A GENERALIZATION OF THE LÖWNER-JOHN'S ELLIPSOID THEOREM

JEAN B. LASSERRE

ABSTRACT. We address the following generalization $\mathbf{P}$ of the Löwner-John ellipsoid problem. Given a (non necessarily convex) compact set $\mathbf{K} \subset \mathbb{R}^n$ and an even integer $d \in \mathbb{N}$, find an homogeneous polynomial g of degree d such that $\mathbf{K} \subset \mathbf{G} := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ and $\mathbf{G}$ has minimum volume among all such sets. We show that $\mathbf{P}$ is a convex optimization problem even if neither $\mathbf{K}$ nor $\mathbf{G}$ are convex! We next show that $\mathbf{P}$ has a unique optimal solution and a characterization with at most $\binom{n+d-1}{d}$ contacts points in $\mathbf{K} \cap \mathbf{G}$ is also provided. This is the analogue for $d > 2$ of the Löwner-John's theorem in the quadratic case $d = 2$, but importantly, we neither require the set $\mathbf{K}$ nor the sublevel set $\mathbf{G}$ to be convex. More generally, there is also an homogeneous polynomial g of even degree d and a point $\mathbf{a} \in \mathbb{R}^n$ such that $\mathbf{K} \subset \mathbf{G_a} := \{\mathbf{x} : g(\mathbf{x} - \mathbf{a}) \leq 1\}$ and $\mathbf{G_a}$ has minimum volume among all such sets (but uniqueness is not guaranteed). Finally, we also outline a numerical scheme to approximate as closely as desired the optimal value and an optimal solution. It consists of solving a hierarchy of convex optimization problems with strictly convex objective function and Linear Matrix Inequality (LMI) constraints.

1. INTRODUCTION

"Approximating" data by relatively simple geometrical objects is a fundamental problem with many important applications and the ellipsoid of minimum volume is the most well-known of the associated computational techniques.

Indeed, in addition to its nice properties from the viewpoint of applications, the ellipsoid of minimum volume is also very interesting from a mathematical viewpoint. Indeed, if $\mathbf{K} \subset \mathbb{R}^n$ is a convex body, computing an ellipsoid of minimum volume that contains $\mathbf{K}$ is a classical and infamous problem which has a unique optimal solution called the *Löwner-John's* ellipsoid. In addition, John's theorem states that the optimal ellipsoid Ω is characterized by s contacts points $u_i \in \mathbf{K} \cap \Omega$, and positive scalars λ_i, $i = 1, \ldots, s$, where s is bounded above by $n(n+3)/2$ in the general case and $s \leq n(n+1)/2$ when $\mathbf{K}$ is symmetric; see e.g. Ball [3, 4]. In particular, and in contrast to other approximation techniques, computing the ellipsoid of

1991 *Mathematics Subject Classification.* 26B15 65K10 90C22 90C25.

Key words and phrases. Homogeneous polynomials; sublevel sets; volume; Löwner-John problem; convex optimization.

minimum volume is a *convex* optimization problem for which efficient techniques are available; see e.g. Calafiore [7] and Sun and Freund [32] for more details. For a nice recent historical survey on the Löwner-John's ellipsoid, the interested reader is referred to the recent paper by Henk [13] and the many references therein.

As underlined in Calafiore [7], *"The problem of approximating observed data with simple geometric primitives is, since the time of Gauss, of fundamental importance in many scientific endeavors"*. For practical purposes and numerical efficiency the most commonly used are polyhedra and ellipsoids and such techniques are ubiquitous in several different area, control, statistics, computer graphics, computer vision, to mention a few. For instance:

• In *robust linear control*, one is interested in outer or inner approximations of the stability region associated with a linear dynamical system, that is, the set of initial states from which the system can be stabilized by some control policy. Typically, the stability region which can be formulated as a semi-algebraic set in the space of coefficients of the characteristic polynomial, is non convex. By using the Hermite stability criterion, it can be described by a parametrized polynomial matrix inequality where the parameters account for uncertainties and the variables are the controller coefficients. Convex inner approximations of the stability region have been proposed in form of polytopes in Nurges [23], ellipsoids in Henrion et al. [15], and more general convex sets defined by Linear Matrix Inequalities (LMIs) in Henrion et al. [17], and Karimi et al. [18].

• In *statistics* one is interested in the ellipsoid ξ of minimum volume covering some given k of m data points because ξ has some interesting statistical properties such as affine equivariance and positive breakdown properties [9]. In this context the center of the ellipsoid is called the minimum volume ellipsoid (MVE) *location* estimator and the associated matrix associated with ξ is called the MVE *scatter* estimator; see e.g. Rousseeuw [31] and Croux et al. [9].

• In *pattern separation*, minimum volume ellipsoids are used for separating two sets of data points. For computing such ellipsoids, convex programming techniques have been used in the early work of Rosen [28] and more modern semidefinite programming techniques in Vandenberghe and Boyd [34]. Similarly, in robust statistics and data mining the ellipsoid of minimum volume covering a finite set of data points identifies *outliers* as the points on its boundary; see e.g. Rousseeuw and Leroy [31]. Moreover, this ellipsoid technique is also scale invariant, a highly desirable property in data mining which is not enjoyed by other clustering methods based on various *distances*; see the discussion in Calafiore [7], Sun and Freund [32] and references therein.

• Other clustering techniques in *computer graphics*, *computer vision* and *pattern recognition*, use various (geometric or algebraic) distances (e.g. the equation error) and compute the best ellipsoid by minimizing an associated non linear least squares criterion (whence the name "least squares fitting ellipsoid" methods). For instance, such techniques have been proposed in computer graphics and computer vision by Bookstein [6] and Pratt [25], in pattern recognition by Rosin [29], Rosin and West [30], Taubin [33], and in another context by Chernousko [8]. When using an algebraic distance (like e.g. the equation error) the geometric interpretation is not clear and the resulting ellipsoid may not be satisfactory; see e.g. an illuminating discussion in Gander et al. [12]. Moreover, in general the resulting optimization problem is not convex and convergence to a global minimizer is not guaranteed.

So optimal data fitting using an ellipsoid of minimum volume is not only satisfactory from the viewpoint of applications but is also satisfactory from a mathematical viewpoint as it reduces to a (often tractable) convex optimization problem with a unique solution having a nice characterization in term of contact points in $\mathbf{K} \cap \Omega$. In fact, reduction to solving a convex optimization problem with a unique optimal solution, is a highly desirable property of any data fitting technique!

A more general optimal data fitting problem. In the Löwner-John problem one restricts to convex bodies $\mathbf{K}$ because for a non convex set $\mathbf{K}$ the optimal ellipsoid is also solution to the problem where $\mathbf{K}$ is replaced with its convex hull $\mathrm{co}(\mathbf{K})$. However, if one considers sets that are more general than ellipsoids, an optimal solution for $\mathbf{K}$ is not necessarily the same as for $\mathrm{co}(\mathbf{K})$, and indeed, in some applications one is interested in approximating as closely as desired a non convex set $\mathbf{K}$. In this case a non convex approximation is sometimes highly desirable as more efficient.

For instance, in the robust control problem already alluded to above, in Henrion and Lasserre [16] we have provided an inner approximation of the stability region $\mathbf{K}$ by the sublevel set $\mathbf{G} = \{\mathbf{x} : g(\mathbf{x}) \leq 0\}$ of a non convex polynomial g. By allowing the degree of g to increase one obtains the convergence $\mathrm{vol}(\mathbf{G}) \to \mathrm{vol}(\mathbf{K})$ which is impossible with the convex polytopes, ellipsoids and LMI approximations described in [23, 15, 17, 18].

So if one considers the more general data fitting problem where $\mathbf{K}$ and/or the (outer) approximating set are allowed to be non convex, can we still infer interesting conclusions as for the Löwner-John problem? Can we also derive a practical algorithm for computing an optimal solution? If the answer is negative, such a generalization would then be useless or sterile (or at least questionable).

The purpose of this paper is to provide results in this direction that can be seen as a non convex generalization of the Lowner-John's problem but, surprisingly, still reduces to solving a convex optimization problem with a unique solution.

Some works have considered generalizations of the Löwner-John's problem. Fo instance, Giannopoulos et al. [10] have extended John's theorem for couples $(\mathbf{K}_1, \mathbf{K}_2)$ of convex bodies when $\mathbf{K}_1$ is *in maximal volume position*

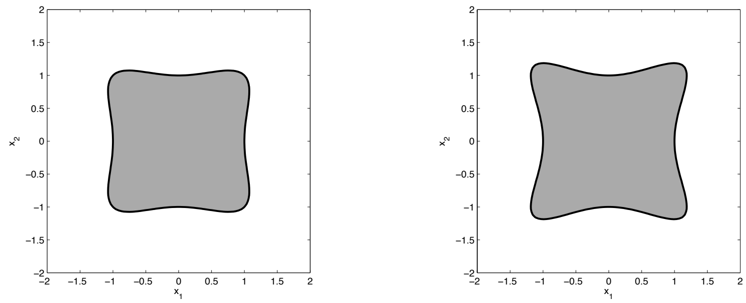

FIGURE 1. $\mathbf{G}_1$ with $x^4 + y^4 - x^2y^2$ and $x^4 + y^4 - 1.4(x^2y^2)$

of $\mathbf{K}_1$ inside $\mathbf{K}_2$, whereas Bastero and Romance [5] refined this result by allowing $\mathbf{K}_1$ to be non-convex.

In this paper we consider a different non convex generalization of the Löwner-John ellipsoid problem, with a more algebraic flavor. Namely, we address the following two problems $\mathbf{P}_0$ and $\mathbf{P}$.

$\mathbf{P}_0$*: Let* $\mathbf{K} \subset \mathbb{R}^n$ *be a compact set (not necessarily convex) and let* d *be an even integer. Find an homogeneous polynomial* g *of degree* d *such that its sublevel set* $\mathbf{G}_1 := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ *contains* $\mathbf{K}$ *and has minimum volume among all such sub level sets with this inclusion property.*

$\mathbf{P}$*: Let* $\mathbf{K} \subset \mathbb{R}^n$ *be a compact set (not necessarily convex) and let* d *be an even integer. Find an homogeneous polynomial* g *of degree* d *and* $\mathbf{a} \in \mathbb{R}^n$ *such that the sublevel set* $\mathbf{G}_1^{\mathbf{a}} := \{\mathbf{x} : g(\mathbf{x} - \mathbf{a}) \leq 1\}$ *contains* $\mathbf{K}$ *and has minimum volume among all such sub level sets with this inclusion property.*

Necessarily g is a nonnegative homogeneous polynomial otherwise $\mathbf{G}_1$ and $\mathbf{G}_1^{\mathbf{a}}$ are not bounded. Of course, when $d = 2$ then g is convex (i.e., $\mathbf{G}_1$ and $\mathbf{G}_1^{\mathbf{a}}$ are ellipsoids) because every nonnegative quadratic form defines a convex function, and g is an optimal solution for problem $\mathbf{P}$ with $\mathbf{K}$ or its convex hull $\mathrm{co}(\mathbf{K})$. That is, one retrieves the Löwner-John's problem. But when $d > 2$ then $\mathbf{G}_1$ and $\mathbf{G}_1^{\mathbf{a}}$ are not necessarily convex. For instance, take $\mathbf{K} = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ where g is some nonnegative homogeneous polynomial such that $\mathbf{K}$ is compact but non convex. Then g is an optimal solution for problem $\mathbf{P}_0$ with $\mathbf{K}$ and cannot be optimal for $\mathrm{co}(\mathbf{K})$; a two-dimensional example is $(x, y) \mapsto g(x, y) := x^4 + y^4 - x^2y^2$:sees Figure 1.

