Robust minimum volume ellipsoids and higher-order polynomial level sets

Amir Ali Ahmadi
ORFE Department
Princeton University
Princeton, NJ
a_a_a@princeton.edu

Dmitry Malioutov
IBM Research
Yorktown Heights, NY
dmalioutov@us.ibm.com

Ronny Luss
IBM Research
Yorktown Heights, NY
rluss@us.ibm.com

Abstract

Minimum volume ellipsoids appear in a variety of important statistics and machine learning problems: anomaly detection, covariance estimation, experiment design, and control. However, many of these applications involve data corrupted by outliers. For example, anomaly detection involves simultaneously learning a representation for typical behavior while excluding a subset of observations as outliers. The most popular approaches for fitting ellipsoids with outliers have been heuristic, e.g. sampling, or iteratively removing influential points and re-fitting the ellipsoid. We follow a different approach by formulating an optimization problem to simultaneously label the outliers and fit a minimum volume ellipsoid to the remaining points. We first consider the combinatorial robust minimum volume ellipsoid problem and establish its complexity. Next, we explore a natural convex relaxation for this problem, which as we show, can be dramatically improved by using a non-convex penalty and an efficient iterative linearization approach. Finally, we extend it to more general higher-order compact polynomial level sets via a sum-of-squares (SOS) formulation using semi-definite programming (SDP).

1 Introduction

The minimum volume ellipsoid (MVE) problem asks to find an ellipsoid of minimum volume that contains a set of given data points in Euclidean space. It has been very popular in a variety of applications in statistics, optimization, and control due to a simple convex formulation allowing efficient algorithms [1]. Furthermore, a variety of extensions have been considered: minimum volume ellipsoid containing a collection of polytopes, a set of other ellipsoids, and more general compact shapes [2, 3]. This enables a general framework to summarize complex shapes and distributions by simple ellipsoids.

Many authors have expressed the need for a robust version of the minimum volume ellipsoid [4, 5, 6] allowing to ignore a fraction of the dataset as outliers and to fit a minimum volume ellipsoid to the remaining points. The challenge of the robust MVE formulation is that the identity of the outliers is not known, and one needs to simultaneously identify the outliers and fit a small volume ellipsoid to the remaining points. Existing approaches to the problem have mostly been heuristic, with the two most popular ones being ellipsoidal trimming that iteratively removes points falling on the boundary [7] and subsampling methods [4, 6]. Exact computations based on combinatorial optimization are available for small to medium datasets [8, 9] but these are not scalable to larger datasets.

We first study the complexity of the robust MVE problem, and show its computational hardness. We explore a convex relaxation based approach following the rich literature on ℓ_1-norm relaxations for sparse approximations. Despite its successes for sparse approximation and compressed sensing, we show that it has poor performance for the robust MVE problem, and furthermore, that using a non-convex log-sum penalty vith a fast iterative linearization [10, 11] provides a very effective and practical way to solve the robust MVE problems.

More generally our goal is to find a concise but flexible description of typical sets of data points [12, 5, 13, 14]. Ellipsoids can be viewed as level sets of positive-definite quadratic functions, and we extend the approach to more flexible shapes given by compact level sets of higher order polynomials. Our approach for optimizing over compact higher-order polynomial level sets is based on sum-of-squares optimization [15] and minimizes a surrogate for the volume.

2 Minimum volume ellipsoids

The problem of finding the minimum volume ellipsoid that contains a collection of points dates back to the work in [7]. Suppose that we have a set of vectors $\{\mathbf{x}_i\}_{i=1}^m$ in $\mathbb{R}^n$. For simplicity assume that we aim to find a zero-centered ellipsoid defined as $\{\mathbf{x} \mid \mathbf{x}^T M \mathbf{x} \leq 1\}$ where $M \succ 0$ is a positive definite $n \times n$ matrix. The volume of this ellipsoid is proportional to the determinant of M^{-1}, leading to a convex formulation for the problem:

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

A variety of efficient algorithms for solving this problem have been proposed, e.g. [1]. The dual of the MVE problem is important in the design of experiments, where it can be interpreted as D-optimal experiment design [16]. A simple multiplicative update algorithm for MVE can be defined via the dual.

