EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 16

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

May 24, 2012

Outline

Generalization of Markov and Chebyshev Inequalities

Convex optimization techniques can be used to generalize classical bounds on the probability of a set. Two examples of this are the *Markov inequality* and the closely related *Chebyshev inequality*. Recall that if x is any random variable on $\mathbb{R}$, then the Markov and Chebyshev inequalities respectively state that $\Pr\{|\mathrm{x}| \geq a\} \leq E[|\mathrm{x}|]/a$ and $\Pr\{|\mathrm{x} - \mu_{\mathrm{x}}| \geq a\} \leq \sigma_{\mathrm{x}}^2/a$ for any $a > 0$, where $\mu_{\mathrm{x}} \triangleq E[\mathrm{x}]$ is the mean of x and $\sigma_{\mathrm{x}}^2 \triangleq E\left[(\mathrm{x} - \mu_{\mathrm{x}})^2\right]$ is the variance of x.

Generalization Assumptions:

- $\mathbf{x}$ is a random variable on $\mathcal{S} \subseteq \mathbb{R}^m$ and we would like to bound $\Pr\{\mathbf{x} \in \mathcal{C}\}$, where $\mathcal{C} \subseteq \mathcal{S}$.
- Our prior knowledge is that $E[f_0(\mathbf{x})] = E[1] = a_0 = 1$ and $E[f_i(\mathbf{x})] = a_i$ for $i = 1, \ldots, n$.

Generalization Approach:

- Construct the linear combination $f(\mathbf{z}) = \sum_{i=0}^{n} v_i f_i(\mathbf{z})$ and ensure that $f(\mathbf{z}) \geq \mathbf{1}_{\mathcal{C}}(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{S}$, where $\mathbf{1}_{\mathcal{C}}(\mathbf{z})$ is the *indicator function* of $\mathcal{C}$ (i.e., $\mathbf{1}_{\mathcal{C}}(\mathbf{z}) = 1$ if $\mathbf{z} \in \mathcal{C}$ and 0 if $z \notin \mathcal{C}$).
- If this holds, then $\Pr\{\mathbf{x} \in \mathcal{C}\} \leq E[f(\mathbf{x})] = \mathbf{a}^T \mathbf{v}$.

Generalized Markov/Chebyshev Bound Problem:

$$\begin{array}{ll} \text{minimize} & v_0 + a_1 v_1 + \cdots + a_n v_n \\ \text{subject to} & f(\mathbf{z}) = \sum_{i=0}^{n} v_i f(\mathbf{z}) \geq 1 \text{ for } \mathbf{z} \in \mathcal{C} \\ & f(\mathbf{z}) = \sum_{i=0}^{n} v_i f(\mathbf{z}) \geq 0 \text{ for } \mathbf{z} \in \mathcal{S}\,,\ \mathbf{z} \notin \mathcal{C} \end{array} .$$

- Yields the best upper bound on $\Pr\{\mathbf{x} \in \mathcal{C}\}$ given the assumptions.
- This problem is always convex as the constraints can be posed in terms of convex functions:

$$g_1(\mathbf{v}) = 1 - \inf_{\mathbf{z} \in \mathcal{C}} f(\mathbf{z}) \leq 0\,,\ g_2(\mathbf{v}) = - \inf_{\mathbf{z} \in \mathcal{S} \setminus \mathcal{C}} f(\mathbf{z}) \leq 0\,.$$

Generalization of Chernoff Bound

Convex optimization can also be used to find an alternate bound on the probability of a set which leads to a generalization of the *Chernoff bound*. Recall that if x is a random variable on $\mathbb{R}$, the Chernoff bound states the following.

$$\Pr\{\mathrm{x} \geq u\} \leq \inf_{\lambda \geq 0} \left\{ E\left[e^{\lambda(\mathrm{x}-u)} \right] \right\} \iff \log \Pr\{\mathrm{x} \geq u\} \leq \inf_{\lambda \geq 0} \left\{ -\lambda u + \log E\left[e^{\lambda \mathrm{x}} \right] \right\} .$$

To generalize this, we assume as before that $\mathbf{x}$ is a random variable on $\mathcal{S} \subseteq \mathbb{R}^m$ and that we would like to bound $\Pr\{\mathbf{x} \in \mathcal{C}\}$, where $\mathcal{C} \subseteq \mathcal{S}$. Then, for some $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$, we consider a function $f : \mathbb{R}^m \to \mathbb{R}$ given by

$$f(\mathbf{z}) = e^{\boldsymbol{\lambda}^T \mathbf{z} + \mu} ,$$

and choose $\boldsymbol{\lambda}$ and μ to ensure that $f(\mathbf{z}) \geq \mathbf{1}_{\mathcal{C}}(\mathbf{z})$ for all $\mathbf{z}$. This leads to the following generalization of the Chernoff bound.