Contribution. Our contribution is to show that problem $\mathbf{P}_0$ and $\mathbf{P}$ are indeed natural generalizations of the Löwner-John ellipsoid problem in the sense that:

- $\mathbf{P}_0$ also has a *unique* solution g^*.

- A characterization of g^* also involves s contact points in $\mathbf{K}\cap\mathbf{G}_1$, where s is now bounded by $\binom{n+d-1}{d}$.

And so when $d=2$ we retrieve the symmetric Löwner-John problem as a particular case. In fact it is shown that $\mathbf{P}_0$ is a convex optimization problem no matter if neither $\mathbf{K}$ nor $\mathbf{G}_1$ are convex!

We use an intermediate and crucial result of independent interest. Namely, the Lebesgue-volume function $g\mapsto f(g) := \mathrm{vol}(\mathbf{G}_1)$ is a strictly convex function of the coefficients of g, which is far from being obvious from its definition.

Concerning the more general problem $\mathbf{P}$, we also show that there is an optimal solution $(g^*,\mathbf{a}^*)$ with again a characterization which involves s contact points in $\mathbf{K}\cap\mathbf{G}_1^{\mathbf{a}^*}$, but uniqueness is not guaranteed.

Again and importantly, in both problems $\mathbf{P}_0$ and $\mathbf{P}$, neither $\mathbf{K}$ nor $\mathbf{G}_1^{\mathbf{a}}$ are required to be convex.

On the computational side. Even though $\mathbf{P}_0$ is a convex optimization problem, it is hard to solve because even if $\mathbf{K}$ would be a finite set of of points (as is the case in statistics applications of the Löwner-John problem) and in contrast to the quadratic case, evaluating the (strictly convex) objective function, its gradient and Hessian can be a challenging problem, especially if the dimension is larger than $n=3$. So this is one price to pay for the generalization of the Löwner-John's ellipsoid problem. (Notice however that if $\mathbf{K}$ is not a finite set of points then even the Löwner-John's ellipsoid problem is also hard to solve because for more general sets $\mathbf{K}$ the inclusion constraint $\mathbf{K}\subset\xi$ (or $\mathrm{conv}(\mathbf{K})\subset\xi$) can be difficult to handle.)

However, we can still approximate as closely as desired the objective function as well as its gradient and Hessian by using the methodology developed in Henrion et al [14], especially when the dimension is small $n=2,3$ (which is the case in several applications in statistics).

Moreover, if $\mathbf{K}$ is a (compact) basic semi-algebraic with an explicit description $\{\mathbf{x} : g_j(\mathbf{x})\geq 0,\ j=1,\ldots,m\}$ for some polynomials $(g_j)\subset\mathbb{R}[\mathbf{x}]$, then we can use powerful positivity certificates from real algebraic geometry to handle the inclusion constraint $\mathbf{G}_1\supset\mathbf{K}$ in the associated convex optimization problem.

Therefore, in this context, we also provide a numerical scheme to approximate the optimal value and the unique optimal solution of $\mathbf{P}_0$ as closely as desired. It consists of solving a hierarchy of convex optimization problems where each problem in the hierarchy has a strictly convex objective function and a feasible set defined by Linear Matrix Inequalities (LMIs).

2. Notation, definitions and preliminary results

2.1. **Notation and definitions.** Let $\mathbb{R}[\mathbf{x}]$ be the ring of polynomials in the variables $\mathbf{x}=(x_1,\ldots,x_n)$ and let $\mathbb{R}[\mathbf{x}]_d$ be the vector space of polynomials of degree at most d (whose dimension is $s(d) := \binom{n+d}{n}$). For every $d\in\mathbb{N}$, let $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : |\alpha|\,(=\sum_i \alpha_i) \leq d\}$, and let $\mathbf{v}_d(\mathbf{x}) = (\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n$, be the vector of monomials of the canonical basis $(\mathbf{x}^\alpha)$ of $\mathbb{R}[\mathbf{x}]_d$.

A polynomial $f \in \mathbb{R}[\mathbf{x}]_d$ is written

$$\mathbf{x} \mapsto f(\mathbf{x}) \,=\, \sum_{\alpha \in \mathbb{N}^n} f_\alpha\, \mathbf{x}^\alpha,$$

for some vector of coefficients $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(d)}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]_d$ is homogeneous of degree d if $f(\lambda\mathbf{x}) = \lambda^d f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\lambda \in \mathbb{R}$.

Let us denote by $\mathbf{P}[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_d$, $d \in \mathbb{N}$, the set of nonnegative and homogeneous polynomials of degree d such that their sublevel set $\mathbf{G}_1 := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ has finite Lebesgue volume (denoted $\mathrm{vol}(\mathbf{G}_1)$). Notice that necessarily d is even and $0 \notin \mathbf{P}[\mathbf{x}]_d$.

For $d \in \mathbb{N}$ and a closed set $\mathbf{K} \subset \mathbb{R}^n$, denote by $C_d(\mathbf{K})$ the convex cone of all polynomials of degree d that are nonnegative on $\mathbf{K}$, and denote by $\mathcal{M}(\mathbf{K})$ the Banach space of signed Borel measures with support contained in $\mathbf{K}$ (equipped with the total variation norm). Let $M(\mathbf{K}) \subset \mathcal{M}(\mathbf{K})$ be the convex cone of finite Borel measures on $\mathbf{K}$.

In the Euclidean space $\mathbb{R}^n$ we denote by $\langle\cdot,\cdot\rangle$ the usual duality bracket.

2.2. **Some preliminary results.** We first have the following result:

Proposition 2.1. *The set $\mathbf{P}[\mathbf{x}]_d$ is a convex cone.*

Proof. Let $g, h \in \mathbf{P}[\mathbf{x}]_d$ with associated sublevel sets $\mathbf{G}_1 = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ and $\mathbf{H}_1 = \{\mathbf{x} : h(\mathbf{x}) \leq 1\}$. For $\lambda \in (0,1)$, consider the nonnegative homogeneous polynomial $\theta := \lambda g + (1-\lambda)h \in \mathbb{R}[\mathbf{x}]_d$, with associated sublevel set

$$\Theta_1 \,:=\, \{\mathbf{x} : \theta(\mathbf{x}) \,\leq\, 1\} \,=\, \{\mathbf{x} : \lambda g(\mathbf{x}) + (1-\lambda)h(\mathbf{x}) \,\leq\, 1\}.$$

Write $\Theta_1 = \Theta_1^1 \cup \Theta_1^2$ where $\Theta_1^1 = \Theta_1 \cap \{\mathbf{x} : g(\mathbf{x}) \geq h(\mathbf{x})\}$ and $\Theta_1^2 = \Theta_1 \cap \{\mathbf{x} : g(\mathbf{x}) < h(\mathbf{x})\}$. Observe that $\mathbf{x} \in \Theta_1^1$ implies $h(\mathbf{x}) \leq 1$ and so $\Theta_1^1 \subset \mathbf{H}_1$. Similarly $\mathbf{x} \in \Theta_1^2$ implies $g(\mathbf{x}) \leq 1$ and so $\Theta_1^2 \subset \mathbf{G}_1$. Therefore $\mathrm{vol}(\Theta_1) \leq \mathrm{vol}(\mathbf{G}_1) + \mathrm{vol}(\mathbf{H}_1) < \infty$. And so $\theta \in \mathbf{P}[\mathbf{x}]_d$. □

With $y \in \mathbb{R}$ and $g \in \mathbb{R}[\mathbf{x}]$ let $\mathbf{G}_y := \{\mathbf{x} : g(\mathbf{x}) \leq y\}$. The following intermediate result is crucial and of independent interest.

Lemma 2.2. *Let $g \in \mathbf{P}[\mathbf{x}]_d$. Then for every $y \geq 0$:*

$$\mathrm{vol}(\mathbf{G}_y) \,=\, \frac{y^{n/d}}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\, d\mathbf{x}. \tag{2.1}$$

Proof. As $g \in \mathbf{P}[\mathbf{x}]_d$, and using homogeneity, $\mathrm{vol}(\mathbf{G}_1) < \infty$ implies $\mathrm{vol}(\mathbf{G}_y) < \infty$ for every $y \geq 0$. Let $f : \mathbb{R} \to \mathbb{R}$ be the function $y \mapsto f(y) := \mathrm{vol}(\mathbf{G}_y)$. Since g is nonnegative, the function f vanishes on $(-\infty, 0]$. Its Laplace transform $\mathcal{L}_f : \mathbb{C} \to \mathbb{C}$ is the function

$$\lambda \mapsto \mathcal{L}_f(\lambda) := \int_0^\infty \exp(-\lambda y) f(y)\, dy, \qquad \Re\,\lambda \,>\, 0.$$

Observe that whenever $\lambda \in \mathbb{R}$ with $\lambda > 0$,

$$\begin{aligned}
\mathcal{L}_f(\lambda) &= \int_0^\infty \exp(-\lambda y) \left(\int_{\{\mathbf{x} : g(\mathbf{x}) \le y\}} d\mathbf{x} \right) dy \\
&= \int_{\mathbb{R}^n} \left(\int_{g(\mathbf{x})}^\infty \exp(-\lambda y) dy \right) d\mathbf{x} \quad \text{[by Fubini's Theorem]} \\
&= \frac{1}{\lambda} \int_{\mathbb{R}^n} \exp(-\lambda g(\mathbf{x}))\, d\mathbf{x} \\
&= \frac{1}{\lambda} \int_{\mathbb{R}^n} \exp(-g(\lambda^{1/(d)} \mathbf{x}))\, d\mathbf{x} \quad \text{[by homogeneity]} \\
&= \frac{1}{\lambda^{1+n/d}} \int_{\mathbb{R}^n} \exp(-g(\mathbf{z}))\, d\mathbf{z} \quad [\text{by } \lambda^{-1/(d)} \mathbf{x} \to \mathbf{z}] \\
&= \frac{\displaystyle\int_{\mathbb{R}^n} \exp(-g(\mathbf{z}))\, d\mathbf{z}}{\Gamma(1+n/d)}\, \mathcal{L}_{y^{n/d}}(\lambda).
\end{aligned}$$

And so, by analyticity of the Laplace transform $\mathcal{L}_f$,

$$f(y) = \frac{y^{n/d}}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\, d\mathbf{x}, \qquad y \ge 0,$$

which is the desired result. □

And we also conclude:

Corollary 2.3. $g \in \mathbf{P}[\mathbf{x}]_d \iff \displaystyle\int_{\mathbb{R}^n} \exp(-g(\mathbf{x})) d\mathbf{x} < \infty.$

Proof. The implication $\Rightarrow$ follows from Lemma 2.2. For the reverse implication consider the function

$$\lambda \mapsto h(\lambda) := \frac{1}{\lambda^{1+n/d}} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\, d\mathbf{x}, \qquad 0 < \lambda \in \mathbb{R}.$$

Proceeding as in the proof of Lemma 2.2 but backward, it follows that h is the Laplace transform of the function $y \mapsto \mathrm{vol}(\mathbf{G}_y)$ and the proof is complete. □

Formula (2.1) relating the Lebesgue volume $\mathbf{G}_1$ with $\int \exp(-g)$ is already proved (with a different argument) in Morozov and Shakirov [20, 21] where the authors want to express the non Gaussian integral $\int \exp(-g)$ in terms of algebraic invariants of g.

Sensitivity analysis and convexity. We now investigate some properties the function $f : \mathbf{P}[\mathbf{x}]_d \to \mathbb{R}$ defined by

$$g \mapsto f(g) := \mathrm{vol}(\mathbf{G}_1) = \int_{\{\mathbf{x} \,:\, g(\mathbf{x}) \le 1\}} d\mathbf{x}, \qquad g \in \mathbf{P}(\mathbf{x}]_d, \tag{2.2}$$

i.e., we now view $\mathrm{vol}(\mathbf{G}_1)$ as a function of the vector $\mathbf{g} = (g_\alpha) \in \mathbb{R}^{\ell(d)}$ of coefficients of g in the canonical basis of homogeneous polynomials of degree d (and $\ell(d) = \binom{n+d-1}{d}$).

