The D$_k$-optimal Design Problem and Its Dual

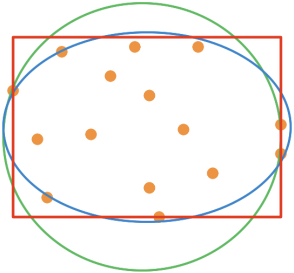
Selin Damla Ahipasaoglu

University of Southampton

13 July 2023

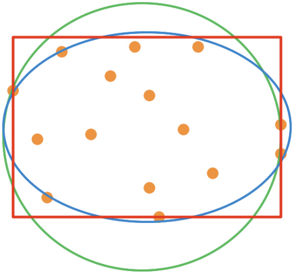
1. Geometry

Approximate sets of points with simple geometric bodies, such as balls, ellipsoids, boxes, or cylinders.



1. Geometry

Approximate sets of points with simple geometric bodies, such as balls, ellipsoids, boxes, or cylinders.



Ellipsoids are prefered to balls (similarly cylinders to boxes) because they are flexible and smooth. We can easily find the optimizer of a linear function over an ellipsoid.

1. Geometry

The set

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

for $\bar{x} \in \mathbf{R}^n$ and $H \succ 0$ is an ellipsoid in $\mathbf{R}^n$ with center $\bar{x}$ and shape defined by H.

We have

$$\text{vol}(\mathcal{E}(\bar{x}, H)) = \text{const}(n)/\sqrt{\det H},$$

and minimizing the volume of $\mathcal{E}(\bar{x}, H)$ is equivalent to minimizing

$$-\ln \det H.$$

1. Geometry: The Fritz John Theorem

Theorem (John, 1948)

For any point set $\mathcal{X} = \{x_1, \ldots, x_m\} \subset \mathbf{R}^n$, *there is an ellipsoid* $\mathcal{E}$, *which satisfies*

$$\bar{x} + \frac{1}{n}\mathcal{E} \subseteq \text{conv}(\mathcal{X}) \subseteq \bar{x} + \mathcal{E},$$

1. Geometry: The Fritz John Theorem

Theorem (John, 1948)

For any point set $\mathcal{X} = \{x_1, \ldots, x_m\} \subset \mathbf{R}^n$*, there is an ellipsoid* $\mathcal{E}$*, which satisfies*

$$\bar{x} + \frac{1}{n}\mathcal{E} \subseteq \text{conv}(\mathcal{X}) \subseteq \bar{x} + \mathcal{E},$$

furthermore if $\mathcal{X} = -\mathcal{X}$*,*

$$\frac{1}{\sqrt{n}}\mathcal{E} \subseteq \text{conv}(\mathcal{X}) \subseteq \mathcal{E}.$$

1. Geometry: The Fritz John Theorem

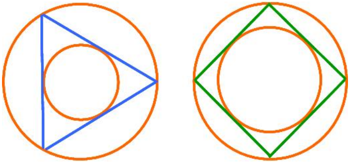
Theorem (John, 1948)

For any point set $\mathcal{X} = \{x_1, \ldots, x_m\} \subset \mathbf{R}^n$*, there is an ellipsoid* $\mathcal{E}$*, which satisfies*

$$\bar{x} + \frac{1}{n}\mathcal{E} \subseteq \mathrm{conv}(\mathcal{X}) \subseteq \bar{x} + \mathcal{E},$$

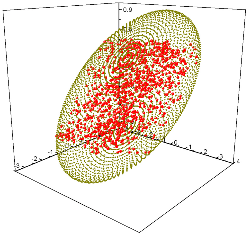
furthermore if $\mathcal{X} = -\mathcal{X}$*,*

$$\frac{1}{\sqrt{n}}\mathcal{E} \subseteq \mathrm{conv}(\mathcal{X}) \subseteq \mathcal{E}.$$



2. Minimum Volume Enclosing Ellipsoids

Given m points $\mathcal{X} := \{x_1, x_2, \ldots, x_m\} \in \mathbf{R}^n$ which span $\mathbf{R}^n$, the Minimum Volume Enclosing Ellipsoid (MVEE) problem finds an ellipsoid which is centered at the origin (wlog), covers all points, and has minimum volume.



2. MVEE Formulation

The MVEE problem can be formulated as follows:

$$
(P) \qquad \begin{array}{lrcl} \min & f(H) & := & -\ln\det H \\ & x_i^T H x_i & \leq & n,\ i = 1, \ldots, m, \\ & H & \succ & 0. \end{array}
$$

Problem (P) is convex, with linear inequality constraints.

