Robust minimum volume ellipsoids and higher order polynomial level sets

Dmitry Malioutov

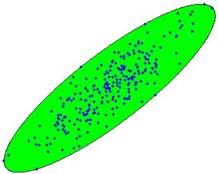
Machine Learning group,
IBM Research
TJ Watson Research Center, NY

Joint work with Amir Ali Ahmadi, Princeton University
and
Ronny Luss, IBM Research

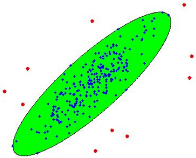
Dec 12, 2014

Overview

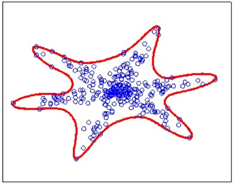
MVE problem: find an ellipsoid of minimum volume that contains a given set of data points in Euclidean space. Many applications.



Robust MVE: allow to ignore a fraction of the points as outliers. Hard problem.
Natural convex relaxation fails. We propose effective non-convex relaxations.



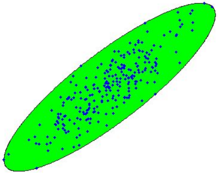
Extend to compact higher-order polynomial level-sets: formulation via Sum of Squares (SOS) programming.



Minimum volume ellipsoids

Overview of minimum volume ellipsoids (MVE)

The MVE problem asks to find an ellipsoid of minimum volume that contains a set of given data points in Euclidean space.



A convex formulation for the minimum volume zero-centered ellipsoid $E = \{x \mid x^T M x \leq 1\}$:

$$\min_{M \succ 0} -\log\det M \quad \text{such that} \quad \mathbf{x}_i^T M \mathbf{x}_i \leq 1, \quad i = 1, ..., m.$$

Applications in statistics, machine learning, control, e.t.c: covariance estimation, anomaly detection, change-point detection, experiment-design.

Overview of minimum volume ellipsoids (MVE)

Allowing an **arbitrary center** is non-convex in this formulation:

$$(\mathbf{x}_i - \boldsymbol{\mu})^T M(\mathbf{x}_i - \boldsymbol{\mu}) \leq 1, \quad i = 1, \ldots, m.$$

However, one can lift the problem to higher dimension: non-centered MVE is equivalent to finding the $d+1$-dim centered MVE for points $\bar{x}_i = [x_i; 1]$. We have $E = \{x \mid [x; 1] \in \bar{E}\}$.

Dual of the MVE: Define $M(\alpha) = \sum \alpha_i \mathbf{x}_i \mathbf{x}_i^T$. Then the dual is:

$$\max_{\alpha} \ \log\det M(\alpha)$$
$$\text{where } \sum \alpha_i = 1, \text{ and } \alpha_i \geq 0.$$

Dual is used for D-optimal experiment design. Multiplicative update solution (Titterington): $\alpha_i^{(n+1)} = \alpha_i^{(n)} \frac{\mathbf{x}_i^T \alpha}{N}$.

Robust MVE

In practice: need to address outliers. E.g., in anomaly detection we have an unlabeled mixture of normal and anomalous data.

Robust MVE: allow to ignore a fraction of the points, and fit MVE to the remaining points:

$$\min_{M \succ 0} -\log\det M \quad \text{such that}$$
$$\mathbf{x}_i^T M \mathbf{x}_i \leq 1 + \xi_i \quad \text{and} \quad \|\boldsymbol{\xi}\|_0 \leq k, \quad i = 1, \ldots, m.$$

Existing algorithms:

- Greedy influential point removal (ellipsoidal trimming).
- Random sampling: sample small subsets of points, fit ellipsoids, and expand.
- Branch and Bound (exact) (exponential complexity).

We will consider robust MVE based on convex relaxations.

Complexity of Robust MVE

We prove the following complexity results about the robust MVE:

Proposition

Given a set of m points in R^n with rational coordinates, and two rational numbers $v > 0$ an $r \in (0, 1)$, it is NP-hard to decide if there exists an ellipsoid of volume $\leq v$ that covers at least a fraction r of the points.

