, we derive bounds on the curvature of f depending explicitly on the problem data. For this we need to define the Gram matrix

$$G := \big[\mathrm{tr}(A_i A_j)\big]_{i,j=1,\ldots,m}. \tag{7}$$

Lemma 5 (Smoothness) *Let* $\lambda := \lambda_{\max}(G)$. *Function* f *is* $\frac{1}{2}\lambda$-*smooth.*

Proof: To prove that f is $\frac{1}{2}\lambda$-smooth, we must show that $\mathbf{u}^T \nabla^2 f(\mathbf{y}) \, \mathbf{u} \leq \frac{1}{2}\lambda\|\mathbf{u}\|_2^2$ for any $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^m$.

Let $\mathbf{y}$ and $\mathbf{u}$ be given, arbitrary. For notational ease, we will write $X := X(\mathbf{y})$ and $A := A(\mathbf{u})$. Let

$$X = V \, \mathrm{diag}(\lambda_i) \, V^T, \quad V = [v_1 \cdots v_n], \quad V^T V = I_n$$

be the spectral decomposition of positive definite symmetric matrix X, so that for all $s \in [0,1]$ it holds

$$X^s = V \, \mathrm{diag}(\lambda_i^s) \, V^T, \quad X^{1-s} = V \, \mathrm{diag}(\lambda_i^{1-s}) \, V^T$$

and hence

$$A X^s A X^{1-s} = A V \, \mathrm{diag}(\lambda_i^s) \, V^T A V \, \mathrm{diag}(\lambda_j^{1-s}) \, V^T.$$

Taking the trace we get

$$\mathrm{tr}(AX^sAX^{1-s}) = \mathrm{tr}(\mathrm{diag}(\lambda_i^s) \, V^T AV \, \mathrm{diag}(\lambda_j^{1-s}) \, V^T AV).$$

Writing out the diagonal product, the (i,i) entry of the right hand side matrix is

$$\sum_{j=1}^n \lambda_i^s \, (V^T AV)_{ij} \, \lambda_j^{1-s} \, (V^T AV)_{ji}.$$

Since A is symmetric, $(V^T AV)_{ji} = (V^T AV)_{ij}$ and hence

$$\mathrm{tr}(AX^sAX^{1-s}) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i^s \, \lambda_j^{1-s} \, (v_i^T A v_j)^2.$$

For $s \in [0, 1]$, let
$$g(s) := \text{tr}(AX^s AX^{1-s}).$$

Using the above expression, we can write
$$g(s) = \sum_{i,j} g_{ij} \exp(s(\log \lambda_i - \lambda_j))$$

for some non-negative coefficients g_{ij}. Since each term $\exp(sa)$ is convex in s, $g(s)$ is convex in s and hence
$$\int_0^1 \text{tr}(AX^s AX^{1-s})ds = \int_0^1 g(s)ds \leq \max(g(0), g(1)) = \text{tr}(A^2 X).$$

Recalling the expression (5) of the second directional derivative it holds
$$\mathbf{u}^T \nabla^2 f(\mathbf{y}) \, \mathbf{u} \leq \text{tr}(A^2(\mathbf{u}) \, X(\mathbf{y})) - [\text{tr}(A(\mathbf{u})X(\mathbf{y}))]^2 \leq \frac{1}{2} \text{tr}((A\mathbf{u})^2) \tag{8}$$

Now let
$$p_k := v_k^T X v_k, \quad a_k := v_k^T A v_k.$$

Then vector $\mathbf{p}$ belongs to the simplex $\{\mathbf{p} \in \mathbb{R}^n : p_k \geq 0, \sum_k p_k = 1\}$ and
$$\text{tr}(A^2 X) = \sum_k p_k a_k^2, \quad \text{tr}(AX) = \sum_k p_k a_k.$$

Hence
$$\mathbf{u}^T \nabla^2 f(\mathbf{y}) \, \mathbf{u} \leq \sum_k p_k a_k^2 - \left(\sum_k p_k a_k\right)^2 =: q(\mathbf{p}). \tag{9}$$

Observe that the hessian $\left(\frac{\partial^2 q(\mathbf{p})}{\partial p_i \partial p_j}\right) = -2(a_i a_j)$ is rank-one negative semidefinite, so that q is concave. In order to get an upper bound on q, let us maximize it on the simplex. Construct the Lagrangian $q(\mathbf{p}) - (\sum_k p_k - 1)\ell$ whose stationarity conditions at a maximizer $\mathbf{p}^*$ are $a_i^2 - 2(\sum_k p_k^* a_k)a_i = \ell$. Let us now prove that amongst all maximizers $\mathbf{p}^*$, we can choose one whose support $\{i : p_i^* > 0\}$ consists of two indices at most. For every index i in the support, it holds $a_i^2 - 2b\,a_i = \lambda$ where $b := \sum_k p_k^* a_k$. This is a quadratic equation in a_i, with non-negative discriminant $(-2b)^2 - 4(-\lambda) = 4(b^2 + \lambda) = 4(b^2 + a_i^2 + 2ba_i) = 4(b - a_i)^2$ so the equation admits at most two distinct real solutions α_1, α_2. This shows that every a_i with $p_i^* > 0$ must be one of the (at most) two roots of that quadratic. In other words, the set $\{a_i : p_i^* > 0\}$ contains at most two distinct values. Group the indices by which root they take $I_1 = \{i : a_i = \alpha_1\}$, $I_2 = \{i : a_i = \alpha_2\}$. Define $r_1 = \sum_{i \in I_1} p_i^*$, $r_2 = \sum_{i \in I_2} p_i^* = 1 - r_1$. Since all a_i in I_1 equal α_1, and all in I_2 equal α_2, one checks that the maximum $q(\mathbf{p}^*) = \sum_k p_k^* a_k^2 - (\sum_k p_k^* a_k)^2 = r_1(\alpha_1)^2 + (1 - r_1)(\alpha_2)^2 - (r_1\alpha_1 + (1 - r_1)\alpha_2)^2$ depends only on r_1 and not on the way it is split among indices. So without loss of generality, and can choose only one index $i_1 \in I_1$ and one index $i_2 \in I_2$ and the corresponding vector $\mathbf{p}^*$ will achieve the same maximum on the simplex. Writing $p_{i_1}^* = t$, $p_{i_2}^* = 1 - t$ one finds $q(\mathbf{p}^*) = \max_{t \in [0,1]} t(1 - t)(\alpha_1 - \alpha_2)^2 = \frac{1}{4}(\alpha_1 - \alpha_2)^2$. Hence on the simplex it holds
$$q(\mathbf{p}) \leq \frac{1}{4}(\max_k a_k - \min_k a_k)^2. \tag{10}$$

14

For any two indices i, j it holds $(a_i - a_j)^2 \leq a_i^2 + a_j^2 \leq 2\sum_k a_k^2$, and hence

$$q(\mathbf{p}) \leq \frac{1}{2}\sum a_k^2.$$

Combining our previous bound (9)

$$\mathbf{u}^T \nabla^2 f(\mathbf{y})\, \mathbf{u} \leq q(\mathbf{p}) \;\leq\; \tfrac{1}{2}\sum_{k=1}^{n} a_k^2 \;=\; \tfrac{1}{2}\,\mathrm{tr}(A(\mathbf{u})^2)$$

with the definition (7) of the Gram matrix

$$\mathrm{tr}(A(\mathbf{u})^2) = \sum_{i,j} u_i\, u_j\, \mathrm{tr}(A_i A_j) = \mathbf{u}^T G\, \mathbf{u},$$

we obtain

$$\mathbf{u}^T \nabla^2 f(\mathbf{y})\, \mathbf{u} \;\leq\; \tfrac{1}{2}\,\mathbf{u}^T G\, \mathbf{u} \;\leq\; \tfrac{1}{2}\,\lambda\,\|\mathbf{u}\|_2^2.$$