Theorem 2.4. *The Lebesgue-volume function $f : \mathbf{P}(\mathbf{x}]_d \to \mathbb{R}$ defined in (2.2) is strictly convex with gradient ∇f and Hessian $\nabla^2 f$ given by:*

$$\frac{\partial f(g)}{\partial g_\alpha} \;=\; \frac{-1}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, \exp(-g(\mathbf{x}))\, d\mathbf{x}, \tag{2.3}$$

for all $\alpha \in \mathbb{N}^n_d$, $|\alpha| = d$.

$$\frac{\partial^2 f(g)}{\partial g_\alpha \partial g_\beta} \;=\; \frac{1}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \mathbf{x}^{\alpha+\beta} \, \exp(-g(\mathbf{x}))\, d\mathbf{x}, \tag{2.4}$$

for all $\alpha, \beta \in \mathbb{N}^n_d$, $|\alpha| = |\beta| = d$.

Moreover, we also have

$$\int_{\mathbb{R}^n} g(\mathbf{x}) \, \exp(-g(\mathbf{x}))\, d\mathbf{x} \;=\; \frac{n}{d} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\, d\mathbf{x}. \tag{2.5}$$

Proof. By Lemma 2.2 $f(g) = \Gamma(1+n(d))^{-1} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))d\mathbf{x}$. Let $p, q \in \mathbf{P}[\mathbf{x}]_d$ and $\alpha \in [0,1]$. By convexity of $u \mapsto \exp(-u)$,

$$\begin{aligned} f(\alpha g + (1-\alpha) q) \;&\leq\; \int_{\mathbb{R}^n} [\, \alpha \exp(-g(\mathbf{x})) + (1-\alpha) \exp(-q(\mathbf{x}))\,]\, d\mathbf{x} \\ &=\; \alpha f(g) + (1-\alpha) f(q), \end{aligned}$$

and so f is convex. Next, in view of the strict convexity of $u \mapsto \exp(-u)$, equality may occur only if $g(\mathbf{x}) = q(\mathbf{x})$ almost everywhere, which implies $g = q$ and which in turn implies strict convexity of f.

To obtain (2.3)-(2.4) one takes partial derivatives under the integral sign, which in this context is allowed. Indeed, write g in the canonical basis as $g(\mathbf{x}) = \sum_{|\alpha|=d} g_\alpha \mathbf{x}^\alpha$. For every $\alpha \in \mathbb{N}^n_d$ with $|\alpha| = d$, let $e_\alpha = (e_\alpha(\beta)) \in \mathbb{R}^{\ell(d)}$ be such that $e_\alpha(\beta) = \delta_{\beta=\alpha}$ (with δ being the Kronecker symbol). Then for every $t \geq 0$,

$$\frac{f(g + te_\alpha) - f(g)}{t} \;=\; \int_{\mathbb{R}^n} \exp(-g) \, \left(\underbrace{\frac{\exp(-t\mathbf{x}^\alpha) - 1}{t}}_{\psi(t,\mathbf{x})} \right) \, d\mathbf{x}.$$

Notice that for every $\mathbf{x}$, by convexity of the function $t \mapsto \exp(-t\mathbf{x}^\alpha)$,

$$\lim_{t \downarrow 0} \psi(t,\mathbf{x}) \;=\; \inf_{t \geq 0} \psi(t,\mathbf{x}) \;=\; \exp(-t\mathbf{x}^\alpha)'_{|t=0} \;=\; -\mathbf{x}^\alpha,$$

because for every $\mathbf{x}$, the function $t \mapsto \psi(t,\mathbf{x})$ is nondecreasing; see e.g. Rockafellar [27, Theorem 23.1]. Hence, the one-sided directional derivative

$f'(g;\mathrm{e}_\alpha)$ in the direction e_α satisfies

$$\begin{aligned} f'(g;e_\alpha) &= \lim_{t\downarrow 0} \frac{f(g+te_\alpha)-f(g)}{t} = \lim_{t\downarrow 0} \int_{\mathbb{R}^n} \exp(-g)\,\psi(t,\mathbf{x})\,d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \exp(-g)\,\lim_{t\downarrow 0}\psi(t,\mathbf{x})\,d\mu(\mathbf{x}) = \int_{\mathbb{R}^n} -\mathbf{x}^\alpha \exp(-g)\,d\mathbf{x}, \end{aligned}$$

where the third equality follows from the Extended Monotone Convergence Theorem [2, 1.6.7]. Indeed for all $t < t_0$ with t_0 sufficiently small, the function $\psi(t,\cdot)$ is bounded above by $\psi(t_0,\cdot)$ and $\int_{\mathbb{R}^n} \exp(-g)\psi(t_0,\mathbf{x})d\mu < \infty$. Similarly, for every $t > 0$

$$\frac{f(g-te_\alpha)-f(g)}{t} = \int_{\mathbb{R}^n} \exp(-g)\,\underbrace{\frac{\exp(t\mathbf{x}^\alpha)-1}{t}}_{\xi(t,\mathbf{x})}\,d\mathbf{x},$$

and by convexity of the function $t \mapsto \exp(t\mathbf{x}^\alpha)$

$$\lim_{t\downarrow 0} \xi(t,\mathbf{x}) = \inf_{t\geq 0} \xi(t,\mathbf{x}) = \exp(t\mathbf{x}^\alpha)'_{|t=0} = \mathbf{x}^\alpha.$$

Therefore, with exactly same arguments as before,

$$\begin{aligned} f'(g;-e_\alpha) &= \lim_{t\downarrow 0} \frac{f(g-te_\alpha)-f(g)}{t} \\ &= \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g)\,d\mathbf{x} = -f'(g;e_\alpha), \end{aligned}$$

and so

$$\frac{\partial f(g)}{\partial g_\alpha} = -\int_{\mathbb{R}^n} \mathbf{x}^\alpha\,\exp(-g)\,d\mathbf{x},$$

for every α with $|\alpha| \leq d$, which yields (2.3). Similar arguments can used for the Hessian $\nabla^2 f(g)$ which yields (2.4).

To obtain (2.5) observe that $g \mapsto h(g) := \int \exp(-g)d\mathbf{x}$, $g \in \mathbf{P}[\mathbf{x}]_d$, is a positively homogeneous function of degree $-n/d$, continuously differentiable. And so combining (2.3) with Euler's identity yields

$$\begin{aligned} -\frac{n}{d}\int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\,d\mathbf{x} &= -\frac{n}{d}\,h(g) \\ &= \langle \nabla h(g), g\rangle \quad \text{[by Euler's identity]} \\ &= -\int_{\mathbb{R}^n} g(\mathbf{x})\,\exp(-g(\mathbf{x}))\,d\mathbf{x}. \end{aligned}$$

□

Notice that convexity of f is not obvious at all from its definition (2.2) whereas it becomes almost transparent when using formula (2.1).

2.3. **The dual cone of $C_d(\mathbf{K})$.** For a convex cone $C\subset\mathbb{R}^n$, the convex cone

$$C^* := \{\mathbf{y} : \langle \mathbf{y},\mathbf{x}\rangle \geq 0 \quad \forall \mathbf{x}\in C\},$$

is called the *dual* cone of C, and if C is closed then $(C^*)^* = C$.

Recall that for a set $\mathbf{K}\subset\mathbb{R}^n$, $C_d(\mathbf{K})$ denotes the convex cone of polynomials of degree at most d which are nonnegative on $\mathbf{K}$. We say that a vector $\mathbf{y}\in\mathbb{R}^{s(d)}$ has a representing measure (or is a d-truncated moment sequence) if there exists a finite Borel measure ϕ such that

$$y_\alpha = \int_{\mathbb{R}^n} \mathbf{x}^\alpha\, d\phi, \qquad \forall \alpha\in\mathbb{N}^n_d.$$

We will need the following characterization the dual cone $C_d(\mathbf{K})^*$.

Lemma 2.5. *Let $\mathbf{K}\subset\mathbb{R}^n$ be compact. For every $d\in\mathbb{N}$, the dual cone $C_d(\mathbf{K})^*$ is the convex cone*

$$\Delta_d := \left\{ \left(\int_{\mathbf{K}} \mathbf{x}^\alpha\, d\phi \right),\ \alpha\in\mathbb{N}^n_d \ :\ \phi\in M(\mathbf{K}) \right\}, \tag{2.6}$$

i.e., the convex cone of vectors of $\mathbb{R}^{s(d)}$ which have a representing measure with support contained in $\mathbf{K}$.

Proof. For every $\mathbf{y}=(y_\alpha)\in\Delta_d$ and $f\in C_d(\mathbf{K})$ with coefficient vector $\mathbf{f}\in\mathbb{R}^{s(d)}$:

$$\langle \mathbf{y},\mathbf{f}\rangle = \sum_{\alpha\in\mathbb{N}^n_d} f_\alpha\, y_\alpha = \sum_{\alpha\in\mathbb{N}^n_d} \int_{\mathbf{K}} f_\alpha\, \mathbf{x}^\alpha\, d\phi = \int_{\mathbf{K}} f\, d\phi \geq 0. \tag{2.7}$$

Since (2.7) holds for all $f\in C_d(\mathbf{K})$ and all $\mathbf{y}\in\Delta_d$, then necessarily $\Delta_d\subseteq C_d(\mathbf{K})^*$ and similarly, $C_d(\mathbf{K})\subseteq\Delta_d^*$. Next,

$$\begin{aligned}
\Delta_d^* &= \left\{ \mathbf{f}\in\mathbb{R}^{s(d)} \ :\ \langle \mathbf{f},\mathbf{y}\rangle \geq 0 \quad \forall \mathbf{y}\in\Delta_d \right\} \\
&= \left\{ f\in\mathbb{R}[\mathbf{x}]_d \ :\ \int_{\mathbf{K}} f\, d\phi \geq 0 \quad \forall \phi\in M(\mathbf{K}) \right\} \\
&\Rightarrow \ \Delta_d^*\subseteq C_d(\mathbf{K}),
\end{aligned}$$

and so $\Delta_d^* = C_d(\mathbf{K})$. Hence the result follows if one proves that Δ_d is closed, because then $C_d(\mathbf{K})^* = (\Delta_d^*)^* = \Delta_d$, the desired result. So let $(\mathbf{y}^k)\subset\Delta_d$, $k\in\mathbb{N}$, with $\mathbf{y}^k\to\mathbf{y}$ as $k\to\infty$. Equivalently, $\int_{\mathbf{K}} \mathbf{x}^\alpha d\phi_k \to y_\alpha$ for all $\alpha\in\mathbb{N}^n_d$. In particular, the convergence $y_0^k\to y_0$ implies that the sequence of measures (ϕ_k), $k\in\mathbb{N}$, is bounded, that is, $\sup_k \phi_k(\mathbf{K}) < M$ for some $M>0$. As $\mathbf{K}$ is compact, the unit ball of $\mathcal{M}(\mathbf{K})$ is sequentially compact in the weak $\star$ topology $\sigma(\mathcal{M}(\mathbf{K}), C(\mathbf{K}))$ where $C(\mathbf{K})$ is the space of continuous functions on $\mathbf{K}$. Hence there is a finite Borel measure $\phi\in M(\mathbf{K})$ and a subsequence (k_i) such that $\int_{\mathbf{K}} g d\phi_{k_i} \to \int_{\mathbf{K}} g d\phi$ as $i\to\infty$, for all $g\in C(\mathbf{K})$. In particular, for every $\alpha\in\mathbb{N}^n_d$,

$$y_\alpha = \lim_{k\to\infty} y^k_\alpha = \lim_{i\to\infty} y^{k_i}_\alpha = \lim_{i\to\infty} \int_{\mathbf{K}} \mathbf{x}^\alpha d\phi_{k_i} = \int_{\mathbf{K}} \mathbf{x}^\alpha d\phi,$$

which shows that $\mathbf{y} \in \Delta_d$, and so Δ_d is closed. □

And we also have:

Lemma 2.6. *Let $\mathbf{K} \subset \mathbb{R}^n$ be with nonempty interior. Then the interior of $\mathbf{C}_d(\mathbf{K})^*$ is nonempty.*

Proof. Since $C_d(\mathbf{K})$ is nonempty and closed, by Faraut and Korány [11, Prop. I.1.4, p. 3]

$$\mathrm{int}(C_d(\mathbf{K})^*) \,=\, \{\mathbf{y} \,:\, \langle \mathbf{y}, \mathbf{g}\rangle > 0, \quad \forall g \in C_d(\mathbf{K}) \setminus \{0\}\},$$

where $\mathbf{g} \in \mathbb{R}^{s(d)}$ is the coefficient of $g \in C_d(\mathbf{K})$, and

$$\mathrm{int}(C_d(\mathbf{K})^*) \,\neq\, \emptyset \Longleftrightarrow C_d(\mathbf{K}) \cap (-C_d(\mathbf{K})) \,=\, \{0\}.$$

But $g \in C_d(\mathbf{K}) \cap (-C_d(\mathbf{K}))$ implies $g \geq 0$ and $g \leq 0$ on $\mathbf{K}$, which in turn implies $g = 0$ because $\mathbf{K}$ has nonempty interior. □

3. Main result

Consider the following problem $\mathbf{P}_0$, a non convex generalization of the Löwner-John ellipsoid problem:

$\mathbf{P}_0$*: Let $\mathbf{K} \subset \mathbb{R}^n$ be a compact set not necessarily convex and d an even integer. Find an homogeneous polynomial g of degree d such that its sublevel set $\mathbf{G}_1 := \{\mathbf{x} \,:\, g(\mathbf{x}) \leq 1\}$ contains $\mathbf{K}$ and has minimum volume among all such sublevel sets with this inclusion property.*

In the above problem $\mathbf{P}_0$, the set $\mathbf{G}_1$ is symmetric and so when $\mathbf{K}$ is a symmetric convex body and $d = 2$, one retrieves the Löwner-John ellipsoid problem in the symmetric case. In the next section we will consider the more general case where $\mathbf{G}_1$ is of the form $\mathbf{G}_1^{\mathbf{a}} := \{\mathbf{x} : g(\mathbf{x} - \mathbf{a}) \leq 1\}$ for some $\mathbf{a} \in \mathbb{R}^n$ and some $g \in \mathbf{P}[\mathbf{x}]_d$.

Recall that $\mathbf{P}[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_d$ is the convex cone of nonnegative homogeneous polynomials of degree d whose sublevel set $\mathbf{G}_1 = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ has finite volume. Recall also that $C_d(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_d$ is the convex cone of polynomials of degree at most d that are nonnegative on $\mathbf{K}$.

We next show that solving $\mathbf{P}$ is equivalent to solving the convex optimization problem:

$$\mathcal{P}: \qquad \rho = \inf_{g \in \mathbf{P}[\mathbf{x}]_d} \left\{ \int_{\mathbb{R}^n} \exp(-g)\, d\mathbf{x} \,:\, 1 - g \,\in\, C_d(\mathbf{K}) \right\}. \tag{3.1}$$

Proposition 3.1. *Problem $\mathbf{P}$ has an optimal solution if and only if problem $\mathcal{P}$ in (3.1) has an optimal solution. Moreover, $\mathcal{P}$ is a finite-dimensional convex optimization problem.*

Proof. By Lemma 2.2,

$$\mathrm{vol}\,(\mathbf{G}_1) \,=\, \frac{1}{\Gamma(1 + n/d)} \int_{\mathbb{R}^n} \exp(-g)\, d\mathbf{x}$$

whenever $\mathbf{G}_1$ has finite Lebesgue volume. Moreover, $\mathbf{G}_1$ contains $\mathbf{K}$ if and only if $1-g\in C_d(\mathbf{K})$, and so $\mathbf{P}$ has an optimal solution $g^*\in\mathbf{P}[\mathbf{x}]_d$ if and only if g^* is an optimal solution of $\mathcal{P}$ (with value $\mathrm{vol}(\mathbf{G}_1)\Gamma(1+n/2(d))$). Now since $g\mapsto\int_{\mathbb{R}^n}\exp(-g)d\mathbf{x}$ is strictly convex (by Lemma 2.4) and both $C_d(\mathbf{K})$ and $\mathbf{P}[\mathbf{x}]_d$ are convex cones, problem $\mathcal{P}$ is a finite-dimensional convex optimization problem. □

We now can state the first result of this paper: Recall that $M(\mathbf{K})$ is the convex cone of finite Borel measures on $\mathbf{K}$.

Theorem 3.2. *Let $\mathbf{K}\subset\mathbb{R}^n$ be compact with nonempty interior and consider the convex optimization problem $\mathcal{P}$ in (3.1).*

(a) *$\mathcal{P}$ has a unique optimal solution $g^*\in\mathbf{P}[\mathbf{x}]_d$.*

(b) *Let $g^*\in\mathbf{P}[\mathbf{x}]_d$ be the unique optimal solution of $\mathcal{P}$. Then there exists a finite Borel measure $\mu^*\in M(\mathbf{K})$ such that*

$$\int_{\mathbb{R}^n}\mathbf{x}^\alpha\exp(-g^*)d\mathbf{x} \;=\; \int_{\mathbf{K}}\mathbf{x}^\alpha\,d\mu^*,\qquad\forall|\alpha|=d \tag{3.2}$$

$$\int_{\mathbf{K}}(1-g^*)\,d\mu^* \;=\; 0;\quad \mu^*(\mathbf{K})=\frac{n}{d}\int_{\mathbb{R}^n}\exp(-g^*)\,d\mathbf{x}. \tag{3.3}$$

In particular, μ^* is supported on the set $V:=\{\mathbf{x}\in\mathbf{K}:g^*(\mathbf{x})=1\}$ ($=\mathbf{K}\cap\mathbf{G}_1^*$) and in fact, μ^* can be substituted with another measure $\nu^*\in M(\mathbf{K})$ supported on at most $\binom{n+d-1}{d}$ contact points of V.

(c) *Conversely, if $g^*\in\mathbb{R}[\mathbf{x}]_d$ is homogeneous with $1-g^*\in C_d(\mathbf{K})$, and there exist points $(\mathbf{x}(i),\lambda_i)\in\mathbf{K}\times\mathbb{R}$, $\lambda_i>0$, $i=1,\ldots,s$, such that $g^*(\mathbf{x}(i))=1$ for all $i=1,\ldots,s$, and*

$$\int_{\mathbb{R}^n}\mathbf{x}^\alpha\exp(-g^*)\,d\mathbf{x}\;=\;\sum_{i=1}^s\lambda_i\,\mathbf{x}(i)^\alpha,\qquad|\alpha|=d,$$

then g^* is the unique optimal solution of problem $\mathcal{P}$.

Proof. (a) As $\mathcal{P}$ is a minimization problem, its feasible set $\{g\in\mathbf{P}[\mathbf{x}]_d:1-g\in C_d(\mathbf{K})\}$ can be replaced by the smaller set

$$F\;:=\;\left\{g\in\mathbf{P}[\mathbf{x}]_d\;:\;\begin{array}{l}\displaystyle\int_{\mathbb{R}^n}\exp(-g(\mathbf{x}))\,d\mathbf{x}\;\leq\;\int_{\mathbb{R}^n}\exp(-g_0(\mathbf{x}))\,d\mathbf{x}\\ 1-g\in\,C_d(\mathbf{K})\end{array}\right\},$$

for some $g_0\in\mathbf{P}[\mathbf{x}]_d$. The set F is a closed convex set since the convex function $g\mapsto\int_{\mathbb{R}^n}\exp(-g)d\mathbf{x}$ is continuous on the interior of its domain.

Next, let $\mathbf{z}=(z_\alpha)$, $\alpha\in\mathbb{N}^n_d$, be a (fixed) element of $\mathrm{int}(C_d(\mathbf{K})^*)$ (hence $z_0>0$). By Lemma 2.6 such an element exists. Then the constraint $1-g\in C_d(\mathbf{K})$ implies $\langle\mathbf{z},1-g\rangle\geq0$, i.e., $\langle\mathbf{z},g\rangle\leq z_0$. On the other hand, being an element of $\mathbf{P}[\mathbf{x}]_d$, g is nonnegative and in particular $g\in C_d(\mathbf{K})$. But then by Corollary I.1.6 in Faraut et Korányi [11, p. 4], the set $\{g\in C_d(\mathbf{K}):\langle\mathbf{z},\mathbf{g}\rangle\leq z_0\}$ is compact (where again $\mathbf{g}\in\mathbb{R}^{s(d)}$ is the coefficient vector of $g\in C_d(\mathbf{K})$). Therefore, the set F is a compact convex set. Finally, since

$g \mapsto \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))d\mathbf{x}$ is strictly convex, it is continuous on the interior of its domain and so it is continuous on F. Hence problem $\mathcal{P}$ has a unique optimal solution $g^* \in \mathbf{P}[\mathbf{x}]_d$.

(b) We may and will consider any homogeneous polynomial g as an element of $\mathbb{R}[\mathbf{x}]_d$ whose coefficient vector $\mathbf{g} = (g_\alpha)$ is such that $g^*_\alpha = 0$ whenever $|\alpha| < d$. And so Problem $\mathcal{P}$ is equivalent to the problem

$$(3.4)\qquad \mathcal{P}' : \quad \begin{cases} \rho = \inf\limits_{g\in\mathbb{R}[\mathbf{x}]_d} & \displaystyle\int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\,d\mathbf{x} \\ \text{s.t.} & g_\alpha = 0, \quad \forall\,\alpha \in \mathbb{N}^n_d;\ |\alpha| < d \\ & 1 - g \,\in\, C_d(\mathbf{K}), \end{cases}$$

where we replaced $g \in \mathbf{P}[\mathbf{x}]_d$ with the equivalent constraints $g \in \mathbb{R}[\mathbf{x}]_d$ and $g_\alpha := 0$ for all $\alpha \in \mathbb{N}^n_d$ with $|\alpha| < d$. Next, doing the change of variable $h = 1 - g$, $\mathcal{P}$' reads:

$$(3.5)\qquad \mathcal{P}' : \quad \begin{cases} \rho = \inf\limits_{h\in\mathbb{R}[\mathbf{x}]_d} & \displaystyle\int_{\mathbb{R}^n} \exp(h(\mathbf{x}) - 1)\,d\mathbf{x} \\ \text{s.t.} & h_\alpha = 0, \quad \forall\,\alpha \in \mathbb{N}^n_d;\ 0 < |\alpha| < d \\ & h_0 = 1 \\ & h \,\in\, C_d(\mathbf{K}), \end{cases}$$

As $\mathbf{K}$ is compact, there exists $\theta \in \mathbf{P}[\mathbf{x}]_d$ such that $1-\theta \in \text{int}(C_d(\mathbf{K}))$, i.e., Slater's condition[1] holds for the convex optimization problem $\mathcal{P}'$. Indeed, choose $\mathbf{x} \mapsto \theta(\mathbf{x}) := M^{-1}\|\mathbf{x}\|^d$ for $M > 0$ sufficiently large so that $1 - \theta > 0$ on $\mathbf{K}$. Hence with $\|g\|_1$ denoting the ℓ_1-norm of the coefficient vector of g (in $\mathbb{R}[\mathbf{x}]_d$), there exists $\epsilon > 0$ such that for every $h \in B(\theta, \epsilon)(:= \{h \in \mathbb{R}[\mathbf{x}]_d : \|\theta - h\|_1 < \epsilon\})$, the polynomial $1 - h$ is (strictly) positive on $\mathbf{K}$.

