Outlier Detection Using the Minimum Volume Ellipsoid

Kathleen Chang

Stanford University

Abstract

The minimum volume ellipsoid is a convex optimization problem that takes a set of points and constructs the minimum volume ellipsoid necessary to span them. Using the dual problem, the Mahalanobis distance, a generalization of the point's position in the distribution of all points in the set, is computed to identify the "most binding" points that can be removed to significantly decrease ellipsoid volume. The applications of this is in statistical analysis of outliers and as a ranking tool in multivariate space. This project implements different algorithms for outlier detection using the minimum volume ellipsoid problem and compares results.

Background

If $C \in \mathbb{R}^n$ is bounded with a nonempty interior, the minimum volume ellipsoid $\mathcal{E}$ covers C. $\mathcal{E}$ is also called the Löwner-John ellipsoid.

$$\mathcal{E} = \{v \,|\, ||Av + b||_2 \leq 1\} \text{ where } A \in \mathcal{S}^n_{++}$$

The convex optimization problem to solve for the minimum volume ellipsoid is (Boyd, 2011):

$$\begin{array}{ll} \min & \log\det A^{-1} \\ \text{s.t.} & \sup_{v \in C} ||Av + b||_2 \leq 1 \end{array} \qquad (1)$$

The dual is obtained by lifting the primal problem to another dimension (Moshtagh, 06).

$$X_{dual} = X_{primal} \cup H$$

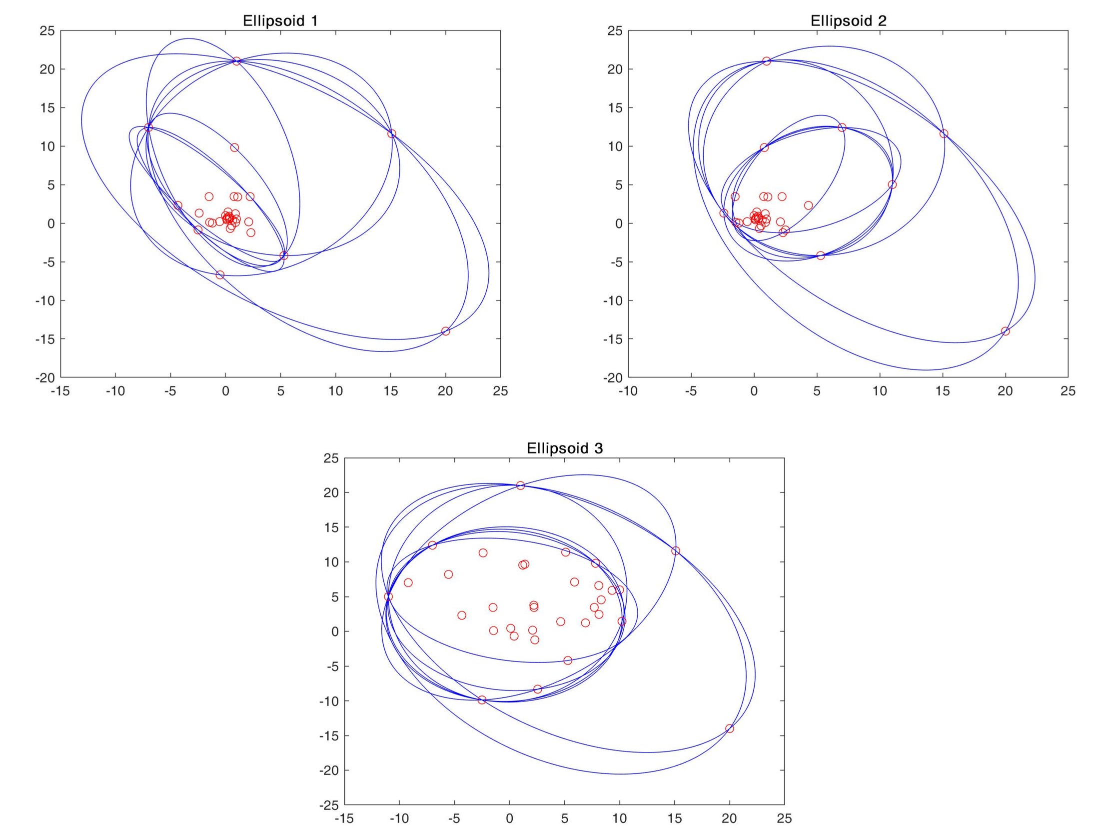
$$H = \{(x, x_{n+1}) \in \mathbb{R}^{n+1} | x_{n+1} = 1\}$$