Since this holds for every direction $\mathbf{u}$ and every $\mathbf{y}$, it follows that f is $\frac{1}{2}\lambda$-smooth on $\mathbb{R}^m$. Note that the bound (10)

$$\mathrm{tr}(A^2(\mathbf{u})\, X(\mathbf{y})) - [\mathrm{tr}(A(\mathbf{u})X(\mathbf{y}))]^2 \leq \frac{1}{4}(\lambda_{\max} A(\mathbf{u}) - \lambda_{\min} A(\mathbf{u}))$$

that we just proved algebraically is a particular case of a more general result in non-commutative probability theory, see [4, Thm. 2]. $\qquad\square$

Lemma 6 (Strong convexity) *Let $y_0 \in \mathbb{R}^m$ be given. On the sublevel set $\mathscr{S} := \{\mathbf{y} \in \mathbb{R}^m : f(\mathbf{y}) \leq f(\mathbf{y}_0)\}$, the function f is α-strongly convex with*

$$\alpha := \frac{\lambda_{\min}(G)}{n^2 \exp f(\mathbf{y}_0)} > 0.$$

In particular if $\mathbf{y}_0 = 0$, f is $\frac{\lambda_{\min}(G)}{n^3}$-strongly convex.

Proof: First note that $\lambda_{\min}(G) > 0$ follows from Assumption 1. For the function

$$\phi(X) := \log \mathrm{tr} \exp X$$

it holds

$$\nabla^2 \phi(X) \;\succeq\; p_{\min}(X)\, I_n, \quad p_{\min}(X) := \min_{i=1,\dots,n} \frac{\exp \lambda_i(X)}{\sum_{j=1}^{n} \exp \lambda_j(X)}.$$

By the chain rule and the variational form of the Hessian of ϕ, one shows $\nabla^2 f(\mathbf{y}) = \mathcal{A}\left[\nabla^2 \phi(A(\mathbf{y}))\right] \mathcal{A}^T$. Since $\mathcal{A}\, I_n\, \mathcal{A}^T = G$, it follows that

$$\nabla^2 f(\mathbf{y}) \;\succeq\; \lambda_{\min}(G)\, p_{\min}(A(\mathbf{y}))\, I_m$$

.

Let $\lambda_{\min}(X) \leq \cdots \leq \lambda_{\max}(X)$ be the eigenvalues of X. Then

$$p_{\min}(X) = \frac{\exp \lambda_{\min}(X)}{\sum_j \exp \lambda_j(X)} \;\geq\; \frac{\exp \lambda_{\min}(X)}{n \,\exp \lambda_{\max}(X)} \;=\; \frac{1}{n \exp(\lambda_{\max}(X) - \lambda_{\min}(X))}.$$

Define the spectral-gap over $\mathscr{S}$:

$$\delta := \sup_{\mathbf{y} \in \mathscr{S}} \Big\{ \lambda_{\max}(A(\mathbf{y})) - \lambda_{\min}(A(\mathbf{y})) \Big\}.$$

Then for all $\mathbf{y} \in \mathscr{S}$, it holds $p_{\min}(A(\mathbf{y})) \geq \frac{1}{n \exp \delta}$, and hence $\lambda_{\min}(\nabla^2 f(\mathbf{y})) \geq \lambda_{\min}(G) \frac{1}{n \exp \delta}$. Taking the infimum over $\mathbf{y} \in \mathscr{S}$ yields

$$\alpha = \inf_{\mathbf{y} \in \mathscr{S}} \lambda_{\min}(\nabla^2 f(\mathbf{y})) \;\geq\; \frac{\lambda_{\min}(G)}{n \exp \delta}. \tag{11}$$

By strong duality for the log-partition function,

$$f(\mathbf{y}^*) = \min_{\mathbf{y}} f(\mathbf{y}) \;=\; \min_{X \in \mathbb{S}_1^n} \mathrm{tr} X \log X \;\geq\; -\log n.$$

On the other hand, since f increases with the spectral-gap, $\delta \;\leq\; f(\mathbf{y}_0) - f(\mathbf{y}^*) \;\leq\; f(y_0) + \log n$. Therefore $\exp \delta \;\leq\; \exp(f(\mathbf{y}_0) + \log n) \;=\; n \exp f(\mathbf{y}_0)$. Substituting into (11) gives the required result. $\qquad\square$

5   Pre-conditioning

For the conditioning κ of f to be as small as possible, Lemmas 5 and 6 indicate that $\lambda_{\max}(G)$ should be small and $\lambda_{\min}(G)$ should be large, where G is the Gram matrix (7) of the A_i. To accelerate the convergence of optimization algorithms to minimize f, we wish to replace A_i by a new set $\hat{A}_i$ whose Gram matrix is perfectly conditioned, i.e. $\lambda_{\max}(G) = \lambda_{\min}(G) = 1$ or equivalently $G = I_m$.

Algorithm 1 Pre-conditioning

Require: $A_1, \ldots, A_m \in \mathbb{S}^n$
Ensure: $\hat{A}_1, \ldots, \hat{A}_m \in \mathbb{S}^n$ with $\operatorname{tr} \hat{A}_i = 0$ and $\operatorname{tr}(\hat{A}_i \hat{A}_j) = \delta_{i-j}$

1: **(Center to traceless)**

$$A_i' \;=\; A_i \;-\; \frac{\operatorname{tr} A_i}{n}\, I_n, \quad i = 1, \ldots, m.$$

2: **(Compute Gram matrix)**

$$G_{ij}' \;=\; \operatorname{tr}\!\left(A_i'\, A_j'\right), \quad i, j = 1, \ldots, m.$$

3: **(Whitening)** Compute the symmetric inverse-square-root $G'^{-1/2}$ via eigendecomposition:

$$G' = U\, D\, U^T, \quad G'^{-1/2} = U\, D^{-1/2}\, U^T.$$

4: **(Form Orthonormal Basis)**

$$\hat{A}_i = \sum_{j=1}^{m} \left(G'^{-1/2}\right)_{ij} A_j', \quad i = 1, \ldots, m.$$

Theorem 2 (Correctness) *The output $\hat{A}_i$ of Algorithm 1 satisfies:*

1. $\operatorname{tr} \hat{A}_i = 0$ *for all i.*

2. $\operatorname{tr}(\hat{A}_i \hat{A}_j) = \delta_{i-j}$ *for all i, j.*

Hence the $\hat{A}_i$ are traceless and orthonormal.

Proof: 1. Tracelessness. Each A_i' is by construction

$$\operatorname{tr} A_i' = \operatorname{tr} A_i - \frac{\operatorname{tr} A_i}{n} \operatorname{tr} I_n = \operatorname{tr} A_k - \operatorname{tr} A_k = 0.$$

Since $\hat{A}_i$ is a linear combination of the A_j', it too is traceless: $\operatorname{tr} \hat{A}_i = \sum_j (G'^{-1/2})_{ij} \operatorname{tr} A_j' = 0$.