Therefore, the unique optimal solution $(1-g^*) =: h^* \in \mathbb{R}[\mathbf{x}]_d$ of $\mathcal{P}$' in (3.5) satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions[2], which for problem (3.5) read:

$$(3.6)\qquad \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(h^*(\mathbf{x}) - 1)\,d\mathbf{x} \;=\; y^*_\alpha, \quad \forall|\alpha| = d$$

$$(3.7)\qquad \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(h^*(\mathbf{x}) - 1)\,d\mathbf{x} + \gamma_\alpha \;=\; y^*_\alpha, \quad \forall\,|\alpha| < d$$

$$(3.8)\qquad \langle h^*, \mathbf{y}^*\rangle \;=\; 0; \quad h^*_0 \;=\; 1;\ h^*_\alpha \;=\; 0, \quad \forall\, 0 < |\alpha| < d$$

[1]Slater's condition holds for an optimization problem $\{\min g_0(\mathbf{x}) : g_k(\mathbf{x}) \le 0,\ k = 1,\dots,m\}$ if there exists $\mathbf{x}_0$ such that $g_k(\mathbf{x}_0) < 0$ for all $k = 1,\dots,m$. If the g_j's are convex, continuously differentiable, and Slater's condition holds then the Karush-Kuhn-Tucker optimality conditions are necessary and sufficient optimality conditions.

[2]For an optimization problem $\{\min g_0(\mathbf{x}) : g_k(\mathbf{x}) \le 0,\ k = 1,\dots,m\}$, the KKT-optimality conditions hold at a feasible point $\mathbf{x}^*$ is there exists $0 \le \lambda \in \mathbb{R}^m$ such that $\nabla(f(\mathbf{x}^*) + \sum_{k=1}^m \lambda_k \nabla g_k(\mathbf{x}^*) = 0$, and $\lambda_k g_k(\mathbf{x}^*) = 0$, $k = 1,\dots,m$.

for some $\mathbf{y}^* = (y^*_\alpha)$, $\alpha \in \mathbb{N}^n_d$, in the dual cone $C_d(\mathbf{K})^* \subset \mathbb{R}^{s(d)}$ of $C_d(\mathbf{K})$, and some vector $\gamma = (\gamma_\alpha)$, $0 < |\alpha| < d$. By Lemma 2.5,

$$C_d(\mathbf{K})^* \,=\, \{\mathbf{y} \in \mathbb{R}^{s(d)} \,:\, \exists \mu \in M(\mathbf{K}) \text{ s.t. } y_\alpha = \int_{\mathbf{K}} \mathbf{x}^\alpha \, d\mu,\; \alpha \in \mathbb{N}^n_d\,\},$$

and so (3.2) is just (3.6) restated in terms of μ^*.

Next, the condition $\langle h^*, \mathbf{y}^*\rangle = 0$ (or equivalently, $\langle 1 - g^*, \mathbf{y}^*\rangle = 0$), reads:

$$\int_{\mathbf{K}} (1 - g^*)\, d\mu^* \,=\, 0,$$

which combined with $1 - g^* \in C_d(\mathbf{K})$ and $\mu^* \in M(\mathbf{K})$, implies that μ^* is supported on $\mathbf{K} \cap \{\mathbf{x} : g^*(\mathbf{x}) = 1\} = \mathbf{K} \cap \mathbf{G}^*_1$.

Next, let $s := \sum_{|\alpha|=d} g^*_\alpha y^*_\alpha \,(= y^*_0)$. From $\langle 1 - g^*, \mu^*\rangle = 0$, the measure $s^{-1}\mu^* =: \psi$ is a probability measure supported on $\mathbf{K} \cap \mathbf{G}^*_1$, and satisfies $\int \mathbf{x}^\alpha d\psi = s^{-1} y^*_\alpha$ for all $|\alpha| = d$.

Hence by a result of Mulholland and Rogers [22], there exists an atomic measure $\nu^* \in M(\mathbf{K})$ supported on $\mathbf{K} \cap \mathbf{G}^*_1$ such that

$$\int_{\mathbf{K}\cap\mathbf{G}^*_1} \mathbf{x}^\alpha d\nu^*(\mathbf{x}) \,=\, \int_{\mathbf{K}\cap\mathbf{G}^*_1} \mathbf{x}^\alpha d\psi(\mathbf{x}) \,=\, s^{-1}\, y^*_\alpha, \qquad \forall\, |\alpha| = d.$$

(See e.g. Anastassiou [1, Theorem 2.1.1, p. 39].) As $g^*(\mathbf{x}) = 1$ on the support of ν^*, we conclude that

$$1 \,=\, s^{-1} \sum_{|\alpha|=d} g^*_\alpha\, y^*_\alpha \,=\, \int_{\mathbf{K}\cap\mathbf{G}^*_1} g^* d\nu^* = \int_{\mathbf{K}\cap\mathbf{G}^*_1} d\nu^*,$$

i.e., ν^* is a probability measure and therefore may be chosen to be supported on at most $\binom{n+d-1}{d}$ points in $\mathbf{K} \cap \mathbf{G}^*_1$; see [1, Theorem 2.1.1]. Hence in (3.2) the measure μ^* can be substituted with the atomic measure $s\,\nu^*$ supported on at most $\binom{n+d-1}{d}$ contact points in $\mathbf{K} \cap \mathbf{G}^*_1$.

To obtain $\mu^*(\mathbf{K}) = \frac{n}{d} \int_{\mathbb{R}^n} \exp(-g^*)$, multiply both sides of (3.6)-(3.7) by h^*_α for every $\alpha \neq 0$, sum up and use $\langle h^*, \mathbf{y}^*\rangle = 0$ to obtain

$$\begin{aligned} -y^*_0 = \sum_{\alpha \neq 0} h^*_\alpha\, y^*_\alpha \;&=\; \int_{\mathbb{R}^n} (h^*(\mathbf{x}) - 1) \exp(h^*(\mathbf{x}) - 1)\, d\mathbf{x} \\ &=\; -\int_{\mathbb{R}^n} g^*(\mathbf{x}) \exp(-g^*(\mathbf{x}))\, d\mathbf{x} \\ &=\; -\frac{n}{d} \int \exp(-g^*(\mathbf{x}))\, d\mathbf{x}, \end{aligned}$$

where we have also used (2.5).

(c) Let $\mu^* := \sum_{i=1}^s \lambda_i \delta_{\mathbf{x}(i)}$ where $\delta_{\mathbf{x}(i)}$ is the Dirac measure at the point $\mathbf{x}(i) \in \mathbf{K}$, $i = 1, \ldots, s$. Next, let $y^*_\alpha := \int \mathbf{x}^\alpha d\mu^*$ for all $\alpha \in \mathbb{N}^n_d$, so that $\mathbf{y}^* \in C_d(\mathbf{K})^*$. In particular $\mathbf{y}^*$ and g^* satisfy

$$\langle 1 - g^*, \mathbf{y}^*\rangle \,=\, \int_{\mathbf{K}} (1 - g^*) d\mu^* \,=\, 0,$$

because $g^*(\mathbf{x}(i)) = 1$ for all $i = 1, \ldots, s$. In other words, the pair $(g^*, \mathbf{y}^*)$ satisfies the KKT-optimality conditions associated with the convex problem $\mathcal{P}$. But since Slater's condition holds for $\mathcal{P}$, those conditions are also sufficient for g^* to be an optimal solution of $\mathcal{P}$, the desired result $\square$

Importantly, notice that neither $\mathbf{K}$ nor $\mathbf{G}_1^*$ are required to be convex.

3.1. **On the contact points.** Theorem 3.2 states that $\mathcal{P}$ (hence $\mathbf{P}_0$) has a unique optimal solution $g^* \in \mathbf{P}[\mathbf{x}]_d$ and one may find contact points $\mathbf{x}^*(i) \in \mathbf{K} \cap \mathbf{G}_1^*$, $i = 1, \ldots, s$, with $s \leq \binom{n+d-1}{d}$, such that

$$y^*_\alpha \,=\, \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, \exp(-g^*(\mathbf{x}))\, d\mathbf{x} \,=\, \sum_{i=1}^s \lambda^*_i \, (\mathbf{x}^*(i))^\alpha, \quad |\alpha| = d, \tag{3.9}$$

for some positive weights λ^*_i. In particular, using the identity (2.5) and $\langle 1 - g^*, \mathbf{y}^* \rangle = 0$, as well as $g^*(\mathbf{x}^*(i)) = 1$ for all i,

$$y^*_0 \,=\, \sum_{|\alpha|=d} y^*_\alpha \, g^*_\alpha \,=\, \frac{n}{d} \int_{\mathbb{R}^n} \exp(-g^*(\mathbf{x}))\, d\mathbf{x} \,=\, \sum_{i=1}^s \gamma^*_i.$$

Next, recall that d is even and let $\mathbf{v}_{d/2} : \mathbb{R}^n \to \mathbb{R}^{s(d/2)}$ be the mapping

$$\mathbf{x} \mapsto \mathbf{v}_{d/2}(\mathbf{x}) = (\mathbf{x}^\alpha), \qquad |\alpha| \,=\, d,$$

i.e., the $s(d/2)$-vector of the canonical basis of monomials $(\mathbf{x}^\alpha)$, $|\alpha| = d/2$, for the homogeneous polynomials of degree d. From (3.6),

$$\int_{\mathbb{R}^n} \mathbf{v}_{d/2}\, \mathbf{v}_{d/2}^T \exp(-g^*)\, d\mathbf{x} \,=\, \sum_{i=1}^s \gamma^*_i \, \mathbf{v}_{d/2}(\mathbf{x}^*(i))\, \mathbf{v}_{d/2}(\mathbf{x}^*(i))^T. \tag{3.10}$$

Hence, when $d = 2$ and $\mathbf{K}$ is symmetric, one retrieves the characterization in John's theorem [13, Theorem 2.1], namely that if the euclidean ball $\xi_n := \{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ is the unique ellipsoid of minimum volume containing $\mathbf{K}$ then there are contact points $(u_i) \subset \xi_n \cap \mathbf{K}$ and positive weights (λ_i), such that $\sum_i \lambda_i u_i u_i^T \,=\, I_n$ (where I_n is the $n \times n$ identity matrix). Indeed in this case, $\mathbf{v}_{d/2}(\mathbf{x}) = \mathbf{x}$, $g^*(\mathbf{x}) = \|\mathbf{x}\|^2$ and $\int_{\mathbb{R}^n} \mathbf{x}\mathbf{x}^T \exp(-\|\mathbf{x}\|^2) d\mathbf{x} = c\, I_n$ for some constant c.

So (3.10) is the analogue for $d > 2$ of the contact-points property in John's theorem and we obtain the following generalization: For d even, let $\|\mathbf{x}\|_d := (\sum_{i=1}^n x_i^d)^{1/d}$ denote the d-norm with associated unit ball $\xi_n^d := \{\mathbf{x} : \|\mathbf{x}\|_d \leq 1\}$.

Corollary 3.3. *If in Theorem 3.2 the unique optimal solution $\mathbf{G}_1^*$ is the d-unit ball ξ_n^d then there are contact points $(x^*(j)) \subset \mathbf{K} \cap \xi_n^d$ and positive weights λ^*_j, $j = 1, \ldots, s$, with $s \leq \binom{n+d-1}{d}$, such that for every $|\alpha| = d$,*

$$\sum_{j=1}^s \lambda^*_j \, (\mathbf{x}^*(j))^\alpha \,=\, \begin{cases} \displaystyle\prod_{i=1}^n \int_{\mathbb{R}} t^{\alpha_i} \exp(-t^d)\, dt & \text{if } \alpha = 2\beta \\ 0 & \text{otherwise.} \end{cases}$$

4. The general case

We now consider the more general case where the set $\mathbf{G}_1$ is of the form $\{\mathbf{x} : g(\mathbf{x}-\mathbf{a}) \leq 1\} =: \mathbf{G}_1^{\mathbf{a}}$ where $\mathbf{a} \in \mathbb{R}^n$ and $g \in \mathbf{P}[\mathbf{x}]_d$.

