Approximate Optimal Designs for Multivariate Polynomial Regression

Fabrice Gamboa

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

Workshop Marseille

30th of April 2018

INSTITUT
de MATHEMATIQUES
de TOULOUSE

Agenda

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,
- ε 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 θ^* (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 ξ ?

Concave matricial criteria

Wish to *maximize* the information matrix with respect to ξ

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

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

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

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

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

Optimal design

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

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

- $\phi_q(M)$ 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

Big problem : description of all possible information matrices

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

Overview

Our frame : polynomial regression

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

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

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

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

Some facts and notations

- $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ 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 $(x^\alpha)_{|\alpha| \leqslant d}$, $|\alpha| := \alpha_1 + \cdots + \alpha_n$.

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

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

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

Overview

Moments, the moment cone and the moment matrix

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

- $\mathcal{M}_+(\mathcal{X})$ is the cone of nonnegative Borel measures supported on $\mathcal{X}$ (the dual of cone of nonnegative elements of $\mathcal{C}(\mathcal{X})$)
- $\mathbf{y} = y_\alpha(\mu) = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ moment sequence of μ
- $\mathcal{M}_d(\mathcal{X})$ moment cone :=convex cone of truncated sequences

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

- $\mathcal{I}$ only depends on $\mathcal{M}_{2d}(\mathcal{X})$

  Recall

$$\mathcal{I} := \left\{ \int_{\mathcal{X}} F(x)F^T(x)dP(x), \ \ P \in \mathbb{P}(\mathcal{X}) \right\}.$$

Main tool

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

Theorem

$\mathcal{M}_d(\mathcal{X}) = \mathcal{P}_d(\mathcal{X})^\star$ *and* $\mathcal{P}_d(\mathcal{X}) = \mathcal{M}_d(\mathcal{X})^\star$

- **Good new :** for $\mathcal{X} = [a, b]$, A non negative polynomial is a sum of two squared polynomial (Lukacs-Markov theorem). So, $\mathcal{M}_d(\mathcal{X})$ is representable using positive semidefinitness of Hankel matrices- So that Optimal design may be solved by SDP
- **Bad new :** no more true in general ! ! !
- **Good new :** when degree increases a non negative polynomial is more and more often a sum of squared polynomials

Our recipe : Increase the dimension, SDP and projection

Good new : when degree goes to infinity warranty to achieve the optimal. In practice finite degree convergence

Approximations of the moment cone

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

$$\mathcal{M}^{\mathrm{SDP}}_{2(d+\delta)}(\mathcal{X}) := \Big\{ \mathbf{y}_{d,\delta} \in \mathbb{R}^{\binom{n+2d}{n}} \; : \; \exists \mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}} \text{ such that } \quad (4)$$

$$\mathbf{y}_{d,\delta} = (y_{\delta,\alpha})_{|\alpha| \leqslant 2d} \text{ and}$$

$$M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0,\; M_{d+\delta-v_j}(g_j \mathbf{y}_\delta) \succcurlyeq 0,\; j = 1, \dots, m \Big\}.$$

Good approximation

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

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

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

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 δ

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 (equivalence theorem in designer words). Then,*

1. $\rho_\delta \to \rho$ *as* $\delta \to \infty$,
2. *For every* α, $|\alpha| \leqslant 2d$, $\lim_{\delta\to\infty} y^\star_{d,\delta,\alpha} = y^\star_\alpha$,
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* $\mathcal{P}^{SOS}_{2(d+\delta)}(\mathcal{X})$ *for some* δ*, then finite convergence takes place.*

$\mathcal{P}^{SOS}_{2(d+\delta)}(\mathcal{X})$ is a set of non negative polynomials of degree less than $2d$ built on sum of squares of polynomial of degrees $2(d+\delta)$ ponderated by the g_j

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 $d = 5$, δ=0). The points match with the known analytic solution to the problem (critical points of the Legendre polynomial)

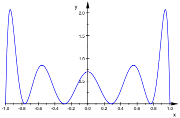

FIGURE: Polynomial $p^\star$ D-optimality, $n = 1$

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

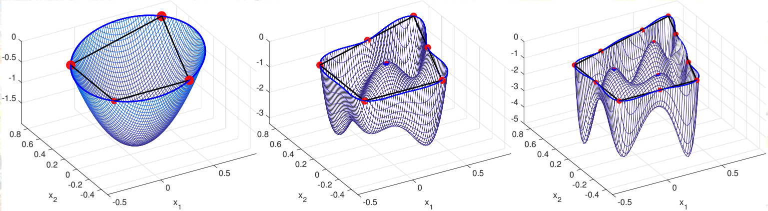

FIGURE: The polynomial $p_d^\star - \binom{2+d}{2}$ where for $d=1$, $d=2$, $d=3$. The red points correspond to the *good* level set

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

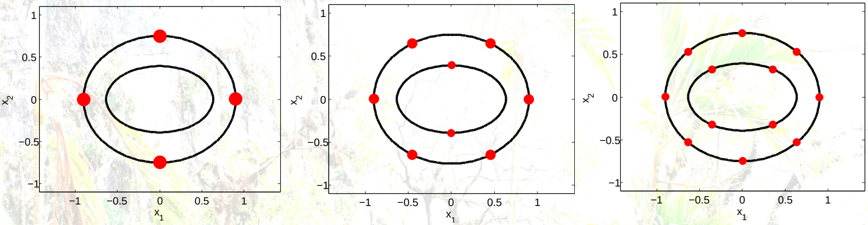

FIGURE: The boundary in bold black. The support of the optimal design measure (red points). Size of the points corresponds to the respective weights for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.

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

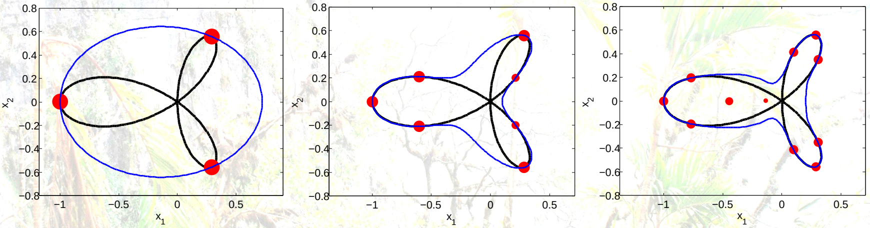

FIGURE: Boundary in bold black The support of the optimal design measure (red points). The *good* level set in thin blue $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right), $\delta = 3$

The 3-dimensional unit sphere

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

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

- D-optimal design

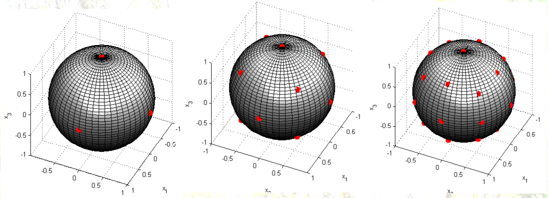

FIGURE: Optimal design in red $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 0$

The 3-dimensional unit sphere+constraint on moments

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

- D-optimal design

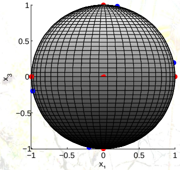

FIGURE: Support points $d = 1$ without constraint in red and constrained in blue

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 and moment problems

- 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.
- K. Schmudgen. The moment problem, Springer. 2017.
- 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.