Duality of ellipsoidal approximations via semi-infinite programming

Filiz Gürtuna*

March 9, 2008

Abstract

In this work, we develop duality of the minimum volume circumscribed ellipsoid and the maximum volume inscribed ellipsoid problems. We present a unified treatment of both problems using convex semi–infinite programming. We establish the known duality relationship between the minimum volume circumscribed ellipsoid problem and the optimal experimental design problem in statistics. The duality results are obtained using convex duality for semi–infinite programming developed in a functional analysis setting.

Key words. Inscribed ellipsoid, circumscribed ellipsoid, minimum volume, maximum volume, duality, semi–infinite programming, optimality conditions, optimal experimental design, D-optimal design, John ellipsoid, Löwner ellipsoid.
Duality of ellipsoidal approximations via semi-infinite programming
AMS(MOS) subject classifications: primary: 90C34, 46B20, 90C30, 90C46, 65K10; secondary: 52A38, 52A20, 52A21, 22C05, 54H15.

*Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA. E-mail: *gurtuna1@math.umbc.edu*. Research partially supported by the National Science Foundation under grant DMS–0411955.

1 Introduction

Inner and outer approximations of complicated sets by simpler ones is a fundamental approach in applied mathematics. In this context, ellipsoids are very useful objects with their remarkable properties such as simple representation, an associated convex quadratic function, and invariance, that is, an affine transformation of an ellipsoid is an ellipsoid. Therefore, ellipsoidal approximations are easier to work with than polyhedral approximations and their invariance property can be a gain over the spherical approximations in some applications such as clustering.

Different criteria can be used for the quality of an approximating ellipsoid, each leading to a different optimization problem. These criteria are given in terms of the "size" of the *shape matrix* that defines the approximating ellipsoid. Among these criteria, ratio of volumes has many applications including optimization, computational geometry, and statistics. This popularity is also a result of the existence of (practical and fast) numerical methods for these problems (see, e.g., [27, 25, 14, 1]). Although ratio of volumes is the most studied criterion (and the most common one in applications), recently other criteria have been considered as well. For instance, the average axis length is optimized in [17] and obtaining a "rounder" ellipsoid, in addition to ratio of volumes, is considered in [5]. We leave these problems as future study.

As we mentioned before, the quality of an approximating ellipsoid is related to the size of its shape matrix. A similar situation arises in the context of *optimal experimental design* problem in statistics, see [19] for an excellent treatment of the subject. Consider the linear regression problem with $y = \Theta^T s + \epsilon$, where ϵ is the error term with mean zero. The design problem is to select points $\{s_i\}$ at which to conduct experiments with the aim of estimating the unknown parameter vector Θ in an "optimal way". The points are to be chosen from a compact set S, the so-called design space, which spans $\mathbb{R}^n$. In [13], Kiefer and Wolfowitz replace the selection of points $\{s_i\}$ by the selection of a probability measure ξ on S. For a design ξ, the *Fisher information matrix*, or the inverse of the covariance matrix, (see, e.g., Section 3.10 in [19]) is given by

$$M(\xi) := \left[\int_S ss^T \, d\xi(s)\right], \tag{1.1}$$

which corresponds also to the *moment matrix*. The quality of a design ξ is measured in terms of the size of $M(\xi)$ and again, different criteria can be used on this purpose. When the optimality criterion is the determinant of $M(\xi)$, this problem is called *D-optimal* design problem and is related to the ellipsoidal approximation problem by duality.

From now on, we focus on the ellipsoidal approximation problem using ratio of volumes as the optimality criterion. More formally, let S be a convex body in $\mathbb{R}^n$, that is, S is a compact convex set with nonempty interior. Among the ellipsoids circumscribing S, we are interested in the one with minimum volume and similarly, among the ellipsoids inscribed in S, we are interested in the one with maximum volume. We refer to these extremal volume ellipsoids as the *minimum volume circumscribed ellipsoid* and the *maximum volume inscribed ellipsoid* of S and denote them by $\mathrm{CE}(S)$ and $\mathrm{IE}(S)$, respectively. For the existence and uniqueness of $\mathrm{CE}(S)$ and $\mathrm{IE}(S)$ along with the necessary and sufficient optimality conditions obtained using semi-infinite programming, we refer the reader to [10, 4, 11, 8].

In this paper, our main interest is in the duality of the extremal volume ellipsoid problems. Duality results play an important role in the development of numerical methods. For instance, the dual structure (or the necessary and sufficient optimality conditions) is exploited by the first-order methods in [12, 14, 1] and the interior point methods in [27, 25].

To our knowledge, a dual problem independent of the primal variables is not known for the maximum volume inscribed ellipsoid problem. This is also the case for our duality results in Section 5. Our approach in Section 5 is similar to [27], except that we consider a convex body defined by a (not necessarily finite) set of hyperplanes.

For the minimum volume ellipsoid problem, the study of duality goes back to 1970s. The first result, conjectured by Silvey, is given by Sibson in the discussion part of [15]. Sibson proves that with center fixed at the origin, the minimum volume circumscribed ellipsoid problem is dual to the D-optimal design problem. In the case of general center, Titterington [26] concludes a similar duality relationship using the duality results of Silvey and Titterington [24] for the "thinnest cylinder problem". The case of general center is also considered in [12] for a convex body Q given by $Q = \text{conv}\{q_1, \ldots, q_m\}$. In [12], Khachiyan and Todd show that finding $\text{CE}(Q)$ is equivalent to finding $\text{CE}(Q')$, where $Q' = \text{conv}\{\mp(q_1, 1), \ldots, \mp(q_m, 1)\}$. In their proof, the volume of $\text{CE}(Q)$ is related to the volume of $\text{CE}(Q')$. To our knowledge, they do not provide a proof of this relationship though. In Section 6.1, we consider convex bodies of the form Q' and arrive at the mentioned relationship between the two volumes as a consequence of our duality results.

Our goal and contribution in this paper is to provide a rigorous, unified, and modern treatment of duality of the extremal volume ellipsoid problems. On this purpose, we develop the convex duality for semi-infinite programming in a functional analysis setting (in Section 3). Although our treatment in Section 3 is not completely new (see, e.g., [23]), it is not a common practice in developing duality results. The convex duality for semi-infinite programming turns out to be a natural way of obtaining the duals of the extremal volume ellipsoid problems.

The organization of the paper is as follows. In Section 2, we present some preliminaries in functional analysis, basically the Riesz Representation Theorem. Then, we develop convex duality for semi-infinite programming in Section 3. The duality results of Section 3 is then used to obtain the dual of the minimum volume circumscribed ellipsoid problem in Section 4 and the dual of the maximum volume inscribed ellipsoid problem in Section 5. In Section 6, we consider some special cases of the extremal volume ellipsoid problems. In the Appendix, we gather some of the results used in the main body of the paper and also present some of the technical proofs.

Notation: We denote by $\mathbb{S}^n$ the set of symmetric $n \times n$ matrices and by e_i the unit vector with one in the ith position. The bracket notation $\langle ., . \rangle$ represents the inner product in $\mathbb{R}^n$ ($\langle u, v \rangle = u^T v$) and the trace inner product in the vector space $\mathbb{R}^{n \times n}$ of $n \times n$ matrices ($\langle X, Y \rangle = \text{tr}(XY^T)$). Given a space X and its dual space X^*, a bilinear form between X and X^* will also be denoted by the bracket notation. The meaning of the bracket notation should be clear from the context. For a set S, ∂S, $\text{conv}(S)$, and $\text{ext}(S)$ represent the boundary, the convex hull, and the extreme points of S, respectively. We denote by $\Lambda(S)$, $\Lambda(S)^+$, $\Xi(S)$, and $\mathcal{C}(S)$ the set of Borel measures, the set of nonnegative Borel measures, the set of probability measures, and the space of continuous functions on S, respectively.

2 Preliminaries

In this section, we present some results in functional analysis and measure theory. One can see, for example, [22, 18] for details. We keep the same notations and definitions throughout the paper.

Let S be a compact metric space and $\mathcal{C}(S)$ the set of (bounded) continuous functions defined on S. This space equipped with the *supremum* norm is a Banach space. We denote the cone of nonnegative functions in $\mathcal{C}(S)$ by $\mathcal{C}(S)^+$. The dual problems we construct in the rest of the paper employ the *dual space* $\mathcal{C}(S)^*$ of $\mathcal{C}(S)$. Recall that $\mathcal{C}(S)^*$ is the space of continuous linear functionals f^* on $\mathcal{C}(S)$. In this paper, $\langle f^*, f\rangle$ represents such a bilinear form. The space $\mathcal{C}(S)^*$ equipped with the supremum norm is also a Banach space. We denote the cone of nonnegative linear functionals in $\mathcal{C}(S)^*$ by $(\mathcal{C}(S)^*)^+$. This is the polar cone of $\mathcal{C}(S)^+$.

