Using the S-Lemma to Design Robust Experiments

Guillaume Sagnol

TU Berlin

Design of Experiments: New Challenges
CIRM Workshop, Marseille, May 3, 2018

Technische Universität Berlin

Design of Experiment

- $X \subset \mathbb{R}^d$: compact design space

An experiment with N trials is defined by a *design*

$$\xi = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ N_1 & \cdots & N_n \end{array} \right\},$$

where

- $\boldsymbol{x}_i \in X$ is the ith *support point* of the design
- $N_i \in \mathbb{N}$ is the replication at the ith design point
- $\sum_{i=1}^{s} N_i = N$.

Design of Experiment

- $X \subset \mathbb{R}^d$: compact design space

When $N \to \infty$, we can consider *approximate designs*:

$$\xi = \left\{ \begin{array}{ccc} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ w_1 & \cdots & w_n \end{array} \right\},$$

where $w_i \in \mathbb{R}_+$ is the proportion of the total number of trials at ith design point, and $\sum_{i=1}^n w_i = 1$.

In this work, we assume that the candidates design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are fixed, so the set of all approximate designs is isomorphic to

$$\mathcal{W} := \{\boldsymbol{w} \geq \boldsymbol{0} : \sum_{i=1}^m w_i = 1\}.$$

The Linear Model

A trial at the design point $\boldsymbol{x} \in X$ provides an observation

$$y = f(\boldsymbol{x})^T \boldsymbol{\theta} + \epsilon,$$

where

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ is an *unknown* vector of parameters;
- $f : X \mapsto \mathbb{R}^m$ is known;
- $\mathbb{E}[\epsilon] = \mathbf{0}, \quad \mathbb{V}[\epsilon] = \sigma^2$ (a known constant), and the noises ϵ, ϵ' of two distinct trials are uncorrelated.

Standard approaches minimize a convex functional of the *information matrix* of the design ξ,

$$M(\xi) := \sum_{i=1}^{s} w_i\, f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_i)^T \in \mathbb{S}_m^+.$$

The Linear Model

A trial at the design point $\boldsymbol{x} \in X$ provides an observation

$$y = f(\boldsymbol{x})^T \boldsymbol{\theta} + \epsilon,$$

where

- $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ is an *unknown* vector of parameters;
- ~~$f : X \mapsto \mathbb{R}^m$ is known;~~
- $\mathbb{E}[\epsilon] = \mathbf{0}, \quad \mathbb{V}[\epsilon] = \sigma^2$ (a known constant), and the noises ϵ, ϵ' of two distinct trials are uncorrelated.

The function f is not always known accurately

Linear model $y = f(\boldsymbol{x})^T\boldsymbol{\theta} + \epsilon$.

1. Error-in-variables Models
   - Instead of observing $y = f(\boldsymbol{x})^T\boldsymbol{\theta} + \epsilon$, the experimenter measures

     $$y = f(\boldsymbol{x} + \boldsymbol{\eta})^T\boldsymbol{\theta} + \epsilon,$$

     where $\boldsymbol{\eta}$ is an unknown noise.
   - Model studied in [Konstantinou & Dette, 2015], for the case of ML estimation and LS estimation.

The function f is not always known accurately

Linear model $y = f(\boldsymbol{x})^T \boldsymbol{\theta} + \epsilon$.

2. The assumed model is *Nonlinear*
   - $y = g(\boldsymbol{x}, \boldsymbol{\theta}) + \epsilon$
   - Standard approach: *local optimal design*. The model is linearized around $\boldsymbol{\theta_0}$, and we compute an optimal design for the linear model

$$y \simeq f(\boldsymbol{x})^T \boldsymbol{\theta} + \epsilon,$$

     where $f(\boldsymbol{x}) := \nabla g(\boldsymbol{x}, \boldsymbol{\theta_0})$
   - But wrong choice of θ_0 leads to an error in the regressor function f.

