Shape-constrained spline estimation of multivariate functions using semidefinite programming

Dávid Papp[1] Farid Alizadeh[1,2]

[1]Rutgers Center for Operations Research

[2]Rutgers, Management Science and Information Systems

ISMP, August 23–28, 2009

Shape-constrained function estimation

- Optimization models involving shape constraints
  1. Regression of a nonnegative function
  2. Convex/concave regression
  3. Design curve stays within a region
  4. Design curve has bounded curvature
  5. Arrival rate estimation for nonhomogeneous Poisson processes
  6. Probability density function estimation
  7. Assessing utility functions
  8. ...
- Common characteristic:
  - The approximating function must be nonnegative, monotone, convex, concave in some (or all) variables
- In this talk: polynomial splines of fixed degree are used

Nonnegative polynomials over $\Delta \subset \mathbb{R}^n$

Nonnegative basis approach:

- Choose a basis of nonnegative polynomials over Δ
  - Eg. Bernstein polynomials: $B_{n,k} = \binom{n}{k} x^k (1-x)^{n-k}$ (for $\Delta = [0,1]$).
- Optimize over nonnegative linear combinations of them
- LP, but not exact (proper subset of nonnegative polys)
- Selection of basis is critical
  - Eg. piecewise cubic Bernstein polynomial splines are dense in the set of bounded functions over $[0,1]$
  - But the standard (monomial) basis would be very bad

Nonnegative polynomials over $\Delta \subset \mathbb{R}^n$

Exact representation: easy in the univariate case

- Polynomial is nonnegative over $\mathbb{R}$ iff it is a finite "sum of squares":

$$p(t) = \sum_{i=1}^{k} q_i(t)^2, \quad \deg q_i \leq (\deg p)/2.$$

- Representable with second-order and semidefinite constraints
- Over $[a, b]$: nonnegative iff it is "weighted sum of squares":

$$p(x) = (x-a)s_1(x) + (b-x)s_2(x), \ \textit{or}$$

$$p(x) = s_1(x) + (x-a)(b-x)s_2(x),$$

where s_1, s_2 are sum of squares

Nonnegative polynomials over $[a, b]$

Theorem (Markov–Lukacs, Nesterov)

Let $p = \sum_{i=0}^{n} p_i x^i$ be a polynomial of odd degree, $n = 2k + 1$, and $a < b$. Then the following are equivalent.

1. $p(x) \geq 0$ *for all* $x \in [a, b]$.
2. $p(x) = (x - a)s_1(x) + (b - x)s_2(x); \quad \deg s_i \leq k, \quad s_i$ *SOS.*
3. *There exist* $(k + 1) \times (k + 1)$ *symmetric matrices* $\mathbf{X} = (x_{ij})_{i,j=0}^{k}$ *and* $\mathbf{Y} = (y_{ij})_{i,j=0}^{k}$ *satisfying*

$$\mathbf{X}, \mathbf{Y} \succcurlyeq 0,$$
$$p_m = \sum_{i+j=m} (-a x_{ij} + b y_{ij}) + \sum_{i+j=m-1} (x_{ij} - y_{ij})$$

for all $m = 0, \ldots, 2k + 1$.

Even degree polynomials: same but different weights.

Multivariate nonnegative polynomials

- Recognizing multivariate nonnegative polynomials is NP-hard, so we cannot expect to optimize over them efficiently
- Using a nonnegative basis might work
- (Weighted) sum of squares approach might work
- Caveat: typically a proper subset
- Bivariate quadratic polys over most "interesting" domains are easy to characterize as WSOS (Miccheli&Pinkus, '89)
- Bivariate, higher degree: much harder

Multivariate nonnegative polynomials

- An example of the kind of theorem we need:

Theorem (Miccheli, Pinkus, '89)

A bivariate quadratic polynomial p is nonnegative over the triangle $\{(x, y) \mid x \geq 0, y \geq 0, 1 - x - y \geq 0\}$ *if and only if*

$$p(x, y) = \alpha xy + \beta x(1 - x - y) + \gamma y(1 - x - y) + s(x, y),$$

where $\alpha, \beta, \gamma \geq 0$, *and* $s(x, y)$ *is SOS of linear functions.*

- Similar theorems hold for other simple polyhedra
- Analogous claims do not hold even for bivariate cubics.

Multivariate nonnegative polynomials

Theorem ("Schmüdgen's Positivstellensatz")

Let $\Delta = \{x \in \mathbb{R}^d \mid r_j(x) \geq 0 \text{ for } j = 1, \ldots, m\}$ *for some* $r_j \in \mathbb{R}[x_1, \ldots, x_d]$ *be compact. Then the polynomial* p *is positive over* Δ *only if it is weighted sum of squares with weights* r_j*:*

$$p(x) = s_0(x) + \sum_{j=1}^{k} w_j(x) s_j(x),$$

where every s_j *is a sum of squares polynomial, and every* w_j *is a product of distinct* r_j*'s.*

- Problem: unlike in the previous theorem, there is no (useful) bound on the degree of s_j.

Nonnegative splines

Univariate case:

- Spline: smooth (up to given order), piecewise polynomial
- Piecewise nonnegativity: semidefinite constraints
  - Degree ≤ 3: simplifies to second-order conic constraint
- Continuity of the function and its derivatives: linear equality constraints
- Scaling is necessary
- For higher degree, the monomial basis in the representation $p = \sum_{i=0}^{n} p_i x^i$ needs to be changed

Nonnegative splines

Multivariate case:

- Same, except that we choose a weighted sum of squares restriction to model nonnegativity
- Models may have many parameters to cross-validate for: the size of the SDP representation is critical
- Splines make good approximation possible even when nonnegativity is not approximated well by SOS polys

Weighted sum of squares polynomials and splines

- A subset of nonnegative polynomials over $\Delta \subset \mathbb{R}^n$:

$$p(x) = \sum_{j=1}^{k} w_j(x) s_j(x), \quad x \in \mathbb{R}^n,$$

  - Weights w_j: arbitrary nonnegative functions over Δ
  - s_j: sums of squares of functions from any (but fixed) finite dimensional functional space (given by its basis)
  - always semidefinite representable (Nesterov)
- Choice of weights and bases: can the obtained splines approximate "interesting" (eg., bounded) functions arbitrarily well, as the subdivision is refined?