The following theorem is a fundamental result in topological measures. It relates $\mathcal{C}(S)^*$ to $\Lambda(S)$, the space of countably additive (scalar) functions defined on the *Borel sets* of S. This relation is such that it preserves the norm (*isometric*) and the ordering. The norm in $\Lambda(S)$ is given by the total variation, that is,

$$\|\lambda\| = \sup \sum_{i=1}^{n} |\lambda(B_i)|, \quad \lambda \in \Lambda(S),$$

where the supremum is taken over all finite sequences $\{B_i\}$ of disjoint subsets of $\mathfrak{B}$ (the Borel σ-field of S) with $B_i \subseteq S$. Note that for a nonnegative measure λ, we have $\|\lambda\| = \lambda(S)$. Because of the isometry, the measures in $\Lambda(S)$ are finite and we refer to $\Lambda(S)$ as the space of *finite* (not necessarily nonnegative) *Borel measures* on S. The reader can find the proof of the theorem in [22] or other functional analysis books.

Theorem 2.1. *(**Riesz Representation Theorem**) Let S be a compact metric space. There is an isomorphism between the* dual space $\mathcal{C}(S)^*$ *of $\mathcal{C}(S)$ and $\Lambda(S)$ such that the corresponding elements $f^* \in \mathcal{C}^*(S)$ and $\lambda \in \Lambda(S)$ satisfy*

$$\langle f^*, f\rangle = \int_S f(s)\, d\lambda(s), \quad \forall\, f \in \mathcal{C}(S). \tag{2.1}$$

Furthermore, this isomorphism is isometric ($\|f^*\| = \|\lambda\|$) and preserves the order.

As an immediate consequence of this theorem, particularly of the order preserving property, we relate the cones of nonnegative functions in $\mathcal{C}(S)^*$ and $\Lambda(S)$.

Corollary 2.2. *Let $(\mathcal{C}(S)^*)^+$ be the polar cone of $\mathcal{C}(S)^+$. Then f^* is in $(\mathcal{C}(S)^*)^+$ if and only if λ is in $\Lambda(S)^+$ (the cone of nonnegative measures in $\Lambda(S)$), where f^* and λ satisfy* (2.1).

3 Duality and optimality conditions for semi–infinite programming

The mathematical apparatus in this paper is the convex duality for semi–infinite programming developed in a functional analysis setting. In this section, we develop this duality for semi–infinite programming by reformulating a convex semi–infinite problem as a cone constrained problem and then using the duality of the cone constrained problem.

Let S be a compact metric space and consider the convex semi–infinite problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x,s) \le 0, \quad \forall\, s \in S. \end{array} \tag{3.1}$$

Here, we assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function and $g : \mathbb{R}^n \times S \to \mathbb{R}$ is convex in x for each fixed s in S and continuous in s for each fixed x in $\mathbb{R}^n$.

Define the operator $G : \mathbb{R}^n \to C(S)$ by

$$G(x)(s) := g(x, s), \quad \forall\, s \in S.$$

Thus, problem (3.1) is equivalent to the cone constrained problem

$$\min\{f(x) :\ G(x) \le 0\} = \min\{f(x) :\ -G(x) \in K\}, \tag{P}$$

where $K = \mathcal{C}(S)^+$. Obviously (P) is a convex programming problem (since G is convex).

A dual problem of (P) is given in the next theorem as a direct consequence of the duality results of Ekeland and Temam [6, pp. 62-68]. Ekeland and Temam obtain the duality results for problems of type (P) as a special case of convex duality.

Theorem 3.1. *Consider problem (P) above and assume additionally that for each g^* in K^*, the bilinear form $\langle g^*, G(x)\rangle$ is lower semi continuous in x. The dual problem of (P) is given by*

$$\max_{g^* \in K^*} \min_{x \in \mathbb{R}^n} f(x) + \langle g^*, G(x)\rangle. \tag{D}$$

Suppose that (P) has a finite optimal value and that there exists an $\hat{x}$ in $\mathbb{R}^n$ satisfying $G(\hat{x}) < 0$ (the Slater's condition). Then, $\bar{x}$ solves (P) if and only if there exists $\bar{g}^*$ in K^* such that $(\bar{x}, \bar{g}^*)$ satisfies

$$f(\bar{x}) + \langle g^*, G(\bar{x})\rangle \le f(\bar{x}) + \langle \bar{g}^*, G(\bar{x})\rangle \le f(x) + \langle \bar{g}^*, G(x)\rangle, \quad \forall\, x \in \mathbb{R}^n,\ g^* \in K^*. \tag{3.2}$$

Remark 3.2. The dual problem (D) in Theorem 3.1 is constructed with respect to the Lagrangian function $L : \mathbb{R}^n \times \mathcal{C}(S)^* \to \mathbb{R}$ given by

$$L(x, g^*) := f(x) + \langle g^*, G(x)\rangle - \chi_{K^*}(g^*), \tag{3.3}$$

where χ_{K^*} is the indicator function of K^*, that is, $\chi_{K^*}(g^*) = 0$ if g^* lies in K^* and ∞ otherwise (we recall that $K^* = (\mathcal{C}(S)^*)^+$). It is easy to see that problem (P) is equivalent to

$$\min_{x \in \mathbb{R}^n} \max_{g^* \in K^*} L(x, g^*).$$

A point $(\bar{x}, \bar{g}^*)$ is called a saddle point of L if it satisfies (3.2). It follows from (3.2) that a point $(\bar{x}, \bar{g}^*)$ is a saddle point of L if and only if *i)* $\bar{x}$ solves (P), *ii)* $\bar{g}^*$ solves (D), and *iii)* (P) and (D) have the same optimal (objective) value. We observe that Theorem 3.1 guarantees the existence of a dual solution under the assumption that (P) has a finite optimal value. Additionally, the result of Theorem 3.1 also reads as follows: $\bar{x}$ solves (P) and $\bar{g}^*$ solves (D) if and only if they are feasible for the corresponding problems and satisfy

$$\bar{x} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} L(x, \bar{g}^*), \tag{3.4}$$

$$\langle \bar{g}^*, G(\bar{x})\rangle = 0. \tag{3.5}$$

To see this, let $(\bar{x}, \bar{g}^*)$ be a saddle point. Then, it follows from the equivalence of the optimal values of (P) and (D) that

$$f(\bar{x}) = \inf_{x \in \mathbb{R}^n} L(x, \bar{g}^*) = \inf_{x \in \mathbb{R}^n} f(x) + \langle \bar{g}^*, G(x)\rangle \le f(\bar{x}) + \langle \bar{g}^*, G(\bar{x})\rangle$$

and we obtain (3.4) and (3.5) since $G(\bar{x}) \leq 0$ and $\bar{g}^* \geq 0$ by virtue of feasibility. Similarly, for $(\bar{x}, \bar{g}^*)$ satisfying the above equations and the feasibility, the second inequality in (3.2) follows from (3.4) and the first one from (3.5) since $\langle g^*, G(\bar{x})\rangle \leq 0$. Hence $(\bar{x}, \bar{g}^*)$ is a saddle point of L.

Before we apply Theorem 3.1 to state our duality results for the semi-infinite problem (3.1), we present a lemma. We will use this lemma to show that the assumptions of Theorem 3.1 hold for problem (3.1) and also to characterize the minimum of $\{f(x) + \langle g^*, G(x)\rangle : x \in \mathbb{R}^n\}$, which appears in the dual problem in Theorem 3.1. The proof of the lemma is similar to the proof of Corollary 5.109 in [2] and we present it here for completeness.

Lemma 3.3. *Consider the setting in problem* (3.1). *Assume additionally that g is differentiable in x for each fixed s in S and that $\nabla_x g(x, s)$ is continuous (jointly in x and s). Let λ be in $\Lambda(S)$. Then, the function I defined by*

$$I(x) := \int_S g(x, s)\, d\lambda(s)$$

is differentiable at x and

$$\nabla_x I(x) = \int_S \nabla_x g(x, s)\, d\lambda(s).$$

Proof. Since g is continuous in s, the integral $I(x)$ above is well-defined. Obviously, I is convex since g is convex in x. Therefore, I is (Fréchet) differentiable at x if and only if the directional derivative of I at x is linear, see [20, Thm. 25.2, pp. 244].