The function f is not always known accurately

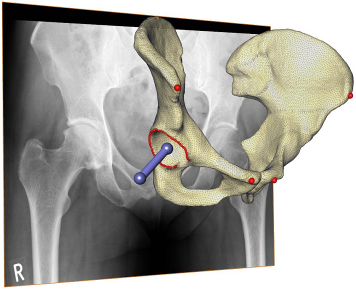
Linear model $y = f(\boldsymbol{x})^T\boldsymbol{\theta} + \epsilon$.

3. Design for computer experiments with a GP surogate.
   - $y = \eta(\boldsymbol{x}) + \epsilon$, where $\eta(\boldsymbol{x})$ is the realization of a Gaussian process with a known semidefinite covariance kernel $K(\cdot, \cdot)$.
   - We can reduce to a linear model by truncating the Karhunen–Loève expansion of the kernel
   - But in practice, the resulting linear model depends on the eigenfunctions of K, which must be estimated using the Nyström approximation, and estimates of Kernel hyperparameters.

The function f is not always known accurately

Linear model $y = f(\boldsymbol{x})^T\boldsymbol{\theta} + \epsilon$.

4. X-ray based Anatomy Reconstruction with Low Radiation Exposure [ongoing work with Jentsch & Weiser]



- Goal: estimation of geometry parameters of the patient's anatomy
- Design: there is a "budget" of exposure to distribute over diffenret projection angles $\boldsymbol{x} \in X$
- Computing the linearized model $f(\boldsymbol{x})$ requires multidimensional integrals, typically approximated with quadratures.

Outline

Robust Linear Model

Linear model in vector form

$$\boldsymbol{y} = A\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where

- $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n, \quad A = \begin{bmatrix} f(\boldsymbol{x}_1)^T \\ f(\boldsymbol{x}_2)^T \\ \vdots \\ f(\boldsymbol{x}_n)^T \end{bmatrix} \in \mathbb{R}^{n \times m}.$
- $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \sigma^2 \operatorname{Diag}(\boldsymbol{w})^{-1}.$
- The matrix A is not known, but assumed to lie in the *ball*
$$\mathcal{A} := \{A_0 + \Delta \mid \|\Delta\| \le \delta\}.$$
- The unknown parameter $\boldsymbol{\theta}$ is assumed to lie in an ellipsoid $\Theta = \{\boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} \le 1\}$.

Estimators for the robust linear model

- Estimators for the robust linear model have been proposed in [El Ghaoui & Lebret 1997, Calafiore & El Ghaoui 2001, Eldar, Ben-Tal & Nemirovski 2005]
- Approaches based on Semidefinite Programming formulations using the *S-Lemma*
- In this talk, we extend this work
  - Goal: simultaneous computation of a robust estimator, and optimal design weights $\boldsymbol{w}$
  - We obtain robust designs for estimation of $\boldsymbol{\theta}$, and for prediction of $f(\boldsymbol{x})^T\boldsymbol{\theta}$ at unsampled locations $\boldsymbol{x}$'s.

A robust criterion

Consider the linear estimator

$$\hat{\boldsymbol{\theta}} = G\boldsymbol{y}$$

We introduce a criterion depending on both the coefficients G and the design weights $\boldsymbol{w} \in \mathcal{W}$:

$$\begin{aligned}
\phi(G, \boldsymbol{w}) &= \sup_{A \in \mathcal{A}} \sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2] \\
&= \sup_{A \in \mathcal{A}} \sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E}[\|G(A\boldsymbol{\theta} + \boldsymbol{\epsilon}) - \boldsymbol{\theta}\|^2] \\
&= \sup_{A \in \mathcal{A}} \sup_{\boldsymbol{\theta} \in \Theta} \|(GA - I)\boldsymbol{\theta}\|^2 + \sigma^2 \operatorname{trace} G \operatorname{Diag}(\boldsymbol{w})^{-1} G^T \\
&= \sup_{A \in \mathcal{A}} \lambda_{\max}\Big((GA - I)^T \Sigma (GA - I)\Big) + \sigma^2 \sum_{i=1}^{n} \frac{\|\boldsymbol{g}_i\|^2}{w_i}
\end{aligned}$$

