

Minimum-Volume Ellipsoids: Applications, Duality, and Algorithms

Mike Todd

July 6, 2007

School of Operations Research and Information Engineering,
Cornell University
<http://legacy.orie.cornell.edu/~miketodd/todd.html>

ICIAM, Zurich, July 2007

1. Löwner-John Ellipsoid

In one of the first papers on modern inequality-constrained nonlinear optimization, **F. John** (1948) considered the **minimum-volume ellipsoid** $E_*(S)$ containing a compact $S \subset \mathbb{R}^n$. We can assume S is convex and full-dimensional.

- Derived necessary conditions, showed them sufficient, but didn't realize the significance of convexity;
- Showed $\exists T \subseteq S$, $|T| \leq n(n+3)/2$, with $E_*(T) = E_*(S)$ (a **core set**); and
- Showed that

$$\frac{1}{n}E_*(S) \subseteq S \subseteq E_*(S).$$

2. LogDet

Let \mathcal{S}^n denote the space of symmetric $n \times n$ real matrices, of dimension $n(n+1)/2$, with \mathcal{S}_+^n and \mathcal{S}_{++}^n its cones of positive semidefinite and positive definite matrices. (We write $X \succeq 0$ or $X \succ 0$.)

Define $F : \mathcal{S}^n \rightarrow \mathbb{R}$ by

$$F(X) := -\ln \det X$$

if X is positive definite, $+\infty$ otherwise.

Note: if $f(x) := -\sum \ln x_j$, then $F = f \circ \lambda$, with $\lambda(X)$ the vector of eigenvalues of X : this is a **spectral function** as studied by A. Lewis.

$$DF(X)[H] = -X^{-1} \bullet H := -\text{Trace} (X^{-1}H),$$

$$D^2F(X)[H, H] = X^{-1}HX^{-1} \bullet H.$$

Hence F is **convex**.

Outline

- Applications
- Geometry and dual problems
- Duality
- Löwner-John ellipsoids again
- The ellipsoid method
- Semidefinite programming
- Algorithms

3. Data Analysis and Computational Geometry

Suppose we are given a finite set S of points in \mathbb{R}^n .

a) **Detecting outliers:**

choose points far from the center of a minimum volume enclosing ellipsoid.

b) **Testing the worth of a cluster $T \subseteq S$:**

measure by the volume of $E_*(T)$.

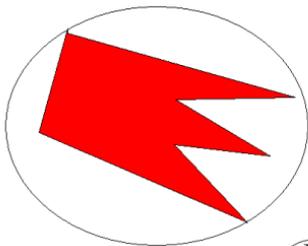
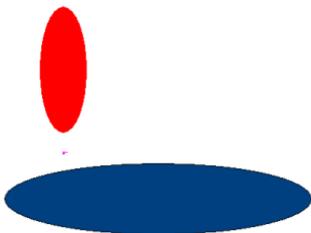
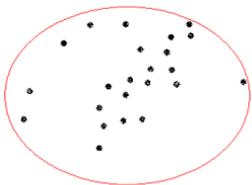
c) **Finding a small representative subset $T \subseteq S$:**

use a core set of $E_*(S)$.

In all cases, desirable linear invariance properties.

Sufficient condition for moving bodies S and T in \mathbb{R}^n not to hit:

$$E_*(S) \cap E_*(T) = \emptyset.$$



4. D-Optimal Design in Statistics

We want to estimate the parameters $\theta \in \mathbb{R}^n$ in a linear model $Y = x^T \theta + \epsilon$, with $x \in \mathbb{R}^n$ a vector of independent variables, drawn from the design space S , compact and spanning \mathbb{R}^n .

If we can choose the design points x_1, \dots, x_m at which to make observations y , we get the (best linear unbiased) estimate of θ :

$$\hat{\theta} := (XX^T)^{-1}Xy,$$

where $X := [x_1, \dots, x_m] \in \mathbb{R}^{n \times m}$. We have

$$E(\hat{\theta}) = \theta, \quad \text{covar}(\hat{\theta}) = (XX^T)^{-1}.$$

D-optimal: choose design measure ξ on S to maximize $\ln \det M(\xi)$, where $M(\xi)$ is the Fisher information matrix $\int xx^T d\xi$.

5. Geometry

The set

$$E(\bar{x}, H) := \{x \in \mathbb{R}^n : (x - \bar{x})^T H (x - \bar{x}) \leq n\}$$

for $\bar{x} \in \mathbb{R}^n$, $H \in \mathcal{S}_{++}^n$, is an ellipsoid in \mathbb{R}^n , with center \bar{x} and shape defined by H . If we write $H^{-1} = LL^T$ and $z = L^{-1}(x - \bar{x})$, we see that

$$E(\bar{x}, H) = \{x = \bar{x} + Lz : \|z\| \leq \sqrt{n}\},$$

so

$$\text{vol}(E(\bar{x}, H)) = |\det L| \cdot (\sqrt{n})^n \cdot \text{vol}(B_n) = \text{const}(n) / \sqrt{\det H},$$

and minimizing the volume of $E(\bar{x}, H)$ is equivalent to minimizing $-\ln \det H$.

