# Approximate Optimal Designs for Multivariate Polynomial Regression


Fabrice Gamboa

Collaboration with: Yohan de Castro, Didier Henrion, Roxana Hess, Jean-Bernard Lasserre


Universität Potsdam

16th of February 2018

INSTITUT de MATHEMATIQUES de TOULOUSE

# Agenda

# Overview

# Short bibliography : Optimal design


- H. Dette and W. J. Studden. The theory of canonical moments with applications in statistics, probability, and analysis, volume 338. John Wiley & Sons, 1997.
- J. Kiefer. General equivalence theory for optimum designs (approximate theory). The annals of Statistics, pages 849–879, 1974.
- I. Molchanov and S. Zuyev. Optimisation in space of measures and optimal design. ESAIM : Probability and Statistics, 8 :12–24, 2004.
- F. Pukelsheim. Optimal design of experiments. SIAM, 2006.
- G. Sagnol and R. Harman. Computing exact D-optimal designs by mixed integer second-order cone programming. The Annals of Statistics, 43(5) :2198–2224, 2015.

# Short bibliography : Optimization


- J. B. Lasserre. Moments, positive polynomials and their applications, volume 1 of Imperial College Press Optimization Series. Imperial College Press, London, 2010.
- J. B. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. Mathematische Zeitschrift, 256(1) :99–112, 2007.
- A. S. Lewis. Convex analysis on the Hermitian matrices. SIAM Journal on Optimization, 6(1) :164–177, 1996.
- L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. SIAM journal on matrix analysis and applications, 19(2) :499–533, 1998.

# Overview

# Classical linear model

What are you dealing with : Regression model

$$F : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}^p, \text{ continuous.}$$

Noisy linear model observed at $t_i \in \mathcal{X}, i = 1, \ldots, N$

$$z_i = \langle \theta^*, F(t_i) \rangle + \varepsilon_i .$$

- $\theta^* \in \mathbb{R}^p$ unknown,
- $\varepsilon$ second order homoscedastic centred white noise

Best choice for $t_i \in \mathcal{X}, i = 1, \ldots, N$ ?

# Information matrix

Normalized inverse covariance matrix of the optimal linear unbiased estimate of $\theta^*$ (Gauss Markov)

$$M(\xi) = \frac{1}{N}\sum_{i=1}^{N} F(t_i)F^T(t_i) = \sum_{i=1}^{l} w_i F(x_i)F^T(x_i).$$

$t_i, i = 1, \cdots, N$ are picked $Nw_i$ times within $x_i, i = 1, \cdots, l, (l < N)$

$$\xi = \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ w_1 & w_2 & \cdots & w_l \end{pmatrix}$$

- $w_j = \frac{n_j}{N}$,
- Simplification of the frame (if not discrete optimisation) $\Rightarrow 0 < w_j < 1, \sum w_j = 1$

Best choice for the design $\xi$ ?

# Concave matricial criteria

Wish to *maximize* the information matrix with respect to $\xi$

$$M(\xi) = \sum_{i=1}^{l} w_i F(x_i) F^T(x_i) = \int_{\mathcal{X}} F(x) F^T(x) d\sigma_\xi .$$

$$\xi = \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ w_1 & w_2 & \cdots & w_l \end{pmatrix} \quad \sigma_\xi(dx) := \sum_{j=1}^{l} w_j \delta_{x_j}(dx).$$

Concave criteria for symmetric $p \times p$ non negative matrix : $\phi_q (q \in [-\infty, 1])$

$$M > 0, \ \phi_q(M) := \begin{cases} (\frac{1}{p} \mathrm{trace}(M^q))^{1/q} & \text{if } q \neq -\infty, 0 \\ \det(M)^{1/p} & \text{if } q = 0 \\ \lambda_{\min}(M) & \text{if } q = -\infty \end{cases}$$

$$\det M = 0, \ \phi_q(M) := \begin{cases} (\frac{1}{p} \mathrm{trace}(M^q))^{1/q} & \text{if } q \in (0, 1] \\ 0 & \text{if } q \in [-\infty, 0]. \end{cases}$$