Outline

The S-Lemma

S-lemma (homogeneous version) [Yakubovich 71]

Let Q_1, and Q_2 be two quadratic forms over $\mathbb{R}^n$ and assume that $\exists \boldsymbol{x}_0 \in \mathbb{R}^n : Q_1(\boldsymbol{x}_0) > 0$. Then, TFAE

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \quad \big(Q_1(\boldsymbol{x}) \geq 0 \implies Q_2(\boldsymbol{x}) \geq 0\big)$$

$$\exists \lambda \geq 0 : \quad Q_2(\boldsymbol{x}) \geq \lambda Q_1(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^n.$$

We can reformulate the S-lemma as follows: Let M_1, M_2 be symmetric matrices of size n, and let:

$$\begin{aligned} v^* := \inf \quad & \boldsymbol{x}^T M_2 \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x}^T M_1 \boldsymbol{x} \geq 0. \end{aligned}$$

Then,

$$v^* \geq 0 \quad \iff \quad \exists \lambda \geq 0 : \; M_2 - \lambda M_1 \succeq 0.$$

Consequence of the S-Lemma

Theorem (Ben-Tal & Nemirovski, 1998)

The linear matrix inequality (with variables M and L)

$$M + L\Delta R + R^T\Delta^T L^T \succeq 0$$

holds for all Δ *such that* $\|\Delta\| \leq \delta$ *iff*

$$\exists \lambda \geq 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0.$$

Consequence of the S-Lemma

$$M+L\Delta R+R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \iff \exists \lambda \geq 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta$$

Consequence of the S-Lemma

$$M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \iff \exists \lambda \geq 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta$$

$$\iff \boldsymbol{y}^T(M + L\Delta R + R^T\Delta^T L^T)\boldsymbol{y} \geq 0, \quad \forall \|\Delta\| \leq \delta, \ \forall \boldsymbol{y} \in \mathbb{R}^n$$

Consequence of the S-Lemma

$$M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \iff \exists \lambda \ge 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$\begin{aligned} & M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \\ \iff & \boldsymbol{y}^T(M + L\Delta R + R^T\Delta^T L^T)\boldsymbol{y} \ge 0, \quad \forall \|\Delta\| \le \delta,\ \forall \boldsymbol{y} \in \mathbb{R}^n \\ \iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\|\Delta\| \le \delta} (L^T\boldsymbol{y})^T \Delta (R\boldsymbol{y}) \ge 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \end{aligned}$$

Consequence of the S-Lemma

$$M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \iff \exists \lambda \geq 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$\begin{aligned}
& M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \\
\iff & \boldsymbol{y}^T (M + L\Delta R + R^T \Delta^T L^T) \boldsymbol{y} \geq 0, \quad \forall \|\Delta\| \leq \delta, \ \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\|\Delta\| \leq \delta} (L^T \boldsymbol{y})^T \Delta (R\boldsymbol{y}) \geq 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2(-\delta \cdot \|L^T \boldsymbol{y}\| \cdot \|R\boldsymbol{y}\|) \geq 0 \quad \forall \boldsymbol{y} \in \mathbb{R}^n
\end{aligned}$$

Consequence of the S-Lemma

$$M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \iff \exists \lambda \ge 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$\begin{aligned}
& M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \\
\iff & \boldsymbol{y}^T (M + L\Delta R + R^T \Delta^T L^T) \boldsymbol{y} \ge 0, \quad \forall \|\Delta\| \le \delta, \ \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\|\Delta\| \le \delta} (L^T \boldsymbol{y})^T \Delta (R\boldsymbol{y}) \ge 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2(-\delta \cdot \|L^T \boldsymbol{y}\| \cdot \|R\boldsymbol{y}\|) \ge 0 \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\{\boldsymbol{z} : \|\boldsymbol{z}\| \le \delta \|R\boldsymbol{y}\|\}} \boldsymbol{y}^T L \boldsymbol{z} \ge 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n
\end{aligned}$$

Consequence of the S-Lemma

$$M+L\Delta R+R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \iff \exists \lambda \geq 0 : \begin{pmatrix} M-\lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