2.1 Robust MVE and its complexity

In many practical applications the data may be corrupted by outliers. We review a combinatorial formulation for the robust MVE problem which allows to ignore up to a fraction r or $k = rm$ vectors and find the minimum volume ellipsoid enclosing the remaining vectors:

$$\min_{M \succ 0} -\log\det M \quad \text{such that} \quad \mathbf{x}_i^T M \mathbf{x}_i \leq 1 + \xi_i \quad \text{and} \quad \|\boldsymbol{\xi}\|_0 \leq k, \;\; i = 1, ..., m. \tag{2}$$

This problem has been described in [4], and an approximate solution based on resampling has been proposed: A collection of subsets of exactly $n + 1$ points are used to fit initial minimum volume ellipsoids containing these subsets. Then each of the small ellipsoids is "inflated" until they contain $n - k$ points, and the best (smallest volume) among them is chosen as the robust MVE solution. Another heuristic solution is based on trimming [7]: at each iteration a non-robust minimum volume ellipsoid is fit, and the points on the boundary are identified as outliers and removed. The procedure continues until the desired number of outliers is removed and has no guarantees to identify the same outliers as robust MVE in (2). An exact combinatorial solution [9] can be made more efficient by branch and bound [8], but these are only practical for relatively small data-sets.

Complexity of the problem In the extended version of this work, we prove the following complexity results about the robust MVE:

Proposition 1 *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 2 *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.*

The 0-volume special case asks if there is a subset of points of size $k = rm$ that lie in a proper subspace [17]. We proceed to study approximate methods for robust MVE.

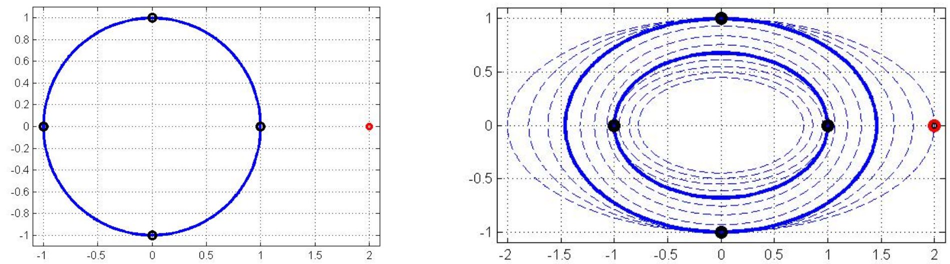

Figure 1: (a) Exact robust MVE solution. (b) The solution path of ℓ_1 MVE as a function of α does not include the correct solution for any α.

3 Convex formulation for robust MVE and its limitations

Following the rich literature on sparse approximations and sparse learning, we approximate the difficult sparsity constraint on outliers (number of non-zeros in $\boldsymbol{\xi}$) by the ℓ_1-norm of $\boldsymbol{\xi}$. While the ℓ_1 relaxation can be used to recover exact sparse solutions in certain problems such as sparse linear regression, we show that it is inappropriate for robust MVE due to the geometry of the ellipsoid, as we discuss below. We propose a significant performance improvement in Section 4 using non-convex penalties.

A natural convex relaxation for (2) replaces the count of non-zero elements of the outlier vector ξ with the ℓ_1-norm [1,2]. For convenience, we penalize the ℓ_1 norm in the objective with regularization parameter α and solve:

$$\min_{M \succ 0} -\log\det M + \alpha \sum \xi_i \tag{3}$$
$$\text{such that} \quad \mathbf{x}_i^T M \mathbf{x}_i \leq 1 + \xi_i, \quad \text{and} \quad \xi_i \geq 0 \, \forall i$$

This is a convex formulation (we will refer to it as ℓ_1-MVE) that can be solved by interior point methods. Alternatively, the dual can be optimized by first-order projected gradient or proximal methods. Despite the overwhelming popularity and strong empirical and theoretical results for using convex relaxations in compressed sensing, sparse log-linear models and related fields [19], the ℓ_1 relaxation is problematic for the robust MVE problem and gives poor results.