Generalized Chernoff Bound:

$$\log \Pr\{\mathbf{x} \in \mathcal{C}\} \leq \inf_{\boldsymbol{\lambda}} \left\{ S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \log E\left[e^{\boldsymbol{\lambda}^T \mathbf{x}} \right] \right\} .$$

- $S_{\mathcal{C}}(\mathbf{y})$ is the *support function* given by $S_{\mathcal{C}}(\mathbf{y}) = \sup \left\{ \mathbf{y}^T \mathbf{w} : \mathbf{w} \in \mathcal{C} \right\}$.
- $\log E\left[e^{\boldsymbol{\lambda}^T \mathbf{x}} \right]$ is the *cumulant generating function* of the distribution.
- Evaluating this bound is, in general, a convex optimization problem.

Minimum Distance Detector Example

To illustrate the probability bounding capabilities possible with convex optimization, we consider a symbol detection example from digital communications.

- A symbol $\mathbf{s} \in \mathbb{R}^n$ from the set $\{\mathbf{s}_1, \ldots, \mathbf{s}_m\} \subseteq \mathbb{R}^n$ (called the *signal constellation*) is transmitted over a noisy channel.
- The received signal is $\mathbf{x} = \mathbf{s} + \mathbf{v}$, where $\mathbf{v}$ is a random variable with $E[\mathbf{v}] = \mathbf{0}_{n \times 1}$ and $E\left[\mathbf{v}\mathbf{v}^T\right] = \sigma^2 \mathbf{I}_n$.
- To estimate which symbol was sent, we use a *minimum distance detector*, which operates as follows. The space $\mathbb{R}^n$ is partitioned into *Voronoi regions* $\mathcal{V}_k$ given by

$$\begin{aligned} \mathcal{V}_k &= \left\{\mathbf{y} \in \mathbb{R}^n : ||\mathbf{y} - \mathbf{s}_k||_2 \leq ||\mathbf{y} - \mathbf{s}_i||_2 \ , \ i \neq k\right\} \ , \ k = 1, \ldots, m \, , \\ &= \left\{\mathbf{y} \in \mathbb{R}^n : 2\left(\mathbf{s}_i - \mathbf{s}_k\right)^T \mathbf{y} \leq ||\mathbf{s}_i||_2^2 - ||\mathbf{s}_k||_2^2 \ , \ i \neq k\right\} \ , \ k = 1, \ldots, m \, . \end{aligned}$$

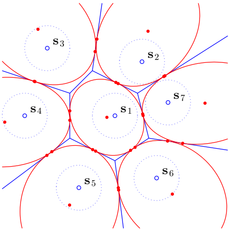
  If $\mathbf{x} \in \mathcal{V}_k$, we estimate the transmitted symbol as $\widehat{\mathbf{s}} = \mathbf{s}_k$.
- If symbol $\mathbf{s}_\ell$ was transmitted, then the probability of correct detection is $\Pr\{\mathbf{s}_\ell + \mathbf{v} \in \mathcal{V}_\ell\}$.
- When the noise $\mathbf{v}$ is Gaussian, then minimum distance decoding is the same as maximum likelihood (ML) decoding.

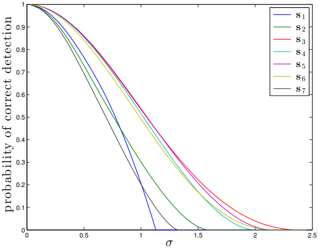
Minimum Distance Detector Example (Continued)

Example: $(m = 7, n = 2)$

Chebyshev bounds: (assuming $E[\mathbf{v}] = \mathbf{0}$ and $E\left[\mathbf{v}\mathbf{v}^T\right] = \sigma^2 \mathbf{I}$)

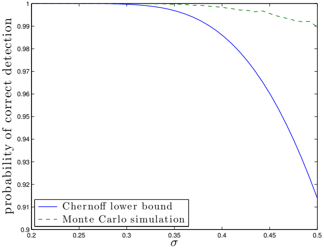
- Left: Constellation, Voronoi regions, and Chebyshev lower bound based ellipses for $\sigma = 1$.
- Right: Chebyshev lower bounds on the probability of correct detection as a function of σ.