This is also an SDP problem. Interior-point methods can be applied to the problem with barrier function $-\ln\det$ on S^n_{++}.

3. The Approximate D-optimal Design Problem

Let $X := [x_1, \ldots, x_m] \in \mathbf{R}^{n \times m}$ and $U := \mathrm{Diag}\,(u)$, the dual problem to the MVEE can be written as

$$(D) \qquad \begin{array}{rrcl} \max & g(u) & := & \ln\det XUX^T \\ & e^T u & = & 1, \\ & u & \geq & 0. \end{array}$$

(D) is the statistical problem of finding a D-optimal design measure on the columns of X, that maximizes the determinant of the Fisher information matrix when estimating all parameters $\theta_1, \ldots, \theta_n$ in the linear model

$$\tilde{y} \approx X^T \theta.$$

4. Duality / Optimality conditions / Equivalence theorem

For any H feasible for (P) and u feasible for (D), we have

$$g(u) \geq f(H).$$

4. Duality / Optimality conditions / Equivalence theorem

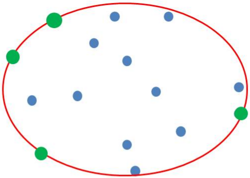
For any H feasible for (P) and u feasible for (D), we have

$$g(u) \geq f(H).$$

Furthermore, optimal solutions $\hat{H}$ and $\hat{u}$ exist and satisfy the following necessary and sufficient conditions:

(a) $\hat{u}_i > 0$ only if $x_i^T \hat{H} x_i = n$

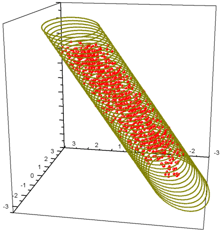
(b) $\hat{H} = (X\hat{U}X^T)^{-1}$.



5. Minimum Area Enclosing Ellipsoidal Cylinders

Given m points $\{x_1, x_2, ..x_m\} \in \mathbf{R}^n$ which span $\mathbf{R}^n$ and $k \leq n$, the Minimum Area Enclosing Ellipsoidal Cylinder (MAEC) problem finds an ellipsoidal cylinder which is centered at the origin, covers all points and has minimum area intersection with

$$\Pi := \left\{ \begin{bmatrix} y \\ z \end{bmatrix} \in \begin{bmatrix} \mathbf{R}^k \\ \mathbf{R}^{n-k} \end{bmatrix} : z = 0 \right\}.$$



6. Geometry

The set

$$\mathcal{C}(E,H) := \{[y;z] \in \mathbf{R}^n : (y+Ez)^T H(y+Ez) \leq k\}$$

for $E \in \mathbf{R}^{k\times(n-k)}$ and $H \succ 0$ is a cylinder in $\mathbf{R}^n$ defined by shape matrix H and axis direction matrix E.

Note that $\mathcal{C}(E,H) \cap \Pi$ is an ellipsoid in $\mathbf{R}^k$ with

$$\text{vol}(\mathcal{C}(E,H) \cap \Pi) = \text{const}(k)/\sqrt{\det H},$$

and minimizing the volume of $\mathcal{C}(E,H) \cap \Pi$ is equivalent to minimizing

$$-\ln\det H.$$

7. MAEC Formulation

The MAEC problem can be formulated as follows:

$$\begin{array}{lrcl} \min & f(H_{YY}) & := & -\ln\det H_{YY} \\ & (y_i + Ez_i)^T H_{YY} (y_i + Ez_i) & \leq & k, \; i = 1, \ldots, m, \end{array}$$

7. MAEC Formulation

The MAEC problem can be formulated as follows:

$$\begin{array}{lrcl} \min & f(H_{YY}) & := & -\ln\det H_{YY} \\ & (y_i + Ez_i)^T H_{YY} (y_i + Ez_i) & \leq & k,\ i = 1, \ldots, m, \end{array}$$

or equivalently

$$(P) \qquad \begin{array}{lrcl} \min & f(H) & := & -\ln\det H_{YY} \\ & x_i^T H x_i & \leq & k,\ i = 1, \ldots, m, \\ & H & \succeq & 0, \end{array}$$

where $H = \begin{pmatrix} H_{YY} & H_{YZ} \\ H_{YZ}^T & H_{ZZ} \end{pmatrix}.$

8. The D_k-optimal Design Problem

The dual problem can be stated as

$$(D) \quad \begin{array}{rcl} \max_{u,K} \; g(u) & := & \ln\det K \\ XUX^T - \overline{K} & := & XUX^T - \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix} \quad \succeq 0 \\ e^T u & = & 1, \\ u & \geq & 0. \end{array}$$