Consider a small computational example in Figure 1 with four points on a unit circle and an outlier at $(2, 0)$. The exact robust MVE solution correctly removes the outlier in plot (a). However, plot (b) demonstrates that the ℓ_1 MVE approximation fails: one can show that the solution path in (3) does not contain the correct solution for any α. The intuition is that the effective penalty on each outlier is different and depends on the geometry of the ellipsoid (i.e. on the eigenvalues of M), and the ℓ_1 MVE stretches the ellipsoid in the direction of the outlier to reduce the ℓ_1 penalty on the outlier.

4 Robust MVE via reweighted-ℓ_1 penalties

While the ℓ_1 norm has been the most widely used approximation to sparsity, many authors have observed that non-convex approximations can have substantially better empirical performance despite the lack of guarantees to reach global optima (initial theoretical results have also started to appear [20]). These non-convex approximations are especially important where the ℓ_1 norm is entirely inappropriate, e.g. optimization over a probability simplex where the ℓ_1 norm is already constrained by the problem, structured Total Least Squares [21], and as we discuss here, for robust MVE.

[1]A very interesting related formulation explores a connection with the conditional value at risk literature in finance and proposes a convex cVaR-relaxation for robust MVE [18].

[2]We have also developed an alternative convex formulation based on an SDP relaxation which provides lower bounds on achievable volume for robust MVE.

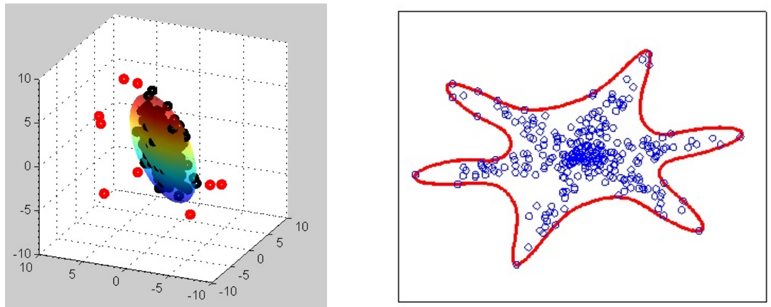

Figure 2: (a) Robust MVE via reweighted-ℓ_1 correctly identifies the outliers. (b) A compact higher-order polynomial level-set obtained via SOS programming provides a more flexible description of a complex set of points.

Specifically, unlike the ℓ_0 norm, the ℓ_1 norm penalizes large coefficients more than small coefficients. It has been suggested [10, 11] that if one had access to the magnitudes of the true sparse solution $\mathbf{x}^*$, and used a *weighted* ℓ_1-norm defined as $\sum w_i|x_i|$ for positive weights $w_i = \frac{1}{x_i^*}$, this would be equivalent to the ℓ_0-norm. In lieu of knowing the true magnitudes, the practical approach is iterative where one first solves an unweighted ℓ_1 problem, and then iteratively sets the weights based on the previous solution, $w_i = \frac{1}{\delta + |\hat{x}_i|}$, where δ is a small positive constant to avoid division by zero. This reweighted-ℓ_1 approach is equivalent to iterative linearization of the non-convex $\log$-sum penalty for sparsity. Dramatic empirical performance improvements in terms of recovering sparse signals are now starting to be supported by emerging theoretical results [22]. We propose to apply the $\log$-sum penalty using iterative reweighting for the robust MVE problem, which we call $\log$-MVE. The resulting algorithm is to first initialize $w_i = 1$, and then to iterate:

(i) SOLVE (3) with a weighted ℓ_1-norm in the objective: $-\log\det M + \alpha \sum_i w_i \xi_i$

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

Typically only a very small number of iterations (5 to 10) is required for convergence. Note that upon reaching a 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. We illustrate the effectiveness of the formulation in Figure 2 where $\log$-MVE successfully identifies the outliers and learns the MVE for a set of points in three dimensions. Note that log-MVE also gets the exact MVE solution for the problem in Figure 1 (a). We are exploring conditions under which this recovery is guaranteed.