Since g is differentiable, it follows from the Mean Value Theorem that the following equality holds for some z strictly between x and $x+td$ (this is represented by $z \in [x, x+td]$):

$$\begin{aligned} g(x + td, s) - g(x + td, s) &= \langle \nabla_x g(z, s), td\rangle \\ &\leq \sup_{\substack{z\in[x,x+td],\\ s\in S}} \|\nabla_x g(z, s)\|\|td\| =: \kappa\|td\|. \end{aligned}$$

Here, the existence of κ follows from the continuity of $\nabla_x g(x, s)$ (jointly in x and s). As a result, $|(g(x + td, s) - g(x, s))/t|$ is bounded from above by an integrable function $\kappa\|d\|$. Hence, it follows from the Lebesgue Dominated Convergence Theorem that

$$\begin{aligned} \lim_{t\downarrow 0} \frac{I(x + td) - I(x)}{t} &= \int_S \lim_{t\downarrow 0} \frac{g(x + td, s) - g(x, s)}{t}\, d\lambda(s) \\ &= \int_S \langle \nabla_x g(x, s), d\rangle\, d\lambda(s) = \langle \int_S \nabla_x g(x, s)\, d\lambda(s), d\rangle. \end{aligned}$$

Lastly, $\int_S \nabla_x g(x, s) d\lambda(s)$ is well-defined since $\nabla_x g(x, s)$ is continuous. This completes the proof. □

In Lemma 3.3, we make additional assumptions on the function g in problem (3.1). However, these assumptions are not restrictive for the applications we consider in this paper.

We remark that Theorem 3.1 holds in a more general setting. One then needs to study the continuity and differentiability of the operator G or of the convex integrals of the form $\int_S g(x, s)\, d\lambda(s)$. This is done in a Banach space setting, for example, in [9].

Next, we state our duality results for the semi–infinite problem (3.1) by applying Theorem 3.1.

Theorem 3.4. *Consider problem* (3.1) *and assume additionally that* g *is differentiable and* $\nabla_x g(x,s)$ *is continuous (jointly in* x *and* s*). A dual of problem* (3.1) *is given by*

$$\begin{aligned} \max \quad & f(x) + \int_S g(x,s)\, d\lambda(s) \\ \text{s.t.} \quad & \nabla_x f(x) + \int_S \nabla_x g(x,s)\, d\lambda(s) = 0, \\ & \lambda \in \Lambda(S)^+,\ x \in \mathbb{R}^n. \end{aligned} \tag{3.6}$$

Suppose that problem (3.1) *has a finite optimal value and that*

$$g(\hat{x}, s) < 0, \quad \forall\, s \in S \tag{3.7}$$

for some $\hat{x}$ *in* $\mathbb{R}^n$*. Then problem* (3.6) *has a solution. Furthermore,* $\bar{x}$ *solves problem* (3.1) *if and only if* $(\bar{\lambda}, \bar{x})$ *solves problem* (3.6) *and the following holds.*

$$g(\bar{x}, s) = 0 \ \bar{\lambda}\text{-almost everywhere (a.e.).} \tag{3.8}$$

Proof. It follows from Corollary 2.2 that $K^* = (\mathcal{C}(S)^*)^+ = \Lambda(S)^+$. Using the Riesz Representation Theorem 2.1, for each g^* in $\Lambda(S)^+$, we have $\langle g^*, G(x)\rangle = \int_S g(x,s)\, d\lambda(s)$ and $\langle g^*, G(x)\rangle$ is continuous by virtue of Lemma 3.3. Then, Theorem 3.1 applies (see also Remark 3.2) and problem (3.6) follows from (D). In writing problem (3.6), we use the Riesz Representation Theorem and the fact that, for λ fixed, the inner minimum in (D) is achieved at a point x if and only if

$$0 = \nabla_x L(x,\lambda) = \nabla_x f(x) + \nabla_x \int_S g(x,s)\, d\lambda(s)$$

since L is a convex function of x. Then, $\nabla_x L(x,\lambda) = \nabla_x f(x) + \int_S \nabla_x g(x,s)\, d\lambda(s)$ by virtue of Lemma 3.3. Now, it follows from Theorem 3.1 (see also (3.4) in Remark 3.2) that $\bar{x}$ solves problem (3.1) if and only if $(\bar{\lambda}, \bar{x})$ solves problem (3.6) and

$$\int_S g(\bar{x}, s)\, d\bar{\lambda}(s) = 0,$$

which is equivalent to (3.8) by a standard result in measure theory. □

The necessary and sufficient optimality conditions for problem (3.1) follows immediately as

Corollary 3.5. *Consider problem* (3.1) *and assume additionally that* g *is differentiable and* $\nabla_x g(x,s)$ *is continuous (jointly in* x *and* s*). Then,* $\bar{x}$ *solves problem* (3.1) *if and only if the following is satisfied for some nonnegative Borel measure* λ *on* S*:*

$$\begin{aligned} & 0 = \nabla_x f(\bar{x}) + \int_S \nabla_x g(\bar{x}, s)\, d\lambda(s), \\ & 0 = g(\bar{x}, s) \ \lambda\text{-a.e.}, \\ & 0 \geq g(\bar{x}, s), \quad \forall\, s \in S. \end{aligned}$$

The last result of this section is on the existence of a solution, which is a discrete measure, to the dual problem of (3.1). Existence of such a discrete measure is important for computational purposes since the dual problem can then be written in finite dimensions. The theorem (except the given bound $m \leq n$) follows from Theorem 1 (or Theorem 2) in [21]. This bound is easy to show using some linear dependence arguments.

Let us denote the *point masses* (or *Dirac measures*) on S by δ_s, that is $\delta_s(B) = 1$ if $s \in B$ and 0 otherwise for any Borel set B of S.

Theorem 3.6. *If* $(\bar{\lambda}, \bar{x})$ *solves the dual problem* (3.6)*, then there exists a solution* $(\lambda, \bar{x})$ *to the same problem, where* λ *is a discrete measure, i.e.,* $\lambda = \sum_{i=1}^{m} \lambda_i \delta_{s_i}$, $\lambda_i \geq 0$ $i = 1, \ldots, m$ *with* $m \leq n$ *and* $\sum_{i=1}^{m} \lambda_i = \bar{\lambda}(S)$.

We remark that in treating the semi-infinite problem (3.1), we first obtained the duality results and then discussed the existence of a discrete measure as a solution to the dual problem. This existence result, in turn, implies the existence of finitely many active constraints in the primal problem. Contrary to this approach, one can first show the existence of a subproblem, which has the same optimal value but which has only finitely many active constraints. This is done using Helly's Theorem, for example, in [16, 3].

4 Dual of the minimum volume circumscribed ellipsoid problem

We recall that the minimum volume circumscribed ellipsoid problem is the problem of finding an ellipsoid of minimum volume circumscribing a convex body S in $\mathbb{R}^n$. We denote such an ellipsoid by $\mathrm{CE}(S)$. In this section, we are interested in constructing a dual of this problem using the duality results given in Section 3, especially Theorem 3.4.

An ellipsoid $\mathcal{E}$ in $\mathbb{R}^m$ is an affine image of the unit ball $B_n := \{u \in \mathbb{R}^n : \|u\| \leq 1\}$, that is,

$$\mathcal{E}(T, c) = T(B_n) + c = \{Tu + c : u \in B_n\} \subset \mathbb{R}^m,$$

where T is any $m \times n$ matrix. Here, c is the center and T is the shape matrix of the ellipsoid. The volume of $\mathcal{E}(T, c)$ is given by $\mathrm{vol}(\mathcal{E}(T, c)) = \det[T]\omega_n$, where ω_n is the volume of the unit ball B_n. In this paper, we are interested in the case when $\mathcal{E}$ has a nonempty interior. Therefore, we assume that T is a nonsingular $n \times n$ matrix. Furthermore, it follows from the polar decomposition of T that T can be assumed, without loss of generality, symmetric. Hence, from now on, we assume that T is symmetric positive definite.

Note that $\mathcal{E}$ is equivalent to

$$\mathcal{E}(T, c) = \{x \in \mathbb{R}^n : T^{-1}(x - c) \in B_n\}.$$

Using change of variables $X := T^{-1}$ and $d := Xc$, the problem of finding $\mathrm{CE}(S)$ can be posed as

$$\begin{array}{ll} \min & -\log\det[X] \\ \mathrm{s.t.} & \langle Xs - d, Xs - d\rangle \leq 1, \quad \forall\, s \in S. \end{array} \tag{4.1}$$

In this problem, the decision variables are $(X, d) \in \mathbb{S}^n \times \mathbb{R}^n$ with X positive definite. Obviously, this is a convex semi-infinite programming problem.

The following theorem is one of the main results in this paper.

