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 \( D_n = \{(X_i, Y_i)\}_{i=1}^n \), empirical loss \( L_n(\cdot) \), function space \( F \), and regularizer \( J(\cdot) \), we want to solve
\[
\hat{f} \leftarrow \text{argmin}_{f \in F, \lambda \geq 0} L_n(f) + \lambda J(f).
\]

### Space of Sum of Squares functions in an RKHS

\[ \kappa : X \times X \rightarrow \mathbb{R} \], corresponding RKHS \( H_0 \), and associated feature map \( \phi : X \rightarrow H_0 \) as defined as \( \phi(x) = (\phi(x))_{i,j \in \mathcal{I}} \).

The space of Sum of Squares (SoS) w.r.t. \( \phi \) is defined as
\[
\mathcal{S} = \{ x \mapsto \phi^\top(x)Q\phi(x) : Q \geq 0 \}.
\]

Any function \( f \in \mathcal{S} \) is nonnegative. \( \mathcal{S} \) is not a subspace of \( H_0 \), but we can construct another RKHS in which \( \mathcal{S} \) is a subspace. Define a new feature map \( \psi : X \rightarrow H'_0 \) as
\[
\psi(x) = (\phi(x) \cdot \phi_j(x))_{i,j \in \mathcal{I}}.
\]

which has the kernel \( \kappa' \)
\[
\kappa'(x,y) = (\psi(x) \cdot \psi(y))_{H'_0} = \sum_{i,j \in \mathcal{I}} \phi_i(x)\phi_j(y) = \sum_{i \in \mathcal{I}} \phi_i(x)\phi_i(y) + \sum_{i \in \mathcal{I}} \phi_i(x)\phi_j(y) = \langle \phi(x), \phi(y) \rangle^2 = \kappa^2(x,y).
\]

Observe that
\[
\mathcal{S} = \{ x \mapsto \phi^\top(x)Q\phi(x) : Q \geq 0 \} = \{ x \mapsto \sum_{i,j \in \mathcal{I}} Q_{ij}\phi_i(x)\phi_j(x) : Q \geq 0 \}
\]
\[
= \{ x \mapsto \sum_{i,j \in \mathcal{I}} Q_{ij}\psi_{ij}(x) : Q \geq 0 \} \subseteq H'_0.
\]

### Representation Theorem

For a particular set \( \{X_i\}_{i=1}^n \), we define \( S_n \):
\[
S_n = \left\{ x \mapsto \sum_{l=1}^n \alpha_l K(x, X_l) : \alpha \in \mathbb{R}^n \right\} \cap \mathcal{S}.
\]

**Theorem 1 (Representation 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
\[
L_n(\hat{f}) + \lambda \|\hat{f}\|^2_{H'} = \inf_{f \in \mathcal{S}} L_n(f) + \lambda \|f\|^2_{H'}.
\]

Moreover, \( \hat{f} \in S_n \).

### Semidefinite Programming Formulation

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

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

\( Q \) 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
\[
eig(Q) = \text{eig}(\Phi \Phi^\top) = \text{eig}(\sqrt{A}\sqrt{A}^\top) = \text{eig}(\sqrt{A}^\top \sqrt{A}) = \text{eig}(GA).
\]

Here \( G = \sqrt{A} \Phi \) is the \( n \times n \) Grammian matrix. We have \( \Phi_{ij} = \sum_{k \in \mathcal{I}} \phi_k(X_i)\phi_k(X_j) = (\phi(X_i), \phi(X_j))^H = \kappa(X_i, X_j) \).

So we can write:
\[
\inf_{f \in \mathcal{S}} L_n(f) + \lambda \|f\|^2_{H'} = \inf_{\alpha \in \mathbb{R}^n} L_n \left( \sum_{l=1}^n \kappa(X_l, \cdot)\alpha_l \right) + \lambda \alpha^\top K\alpha \quad \text{s.t.} \quad G \alpha \succeq 0.
\]

For squared loss, we get the following semidefinite program:
\[
\min_{\lambda, \alpha \in \mathbb{R}^n} \left[ \frac{1}{2} \sum_{i \in \mathcal{I}} \alpha_i^2 + \frac{1}{2} \sum_{i \in \mathcal{I}} \alpha_i \left[ \lambda + 2\alpha_i^2 + \lambda \alpha_i \right] \right] \geq 0.
\]

\[
\text{s.t.} \quad G \alpha \succeq 0.
\]

Illustrations

- Learning Positive Functions in a Hilbert Space

- J. Andrew Bagnell and Amir-massoud Farahmand

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

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