The dual is

$$\begin{array}{ll} \max & g(u) = \log\det(XUX^T) \\ \text{s.t.} & \mathbf{1}^T u = 1 \\ u \geq 0 & \end{array} \qquad (2)$$

Ellipsoid Peeling

- Select points to remove by repeatedly solving (1) and peeling point corresponding to maximum dual variable



Mahalanobis Distance

$$M(x) = \sqrt{(\vec{x} - \vec{\mu})^T \mathbf{S}^{-1} (\vec{x} - \vec{\mu})}$$

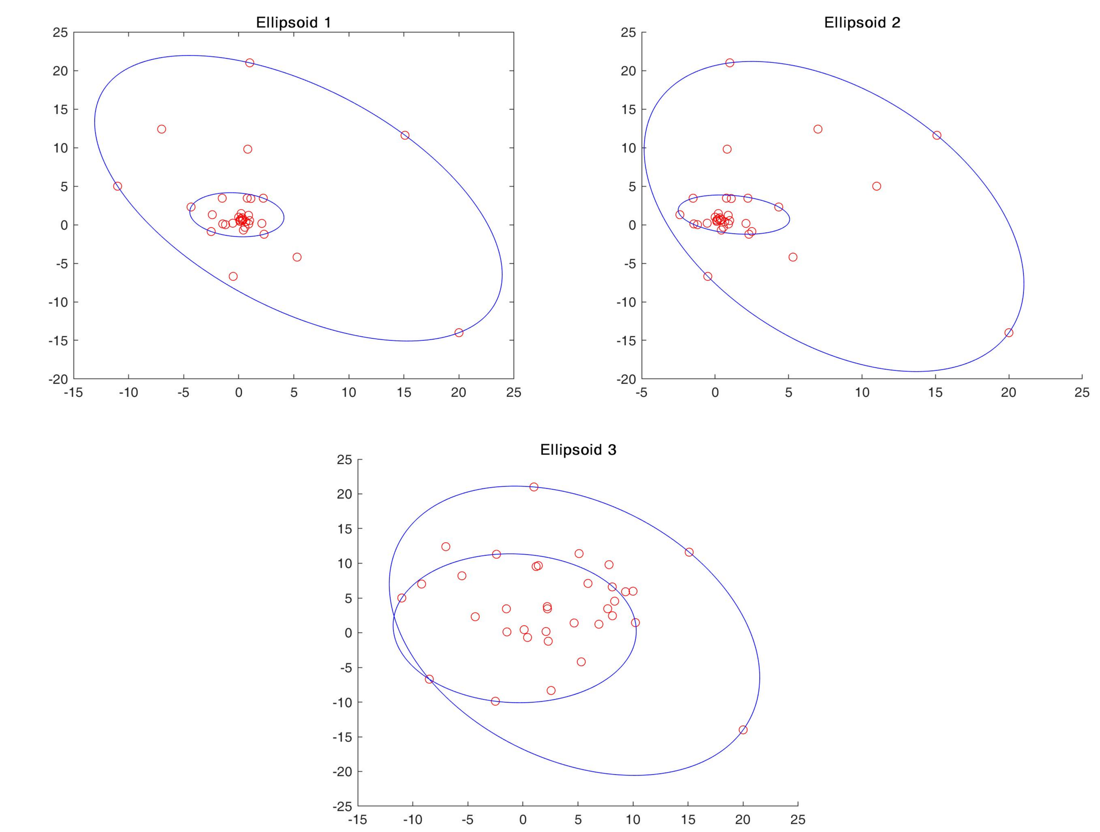
- Generalization to multidimensional space of a point's position $\vec{x}$ in a distribution using covariance matrix **S**
- Distribution $\vec{\mu}$ has a Mahalanobis distance is 0

$$M(z, \mathcal{E}) = \inf\{t \geq 0 \,|\, z \in c + t(\mathcal{E} - c)\}$$

- Factor to scale an ellipsoid $\mathcal{E}$ about its center such that a point z lies on the boundary of the ellipsoid
- u lies on the unit simplex making it a probability distribution over the set X
- By solving the dual problem, we can find the Mahalanobis distance for each point x_k as $x_k(XUX^T)^{-1}x_k$

