Optimal experimental design with polynomial optimization

Didier HENRION

with Yohann DE CASTRO, Fabrice GAMBOA
and Jean-Bernard LASSERRE

IMA Minneapolis, January 2016

Outline

1. Optimal linear regression and semidefinite optimization
2. Polynomial regression and moment optimization
3. Moment relaxations and semidefinite optimization
4. Examples

Estimate parameters $p \in \mathbb{R}^n$ from **linear** measurements

$$q_k = a_k^T p + \epsilon_k \in \mathbb{R}, \quad k = 1, \ldots, N$$

where a_k are test vectors and ϵ_k is a Gaussian measurement noise

Test vectors chosen **randomly** amongst finitely many linearly independent vectors in $X := \{x_1, \ldots, x_m\} \subset \mathbb{R}^n$ with $m \ll N$ and

$$\text{Prob}\,(a_k = x_i) = w_i$$

with given weights $w_1 \geq 0, \ldots, w_m \geq 0$, $w_1 + \cdots + w_m = 1$

In matrix form

$$q \approx A^T p$$

with full row rank $A \in \mathbb{R}^{n \times N}$ having columns a_k

Least squares estimator

$$\min_{p} \|q - A^T p\|_2^2$$

yields necessary and sufficient optimality conditions $AA^T p = Aq$ and optimal solution

$$\widehat{p} = (AA^T)^{-1} Aq$$

Error covariance matrix w.r.t measurement noise

$$E(\widehat{p} - p)(\widehat{p} - p)^T = (AA^T)^{-1} = \left(\sum_{k=1}^{N} a_k a_k^T \right)^{-1}$$

should be as **small** as possible, with

$$E(a_k a_k^T) = \sum_{i=1}^{m} w_i x_i x_i^T$$

Optimal linear regression = given atoms x_i, find weights w_i making as "large" as possible the information matrix

$$M(w) := \sum_{i=1}^{m} w_i x_i x_i^T$$

Various spectral criteria to maximize over the simplex:
$\min \mathrm{eig}\, M(w)$ (E-optimal), $\mathrm{trace}\, M(w)$ (A-optimal), $\mathrm{cond}\, M(w)$ (K-optimal) or $\det M(w)$ (**D-optimal**)

Analytic centering problem of **semidefinite programming**

$$\begin{aligned} \max_{w} \ & \log\det M(w) \\ \text{s.t.} \ & w_1 + \cdots + w_m = 1 \\ & w_1 \geq 0, \ldots, w_m \geq 0 \\ & M(w) \succeq 0 \end{aligned}$$

see e.g. [Vandenberghe, Boyd, Wu - SIMAX 1998]

Outline

1. Optimal linear regression and semidefinite optimization

2. **Polynomial regression and moment optimization**

3. Moment relaxations and semidefinite optimization

4. Examples

Now suppose we estimate p from **polynomial** measurements

$$q(x) \approx \sum_{\alpha} p_\alpha a_\alpha(x)$$

where x is a random variable in a compact set $X \subset \mathbb{R}^n$ whose law is the **probability measure** $\mu \in \mathscr{P}(X)$

Basis elements $a := (a_\alpha)_{\alpha \in \mathbb{N}^n,\, |\alpha| := \alpha_1 + \cdots + \alpha_n \leq d} \subset \mathbb{R}[x]$ should span the vector space of n-variate polynomials of degree up to d

Linear regression is a particular case for which $d = 1$, $X = \{x_1, \ldots, x_m\} \subset \mathbb{R}^n$ and atomic probability distribution

$$\mu = \sum_{i=1}^{m} w_i \delta_{x_i}$$

Least squares estimator

$$\min_{p} \int_X (q(x) - a(x)^T p)^2 \mu(dx)$$

yields necessary and sufficient optimality conditions

$$\int_X a(x)a(x)^T \mu(dx) p = \int_X q(x)a(x)\mu(dx)$$

and optimal solution

$$\hat{p} = \left(\int_X a(x)a(x)^T \mu(dx)\right)^{-1} \int_X q(x)a(x)\mu(dx)$$

with information matrix, or **moment matrix**

$$M_d(\mu) := \int_X a(x)a(x)^T \mu(dx)$$

to be "maximized" with respect to measure $\mu \in \mathscr{P}(X)$

Since measurements are polynomial $a = (a_\alpha)_{|\alpha| \leq d} \subset \mathbb{R}[x]$ we can express the moment matrix

$$M_d(\mu) = \int_X a(x)a(x)^T \mu(dx)$$

as a function of the **moments** of measure $\mu \in \mathscr{P}(X)$

$$y_\alpha := \int_X a_\alpha(x)\mu(dx) \in \mathbb{R}$$

and we write

$$M_d(\mu) = M_d(y) = \sum_{|\alpha| \leq 2d} y_\alpha H_\alpha$$

where H_α are given symmetric matrices, $\alpha \in \mathbb{N}^n$

Moment matrix now to be "maximized" with respect to y

Moment formulation of D-optimal design for polynomial regression of degree d

$$\begin{aligned} \max_{y} \ & \log\det M_d(y) \\ \text{s.t. } & y = \int_X a(x)\mu(dx) \text{ for some } \mu \in \mathscr{P}(X) \end{aligned}$$

