Finding Global Minima via Kernel Approximations

Ulysse Marteau-Ferey

INRIA - Ecole Normale Supérieure, Paris, France

Joint work with Alessandro Rudi and Francis Bach

LAAS - February 16, 2020

Outline

- **Global optimization using function values**
- **Representing non-negative functions**
- **Global optimization through sampling**
- **Conclusions and extensions**

Global optimization using function values

- Global optimization
- Optimal algorithms
- Optimal rates
- Convex reformulation

Global optimization

- **Zero-th order minimization**

$$\min_{x \in \Omega} f(x)$$

  - $\Omega \subset \mathbb{R}^d$ simple compact subset (e.g., $[-1, 1]^d$)
  - f with some bounded derivatives
  - access to function values

Global optimization

- **Zero-th order minimization**

$$\min_{x \in \Omega} f(x)$$

  - $\Omega \subset \mathbb{R}^d$ simple compact subset (e.g., $[-1, 1]^d$)
  - f with some bounded derivatives
  - access to function values

- **No convexity assumption**

Global optimization

- **Zero-th order minimization**

$$\min_{x \in \Omega} f(x)$$

  - $\Omega \subset \mathbb{R}^d$ simple compact subset (e.g., $[-1, 1]^d$)
  - f with some bounded derivatives
  - access to function values

- **No convexity assumption**

- **Many applications**
  - e.g., hyperparameter optimization in machine learning

Optimal algorithms

- **Goal**: Find $\hat{x} \in \Omega$ such that $f(\hat{x}) - \min_{x \in \Omega} f(x) \leqslant \varepsilon$
  - Lowest number of function calls
  - Worst-case guarantees over all functions f in some convex set $\mathcal{F}$

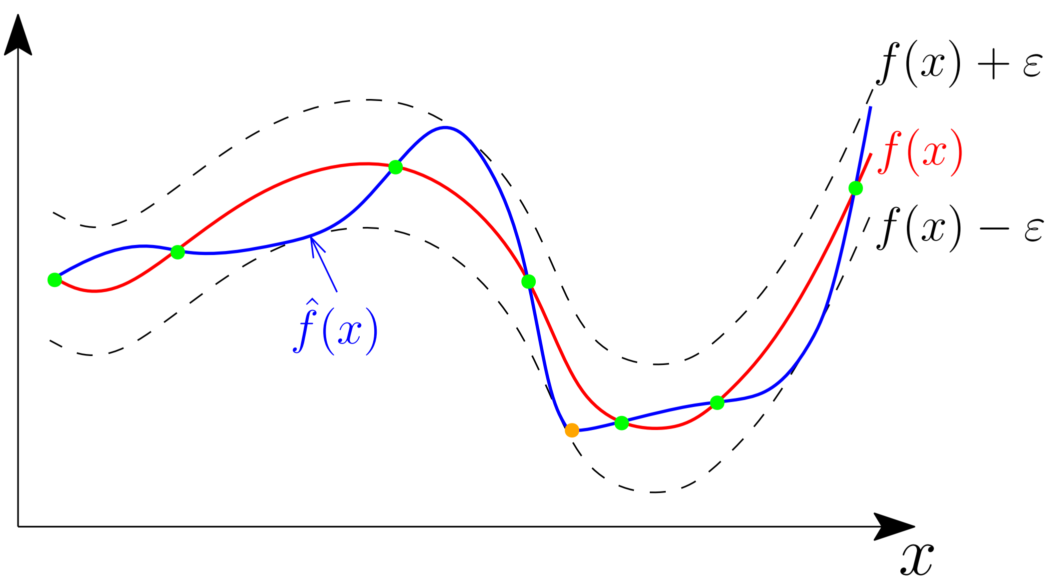
$$\sup_{f \in \mathcal{F}} \Big\{ f(\hat{x}) - \min_{x \in \Omega} f(x) \Big\} \leqslant \varepsilon$$

Optimal algorithms

- **Goal**: Find $\hat{x} \in \Omega$ such that $f(\hat{x}) - \min_{x \in \Omega} f(x) \leqslant \varepsilon$
  - Lowest number of function calls
  - Worst-case guarantees over all functions f in some convex set $\mathcal{F}$

$$\sup_{f \in \mathcal{F}} \left\{ f(\hat{x}) - \min_{x \in \Omega} f(x) \right\} \leqslant \varepsilon$$

- **Equivalence to uniform function approximation** (Novak, 2006)
  - Simplest algorithm: approximate f by $\hat{f}$ and minimize $\hat{f}$



Optimal rates

- **Optimal worst-case performance over** $\mathcal{F}$ (Novak, 2006)
  - $n =$ number of function evaluations
  - $\mathcal{F} =$ Lipschitz-continuous functions: $n \propto \varepsilon^{-d}$
  - $\mathcal{F} = m$ bounded derivatives: $n \propto \varepsilon^{-d/m}$

- **Smoothness to circumvent the curse of dimensionality**
  - NB: constants may depend (exponentially) in d

Optimal rates

- **Optimal worst-case performance over $\mathcal{F}$** (Novak, 2006)
  - $n =$ number of function evaluations
  - $\mathcal{F} =$ Lipschitz-continuous functions: $n \propto \varepsilon^{-d}$
  - $\mathcal{F} = m$ bounded derivatives: $n \propto \varepsilon^{-d/m}$
- **Smoothness to circumvent the curse of dimensionality**
  - NB: constants may depend (exponentially) in d
- **Algorithms have exponential running-time complexity**
  - "Approximate then optimize"