Proof.

$$\begin{aligned}
& M + L\Delta R + R^T\Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \leq \delta \\
\iff & \boldsymbol{y}^T(M + L\Delta R + R^T\Delta^T L^T)\boldsymbol{y} \geq 0, \quad \forall \|\Delta\| \leq \delta, \ \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M\boldsymbol{y} + 2 \inf_{\|\Delta\|\leq\delta} (L^T\boldsymbol{y})^T\Delta(R\boldsymbol{y}) \geq 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M\boldsymbol{y} + 2(-\delta \cdot \|L^T\boldsymbol{y}\| \cdot \|R\boldsymbol{y}\|) \geq 0 \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M\boldsymbol{y} + 2 \inf_{\{\boldsymbol{z}:\|\boldsymbol{z}\|\leq\delta\|R\boldsymbol{y}\|\}} \boldsymbol{y}^T L\boldsymbol{z} \geq 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \left((\|\boldsymbol{z}\| \leq \delta\|R\boldsymbol{y}\|) \implies (\boldsymbol{y}^T M\boldsymbol{y} + 2\boldsymbol{y}^T L\boldsymbol{z} \geq 0) \right)
\end{aligned}$$

Consequence of the S-Lemma

$$M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \iff \exists \lambda \ge 0 : \begin{pmatrix} M - \lambda\delta^2 R^T R & L \\ L^T & \lambda I \end{pmatrix} \succeq 0$$

$$\begin{aligned}
& M + L\Delta R + R^T \Delta^T L^T \succeq 0, \quad \forall \|\Delta\| \le \delta \\
\iff & \boldsymbol{y}^T (M + L\Delta R + R^T \Delta^T L^T) \boldsymbol{y} \ge 0, \quad \forall \|\Delta\| \le \delta,\ \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\|\Delta\| \le \delta} (L^T \boldsymbol{y})^T \Delta (R\boldsymbol{y}) \ge 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2(-\delta \cdot \|L^T \boldsymbol{y}\| \cdot \|R\boldsymbol{y}\|) \ge 0 \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \boldsymbol{y}^T M \boldsymbol{y} + 2 \inf_{\{\boldsymbol{z} : \|\boldsymbol{z}\| \le \delta \|R\boldsymbol{y}\|\}} \boldsymbol{y}^T L \boldsymbol{z} \ge 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n \\
\iff & \left((\|\boldsymbol{z}\| \le \delta \|R\boldsymbol{y}\|) \implies (\boldsymbol{y}^T M \boldsymbol{y} + 2\boldsymbol{y}^T L \boldsymbol{z} \ge 0) \right) \\
\iff & \left[\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix}^T \begin{pmatrix} \delta^2 R^T R & \\ & -I \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \ge 0 \right. \\
& \qquad \left. \implies \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix}^T \begin{pmatrix} M & L \\ L^T & \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \ge 0 \right]
\end{aligned}$$

Outline

SDP formulation

Recall that we want to minimize

$$\phi(G, \boldsymbol{w}) = \sup_{A \in \mathcal{A}} \quad \lambda_{\max}\Big((GA - I)^T \Sigma (GA - I)\Big) + \sigma^2 \sum_{i=1}^{n} \frac{\|\boldsymbol{g}_i\|^2}{w_i}.$$

This is the same as minimizing $t + \sigma^2 u$ under the constraints

- $\lambda_{\max}\Big((GA - I)^T \Sigma (GA - I)\Big) \leq t, \quad \forall A \in \mathcal{A}$
- $\sum_{i=1}^{n} \frac{\|\boldsymbol{g}_i\|^2}{w_i} \leq u.$

SDP formulation

- $\lambda_{\max}\Big((GA - I)^T \Sigma (GA - I)\Big) \leq t, \quad \forall A \in \mathcal{A}$

Using a Schur-complement, this can be rewritten as

$$\begin{pmatrix} tI & (GA - I)^T \\ (GA - I) & \Sigma^{-1} \end{pmatrix} \succeq 0, \qquad \forall A \in \mathcal{A}.$$

With $A = A_0 + \Delta$, we obtain:

$$\underbrace{\begin{pmatrix} tI & (GA_0 - I)^T \\ (GA_0 - I) & \Sigma^{-1} \end{pmatrix}}_{M} + \underbrace{\begin{pmatrix} 0 \\ G \end{pmatrix}}_{L} \Delta \underbrace{(I\, 0)}_{R} + \begin{pmatrix} I \\ 0 \end{pmatrix} \Delta^T (0\, G^T) \succeq 0,$$

which has the desired form to apply the S-lemma for robust LMIs.