Finite-dimensional convex optimization problem

Essential difficulty = representing a measure with its moments

Outline

1. Optimal linear regression and semidefinite optimization
2. Polynomial regression and moment optimization
3. **Moment relaxations and semidefinite optimization**
4. Examples

Let $p_0, p_1, \ldots, p_m \in \mathbb{R}[x]$ with $p_0(x) = 1$ and $p_1(x) = R^2 - \sum_{i=1}^n x_i^2$ and let $\ell_y(\sum_\alpha p_\alpha a_\alpha(x)) := \sum_\alpha p_\alpha y_\alpha$ be the Riesz functional

Given the compact basic semialgebraic set

$$X := \{x \in \mathbb{R}^n \,:\, p_1(x) \geq 0, \ldots, p_m(x) \geq 0\}$$

the finite-dimensional convex **moment cone**

$$\mathscr{M}_d(X) := \{(y_\alpha)_{|\alpha| \leq 2d} \,:\, y_\alpha = \int a_\alpha \mu \text{ for some } \mu \in \mathscr{P}(X)\}$$

is approximated with projections of the **semidefinite cone**

$$\begin{aligned}\mathscr{M}_r^{\mathrm{sdp}}(X) := \{(y_\alpha)_{|\alpha| \leq 2d} \,:\, & \ell_y(p_i q^2) \geq 0 \text{ for all } q \in \mathbb{R}[x] \\ & \text{s.t. } \deg p_i q \leq 2r,\ i = 0, 1, \ldots, m\}\end{aligned}$$

$$\mathscr{M}_d(X) \subset \cdots \subset \mathscr{M}_{d+1}^{\mathrm{sdp}}(X) \subset \mathscr{M}_d^{\mathrm{sdp}}(X), \quad \mathscr{M}_d(X) = \overline{\cap_{r=d}^{\infty} \mathscr{M}_r^{\mathrm{sdp}}(X)}$$

D-optimal polynomial regression of degree d

$$p^* = \max_{y} \ \log\det M_d(y) \quad \text{s.t. } y \in \mathscr{M}_d(X)$$

solved by hierarchy of moment SDP relaxations

$$\boxed{p_r^* = \max_{y} \ \log\det M_d(y) \quad \text{s.t. } y \in \mathscr{M}_r^{\mathsf{sdp}}(X)}$$

for increasing relaxation orders $r = d, d+1, \ldots$

It holds $p_d^* \leq p_{d+1}^* \leq \cdots \leq p_\infty^* = p^*$ and we expect $p_r^* = p^*$ for finite and small r

Let $y^* := (y^*_\alpha)_{|\alpha|\leq 2r}$ be a solution of the relaxation of order r

We want a measure $\mu \in \mathscr{P}(X)$ with these moments

Following [Jiawang Nie. Found Comput Math. 2014]
we solve a **truncated moment** SDP problem

$$\begin{aligned} \min_{y}\ & \ell_y(q) \\ \text{s.t. } & (y_\alpha)_{|\alpha|\leq 2r} = y^* \\ & y \in \mathscr{M}_s^{\mathsf{sdp}}(X) \end{aligned}$$

for a generic (random) SOS polynomial $q \in \mathbb{R}[x]$ and
for increasing relaxation orders $s = r, r+1, \ldots$

Atomic measure recovered by flat extension

$$\operatorname{rank} M_s(y) = \operatorname{rank} M_{s+1}(y) = \operatorname{rank} M_{s+2}(y) = \cdots$$

Outline

1. Optimal linear regression and semidefinite optimization
2. Polynomial regression and moment optimization
3. Moment relaxations and semidefinite optimization
4. **Examples**

Univariate unit interval

Regression for degree d polynomial measurements

$$\sum_{\alpha=0,1,\ldots,d} p_\alpha x^\alpha$$

on the interval $X = [-1, 1]$

D-optimal measure $\mu \in \mathscr{P}(X)$ is atomic uniformly supported on the end points -1, $+1$ and the critical points of the **Legendre polynomial** of degree d