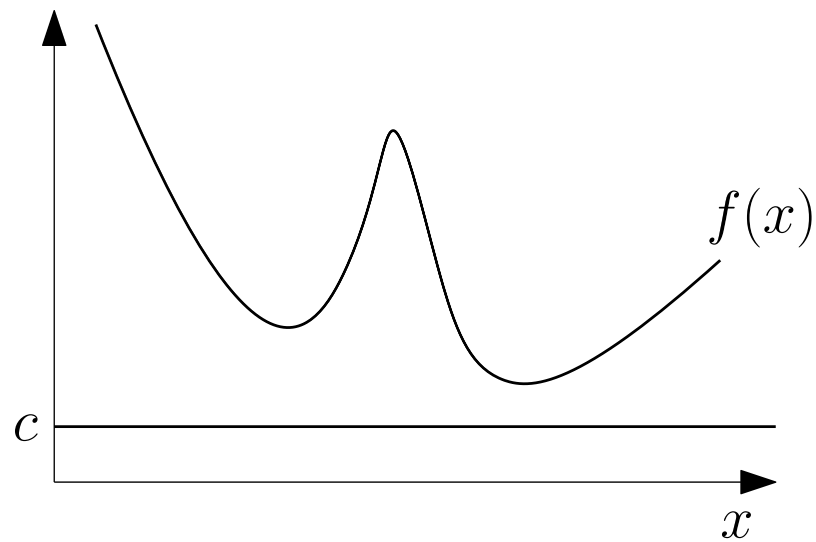
Optimal rates

- **Optimal worst-case performance over $\mathcal{F}$** (Novak, 2006)
  - $n =$ number of function evaluations
  - $\mathcal{F} =$ Lipschitz-continuous functions: $n \propto \varepsilon^{-d}$
  - $\mathcal{F} = m$ bounded derivatives: $n \propto \varepsilon^{-d/m}$
- **Smoothness to circumvent the curse of dimensionality**
  - NB: constants may depend (exponentially) in d
- **Algorithms have exponential running-time complexity**
  - "Approximate then optimize"
- **Algorithms with polynomial-time complexity in n?**
  - "Approximate and optimize"

Reformulations

- **Equivalent convex problem**

$$\min_{x\in\Omega} f(x) = \sup_{c\in\mathbb{R}} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) - c \geqslant 0$$

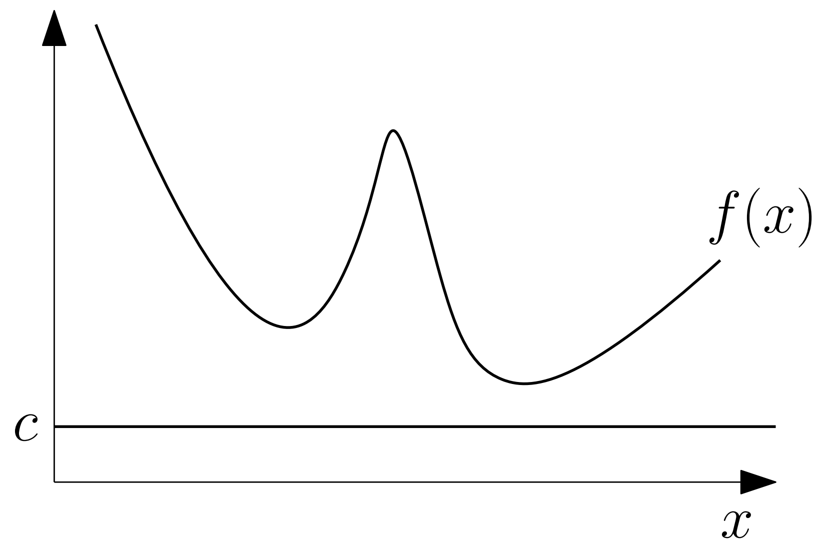


  - All optimization problems are convex!

Reformulations

- **Equivalent convex problem**

$$\min_{x \in \Omega} f(x) = \sup_{c \in \mathbb{R}} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) - c \geqslant 0$$



  - All optimization problems are convex!

- **Need to represent non-negative functions** (such as $f(x) - c$)

Representing non-negative functions

- Generic representation in a PSD cone
- The case of polynomials
- Infinite-dimensional representations using RKHS

Representing non-negative functions

- **Assumption**: $g(x)$ can be represented as $g(x) = \langle \phi(x), G\phi(x) \rangle$ with G symmetric operator

- **Example**: set of polynomials of degree $2r$
  with $\phi(x)$ composed of monomials of degree r, of dimension $\binom{d+r}{r}$

$$\phi(x) = (1, x_1, ..., x_d, x1x2, ..., x_1^r, ..., x_d^r)$$

Representing non-negative functions

- **Assumption**: $g(x)$ can be represented as $g(x) = \langle \phi(x), G\phi(x) \rangle$ with G symmetric operator

- **Example**: set of polynomials of degree $2r$
with $\phi(x)$ composed of monomials of degree r, of dimension $\binom{d+r}{r}$

$$\phi(x) = (1, x_1, ..., x_d, x1x2, ..., x_1^r, ..., x_d^r)$$

- **Non-negativity through "sums-of-squares"**
  - If $G \succcurlyeq 0$, then $\forall x \in \Omega,\ g(x) = \langle \phi(x), G\phi(x) \rangle \geqslant 0$
  - Sum of squares : eigen-decomposition $G = \sum_{i \in I} \lambda_i \ h_i \otimes h_i$,

$$g(x) = \sum_{i \in I} \lambda_i \langle \phi(x), (h_i \otimes h_i)\phi(x) \rangle = \sum_{i \in I} \lambda_i \langle \phi(x), h_i \rangle^2$$

Global optimization with sums of square polynomials

- **Replace** $f - c \geqslant 0$ by $f - c = \langle \phi(x), A\phi(x) \rangle$ with $A \succcurlyeq 0$

- **Sum-of-squares optimization** (Lasserre, 2001)

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathbb{R}^d,\ f - c = \langle \phi(x), A\phi(x) \rangle \qquad \text{(P)}$$