# Optimal design

Wish to maximize $\phi_q(M(\xi))$ ( with respect to $\xi$)

$$\xi = \begin{pmatrix} x_1 & x_2 & \cdots & x_l \\ w_1 & w_2 & \cdots & w_l \end{pmatrix} \quad \sigma_\xi(dx) := \sum_{j=1}^{l} w_j \delta_{x_j}(dx).$$

- $\phi_q(M)$ concave with respect to $M$ and positively homogenous,
- $\phi_q(M)$ is isotonic with respect to Loewner ordering

Main idea : extend the optimization problem to all probability measures

$$M(P) = \int_{\mathcal{X}} F(x)F^T(x)dP(x), \quad P \in \mathbb{P}(\mathcal{X}).$$

Within the solutions build one with finite support

# Overview

# Our frame : polynomial regression

- $\mathbb{R}[x]$ real polynomials in the variables $x = (x_1, \dots, x_n)$
- $d \in \mathbb{N}$ $\mathbb{R}[x]_d := \{p \in \mathbb{R}[x] : \deg p \leqslant d\}$ $\deg p :=$ total degree of $p$
  Assumption $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_p) \in (\mathbb{R}[x]_d)^p$.
- $\mathcal{X} \subset \mathbb{R}^n$ is a given closed basic semi-algebraic set

$$\mathcal{X} := \{x \in \mathbb{R}^m : g_j(x) \geqslant 0,\ j = 1, \dots, m\} \tag{1}$$

$g_j \in \mathbb{R}[x],\ \deg g_j = d_j,\ j = 1, \dots, m$, and $\mathcal{X}$ compact e.g :

$$g_1(x) := R^2 - \|x\|^2$$

# Overview

# Some facts and notations

- $x_1^{\alpha_1}\cdots x_n^{\alpha_n}$, $\alpha=(\alpha_1,\dots,\alpha_n)\in\mathbb{N}^n$ basis of $\mathbb{R}[x]$ ($x^\alpha := x_1^{\alpha_1}\cdots x_n^{\alpha_n}$)
- $\mathbb{R}[x]_d$ has dimension $s(d) := \binom{n+d}{n}$. Basis $(x^\alpha)_{|\alpha|\leqslant d}$, $|\alpha| := \alpha_1+\cdots+\alpha_n$.

$$\mathbf{v}_d(x) := (\underbrace{1}_{\text{degree } 0}, \underbrace{x_1,\dots,x_n}_{\text{degree } 1}, \underbrace{x_1^2, x_1x_2,\dots,x_1x_n, x_2^2,\dots,x_n^2}_{\text{degree } 2},\dots,\underbrace{\dots,x_1^d,\dots,x_n^d}_{\text{degree } d})^\top$$

- There exists a unique matrix $\mathfrak{A}$ of size $p\times\binom{n+d}{n}$ such that

$$\forall x\in\mathcal{X},\quad \mathbf{F}(x) = \mathfrak{A}\,\mathbf{v}_d(x)\,. \tag{2}$$

- $\mathcal{M}_+(\mathcal{X})$ is the cone of nonnegative Borel measures supported on $\mathcal{X}$ (the dual of cone of nonnegative elements of $\mathcal{C}(\mathcal{X})$)

# Moments, the moment cone and the moment matrix

$$\mu \in \mathcal{M}_+(\mathcal{X})\alpha \in \mathbb{N}^n, y_\alpha = y_\alpha(\mu) = \int_{\mathcal{X}} x^\alpha d\mu$$