2. Orthonormality. Define the centered Gram matrix G'. Then

$$\operatorname{tr}(\hat{A}_i \hat{A}_j) = \sum_{p,q} (G'^{-1/2})_{ip} (G'^{-1/2})_{jq} \operatorname{tr}(A_p' A_q') = \left[G'^{-1/2} G' G'^{-1/2}\right]_{ij} = \delta_{i-j}.$$

This shows $(\hat{A}_i)$ are orthonormal in the Frobenius inner product. $\qquad\square$
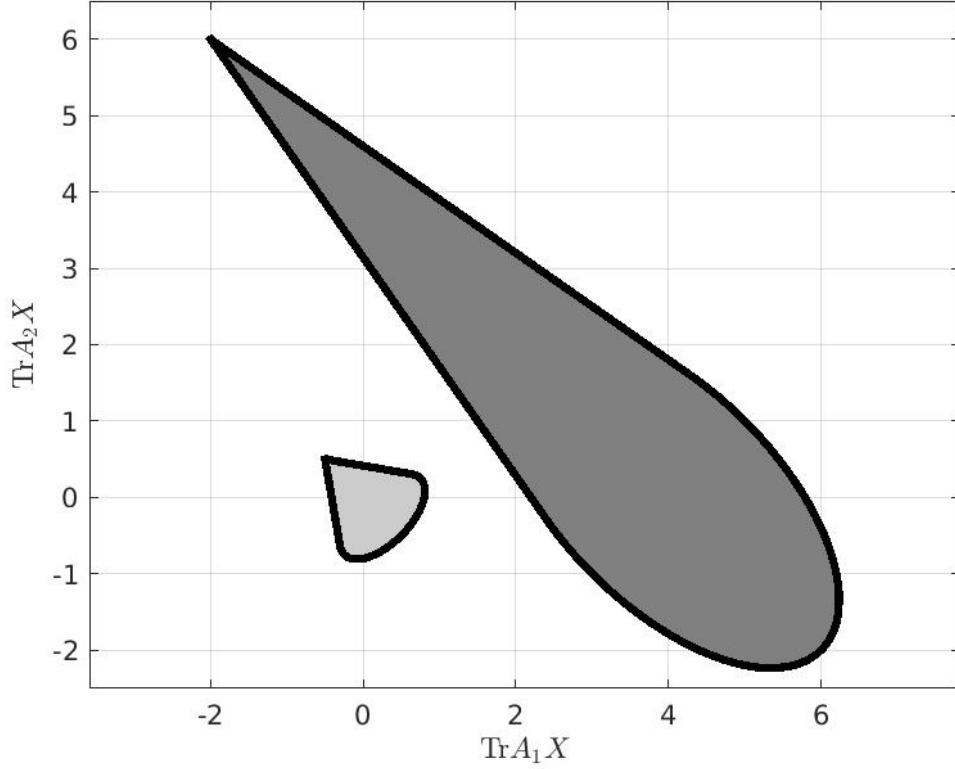
Figure 5: Moment bodies before (dark gray) and after (light gray) pre-conditioning Algorithm 1.

Example 4 *Let $n = 3$, $m = 2$ and*

$$A_1 = \begin{pmatrix} 6 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad A_2 = \frac{1}{2} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

whose Gram matrix has eigenvalues 28 and 64. Step 1 of Algorithm 1 yields the traceless matrices

$$A_1' = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -4 \end{pmatrix}, \quad A_2' = \frac{1}{2} \begin{pmatrix} -4 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

whose Gram matrix computed in step 2 has eigenvalues 4 and 64. Finally, step 4 yields the traceless orthonormal matrices

$$\hat{A}_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \hat{A}_2 = \frac{1}{2} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

corresponding to Example 3, whose Gram matrix is identity. The corresponding moment bodies, before and after the application of Algorithm 1, are represented in Figure 5. The same pre-conditioned moment body is also represented on Figures 1 and 4.

18

We remark that the idea of using the Cholesky factor of the Gram matrix of the linear map was already exploited in the context of projection methods for semidefinite optimization, see [19] and [12, section 3.2].

6 Refined geometric analysis

In this section we assume that after the application of Algorithm 1 the matrices A_i are traceless and orthonormal, i.e.

$$\mathrm{tr} A_i = 0, \quad \mathrm{tr}(A_i A_j) = \delta_{i-j}, \quad i, j = 1, \dots, m.$$

or equivalently $\mathcal{A}(I_n) = 0$, $\mathcal{A} \circ \mathcal{A}^T = I_m$. The corresponding moment body $\mathcal{M}$ is normalized[3], and we now report some of its geometric properties.

Lemma 7 *The radius of $\mathcal{M}$ is*

$$\mathrm{rad}\,\mathcal{M} := \max_{X \in \mathbb{S}_1^n} \|A(X)\|_2 \;=\; \sqrt{\frac{n-1}{n}} < 1.$$

Proof: Since A_i are traceless and orthonormal, it holds $\mathcal{A}(\frac{1}{n} I_n) = 0$ and hence $\mathcal{M}$ is centered at the origin. The map $\mathcal{A}$ acts on the traceless part of X:

$$\mathcal{A}(X) = \mathcal{A}(X - \frac{\mathrm{tr} X}{n} I_n).$$

Furthermore, since the map $\mathcal{A}$ (from the space of matrices with the Frobenius norm to $\mathbb{R}^m$ with the Euclidean norm) corresponds to an orthogonal projection, its operator norm is at most one. This implies the inequality $\|\mathcal{A}(Y)\|_2 \leq \|Y\|_F$ for any Y. Hence, for any matrix $X \in \mathbb{S}_1^n$:

$$\|\mathcal{A}(X)\|_2^2 = \|\mathcal{A}(X - \tfrac{1}{n} I_n)\|_2^2 \leq \|X - \tfrac{1}{n} I_n\|_F^2.$$

The radius is therefore bounded by the maximum value of the term on the right. We have

$$\|X - \tfrac{1}{n} I\|_F^2 = \|X\|_F^2 - \frac{2}{n} \mathrm{tr} X + \frac{1}{n} \leq 1 - \frac{2}{n} + \frac{1}{n} = \frac{n-1}{n},$$

where we used $\|X\|_F^2 = \mathrm{tr}(X^2) \leq \mathrm{tr}(X) = 1$. This proves that $\mathrm{rad}\,\mathcal{M} \leq \sqrt{\frac{n-1}{n}}$. As this bound is achievable, the equality holds. □

The support function of $\mathcal{M}$ is $h : \mathscr{U} \to \mathbb{R}_+, \mathbf{u} \mapsto h(\mathbf{u}) := \sup_{\mathbf{x} \in \mathcal{M}} \mathbf{b}^T \mathbf{x}$. The width of $\mathcal{M}$ in direction $\mathbf{u}$ is $w(\mathbf{u}) := h(\mathbf{u}) - h(-\mathbf{u})$. The minimal width or thickness, is $\mathrm{thick}\,\mathcal{M} := \inf_{\mathbf{u} \in \mathscr{U}} w(\mathbf{u})$. The maximal width or diameter, is $\mathrm{diam}\,\mathcal{M} := \sup_{\mathbf{u} \in \mathscr{U}} w(\mathbf{u})$.

---

[3]A Traceless Orthonormal Moment Body can be called a TOMB, evoking a solid, well-defined shape, where all the traceless moments rest.