  – Equivalent to original problem if $f - f_*$ is SoS
  – Polynomial constraints can be added (and help !)
  – If not sum of squares : **hierarchies** (P_i) of degree $i \to \infty$

- **Guarantees**

  – the hierarchy converges to the optimum value
  – Lower bound when solving P_i : $f_i \leq f_*$.
  – Solve a finite dimensional SDP

Global optimization with sums of square polynomials

- **Replace** $f - c \geqslant 0$ by $f - c = \langle \phi(x), A\phi(x) \rangle$ with $A \succcurlyeq 0$

- **Sum-of-squares optimization** (Lasserre, 2001)

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} \quad c \quad \text{such that} \quad \forall x \in \mathbb{R}^d,\ f - c = \langle \phi(x), A\phi(x) \rangle \qquad \text{(P)}$$

  - Equivalent to original problem if $f - f_*$ is SoS
  - Polynomial constraints can be added (and help !)
  - If not sum of squares : **hierarchies** (P_i) of degree $i \to \infty$

- **Drawbacks**
  - Not all non-negative polynomials are SoS if $d \geq 1$
  - No guarantee a priori on the degree i needed for a given precision
  - Only for polynomials

Representing more general functions with RKHS

- **Reproducing Kernel Hilbert Space (RKHS)** :
  - Hilbert space of functions $g \in \mathcal{H},\ g : \mathbb{R}^d \to \mathbb{R}$
  - Reproducing property : $g(x) = \langle g, \phi(x) \rangle_{\mathcal{H}}$
  - Kernel : $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (computable)

Representing more general functions with RKHS

- **Reproducing Kernel Hilbert Space (RKHS)** :
  - Hilbert space of functions $g \in \mathcal{H}$, $g : \mathbb{R}^d \to \mathbb{R}$
  - Reproducing property : $g(x) = \langle g, \phi(x) \rangle_{\mathcal{H}}$
  - Kernel : $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (computable)

- **Example : Sobolev spaces** (Berlinet and Thomas-Agnan, 2011)
  - Sobolev spaces $H^s(\Omega)$ with $\Omega \subset \mathbb{R}^d$, $s > d/2$

$$\langle f, g \rangle_{H^s(\Omega)} = \sum_{|\alpha| \leq s} \int_{\Omega} \partial^\alpha f \; \partial^\alpha g$$

  - $s = d/2 + 1/2$: $k(x, y) = \exp(-\|x - y\|)$

Representing more general functions with RKHS

- **Reproducing Kernel Hilbert Space (RKHS)** :
  - Hilbert space of functions $g \in \mathcal{H}, \; g : \mathbb{R}^d \to \mathbb{R}$
  - Reproducing property : $g(x) = \langle g, \phi(x) \rangle_{\mathcal{H}}$
  - **Kernel** : $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (computable)

- **Everything can be expressed using only the kernel function k**
  - k is known for all Sobolev spaces with $s > d/2$.
  - $s = d/2 + 1/2 : \; k(x, y) = \exp(-\|x - y\|)$

Non-negative functions with RKHS

- **RKHS** $\mathcal{H}$:
  - Feature map $\phi(x) \in \mathcal{H}$ $(g(x) = \langle g, \phi(x) \rangle_{\mathcal{H}})$
  - Kernel : $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (computable)
- **RKHS Sum of Squares : (Marteau-Ferey et al., 2020)**
  - $g_A(x) = \langle \phi(x), A\phi(x) \rangle, \quad A \succeq 0$
  - Eigendecomposition $A = \sum_{i \in I} \lambda_i \; h_i \otimes h_i$

    SoS : $g_A(x) = \sum_{i \in I} \lambda_i \; h_i(x)^2, \; h_i \in \mathcal{H}$

Non-negative functions with RKHS

- **RKHS** $\mathcal{H}$:
  - Feature map $\phi(x) \in \mathcal{H}$ $(g(x) = \langle g, \phi(x) \rangle_{\mathcal{H}})$
  - **Kernel** : $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (computable)

- **RKHS Sum of Squares : (Marteau-Ferey et al., 2020)**
  - $g_A(x) = \langle \phi(x), A\phi(x) \rangle, \quad A \succeq 0$
  - Eigendecomposition $A = \sum_{i \in I} \lambda_i \; h_i \otimes h_i$
    SoS : $g_A(x) = \sum_{i \in I} \lambda_i \; h_i(x)^2, \; h_i \in \mathcal{H}$

- **Everything can be expressed using the kernel k**

  - Sum of squares of sobolev functions.

Global optimization through sampling

- Global optimization : formulation and theoretical result in the RKHS
- Controlled approximation through sampling
- Algorithms and illustrations

Going infinite-dimensional (Rudi, Marteau-Ferey, and Bach, 2020)

$$\sup_{c\in\mathbb{R},\ A\succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) - c = \langle \phi(x), A\phi(x)\rangle$$

- Sobolev space $\mathcal{H}$:
  - $s > d/2$ squared-integrable derivative
  - RKHS with feature map $\phi(x)$ and kernel $k(x, x')$
  - RKHS SoS : $g_A(x) = \langle \phi(x), A\phi(x)\rangle_{\mathcal{H}},\ A \succeq 0$

Going infinite-dimensional (Rudi, Marteau-Ferey, and Bach, 2020)

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) = c + \langle \phi(x), A\phi(x) \rangle$$

- **Sobolev space** $\mathcal{H}$
  - $s > d/2$ squared-integrable derivative
  - RKHS with feature map $\phi(x)$ and kernel $k(x, x')$
  - RKHS SoS : $g_A(x) = \langle \phi(x), A\phi(x) \rangle_{\mathcal{H}},\ A \succeq 0$
- **Theorem**: $\exists A_* \succcurlyeq 0$ such that $\forall x \in \Omega,\ f(x) = f_* + \langle \phi(x), A_*\phi(x) \rangle$
  - If f has isolated strict-second order minima in $\mathring{\Omega}$, and f is $(s+3)$-times differentiable

