Applications (II)

Lijun Zhang
zlj@nju.edu.cn
http://cs.nju.edu.cn/zlj

Outline

- Experiment Design
  - The Relaxed Problem
  - Scalarization
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Statistical Estimation

- Estimate a Vector

$$y_i = a_i^\top x + w_i, i = 1, \dots, m$$

  - w_i is independent Gaussian random variables with zero mean and unit variance
  - $a_1, \dots, a_m$ span $\mathbf{R}^n$

- Maximum Likelihood Estimate
  - Least-Squares Approximation

$$\min \quad \|Ax - y\|_2^2 = \sum_{i=1}^{m} \left(a_i^\top x - y_i\right)^2$$

Statistical Estimation

Estimate a Vector

$$y_i = a_i^\top x + w_i, i = 1, \dots, m$$

- w_i is independent Gaussian random variables with zero mean and unit variance
- $a_1, \dots, a_m$ span $\mathbf{R}^n$

Maximum Likelihood Estimate

- Least-Squares Approximation

$$\hat{x} = \left(\sum_{i=1}^{m} a_i a_i^\top \right)^{-1} \sum_{i=1}^{m} y_i a_i$$

Statistical Estimation

Estimation Error

$$e = \hat{x} - x$$

- Zero mean, covariance matrix

$$E = \mathbf{E} e e^\top = \left(\sum_{i=1}^{m} a_i a_i^\top \right)^{-1}$$

- E characterizes the accuracy of the estimation
- α-confidence level ellipsoid for x

$$\mathcal{E} = \{ z \,|\, (z - \hat{x})^\top E^{-1} (z - \hat{x}) \le \beta \}$$

  - ✓ β is a constant that depends on n and α

Experiment Design

- Setting
  - We are allowed to choose $a_1, \ldots, a_m$
- Goal
  - Choose $a_1, \ldots, a_m$ such that

$$E = \mathbf{E} e e^\top = \left(\sum_{i=1}^{m} a_i a_i^\top \right)^{-1}$$

is small

- A Special Case of Active Learning

Experiment Design

The Basic Problem

- The menu of possible choices for experiments $v_1, \dots, v_p$
- The total number m of experiments to be carried out
- Let m_j denote the number of experiments that v_j was choose

$$m_1 + \cdots + m_p = m$$

$$E = \left(\sum_{i=1}^{m} a_i a_i^\top \right)^{-1} = \left(\sum_{j=1}^{p} m_j v_j v_j^\top \right)^{-1}$$

Experiment Design

The Basic Problem

- The menu of possible choices for experiments $v_1, \ldots, v_p$
- The total number m of experiments to be carried out
- Let m_j denote the number of experiments that v_j was choose
- Decide the value of m_j to make the error covariance E small

Experiment Design

The Basic Problem

$$\min(\text{w.r.t. } \mathbf{S}_+^n) \quad E = \left(\sum_{j=1}^{p} m_j v_j v_j^\top\right)^{-1}$$

$$\text{s.t.} \quad m_i \geq 0, m_1 + \cdots + m_p = m$$

$$m_i \in \mathbf{Z}$$

- Variable are integers $m_1, \ldots, m_p$
- A vector optimization problem over the positive semidefinite cone
- A hard combinatorial problem

Outline

- Experiment Design
  - The Relaxed Problem
  - Scalarization
- Projection
  - Projection on a Set
  - Projection on a Convex Set