Lemma 8 *It holds $h(\mathbf{u}) = \lambda_{\max}(A(\mathbf{u}))$ and hence $w(\mathbf{u}) = \lambda_{\max}(A(\mathbf{u})) - \lambda_{\min}(A(\mathbf{u}))$ is the spectral gap along direction $\mathbf{u}$. Moreover*

$$\operatorname{diam}\mathscr{M} = \sqrt{2}, \quad \operatorname{thick}\mathscr{M} = \sqrt{\frac{n}{\lfloor n/2 \rfloor \lceil n/2 \rceil}} = \begin{cases} \frac{2}{\sqrt{n}} & \text{for even } n = 2k, \\ \sqrt{\frac{2k+1}{k(k+1)}} & \text{for odd } n = 2k+1. \end{cases}$$

Proof: (Sketch) For any unit vector $\mathbf{u} \in \mathscr{U}$,

$$\sup_{\mathbf{x}\in\mathscr{M}} \mathbf{u}^T\mathbf{x} = \sup_{X\in\mathbb{S}_1^n} \mathbf{u}^T A(X) = \sup_{X\in\mathbb{S}_1^n} \operatorname{tr}(A(\mathbf{u})X) = \lambda_{\max}(A(\mathbf{u}))$$

and similarly for the infimum. The maximal spectral gap of a traceless Frobenius-unit matrix is achieved by a rank-2 matrix with eigenvalues $\pm\frac{1}{\sqrt{2}}$, giving $\sqrt{2}$. The minimal spectral gap occurs when the positive and negative eigenvalues are as evenly distributed as possible, leading to the stated formula in terms of $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$. □

We observe that the diameter $\sqrt{2}$ is less than twice the radius $2\sqrt{\frac{n-1}{n}}$, reflecting the fact that $\mathscr{M}$ is not centrally symmetric.

Let

$$\mathbf{y}^* := \arg\min_{\mathbf{y}\in\mathbb{R}^m} f(\mathbf{y})$$

denote the minimizer of f, which is unique from Lemma 3. The geometric properties of $\mathscr{M}$ allow us to bound the value of f at $\mathbf{y}^*$, as well as the norm of $\mathbf{y}^*$ itself. Let Assumption 2 hold for the remainder of this section, and let

$$\beta := 1 - \|\mathbf{b}\|_2 \in (0,1].$$

We can bound the minimum and the norm of the minimizer. Tighter bounds can be obtained, but their expressions are slightly more involved.

Lemma 9 *It holds*

$$\log\beta \;\leq\; f(\mathbf{y}^*) \;\leq\; \log n, \qquad \|\mathbf{y}^*\|_2 \;\leq\; \sqrt{n}\,\log\frac{1}{\beta}.$$

Proof: Since $\mathbf{b} \in \mathscr{M}$ and by Lemma 7 the largest norm of any point in $\mathscr{M}$ is $\operatorname{rad}\mathscr{M} = \sqrt{\frac{n-1}{n}} < 1$, it follows that $\beta > 0$. Moreover, the distance of $\mathbf{b}$ to the boundary of $\mathscr{M}$ is larger than the distance of $\mathbf{b}$ to the unit sphere $\mathscr{U}$, equal to β. By strong duality,

$$f(\mathbf{y}^*) = \min_{X\in\mathbb{S}_1^n} \operatorname{tr}(X\log X) \;\geq\; \log\lambda_{\min}(X^*) \;\geq\; \log\beta.$$

On the other hand, evaluating at $y = 0$ gives

$$f(\mathbf{y}^*) \;\leq\; f(0) = \log\operatorname{tr}\exp A(0) = \log n.$$

Hence the first two inequalities.

Next, since $\mathcal{A}^T$ is an isometry onto the traceless subspace, it holds $\|\mathbf{y}^*\|_2 = \|A(\mathbf{y}^*)\|_F$. But

$$A(\mathbf{y}^*) = \log X^* - \tfrac{1}{n}\operatorname{tr}\log X^*\, I_n,$$

20

and the spectrum of X^* lies in $[\beta, 1]$. Therefore each eigenvalue of $\log X^*$ lies in $[\log \beta, 0]$, so the centered spectrum lies in an interval of length $-\log \beta$. Hence

$$\|\mathbf{y}^*\|_2^2 = \|\log X^* - \tfrac{1}{n}\text{tr} \log X^* \, I_n\|_F^2 = \sum_i (\log \lambda_i - \tfrac{1}{n}\sum_i \log \lambda_i)^2 \leq n \left(-\log \beta\right)^2.$$

$\square$

Theorem 3 *Function f is $\frac{1}{2}$-smooth.*

Proof: The global smoothness constant is just an application of Lemma 5 when $G = I_m$. $\square$

Theorem 4 *Function f is $\frac{1}{n^3}$-strongly convex and β-strongly convex around its minimizer.*

Proof: The sublevel strong convexity modulus is just an application of Lemma 6 when $G = I_m$ and $\mathbf{y}_0 = 0$, since then $\exp f(\mathbf{y}_0) = n$. As shown in the proof of Lemma 9, at the minimizer $\mathbf{y}^*$, it holds $\lambda_{\min}(X(\mathbf{y}^*)) \geq \beta$, and hence $\nabla^2 f(\mathbf{y}^*) \succeq \lambda_{\min}(X(\mathbf{y}^*))I_m \geq \beta I_m$. $\square$

Lemma 10 *On the sublevel set $\{\mathbf{y} : f(\mathbf{y}) \leq f(\mathbf{y}_0)\}$, all eigenvalues $\lambda_i(A(\mathbf{y}))$ are uniformly bounded:*

$$|\lambda_i(A(\mathbf{y}))| \leq \frac{n-1}{n}(f(\mathbf{y}_0) + \log n)$$

and in particular if $\mathbf{y}_0 = 0$ this simplifies to

$$|\lambda_i(A(\mathbf{y}))| \leq \frac{2(n-1)\log n}{n}.$$

Proof: The sum of the eigenvalues of $A(\mathbf{y})$ is zero: $\text{tr}A(\mathbf{y}) = \sum_{i=1}^n \lambda_i(A(\mathbf{y})) = 0$. Define the spectral gap $\delta(\mathbf{y}) := \lambda_{\max}(A(\mathbf{y})) - \lambda_{\min}(A(\mathbf{y}))$. By standard log-partition duality one shows $\delta(\mathbf{y}) \leq f(\mathbf{y}) - f(\mathbf{y}^*) \leq f(\mathbf{y}_0) - \min f \leq f(\mathbf{y}_0) + \log n$, using $f(\mathbf{y}^*) \geq -\log n$ for the minimizer $\mathbf{y}^*$. Hence for all $\mathbf{y}$ in the sublevel set $\{\mathbf{y} : f(\mathbf{y}) \leq f(\mathbf{y}_0)\}$, it holds $\delta(\mathbf{y}) \leq f(\mathbf{y}_0) + \log n$. Moreover, from the zero-trace condition $0 = \sum_{i=1}^n \lambda_i = \geq \lambda_{\max} + (n-1)\lambda_{\min}$ so $\lambda_{\max}(A(\mathbf{y})) \leq -(n-1)\lambda_{\min}(A(\mathbf{y}))$. Combine with $\delta = \lambda_{\max} - \lambda_{\min}$ to get $\lambda_{\max} \leq \frac{n-1}{n}\delta$ and $\lambda_{\min} \geq -\frac{n-1}{n}\delta$. Substituting the upper bound on δ yields the claimed uniform spectral bound. $\square$

7 Complexity analysis

Now we are fully equipped to analyse the convergence and computational complexity of L-BFGS for minimizing f with normalized data.