Theorem 4.1. *Let S be a convex body in $\mathbb{R}^n$. A dual of problem (4.1) is given by*

$$\max_{\xi \in \Xi(S)} \frac{1}{2} \log\det \left[\int_S (s-\tilde{s})(s-\tilde{s})^T \, d\xi(s)\right] + \frac{n}{2}\log n, \tag{4.2}$$

where $\tilde{s} = \int_S s \, d\xi(s)$ and $\Xi(S)$ is the set of probability measures on S. Problem (4.2) has a solution. Furthermore, $(\bar{X}, \bar{d})$ solves problem (4.1) if and only if $\bar{\xi}$ solves problem (4.2) and the following is satisfied:

$$\|\bar{X}s - \bar{d}\| = 1 \;\; \bar{\xi}\text{-a.e.}. \tag{4.3}$$

Also, the two solutions are related by

$$\bar{X}^{-2} = n \int_S (s-\tilde{s})(s-\tilde{s})^T \, d\bar{\xi}(s), \qquad \bar{d} = \bar{X}\tilde{s}. \tag{4.4}$$

Proof. The proof uses Theorem 3.4. We claim that the assumptions of Theorem 3.4 are satisfied: It is obvious that problem (4.1) is a special case of problem (3.1). Also, $\langle Xs - d, Xs - d\rangle$ is differentiable in (X, d) on S and its derivative is continuous (jointly in (X, d) and s), see Lemma 7.1 in Appendix. Now, the existence of a solution to problem (4.1) follows from the continuity of the objective function and the compactness of the feasible set (see, e.g., [10, 8]). Lastly, since S is a convex body, there exists a ball of radius R centered at the origin containing S. Hence, (3.7) is satisfied. This proves the claim.

It follows from Theorem 3.4, that the problem (with variables λ, X, and d)

$$\begin{aligned} \max \quad & -\log\det[X] + \int_S (\langle Xs - d, Xs - d\rangle - 1) \, d\lambda(s) \\ \text{s.t.} \quad & -\nabla_X \log\det[X] + \int_S \nabla_X (\langle Xs - d, Xs - d\rangle - 1) \, d\lambda(s) = 0, \\ & \int_S \nabla_d (\langle Xs - d, Xs - d\rangle - 1) \, d\lambda(s) = 0, \\ & \lambda \in \Lambda(S)^+ \end{aligned} \tag{4.5}$$

is dual of problem (4.1) and has a solution. Furthermore, $(\bar{X}, \bar{d})$ solves problem (4.1) if and only if $(\bar{\lambda}, \bar{X}, \bar{d})$ solves problem (4.5) and the following is satisfied:

$$\|\bar{X}s - \bar{d}\| = 1 \;\; \bar{\lambda}\text{-a.e.}. \tag{4.6}$$

In the rest of the proof, we rewrite problem (4.5) in a simpler form by eliminating the variables X and d.

We have $\nabla_X \log\det[X] = X^{-1}$ and $\nabla_X(\langle Xs - d, Xs - d\rangle)$ is calculated in Lemma 7.1 in Appendix. Therefore, the equality constraints in problem (4.5) are equal to

$$\begin{aligned} X^{-1} &= \int_S (Xss^T + ss^TX - sd^T - ds^T) \, d\lambda(s), \\ 0 &= \int_S (Xs - d) \, d\lambda(s). \end{aligned} \tag{4.7}$$

Let us define

$$k := \lambda(S), \quad \xi := \lambda/k, \quad \tilde{s} := \int_S s \, d\xi(s).$$

We note that ξ is a probability measure on S. Using change of variables, with k and ξ, the second equation in (4.7) becomes $d = X\tilde{s}$. Substituting this expression for d into the first equation in (4.7), we arrive at

$$X^{-1} = k(XM + MX),$$

where

$$M := \int_S ss^T \, d\xi(s) - \tilde{s}\tilde{s}^T = \int_S (s - \tilde{s})(s - \tilde{s})^T \, d\xi(s).$$

As M is symmetric positive semi-definite, it follows from Lemma 7.2 in Appendix that

$$X^{-2} = 2k \int_S (s - \tilde{s})(s - \tilde{s})^T \, d\xi(s). \tag{4.8}$$

Next, we incorporate the expressions for X^{-2} and d into problem (4.5) and end up with

$$\begin{array}{ll} \max & \dfrac{n}{2}\log(2k) + \dfrac{1}{2}\log\det \displaystyle\int_S (s - \tilde{s})(s - \tilde{s})^T \, d\xi(s) + k \int_S \langle X(s - \tilde{s}), X(s - \tilde{s}) \rangle \, d\xi(s) - k \\ \text{s.t.} & \xi \in \Xi(S),\ k > 0, \end{array} \tag{4.9}$$

where $\Xi(S)$ is the set of probability measures on S. We claim that

$$\int_S \langle X(s - \tilde{s}), X(s - \tilde{s}) \rangle \, d\xi(s) = n/(2k).$$

To see this, postmultiply and premultiply (4.8) by X and take the trace of both sides of the resulting equality to find

$$\begin{aligned} n = 2k\,\mathrm{tr}\Big(\int_S X(s - \tilde{s})(s - \tilde{s})^T X \, d\xi(s)\Big) &= 2k \int_S \mathrm{tr}(X(s - \tilde{s})(s - \tilde{s})^T X) \, d\xi(s) \\ &= 2k \int_S \langle X(s - \tilde{s}), X(s - \tilde{s}) \rangle \, d\xi(s). \end{aligned}$$

Now, we observe that in problem (4.9), we can maximize with respect to k to find $\bar{k} = n/2$. Thus, problem (4.9) reduces to problem (4.2).

As we mentioned at the beginning of the proof, the rest of the theorem follows from Theorem 3.4 and from the equivalence of the condition (4.6) to $\|\bar{X}s - \bar{d}\| = 1$ $\bar{\xi}$-a.e. (since $\bar{\xi} = \bar{\lambda}/\bar{k}$). This completes the proof. □

Remark 4.2. Let $\mathrm{ext}(S)$ denote the set of extreme points of S. We have

$$\begin{aligned} \|\bar{X}s - \bar{d}\| = 1 \ \bar{\xi}\text{-a.e.} \quad &\Leftrightarrow \quad s \in \partial S \cap \partial\mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d}) \ \bar{\xi}\text{-a.e.} \\ &\Leftrightarrow \quad s \in \mathrm{ext}(S) \ \bar{\xi}\text{-a.e.}. \end{aligned}$$

The first equivalence above follows from the definition of the ellipsoid and the feasibility. For the second equivalence, we argue as follows: Let $s \in \partial S \cap \partial\mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d})$. Note that $\partial\mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d}) = \mathrm{ext}(\mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d}))$. If $s \notin \mathrm{ext}(S)$, then there exist $y, z \in S$, $y \neq z$, such that s lies in the interior of the line segment $[y, z]$. However, $s \in [y, z] \subseteq \mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d})$, contradicting the fact that $s \in \mathrm{ext}(\mathcal{E}(\bar{X}^{-1}, \bar{X}^{-1}\bar{d}))$.

Remark 4.3. The matrix appearing in the dual problem (4.2) is the Fisher information matrix (or the inverse of the covariance matrix) for the linear regression problem with $y = \Theta^T s + \theta_0 + \epsilon$ when one is interested in estimating only the parameter vector Θ (see, e.g., [26] or Section 3.10 in [19]).

We observe that finding $\tilde{s}$ in the dual problem (4.2) can be hard. For example, for the Lebesgue measure, $\tilde{s}$ corresponds to the center of gravity of S. On the other hand, it follows from Theorem 3.6 that problem (4.2) can be posed as

$$\max_{\lambda \in \mathbb{R}^k} \{\log\det[\sum_{i=1}^{k} \lambda_i (s_i - \tilde{s})(s_i - \tilde{s})^T] : \sum_{i=1}^{k} \lambda_i = 1,\ \lambda \geq 0\},$$

where $s_i \in S$, $\tilde{s} = \sum_{i=1}^{k} \lambda_i s_i$ and $k = n(n+1)/2$. For a general set S, this problem can still be difficult to solve. However, when S is given as the convex hull of a number of points, the problem becomes easier as there is no need to find $\{s_i\}$. The special algorithms for this problem are mostly designed for this type of S.

5 Dual of the maximum volume inscribed ellipsoid problem

We recall that the maximum volume inscribed ellipsoid problem is the problem of finding an ellipsoid of maximum volume inscribed in a convex body S in $\mathbb{R}^n$. We denote such an ellipsoid by IE(S). In this section, we are interested in constructing a dual of this problem using the duality results given in Section 3, especially Theorem 3.4.