The Relaxed Problem

- Introduce $\lambda_i = m_i/m$

$$\min(\text{w.r.t. } \mathbf{S}_+^{\mathrm{n}}) \quad E = \left(\sum_{j=1}^{p} m_j v_j v_j^{\top}\right)^{-1}$$

$$\text{s.t.} \quad m_i \geq 0, m_1 + \cdots + m_p = m$$

$$m_i \in \mathbf{Z}$$

$$\min(\text{w.r.t. } \mathbf{S}_+^{\mathrm{n}}) \quad E = \frac{1}{m}\left(\sum_{j=1}^{p} \lambda_j v_j v_j^{\top}\right)^{-1}$$

$$\text{s.t.} \quad \lambda_i \geq 0, \lambda_1 + \cdots + \lambda_p = 1$$

$$\lambda_i = \frac{m_i}{m}, m_i \in \mathbf{Z}$$

The Relaxed Problem

- When m is large, a good approximate solution can be found by relaxing $\lambda_i = m_i/m$

$$\min(\text{w.r.t. } \mathbf{S}_+^{\mathrm{n}}) \quad E = \frac{1}{m}\left(\sum_{j=1}^{p} \lambda_j v_j v_j^{\top}\right)^{-1}$$

$$\text{s.t.} \qquad \lambda_i \geq 0, \lambda_1 + \cdots + \lambda_p = 1$$

  - The relaxed experiment design problem
  - A convex optimization problem
  - Provide a lower bound on the optimal value of the combinatorial one

The Relaxed Problem

- Let λ_i be the solution of the relaxed problem
- We can find a approximation solution by

$$m_i = \text{round}(m\lambda_i), \qquad i = 1, \ldots, p$$

- Correspond to this choice of $m_1, \ldots, m_p$ is the vector

$$\tilde{\lambda}_i = \frac{1}{m}\text{round}(m\lambda_i), \qquad i = 1, \ldots, p$$

- When m is large

$$\lambda \approx \tilde{\lambda}, \qquad \text{since } \left|\lambda_i - \tilde{\lambda}_i\right| \leq \frac{1}{2m}, i = 1, \ldots, p$$

Outline

- Experiment Design
  - The Relaxed Problem
  - Scalarization

- Projection
  - Projection on a Set
  - Projection on a Convex Set

Scalarization

- D-optimal Design
  - Minimize the determinant of the error covariance matrix E

$$\begin{array}{ll} \min & \log\det\left(\sum_{i=1}^{p}\lambda_i v_i v_i^\top\right)^{-1} \\ \text{s.t.} & \lambda \succcurlyeq 0, \mathbf{1}^\top\lambda = 1 \end{array}$$

  - Minimize the volume of the resulting confidence ellipsoid
  - A convex optimization problem

Scalarization

- E-optimal Design
  - Minimize the norm of the error covariance matrix, i.e., the maximum eigenvalue of E

$$\begin{aligned} \min \quad & \left\| \left(\sum_{i=1}^{p} \lambda_i v_i v_i^\top \right)^{-1} \right\| \\ \text{s.t.} \quad & \lambda \succcurlyeq 0, 1^\top \lambda = 1 \end{aligned}$$

  - Minimize the diameter of the confidence ellipsoid
  - A convex optimization problem

Scalarization

- E-optimal Design
  - Minimize the norm of the error covariance matrix, i.e., the maximum eigenvalue of E

$$\begin{array}{ll} \min & \left\| \left(\sum_{i=1}^{p} \lambda_i v_i v_i^\top \right)^{-1} \right\| \\ \text{s.t.} & \lambda \succcurlyeq 0, \mathbf{1}^\top \lambda = 1 \end{array} \quad \overset{\text{SDP}}{\Longrightarrow} \quad \begin{array}{ll} \max & t \\ \text{s.t.} & \sum_{i=1}^{p} \lambda_i v_i v_i^\top \succcurlyeq tI \\ & \lambda \succcurlyeq 0, \mathbf{1}^\top \lambda = 1 \end{array}$$

  - Minimize the diameter of the confidence ellipsoid
  - A convex optimization problem

Scalarization

- A-optimal Design
  - Minimize the trace of the error covariance matrix E

$$\begin{array}{ll} \min & \operatorname{tr}\left(\sum_{i=1}^{p} \lambda_i v_i v_i^{\top}\right)^{-1} \\ \text{s.t.} & \lambda \succcurlyeq 0, 1^{\top}\lambda = 1 \end{array}$$