For every $\mathbf{a} \in \mathbb{R}^n$ and $g \in \mathbb{R}[\mathbf{x}]_d$ (with coefficient vector $\mathbf{g} \in \mathbb{R}^{s(d)}$) define the polynomial $g_{\mathbf{a}} \in \mathbb{R}[\mathbf{x}]_d$ by $\mathbf{x} \mapsto g_{\mathbf{a}}(\mathbf{x}) := g(\mathbf{x}-\mathbf{a})$ and its sublevel set $\mathbf{G}_1^{\mathbf{a}} := \{\mathbf{x} : g_{\mathbf{a}}(\mathbf{x}) \leq 1\}$. The polynomial $g_{\mathbf{a}}$ can be written

$$g_{\mathbf{a}}(\mathbf{x}) \,=\, g(\mathbf{x}-\mathbf{a}) \,=\, \sum_{\alpha\in\mathbb{N}^n_d} p_\alpha(\mathbf{a},\mathbf{g})\,\mathbf{x}^\alpha, \tag{4.1}$$

where $\mathbf{g} \in \mathbb{R}^{s(d)}$ and the polynomial $p_\alpha \in \mathbb{R}[\mathbf{a},\mathbf{g}]$ is linear in $\mathbf{g}$, for every $\alpha \in \mathbb{N}^n_d$.

Consider the following generalization of $\mathbf{P}_0$:

P: *Let $\mathbf{K} \subset \mathbb{R}^n$ be a compact set not necessarily convex and $d \in \mathbb{N}$ an even integer. Find an homogeneous polynomial g of degree d and a point $\mathbf{a} \in \mathbb{R}^n$ such that the sublevel set $\mathbf{G}_{\mathbf{a}} := \{\mathbf{x} \,:\, g(\mathbf{x}-\mathbf{a}) \leq 1\}$ contains $\mathbf{K}$ and has minimum volume among all such sublevel sets with this inclusion property.*

When $d = 2$ one retrieves the general (non symmetric) Löwner-John ellipsoid problem. For $d > 2$, an even more general problem would be to find a (non homogeneous) polynomial g of degree d such that $\mathbf{K} \subset \mathbf{G} = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ and $\mathbf{G}_1$ has minimum volume among all such set $\mathbf{G}$ with this inclusion property. However when g is not homogeneous we do not have an analogue of Lemma 2.2 for the Lebesgue-volume vol$(\mathbf{G})$.

So in view of (4.3), one wishes to solve the optimization problem

$$\mathcal{P}: \quad \rho = \min_{\mathbf{a}\in\mathbb{R}^n, g\in\mathbf{P}[\mathbf{x}]_d} \{\mathrm{vol}(\mathbf{G}_1^{\mathbf{a}}) \,:\, 1 - g_{\mathbf{a}} \in C_d(\mathbf{K})\}, \tag{4.2}$$

a generalization of (3.1) where $\mathbf{a} = 0$. Let $\mathbf{K}-\mathbf{a}$ denotes the set $\{\mathbf{x}-\mathbf{a} : \mathbf{x} \in \mathbf{K}\}$, and observe that whenever $g \in \mathbf{P}[\mathbf{x}]_d$,

$$\mathrm{vol}(\mathbf{G}_1^{\mathbf{a}}) \,=\, \mathrm{vol}(\mathbf{G}_1^0) \,=\, \mathrm{vol}(\mathbf{G}_1) \,=\, \frac{1}{\Gamma(1+n/d)} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\,d\mathbf{x}. \tag{4.3}$$

Theorem 4.1. *Let $\mathbf{K} \subset \mathbb{R}^n$ be compact with nonempty interior and consider the optimization problem $\mathcal{P}$ in (4.2).*

(a) *$\mathcal{P}$ has an optimal solution $(\mathbf{a}^*, g^*) \in \mathbb{R}^n \times \mathbf{P}[\mathbf{x}]_d$.*

(b) *Let $(\mathbf{a}^*, g^*) \in \mathbb{R}^n \times \mathbf{P}[\mathbf{x}]_d$ be an optimal solution of $\mathcal{P}$. Then there exists a finite Borel measure $\mu^* \in M(\mathbf{K}-\mathbf{a}^*)$ such that*

$$\int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g^*)d\mathbf{x} \;=\; \int_{\mathbf{K}-\mathbf{a}^*} \mathbf{x}^\alpha\, d\mu^*, \qquad \forall |\alpha| = d \tag{4.4}$$

$$\int_{\mathbf{K}-\mathbf{a}^*} (1-g^*)\, d\mu^* \;=\; 0; \quad \mu^*(\mathbf{K}-\mathbf{a}^*) = \frac{n}{d}\int_{\mathbb{R}^n} \exp(-g^*)\, d\mathbf{x}. \tag{4.5}$$

In particular, μ^* is supported on the set $V := \{\mathbf{x} \in \mathbf{K} : g^*(\mathbf{x} - \mathbf{a}^*) = 1\}$ ($= \mathbf{K} \cap \mathbf{G}_1^{\mathbf{a}^*}$) and in fact, μ^* can be substituted with another measure $\nu^* \in M(\mathbf{K})$ supported on at most $\binom{n+d-1}{d}$ contact points of V.

Proof. First observe that (4.2) reads

$$\mathcal{P}: \quad \min_{\mathbf{a}\in\mathbb{R}^n} \left\{ \min_{g\in\mathbf{P}[\mathbf{x}]_d} \{\mathrm{vol}(\mathbf{G}_1^{\mathbf{a}}) \,:\, 1 - g_{\mathbf{a}} \in C_d(\mathbf{K})\} \right\}, \tag{4.6}$$

and notice that the constraint $1 - g_{\mathbf{a}} \in C(\mathbf{K})$ is the same as $1 - g \in C(\mathbf{K} - \mathbf{a})$. And so for every $\mathbf{a} \in \mathbb{R}^n$, the inner minimization problem

$$\min_{g\in\mathbf{P}[\mathbf{x}]_d} \{\mathrm{vol}(\mathbf{G}_1^{\mathbf{a}}) \,:\, 1 - g_{\mathbf{a}} \in C_d(\mathbf{K})\}$$

of (4.6) reads

$$\rho_{\mathbf{a}} \;=\; \min_{g\in\mathbf{P}[\mathbf{x}]_d} \{\mathrm{vol}(\mathbf{G}_1) \,:\, 1 - g \in C_d(\mathbf{K} - \mathbf{a})\}. \tag{4.7}$$

From Theorem 3.2 (with $\mathbf{K} - \mathbf{a}$ in lieu of $\mathbf{K}$), problem (4.7) has a unique minimizer $g^* \in \mathbf{P}[\mathbf{x}]_d$ with value $\rho_{\mathbf{a}} = \int_{\mathbb{R}^n} \exp(-g^*) d\mathbf{x} = \int_{\mathbb{R}^n} \exp(-g^*_{\mathbf{a}}) d\mathbf{x}$.

Therefore, in a minimizing sequence $(\mathbf{a}_\ell, g^*_\ell) \subset \mathbb{R}^n \times \mathbf{P}[\mathbf{x}]_d$, $\ell \in \mathbb{N}$, for problem $\mathcal{P}$ in (4.2) with

$$\rho = \lim_{\ell\to\infty} \int_{\mathbb{R}^n} \exp(-g^*_\ell) d\mathbf{x},$$

we may and will consider that for every ℓ, the homogeneous polynomial $g^*_\ell \in \mathbf{P}[\mathbf{x}]_d$) solves the inner minimization problem (4.7) with $\mathbf{a}_\ell$ fixed.

As observed in the proof of Theorem 3.2, there is $\mathbf{z} \in \mathrm{int}(C_d(\mathbf{K})^*)$ such that $\langle 1 - g^*_{\mathbf{a}_\ell}, \mathbf{z} \rangle \geq 0$ and by Corollary I.1.6 in Faraut et Korányi [11], the set $\{h \in C_d(\mathbf{K}) : \langle \mathbf{z}, h \rangle \leq z_0\}$ is compact.

Also, $\mathbf{a}_\ell$ can be chosen with $\|\mathbf{a}_\ell\| \leq M$ for all ℓ (and some M), otherwise the constraint $1 - g_{\mathbf{a}_\ell} \in C_d(\mathbf{K})$ would impose a much too large volume $\mathrm{vol}(\mathbf{G}_1^{\mathbf{a}_\ell})$.

Therefore, there is a subsequence (ℓ_k), $k \in \mathbb{N}$, and a point $(\mathbf{a}^*, \theta^*) \in \mathbb{R}^n \times C_d(\mathbf{K})$ such that

$$\lim_{k\to\infty} \mathbf{a}_{\ell_k} \;=\; \mathbf{a}^*; \qquad \lim_{k\to\infty} (g^*_{\mathbf{a}_{\ell_k}})_\alpha \;=\; \theta^*_\alpha, \quad \forall \alpha \in \mathbb{N}^n_d.$$

Recall the definition (4.1) of $g^*_{\mathbf{a}^*_\ell}(\mathbf{x}) = g^*_\ell(\mathbf{x} - \mathbf{a}^*_\ell)$ for some homogeneous polynomial $g^*_\ell \in \mathbf{P}[\mathbf{x}]_d$ with coefficient vector $\mathbf{g}^*_\ell$, i.e.,

$$(g^*_{\mathbf{a}^*_\ell})_\alpha \;=\; p_\alpha(\mathbf{a}^*_\ell, \mathbf{g}^*_\ell), \qquad \forall \alpha \in \mathbb{N}^n_d,$$

for some polynomials $(p_\alpha) \subset \mathbb{R}[\mathbf{a}, \mathbf{g}]$, $\alpha \in \mathbb{N}^n_d$. In particular, for every $\alpha \in \mathbb{N}^n_d$ with $|\alpha| = d$, $p_\alpha(\mathbf{a}^*_\ell, \mathbf{g}^*_\ell) = (g^*_\ell)_\alpha$. And so for every $\alpha \in \mathbb{N}^n_d$ with $|\alpha| = d$,

$$\theta^*_\alpha \;=\; \lim_{k\to\infty} \;=\; (g^*_{\ell_k})_\alpha.$$

If we define the homogeneous polynomial g^* of degree d by $(g^*)_\alpha = \theta^*_\alpha$ for every $\alpha \in \mathbb{N}^n_d$ with $|\alpha| = d$, then

$$\lim_{k\to\infty} (g^*_{\mathbf{a}^*_{\ell_k}})_\alpha \,=\, \lim_{k\to\infty} p_\alpha(\mathbf{a}^*_{\ell_k}, \mathbf{g}^*_{\ell_k}), =\, p_\alpha(\mathbf{a}^*, \mathbf{g}^*), \quad \forall \alpha \in \mathbb{N}^n_d.$$

This means that for every $\alpha \in \mathbb{N}^n_d$

$$\theta^*(\mathbf{x}) \,=\, g^*(\mathbf{x} - \mathbf{a}^*), \quad \mathbf{x} \in \mathbb{R}^n.$$

In addition, one has the pointwise convergence $g^*_\ell(\mathbf{x}) \to g^*(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Therefore, by Fatou's Lemma (see e.g. Ash [2]),

$$\rho \,=\, \lim_{\ell\to\infty} \int_{\mathbb{R}^n} \exp(-g^*_\ell)\, d\mathbf{x} \,\geq\, \int_{\mathbb{R}^n} \liminf_{\ell\to\infty} \exp(-g^*_\ell)\, d\mathbf{x} \,=\, \int_{\mathbb{R}^n} \exp(-g^*)\, d\mathbf{x},$$

which proves that $(\mathbf{a}^*, g^*)$ is an optimal solution of (4.2).