- $\mathbf{y} = y_\alpha(\mu) = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ moment sequence of $\mu$
- $\mathcal{M}_d(\mathcal{X})$ moment cone=convex cone of truncated sequences

$$\mathcal{M}_d(\mathcal{X}) := \Big\{ \mathbf{y} \in \mathbb{R}^{\binom{n+d}{n}} : \exists\, \mu \in \mathcal{M}_+(\mathcal{X}) \text{ s.t. } y_\alpha = \int_{\mathcal{X}} x^\alpha \, d\mu, \ \forall \alpha \in \mathbb{N}^n, \ |\alpha| \leqslant d \Big\}. \tag{3}$$

- $\mathcal{P}_d(\mathcal{X})$ convex cone of nonnegative polynomials of degree $\leqslant d$ (may be also viewed as a vector of coefficients)
- $\mathcal{M}_d(\mathcal{X}) = \mathcal{P}_d(\mathcal{X})^\star$ and $\mathcal{P}_d(\mathcal{X}) = \mathcal{M}_d(\mathcal{X})^\star$

If $\mathcal{X} = [a, b]$, $\mathcal{M}_d(\mathcal{X})$ is representable using positive semidefinitness of Hankel matrices. No more true in general

# Generalized Hankel matrix I

Sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ is given, build linear form $L_\mathbf{y} : \mathbb{R}[x] \to \mathbb{R}$ mapping a polynomials $f = \sum_{\alpha \in \mathbb{N}^n} f_\alpha x^\alpha$ to $L_\mathbf{y}(f) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha y_\alpha$.

**Theorem**

*A sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ has a representing measure $\mu$ supported on $\mathcal{X}$ if and only if $L_\mathbf{y}(f) \geqslant 0$ for all polynomials $f \in \mathbb{R}[x]$ nonnegative on $\mathcal{X}$.*

The *generalized Hankel matrix* associated to a truncated sequence $\mathbf{y} = (y_\alpha)_{|\alpha| \leqslant 2d}$

$$M_d(\mathbf{y})(\alpha, \beta) = L_\mathbf{y}(x^\alpha x^\beta) = y_{\alpha+\beta}, \quad (|\alpha|, |\beta| \leqslant d).$$

- $(M_d(\mathbf{y})(\alpha, \beta))$ is symmetric
- $(M_d(\mathbf{y})(\alpha, \beta))$ is linear in $\mathbf{y}$
- If $\mathbf{y}$ has a representing measure, then $M_d(\mathbf{y}) \succcurlyeq 0$

# Generalized Hankel matrix II

Generalized Hankel matrix associated to polynomial $f = \sum_{|\alpha|\leqslant r} f_\alpha x^\alpha \in \mathbb{R}[x]_r$ and a sequence $\mathbf{y} = (y_\alpha)_{|\alpha|\leqslant 2d+r}$

$$M_d(f\mathbf{y})(\alpha,\beta) = L_{\mathbf{y}}(f(x)\,x^\alpha x^\beta) = \sum_{|\gamma|\leqslant r} f_\gamma y_{\gamma+\alpha+\beta}, \quad (|\alpha|,|\beta| \leqslant d)$$

- If $\mathbf{y}$ has a representing measure $\mu$, and Supp$\mu \subset \{x \in \mathbb{R}^n : f(x) \geqslant 0\}$ then $M_d(f\mathbf{y}) \succcurlyeq 0$
- The converse statement holds in the infinite case

$\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n}$ has a representing measure $\mu \in \mathcal{M}_+(\mathcal{X})$

$$\Leftrightarrow \forall d \in \mathbb{N}, \quad M_d(\mathbf{y}) \succcurlyeq 0 \text{ and } M_d(g_j\mathbf{y}) \succcurlyeq 0, \quad j = 1,\ldots,m$$