  - Minimize the dimensions of the enclosing box around the confidence ellipsoid
  - A convex optimization problem

Scalarization

□ A-optimal Design

- Minimize the trace of the error covariance matrix E
  - ✓ SDP

$$
\begin{aligned}
\min \quad & 1^\top u \\
\text{s.t.} \quad & \begin{bmatrix} \sum_{i=1}^{p} \lambda_i v_i v_i^\top & e_k \\ e_k^\top & u_k \end{bmatrix} \succcurlyeq 0, k = 1, \dots, n \\
& \lambda \succcurlyeq 0, 1^\top \lambda = 1
\end{aligned}
$$

- Minimize the dimensions of the enclosing box around the confidence ellipsoid
- A convex optimization problem

Optimal Experiment Design and Duality

□ The Dual of D-optimal Design

$$\begin{aligned} \max \quad & \log\det W + n\log n \\ \text{s.t.} \quad & v_i^\top W v_i \leq 1, i = 1, \dots, p \end{aligned}$$

- $W \in \mathbf{S}^n$ and domain $\mathbf{S}_{++}^n$
- W^* determines the minimum volume ellipsoid $\{x | x^\top W^* x \leq 1\}$ that contains $v_1, \dots, v_p$
- Complementary Slackness

$$\lambda_i^*\left(1 - v_i^\top W^* v_i\right) = 0, i = 1, \dots, p$$

- The optimal design only uses the experiments v_i which lie on the surface of the minimum volume ellipsoid

Optimal Experiment Design and Duality

- The Dual of E-optimal Design

$$\begin{array}{ll} \max & \operatorname{tr} W \\ \text{s.t.} & v_i^\top W v_i \leq 1, i = 1, \dots, p \\ & W \succcurlyeq 0 \end{array}$$

  - $W \in \mathbf{S}^n$

- The Dual of A-optimal Design

$$\begin{array}{ll} \max & \left(\operatorname{tr} W^{1/2}\right)^2 \\ \text{s.t.} & v_i^\top W v_i \leq 1, i = 1, \dots, p \end{array}$$

  - $W \in \mathbf{S}^n$ and domain $\mathbf{S}_+^n$

Example

- A Problem with $x \in \mathbf{R}^2$, and $p = 20$

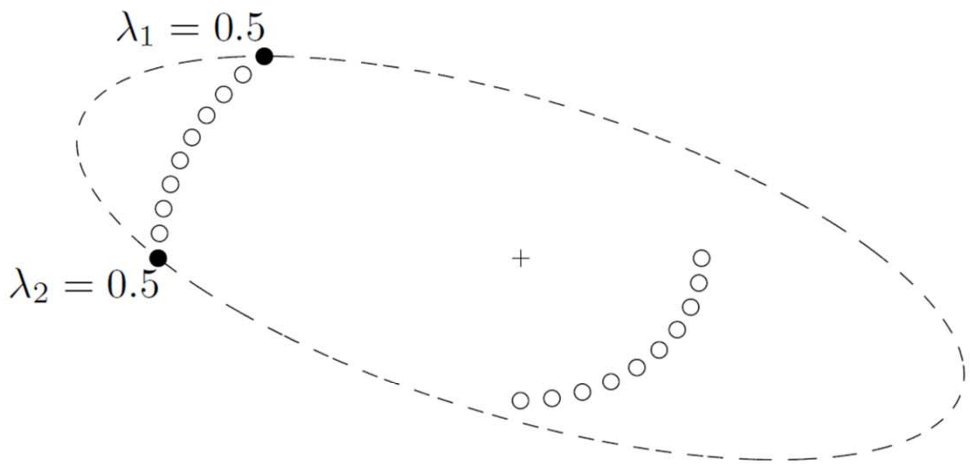

Figure 7.9 Experiment design example. The 20 candidate measurement vectors are indicated with circles. The D-optimal design uses the two measurement vectors indicated with solid circles, and puts an equal weight $\lambda_i = 0.5$ on each of them. The ellipsoid is the minimum volume ellipsoid centered at the origin, that contains the points v_i.