In addition $g^* \in \mathbf{P}[\mathbf{x}]_d$ is an optimal solution of the inner minimization problem in (4.7) with $\mathbf{a} := \mathbf{a}^*$. Otherwise an optimal solution $h \in \mathbf{P}[\mathbf{x}]_d$ of (4.7) with $\mathbf{a} = \mathbf{a}^*$ would yield a solution $(\mathbf{a}^*, h)$ with associated cost $\int_{\mathbb{R}^n} \exp(-h)$ strictly smaller than ρ, a contradiction.

Hence by Theorem 3.2 (applied to problem (4.7)) there is a finite Borel measure $\mu^* \in M(\mathbf{K} - \mathbf{a}^*)$ such that

$$\begin{aligned} \int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g^*) d\mathbf{x} \;&=\; \int_{\mathbf{K}-\mathbf{a}^*} \mathbf{x}^\alpha\, d\mu, \qquad \forall |\alpha| = d \\ \int_{\mathbf{K}-\mathbf{a}^*} (1 - g^*)\, d\mu^* \;&=\; 0; \quad \mu(\mathbf{K} - \mathbf{a}^*) = \frac{n}{d} \int_{\mathbb{R}^n} \exp(-g^*)\, d\mathbf{x}. \end{aligned}$$

And so μ^* is supported on the set

$$V_{\mathbf{a}^*} \,=\, \{\mathbf{x} \in \mathbf{K} - \mathbf{a}^* \,:\, g^*(\mathbf{x}) = 1\} \,=\, \{\mathbf{x} \in \mathbf{K} \,:\, g^*(\mathbf{x} - \mathbf{a}^*) = 1\}.$$

Invoking again [1, Theorem 2.1.1, p. 39], there exists an atomic measure $\nu^* \in M(\mathbf{K} - \mathbf{a}^*)$ supported on $\mathbf{K} - \mathbf{a}^* \cap \mathbf{G}^*_1$ such □

5. A computational procedure

Even though $\mathcal{P}$ in (3.1) is a finite-dimensional convex optimization problem, it is hard to solve for mainly two reasons:

- From Theorem 2.4, the gradient and Hessian of the (strictly) convex objective function $g \mapsto \int \exp(-g)$ requires evaluating integrals of the form

$$\int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g(\mathbf{x}))\, d\mathbf{x}, \quad \forall \alpha \in \mathbb{N}^n_d,$$

  a difficult and challenging problem. (And with $\alpha = 0$ one obtains the value of the objective function.)
- The convex cone $C_d(\mathbf{K})$ has no *exact* and *tractable* representation to efficiently handle the constraint $1 - g \in C_d(\mathbf{K})$ in an algorithm for solving problem (3.1).

However, below we outline a numerical scheme to approximate to any desired ϵ-accuracy (with $\epsilon > 0$):

- the optimal value ρ of (3.1),
- the unique optimal solution $g^* \in \mathbf{P}[\mathbf{x}]_d$ of $\mathcal{P}$ obtained in Theorem 3.2.

5.1. **Concerning gradient and Hessian evaluation.** To approximate the gradient and Hessian of the objective function we will use the following result:

Lemma 5.1. *Let $g \in \mathbf{P}[\mathbf{x}]_d$ and let $\mathbf{G}_1 = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$. Then for every $\alpha \in \mathbb{N}^n$*

$$\int_{\mathbb{R}^n} \mathbf{x}^\alpha \exp(-g)\, d\mathbf{x} \,=\, \Gamma\left(1 + \frac{n + |\alpha|}{d}\right) \int_{\mathbf{G}_1} \mathbf{x}^\alpha \, d\mathbf{x}. \tag{5.1}$$

The proof being identical to that of Lemma 2.2 is omitted. So Lemma 5.1 relates in a very simple and explicit manner all moments of the Borel measure with density $\exp(-g)$ on $\mathbb{R}^n$ with those of the Lebesgue measure on the sublevel set $\mathbf{G}_1$.

It turns out that in Henrion et al. [14] we have provided a hierarchy of semidefinite programs[3] a convex optimization problem to approximate as closely as desired, any finite moment sequence (z_α), $\alpha \in \mathbb{N}^n_\ell$, defined by

$$z_\alpha \,=\, \int_\Omega \mathbf{x}^\alpha \, d\mathbf{x}, \quad \alpha \in \mathbb{N}^n_\ell.$$

where Ω is a compact basic semi-algebraic set of the form $\{\mathbf{x} : \theta_j(\mathbf{x}) \geq 0, j = 1, \ldots, m\}$ for some polynomials $(\theta_j) \subset \mathbb{R}[\mathbf{x}]$. For more details the interested reader is referred to [14].

Hence in any minimization algorithm for solving $\mathcal{P}$, and given a current iterate $g \in \mathbf{P}[\mathbf{x}]_d$, one may approximate as closely as desired the value at g of the objective function as well as its gradient and Hessian by using the methodology described in [14, 19].

5.2. **Concerning the convex cone $C_d(\mathbf{K})$.** We here assume that the compact (and non necessarily convex) set $\mathbf{K} \subset \mathbb{R}^n$ is a basic semi-algebraic set defined by

$$\mathbf{K} \,=\, \{\mathbf{x} \in \mathbb{R}^n_+ \,:\, w_j(\mathbf{x}) \geq 0,\, j = 1, \ldots, s\}, \tag{5.2}$$

for some given polynomials $(w_j) \subset \mathbb{R}[\mathbf{x}]$. Denote by $\Sigma_k \subset \mathbb{R}[\mathbf{x}]_{2k}$ the convex cone of SOS (sum of squares) polynomials of degree at most $2k$, and let w_0 be the constant polynomial equal to 1, and $v_j := \lceil \deg(w_j)/2 \rceil$, $j = 0, \ldots, s$.

[3]A semidefinite program is a finite-dimensional convex optimization problem which in canonical form reads: $\min_{\mathbf{x}} \{\mathbf{c}^T\mathbf{x} : \mathbf{A}_0 + \sum_{k=1}^t \mathbf{A}_k x_k \succeq 0\}$, where $\mathbf{c} \in \mathbb{R}^t$, the $\mathbf{A}_k$'s are real symmetric matrices, and the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ is positive semidefinite. Importantly, up to arbitrary fixed precision it can be solved in time polynomial in the input size of the problem.

With k fixed, arbitrary, we now replace the condition $1-g\in C_d(\mathbf{K})$ with the stronger condition $1-g\in\mathcal{C}_k$ $(\subset C_d(\mathbf{K}))$ where

$$\mathcal{C}_k \,=\, \left\{\sum_{j=0}^s \sigma_j\, w_j \;:\; \sigma_j\in\Sigma_{k-v_j},\, j=0,1,\ldots,s\right\}. \tag{5.3}$$

It turns out that membership in $\mathcal{C}_k$ translates into Linear Matrix Inequalities[4] (LMIs) on the coefficients of the polynomials g and σ_j's; see e.g. [19]. If $\mathbf{K}$ has nonempty interior then the convex cone $\mathcal{C}_k$ is closed.

Assumption 1 (Archimedean assumption)**.** *There exist $M>0$ and $k\in\mathbb{N}$ such that the quadratic polynomial $\mathbf{x}\mapsto M-\|\mathbf{x}\|^2$ belongs to $\mathcal{C}_k$.*

Notice that Assumption 1 is not restrictive. Indeed, $\mathbf{K}$ being compact, if one knows an explicit value $\mathbf{M}>0$ such that $\mathbf{K}\subset\{\mathbf{x}:\|\mathbf{x}\|<M\}$, then its suffices to add to the definition of $\mathbf{K}$ the redundant quadratic constraint $w_{s+1}(\mathbf{x})\geq 0$, where $w_{s+1}(\mathbf{x}):=M^2-\|\mathbf{x}\|^2$.

Under Assumption 1, $C_d(\mathbf{K})=\overline{\bigcup_{k=0}^{\infty}\mathcal{C}_k}$, that is, the family of convex cones $(\mathcal{C}_k)$, $k\in\mathbb{N}$, provide a converging sequence of (nested) *inner approximations* of the larger convex cone $C_d(\mathbf{K})$.

5.3. **A numerical scheme.** In view of the above it is natural to consider the following hierarchy of convex optimization problems $(\mathcal{P}_k)$, $k\in\mathbb{N}$, where for each fixed k:

$$\begin{array}{rll} \rho_k=\displaystyle\min_{g,\sigma_j} & \displaystyle\int_{\mathbb{R}^n}\exp(-g)\,d\mathbf{x} & \\ \text{s.t.} & 1-g=\displaystyle\sum_{j=0}^s\sigma_j w_j & \\ & g_\alpha=0,\quad\forall|\alpha|<d & \\ & g\in\mathbb{R}[\mathbf{x}]_d;\;\sigma_j\in\Sigma_{k-v_j},\;j=0,\ldots,s. & \end{array} \tag{5.4}$$

Of course the sequence (ρ_k), $k\in\mathbb{N}$, is monotone non increasing and $\rho_k\geq\rho$ for all k. Moreover, for each fixed $k\in\mathbb{N}$, $\mathcal{P}_k$ is a convex optimization problem which consists of minimizing a strictly convex function under LMI constraints.

From Corollary 2.3, $\int_{\mathbb{R}^n}\exp(-g)d\mathbf{x}<\infty$ if and only if $g\in\mathbf{P}[\mathbf{x}]_d$ and so the objective function also acts as a barrier for the convex cone $\mathbf{P}[\mathbf{x}]_d$. Therefore, to solve $\mathcal{P}_k$ one may use first-order or second-order (local minimization) algorithms, starting from an initial guess $g_0\in\mathbf{P}[\mathbf{x}]_d$. At any

[4] A Linear Matrix Inequality (LMI) is a constraint of the form $\mathbf{A}_0+\sum_{\ell=1}^t\mathbf{A}_\ell x_\ell\succeq 0$ where each $\mathbf{A}_\ell$, $\ell=0,\ldots,t$, is a real symmetric matrix and the notation $\mathbf{A}\succeq 0$ stands for $\mathbf{A}$ is positive semidefinite. Equivalently $\mathbf{A}(\mathbf{x})\succeq 0$ where $\mathbf{A}(\mathbf{x}):=\mathbf{A}_0+\sum_{\ell=1}^t\mathbf{A}_\ell x_\ell$, a real symmetric matrix whose each entry is affine in $\mathbf{x}\in\mathbb{R}^t$. An LMI always define a convex set, i.e., the set $\{\mathbf{x}\in\mathbb{R}^t:\mathbf{A}(\mathbf{x})\succeq\mathbf{A}_0\}$ is convex.

current iterate $g \in \mathbf{P}[\mathbf{x}]_d$ of such an algorithm one may use the methodology described in [14] to approximate the objective function $\int \exp(-g)$ as well as its gradient and Hessian.

