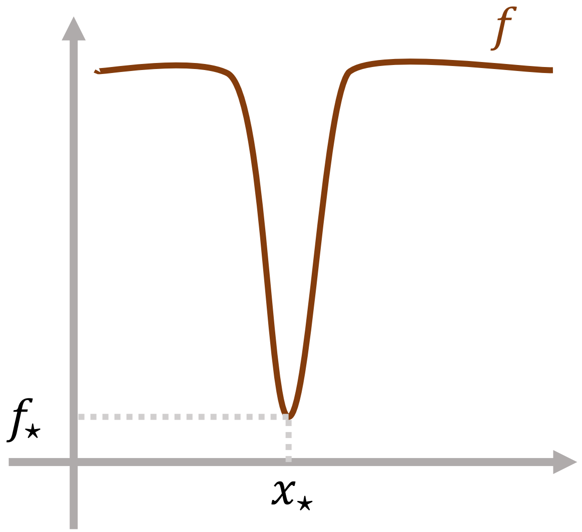
NEURAL INFORMATION PROCESSING SYSTEMS

GloptiNets: Scalable Non-Convex Optimization with Certificates

Gaspard Beugnot, Julien Mairal, Alessandro Rudi

Inria cnrs ENS ÉCOLE NORMALE SUPÉRIEURE

What can we expect from nonconvex optimization?



⛔ High norm $\Longrightarrow$ high difficulty.

Easy to devise **arbitrarily hard** problems.

Smoothness to the rescue:
Smooth function can be approximated in

$$n^{-\frac{s}{d}} \ll n^{-\frac{1}{d}}$$

⚠️ Constant can be exponential in d

💡 Nonconvex optimization is hard, but **smoothness allows to escape the curse of dimensionality**

Key ingredient: **kernel Sum-of-Squares** (k-SoS)

$$g(x) = \phi(x)^\top G\, \phi(x), \qquad G \succcurlyeq 0$$

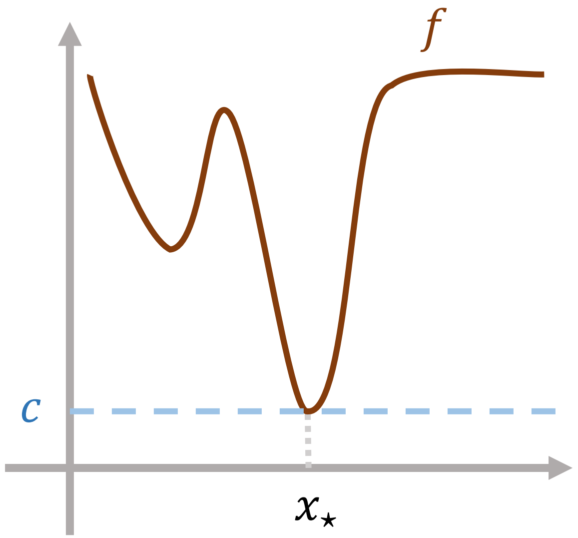
Analogous of linear kernel function: $f(x) = \omega^\top \phi(x)$
Good properties:

- Positive everywhere by design
- Convex in the parameters
- Universal approximators

Dense set of inequalities ?

$\forall x \in \mathcal{X}, \quad f(x) \geq c$

$\forall x \in \mathcal{X}, \quad f(x) - c = g(x),$
$g \in$ k−SoS



Global optimization

$$f_\star = \inf_{x\in\mathcal{X}} f(x)$$
$$= \sup_{c\in\mathbb{R}} c \quad \text{s.t.} \quad \forall x \in \mathcal{X}, f(x) \geq c$$

Applications
Optimal Transport, Density modeling, black-box optimization, Kalman filtering…

Marteau-Ferey et al. NeurIPS 2020

General recipe for nonconvex optimization

$f_\star = \inf_{x\in\mathcal{X}} f(x)$

$f_\star = \sup_{c\in\mathbb{R}} c \quad \text{s.t.} \quad \forall x \in \mathcal{X}, f(x) \geq c$ — Definition

$f_\star = \sup_{c\in\mathbb{R}, g\geq 0} c \quad \text{s.t.} \quad \forall x \in \mathcal{X}, f(x) - c = g(x)$ — Dense set of equality

$f_\star = \sup_{c\in\mathbb{R}, g\geq 0} c - \|f - c - g\|_{\mathcal{L}_\infty(\mathcal{X})}$ — Penalized version

$f_\star \geq \sup_{c\in\mathbb{R}, g\in\mathcal{G}} c - \|f - c - g\|_{\mathcal{L}_\infty(\mathcal{X})}$ — Strengthen constraint

$f_\star \geq \sup_{c\in\mathbb{R}, g\in\mathcal{G}} c - \|f - c - g\|_F$ — Strenthen upper bound — $\|u\|_{\mathcal{L}_\infty(\mathcal{X})} \leq \|u\|_F$

Bound on the minimum is valid for ***any*** c, g !!!