# Approximations of the moment cone

Set $v_j := \lceil d_j/2 \rceil$, $j = 1, \ldots, m$, (half the degree of the $g_j$). For $\delta \in \mathbb{N}$, $\mathcal{M}_{2d}(\mathcal{X})$ can be approximate by

$$\mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}) := \Big\{ \mathbf{y}_{d,\delta} \in \mathbb{R}^{\binom{n+2d}{n}} \ : \ \exists \mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}} \text{ such that } \quad (4)$$
$$\mathbf{y}_{d,\delta} = (y_{\delta,\alpha})_{|\alpha| \leqslant 2d} \text{ and}$$
$$M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0,\ M_{d+\delta-v_j}(g_j \mathbf{y}_\delta) \succcurlyeq 0,\ j = 1, \ldots, m \Big\}.$$

Good approximation

- $\mathcal{M}_{2d}(\mathcal{X}) \subseteq \cdots \subseteq \mathcal{M}^{\mathsf{SDP}}_{2(d+2)}(\mathcal{X}) \subseteq \mathcal{M}^{\mathsf{SDP}}_{2(d+1)}(\mathcal{X}) \subseteq \mathcal{M}^{\mathsf{SDP}}_{2d}(\mathcal{X})$.
- This hierarchy converges $\mathcal{M}_{2d}(\mathcal{X}) = \overline{\bigcap_{\delta=0}^{\infty} \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})}$

# Overview

# Optimal design moment formulation

For $\ell \leqslant \binom{n+2d}{n}$ and $W$ the simplex Carathéodory's theorem gives

$$\left\{\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1\right\} = \left\{\mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}} : y_\alpha = \int_{\mathcal{X}} x^\alpha d\mu \quad \forall |\alpha| \leqslant 2d, \right.$$

$$\left. \mu = \sum_{i=1}^{\ell} w_i \delta_{x_i},\ x_i \in \mathcal{X},\ w \in \mathcal{W} \right\},$$

Our design problem becomes : Step 1 find $y^*$ solution

$$\begin{aligned} \rho = \max\ & \phi_q(M) \\ \text{s.t. } M = & \sum_{|\gamma| \leqslant 2d} A_\gamma y_\gamma \succcurlyeq 0, \\ & \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}),\ y_0 = 1, \end{aligned} \tag{5}$$

Step 2 find an atomic measure $\mu^*$ representing $\mathbf{y}^*$

# Equivalence Theorem

## Theorem (Equivalence theorem)

*Let* $q \in (-\infty, 1)$, *Step 1 is is a convex optimization problem with a unique optimal solution* $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$. *Denote by* $p_d^\star$ *the polynomial*

$$x \mapsto p_d^\star(x) := \mathbf{v}_d(x)^\top M_d(\mathbf{y}^\star)^{q-1}\mathbf{v}_d(x) = \|M_d(\mathbf{y}^\star)^{\frac{q-1}{2}}\mathbf{v}_d(x)\|_2^2. \qquad (6)$$

$\mathbf{y}^\star$ *is the vector of moments of a discrete measure* $\mu^\star$ *supported on at least* $\binom{n+d}{n}$ *and at most* $\binom{n+2d}{n}$ *points in the set (Step 2)*

$$\Omega := \left\{x \in \mathcal{X} : \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) = 0\right\}.$$

$\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ *is the unique solution to Step 1 iff*
$\mathbf{y}^\star \in \left\{\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1\right\}$ *and* $p^\star := \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star \geqslant 0$ *on* $\mathcal{X}$.

# Overview

# The SDP relaxation scheme

Ideal moment Problem on $\mathcal{M}_{2d}(\mathcal{X})$ is not SDP representable $\Rightarrow$ Use the outer approximations of Step 1

$$\rho_\delta = \max_{\mathbf{y}} \quad \phi_q(M_d(\mathbf{y}))$$
$$\text{s.t.} \quad \mathbf{y} \in \mathcal{M}^{\mathrm{SDP}}_{2(d+\delta)}(\mathcal{X}),\ y_0 = 1. \qquad (\forall \delta > 0,\ \rho_\delta \geqslant \rho.)$$

**Theorem (Equivalence theorem for SDP relaxation)**

*Let* $q \in (-\infty, 1)$,

1. *SDP relaxed Problem has a unique optimal solution* $\mathbf{y}^\star \in \mathbb{R}^{\binom{n+2d}{n}}$.
2. $M_d(\mathbf{y}^\star)$ *is positive definite.* $p^\star := \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star$ *is non-negative on* $\mathcal{X}$ *and* $L_{\mathbf{y}^\star}(p^\star) = 0$.