Going infinite-dimensional (Rudi, Marteau-Ferey, and Bach, 2020)

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) = c + \langle \phi(x), A\phi(x) \rangle$$

- **Sobolev space** $\mathcal{H}$
  - $s > d/2$ squared-integrable derivative
  - RKHS with feature map $\phi(x)$ and kernel $k(x, x')$
  - RKHS SoS : $g_A(x) = \langle \phi(x), A\phi(x) \rangle_{\mathcal{H}},\ A \succeq 0$

- **Theorem**: $\exists A_* \succcurlyeq 0$ such that $\forall x \in \Omega,\ f(x) = f_* + \langle \phi(x), A_*\phi(x) \rangle$
  - If f has isolated strict-second order minima in $\mathring{\Omega}$, and f is $(s+3)$-times differentiable

$\Rightarrow$ Equivalent to original problem, but infinite-dimensional

Controlled approximation through sampling

- **Subsample** n **points** $x_1, \dots, x_n \in \Omega$ **and solve**

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c - \lambda \operatorname{tr}(A) \text{ such that } \forall i \in \{1, \dots, n\},\ f(x_i) = c + \langle \phi(x_i), A\phi(x_i) \rangle$$

Controlled approximation through sampling

- **Subsample n points $x_1, \dots, x_n \in \Omega$ and solve**

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c - \lambda \operatorname{tr}(A) \ \text{ such that } \ \forall i \in \{1, \dots, n\},\ f(x_i) = c + \langle \phi(x_i), A\phi(x_i) \rangle$$

- **Finite-dimensional formulation through representer theorem (kernels !)**
  - Marteau-Ferey, Bach, and Rudi (2020)
  - SDP of dimension n :

$$\sup_{c \in \mathbb{R},\ B \succcurlyeq 0, B \in \mathbb{R}^{n \times n}} c - \lambda \operatorname{tr}(B) \ \text{ st } \forall i \in \{1, \dots, n\},\ f(x_i) = c + \Phi_i^\top B \Phi_i$$

- **Solvable in polynomial time** with precision ϵ in $O(n^{3.5} \log \frac{1}{\epsilon})$

Controlled approximation through sampling

- **Subsample n points $x_1, \dots, x_n \in \Omega$ and solve**

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c - \lambda \operatorname{tr}(A) \ \text{ such that } \ \forall i \in \{1, \dots, n\},\ f(x_i) = c + \langle \phi(x_i), A\phi(x_i) \rangle$$

- **Theorem** (Rudi, Marteau-Ferey, and Bach, 2020) Let $\hat{c}, \hat{A}$ be the result of the algorithm. Up to logarithmic terms :
  - if $n = \Omega\big(\varepsilon^{-d/(m-d/2-3)}\big)$ and the samples $(x_1, ..., x_n)$ are taken randomly from Ω, and if $\lambda = \varepsilon$, then it holds with probability at least $1 - \delta$:

$$|\hat{c} - f_*| \leq \varepsilon \log \tfrac{1}{\delta}$$

  - in that case, the complexity is $\Omega\big(\varepsilon^{-3d/(m-d/2-3)} \log \frac{1}{\epsilon}\big)$

- **Extension**: possible to find the minimizer $\hat{x}$.

Controlled approximation through sampling

- **Subsample n points $x_1, \dots, x_n \in \Omega$ and solve**

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c - \lambda \operatorname{tr}(A) \ \text{ such that } \ \forall i \in \{1, \dots, n\},\ f(x_i) = c + \langle \phi(x_i), A\phi(x_i) \rangle$$

- **Approximation guarantees** (Rudi, Marteau-Ferey, and Bach, 2020)
  - With random samples, $n \approx \varepsilon^{-d/(m-d/2-3)}$ (up to logarithmic terms)
  - To be compared to optimal rate $n \approx \varepsilon^{-d/(m-d/2)}$
  - Constraint $m \geqslant \frac{d}{2} + 3$ can be lifted

Controlled approximation through sampling

- **Subsample n points $x_1, \dots, x_n \in \Omega$ and solve**

$$\sup_{c\in\mathbb{R},\ A\succcurlyeq 0} \ c - \lambda \operatorname{tr}(A) \ \text{ such that } \forall i \in \{1,\dots,n\},\ f(x_i) = c + \langle \phi(x_i), A\phi(x_i)\rangle$$

- **Approximation guarantees** (Rudi, Marteau-Ferey, and Bach, 2020)
  - With random samples, $n \approx \varepsilon^{-d/(m-d/2-3)}$ (up to logarithmic terms)
  - To be compared to optimal rate $n \approx \varepsilon^{-d/(m-d/2)}$
  - Constraint $m \geqslant \frac{d}{2} + 3$ can be lifted

- **Subsampling inequalities as $f(x_i) \geqslant c$ directly?**
  - cannot improve on $n \approx \varepsilon^{-d}$

Final algorithm

- **Input:** $f : \mathbb{R}^d \to \mathbb{R}$, $\Omega \subset \mathbb{R}^d, n \geqslant 0, \lambda > 0, s > d/2$

1. **Sampling:** $\{x_1, \ldots, x_n\}$ sampled i.i.d. uniformly on Ω

2. **Feature computation**

   – Set $f_j = f(x_j)$, $\forall j \in \{1, \ldots, n\}$
   – Compute $K_{ij} = k(x_i, x_j)$ for k Sobolev kernel of smoothness s
   – Set $\Phi_j \in \mathbb{R}^n$ computed using a cholesky decomposition of K $\forall j \in \{1, \ldots, n\}$.