Example

- A Problem with $x \in \mathbf{R}^2$, and $p = 20$

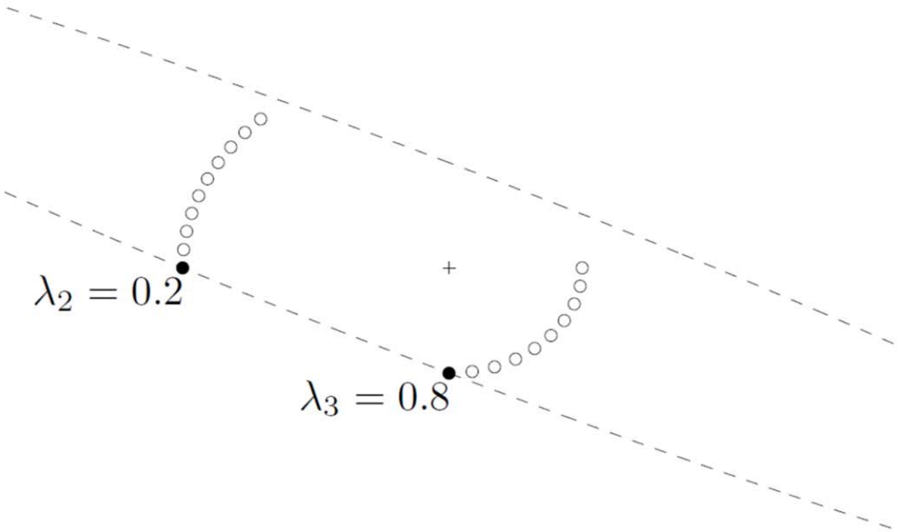

Figure 7.10 The E-optimal design uses two measurement vectors. The dashed lines are (part of) the boundary of the ellipsoid $\{x \mid x^T W^\star x \leq 1\}$ where $W^\star$ is the solution of the dual problem (7.30).

Example

- A Problem with $x \in \mathbf{R}^2$, and $p = 20$

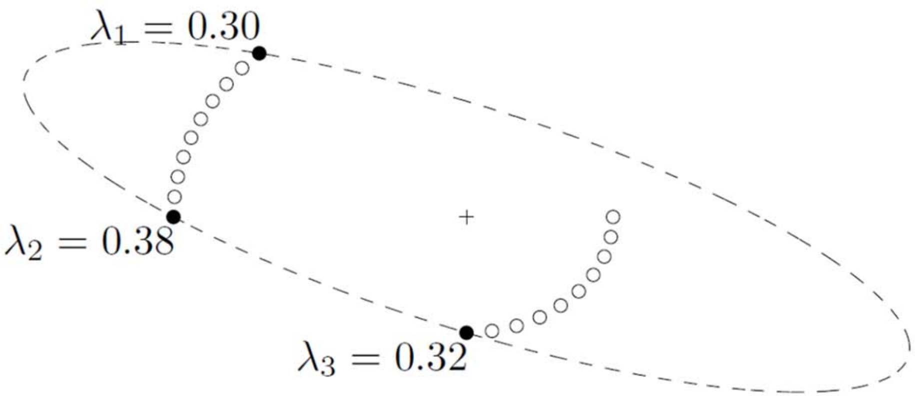

Figure 7.11 The A-optimal design uses three measurement vectors. The dashed line shows the ellipsoid $\{x \mid x^T W^\star x \leq 1\}$ associated with the solution of the dual problem (7.31).