SDP formulation

- $\sum_{i=1}^{n} \frac{\|\boldsymbol{g}_i\|^2}{w_i} \leq u$

To handle these constraints with LMIs, we introduce a variable v_i for each summand:

$$\sum_{i=1}^{n} \frac{\|\boldsymbol{g}_i\|^2}{w_i} \leq u \iff \exists \boldsymbol{v} \geq \boldsymbol{0} : \begin{cases} \|\boldsymbol{g}_i\|^2 \leq w_i v_i, \quad \forall i \in \{1, \dots, n\} \\ \sum_{i=1}^{m} v_i \leq u \end{cases}$$

Then, it is well known that each constraint $\|\boldsymbol{g}_i\|^2 \leq w_i v_i$ can be reformulated as the equivalent second-order cone constraint $\left\| \begin{pmatrix} 2\boldsymbol{g}_i \\ w_i - v_i \end{pmatrix} \right\| \leq w_i + v_i$, or as the LMI

$$\begin{pmatrix} v_i & \boldsymbol{g}_i^T \\ \boldsymbol{g}_i & w_i I \end{pmatrix} \succeq 0.$$

SDP formulation

Putting all together, we obtain the following SDP to minimize $\phi(G, \boldsymbol{w})$:

$$\begin{aligned}
\underset{G=[\boldsymbol{g}_1,\cdots,\boldsymbol{g}_m],\boldsymbol{w},\lambda,\boldsymbol{v}}{\textbf{minimize}} \quad & t + \sigma^2 \sum v_i \\
s.t. \quad & \begin{pmatrix} (t-\lambda\delta^2)I & (GA_0 - I)^T & 0 \\ (GA_0 - I) & \Sigma^{-1} & G \\ 0 & G^T & \lambda I \end{pmatrix} \succeq 0 \\
& \lambda \geq 0 \\
& \begin{pmatrix} v_i & \boldsymbol{g}_i^T \\ \boldsymbol{g}_i & w_i I \end{pmatrix} \succeq 0, \qquad \forall i \in \{1, \ldots, n\} \\
& \boldsymbol{w}, \boldsymbol{v} \geq \boldsymbol{0} \\
& \sum_{i=1}^{m} w_i = 1.
\end{aligned}$$

Possible extensions to the SDP model

We can also obtain similar tractable SDPs when

- *A* varies in a "scaled ball"

$$\mathcal{A} = \{A_0 + \Delta | \, \|S\Delta T\| \leq \delta\},$$

  for some invertible matrices $S \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{m\times m}$.

Possible extensions to the SDP model

We can also obtain similar tractable SDPs when

- A varies in a "scaled ball"

  $$\mathcal{A} = \{A_0 + \Delta \mid \|S\Delta T\| \leq \delta\},$$

  for some invertible matrices $S \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{m\times m}$.
- There is a prior $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, and we minimize

  $$\Phi_E(G, \boldsymbol{w}) = \sup_{A\in\mathcal{A}} \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\epsilon}}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2]$$

  (integrate over $\boldsymbol{\theta}$ instead of taking the worst-case).

Possible extensions to the SDP model

We can also obtain similar tractable SDPs when

- A varies in a "scaled ball"

$$\mathcal{A} = \{A_0 + \Delta | \ \|S\Delta T\| \leq \delta\},$$

  for some invertible matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$.

- There is a prior $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, and we minimize

$$\Phi_E(G, \boldsymbol{w}) = \sup_{A \in \mathcal{A}} \ \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\epsilon}}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2]$$

  (integrate over $\boldsymbol{\theta}$ instead of taking the worst-case).

