Sum of squares with Reproducing Kernel Hilbert Spaces, a path to global optimisation of regular functions

Alessandro RUDI, École Normale Supérieure (PSL) / Inria - Paris

Ulysse MARTEAU-FEREY, École Normale Supérieure (PSL) / Inria - Paris

Francis BACH, École Normale Supérieure (PSL) / Inria - Paris

We consider the global minimization of smooth functions based solely on function evaluations. Algorithms that achieve the optimal number of function evaluations for a given precision level typically rely on explicitly constructing an approximation of the function which is then minimized with algorithms that have exponential running-time complexity. In this paper, we consider an approach that jointly models the function to approximate and finds a global minimum. This is done by using infinite sums of square smooth functions and has strong links with polynomial sum-of-squares hierarchies. Leveraging recent representation properties of reproducing kernel Hilbert spaces, the infinite-dimensional optimization problem can be solved by subsampling in time polynomial in the number of function evaluations, and with theoretical guarantees on the obtained minimum.

More formally, let f be a m times differentiable function on a subset Ω of $\mathbb{R}^d$ for a given integer d. Let f_* be the minimum of f on Ω. We consider the problem of estimating f_* given n points $(x_1, ..., x_n) \in \Omega^n$ based only on the evaluations of f at these points : $(f(x_i))_{1 \leq i \leq n}$. To design our algorithm, we start by proving that $f - f_*$ can be written as a sum of squares of functions f_j which belong to a Sobolev space $H^s(\Omega)$ of functions with s square-integrable derivatives for a suitable s.

Théorème 1 (Theorem 3 in [1]). *If f has isolated strict minima inside Ω, then for any $s > m + d/2$, $f - f_*$ can be written as a finite sum of squares*

$$\forall x \in \Omega,\ f(x) = f_* + \sum_{j \in J} f_j(x)^2, \tag{1}$$

where the f_j belong to $H^s(\Omega)$, which is a reproducing kernel Hilbert space.

Our algorithm works by finding a sum of squares decomposition as in (1) but which holds only on the n points $(x_i)_{1 \leq i \leq n}$, i.e. by solving the following problem for a suitable regularizer $\mathcal{R}$:

$$\min_{c \in \mathbb{R},\ f_j \in H^s(\Omega)} -c + \mathcal{R}((f_j)_{1 \leq i \leq n}), \qquad s.t. \qquad f(x_i) = c + \sum_{j \in J} f_j(x_i)^2,\ 1 \leq i \leq n. \tag{2}$$

Using the reproducing kernel properties of $H^s(\Omega)$,(2) can be formulated as a semi-definite program and solved with a computational cost of $O(n^{3.5})$ in time and $O(n^2)$ in memory. Moreover,we prove the following theorem showing that the optimal c_* is a good approximation of f_*.

Théorème 2 (Theorem 6 in [1]). *If $x_1, ..., x_n$ are sampled uniformly, there exists a constant C depending only on d, m and Ω such that with high probability,*

$$\|f_* - c_*\| \leq C \|f\|^2_{C^m(\Omega)}\ n^{-m/d + 1/2 + 3/d}.$$

This rate is nearly optimal in the case of Sobolev functions and more generally makes the proposed method particularly suitable for functions which have a large number of derivatives. Indeed, when m is in the order of d, the convergence rate to the global optimum does not suffer from the curse of dimensionality, which affects only the worst case constant.

[1] A. Rudi, U. Marteau-Ferey, F. Bach. *Finding global minima via kernel approximations*, 2020.

Contact : `ulysse.marteau-ferey@inria.fr`