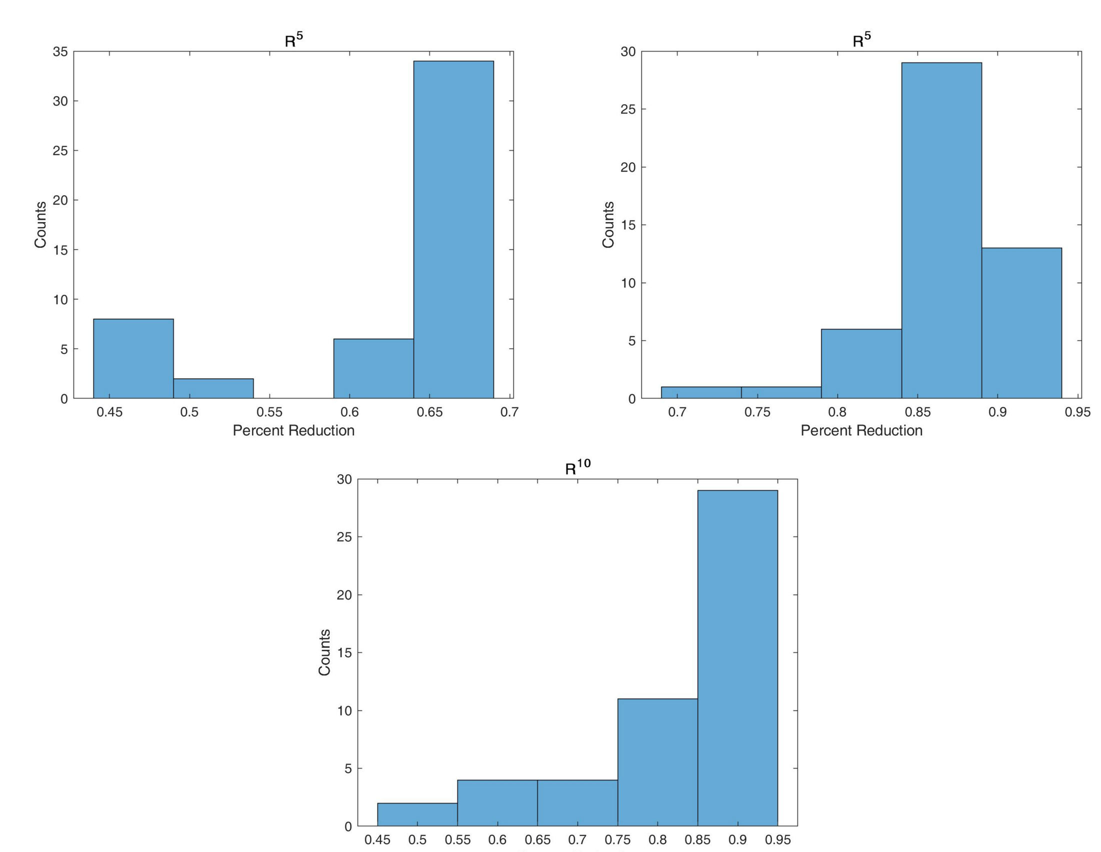
Ellipsoid Ordering

- Select points to remove by solving (2) and removing largest Mahalanobis distance $\xi_k = x_k(XUX^T)x_k$ associated with x_k.



Random Ellipsoid Ordering

- Perform 50 trials of random ellipsoid ordering (h = 27) on 35 points in $\mathbb{R}^2, \mathbb{R}^5$, and $\mathbb{R}^{10}$
- Select number of basis vectors based on the dimension of points
- Trade-off between dimensionality of data and range of percent volume reduction
- Effectiveness of reducing ellipsoid volume depends on initial basis vectors



Comparison of Methods

Ellipsoid Peeling

Ellipsoid $\in \mathbb{R}^2$	Original Volume	Reduced Volume	% Reduction
1	289.65	32.36	88.8%
2	234.44	29.69	87.3%
3	331.81	99.76	69.9%

Ellipsoid Ordering

Ellipsoid $\in \mathbb{R}^2$	Original Volume	Reduced Volume	% Reduction
1	289.65	12.12	95.8%
2	234.44	9.38	96%
3	331.81	118.34	64.3%

Random Ellipsoid Ordering

Ellipsoid	$\mathbb{R}^2$	$\mathbb{R}^5$	$\mathbb{R}^{10}$
Mean % Reduction	60.81%	86.41%	83.03%

- Ellipsoid ordering reduces volume more than ellipsoid peeling
- Random ellipsoid ordering does not outperform either ellipsoid peeling or ordering in $\mathbb{R}^2$ case
- Larger volume reduction in non-uniformly distributed sets

References

[1] Moshtagh, Nima. (2006). Minimum volume enclosing ellipsoid.

[2] Ahipaşaoğlu, S.D. Fast algorithms for the minimum volume estimator. J Glob Optim 62, 351–370 (2015). https://doi.org/10.1007/s10898-014-0233-8

[3] Boyd, S., & Vandenberghe, L. (2022). Additional Exercises for Convex Optimization.

[4] Boyd, S. P., & Vandenberghe, L. (2011). Convex optimization. Cambridge Univ. Pr.