GloptiNets:

$\mathcal{G}$ = k-SoS model

$$\|u\|_F = \sum_{\omega\in\mathbb{Z}^d} |\hat{f}_\omega|$$

Probabilistic estimate of the F-Norm

$$|f(x)| = \left|\sum_{\omega\in\mathbb{Z}^d} \hat{f}_\omega e^{2\pi i \omega\cdot x}\right| \leq \sum_{\omega\in\mathbb{Z}^d} |\hat{f}_\omega| = \sum_{\omega\in\mathbb{Z}^d} \frac{|\hat{f}_\omega|}{\hat{\lambda}_\omega} \times \hat{\lambda}_\omega = \mathbb{E}_{\omega\sim\hat{\lambda}}\left[\frac{|\hat{f}_\omega|}{\hat{\lambda}_\omega}\right] = \|f\|_F$$

$\frac{|\hat{f}_\omega|}{\hat{\lambda}_\omega}$: unbiased estimate of $\|f\|_F$!! **Variance**?

$$\mathrm{Var}\frac{|\hat{f}_\omega|}{\hat{\lambda}_\omega} \leq \mathbb{E}_{\omega\sim\hat{\lambda}}\left[\left(\frac{|\hat{f}_\omega|}{\hat{\lambda}_\omega}\right)^2\right] = \sum_{\omega\in\mathbb{Z}^d} \frac{|\hat{f}_\omega|^2}{\hat{\lambda}_\omega} = \|f\|^2_{\mathcal{H}_\lambda}$$

$\mathcal{H}_\lambda$: RKHS associated with $\hat{\lambda}_\omega$

$$K(x,y) = \sum_{\omega\in\mathbb{Z}^d} \hat{\lambda}_\omega \, e^{2\pi i \omega\cdot(x-y)}$$

Unbiased estimator
\+ variance
\+ Chebychev bound / Median-of-Means / …
= **Bound on F norm with proba $1 - \delta$**

GloptiNets

Requirements
$\hat{f}_\omega$: Fourier[1] coeff of f
$\|f\|_{\mathcal{H}_\lambda}$: norm of f

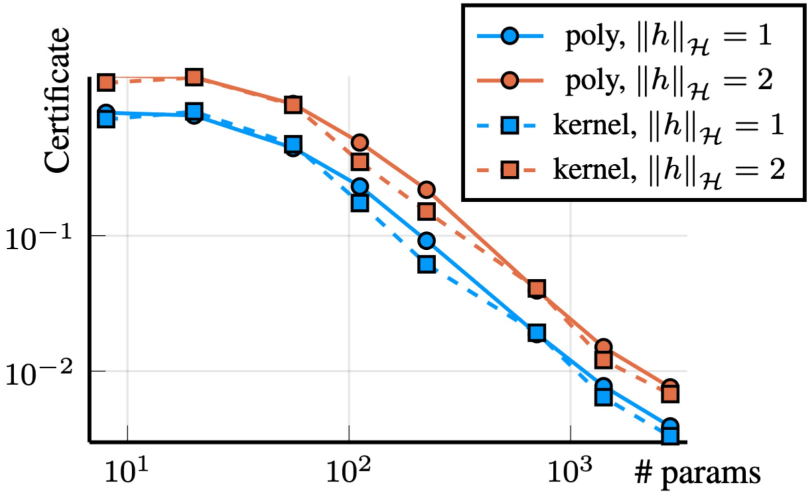
Input
h: smooth function on $(-1,1)^d$
$\hat{x}$: candidate
δ: confidence

Output
Certificate ϵ_δ s.t.
$|h(\hat{x}) - h(x_\star)| \leq \epsilon_\delta$
With proba. $1 - \delta$.

Key ideas
No *a priori* certificates; but an *a posteriori* guarantee.
Leverage the good empirical optimization of **overparametrized functions** with **GPU computations**

Solve $L(g) = c - \|f - c - g\|_F \leq f_\star$ with g an OP k-SoS model

Experiments
No alternative we are aware of, except when f is a *polynomial*.
- Complexity only depends on the **norm of f**
- The **bigger** the model, the **tighter** the certificate



Certificate valid for any c, g...

So minimizing the certificate gap is a non-convex problem...

I But you can leverage familiar OP models which are empirically very good!

1. Or Chebychev coeff. = Fourier coeff of $f(\cos 2\pi\ \cdot)$