5 SOS polynomial level sets

In the full version of this work, we extend the methodology presented here from ellipsoids to level sets of higher degree polynomials. As Figure 2 (b) demonstrates, this could be very useful in situations where the complex pattern of the data points cannot be captured by a simple shape like an ellipsoid. The red curve in the Figure is the level set of a degree-6 polynomial function that has been found via *sum of squares optimization* [15]. These techniques allow us to place various shape constraints on the level set; for example compactness as in Figure 2 (b), or convexity, which may be desired in other situations. Such constraints are not possible using the kernel-ellipsoid formulation [5]. The full version of our paper also presents extensions where we use polynomial level sets to do robust coverage, i.e., cover only a specified fraction of the data points.

6 Conclusions

We have considered the robust MVE problem, and commented on its complexity. We first explored a convex formulation for robust MVE, and found that it suffers from rotational asymmetry of the penalty. A non-convex $\log$-sum penalty using an iterative linearization approach alleviates this issue and provides a powerful formulation for robust MVE. We also proposed an extension with level sets of higher-order polynomials.

References

[1] P. Sun and R.M. Freund, "Computation of minimum-volume covering ellipsoids," *Operations Research*, vol. 52, no. 5, pp. 690–706, 2004.

[2] P. Kumar and E.A. Yildirim, "Minimum-volume enclosing ellipsoids and core sets," *Journal of Optimization Theory and Applications*, vol. 126, no. 1, pp. 1–21, 2005.

[3] L. Pronzato and E. Walter, "Minimum-volume ellipsoids containing compact sets: application to parameter bounding," *Automatica*, vol. 30, no. 11, pp. 1731–1739, 1994.

[4] P. J. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications, Vol. B*, pp. 283–297. Reidel, 1985.

[5] 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.

[6] S. Van Aelst and P. Rousseeuw, "Minimum volume ellipsoid," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 71–82, 2009.

[7] D. M. Titterington, "Estimation of correlation coefficients by ellipsoidal trimming," *Journal of the Royal Statistical Society. Series C, Applied Statistics.*, pp. 227–234, 1978.

[8] J. A. Candela, "Exact iterative computation of the multivariate minimum volume ellipsoid estimator with a branch and bound algorithm," in *COMPSTAT*. Springer, 1996, pp. 175–180.

[9] R. D. Cook, D. M. Hawkins, and S Weisberg, "Exact iterative computation of the robust multivariate minimum volume ellipsoid estimator," *Statistics & probability letters*, vol. 16, no. 3, pp. 213–218, 1993.

[10] 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.

[11] 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.

[12] V. Barnett, "The ordering of multivariate data," *Journal of the Royal Statistical Society. Series A*, pp. 318–355, 1976.

[13] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[14] C. Bennett and Campbell K. P., "A linear programming approach to novelty detection," in *Advances in Neural Information Processing Systems*. MIT Press, 2001, vol. 13, p. 395.

[15] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical programming*, vol. 96, no. 2, pp. 293–320, 2003.

[16] D. M. Titterington, "Optimal design: Some geometrical aspects of d-optimality," *Biometrika*, vol. 62, no. 2, pp. 313–320, 1975.

[17] V. Guruswami and P. Raghavendra, "Hardness of solving sparse overdetermined linear systems: A 3-query PCP over integers," *ACM Transactions on Computation Theory (TOCT)*, vol. 1, no. 2, pp. 6, 2009.

[18] 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.

[19] P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*, Springer, 2011.

[20] P. Loh and M.J. Wainwright, "Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima," in *Advances in Neural Information Processing Systems*, 2013, pp. 476–484.

[21] D. Malioutov and N. Slavov, "Convex total least squares," in *Int. Conf. Machine Learning (ICML)*, 2014.

[22] D. Needell, "Noisy signal recovery via iterative reweighted l1-minimization," in *Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 113–117.