3. **Solve** $\max\limits_{c \in \mathbb{R}, B \succcurlyeq 0} c - \lambda \operatorname{tr}(B)$ s. t. $\forall j \in \{1, \ldots, n\}, f_j - c = \Phi_j^\top B \Phi_j$

- **Output:** c proxy for f_*

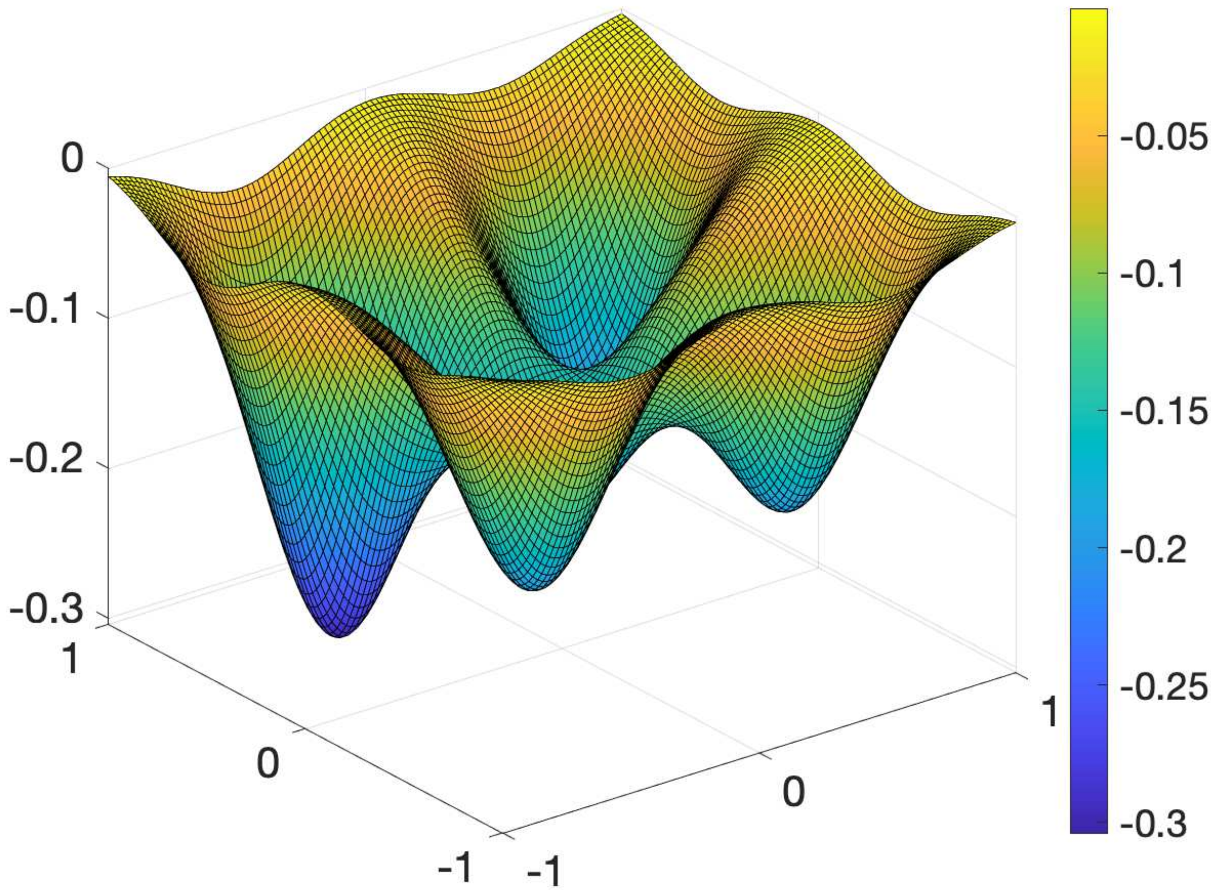
- One can extend the algorithm in order to compute a proxy of the minimizer

Opposite properties from Polynomial SoS !

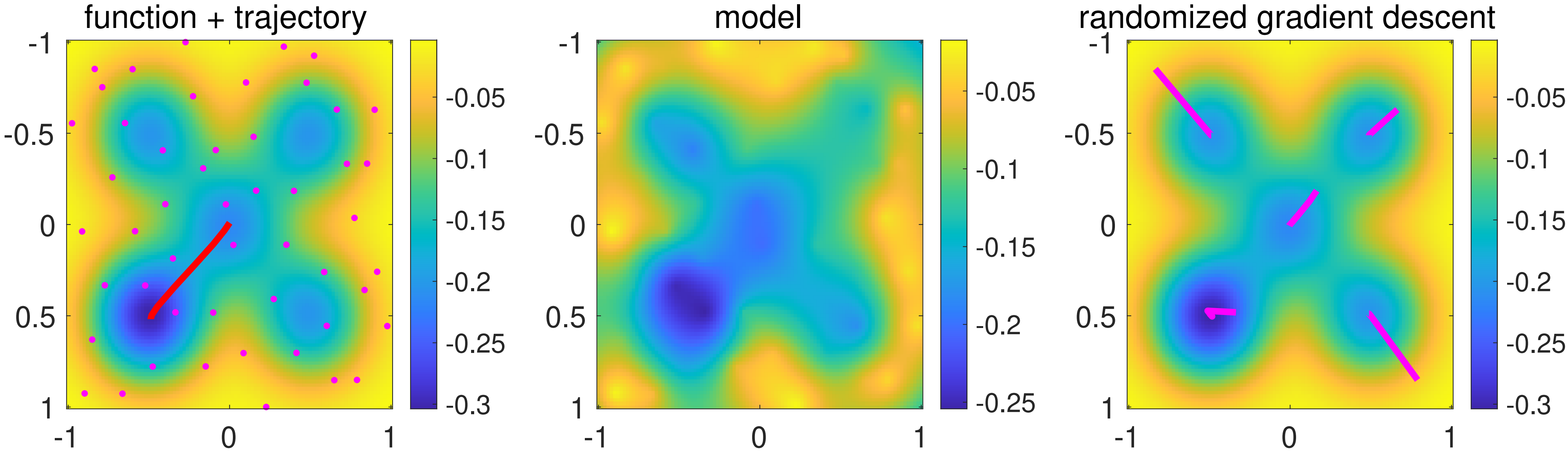
- "Always" possible to write a non-negative function as a RKHS SoS
- Bounds on the number of samples needed for a given precision
- Finite dimensional SDP with bounded complexity $O(n^{3.5} \log \frac{1}{\epsilon})$
- Breaks the curse of dimensionality (needs $\epsilon^{-d/m}$ samples) for smooth enough functions
- For the moment, **no certificate bound on the result of the algorithm**