We assume that S contains the origin in its interior. This assumption is not restrictive for our problem since the volume ratios are invariant under affine transformations. Then, we can represent S as the intersection of halfspaces

$$S = \{s : \langle a, s\rangle \leq 1,\ \forall\, a \in A\},$$

where

$$A := \{d/\delta(d\,|S) : \ d \in \partial B_n\}, \quad \delta(d\,|S) := \sup_{s \in S} \langle d, s\rangle. \tag{5.1}$$

Clearly, $\delta(.\,|S)$, the Minkowski support function of S, is defined everywhere on $\mathbb{R}^n$. It is positive since S contains the origin in its interior and continuous since it is convex being the maximum of linear functions indexed by s. Obviously, A is compact.

We remark that if S is given by $\{x : \langle a, x\rangle \leq b,\ (a, b) \in I\}$, where I is an index set, then we can set $A = \{a/b : \ (a, b) \in I\}$. We use the form $A = \{d/\delta(d\,|S) : \ d \in \partial B_n\}$ explicitly only in the proof of Corollary 5.2.

We use the representation $\mathcal{E}(T, c) = T(B_n) + c$ for an ellipsoid in $\mathbb{R}^n$, where T is an $n \times n$ symmetric positive definite matrix. Then,

$$\mathcal{E}(T, c) \subseteq S \quad \Leftrightarrow \quad \max_{u \in B_n} \langle a, Tu + c\rangle \leq 1,\ \forall\, a \in A,$$

that is, $\|Ta\| + \langle a, c\rangle \leq 1$ for each a in A. As a result, we can formulate the IE(S) problem as

$$
\begin{aligned}
\min \quad & -\log\det[T] \\
\text{s.t.} \quad & \|Ta\| + \langle a, c\rangle \le 1, \quad \forall\, a \in A,
\end{aligned}
\tag{5.2}
$$

where the decision variables are $(T, c) \in \mathbb{S}^n \times \mathbb{R}^n$ with T positive definite. This is clearly a convex semi–infinite programming problem.

The next theorem is another main result in this paper.

Theorem 5.1. *Let S be a convex body in $\mathbb{R}^n$ containing the origin in its interior and A be given as in* (5.1)*. A dual of problem* (5.2) *is given by*

$$
\begin{aligned}
\max \quad & \frac{1}{2}\log\det\left[\int_A aa^T/\|Ua\|\, d\xi(a)\right] + n\log n \\
\text{s.t.} \quad & U^{-2} = \int_A \frac{aa^T}{\|Ua\|}\, d\xi(a), \\
& 0 = \int_A a\, d\xi(a), \\
& \xi \in \Xi(A),
\end{aligned}
\tag{5.3}
$$

where $\Xi(A)$ is the set of probability measures on A. Problem (5.3) *has a solution. Furthermore, $(\bar{T}, \bar{c})$ solves problem* (5.2) *if and only if $(\bar{\xi}, \bar{U})$ solves problem* (5.3) *and the following holds:*

$$
\|\bar{T}a\| + \langle a, \bar{c}\rangle = 1 \;\; \bar{\xi}\text{-a.e.}, \quad \bar{U} = n\bar{T}. \tag{5.4}
$$

Proof. The proof uses Theorem 3.4. We claim that the assumptions of Theorem 3.4 are satisfied: It is obvious that problem (5.2) is a special case of problem (3.1). Also, $\|Ta\|+\langle a, c\rangle$ is differentiable in (T, c) on A and its derivative is continuous (jointly in (T, c) and a), see Lemma 7.1 in Appendix. Now, the existence of a solution to problem (5.2) follows from the continuity of the objective function and the compactness of the feasible set (see, e.g., [4, 8]). Lastly, since S is a convex body, there exists a ball of radius R contained in S. Hence, (3.7) is satisfied. This proves the claim.

It follows from Theorem 3.4, that the problem (with variables λ, T, and c)

$$
\begin{aligned}
\max \quad & -\log\det[T] + \int_A (\|Ta\| + \langle a, c\rangle - 1)\, d\lambda(a) \\
\text{s.t.} \quad & -T^{-1} + \int_A \frac{Taa^T + aa^T T}{2\|Ta\|}\, d\lambda(a) = 0, \\
& \int_A a\, d\lambda(a) = 0, \\
& \lambda \in \Lambda(A)^+
\end{aligned}
\tag{5.5}
$$

is dual of problem (5.2) and has a solution. Furthermore, $(\bar{T}, \bar{c})$ solves problem (5.2) if and only if $(\bar{\lambda}, \bar{T}, \bar{c})$ solves problem (5.5) and the following is satisfied:

$$
\|\bar{T}a\| + \langle a, \bar{c}\rangle = 1 \;\; \bar{\lambda}\text{-a.e.}. \tag{5.6}
$$

We note that in problem (5.5), the first and the second equality constraints follow from

$$
\begin{aligned}
0 &= -\nabla_T \log\det[T] + \int_A \nabla_T(\|Ta\| + \langle a, c\rangle - 1)\, d\lambda(a), \\
0 &= \int_A \nabla_c(\|Ta\| + \langle a, c\rangle - 1)\, d\lambda(a),
\end{aligned}
$$

respectively, using $\nabla_T \log\det[T] = T^{-1}$ and the derivative of $\|Ta\|$ as calculated in Lemma 7.1 in Appendix.

Let us define

$$k := \lambda(A), \quad \xi := \lambda/k, \quad U := kT$$

so that ξ is a probability measure on A. Then, problem (5.5) can equivalently be written as

$$\begin{aligned} \max \quad & n\log k - \log\det[U] + \int_A (\|Ua\| + k\langle a, c\rangle - k)\, d\xi(a) \\ \text{s.t.} \quad & -U^{-1} + \int_A \frac{Uaa^T + aa^TU}{2\|Ua\|}\, d\xi(a) = 0, \\ & \int_A a\, d\xi(a) = 0, \\ & \xi \in \Xi(A)^+,\ k > 0, \end{aligned} \tag{5.7}$$

where $\Xi(A)$ is the set of probability measures on A. In the rest of the proof, we rewrite problem (5.7) in a simpler form by trying to eliminate the variables U and c.

The first constraint in problem (5.7) implies

$$U^{-2} = \int_A \frac{aa^T}{\|Ua\|}\, d\xi(a) \tag{5.8}$$

by virtue of Lemma 7.2 in Appendix. Next, we incorporate the expression for U^{-2} and $\int_A a\, d\xi(a) = 0$ into problem (5.7) and end up with

$$\begin{aligned} \max \quad & n\log k + \frac{1}{2}\log\det[\int_A \frac{aa^T}{\|Ua\|}\, d\xi(a)] + \int_A \|Ua\|\, d\xi(a) - k \\ \text{s.t.} \quad & -U^{-2} + \int_A \frac{aa^T}{\|Ua\|}\, d\xi(a) = 0, \\ & \int_A a\, d\xi(a) = 0, \\ & \xi \in \Xi(A),\ k > 0. \end{aligned} \tag{5.9}$$

We claim that $\int_A \|Ua\|\, d\xi(a) = n$. To see this, postmultiply and premultiply (5.8) by U and take the trace of both sides of the resulting equality to find

$$n = \mathrm{tr}(\int_A \frac{Uaa^TU}{\|Ua\|}\, d\xi(a)) = \int_A \mathrm{tr}(\frac{Uaa^TU}{\|Ua\|})\, d\xi(a) = \int_A \|Ua\|\, d\xi(a). \tag{5.10}$$

Now, we observe that in problem (5.9), we can maximize with respect to k to find $\bar{k} = n$. Thus problem (5.9) reduces to problem (5.3).

As we mentioned at the beginning of the proof, the rest of the theorem follows from Theorem 3.4. This completes the proof. □

In the next corollary, we write the necessary and sufficient optimality conditions for problem (5.2) in another form. The proof follows from Theorem 5.1 using change of variables and change of measures. Since the proof is technical, we present it in Appendix 7.

Corollary 5.2. *Let S be a convex body in $\mathbb{R}^n$ containing the origin in its interior and A be given as in* (5.1). *Then $(\bar{T}, \bar{c})$ solves problem* (5.2) *if and only if for some probability measure μ on S the following holds:*

$$\bar{T}^2 = n \int_S (s - \bar{c})(s - \bar{c})^T \, d\mu(s), \quad \bar{c} = \int_S s \, d\mu(s),$$
$$s \in \partial S \cap \partial \mathcal{E}(\bar{T}, \bar{c}) \ \mu\text{-a.e.}, \quad \mathcal{E}(\bar{T}, \bar{c}) \subseteq S.$$

6 Special Cases

In this section, we consider some special cases of problems (4.1) and (5.2). Mainly, we use Theorems 4.1 and 5.1 directly or follow their proofs.