Legendre polys are orthogonal w.r.t. Lebesgue measure on $[-1, 1]$

GloptiPoly YALMIP Matlab script

```
r = 5; % half degree
mpol x;
P  = msdp(1-x^2>=0,r);
[F,h,y] = myalmip(P);
M = sdpvar(F(1)); % moment matrix
sol = optimize(F,-geomean(M)); % solve maxdet SDP
ystar = [1;double(y)]; % vector of moments

% recover the atoms
R = msdp(mom(x.^((0:2*r)'))==ystar, 1-x^2>=0, r+1);
[stat,obj] = msol(R);
double(x)
```

3D unit sphere

Regression for degree d polynomial measurements

$$\sum_{|\alpha|\leq d} p_\alpha x^\alpha$$

on the unit sphere $X = \{x \in \mathbb{R}^3 \, : \, x_1^2 + x_2^2 + x_3^2 = 1\}$

First step: max-det moment relaxation of order $r \geq d$

$$\boxed{\begin{aligned} p_r^* \; = \; \max_{y} \;\; & \log\det M_d(y) \\ \text{s.t.} \;\; & y \in \mathscr{M}_r^{\mathsf{sdp}}(X) \end{aligned}}$$

For $d = 1$ and $r = 1$ we obtain

$$y^*_{000} = 1.0000, \quad y^*_{200} = y^*_{020} = y^*_{002} = 0.3333$$

and all other entries of $y^* \in \mathbb{R}^{10}$ are zero

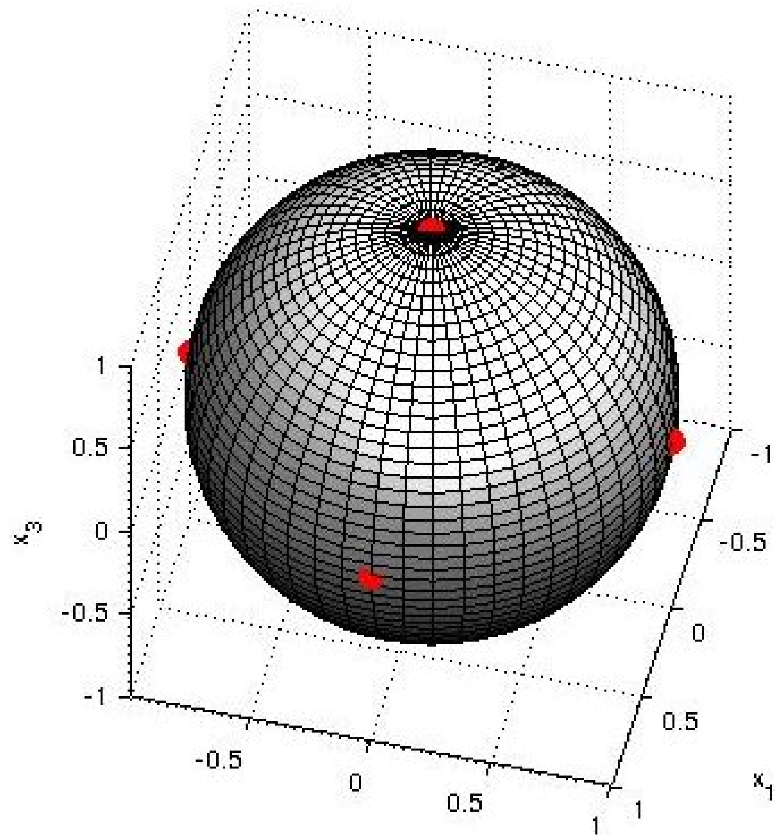
Second step: solve truncated moment relaxation of order $s \geq r$

$$\begin{aligned} \min_{y} \ & \ell_y(q) \\ \text{s.t.} \ & (y_\alpha)_{|\alpha| \leq 2r} = y^* \\ & y \in \mathscr{M}_s^{\mathsf{sdp}}(X) \end{aligned}$$