6. Dual Problems

So ξ is D-optimal iff ξ **maximizes** the volume of $E(0, M(\xi)^{-1})$.
Silvey (1972) asked: is this related to **minimizing** the volume of a central ellipsoid containing S ?

Sibson (1972): **Yes**, using convex duality!

$$(P) \quad \min_{H \succ 0} \quad -\ln \det H \\ x^T H x \leq n \text{ for all } x \in S.$$

$$(D) \quad \max \quad \ln \det M(\xi) \\ \xi \text{ a probability measure on } S.$$

7. Weak Duality

$$\begin{array}{ll} \min_{H \succ 0} & -\ln \det H \\ (P) & x^T H x \leq n, \forall x \in S \end{array} \quad \begin{array}{ll} \max & \ln \det M(\xi) \\ (D) & \xi \text{ prob. meas. on } S. \end{array}$$

Suppose H and ξ are feasible for (P) and (D) respectively.

Then

$$n \geq \int_S x^T H x d\xi = \int_S (x x^T \bullet H) d\xi = M(\xi) \bullet H.$$

So,

$$\begin{aligned} -\ln \det H - \ln \det M(\xi) &= -\ln \det M(\xi) H \\ &= -n \ln (\Pi_i \lambda_i(M(\xi) H))^{1/n} \\ &\geq -n \ln \frac{\sum_i \lambda_i(M(\xi) H)}{n} \\ &= -n \ln \frac{M(\xi) \bullet H}{n} \geq 0. \end{aligned}$$

8. Consequences

Equality holds iff

- ξ is supported on $\{x \in S : x^T H x = n\}$, and
- $M(\xi)H$ is (a multiple of) the identity matrix.

So, assuming strong duality, we find that if ξ is D-optimal, then $M(\xi)^{-1}$ is optimal for (P).

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

$$\begin{aligned} \max\{|v^T x| : x \in \frac{1}{\sqrt{n}}E(0, H)\} &= \sqrt{v^T M(\xi)v} \\ &= \sqrt{\int_S (v^T x)^2 d\xi} \leq \max_{x \in S} |v^T x| \end{aligned}$$

and hence $\frac{1}{\sqrt{n}}E(0, H) \subseteq \text{conv}\{S \cup (-S)\} \subseteq E(0, H)$; so we have the well-known Löwner-John property of the minimum-volume (central) ellipsoid enclosing a centrally symmetric set.

9. Löwner-John Ellipsoids, Again

Problem (P) seeks the smallest **central** ellipsoid circumscribing S . What if, as did Löwner and John, we seek the smallest ellipsoid **with arbitrary center**? Unfortunately, the set of (\bar{x}, H) with $(x - \bar{x})^T H (x - \bar{x}) \leq n$ for all $x \in S$ is **not convex**. But we can proceed as follows.

Let $S' := \{(1; x) : x \in S\} \subseteq \mathbb{R}^{1+n}$, and let E' be the minimum-volume central ellipsoid containing S' . Then

$$E := \{x \in \mathbb{R}^n : (1; x) \in E'\}$$

is the minimum-volume ellipsoid containing S . So, at the expense of increasing the dimension by 1, we can reduce the general case to the central one. This also allows one to prove the **general Löwner-John property** that $\frac{1}{n}E_*(S) \subseteq S \subseteq E_*(S)$.

10. The Ellipsoid Method

In 1976, Yudin and Nemirovski introduced the ellipsoid method for convex programming (also independently Shor), and in 1979, Khachiyan famously used it to show $LP \in P$. The key step at each iteration was to replace the current ellipsoid $E = E(\bar{x}, H) \subseteq \mathbb{R}^n$ by the minimum-volume ellipsoid E_+ containing

$$E_{01} := \{x \in E : a^T x \geq a^T \bar{x}\}$$

or more generally

$$E_{\alpha\beta} := \{x \in E : a^T x - a^T \bar{x} \in [\alpha, \beta]\},$$

where $-1 \leq \alpha \leq \beta \leq 1$ and $a \in \mathbb{R}^n$ is normalized so that $a^T H^{-1} a = 1/n$.

The Ellipsoid Method, 2

Obtaining a plausible such ellipsoid E_+ is not too hard, but proving it has minimum volume was messy and ad hoc. Using the general method above with $S = E_{\alpha\beta}$ gives a much cleaner approach. The optimal measure ξ puts some of the measure uniformly distributed on

$$E_\alpha := \{x : (x - \bar{x})^T H (x - \bar{x}) = n, a^T x = a^T \bar{x} + \alpha\}$$

and the rest uniformly distributed on E_β .

We can similarly derive the minimum-volume ellipsoid containing the intersection of two balls.