Chernoff bounds: (assuming the noise is Gaussian)

- Chernoff lower bound (solid line) and Monte Carlo simulation (dashed line) for the probability of correct detection of $\mathbf{s}_1$.



Introduction to Experiment Design

Suppose we want to estimate a vector $\mathbf{x} \in \mathbb{R}^n$ given m measurments of the form

$$y_i = \mathbf{a}_i^T \mathbf{x} + w_i \,,\; i = 1, \ldots, m \,.$$

Here, we assume that the measurement errors w_i are i.i.d. with distribution $\mathcal{N}(0,1)$.

- The ML estimate is the least-squares (LS) solution

$$\widehat{\mathbf{x}} = \left(\sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^T \right)^{-1} \sum_{i=1}^{m} y_i \mathbf{a}_i \,.$$

- The error $\mathbf{e} = \widehat{\mathbf{x}} - \mathbf{x}$ has zero mean and covariance matrix

$$\mathbf{E} = E\left[\mathbf{e}\mathbf{e}^T\right] = \left(\sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^T \right)^{-1} \,.$$

- The α-confidence level ellipsoid for $\mathbf{x}$ is given by

$$\mathcal{E} = \left\{ \mathbf{z} : (\mathbf{z} - \widehat{\mathbf{x}})^T \, \mathbf{E}^{-1} \, (\mathbf{z} - \widehat{\mathbf{x}}) \leq \beta \right\} \,,$$

  where β is a constant that depends on α, m, and n.

Experiment Design: In *experiment design*, we choose $\mathbf{a}_i \in \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ (a set of possible test vectors) to make $\mathbf{E}$ 'small' in some sense.

Experiment Design Problem and its Relaxation

Let m_k for $k = 1, \ldots, p$ denote the number of experiments $\mathbf{a}_i$ equal to $\mathbf{v}_k$. Then, we have $m_1 + \cdots + m_p = m$, and also

$$\mathbf{E} = \left(\sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^T\right)^{-1} = \left(\sum_{k=1}^{p} m_k \mathbf{v}_k \mathbf{v}_k^T\right)^{-1} .$$

From this, we get the following formulation of the experiment design problem.

Vector/Matrix Optimization Formulation of the Experiment Design Problem:

$$\begin{array}{ll} \text{minimize (with respect to } \mathbb{S}_+^n) & \mathbf{E} = \left(\sum_{k=1}^{p} m_k \mathbf{v}_k \mathbf{v}_k^T\right)^{-1} \\ \text{subject to} & m_k \geq 0 \, , \; m_1 + \cdots + m_p = m \\ & m_k \in \mathbb{Z} \end{array} .$$

- This is an NP-complete combinatorial problem due to the integer constraint $m_k \in \mathbb{Z}$.

When $m \gg p$, then if $\lambda_k \triangleq m_k/m$ denotes the fraction of experiments equal to $\mathbf{v}_k$, we can approximately treat λ_k as a continuous real variable. This yields the following convex relaxation.

Relaxed Experiment Design Problem:

$$\begin{array}{ll} \text{minimize (with respect to } \mathbb{S}_+^n) & \mathbf{E} = (1/m)\left(\sum_{k=1}^{p} \lambda_k \mathbf{v}_k \mathbf{v}_k^T\right)^{-1} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \, , \; \mathbf{1}^T \boldsymbol{\lambda} = 1 \end{array} .$$

- We can obtain a suboptimal solution to the original problem by setting $m_k = \text{round}(m\lambda_k)$.
- There are multiple ways to scalarize the problem, which lead to different optimal designs.

Determinant (*D*)-Optimal Design

In *determinant* or *D*-optimal design, the scalarization used for the relaxed experiment design problem is such that the (logarithm of the) determinant of $\mathbf{E}$ is minimized.