Illustration

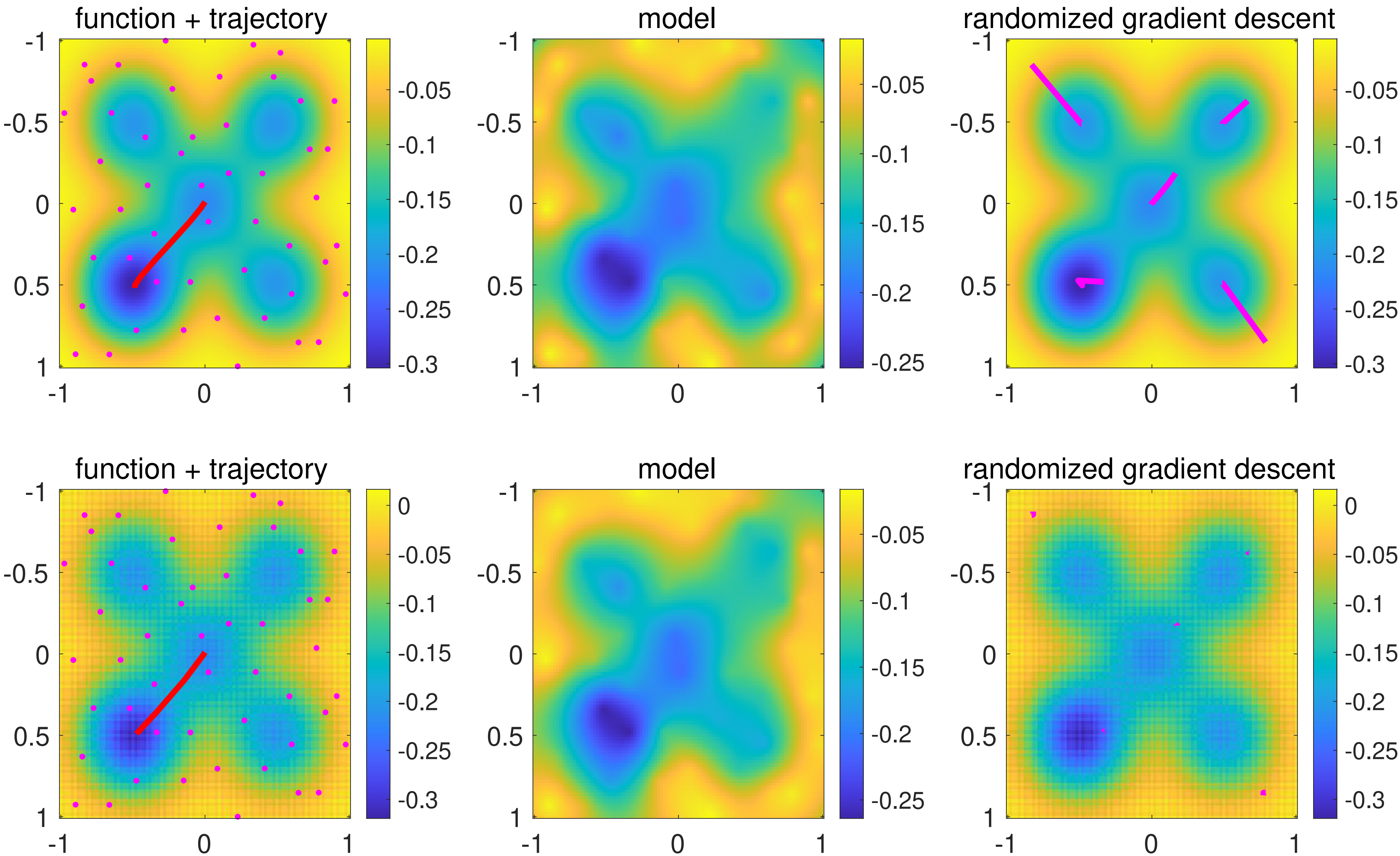
- **Minimization of two-dimensional function**