6.1 CE(S) problem when S is symmetrization of a "lifted" convex body

Consider the minimum volume circumscribed ellipsoid problem (4.1) with S given by

$$S = \text{conv}(\hat{V}, -\hat{V}), \text{ where } \hat{V} = \{(v, 1)^T \in \mathbb{R}^n : v \in V\}. \tag{6.1}$$

Here, we assume that V is a convex body in R^{n-1}. We call $\hat{V}$ a *lifted convex body* since it is obtained by lifting a convex body from R^{n-1} to R^n. The symmetrization of $\hat{V}$, that is, $\text{conv}(\hat{V}, -\hat{V})$ is called the *Elfving set* of $\hat{V}$ in the optimal design literature (see, e.g., Section 2.9 in [19]).

The following theorem is an important result in this paper.

Theorem 6.1. *Let S be given as in* (6.1). *Finding* CE(S) *is equivalent to finding* CE(V) *in the sense that*

$$\text{CE}(S) = \mathcal{E}(\bar{X}^{-1}, 0) \quad \Leftrightarrow \quad \text{CE}(V) = \mathcal{E}((\frac{n}{n-1}\bar{Y})^{-1/2}, \bar{c}), \tag{6.2}$$

where $\bar{X}^2 = \begin{bmatrix} \bar{Y} & -\bar{Y}\bar{c} \\ -\bar{c}^T\bar{Y} & (1/n + \bar{c}^T\bar{Y}\bar{c}) \end{bmatrix}$. *Furthermore,*

$$\text{vol}(CE(S)) = \det[\bar{Y}]^{-1/2}\sqrt{n}\,\omega_n,$$
$$\text{vol}(CE(V)) = \det[\bar{Y}]^{-1/2}(\frac{n}{n-1})^{-(n-1)/2}\,\omega_{n-1}.$$

Proof. Consider finding the ellipsoid CE(S). As S is symmetric about the origin, the optimal center of CE(S) is the origin, see, for example, [8]. Now, it follows from Theorem 3.4 (similar to the proof of Theorem 4.1) that $\bar{X}$ solves the CE(S) problem, that is, $\text{CE}(S) = \mathcal{E}(\bar{X}^{-1}, 0)$ if and only if $\bar{\xi}$ solves the dual problem

$$\max_{\xi \in \Xi(S)} \log\det[\int_S ss^T \, d\xi(s)\,] \tag{6.3}$$

and the following is satisfied:

$$s \in \text{ext}(S) \ \ \bar{\xi}\text{-a.e.}. \tag{6.4}$$

Furthermore, the two solutions are related by

$$I = n \int_S \bar{X}^2 s s^T \, d\bar{\xi}(s). \tag{6.5}$$

For the condition $s \in \text{ext}(S)$ $\bar{\xi}$-*a.e.*, see Remark 4.2.

We claim that the dual problem (6.3) is equivalent to

$$\max_{\xi \in \Xi(V)} \log \det \left[\int_V (v - \tilde{v})(v - \tilde{v})^T \, d\xi(v) \right], \tag{6.6}$$

where $\tilde{v} := \int_V v \, d\xi(v)$ and the conditions (6.5) and (6.4) to

$$I = n \int_V \bar{X}^2 \begin{bmatrix} vv^T & v \\ v^T & 1 \end{bmatrix} d\bar{\xi}(v), \tag{6.7}$$

$$v \in \text{ext}(V) \;\; \bar{\xi}\text{-a.e.}. \tag{6.8}$$

To see this, we observe that

$$\text{ext}(S) = \{(\delta v, \delta)^T : v \in \text{ext}(V), \, \delta \in \{-1, 1\}\}.$$

Therefore, we have for each $s \in \text{ext}(S)$

$$ss^T = \begin{bmatrix} \delta^2 vv^T & \delta^2 v \\ \delta^2 v^T & \delta^2 \end{bmatrix} = \begin{bmatrix} vv^T & v \\ v^T & 1 \end{bmatrix}.$$

In fact, without loss of generality, we can replace a probability $\xi \in \Xi(S)$ by a probability $\xi \in \Xi(V)$. Consider the simple case of discrete probabilities. If ξ assigns probabilities p_i and q_i to $(v_i, 1)^T$ and $(-v_i, -1)^T$, respectively, then we can assign the probability $p_i + q_i$ to v_i and the conditions remain the same. Similarly, if v_i is assigned a probability of p_i, then we can assign equal probabilities of $p_i/2$ to $(v_i, 1)^T$ and $(-v_i, -1)^T$ and the conditions remain the same. Therefore, by replacing the probability $\xi \in \Xi(S)$ by $\xi \in \Xi(V)$, we arrive at the conditions (6.7) and (6.8).

Lastly, we rewrite the dual problem (6.3) observing that

$$\begin{aligned} \det[\int_V \begin{bmatrix} vv^T & v \\ v^T & 1 \end{bmatrix} d\xi(v)\,] &= \det \begin{bmatrix} \int_V vv^T \, d\xi(v) & \int_V v \, d\xi(v) \\ \int_V v^T \, d\xi(v) & 1 \end{bmatrix} \\ &= \det \left[\int_V vv^T \, d\xi(v) - \tilde{v}\tilde{v}^T \right] \\ &= \det \left[\int_V (v - \tilde{v})(v - \tilde{v})^T \, d\xi(v) \, \right]. \end{aligned}$$

Here, the second equality follows from the Schur complement and the last one from the definition $\tilde{v} := \int_V v \, d\xi(v)$. This proves the claim. We remark that problem (6.6) is the dual of the CE(V) problem.

Now, without loss of any generality, we use the representation $\bar{X}^2 = \begin{bmatrix} \bar{Y} & -\bar{Y}\bar{c} \\ -\bar{c}^T\bar{Y} & (\bar{b} + \bar{c}^T\bar{Y}\bar{c}) \end{bmatrix}$. Then (6.7) becomes

$$I = n \int_V \bar{Y}(v - \bar{c})v^T \, d\bar{\xi}(v), \qquad 0 = n \int_V \bar{Y}(v - \bar{c}) \, d\bar{\xi}(v), \tag{6.9}$$

$$0 = n \int_V (\bar{b} + \bar{c}^T\bar{Y}(c - v))v^T \, d\bar{\xi}(v), \qquad 1 = n \int_V (\bar{b} + \bar{c}^T\bar{Y}(\bar{c} - v)) \, d\bar{\xi}(v). \tag{6.10}$$

Next, we claim that the system (6.9)-(6.10) is equivalent to

$$I = n \int_V \bar{Y}(v - \tilde{v})(v - \tilde{v})^T \, d\bar{\xi}(v), \qquad \bar{c} = \tilde{v}, \qquad \bar{b} = 1/n, \tag{6.11}$$

where $\tilde{v} = \int_V v \, d\bar{\xi}(v)$: first, it follows from the second equation in (6.9) that $\bar{c} = \tilde{v}$. Then, it is easy to show that the first equation in (6.9) is equivalent to the first equation in (6.11) and that the second equation in (6.9) holds automatically. Also, it follows from the second equations in (6.9) and (6.10) that $\bar{b} = 1/n$. Therefore, using the Schur complement of $\bar{X}^2$ in $\det[\bar{X}^2] = \bar{b} \det[\bar{Y}]$, we have

$$\mathrm{vol}(CE(S)) = \det[\bar{X}]^{-1} \omega_n = \det[\bar{Y}]^{-1/2} \sqrt{n} \, \omega_n.$$

Now, as we mentioned before, problem (6.6) is the dual of the CE(V) problem. Let $(\bar{Z}, \bar{d})$ be the solution to the CE(V) problem, that is, CE(V) $= \mathcal{E}(\bar{Z}^{-1}, \bar{Z}^{-1}d)$. Then, it follows from Theorem 4.1 that $(\bar{Z}, \bar{d})$ solves the CE(V) problem if and only if $v \in \mathrm{ext}(V)$ $\bar{\xi}$-*a.e.* and

$$I = (n-1) \int_V \bar{Z}^2 (v - \tilde{v})(v - \tilde{v})^T \, d\bar{\xi}(v), \quad \bar{d} = \bar{Z}\tilde{v}.$$

Comparing these equalities to the first two equalities in (6.11), we conclude that

$$\bar{Z}^{-2} = (n-1)\bar{Y}^{-1}/n, \quad \bar{d} = \bar{Z}\bar{c}.$$

This establishes the equivalence relationship in the theorem and also yields

$$\mathrm{vol}(CE(V)) = \det[\bar{Y}]^{-1/2} (\frac{n}{n-1})^{-(n-1)/2} \omega_{n-1}.$$

□

6.2 Minimum volume circumscribed ball problem

Consider the minimum volume circumscribed ellipsoid problem (4.1) when the ellipsoids are restricted to balls. This problem (and its dual) is studied in [7].