D-Optimal Design Problem: (Primal)

$$\begin{array}{ll} \text{minimize} & \log\det\left(\sum_{k=1}^{p}\lambda_k \mathbf{v}_k \mathbf{v}_k^T\right)^{-1} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0}\,,\ \mathbf{1}^T\boldsymbol{\lambda} = 1 \end{array} .$$

Interpretation: This minimizes the *volume* of the confidence ellipsoids.

D-Optimal Design Problem: (Dual)

$$\begin{array}{ll} \text{maximize} & \log\det\mathbf{W} + n\log n \\ \text{subject to} & \mathbf{v}_k^T\mathbf{W}\mathbf{v}_k \le 1\,,\ k = 1,\ldots,p \end{array} .$$

Interpretation: $\left\{\mathbf{x} : \mathbf{x}^T\mathbf{W}^\star\mathbf{x} \le 1\right\}$ is the minimum volume ellipsoid centered at the origin, that includes all test vectors $\mathbf{v}_k$.

Complementary Slackness:

For primal optimal $\boldsymbol{\lambda}^\star$ and dual optimal $\mathbf{W}^\star$, we have

$$\lambda_k^\star\left(1 - \mathbf{v}_k^T\mathbf{W}^\star\mathbf{v}_k\right) = 0\,,\ k = 1,\ldots,p\,.$$

The optimal experiment only uses vectors $\mathbf{v}_k$ on the boundary of the ellipsoid defined by $\mathbf{W}^\star$.

Eigenvalue (*E*)-Optimal Design

In *eigenvalue* or *E*-optimal design, the scalarization used for the relaxed experiment design problem is such that the maximum eigenvalue of $\mathbf{E}$ is minimized.

E-Optimal Design Problem: (Primal)

$$\begin{array}{ll} \text{minimize} & \left\| \left(\sum_{k=1}^{p} \lambda_k \mathbf{v}_k \mathbf{v}_k^T \right)^{-1} \right\|_2 \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \,,\ \mathbf{1}^T \boldsymbol{\lambda} = 1 \end{array} \iff \begin{array}{ll} \text{maximize} & t \\ \text{subject to} & \sum_{k=1}^{p} \lambda_k \mathbf{v}_k \mathbf{v}_k^T \succeq t\mathbf{I} \\ & \boldsymbol{\lambda} \succeq \mathbf{0} \,,\ \mathbf{1}^T \boldsymbol{\lambda} = 1 \end{array} .$$

Interpretation: This minimizes the *diameter* of the confidence ellipsoids.

E-Optimal Design Problem: (Dual)

$$\begin{array}{ll} \text{maximize} & \operatorname{tr} \mathbf{W} \\ \text{subject to} & \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \leq 1 \,,\ k = 1, \ldots, p \\ & \mathbf{W} \succeq \mathbf{0} \end{array} .$$

Complementary Slackness:

For primal optimal $\boldsymbol{\lambda}^\star$ and dual optimal $\mathbf{W}^\star$, we have

$$\lambda_k^\star \left(1 - \mathbf{v}_k^T \mathbf{W}^\star \mathbf{v}_k \right) = 0 \,,\ k = 1, \ldots, p \,.$$

The optimal experiment only uses vectors $\mathbf{v}_k$ on the boundary of the ellipsoid defined by $\mathbf{W}^\star$.

Average (A)-Optimal Design

In *average* or A-optimal design, the scalarization used for the relaxed experiment design problem is such that the trace of $\mathbf{E}$ is minimized.

A-Optimal Design Problem: (Primal)

$$\begin{array}{ll} \text{minimize} & \operatorname{tr}\left(\sum_{k=1}^{p} \lambda_k \mathbf{v}_k \mathbf{v}_k^T\right)^{-1} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0}\,,\ \mathbf{1}^T\boldsymbol{\lambda} = 1 \end{array} \iff \begin{array}{ll} \text{maximize} & \mathbf{1}^T\mathbf{u} \\ \text{subject to} & \begin{bmatrix} \sum_{k=1}^{p} \lambda_k \mathbf{v}_k \mathbf{v}_k^T & \mathbf{e}_i \\ \mathbf{e}_i^T & u_i \end{bmatrix} \succeq \mathbf{0}\ \forall\, i \\ & \boldsymbol{\lambda} \succeq \mathbf{0}\,,\ \mathbf{1}^T\boldsymbol{\lambda} = 1 \end{array} .$$

A-Optimal Design Problem: (Dual)

$$\begin{array}{ll} \text{maximize} & \left(\operatorname{tr} \mathbf{W}^{1/2}\right)^2 \\ \text{subject to} & \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \le 1\,,\ k = 1, \ldots, p \end{array} .$$

There is an implicit constraint that $\mathbf{W} \in \mathbb{S}_+^n$ here.