Theorem 5.2. *Let* $\mathbf{K}$ *in (5.2) be compact with nonempty interior and let Assumption 1 hold. Then there exists* k_0 *such that for every* $k \geq k_0$*, problem* $\mathcal{P}_k$ *in (5.4) has a unique optimal solution* $g_k^* \in \mathbf{P}[\mathbf{x}]_d$*.*

Proof. Firstly, $\mathcal{P}_k$ has a feasible solution for sufficiently large k. Indeed consider the polynomial $\mathbf{x} \mapsto g_0(\mathbf{x}) = \sum_{i=1}^n x_i^d$ which belongs to $\mathbf{P}[\mathbf{x}]_d$. Then as $\mathbf{K}$ is compact, $M - g_0 > 0$ on $\mathbf{K}$ for some M and so by Putinar's Positivstellensatz [26], $1 - g_0/M \in \mathcal{C}_k$ for some k_0 (and hence for all $k \geq k_0$). Hence g_0/M is a feasible solution for $\mathcal{P}_k$ for all $k \geq k_0$. Of course, as $\mathcal{C}_k \subset C_d(\mathbf{K})$, every feasible solution $g \in \mathbf{P}[\mathbf{x}]_d$ satisfies $0 \leq g \leq 1$ on $\mathbf{K}$. So proceeding as in the proof of Theorem 3.2 and using the fact that $\mathcal{C}_k$ is closed, the set

$$\{g \in \mathbf{P}[\mathbf{x}]_d \cap \mathcal{C}_k : \int_{\mathbb{R}^n} \exp(-g)d\mathbf{x} \leq \int_{\mathbb{R}^n} \exp(-\frac{g_0}{M})d\mathbf{x}\},$$

is compact. And as the objective function is strictly convex, the optimal solution $g_k^* \in \mathbf{P}[\mathbf{x}]_d \cap \mathcal{C}_k$ is unique (but the representation of $1 - g_k^*$ in (5.4) is not unique in general). □

Remark 5.3. *If desired one may also impose* g *to be convex by simply requiring* $\mathbf{z}^T \nabla^2 g(\mathbf{x}) \mathbf{z} \geq 0$ *for all* $(\mathbf{x}, \mathbf{z})$*.*

We now consider the asymptotic behavior of the solution of (5.4) as $k \to \infty$.

Theorem 5.4. *Let* $\mathbf{K}$ *in (5.2) be compact with nonempty interior and let Assumption 1 hold. If* ρ *(resp.* ρ_k*) is the optimal value of* $\mathcal{P}$ *(resp.* $\mathcal{P}_k$*) then* $\rho = \lim_{k\to\infty} \rho_k$*. Moreover, for every* $k \geq k_0$*, let* $g_k^* \in \mathbf{P}[\mathbf{x}]_d$ *be the unique optimal solution of* $\mathcal{P}_k$*. Then as* $k \to \infty$*,* $g_k^* \to g^*$ *where* g^* *is the unique optimal solution of* $\mathcal{P}$*.*

Proof. By Theorem 3.2, $\mathcal{P}$ has a unique optimal solution $g^* \in \mathbf{P}[\mathbf{x}]_d$. Let $\epsilon > 0$ be fixed, arbitrary. As $1 - g^* \in C_d(\mathbf{K})$, the polynomial $1 - g/(1+\epsilon)$ is strictly positive on $\mathbf{K}$, and so by Putinar's Positivstellensatz [26], $1 - g^*/(1+\epsilon)$ belongs to $\mathcal{C}_k$ for all $k \geq k_\epsilon$ for some integer k_ϵ. Hence the polynomial $g^*/(1+\epsilon) \in \mathbf{P}[\mathbf{x}]_d$ is a feasible solution of $\mathcal{P}_k$ for all $k \geq k_\epsilon$. Moreover, by homogeneity,

$$\begin{aligned}\int_{\mathbb{R}^n} \exp(-\frac{g^*}{1+\epsilon})\,d\mathbf{x} &= (1+\epsilon)^{n/d} \int_{\mathbb{R}^n} \exp(-g^*)\,d\mathbf{x} \\ &= (1+\epsilon)^{n/d} \rho.\end{aligned}$$

This shows that $\rho_k \leq (1+\epsilon)^{n/d}\rho$ for all $k \geq k_\epsilon$. Combining this with $\rho_k \geq \rho$ and the fact that $\epsilon > 0$ was arbitrary, yields the convergence $\rho_k \to \rho$ as $k \to \infty$.

Next, let $\mathbf{y} \in \mathrm{int}(C_d(\mathbf{K})^*)$ be as in the proof of Theorem 3.2. From $1 - g_k^* \in \mathcal{C}_k$ we also obtain $\langle \mathbf{y}, 1 - g_k^* \rangle \geq 0$, i.e.,

$$y_0 \geq \langle \mathbf{y}, g_k^* \rangle, \quad \forall k \geq k_0,$$

Recall that the set $\{g \in C_d(\mathbf{K}) : \langle \mathbf{y}, g \rangle \leq y_0\}$ is compact. Therefore there exists a subsequence (k_ℓ), $\ell \in \mathbb{N}$, and $\tilde{g} \in C_d(\mathbf{K})$ such that $g_{k_\ell}^* \to \tilde{g}$ as $\ell \to \infty$. In particular, $1 - \tilde{g} \in C_d(\mathbf{K})$ and $\tilde{g}_\alpha = 0$ whenever $|\alpha| < d$ (i.e., $\tilde{g}$ is homogeneous of degree d). Moreover, one also has the pointwise convergence $\lim_{\ell\to\infty} g_{k_\ell}^*(\mathbf{x}) = \tilde{g}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Hence by Fatou's lemma,

$$\begin{aligned} \rho \,=\, \lim_{\ell\to\infty} \rho_{k_\ell} &= \lim_{\ell\to\infty} \int_{\mathbb{R}^n} \exp(-g_{k_\ell}^*(\mathbf{x}))d\mathbf{x} \\ &\geq \int_{\mathbb{R}^n} \liminf_{\ell\to\infty} \exp(-g_{k_\ell}^*(\mathbf{x}))d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \exp(-\tilde{g}(\mathbf{x}))d\mathbf{x} \,=\, \rho, \end{aligned}$$

which proves that $\tilde{g}$ is an optimal solution of $\mathcal{P}$, and by uniqueness of the optimal solution, $\tilde{g} = g^*$. As (g_{k_ℓ}), $\ell \in \mathbb{N}$, was an arbitrary converging subsequence, the whole sequence (g_k^*) converges to g^*. □

6. Conclusion

We have considered non convex generalizations $\mathbf{P}_0$ and $\mathbf{P}$ of the Löwner-John ellipsoid problem where we now look for an homogeneous polynomial g of (even) degree $d > 2$. Importantly, neither $\mathbf{K}$ not the sublevel set $\mathbf{G}$ associated with g are required to be convex. However both $\mathbf{P}_0$ and $\mathbf{P}$ have an optimal solution (unique for $\mathbf{P}_0$) and a characterization in terms of contact points in $\mathbf{K} \cap \mathbf{G}$ is also obtained as in the Löwner-John ellipsoid theorem. Crucial is the fact that the Lebesgue volume $\mathbf{G}$ is a strictly convex function of the coefficients of g. This latter fact also permits to define a hierarchy of convex optimization problems to approximate as closely as desired the optimal solution of $\mathbf{P}_0$.

References

[1] G.A. Anastassiou; *Moments in Probability and Approximation Theory*, Longman Scientific & Technical, UK, 1993.

[2] R.B. Ash. *Real Analysis and Probability*, Academic Press Inc., Boston, 1972.

[3] K. Ball. Ellipsoids of maximal volume in convex bodies, *Geom. Dedicata* **41** (1992), pp. 241–250.

[4] K. Ball. Convex geometry and functional analysis, In *Handbook of the Geometry of Banach Spaces I*, W.B. Johnson and J. Lindenstrauss (Eds.), North Holland, Amsterdam 2001, pp. 161–194.

[5] J. Bastero and M. Romance. John's decomposition of the identity in the non-convex case, *Positivity* **6** (2002), pp. 1–16.

[6] F.L. Bookstein. Fitting conic sections to scattered data, *Comp. Graph. Image. Process.* **9** (1979), pp. 56–71.

[7] G. Calafiore. Approximation of n-dimnsional data using spherical and ellipsoidal primitives, *IEEE Trans. Syst. Man. Cyb.* **32** (2002), pp. 269–2768.

[8] F.L. Chernousko. Guaranteed estimates of undetermined quantities by means of ellipsoids, *Sov. Math. Dodkl.* **21** (1980), pp. 396–399.

[9] C. Croux, G. Haesbroeck and P.J. Rousseeuw. Location adjustment for the minimum volume ellipsoid estimator, *Stat. Comput.* **12** (2002), pp. 191–200.

[10] A. Giannopoulos, I. Perissinaki and A. Tsolomitis. A. John's theorem for an arbitrary pair of convex bodies. *Geom. Dedicata* **84** (2001), pp. 6379.

[11] J. Faraut and A. Korányi. *Analysis on Symmetric Cones*, Clarendon Press, Oxford, 1994.

[12] W. Gander, G.H. Golub and R. Strebel. Least-squares fitting of circles and ellipses, *BIT* **34** (1994), pp. 558–578.

[13] M. Henk. Löwner-John ellipsoids, *Documenta Math.*, to appear.

[14] D. Henrion, J.B. Lasserre and C. Savorgnan. Approximate volume and integration of basic semi-algebraic sets, SIAM Review **51** (2009), pp. 722–743.

[15] D. Henrion, D. Peaucelle, D. Arzelier and M.Sebek. Ellipsoidal approximation of the stability domain of a polynomial, *IEEE Trans. Aut. Control* **48** (2003), pp. 2255–2259.

[16] D. Henrion and J.B. Lasserre. Inner approximations for polynomial matrix inequalities and robust stability regions, *IEEE Trans. Aut. Control* **57** (2012), pp. 1456–1467.

[17] D. Henrion, M. Sebek and V. Kucera. Positive polynomials and and robust stabilization with fixed-order controllers, *IEEE Trans. Aut. Control* **48** (2003), pp. 1178–1186.

[18] A. Karimi, H. Khatibi and R. Longchamp. Robust control of polytopic systems by convex optimization, *Automatica* **43** (2007), pp. 1395–1402.

[19] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College, London, 2009.

[20] A. Morosov and S. Shakirov. New and old results in resultant theory, `arXiv:0911.5278v1`, 2009.

[21] A. Morosov and S. Shakirov. Introduction to integral discriminants, J. High Energy Phys. **12** (2009), `arXiv:0911.5278v1`, 2009.

[22] H.P. Mulholland and C.A. Rogers, Representation theorems for distribution functions, Proc. London Math. Soc. **8** (1958), pp. 177–223.

[23] U. Nurges. Robust pole assignment via reflection coefficientsof polynomials, *Automatica* **42** (2006), pp. 1223–1230.

[24] J. O'Rourke and N.I. Badler. Decomposition of three-dimensional objets into spheres, *IEEE Trans. Pattern Anal. Machine Intell.* **1** (1979), pp. 295–305.

[25] V. Pratt. Direct least squares fittingof algebraic surfaces, *ACM J. Comp. Graph.* **21** (1987).

[26] M. Putinar. Positive polynomials on compact semi-algebraic sets, Indiana Univ. Math. J. **42** (1993), pp. 969–984

[27] R.T. Rockafellar. *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[28] J.B. Rosen. Pattern separation by convex programming techniques, *J. Math. Anal. Appl.* **10** (1965), pp. 123–1324.

[29] P.L. Rosin. A note on the least squares fitting of ellipses, *Pattern Recog. Letters* **14** (1993), pp. 799–808.

[30] P.L. Rosin and G.A. West. Nonparametric segmentation of curves into various representations, *IEEE Trans. Pattern Anal. Machine Intell.* **17** (1995), pp. 1140–1153.

[31] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*, John Wiley, New York (1987).

[32] P. Sun and R. Freund. Computation of minimum-volume covering ellipsoids, *Oper. Res.* **52** (2004), pp. 690–706.
[33] G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to to edge and range image segmentation, *IEEE Trans. Pattern Anal. Machine Intell.* **13** (1991), pp. 1115–1138.
[34] L. Vandenberghe and S. Boyd. Semidefinite programming, *SIAM Rev.* **38** (1996), pp. 49–95.

LAAS-CNRS and Institute of Mathematics, University of Toulouse, 7 Avenue du Colonel Roche, 310777 Toulouse Cedex 4, France.
E-mail address: lasserre@laas.fr