Illustration

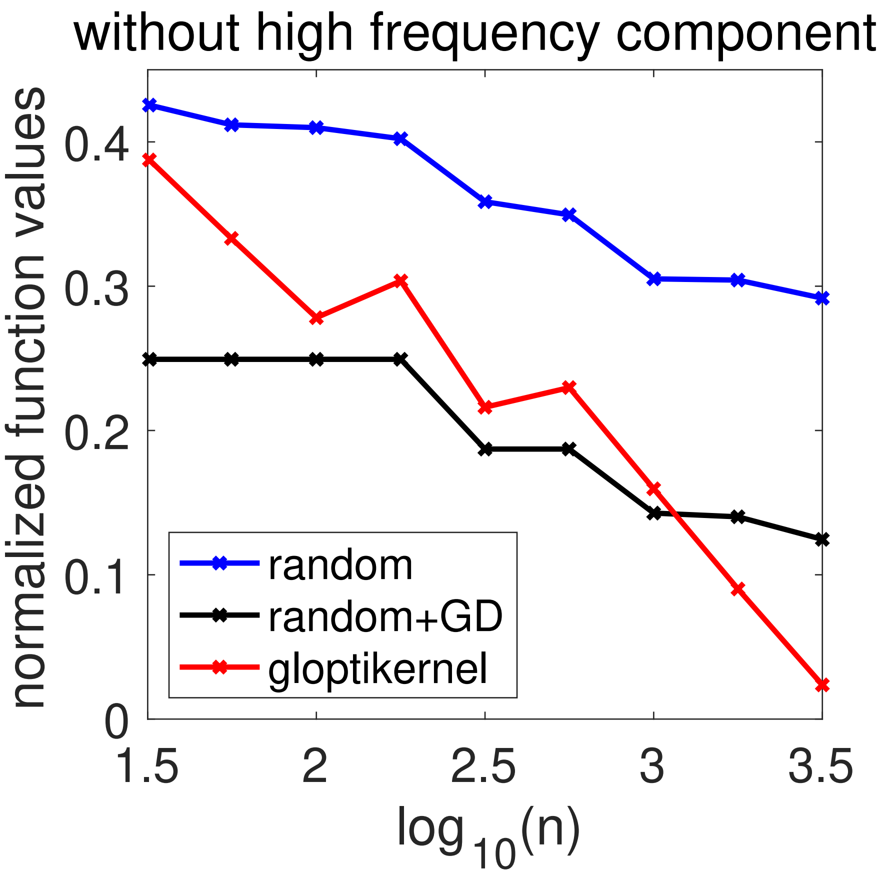


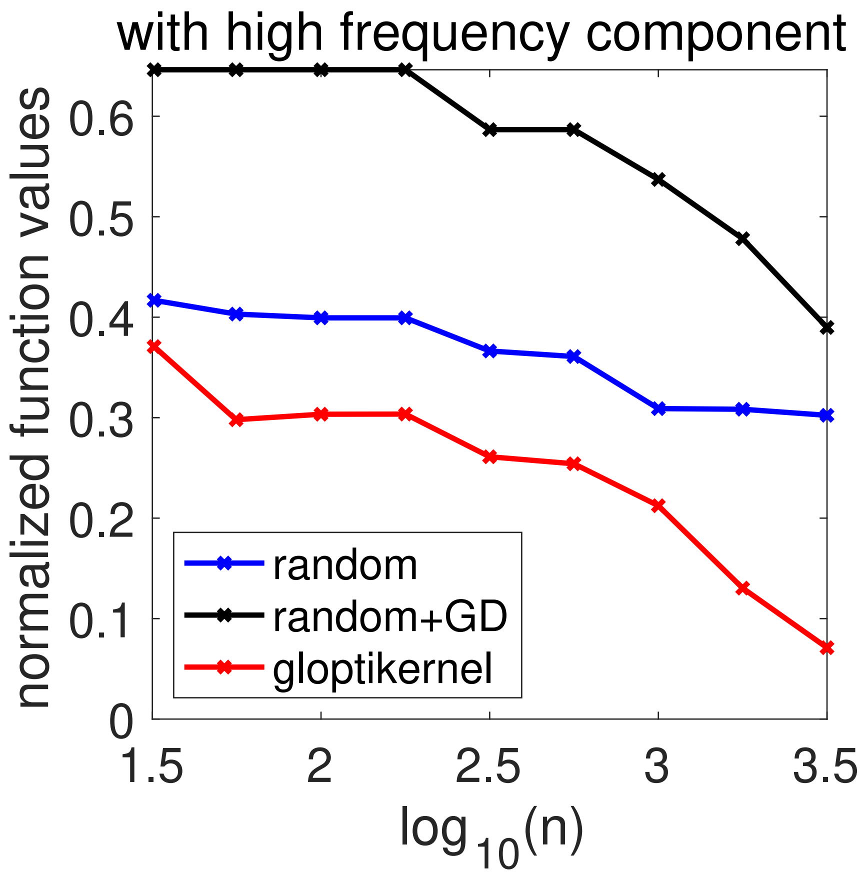
Illustration



Illustration

- **Minimization of eight-dimensional function**





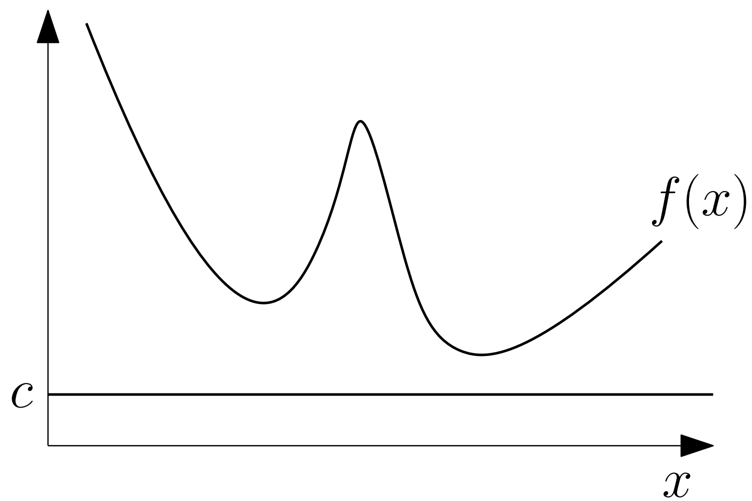
Extensions

- Duality
- Extionsions
- Conclusion

Duality

- **Primal problem**

$$\min_{x\in\Omega} f(x) = \sup_{c\in\mathbb{R}} c \quad \text{such that} \quad \forall x\in\Omega,\ f(x)-c \geqslant 0$$