Theorem 5 *Under Assumption 2, let $\mathbf{y}_k$ denote the L-BFGS iterates (with exact line-search). Given $\epsilon > 0$, in order to guarantee $\|\nabla f(\mathbf{y}_k)\| \leq \epsilon$ it suffices to take*

$$k \geq n^2 \exp f(\mathbf{y}_0) \log(\frac{1}{\epsilon} \sqrt{f(\mathbf{y}_0) - \log \beta}).$$

In particular if $\mathbf{y}_0 = 0$ this simplifies to

$$k \geq n^3 \log(\frac{1}{\epsilon} \sqrt{\log n - \log \beta}).$$

Proof: Since f is α-strongly convex and λ-smooth on $\mathbb{R}^m$, a standard result (e.g. for gradient descent with step $1/\lambda$) gives $f(\mathbf{y}_{k+1}) - f(\mathbf{y}^*) \leq (1 - \frac{\alpha}{\lambda})(f(\mathbf{y}_k) - f(\mathbf{y}^*))$, and by induction $f(\mathbf{y}_k) - f(\mathbf{y}^*) \leq (1 - \frac{\alpha}{\lambda})^k (f(\mathbf{y}_0) - f(\mathbf{y}^*))$. Moreover using the inequality $1 - t \leq e^{-t}$ for $0 < t < 1$, we obtain $f(\mathbf{y}_k) - f(\mathbf{y}^*) \leq \exp(-\frac{\alpha}{\lambda}k)(f(\mathbf{y}_0) - f(\mathbf{y}^*))$. By Lemma 9 we have the bound $f(\mathbf{y}^*) \geq \log \beta$, so $f(\mathbf{y}_0) - f(\mathbf{y}^*) \leq f(\mathbf{y}_0) - \log \beta$. Hence $f(\mathbf{y}_k) - f(\mathbf{y}*) \leq \exp(-\frac{\alpha}{\lambda}k)(f(\mathbf{y}_0) - \log \beta)$. By λ-smoothness, for any $\mathbf{y}$ we have $\|\nabla f(\mathbf{y})\|_2^2 \leq 2\lambda(f(\mathbf{y}) - f(\mathbf{y}^*))$. Applying this at $\mathbf{y} = \mathbf{y}_k$ gives $\|\nabla f(\mathbf{y}_k)\|_2 \leq \sqrt{2\lambda(f(\mathbf{y}_k) - f(\mathbf{y}^*))} \leq \sqrt{2\lambda(f(\mathbf{y}_0) - \log \beta)} \exp(-\frac{\alpha}{2\beta}k)$. To ensure $\|\nabla f(\mathbf{y}_k)\|_2 \leq \epsilon$, we require $k \geq \frac{2\lambda}{\alpha} \log(\frac{1}{\epsilon} \sqrt{2\lambda(f(\mathbf{y}_0) - \log \beta)})$. The final expressions are obtained by letting $\lambda = \frac{1}{2}$ (Theorem 3) and $\alpha = \frac{1}{n^2} \exp(-f(\mathbf{y}_0))$ (Theorem 4). $\square$

Theorem 6 *The cost of one iteration for L-BFGS with memory m_{hist} and exact line-search is $O(n^3 + mn^2 + m_{\text{hist}}^2)$.*

Proof: Each iteration involves:

- Matrix exponential and normalization: diagonalize $A(\mathbf{y})$ in $O(n^3)$ to form $X = \exp_1 A(\mathbf{y})$.

- Gradient evaluation: compute $A(X) - \mathbf{b}$, requiring m inner-products $\text{tr}(A_i X)$ at $O(n^2)$ each, total $O(mn^2)$.

- Two-loop recursion: update the L-BFGS direction in $O(m_{\text{hist}}^2)$.

$\square$

8 Feasibility versus infeasibility

Let us now relax Assumption 2 and distinguish two cases.

8.1 Weak feasibility

Weak feasibility means $\mathbf{b} \notin \text{int}\,\mathscr{M}$ but $\mathbf{b} \in \mathscr{M}$, i.e. $\mathbf{b}$ lies along the boundary of the moment body. Then f remains convex and finite for all $\mathbf{y}$. It grows in $O(\log \mathbf{y})$ along the unique

supporting direction $\mathbf{u}$ with $\lambda_{\max}(A(\mathbf{u})) = \mathbf{b}^T \mathbf{u}$, and linearly in all other directions. The Hessian of f is positive semidefinite but degenerates as $\|\mathbf{y}\| \to \infty$ in direction $\mathbf{u}$. No finite minimizer exists and L-BFGS iterates drift off along $\mathbf{u}$. The gradient norm decays only in $O(1/\|\mathbf{y}\|)$, so convergence stalls. Sublevel sets $\{\mathbf{y} : f(\mathbf{y}) \leq f(\mathbf{y}_0)\}$ are unbounded in the direction $\mathbf{u}$.

Corollary 1 *Fix a tolerance $\epsilon > 0$. At iteration k, if $\|\nabla f(\mathbf{y}_k)\|_2 \leq \epsilon$, then setting $X_k = \exp_1 A(\mathbf{y}_k)$ yields a matrix $X_k \in \mathbb{S}_1^n$ satisfying $\|A(X_k) - b\|_2 \leq \epsilon$, therefore certifying that $\mathbf{b}$ lies within distance ϵ of $\mathcal{M}$.*

Proof: From Lemma 1 it holds $\nabla f(\mathbf{y}) = A(X(\mathbf{y})) - \mathbf{b}$, independently of the location of $\mathbf{b}$. Hence

$$\|\nabla f(\mathbf{y})\|_2 = \|A(X(\mathbf{y})) - \mathbf{b}\|_2 \geq \min_{X \succeq 0,\, \mathrm{tr} X = 1} \|A(X) - \mathbf{b}\|_2 = \min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{x} - \mathbf{b}\|_2$$

since $X(\mathbf{y})$ is one particular feasible point in the minimum defining the distance of $\mathbf{b}$ to $\mathcal{M}$. In particular, if $\|\nabla f(\mathbf{y})\|_2 \leq \varepsilon$ then $\min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{x} - \mathbf{b}\|_2 \leq \varepsilon$, i.e. b lies within ε of $\mathcal{M}$. $\qquad\square$

Corollary 2 *At iteration k, if $\|\mathbf{y}_k\| > \sqrt{n} \log \frac{1}{\beta}$, then $\mathbf{b} \notin \mathrm{int}\,\mathcal{M}$.*

Proof: By Lemma 9, any interior feasible $\mathbf{b}$ forces the sublevel set $\{\mathbf{y} : f(\mathbf{y}) \leq f(\mathbf{y}_0)\}$ to lie inside the ball $\{\mathbf{y} : \|\mathbf{y}\| \leq \sqrt{n} \log \frac{1}{\beta}\}$. $\qquad\square$

8.2 Infeasibility

If $\mathbf{b} \notin \mathcal{M}$ then f is convex and unbounded below. There exists $\mathbf{u} \in \mathcal{U}$ so that for all large t, $f(t\mathbf{u}) \approx t(\lambda_{\max}(A(\mathbf{u})) - \mathbf{b}^T \mathbf{u}) \to -\infty$.