Example

- A Problem with $x \in \mathbf{R}^2$, and $p = 20$

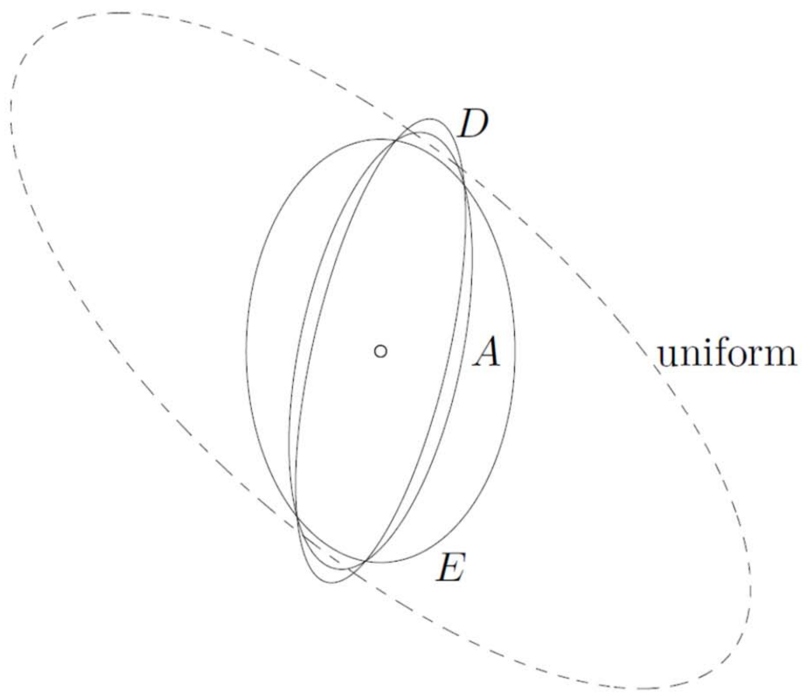

Figure 7.12 Shape of the 90% confidence ellipsoids for D-optimal, A-optimal, E-optimal, and uniform designs.

Outline

- Experiment Design
  - The Relaxed Problem
  - Scalarization
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Projection on a Set

- The distance of a point $x_0 \in \mathbf{R}^n$ to a closed set $C \subseteq \mathbf{R}^n$, in the norm $\|\cdot\|$

$$\operatorname{dist}(x_0, C) = \inf\{\|x_0 - x\| \mid x \in C\}$$

  - The infimum is always achieved
- Projection of x_0 on C
  - Any point $z \in C$ which is closest to x_0

$$\|z - x_0\| = \operatorname{dist}(x_0, C)$$

  - Can be more than one projection of x_0 on C
  - If C is closed and convex, and the norm is strictly convex, there is exactly one

Projection on a Set

- The distance of a point $x_0 \in \mathbf{R}^n$ to a closed set $C \subseteq \mathbf{R}^n$, in the norm $\|\cdot\|$

$$\operatorname{dist}(x_0, C) = \inf\{\|x_0 - x\| \mid x \in C\}$$

  - The infimum is always achieved

- $P_C: \mathbf{R}^n \longrightarrow \mathbf{R}^n$ to denote the projection of x_0 on C

$$P_C(x_0) \in C, \|x_0 - P_C(x_0)\| = \operatorname{dist}(x_0, C)$$

$$P_C(x_0) = \operatorname{argmin}\{\|x - x_0\| \mid x \in C\}$$

  - We refer to P_C as projection on C

Example

- Projection on the Unit Square in $\mathbf{R}^2$
  - Consider the boundary of the unit square in $\mathbf{R}^2$, i.e., $C = \{x \in \mathbf{R}^2 | \|x\|_\infty = 1\}$, take $x_0 = 0$
  - In the ℓ_1-norm, the four points $(1,0)$, $(0,-1)$, $(-1,0)$, and $(0,1)$ are closest to $x_0 = 0$, with distance 1, so we have $\mathrm{dist}(x_0, C) = 1$ in the ℓ_1-norm
  - In the ℓ_∞-norm, all points in C lie at a distance 1 from x_0, and $\mathrm{dist}(x_0, C) = 1$