In fact, an even stronger statement is true:

Proposition

For any ϵ, $\delta \in (0, 1/2)$, given a set of m points in R^n with rational coordinates and a rational number $v > 0$, it is NP-hard to distinguish between the following cases: (i) there exists an ellipsoid of volume $\leq v$ that covers a fraction $(1 - \epsilon)$ of the points, and (ii) no ellipsoid of volume $\leq v$ can cover even a fraction δ of the points.

Natural convex relaxation for robust MVE

Motivated by the rich literature on ℓ_1 relaxations for sparse approximations, we first attempt an ℓ_1-formulation (ℓ_1-MVE):

$$\min_{M \succ 0} -\log\det M + \lambda \sum \xi_i$$
$$\text{such that} \quad \mathbf{x}_i^T M \mathbf{x}_i \le 1 + \xi_i, \quad \text{and} \quad \xi_i \ge 0 \ \forall i$$

The regularization parameter λ trades off sparsity of the errors vs. the volume. Convex problem. Variety of efficient solvers.

ℓ_1-MVE formulation does not give lower bounds on robust-MVE volume. We also develop an SDP formulation that provides such bounds (see appendix): i.e.
no ellipsoid that covers more than a fraction r of points can have volume less than v^*.

Limitations of the convex relaxation

The ℓ_1 relaxation gives very poor solutions for robust MVE. [1]

Intuitively: the effective penalty on each outlier depends on the geometry of the ellipsoid (i.e. on the eigenvalues of M). The ℓ_1-MVE stretches the ellipsoid in the direction of the outlier to reduce the ℓ_1 penalty on that outlier.

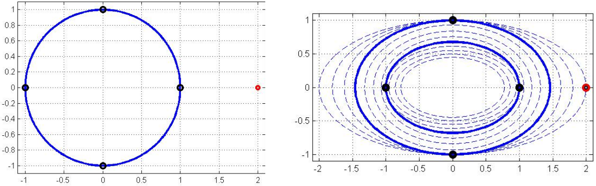


Figure : (a) Exact robust MVE solution. (b) The solution path of ℓ_1 MVE as a function of λ does not include the correct solution for any λ.

[1] ℓ_1-relaxations also fail for other sparse approximation problems: sparse-Markowitz portfolios, Total Least Squares (Malioutov et al., 2014), etc.

Reweighted-ℓ_1 MVE relaxation

Limitation of ℓ_1 norm: penalizes large coefficients more than small coefficients.

Weighted ℓ_1-norm: $\sum w_i |x_i|$. Defining $w_i = \frac{1}{|x_i^*|}$, where x^* is the unknown optimal solution would be equivalent to the ℓ_0-norm.

Practical solution: $w_i^{(n+1)} = \frac{1}{\delta + |\hat{x}_i^{(n)}|}$, with small $\delta > 0$.

Reweighted-ℓ_1 approach is equivalent to iterative linearization of the non-convex log-sum penalty for sparsity:[2]

$$\min_{M \succ 0} - \log \det M + \lambda \sum \log(\xi_i + \delta)$$

$$\text{such that} \quad \mathbf{x}_i^T M \mathbf{x}_i \leq 1 + \xi_i, \quad \text{and} \quad \xi_i \geq 0 \ \forall i$$

[2] Faster solution via iterative log-thresholding (Malioutov, Aravkin, 2014)

Experiments with RW-ℓ_1 MVE

(i) SOLVE (1) with a weighted ℓ_1-norm in the objective:
$-\log\det M + \lambda \sum_i w_i \xi_i$

(ii) UPDATE the weights $w_i = \frac{1}{\delta + |\hat{x}_i|}$.

Typically only a few iterations (< 10) needed for convergence.
At fixed point: $\sum_i w_i |x_i| \approx \sum_i \frac{|\hat{x}_i|}{\delta + |\hat{x}_i|} \approx \|\hat{x}\|_0$. This avoids the dependence on the geometry of the ellipsoid that plagues ℓ_1-MVE.

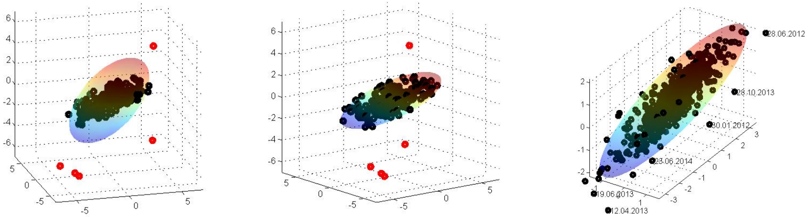


Figure : (a) ℓ_1-MVE. (b) RW-ℓ_1-MVE correctly identifies the outliers. (c) Oil-markets anomaly detection.