Lemma 11 *If $f(\mathbf{y}) < 0$ for some $\mathbf{y} \in \mathbb{R}^m$, then $b \notin \mathcal{M}$ and the vector $\mathbf{y}/\|\mathbf{y}\| \in \mathcal{U}$ is a certificate of infeasibility.*

Proof: The support function of $\mathcal{M}$ is $h(\mathbf{u}) := \max_{X \in \mathbb{S}_1^n} \mathbf{u}^T A(X) = \lambda_{\max}(A(\mathbf{u}))$. By convex separation: $\mathbf{b} \notin \mathcal{M}$ if and only if $\mathbf{b}^T \mathbf{u} > \lambda_{\max}(A(\mathbf{u}))$. On the other hand, since $\log \mathrm{tr} \exp X \geq \lambda_{\max}(X)$ for all $X \in \mathbb{S}^n$, we have $f(\mathbf{y}) = \log \mathrm{tr} \exp A(\mathbf{y}) - \mathbf{b}^T \mathbf{y} \geq \lambda_{\max}(A(\mathbf{y})) - \mathbf{b}^T \mathbf{y}$. Hence if $f(y) < 0$ then $\mathbf{b}^T \mathbf{y} - \lambda_{\max}(A(\mathbf{y})) > 0$ and the normalized vector $\mathbf{y}/\|\mathbf{y}\| \in \mathcal{U}$ yields a strict separating hyperplane certifying $b \notin M$. $\qquad\square$

Corollary 3 *If $\mathbf{b} \notin \mathcal{M}$, then $\inf_{\mathbf{y}} f(\mathbf{y}) = -\infty$. Moreover, when L-BFGS with exact line-search is applied to f, the iterates satisfy $f(\mathbf{y}_{k+1}) < f(\mathbf{y}_k)$ for all k, and hence there exists a finite index k for which $f(\mathbf{y}_k) < 0$.*

Proof: Since $\mathbf{b} \notin \mathscr{M}$, by Lemma 11 there is a vector $\mathbf{u}$ with $\lambda_{\max}(A(\mathbf{u})) < \mathbf{b}^T \mathbf{u}$, whence $f(\mathbf{u}) \leq \lambda_{\max}(A(\mathbf{u})) - \mathbf{b}^T \mathbf{u} < 0$ and $\sup_{t>0} f(t\mathbf{u}) \to -\infty$, so $\inf f = -\infty$.

Under exact line-search L-BFGS is a descent method: at each step, provided $\nabla f(\mathbf{y}_k) \neq 0$, the value of f strictly decreases: $f(\mathbf{y}_{k+1}) < f(\mathbf{y}_k)$. From Lemma 1, no stationary point exists when $\mathbf{b} \notin \mathscr{M}$, so $\nabla f(\mathbf{y}_k) \neq 0$ for all k. Hence $f(\mathbf{y}_k)$ is strictly decreasing and unbounded below. Since $f(\mathbf{y}_0)$ is finite, there must be some finite k at which $f(\mathbf{y}_k)$ crosses zero, i.e. $f(\mathbf{y}_k) < 0$. $\qquad\square$

Corollary 4 *At iteration k, if $f(\mathbf{y}_k) < 0$ then $\mathbf{y}_k/\|\mathbf{y}_k\|$ is a certificate of infeasibility implying $\mathbf{b} \notin \mathscr{M}$, and the algorithm may terminate.*

Finally, straightforward sufficient conditions for infeasibility can be derived from Lemma 7: if $\|\mathbf{b}\| > \frac{n-1}{n}$ then $\mathbf{b} \notin \mathscr{M}$. Similarly, componentwise if $b_i \notin [\min_{X \in \mathbb{S}_1^n} \operatorname{tr}(A_i X), \max_{X \in \mathbb{S}_1^n} \operatorname{tr}(A_i X)] = [\lambda_{\min}(A_i), \lambda_{\max}(A_i)]$ for some $i = 1, \ldots, m$ then $\mathbf{b} \notin \mathscr{M}$.

9 Block-separable problems

We can consider a block separable version of the maximum entropy primal

$$\max_X \ \sum_{j=1}^p \Big[\operatorname{tr}(X_j) \ - \ \operatorname{tr}(X_j \log X_j) \Big]$$

$$\text{s.t.} \ \ \sum_{j=1}^p \operatorname{tr}\Big(A_{ij} X_j\Big) \ = \ b_i \quad i = 1, \ldots, m,$$

$$X \in \bigoplus_{j=1}^p \mathbb{S}^{n_j} \cap \mathbb{S}_1^{\sum_{j=1}^p n_j}$$

where the unknown matrix X is block diagonal with positive semidefinite blocks $X_j \in \mathbb{S}^{n_j}$ whose traces sum up to one. Its Lagrangian separates over j, and as in the proof of Theorem 1 one shows that the dual can be written as the unconstrained minimization

$$\min_{\mathbf{y} \in \mathbb{R}^m} \log \sum_{j=1}^p \operatorname{tr} \exp \sum_{i=1}^m y_i A_{ij} - \mathbf{b}^\top \mathbf{y}.$$

10 Numerical experiments

We constructed a basic Matlab implementation[4] of L-BFGS that takes as input matrix A of size m-by-n^2 and a vector $\mathbf{b}$ of size m, and returns a vector $\mathbf{y}$ of size m minimizing f:

```
y = maxentmom(A,b);
```

---

[4]Available for download at homepages.laas.fr/henrion/software/maxentmom/maxentmom.m

The algorithm calls the following function which evaluates f and its gradient:

```
function [val, grad] = logpart(A,b,y)
% A : matrix of size m by n^2
% b, y : vectors of size m
n = sqrt(size(A,2));
[V,D] = eig(reshape(A'*y,n,n));
X = V * diag(exp(diag(D))) * V';
t = trace(X);
val = log(t) - y'*b; % f(y)
grad = (A*X(:))/t - b; % grad f(y)
end
```

Alternatively we can use HANSO [26] which is a Matlab implementation of L-BFGS also aimed at non-smooth non-convex problems.

Convergence of iterate $\mathbf{y}_k$ occurs when the norm of the residual $\mathcal{A}(\exp_1 \mathcal{A}^T(\mathbf{y}_k)) - \mathbf{b}$ (i.e. the gradient of f at $\mathbf{y}_k$) is smaller than some a priori given expected accuracy, typically 10^{-8}. This is a relative accuracy when the data is normalized via Algorithm 1, since the norm of $\mathbf{b}$ is less than one by Lemma 7.

10.1 Toy problem

Let us illustrate the behavior of `maxentmom` on our toy planar moment body of Example 3. On Figure 6 are represented 10 typical trajectories $\mathcal{A}(\exp_1 \mathcal{A}^T(\mathbf{y}_k))$ for 10 different target vectors $\mathbf{b}$ chosen close to the boundary of the moment body, with the same initial condition $\mathbf{y}_0 = 0$. Iterates are represented by black dots, and typically 7 iterations suffice to reach the target vector at accuracy 10^{-8}.

10.2 Medium scale problems

Our implementation largely outperforms the state-of-the-art second-order interior-point semidefinite solver of MOSEK. For example, on our standard laptop, with $m = n = 300$ and accuracy 10^{-8}, `maxentmom` takes 0.2s and 7 iterations to solve a random problem, compared to 180s with MOSEK. Random problems are generated as follows. We apply pre-conditioning algorithm 1 on a normally distributed random map A, we let X be the normalized exponential of a normally distributed random symmetric matrix, and we choose $\mathbf{b} = \mathcal{A}(X)$.