Example

- Projection onto Rank-k Matrices
  - The set of $m \times n$ matrices with rank less than or equal to k

$$C = \{X \in \mathbf{R}^{m \times n} | \operatorname{rank} X \leq k\}$$

with $k \leq \min\{m, n\}$

  - The Projection of $X_0 \in \mathbf{R}^{m \times n}$ on C in $\|\cdot\|_2$
    - ✓ SVD of X_0

$$X_0 = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$$

$$P_C(x_0) = \sum_{i=1}^{\min\{k,r\}} \sigma_i u_i v_i^\top$$

Outline

- Experiment Design
  - The Relaxed Problem
  - Scalarization
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Projection on a Convex Set

- C is Convex
  - Represent C by a set of linear equalities and convex inequalities

$$Ax = b, \qquad f_i(x) \le 0, i = 1, \dots, m$$

- Projection of x_0 on C

$$\begin{aligned} \min \quad & \|x - x_0\| \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & Ax = b \end{aligned}$$

  - A convex optimization problem
  - Feasible if and only if C is nonempty

Example

Euclidean Projection on a Polyhedron

- Projection of x_0 on $C = \{x | Ax \preccurlyeq b\}$

$$\begin{aligned} \min \quad & \|x - x_0\|_2 \\ \text{s.t.} \quad & Ax \preccurlyeq b \end{aligned}$$

- Projection of x_0 on $C = \{x | a^\top x = b\}$

$$P_C(x_0) = x_0 + \frac{(b - a^\top x_0)a}{\|a\|_2^2}$$

- Projection of x_0 on $C = \{x | a^\top x \leq b\}$

$$P_C(x_0) = \begin{cases} x_0 + \dfrac{(b - a^\top x_0)a}{\|a\|_2^2}, & a^\top x_0 > b \\ x_0, & a^\top x_0 \leq b \end{cases}$$

Example

- Euclidean Projection on a Polyhedron
  - Projection of x_0 on $C = \{x | l \preccurlyeq x \preccurlyeq u\}$

$$P_C(x_0)_k = \begin{cases} l_k, & x_{0k} \le l_k \\ x_{0k}, & l_k \le x_{0k} \le u_k \\ u_k, & u_k \le x_{0k} \end{cases}$$

- Property of Euclidean Projection
  - C is Convex

$$\|P_C(x) - P_C(x)\|_2 \le \|x - y\|_2$$

for all x, y

Example

Euclidean Projection on a Proper Cone

- Projection of x_0 on a proper cone K

$$\begin{array}{ll} \min & \|x - x_0\|_2 \\ \text{s.t.} & x \succcurlyeq_K 0 \end{array}$$

- KKT Conditions

$$x \succcurlyeq_K 0, \qquad x - x_0 = z, \qquad z \succcurlyeq_{K^*} 0, \qquad z^\top x = 0$$

- Introduce $x_+ = x$ and $x_- = z$

$$x_0 = x_+ - x_-, \qquad x_+ \succcurlyeq_K 0, \qquad x_- \succcurlyeq_{K^*} 0, \qquad x_+^\top x_- = 0$$

- Decompose x_0 into two orthogonal elements
  - ✓ One nonnegative with respect to K
  - ✓ The other nonnegative with respect to K^*

Example

- $K = \mathbf{R}_+^n$

$$P_K(x_0)_k = \max\{x_{0k}, 0\}$$

  - Replace each negative component with 0

- $K = \mathbf{S}_+^n$

$$P_K(X_0) = \sum_{i=1}^{n} \max\{0, \lambda_i\} v_i v_i^\top$$

  - The eigendecomposition of X_0 is $X_0 = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$
  - Drop terms associated with negative eigenvalues

Summary

- Experiment Design
  - The Relaxed Problem
  - Scalarization
- Projection
  - Projection on a Set
  - Projection on a Convex Set