If $\mathbf{y}^\star$ is coming from a measure $\mu^\star$ then $\rho_\delta = \rho$ and $\mathbf{y}^\star_{d,\delta}$ is the solution of unrelaxed Step 1

# Asymptotics on $\delta$

## Theorem

*Let* $q \in (-\infty, 1)$, $\mathbf{y}^\star_{d,\delta}$ *optimal solution of relaxed,* $p^\star_{d,\delta} \in \mathbb{R}[x]_{2d}$ *dual polynomial associated. Then,*

1. $\rho_\delta \to \rho$ *as* $\delta \to \infty$,
2. *For every* $\alpha$, $|\alpha| \leqslant 2d$, $\lim_{\delta \to \infty} y^\star_{d,\delta,\alpha} = y^\star_\alpha$,
3. $p^\star_{d,\delta} \to p^\star_d$ *as* $\delta \to \infty$,
4. *If the dual polynomial* $p^\star := \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p^\star_d$ *to the unrelaxed Step 1 belongs to* $\mathcal{P}^{\mathrm{SOS}}_{2(d+\delta)}(\mathcal{X})$ *for some* $\delta$*, then finite convergence takes place.*

$\mathcal{P}^{\mathrm{SOS}}_{2(d+\delta)}(\mathcal{X})$ is the set of sum of squares of polynomial of degrees less than $d + \delta$

# Overview

# Example I : interval

$\mathcal{X} = [-1, 1]$, polynomial regression model $\sum_{j=0}^{d} \theta_j x^j$

- D-optimal design : for $d = 5$ and $\delta = 0$ we obtain the sequence $\mathbf{y}^\star \approx (1, 0, 0.56, 0, 0.45, 0, 0.40, 0, 0.37, 0, 0.36)^\top$. Recover the corresponding atomic measure from the sequence $\mathbf{y}^\star$ : supported by -1, -0.765, -0.285, 0.285, 0.765 and 1 (for $d = 5$, $\delta$=0). The points match with the known analytic solution to the problem (critical points of the Legendre polynomial)

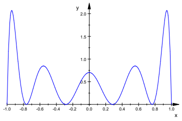


FIGURE: Polynomial $p^\star$ D-optimality, $n = 1$

# Example II : Wynn's polygon

Polygon given by the vertices $(-1,-1)$, $(-1,1)$, $(1,-1)$ and $(2,2)$, scaled to fit the unit circle, *i.e.*, we consider the design space

$$\mathcal{X} = \left\{ x \in \mathbb{R}^2 : x_1, x_2 \geqslant -\tfrac{1}{4}\sqrt{2},\ x_1 \leqslant \tfrac{1}{3}(x_2+\sqrt{2}),\ x_2 \leqslant \tfrac{1}{3}(x_1+\sqrt{2}),\ x_1^2+x_2^2 \leqslant 1 \right\}.$$

- D-optimal design

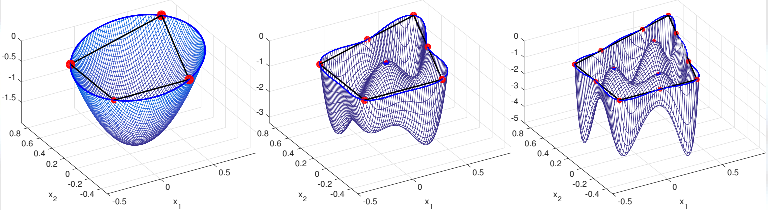


FIGURE: The polynomial $p_d^\star - \binom{2+d}{2}$ where for $d=1$ , $d=2$, $d=3$. The red points correspond to the *good* level set