For $s = 3$ we obtain the flat extension

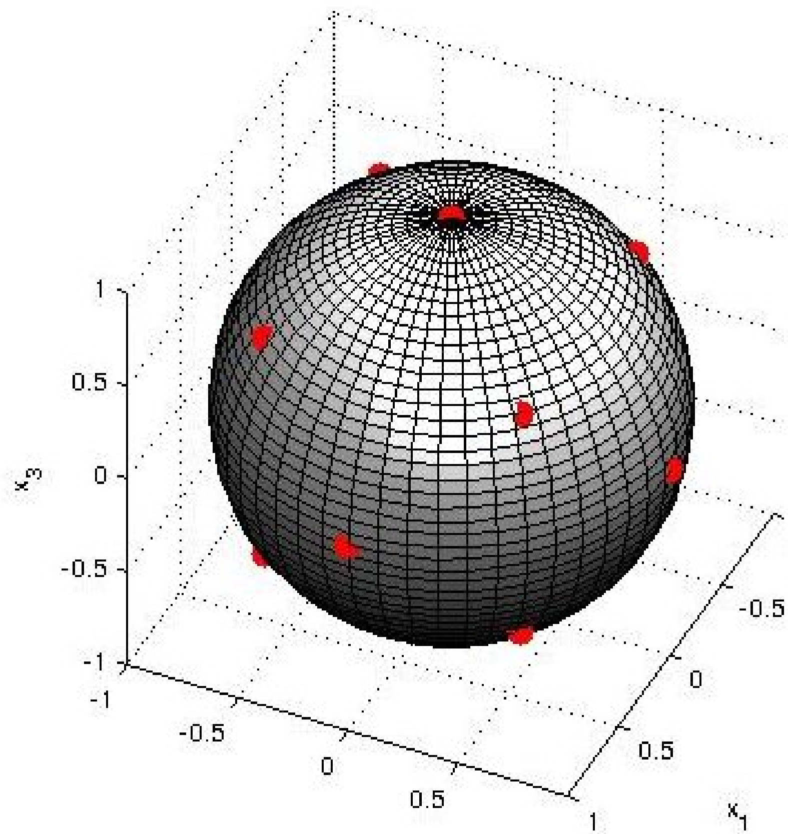
$$\operatorname{rank} M_2(y) = \operatorname{rank} M_3(y) = 6$$

and the 6 atoms $\{(\pm 1, 0, 0),\ (0, \pm 1, 0),\ (0, 0, \pm 1)\} \subset X$ on which the optimal measure $\mu \in \mathscr{P}(X)$ is uniformly supported

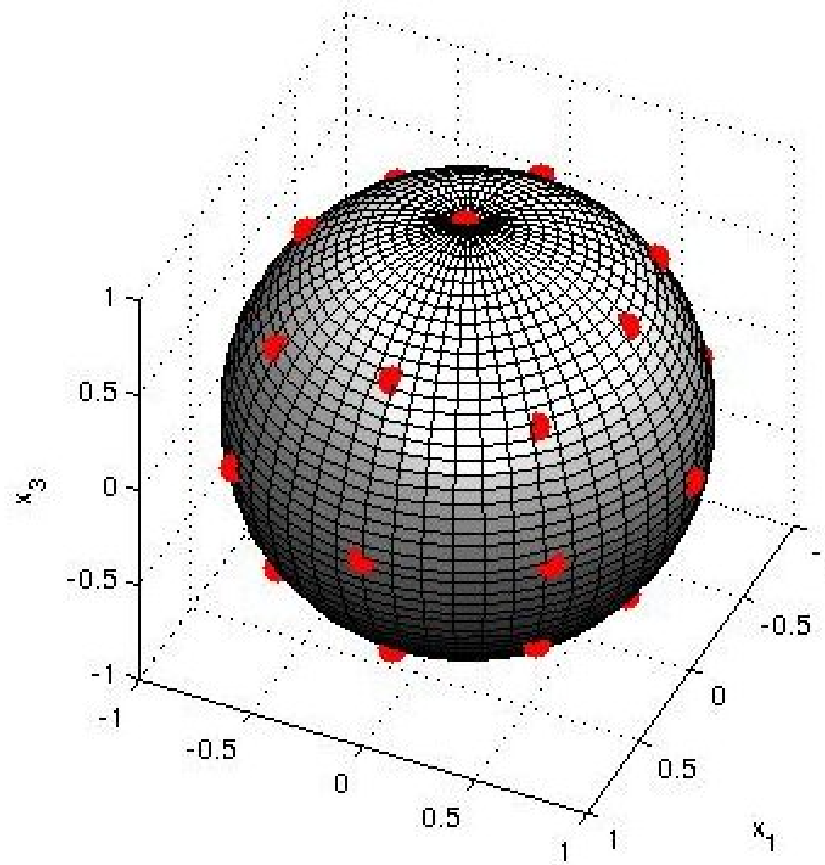


1
0.5
0
-0.5
-1
x3
-0.5
0
0.5
1
-1
-0.5
0
0.5
1
x1

For quadratic regressions ($d = 2$) we obtain (with $r = 2$ and $s = 5$) an optimal measure supported on 14 atoms distributed evenly on the sphere



For cubic regressions ($d = 3$) we obtain (with $r = 3$ and $s = 7$) an optimal measure supported on 26 atoms distributed evenly on the sphere



Summary

Optimal design with polynomial regression on semialgebraic sets

Instead of fixing the (atomic) support of the measure and optimizing over the weights, optimize directly over the measure

Measure optimal design problem = moment problem

Use Lasserre's hierarchy to relax the moment problem

Use Nie's strategy to retrieve atomic measure