- We want to predict $\eta(\boldsymbol{x}) = f(\boldsymbol{x})^T \boldsymbol{\theta}$ over X, we search a linear predictor of the form $\hat{\eta}(\boldsymbol{x}) = f_0(\boldsymbol{x})^T G \boldsymbol{y}$, and we minimize

$$\Phi_\mu(G, \boldsymbol{w}) = \sup_{A \in \mathcal{A}} \ \sup_{\boldsymbol{\theta} \in \Theta} \int_{\boldsymbol{x} \in X} \mathbb{E}_{\boldsymbol{\epsilon}}[\|\eta(\boldsymbol{x}) - \hat{\eta}(\boldsymbol{x})\|^2] \ d\mu(\boldsymbol{x})$$

Outline

Error-in-Variables Model

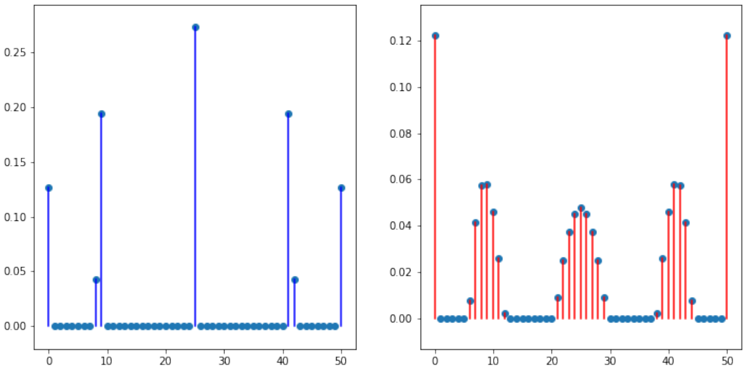
We consider an Error-In-Variable models for polynomial regression.

- X = $[-1, 1]$, discretized with $n = 51$ points.
- We assume the observed function is a polynomial of degree 4, so $m = 5$;
- As basis functions we take the first Legendre polynomials: $P(x) = \sum_{i=1}^{5} \theta_i L_i(x)$, i.e.,

$$f_0(x) = [L_1(x), L_2(x), \dots, L_5(x)]^T.$$

- When the experimenter wants to observe $P(x)$, he/she gets a (noisy) observation of $P(x + \eta)$ instead, where $\eta \sim \mathcal{N}(0, 0.03^2)$.
- We generate a *true* $\boldsymbol{\theta} \sim \mathcal{N}(0, I)$

Optimal designs



Left: Bayesian $A-$optimal design, computed with prior

$$\boldsymbol{\theta} \sim \mathcal{N}(0, I)$$

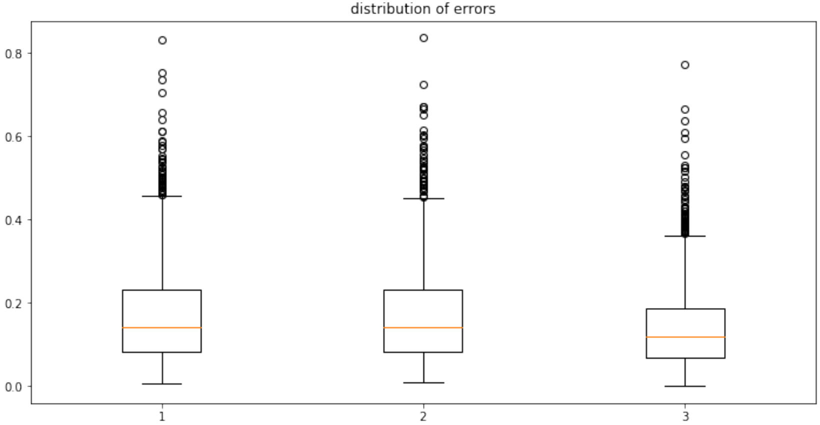
Right: Robust design, for $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^m | \, \|\boldsymbol{\theta}\| \leq 1\}$ (i.e., $\Sigma = I$) and robustness level $\delta = 0.1$.