11. Semidefinite Programming

The semidefinite programming problem is

$$\min\{C \bullet X : A_i \bullet X = b_i, i = 1, \dots, m, X \succeq 0\},$$

where C and all the A_i s lie in \mathcal{S}^n . Interior-point methods can be applied to this problem by using the barrier function $-\ln \det$ on \mathcal{S}_{++}^n , as shown by Nesterov-Nemirovski and Alizadeh.

Thus minimum-volume ellipsoid problems are related to analytic center problems for SDP.

Boyd, Vandenberghe and Wu (1998), describing an interior-point method for the **max det** problem, also surveyed its wide ranging applications; also Sun and Freund (2004).

12. Algorithms

Khachiyan (1996) developed/analyzed an algorithm to compute a $(1 + \epsilon)n$ -rounding of the convex hull S of m points x_1, \dots, x_m in \mathbb{R}^n , i.e.,

$$\frac{1}{(1 + \epsilon)n}E \subseteq S \subseteq E,$$

in $O(mn^2(\epsilon^{-1} + \ln n + \ln \ln m))$ arithmetic operations, and hence an η -approximate Löwner-John ellipsoid, i.e.,

$$S \subseteq E, \quad \text{vol}(E) \leq (1 + \eta)\text{vol}(E_*(S)),$$

in $O(mn^2(n\eta^{-1} + \ln n + \ln \ln m))$ arithmetic operations.

The latter contrasts with a variant of the ellipsoid method that needs $O((mn^6 + n^8) \ln(nR/\eta r))$ operations, and an interior-point method that needs $O(m^{3.5} \ln(mR/\eta r))$ operations.

Algorithms, 2

In fact, Khachiyan's method coincides with that of Fedorov (1972) and is closely related to that of Wynn (1970) for the D-optimal design problem. It also coincides with a central shallow-cut ellipsoid method for the polar of S , and is the Frank-Wolfe (1956) method specialized for (D) .

Assuming that S is centrally symmetric, Khachiyan applies "barycentric coordinate ascent" to the problem

$$(D) \quad \max \ln \det X \text{Diag}(u) X^T, \quad e^T u = 1, u \geq 0,$$

starting with $u = (1/m)e$ (uniform distribution).

At each iteration, compute $\max_i x_i^T (X \text{Diag}(u) X^T)^{-1} x_i$ and stop if the max is at most $(1 + \epsilon)n$.

Else update $u \leftarrow (1 - \delta)u + \delta e_i$ for the optimal $\delta > 0$.

Algorithms, 3

Kumar and Yıldırım (2005), with applications in computational geometry with $m \gg n$ in mind, suggest a different initialization, eliminate the $\ln \ln m$ term, and obtain a **small core set**.

Todd and Yıldırım (2005) also compute **min** $x_j^T (X \text{Diag}(u) X^T)^{-1} x_j$ over j with u_j positive, and sometimes update

$$u \leftarrow (1 + \delta)u - \delta e_j$$

for the optimal $\delta > 0$.

This coincides with a method of **Atwood (1973)** and with the **Frank-Wolfe method with Wolfe's away steps (1970)**!

Their algorithm has the same complexity as the Kumar-Yıldırım method, and is likely to produce an even smaller core set.

Algorithms, 4

The optimality conditions for (D) are

$$x_i^T (X \text{Diag}(u) X^T)^{-1} x_i \leq n, \text{ for all } i,$$

with equality if $u_i > 0$.

The FW-K method computes u such that

$$x_i^T (X \text{Diag}(u) X^T)^{-1} x_i \leq (1 + \epsilon)n, \text{ for all } i.$$

The WA-TY method computes u such that

$$x_i^T (X \text{Diag}(u) X^T)^{-1} x_i \leq (1 + \epsilon)n, \text{ for all } i,$$

$$x_j^T (X \text{Diag}(u) X^T)^{-1} x_j \geq (1 - \epsilon)n, \text{ for all } j, u_j > 0.$$

13. Discussion

It appeared from computational experiments with small ϵ (10^{-10} is feasible compared with the expected 10^{-1} or 10^{-2}), that the number of iterations of the WA-TY algorithm grows with $O(\ln(1/\epsilon))$ not $O(1/\epsilon)$.

Ahipasaoglu-Sun-Todd (2006) have recently shown why this is the case: they prove a **local linear convergence rate**, but unfortunately not an improved global complexity bound.

For example, we can find an approximate rounding with $\epsilon = 10^{-10}$ for **1000 points in 100-dimensional space in under 4 seconds** (my laptop with a 1.5GHz Intel Pentium). The algorithm is even faster if we use a method of Harman-Prinzato (2005) to eliminate unnecessary points.

14. Conclusions

The minimum-volume ellipsoid problem and $-\ln \det$ minimization are ubiquitous in statistics, computational geometry, and optimization.

Interior-point methods led to a **revolution in optimization**, but their work per iteration is often very high. For large-scale problems, where high accuracy is not required, **sophisticated first-order methods can be very successful** (cf. recent work of Nesterov, Ben-Tal and Nemirovski, and others).