In problem (4.1), let $X = rI$. Then, the minimum volume circumscribed ball problem is formulated as

$$\begin{array}{ll} \min & -n \log r \\ \mathrm{s.\,t.} & \langle rs - d, rs - d \rangle \leq 1, \quad \forall \, s \in S. \end{array} \tag{6.12}$$

In this problem, the decision variables are $(r, d) \in \mathbb{R} \times \mathbb{R}^n$ with r positive. The dual of this problem follows easily, similar to the proof of Theorem 4.1, and is given in the following corollary.

Corollary 6.2. *Let S be a convex body in $\mathbb{R}^n$. A dual of problem* (6.12) *is given by*

$$\max_{\xi \in \Xi(S)} \left[\int_S \langle s - \tilde{s}, s - \tilde{s} \rangle \, d\xi(s) \right], \tag{6.13}$$

where $\tilde{s} = \int_S s \, d\xi(s)$ and $\Xi(S)$ is the set of probability measures on S. Problem (6.13) *has a solution. Furthermore, $(\bar{r}, \bar{d})$ solves problem* (6.12) *if and only if $\bar{\xi}$ solves problem* (6.13) *and the following is satisfied:*

$$\|\bar{r}s - \bar{d}\| = 1 \;\; \bar{\xi}\text{-}a.e..$$

The two solutions are related by

$$\bar{r}^{-2} = \int_S \langle s - \tilde{s}, s - \tilde{s}\rangle \, d\bar{\xi}(s), \qquad \bar{d} = \bar{r}\tilde{s}.$$

Problem (6.13) can be interpreted as finding a probability measure on the given set S with the maximum variance. This dual problem and its interpretation was suggested to the author by O. Güler.

6.3 Maximum volume inscribed ball problem

Consider the maximum volume inscribed ellipsoid problem (5.2) when the ellipsoids are restricted to balls.

In problem (5.2), let $T = rI$. Then, the maximum volume inscribed ball problem is formulated as

$$\begin{aligned} \min \quad & -n \log r \\ \text{s.t.} \quad & r\|a\| + \langle a, c\rangle \leq 1, \quad \forall\, a \in A. \end{aligned} \tag{6.14}$$

In this problem, the decision variables are $(r, c) \in \mathbb{R} \times \mathbb{R}^n$ with r positive. Again, the dual of this problem follows easily, similar to the proof of Theorem 5.1, and is given in the following corollary.

Corollary 6.3. *Let S be a convex body in $\mathbb{R}^n$ containing the origin in its interior and A be given as in* (5.1)*. A dual of problem* (6.14) *is given by*

$$\begin{aligned} \max \quad & \int_A \|a\| \, d\xi(a) \\ \text{s.t.} \quad & 0 = \int_A a \, d\xi(a), \\ & \xi \in \Xi(A), \end{aligned}$$

where $\Xi(A)$ is the set of probability measures on A. Problem (6.3) *has a solution. Furthermore, $(\bar{r}, \bar{c})$ solves problem* (6.14) *if and only if $\bar{\xi}$ solves problem* (6.3) *and the following is satisfied:*

$$\bar{r}\|a\| + \langle a, \bar{c}\rangle = 1 \;\; \bar{\xi}\text{-}a.e..$$

The two solutions are related by $\bar{r}^{-1} = \int_A \|a\| \, d\bar{\xi}(a)$.

Acknowledgements

The author would like to thank Professor Osman Güler for helpful discussions on the subject. This research is partially supported by the National Science Foundation under grant DMS-0411955 (principle investigator: Osman Güler).

7 Appendix

In this appendix, we collect for completeness some results used in the main body of the paper.

First, we compute the gradients of $f(A) = \langle Ax - b, Ax - b\rangle$ and $g(A) = \|Ax\|$.

Lemma 7.1. *The gradients of $f(A) = \langle Ax - b, Ax - b\rangle$ and $g(A) = \|Ax\|$ at a symmetric positive definite matrix A are given by*

$$\nabla f(A) = (Ax - b)x^T + x(Ax - b)^T, \quad \nabla g(A) = \frac{Axx^T + xx^T A}{2\|Ax\|}.$$

Proof. We expand the Taylor series of the function $f(A)$ in a given direction $D \in \mathbb{S}^{n\times n}$.

$$\begin{aligned}\Delta f :&= f(A + tD) - f(A) = \langle (A + tD)x - b, (A + tD)x - b\rangle - \langle Ax - b, Ax - b\rangle \\ &= t\langle Dx, Ax - b\rangle + t\langle Ax - b, Dx\rangle + t^2\langle Dx, Dx\rangle \\ &= t\,\mathrm{tr}(x^T D(Ax - b) + (Ax - b)^T Dx) + t^2\,\mathrm{tr}(x^T DDx) \\ &= t\,\mathrm{tr}((Ax - b)x^T D + x(Ax - b)^T D) + \frac{t^2}{2}\,\mathrm{tr}(2xx^T DD) \\ &= t\langle (Ax - b)x^T + x(Ax - b)^T, D\rangle + \frac{t^2}{2}\langle 2xx^T D, D\rangle.\end{aligned}$$

Now for $b = 0$ we obtain $\nabla\|Ax\|^2 = Axx^T + xx^T A$, from which we obtain the formula for $\nabla g(A)$ using the chain rule. □

The following lemma is given in [27] and since its proof is simple, we include it here.

Lemma 7.2. *Let the symmetric matrices C and X be positive semi-definite and positive definite, respectively. Then $X^{-1} = (XC + CX)/2$ holds if and only if $X = C^{-1/2}$.*

Proof. Obviously if $X = C^{-1/2}$ then $X^{-1} = (XC + CX)/2$ holds. Now if $X^{-1} = (XC + CX)/2$ holds, then first C must be positive definite. Substitute the eigenvalue decomposition $X = U\Sigma U^T$ of X into $X^{-1} = (XC + CX)/2$ to obtain $\Sigma^{-1} = (D\Sigma + \Sigma D)/2$, where $D = U^T CU$. This implies

$$\frac{1}{2}D_{ij}(\Sigma_{ii} + \Sigma_{jj}) = \begin{cases} 0 & \text{if } i \neq j \\ 1/\Sigma_{ii} & \text{if } i = j \end{cases},$$

from which it follows that $\Sigma^{-2} = D$ and hence $C = UDU^T = X^{-2}$. □

Lastly, we give the proof of Corollary 5.2.

Proof. (***of Corollary 5.2***) It follows from Theorem 5.1 (see also Corollary 3.5) that $(\bar{T}, \bar{c})$ solves problem (5.2) if and only if the following holds for some probability measure ξ on A:

$$\begin{aligned}\bar{T}^{-2} &= n\int_A \frac{aa^T}{\|\bar{T}a\|}\,d\xi(a), \qquad 0 = \int_A a\,d\xi(a), \\ 1 &= \|\bar{T}a\| + \langle a, \bar{c}\rangle \;\; \xi\text{-a.e.}, \quad 1 \geq \|\bar{T}a\| + \langle a, \bar{c}\rangle, \quad \forall\, a \in A.\end{aligned}$$

Let us define a measure ν on the Borel sets B of A by

$$\nu(B) := \int_B \|\bar{T}a\| \, d\xi(a).$$

We note that ν is a probability measure on A since $\nu(A) = \int_A \|\bar{T}a\| \, d\xi(a) = 1$ by virtue of Theorem 5.1 and (5.10). Then, using change of measures, we write the above conditions as

$$\bar{T}^{-2} = n \int_A \frac{aa^T}{\|\bar{T}a\|^2} \, d\nu(a), \qquad 0 = \int_A \frac{a}{\|\bar{T}a\|} \, d\nu(a), \tag{7.1}$$

$$1 = \|\bar{T}a\| + \langle a, \bar{c} \rangle \;\; \nu\text{-a.e.}, \quad 1 \geq \|\bar{T}a\| + \langle a, \bar{c} \rangle, \quad \forall \, a \in A. \tag{7.2}$$

Now, we define $\phi : \partial\mathcal{E}(\bar{T}, \bar{c}) \to A$ and correspondingly $\phi^{-1} : A \to \partial\mathcal{E}(\bar{T}, \bar{c})$ by

$$\phi(s) := \frac{\bar{T}^{-2}(s - \bar{c})}{\delta(\bar{T}^{-2}(s - \bar{c}) \,|S)}, \quad \phi^{-1}(a) := \frac{\bar{T}^2 a}{\|\bar{T}a\|} + \bar{c}.$$

We have $\phi(s) \in A$ since $\langle \phi(s), x \rangle \leq 1$ for each x in S, by definition of $\delta(. \,|S)$. It is obvious that $\phi^{-1}(a) \in \partial\mathcal{E}(\bar{T}, \bar{c})$. Also it is easy to verify, by direct substitution, that $\phi(\phi^{-1}(a)) = a$ and $\phi^{-1}(\phi(s)) = s$.