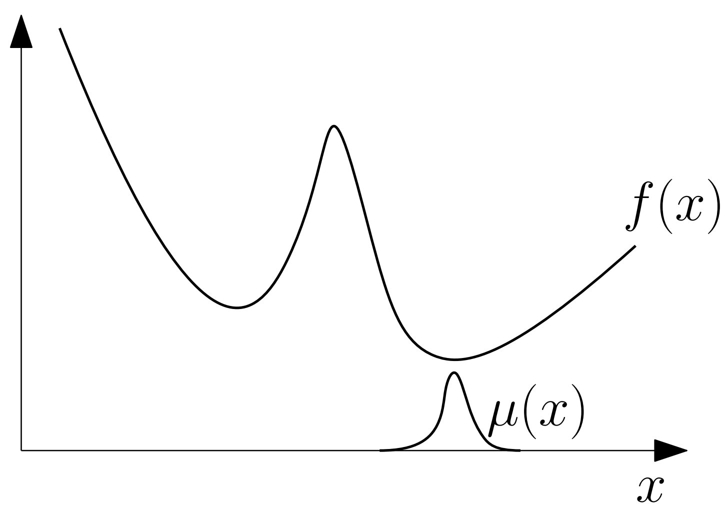
Duality

- **Primal problem**

$$\min_{x\in\Omega} f(x) = \sup_{c\in\mathbb{R}} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) - c \geqslant 0$$

- **Dual problem on probability measures**

$$\inf_{\mu\in\mathbb{R}^\Omega} \int_\Omega \mu(x) f(x) dx \quad \text{such that} \quad \int_\Omega \mu(x) dx = 1,\ \forall x \in \Omega,\ \mu(x) \geqslant 0$$



Duality with sums-of-squares

- **Primal problem**

$$\min_{x \in \Omega} f(x) = \sup_{c \in \mathbb{R},\ A \succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \Omega,\ f(x) - c = \langle \phi(x), A\phi(x) \rangle$$

- **Dual problem on signed measures**

$$\inf_{\mu \in \mathbb{R}^{\Omega}} \int_{\Omega} \mu(x) f(x) dx \quad \text{s. t.} \quad \int_{\Omega} \mu(x) dx = 1,\ \int_{\Omega} \mu(x) \phi(x) \otimes \phi(x) \succcurlyeq 0$$

  – Extension of results on polynomials (Lasserre, 2020)

Extension - I

- **Generic constrained optimization problem**

$$\inf_{\theta \in \Theta} \; F(\theta) \quad \text{such that} \quad \forall x \in \Omega, \; g(\theta, x) \geqslant 0$$

Extension - I

- **Generic constrained optimization problem**

$$\inf_{\theta \in \Theta} F(\theta) \quad \text{such that} \quad \forall x \in \Omega,\ g(\theta, x) \geqslant 0$$

- **Sums-of-squares reformulation**

$$\inf_{\theta \in \Theta,\ A \succcurlyeq 0} F(\theta) \quad \text{such that} \quad \forall x \in \Omega,\ g(\theta, x) = \langle \phi(x), A\phi(x) \rangle$$

  – Requires penalization by $\mathrm{tr}(A)$ and subsampling
  – Need representation as sums-of-squares to benefit from smoothness
  – Can be done in the primal or the dual

Extension - I

- **Generic constrained optimization problem**

$$\inf_{\theta \in \Theta} \; F(\theta) \quad \text{such that} \quad \forall x \in \Omega, \; g(\theta, x) \geqslant 0$$

- **Sums-of-squares reformulation**

$$\inf_{\theta \in \Theta, \; A \succcurlyeq 0} F(\theta) \quad \text{such that} \quad \forall x \in \Omega, \; g(\theta, x) = \langle \phi(x), A\phi(x) \rangle$$

  – Requires penalization by $\mathrm{tr}(A)$ and subsampling
  – Need representation as sums-of-squares to benefit from smoothness
  – Can be done in the primal or the dual

- **Application to optimal transport** (Vacher, Muzellec, Bach, Rudi, Vialard, 2021)

Extension - II

- **Constrained optimization problem**

$$\inf_{x \in \mathbb{R}^d} f(x) \quad \text{such that} \quad \forall x \in \Omega, \; g(x) \geqslant 0$$

Extension - II

- **Constrained optimization problem**

$$\inf_{x \in \mathbb{R}^d} f(x) \quad \text{such that} \quad \forall x \in \Omega,\ g(x) \geqslant 0$$

- **Sums-of-squares reformulation**

$$\sup_{c \in \mathbb{R},\ A \succcurlyeq 0,\ B \succcurlyeq 0} \quad c$$

$$\text{such that} \quad \forall x \in \Omega,\ f(x) = c + \langle \phi(x), A\phi(x) \rangle + g(x) \langle \phi(x), B\phi(x) \rangle$$

– Extension of results on polynomials (Lasserre, 2001)

Conclusion

- **Global optimization through kernel approximations**
  - Joint optimization and approximation
  - infinite-dimensional sums-of-squares representation
  - Controlled subsampling with guarantees

Conclusion

- **Global optimization through kernel approximations**
  - Joint optimization and approximation
  - infinite-dimensional sums-of-squares representation
  - Controlled subsampling with guarantees

- **Further extensions**
  - Efficient algorithms below $O(n^3)$ complexity
  - Other infinite-dimensional convex optimization problems

Conclusion

- **Global optimization through kernel approximations**
  - Joint optimization and approximation
  - infinite-dimensional sums-of-squares representation
  - Controlled subsampling with guarantees

- **Further extensions**
  - Efficient algorithms below $O(n^3)$ complexity
  - Other infinite-dimensional convex optimization problems

- **See** `arxiv.org/abs/2012.11978` and `francisbach.com/`

References

Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2011.

Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

Jean-Bernard Lasserre. The moment-SOS hierarchy and the Christoffel-Darboux kernel. Technical Report 2011.08566, arXiv, 2020.

Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi. Non-parametric models for non-negative functions. *Advances in Neural Information Processing Systems*, 33, 2020.

Erich Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349. Springer, 2006.

Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations. Technical Report 2012.11978, arXiv, 2020.