Comparisons of designs

- We run tests with $N_1 = 10$ randomly generated polynomials.
- For each, we generate $N_2 = 100$ observartion vector $\boldsymbol{y}$ for randomly generated offsets of the observation location $x_i' = x_i + \eta_i$.
- We compute an estimate of $\boldsymbol{\theta}$:
  - For the Bayesian $A-$optimal design, with LS estimation ignoring the x_i-offsets.
  - For the Bayesian $A-$optimal design, with a robust estimator [El Ghaoui & Lebret]
  - For the robust design, with the robust linear estimator $\boldsymbol{\theta} = G\boldsymbol{y}$.

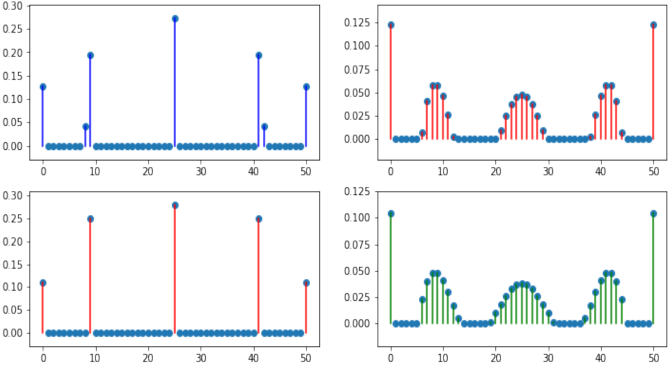
Comparisons of designs

Box plots of squared estimation error



A-optimal design LS estimator	A-optimal design robust estimator	robust design linear robust estimator

Estimation vs. prediction



estimation

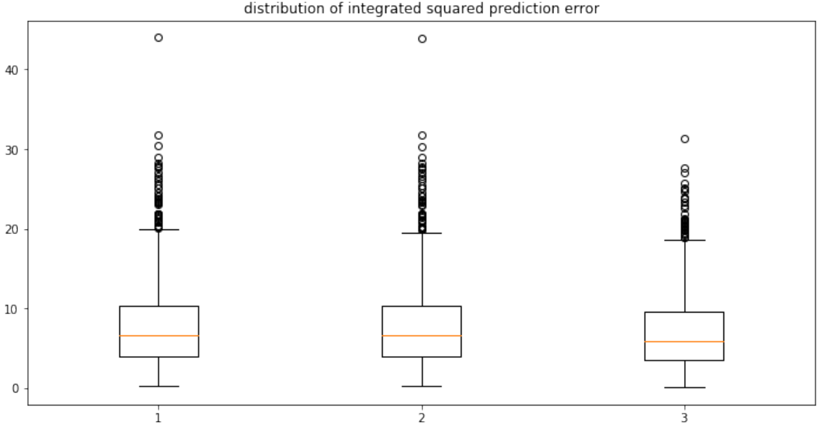
prediction

A-/I- optimal

robust designs

Comparisons of designs

Box plots of integrated squared prediction error



G-optimal design LS predictor	G-optimal design robust predictor	robust design linear robust predictor

Conclusion & Perspectives

- Tools of robust optimal control used to compute robust optimal designs

TO DO:

- Study influence of robustness level δ
- Can we formulate an *equivalence theorem* for the robust criterion, in particular for the case where X is not discretized?
- Analytical computation of robust design measure in simple cases
- Real-world applications (e.g. X-ray imaging)

References

A few references on robust estimation:

- L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data", *SIAM Journal on matrix analysis and applications*, **18.4** (1997): 1035-1064.
- G. Calafiore and L. El Ghaoui, "Robust maximum likelihood estimation in the linear model", *Automatica*, **37.4** (2001): 573-580.
- Y.C. Eldar, A. Ben-Tal and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties", *IEEE Transactions on Signal Processing*, **53.1** (2005): 168-181.
- I. Pólik and T. Terlaky, "A survey of the S-lemma", *SIAM review*, **49.3** (2007): 371-418.