# Example III : Ring of ellipses

## An ellipse with a hole in the form of a smaller ellipse

$$\mathcal{X} = \{x \in \mathbb{R}^2 : 9x_1^2 + 13x_2^2 \leqslant 7.3,\ 5x_1^2 + 13x_2^2 \geqslant 2\}.$$

- D-optimal design

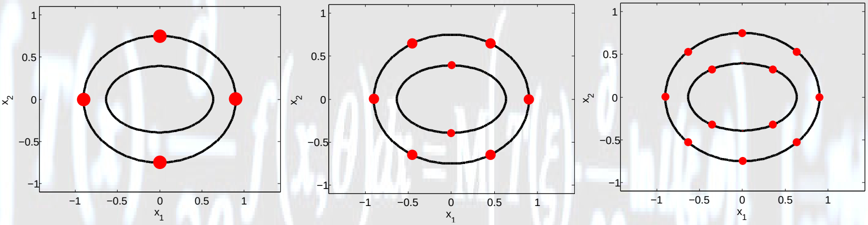


FIGURE: The boundary in bold black. The support of the optimal design measure (red points). Size of the points corresponds to the respective weights for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.

# Example IV : Folium

Zero set of $f(x) = -x_1(x_1^2 - 2x_2^2)(x_1^2 + x_2^2)^2$ is a curve with a triple singular point at the origin called a folium,

$$\mathcal{X} = \{x \in \mathbb{R}^2 : f(x) \geqslant 0,\ x_1^2 + x_2^2 \leqslant 1\}.$$

- D-optimal design

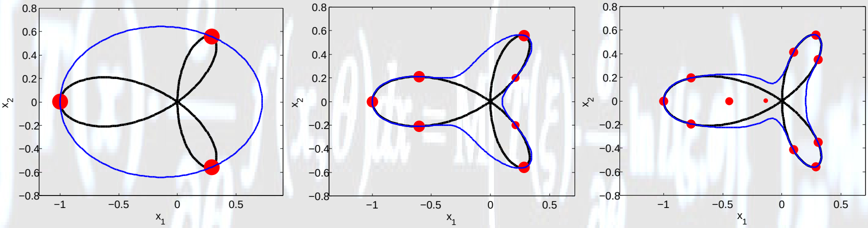


FIGURE: Boundary in bold black The support of the optimal design measure (red points). The *good* level set in thin blue $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right), $\delta = 3$

# The 3-dimensional unit sphere

## Polynomial regression $\sum_{|\alpha|\leqslant d} \theta_\alpha x^\alpha$ on the unit sphere

$$\mathcal{X} = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}.$$

- $\mathrm{D}$-optimal design

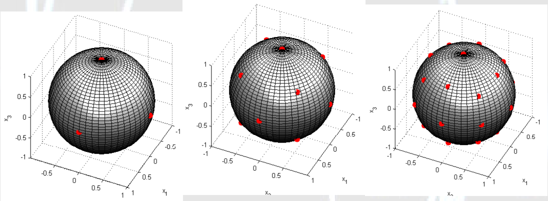

FIGURE: Optimal design in red $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 0$

# The 3-dimensional unit sphere+constraint on moments

Fix $y_{020} := 2\omega$, $y_{002} := \omega$, $y_{110} := 0.01\omega$ and $y_{101} := 0.95\omega$. $\omega$ chosen such that the problem is feasible

- D-optimal design

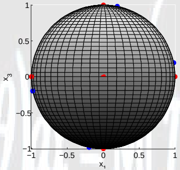


FIGURE: Support points $d = 1$ without constraint in red and constrained in blue