# Optimal second-order method complexity without function evaluations

Serge Gratton [1,2] Sadok Jerad [1,2] Philippe L.Toint [3]

[1]IRIT-APO

[2]Toulouse INP-ENSEEIHT

[3]University of Namur

Saint-Girons, June 20, 2022

# Problem and motivations

Approximately solving the non convex problem

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

without evaluating $f$.

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

Université Fédérale
Toulouse Midi-Pyrénées

# Problem and motivations

Approximately solving the non convex problem

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

without evaluating $f$.

Why devise an algorithm that do not evaluate the function ?

# Problem and motivations

Approximately solving the non convex problem

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

without evaluating $f$.

Why devise an algorithm that do not evaluate the function ?

- Stochastic algorithms depend heavily on accurate function values [BGMT21].
- Empirical success in modern Machine Learing (Adagrad [DHS11], Adam[KB14]).

# Adaptive Regularization Methods

For a given $\epsilon \in (0, 1]$, we try to find an approximate first-order point $x_\epsilon$

$$\|\nabla_x f(x_\epsilon)\| \leq \epsilon,$$

Adaptive regularization methods exhibits the best complexity rate at $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ for second order methods [CGT11].

# Adaptive Regularization Methods

For a given $\epsilon \in (0,1]$, we try to find an approximate first-order point $x_\epsilon$

$$\|\nabla_x f(x_\epsilon)\| \leq \epsilon,$$

Adaptive regularization methods exhibits the best complexity rate at $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ for second order methods [CGT11].

Can we devise a second order adaptive algorithm objective free function with a $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ complexity?

# OFFO-AR2

Affirmative

We devise an OFFO[1]-AR2 that uses **only** deterministic derivatives with an optimal complexity at $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ rate.

[1]Objective Free Function Optimization

# OFFO-AR2

## Affirmative

We devise an OFFO[1]-AR2 that uses **only** deterministic derivatives with an optimal complexity at $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ rate.

**Numerical Results**

- No improvement when very limited noise is present.
- Clear benefit when the computed derivatives are heavily-noised.



[1] Objective Free Function Optimization

Thank you for your attention and see you at the poster corner. 😉

# References I

Stefania Bellavia, Gianmarco Gurioli, Benedetta Morini, and Ph.L. Toint, *Adaptive regularization for nonconvex optimization using inexact function values and randomly perturbed derivatives*, Journal of Complexity **68** (2021), 101591.

Coralia Cartis, Nicholas Gould, and Philippe Toint, *Adaptive cubic regularisation methods for unconstrained optimization. part ii: Worst-case function- and derivative-evaluation complexity*, Math. Program. **130** (2011), 295–319.

John Duchi, Elad Hazan, and Yoram Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, Journal of Machine Learning Research **12** (2011), no. 61, 2121–2159.

Diederik P Kingma and Jimmy Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014).

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

Université Fédérale
Toulouse Midi-Pyrénées