(D) is the statistical problem of finding a D_k-optimal design measure on the columns of X, that maximizes the determinant of a Schur Complement in the Fisher information matrix which is related to estimating the first k parameters $\theta_1, \ldots, \theta_k$ in the linear model

$$\tilde{y} \approx X^T \theta.$$

9. Weak Duality

Suppose H, u and K are feasible in (P) and (D) respectively. Then

$$0 \leq H \bullet (XUX^T - \overline{K}) = \sum_i u_i x_i^T H x_i - H \bullet \overline{K} \leq ke^T u - H_{YY} \bullet K.$$

Hence we have

$$\begin{array}{rcl} -\ln\det H_{YY} - \ln\det K & = & -\ln\det H_{YY}K \\ = -k\ln(\Pi_{i=1}^k \lambda_i(H_{YY}K))^{1/k} & \geq & -k\ln\left(\frac{\sum_{i=1}^n \lambda_i(H_{YY}K)}{k}\right) \\ \geq -k\ln\left(\frac{k}{k}\right) & \geq & 0. \end{array}$$

10. Optimality Conditions / Equivalence Theorem

We have strong duality if

(a) $H \bullet (XUX^T - \overline{K}) = 0$

(b) $u_i > 0$ only if $x_i^T H x_i = (y_i + Ez_i)^T K^{-1} (y_i + Ez_i) = k$

(c) $H_{YY} = K^{-1}$.

For optimal u, condition (a) implies $E(ZUZ^T) = -(YUZ^T)$ and $K = YUY^T - E(ZUZ^T)E^T$.

We say u is an ϵ-approximate optimal solution if

(a) $(y_i + Ez_i)^T K^{-1} (y_i + Ez_i) \leq (1+\epsilon)k$, $i = 1, \ldots, m$

(b) $u_i > 0$ implies $(y_i + Ez_i)^T K^{-1} (y_i + Ez_i) \geq (1-\epsilon)k$.

11. A Frank-Wolfe Type Algorithm

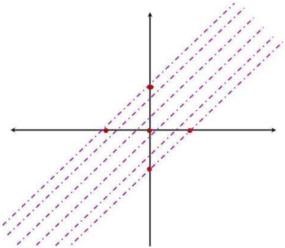
Using $u_+ := (1 - \tau)u + \tau e_i$ and rank-one update formulas lead to an algorithm:

1. Find a feasible u, E and K and calculate $w^k(u)$ where $\frac{\partial g(u)}{\partial u_i} = w_i^k(u) := (y_i + Ez_i)^T K^{-1}(y_i + Ez_i)$.
2. Check for ϵ-approximate optimality.
3. i is chosen to maximize the improvement in the objective function or optimality conditions.
4. Update u to u_+, where step size τ is a solution of a quadratic equation.
5. Update E, K and w^k and go to step 2.

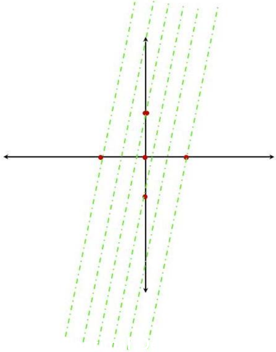
12. Why is D$_k$ criterion harder than D?

Example: Let $X = \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $k = 1$, and $u = [0, 0, 1]$. We have $XUX^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $E(ZUZ^T) = -(YUZ^T)$ becomes $E0 = 0$. For $|E| \leq 1$, this cylinder contains $\mathcal{X}$

$|E| = 1$

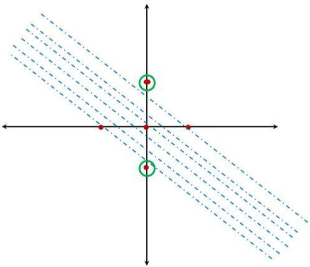


$0 < |E| < 1$



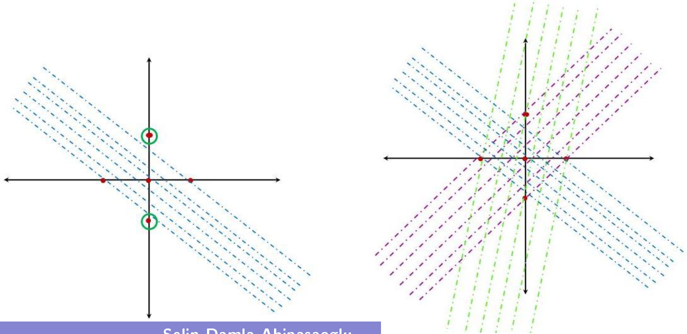
12. Why is D$_k$ criterion harder than D?