Using the substitution $a = \phi(s)$, the first equality in (7.1) yields

$$\begin{aligned} \bar{T}^{-2} = n \int_A \frac{aa^T}{\|\bar{T}a\|^2} \, d\nu(a) &= n \int_{\partial\mathcal{E}(\bar{T},\bar{c})} \frac{\phi(s)\phi(s)^T}{\|\bar{T}\phi(s)\|^2} \, d\mu(s) \\ &= n \int_{\partial\mathcal{E}(\bar{T},\bar{c})} \frac{\bar{T}^{-2}(s - \bar{c})(s - \bar{c})^T \bar{T}^{-2}}{\|\bar{T}^{-1}(s - \bar{c})\|^2} \, d\mu(s), \end{aligned}$$

where μ is the probability measure on $\partial\mathcal{E}(T, c)$ defined by $\mu(\phi^{-1}(B)) = \nu(B)$ for the Borel sets B of A. Since $s \in \partial\mathcal{E}(\bar{T}, \bar{c}))$, we have $\|\bar{T}^{-1}(s - \bar{c})\| = 1$ and the equation above implies that

$$\bar{T}^2 = n \int_{\partial\mathcal{E}(\bar{T},\bar{c})} (s - \bar{c})(s - \bar{c})^T \, d\mu(s).$$

In a similar fashion, it follows from the second equation in (7.1) that

$$0 = \int_A \frac{a}{\|\bar{T}a\|} \, d\nu(a) = \int_{\partial\mathcal{E}(\bar{T},\bar{c})} \frac{\phi(s)}{\|\bar{T}\phi(s)\|} \, d\mu(s) = \int_{\partial\mathcal{E}(\bar{T},\bar{c})} \bar{T}^{-2}(s - \bar{c}) \, d\mu(s).$$

Here, the last equality follows from $\|\bar{T}^{-1}(s - \bar{c})\| = 1$. Hence, we obtain

$$\bar{c} = \int_{\partial\mathcal{E}(\bar{T},\bar{c})} s \, d\mu(s).$$

Now, we claim that the first condition in (7.2) is equivalent to

$$s \in \partial S \cap \partial\mathcal{E}(\bar{T}, \bar{c}) \;\; \mu\text{-a.e..}$$

To see this we consider

$$\begin{aligned} 0 &= \int_A (\|\bar{T}a\| + \langle a, \bar{c} \rangle - 1) \, d\nu(a) = \int_{\partial\mathcal{E}(\bar{T},\bar{c})} (\|\bar{T}\phi(s)\| + \langle \phi(s), \bar{c} \rangle - 1) \, d\mu(s) \\ &= \int_{\partial\mathcal{E}(\bar{T},\bar{c})} (\frac{1 + \langle \bar{T}^{-2}(s - \bar{c}), \bar{c} \rangle}{\delta(\bar{T}^{-2}(s - \bar{c}) \,|S)} - 1) \, d\mu(s) = \int_{\partial\mathcal{E}(\bar{T},\bar{c})} (\frac{\langle s, \bar{T}^{-2}(s - \bar{c}) \rangle}{\delta(\bar{T}^{-2}(s - \bar{c})| \, S)} - 1) \, d\mu(s), \end{aligned}$$

where the first equality follows from the first condition in (7.2) and the feasibility, the second and the third equations from change of measure and substitution, and the last one from the fact that $\|\bar{T}^{-1}(s - \bar{c})\|^2 = 1$. We have

$$\langle s, T^{-2}(s-c)\rangle \leq \delta(T^{-2}(s-c)\,|S)$$

by definition of δ since s is contained in S by virtue of feasibility. This proves the claim.

Lastly, it follows from the condition $s \in \partial S \cap \partial \mathcal{E}(\bar{T}, \bar{c})$ μ-a.e. that all the integrals (with measure μ) that we obtain above are equivalent to the ones given in the Corollary. This completes the proof. □

References

[1] D. Ahipasaoglu, P. Sun, and M. Todd, *Linear convergence of a modified Frank-Wolfe algorithm for computing minimum volume enclosing ellipsoids*, *Optim. Methods Softw.*, (2007). accepted for publication.

[2] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, *Springer Series in Operations Research*, Springer-Verlag, New York, 2000.

[3] J. M. Borwein, *Direct theorems in semi-infinite convex programming*, Math. Programming, 21 (1981), pp. 301–318.

[4] L. Danzer, D. Laugwitz, and H. Lenz, *Über das Löwnersche Ellipsoid und sein Analogon unter den einem Eikörper einbeschriebenen Ellipsoiden*, Arch. Math., 8 (1957), pp. 214–219.

[5] A. N. Dolia, T. De Bie, C. J. Harris, J. Shawe-Taylor, and D. M. Titterington, *The minimum volume covering ellipsoid estimation in kernel-defined feature spaces*, Proceedings of the 17th European Conference on Machine Learning (ECML 2006), Berlin, (September 2006).

[6] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland Publishing Co., Amsterdam, 1976. Translated from the French, Studies in Mathematics and its Applications, Vol. 1.

[7] D. J. Elzinga and D. W. Hearn, *The minimum covering sphere problem*, Management Sci., 19 (1972), pp. 96–104.

[8] O. Güler and F. Gürtuna, *The extremal volume ellipsoids of convex bodies, their symmetry properties, and their determination in some special cases.* Technical Report TR2007-7, Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA, September/2007.

[9] A. D. Ioffe and V. L. Levin, *Subdifferentials of convex functions*, Trudy Moskov. Mat. Obšč., 26 (1972), pp. 3–73.

[10] F. John, *Extremum problems with inequalities as subsidiary conditions*, in Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, Interscience Publishers, Inc., New York, N. Y., 1948, pp. 187–204.

[11] F. Juhnke, *Embedded maximal ellipsoids and semi-infinite optimization*, Beiträge Algebra Geom., 35 (1994), pp. 163–171.

[12] L. G. Khachiyan and M. J. Todd, *On the complexity of approximating the maximal inscribed ellipsoid for a polytope*, Math. Programming, 61 (1993), pp. 137–159.

[13] J. Kiefer and J. Wolfowitz, *Optimum designs in regression problems*, Ann. Math. Statist., 30 (1959), pp. 271–294.

[14] P. Kumar and E. A. Yildirim, *Minimum-volume enclosing ellipsoids and core sets*, J. Optim. Theory Appl., 126 (2005), pp. 1–21.

[15] P. J. Laycock, *Convex loss applied to design in regression problems*, J. Roy. Statist. Soc. Ser. B, 34 (1972), pp. 148–170, 170–186. With discussion by M. J. Box, P. Whittle, S. D. Silvey, A. A. Greenfield, Agnes M. Herzberg, J. Keifer, D. R. Cox, Lynda V. White, A. C. Atkinson, R. J. Brooks, Corwin L. Atwood and Robin Sibson, and replies by Henry P. Wynn and P. J. Laycock.

[16] V. L. Levin, *The application of E. Helly's theorem in convex programming, problems of best approximation, and related questions*, Mat. Sb. (N.S.), 79 (121) (1969), pp. 250–263.

[17] J. Nie and J. W. Demmel, *Minimum ellipsoid bounds for solutions of polynomial systems via sum of squares*, J. Global Optim., 33 (2005), pp. 511–525.

[18] K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.

[19] F. Pukelsheim, *Optimal design of experiments*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1993. A Wiley-Interscience Publication.

[20] R. T. Rockafellar, *Convex analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[21] W. W. Rogosinski, *Moments of non-negative mass*, Proc. Roy. Soc. London Ser. A, 245 (1958), pp. 1–27.

[22] W. Rudin, *Real and complex analysis*, McGraw-Hill Book Co., New York, third ed., 1987.

[23] A. Shapiro, *On duality theory of convex semi-infinite programming*, Optimization, 54 (2005), pp. 535–543.

[24] S. D. Silvey and D. M. Titterington, *A geometric approach to optimal design theory*, Biometrika, 60 (1973), pp. 21–32.

[25] P. Sun and R. M. Freund, *Computation of minimum-volume covering ellipsoids*, Oper. Res., 52 (2004), pp. 690–706.

[26] D. M. Titterington, *Optimal design: some geometrical aspects of D-optimality*, Biometrika, 62 (1975), pp. 313–320.

[27] Y. Zhang and L. Gao, *On numerical solution of the maximum volume ellipsoid problem*, SIAM J. Optim., 14 (2003), pp. 53–76 (electronic).