Weighted sum of squares polynomials and splines

Theorem (Alizadeh, P.)

Consider a fixed set of WSOS polynomials $\mathcal{W}$ over a compact Δ. Let $f_1, \ldots, f_k \in \mathcal{W}$ such that for some $M \in \mathbb{R}$, $M - f_i \in \mathcal{W}$, too. Then the piecewise $\mathcal{W}$-splines are dense (wrt. to the uniform norm topology) in the cone of nonnegative $\mathcal{F}$-splines, where $\mathcal{F} = \mathrm{span}(\{f_1, \ldots, f_k\})$.

- If $\mathcal{F}$ is large enough (Bernstein polynomials of a given degree, etc.), then Theorem implies that the piecewise WSOS splines are dense in the cone of bounded nonnegative functions.
- "Rich" WSOS systems with very compact SDP representation can be found this way
  - Example (bicubic nonnegative polys over $\Delta = [0,1]^2$):
    - weights: {xy, x(1-y), (1-x)y, (1-x)(1-y)},
    - basis for each weight: $\{xy, x(1-y), (1-x)y, (1-x)(1-y)\}$

Nonparametric regression

Input

- data points $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, \ldots, n$
- function value z_i corresponding to each (x_i, y_i)

Objective

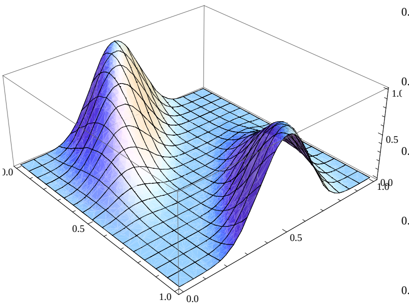
- find a nonnegative spline f of degree m that minimizes the mean squared error $n^{-1} \sum_{i=1}^{n} (z_i - f(x_i, y_i))^2$
- alternative: sum of absolute differences
- smoothing term could be added:

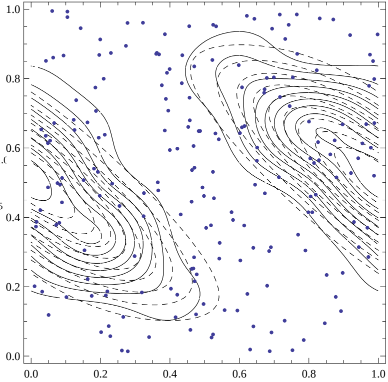
$$+\lambda \sum_{\alpha_1+\alpha_2=m} \frac{m!}{\alpha_1!\alpha_2!} \iint \left(\frac{\partial^m f}{\partial x^{\alpha_1} \partial y^{\alpha_2}} \right)^2 \mathrm{d}x\mathrm{d}y$$

Result in SDPs.

Nonparametric regression - example

- Test function: $t(x, y) = \sin(\pi(x + y))^4 \sin(\pi x)^2$
- 200 samples; large, Gaussian noise added
- weights: xy, x(1-y), (1-x)y, (1-x)(1-y),
- bases: $\{xy, x(1-y), (1-x)y, (1-x)(1-y)\}$





Estimation of posterior probabilities

[Villalobos, Wahba, '87]

Input

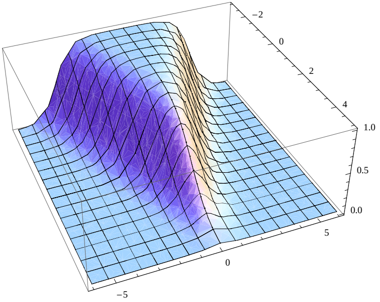
- data points $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, \ldots, n$
- each belongs to population A or B

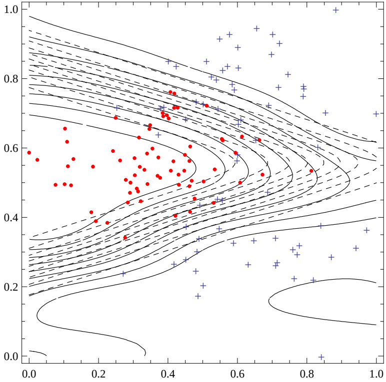
Objective

- estimate the probability $p(x, y)$ that a new point (x, y) belongs to population A
- can be modeled as a regression problem with constraints $0 \leq p(x, y) \leq 1$

Estimation of posterior probabilities - example

- test distribution: mixture of normals
- 70 samples each
- weights: xy, x(1-y), (1-x)y, (1-x)(1-y),
- bases: $\{xy, x(1-y), (1-x)y, (1-x)(1-y)\}$





Summary and outlook

Summary

- When using splines for function estimation, shape constraints can be modeled by nonnegative polynomial constraints
- Such constraints are replaced by weighted sum of squares constraints
- Models involve semidefinite constraints, whose number and size must be kept small for efficiency
- Initial computational results are promising

Summary and outlook

Future work

- Univariate: “spread the word”
- Multivariate: better characterizations for nonnegative polynomials of small degree
- Convexity: the multivariate case needs characterizations of PSD matrix valued polynomials
  - partial results to come on ISMP
- Computational issue: need solver for SDP with arbitrary convex objective function
  - maximum (log-)likelihood estimation, etc.