Extension to
higher-order polynomial level sets

Higher order polynomial level sets

Ellipsoids are sublevel sets of quadratic functions: $\{x \mid q(x) \leq 1\}$, where $q(x) \triangleq (x-\mu)^T M(x-\mu)$.
More flexible: sublevel sets of higher order (degree d) polynomials: $\{x \mid p(x) \leq 1\}$, where $p(x) = \sum_{\alpha:|\alpha|\leq d} a_\alpha x^\alpha = \sum_\alpha a_\alpha x_1^{\alpha_1} \ldots x_n^{\alpha_n}$.

Constraints $p(x_i) \leq 1$ for all i are linear in the coefficients a_α.

- We minimize a proxy for the volume as a heuristic.
- Impose compactness and convexity via *SOS* formulation.

Consider the set of positive semi-definite (p.s.d.) polynomials $p(x) \geq 0$. This is a convex set, but NP-hard to optimize over[3].

Sum of squares (SOS) approximation: $p(x) = \sum_i p_i(x)^2$. If $p(x)$ is SOS, then $p(x)$ is p.s.d. Converse not true in general.

---

[3] Ahmadi et. al, 2013

Sum of Squares (SOS) polynomials

For simplicity, we first assume that $p(x)$ is *homogeneous* (all monomials have the same degree).

Then compactness of $\{x \mid p(x) \leq 1\}$ is equivalent to $p(x) > 0$ for all x, i.e. $p(x)$ is p.d. SOS sufficient condition for p.d.:

$p(x) - \epsilon(x_1^2 + \ldots + x_n^2)^{d/2}$ is SOS $\Longrightarrow$ $p(x)$ is positive definite where ϵ is a small constant.

SDP formulation for SOS: Suppose $p(x)$ is degree d. Collect monomials up-to power $d/2$ into vector $z(x)$. Then $p(x)$ is SOS iff $p(x) = z(x)^T M \; z(x)$ for some p.s.d. matrix $M \succeq 0$.

SOS formulation with compactness and convexity

Heuristic for **minimizing the volume** of the sublevel set:[4]

$$
\begin{array}{ll}
\text{minimize}_{p,\beta} & \beta \\
\text{subject to} & p(x_i) \leq \beta, \quad i = 1, \ldots, m \\
& p(x) - \epsilon(x_1^2 + \ldots + x_n^2)^{d/2} \text{ is SOS} \\
& \int_{x_1^2+\ldots+x_n^2=1} p(x) = 1
\end{array}
$$

The integral $\int_{S^n} p(x) = 1$ over the unit sphere S^n reduces to a single linear constraint on the coefficients of $p(x)$.

Convexity: $p(x)$ convex is sufficient for $\{x \mid p(x) \leq 1\}$ to be convex. However, NP-hard to enforce (or even check) for $d > 2$.

SOS-convexity:
$p(x)$ is SOS-convex if $g(x, y) = y^T H(x) y$ is SOS.
$p(x)$ is SOS-convex $\implies$ $p(x)$ is convex.

---

[4] Another approximation for volume (Magnani, Lall, Boyd, 2005): $\min -\log\det M$, where M appears in SOS-convex constraint.

Experiments with SOS-poly level sets

Robust versions can be formulated in the same manner as for MVE, by allowing sparse errors $p(x_i) \leq 1 + \xi_i$, $i = 1, ..., m$.

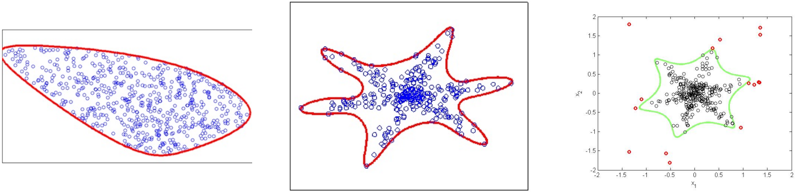


Figure : (a) Non-convex compact polynomial level set. (b) Convex compact polynomial level-set. (c) Robust polynomial level-set.