Complementary Slackness:

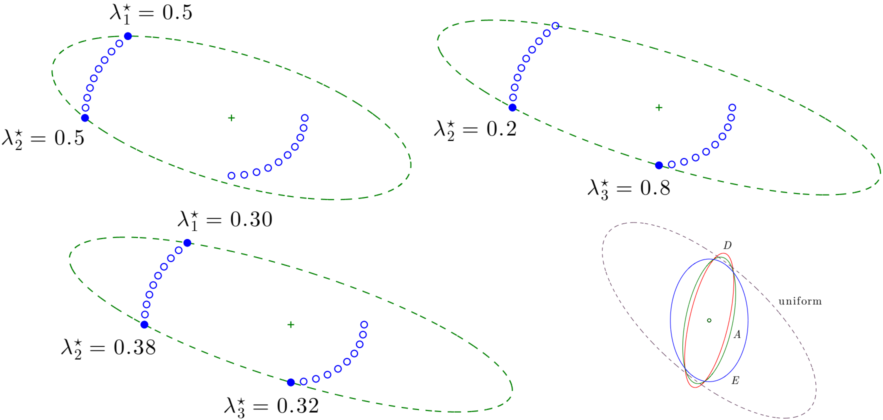
For primal optimal $\boldsymbol{\lambda}^\star$ and dual optimal $\mathbf{W}^\star$, we have

$$\lambda_k^\star \left(1 - \mathbf{v}_k^T \mathbf{W}^\star \mathbf{v}_k\right) = 0\,,\ k = 1, \ldots, p\,.$$

The optimal experiment only uses vectors $\mathbf{v}_k$ on the boundary of the ellipsoid defined by $\mathbf{W}^\star$.

Experiment Design Example

$\mathbf{x} \in \mathbb{R}^2$, $p = 20$



- Top left: *D*-optimal design $\boldsymbol{\lambda}^\star$ and $\mathbf{W}^\star$.
- Top right: *E*-optimal design $\boldsymbol{\lambda}^\star$ and $\mathbf{W}^\star$.
- Bottom left: *A*-optimal design $\boldsymbol{\lambda}^\star$ and $\mathbf{W}^\star$.
- Bottom right: Shape of 90% confidence ellipsoids for *D*-optimal, *E*-optimal, and *A*-optimal, and uniform designs.

Extensions to Experiment Design Problem

Resource limits:

- Associate a cost c_k for experiment $\mathbf{v}_k$. This could represent the economic cost or time required to carry out $\mathbf{v}_k$.
- The total cost is then
$$m_1 c_1 + \cdots + m_p c_p = m\mathbf{c}^T \boldsymbol{\lambda}\,.$$
- Can add a limit on the total cost with the affine inequality constraint $m\mathbf{c}^T \boldsymbol{\lambda} \leq B$, where B is a budget.

Multiple measurements per experiment: $\mathbf{v}_k \in \mathbb{R}^{n \times i_k}$, with
$$\mathbf{v}_k = \left[\begin{array}{ccc} \mathbf{u}_{k,1} & \cdots & \mathbf{u}_{k,i_k} \end{array} \right] ,$$
where i_k is the number of scalar measurements obtained when experiment $\mathbf{v}_k$ is carried out.

- Can model discounts or time savings associated with performing groups of measurements simultaneously.
- For example, if the cost of making measurements $\mathbf{v}_1$ and $\mathbf{v}_2$ together is less than the sum of making them separately, we take $\mathbf{v}_3$ to be
$$\mathbf{v}_3 = \left[\begin{array}{cc} \mathbf{v}_1 & \mathbf{v}_2 \end{array} \right] ,$$
and assign costs c_1, c_2, and c_3 such that $c_3 < c_1 + c_2$.