10.3 Larger scale problems

Our rudimentary implementation stands the comparison with SDPNAL+1.0, a state of the art large-scale semidefinite solver based on semismooth Newton-CG augmented Lagrangian [36, 31], see Figure 7 which corresponds to normalized randomly generated instances as described in the previous section. For $m = n = 1000$ and expected accuracy 10^{-8}, `maxentmom`
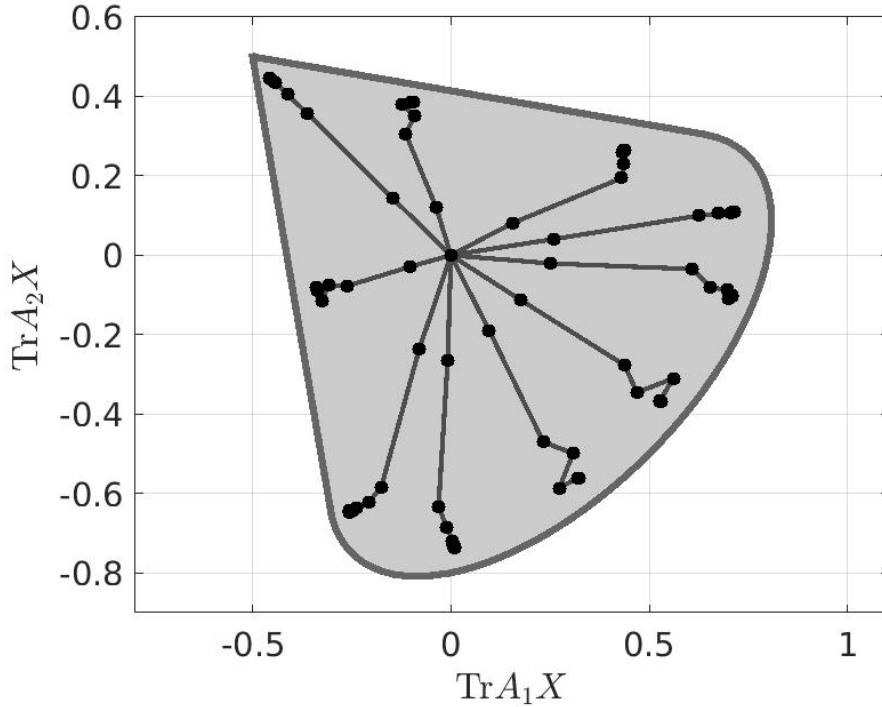
Figure 6: Typical iterates (black dots) starting from the origin and reaching various target vectors near the boundary (dark gray) of the moment body (light gray).

solves a randomly generated moment body membership problem in less than 5s. For larger problems, it is too costly to store the linear map as a single matrix A of size m-by-n^2, and other storage and matrix vector multiplication strategies must be followed. For illustration, when $m = n = 1000$, storing a double precision matrix A requires almost 8 gigabytes.

11 Conclusion

Motivated by pre-conditioning strategies for semidefinite optimization, this paper reports on a specific problem class whose geometry is simple enough to allow for a comprehensive analysis. We consider the moment body membership oracle problem, which consists of determining whether a given vector of size m belongs to a given linear projection of the spectraplex, the compact convex set of unit trace positive semdefinite matrices of size n-by-n. Inspired by maximum entropy techniques from quantum information theory, we propose to solve the problem by minimizing on the whole m-dimensional space a dual smooth strictly convex log-partition function. Geometric curvature analysis reveals how key input data quantities can be modified to improve the problem conditioning. After pre-conditioning, we can solve the convex dual problem with L-BFGS, a widely used first-order algorithm approximating second-order information with limited gradient evaluation and storage. Numerical experiments on a rudimentary Matlab implementation show that the approach largely outperforms second-order interior-point methods, while standing the comparison with state-
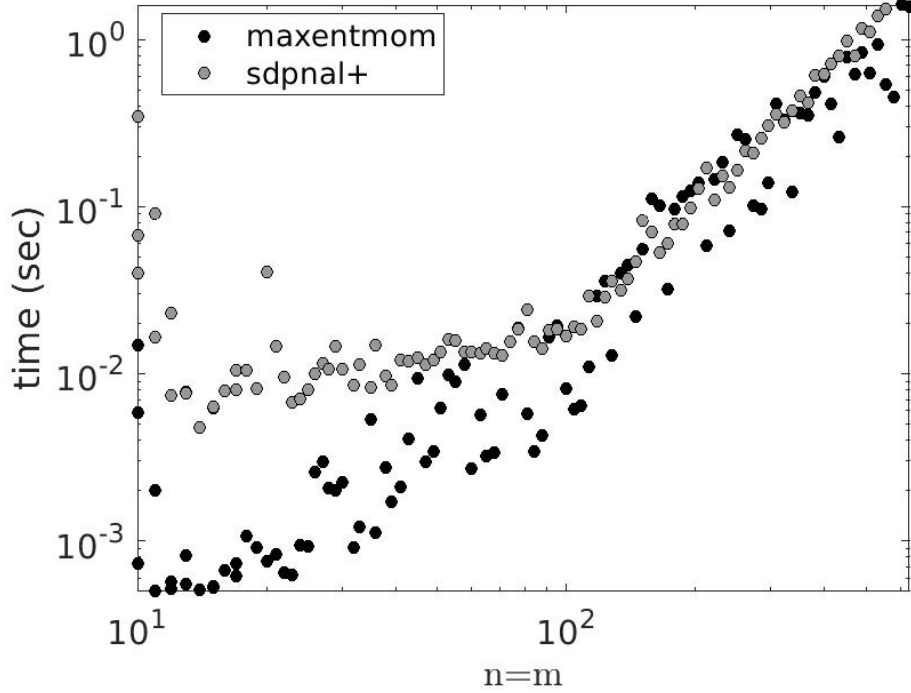
Figure 7: Comparative computational times (logarithmic scales).

of-the-art first-order algorithms for large-scale semidefinite optimization. Fully dense (i.e. non-sparse) problems of size $n = m = 1000$ can be solved to 8 significant digits in a few seconds on a standard laptop, the only limitation being the memory requirements. For this problem class, it means that the bottleneck is now essentially concentrated into efficient gradient computation and storage, consistently with the recent developments reported e.g. in [37, 20].

Polynomial SOS decompositions are particular cases of the moment body membership oracle, where X is the Gram matrix representing a polynomial $p(\mathbf{x}) = \phi^T(\mathbf{x})X\phi(\mathbf{x})$ as a quadratic form w.r.t. some basis vector ϕ. The linear map $\mathcal{A}(X) = \mathbf{b}$ matches X with p expressed as a coefficient vector $\mathbf{b}$ in some basis. It can be normalized since $\int p(\mathbf{x})d\mu(\mathbf{x}) = \text{tr}(X \int \phi(\mathbf{x})\phi^T(\mathbf{x})d\mu(\mathbf{x})) = \text{tr}X$ whenever ϕ is an orthonormal basis with respect to the inner product induced by μ. Memberships in truncated quadratic modules, also called weighted SOS decompositions, can also be modeled as particular moment body problems. They are at the core of the moment-SOS hierarchy for polynomial optimization [10]. It would be interesting to derive specific curvature properties to pre-condition these problems in the same way we did it for general moment bodies. Relationships with the dual certificates of truncated quadratic module membership investigated in [7] are also worth investigating, especially since these dual certificates allow to construct SOS representations with rational coefficients.

In the context of semidefinite relaxations of combinatorial optimization problems, the trace one constraint holds for the first relaxation of the moment-SOS hierarchy. This constant trace

property was exploited in [8] in the context of spectral bundle methods. It was generalized in [16, 17] where it was shown that every every polynomial optimization problem on a compact semialgebraic set has an equivalent equality constrained formulation on an sphere (possibly after adding some artificial variables), and hence a constant trace moment relaxation.