An alternative formulation for level-sets of higher order polynomials is through kernel-MVE (Dolia et al., 2007). However, it does not allow enforcing compactness and convexity.

Summary and Conclusion

Talk summary:

- Reviewed the robust minimum volume ellipsoid problem
- Established its computational complexity
- Studied convex relaxations and showed their limitations
- Proposed a reweighted-ℓ_1 approach for robust-MVE
- Extended the framework to higher-order polynomial level-sets via sum of squares (SOS) programming

Directions for future work:

- Fast algorithms
- Polynomials with sparse coefficients

Thank you!

Appendix

SDP lower bound

ℓ_1-MVE formulation does not give lower bounds on robust-MVE volume. These can be obtained via an SDP formulation:

An equivalent formulation of robust MVE is (for large C):

$$\begin{array}{ll} \min_{M\succ 0} & -\log\det(M) \\ \text{subject to} & x_i^T M x_i \leq 1 + C\xi_i, \\ & \xi_i(1-\xi_i) = 0, \sum_i \xi_i \leq k \end{array}$$

Define $Y = [\xi^T, 1]^T[\xi^T, 1]$ and another equivalent formulation is:

$$\begin{array}{ll} \min_{M\succ 0} & -\log\det(M) \\ \text{subject to} & x_i^T M x_i \leq 1 + CY_{ii}, Y_{n+1,n+1} = 1 \\ & Y_{n+1,i} = Y_{ii}, \sum_i Y_{ii} \leq k, Y \succeq 0, \text{rank}(Y) = 1 \end{array}$$

and if we drop the rank constraint, we get a convex lower bound.

SOS formulation with convexity

Suppose we need sublevel-sets $\{x \mid p(x) \leq 1\}$ to be convex.
Sufficient condition: $p(x)$ is convex. NP-hard to enforce for $d > 2$.

Instead we impose that $p(x)$ is **SOS-convex**:

- $p(x)$ is SOS-convex if the Hessian $H(x)$ is an SOS-matrix.
- $H(x)$ is an SOS-matrix if $y^T H(x) y$ is SOS in lifted dimension $z = (x^T, y^T)^T$
- $p(x)$ is SOS-convex $\implies$ $p(x)$ is convex.

The heuristic for minimizing the volume of convex sublevel set is

$$\begin{array}{ll} \text{minimize}_{p,\beta} & \beta \\ \text{subject to} & p(x_i) \leq \beta, \quad i = 1, \ldots, m \\ & p(x) - \epsilon(x_1^2 + \ldots + x_n^2)^{d/2} \text{is SOS-convex} \\ & \int_{x_1^2+\ldots+x_n^2=1} p(x) = 1 \end{array}$$

References

- D. M. Titterington, "Estimation of correlation coefficients by ellipsoidal trimming," *Journal of the Royal Statistical Society. Series C, Applied Statistics.*, pp. 227–234, 1978.
- P. J. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications, Vol. B*, pp. 283–297. Reidel, 1985.
- A. N. Dolia, C. J. Harris, J. S. Shawe-Taylor, and D. M. Titterington, "Kernel ellipsoidal trimming," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 309–324, 2007.
- M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *IEEE American Control Conference*, 2001.
- E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. of Fourier Analysis and Applic.*, vol. 14, no. 5, pp. 877–905, 2008.
- P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical programming*, vol. 96, no. 2, pp. 293–320, 2003.
- J. Gotoh and A. Takeda, "Conditional minimum volume ellipsoid with application to multiclass discrimination," *Computational Optim. and Applications*, vol. 41, no. 1, pp. 27–51, 2008.
- D. Malioutov and N. Slavov, "Convex total least squares," in *Int. Conf. Machine Learning (ICML)*, 2014.
- A. A. Ahmadi and P. A. Parrilo, "A complete characterization of the gap between convexity and SOS-convexity," *SIAM Journal on Optimization*, 2013.
- D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in *Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 113–117.
- A. A. Ahmadi, A. Olshevsky, P. A. Parrilo and J. N. Tsitsiklis, "NP-hardness of Deciding Convexity of Quartic Polynomials and Related Problems," *Mathematical Programming*, 2013.
- D. Malioutov and A. Aravkin, "Iterative Log Thresholding," in *ICASSP*, 2014.