Learning Positive Functions in a Hilbert Space

J. Andrew Bagnell[1] and Amir-massoud Farahmand[2,1]

1: The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

2: Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

High-Level Summary

- **Problem:** Given a loss function and data, design an estimator with a guaranteed positive output.
- Approach #1 (Trivial): Use any regular estimator, but truncate negative values.
- **Approach #2:** Extend the concept of Sum of Squares, which guarantees non-negativity, to reproducing kernel Hilbert spaces (RKHS) and define the estimator in that space.

Given $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$, empirical loss $L_n(\cdot)$, function space $\mathcal{F}$, and regularizer $J(\cdot)$, we want to solve

$$\hat{f} \leftarrow \operatorname*{argmin}_{f \in \mathcal{F}, f \geq 0} L_n(f) + \lambda J(f).$$

Sum of Squares Polynomials

Real polynomial $p(x)$ is called *Positive semidefinite* or nonnegative if $p(x) \geq 0 \; \forall x$.

Sum of Squares (SoS): If there exist some other polynomials $q_1(x), q_2(x), \dots$ such that $p(x) = \sum_i q_i^2(x)$.

- Verifying SoS is computationally feasible (through Semidefinite Programming), but verifying PSD is not (NP-Hard).
- For univariate polynomials: SoS and PSD are equivalent.
- For multivariate polynomials: Not all PSD polynomials can be written as a SoS, but they are an important subclass with some denseness properties.

Let us define the vector of monomials: $\phi(x) = \left[1, x, \dots, x^d\right]^\top$. Let $p(x) = \sum_{i=1}^m q_i^2(x)$ (SoS representation). We can write $\bar{q}(x) \triangleq \left[q_1(x) \cdots q_m(x)\right]^\top = V\phi(x)$, and

$$p(x) = \sum_{i=1}^m q_i^2(x) = \bar{q}^\top(x)\bar{q}(x) = \phi^\top(x) \underbrace{V^\top V}_{\triangleq Q} \phi(x).$$

Key insight: ϕ does not have to be the features defined by monomials.

Space of Sum of Squares functions in an RKHS

$\mathrm{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, corresponding RKHS $\mathcal{H}$, and associated feature map $\phi : \mathcal{X} \to \mathcal{H}$, defined as $\phi(x) = (\phi_i(x))_{i \in \mathcal{I}}$.

The space of Sum of Squares (SoS) w.r.t. ϕ is defined as

$$\mathcal{S} \triangleq \left\{ x \mapsto \phi^\top(x) Q \phi(x) \, : \, Q \succeq 0 \right\}.$$

Any function $f \in \mathcal{S}$ is nonnegative.

$\mathcal{S}$ is not a subspace of $\mathcal{H}$, but we can construct another RKHS in which $\mathcal{S}$ is a subspace. Define a new feature map $\psi : \mathcal{X} \to \mathcal{H}'_0$ as

$$\psi(x) = (\phi_i(x) \cdot \phi_j(x))_{i,j \in \mathcal{I}}.$$

which has the kernel K'

$$\mathrm{K}'(x,y) \triangleq \langle \psi(x) \, , \, \psi(y) \rangle_{\mathcal{H}'_0} = \sum_{i,j \in \mathcal{I}} \phi_i(x)\phi_j(x) \; \phi_i(y)\phi_j(y) =$$

$$= \sum_{i \in \mathcal{I}} \phi_i(x)\phi_i(y) \sum_{j \in \mathcal{I}} \phi_j(x)\phi_j(y) = \langle \phi(x) \, , \, \phi(y) \rangle^2 = \mathrm{K}^2(x,y).$$

Observe that

$$\mathcal{S} = \left\{ x \mapsto \phi^\top(x) Q \phi(x) \, : \, Q \succeq 0 \right\} = \left\{ x \mapsto \sum_{i,j \in \mathcal{I}} Q_{ij} \phi_i(x) \phi_j(x) \, : \, Q \succeq 0 \right\}$$

$$= \left\{ x \mapsto \sum_{i,j \in \mathcal{I}} Q_{ij} \psi_{I(i,j)}(x) \, : \, Q \succeq 0 \right\} \subset \mathcal{H}'.$$

Representer Theorem

For a particular set $\{X_i\}_{i=1}^n$, we define $\mathcal{S}_n$:

$$\mathcal{S}_n = \left\{ x \mapsto \sum_{l=1}^n \alpha_l \mathrm{K}'(x, X_i) \, : \, \alpha \in \mathbb{R}^n \right\} \cap \mathcal{S}.$$

Theorem 1 (Representer Theorem)**.** *Let L_n be a convex empirical loss function. Then for all $\lambda > 0$, there exists a unique solution $\hat{f} \in \mathcal{S}$ satisfying*

$$L_n(\hat{f}) + \lambda \left\| \hat{f} \right\|_{\mathcal{H}'}^2 = \inf_{f \in \mathcal{S}} L_n(f) + \lambda \left\| f \right\|_{\mathcal{H}'}^2.$$

Moreover, $\hat{f} \in \mathcal{S}_n$.

Semidefinite Programming Formulation

By representer theorem: $f(x) = \sum_{l=1}^n \alpha_l \mathrm{K}'(X_i, x)$ under the condition that the function has an SoS representation, i.e., $f(x) = \phi(x)^\top Q \phi(x)$ for some $Q \succeq 0$.

Define a $d \times n$ matrix $\Phi = [\phi(X_1) \cdots \phi(X_n)]$ and an $n \times n$ diagonal matrix $A = \mathrm{diag}(\alpha) = \mathrm{diag}(\alpha_1, \dots, \alpha_n)$. We have $Q = \Phi A \Phi^\top$.

Q is $d \times d$, but has rank n, which can be much smaller than d. The constraint on PSDness of Q can be written as

$$\mathrm{eig}(Q) = \mathrm{eig}(\Phi A \Phi^\top) = \mathrm{eig}(\Phi \sqrt{A} \sqrt{A} \Phi^\top) = \mathrm{eig}(\sqrt{A} \underbrace{\Phi^\top \Phi}_{\triangleq G} \sqrt{A}) = \mathrm{eig}(GA).$$

Here $G = \Phi^\top \Phi$ is the $n \times n$ Grammian matrix. We have $\Phi_{ij} = \sum_{k \in \mathcal{I}} \phi_k(X_i) \phi_k(X_j) = \langle \phi(X_i) \, , \, \phi(X_j) \rangle_{\mathcal{H}} = \mathrm{K}(X_i, X_j)$.

So we can write:

$$\inf_{f \in \mathcal{S}} L_n(f) + \lambda \|f\|_{\mathcal{H}'}^2 = \inf_{\alpha \in \mathbb{R}^n} L_n \left(\sum_{l=1}^n \mathrm{K}'(X_l, \cdot) \alpha_l \right) + \lambda \alpha^\top K' \alpha.$$

$$\text{s.t.} \qquad G \, \mathrm{diag}(\alpha) \succeq 0$$

For squared loss, we get the following semidefinite program:

$$\min_{t, \alpha \in \mathbb{R}^n} \; t$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{I}_{n \times n} & L\alpha & 0_{n \times n} \\ \alpha^\top L^\top & t + 2\alpha^\top K'^\top Y & 0_{1 \times n} \\ 0_{n \times n} & 0_{n \times 1} & G \, \mathrm{diag}(\alpha) + \mathrm{diag}(\alpha) G \end{bmatrix}_{2n+1 \times 2n+1} \succeq 0.$$

Illustrations