A natural extension of our approach consists of minimizing a linear function on the moment body, i.e. given a matrix $C \in \mathbb{S}^n$, solving the semidefinite optimization problem

$$\min_{X \in \mathbb{S}_1^n} \operatorname{tr}(CX) \quad \text{s.t.} \quad \mathcal{A}(X) = \mathbf{b}.$$

For a given regularization parameter $\mu > 0$, to a primal entropic problem

$$\min_{X \in \mathbb{S}_1^n} \operatorname{tr}(CX) - \mu \operatorname{tr}(X - X \log X) \quad \text{s.t.} \quad \mathcal{A}(X) = \mathbf{b}$$

corresponds a dual log-partition problem

$$\max_{\mathbf{y} \in \mathbb{R}^m} \mathbf{b}^T \mathbf{y} - \mu \, \log \operatorname{tr} \exp(-\tfrac{1}{\mu}(C - A(\mathbf{y}))).$$

One then follows a primal admissible central path

$$X_\mu^* = \exp_1(-\frac{1}{\mu}(C - A(\mathbf{y}_\mu^*))) \in \mathcal{A}^{-1}(\mathbf{b})$$

parametrized by dual optimal solutions $\mathbf{y}_\mu^*$ and we let $\mu \to 0^+$. A detailed analysis of convergence of this method remains to be done. Note that the idea was followed recently in [15, 6], but without the trace-one restriction. Consequently, the dual function there is the much less regular partition function, which is the exponential of the log-partition function. This may explain why the semidefinite optimization experiments reported in [6] are somewhat disappointing. Whether more convincing and scalable numerical results can be obtained with the log-partition function remains however to be seen.

Acknowledgement

References

[1] S.-I. Amari, H. Nagaoka. Methods of information geometry. Translations of Mathematical Monographs. Translations of Mathematical Monographs, 191, Amer. Math. Soc., 2000. Translated from the Japanese original of 1993.

[2] I. Bengtsson, K. Życzkowski. Geometry of quantum states. Cambridge University Press, 2nd edition, 2017.

[3] A. Ben-Tal, A. Nemirovski. Lectures on modern convex optimization: analysis, algorithms, and engineering applications. SIAM, 2001.

[4] R. Bhatia, C. Davis. A better bound on the variance. Amer. Math. Monthly 107(4):353-357, 2000.

[5] S. Boyd and L. Vandenberghe. Convex optimization. Cambridge Univ. Press, 2004.

[6] S. P. Chhatoi, J. B. Lasserre. Shannon- and von Neumann-entropy regularizations of linear and semidefinite programs. arXiv:2503.23815, 2025.

[7] M. M. Davis, D. Papp. Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets. SIAM J. Optim. 32(4):2461-2492, 2022.

[8] C. Helmberg, F. Rendl. A spectral bundle method for semidefinite programming. SIAM J. Optim. 10(3), 673–696, 2000.

[9] D. Henrion. Semidefinite geometry of the numerical range. Elec. J. Lin. Alg. 20:322-332, 2010.

[10] D. Henrion, M. Korda, J. B. Lasserre. The moment-SOS hierarchy. World Scientific, 2020.

[11] D. Henrion, J. Malick. Projection methods for conic feasibility problems, applications to polynomial sum-of-squares decompositions, Optim. Methods and Software, Vol. 26, No. 1, pp. 23-46, 2011.

[12] D. Henrion, J. Malick. Projection methods in convex optimization. In M. Anjos and J. B. Lasserre (Editors). Handbook of semidefinite, cone and polynomial optimization. Springer, 2012.

[13] S.-Y. Hou, Z. Wu, J. Zeng, N. Cao, C. Cao, Y. Li, B. Zeng. Maximum entropy methods for quantum state compatibility problems. Adv. Quantum Technol. 2400172, 2024.

[14] S. Jarov, M. Van Raamsbonk. Mapping the space of quantum expectation values. `arXiv:2310.13111`, 2023.

[15] M. Lindsey. Fast randomized entropically regularized semidefinite programming. arXiv:2303.12133, 2023.

[16] N. H. A. Mai, J. B. Lasserre, V. Magron. A hierarchy of spectral relaxations for polynomial optimization. Math. Prog. Comp. 15:651–701, 2023.

[17] N. H. A. Mai, J. B. Lasserre, V. Magron, J. Wang. Exploiting constant trace property in large-scale polynomial optimization. ACM Trans. Math. Software 48(4):1-39, 2022.

[18] J. Malick. A dual approach to semidefinite least-squares problems. SIAM J. Matrix Anal. Appl. Vol. 26, No. 1, pp. 272-284, 2004.

[19] J. Malick, J. Povh, F. Rendl, and A. Wiegele. Regularization methods for semidefinite programming. SIAM J. Optim 20(1):336–356, 2009.

[20] R. D. C. Monteiro, A. Sujanani, D. Cifuentes. A low-rank augmented Lagrangian method for large-scale semidefinite programming based on a hybrid convex-nonconvex approach. arXiv:2401.12490, 2024.

[21] J. Nie. Moment and polynomial optimization, SIAM, 2023.

[22] J. Niño-Cortes, C. Vinzant. The convex algebraic geometry of higher-rank numerical ranges. arXiv:2410.21625, 2024.

[23] Y. Nesterov. Introductory lectures on convex optimization: a basic course. Kluwer Academic Publishers, 2004.

[24] Y. Nesterov, A. Nemirovskii. Interior-point polynomial algorithms in convex programming. SIAM, 1994.

[25] J. Nocedal and S. J. Wright. Numerical optimization, 2nd edition. Springer, 2006.

[26] M. L. Overton. HANSO: Hybrid Algorithm for Non-Smooth Optimization (Version 3.0) [Software], 2021.

[27] G. Pataki. Characterizing bad semidefinite programs: normal forms and short proofs. SIAM Review 61(4):839-859, 2019.

[28] D. Pavlov, B. Sturmfels, S. Telen. Gibbs manifolds. Information Geometry 7:691–71, 2024.

[29] D. Plaumann, R. Sinn, S. Weis. Kippenhahn's theorem for joint numerical ranges and quantum states. SIAM J. Appl. Alg. Geom. 5(1):86-113, 2021.

[30] J. Renegar. Linear programming, complexity theory and elementary functional analysis. Math. Prog. 70:279-351, 1995.

[31] D. F. Sun, K. C. Toh, Y. C. Yuan, X. Y. Zhao, SDPNAL+: A Matlab software for semidefinite programming with bound constraints (version 1.0), Optim. Methods and Software, 35:87–115, 2020.

[32] T. Theobald. Real algebraic geometry and optimization. AMS, 2024.

[33] S. Weis. Information topologies on non-commutative state spaces. J. Convex Anal. 21(2):339–399, 2014.

[34] E. H. Wichmann. Density matrices arising from incomplete measurements. J. Math. Phys. 4:884–896, 1963.

[35] S. J. Wright. Primal–Dual Interior-Point Methods. SIAM, 1997.

[36] L. Q. Yang, D. F. Sun, K. C. Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, Math. Prog. Comp. 7:331-366, 2015.

[37] A. Yurtsever, J. A. Tropp, O. Fercoq, M. Udell, V. Cevher. Scalable semidefinite programming. SIAM J. Math. Data Sci. 3(1), 2021

[38] R. M. Wilcox. Exponential operators and parameter differentiation in quantum physics. J. Math. Phys. 8(4):962-982, 1967.