Example: Let $X = \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $k = 1$, and $u = [0, 0, 1]$. We have $XUX^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $E(ZUZ^T) = -(YUZ^T)$ becomes $E0 = 0$. For $|E| \leq 1$, this cylinder contains $\mathcal{X}$, but for $|E| > 1$, it does not:



12. Why is D$_k$ criterion harder than D?

Example: Let $X = \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $k = 1$, and $u = [0, 0, 1]$. We have $XUX^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $E(ZUZ^T) = -(YUZ^T)$ becomes $E0 = 0$. For $|E| \leq 1$, this cylinder contains $\mathcal{X}$, but for $|E| > 1$, it does not:

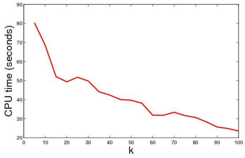


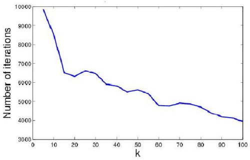
12. Why is D_k criterion harder than D?

- For a given iterate u, when ZUZ^T is not pd, it is hard to choose a matrix E which satisfies $E(ZUZ^T) = -(YUZ^T)$.
- Computational and theoretical complications.
- Modify the algorithm so that ZUZ^T never becomes singular until the last iteration.
- Unlike MVEE, choosing the right pivot is not trivial.
- Choosing a warm-start is hard because of the unknown direction of the axis.

12. Why is D$_k$ criterion harder than D?

- For a given iterate u, when ZUZ^T is not pd, it is hard to choose a matrix E which satisfies $E(ZUZ^T) = -(YUZ^T)$.
- Computational and theoretical complications.
- Modify the algorithm so that ZUZ^T never becomes singular until the last iteration.
- Unlike MVEE, choosing the right pivot is not trivial.
- Choosing a warm-start is hard because of the unknown direction of the axis.

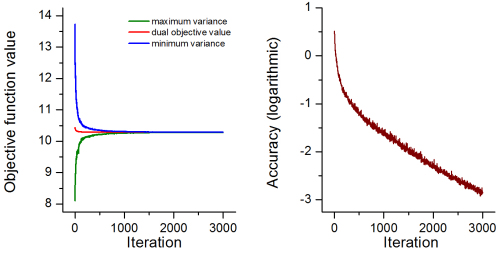




13. Complexity Analysis

Assuming $ZUZ^T \succ 0$, $w(u) < C_1$, $w^k(u) < C_2$, we have:

- $\mathcal{O}(k(m + \ln k + k \ln \ln m + \epsilon^{-1}))$ iterations.
- Each iteration takes $\mathcal{O}(nm)$ operations.
- Local linear convergence similar to D-optimal design: $\mathcal{O}(\tilde{Q} + \tilde{M} \log(\epsilon^{-1}))$ iterations under technical assumptions.
- Away steps are necessary for rapid convergence.



14. Computational Study

Table: Geometric Mean of Running Time and Average Number of Iterations Required by the Algorithm to Obtain an ϵ-Approximate Solution

Dimensions				With Away Steps	
k	n	m	$-\log_{10}\epsilon$	iter	time (sec.)
5	20	20000	6	227	1.125
10	20	20000	6	162	0.7344
15	20	20000	6	136	0.6562
10	50	50000	6	1358	23.375
25	50	50000	6	2132	34.3438
40	50	50000	6	651	9.6719
20	100	100000	6	773	55.7812
50	100	100000	6	997	62.2188
80	100	100000	6	677	37.4219

Conclusions

- First-order methods are very effective and actually necessary to handle very large instances.
- Modification is necessary and very practical for $ZUZ^T \not\succ 0$.
- A good warm-start strategy can be helpful.
- Can we prove any non-trivial core-set results?
- Identify and eliminate non-support points?

Conclusions

- First-order methods are very effective and actually necessary to handle very large instances.
- Modification is necessary and very practical for $ZUZ^T \not\succ 0$.
- A good warm-start strategy can be helpful.
- Can we prove any non-trivial core-set results?
- Identify and eliminate non-support points?

THANK YOU :)