The Curse of Unrolling: Rate of Differentiating Through Optimization

Damien Scieur, Quentin Bertrand,
Gauthier Gidel, Fabian Pedregosa

Estimating a Jacobian

Compute the Jacobian of an optimization problem's solution:

$$\partial \boldsymbol{x}_\star(\boldsymbol{\theta}), \text{ where } \boldsymbol{x}_\star(\boldsymbol{\theta}) = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^d} f(\boldsymbol{x}, \boldsymbol{\theta}).$$

Useful for bi-level optimization but also sensitivity analysis, explainable AI, …

This paper study how good unrolling is to estimate the Jacobian.

Unrolling to compute the Jacobian

Use backpropagation through a solver to compute the Jacobian.

$$x_{t+1}(\theta) = x_t(\theta) - \nabla f(x_t(\theta), \theta)$$

$$\partial_\theta x_{t+1}(\theta) = (Id - \mu H(\theta))\partial_\theta x_t(\theta) - \mu \partial_\theta \nabla f(x_t(\theta), \theta)$$

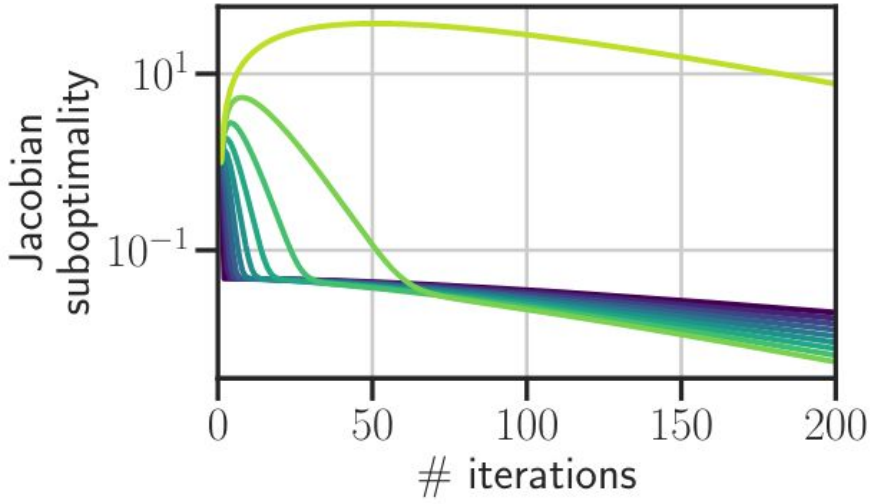
Residual Polynoms

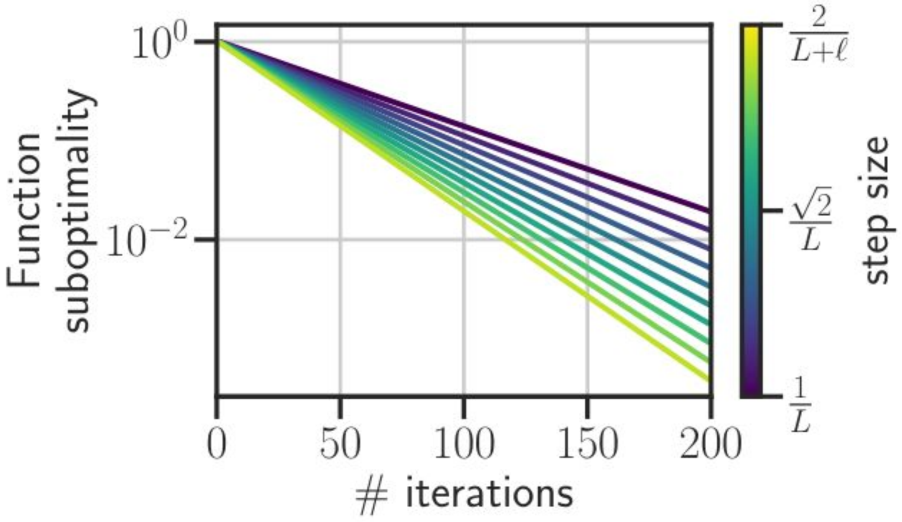
$$\boldsymbol{x}_t(\boldsymbol{\theta}) - \boldsymbol{x}_\star(\boldsymbol{\theta}) = P_t(\boldsymbol{H}(\boldsymbol{\theta}))(\boldsymbol{x}_0(\boldsymbol{\theta}) - \boldsymbol{x}_\star(\boldsymbol{\theta}))\,.$$

$$\begin{aligned}\partial\boldsymbol{x}_t(\boldsymbol{\theta}) - \partial\boldsymbol{x}_\star(\boldsymbol{\theta}) = &\big(P_t(\boldsymbol{H}(\boldsymbol{\theta})) - P_t'(\boldsymbol{H}(\boldsymbol{\theta}))\boldsymbol{H}(\boldsymbol{\theta})\big)(\partial\boldsymbol{x}_0(\boldsymbol{\theta}) - \partial\boldsymbol{x}_\star(\boldsymbol{\theta}))\\ &+ P_t'(\boldsymbol{H}(\boldsymbol{\theta}))\partial_{\boldsymbol{\theta}}\nabla f(\boldsymbol{x}_0(\boldsymbol{\theta}), \boldsymbol{\theta})\,.\end{aligned}$$

Main result

The best algorithm to estimate the Jacobian is not always the best to solve the original problem.





Other results

They characterize a burning-phase when the step-size is too large.

They propose a “best” algorithm based on Sobolev polynomials.

They evaluate this on synthetic and real data.

Limitations

This work mainly work for quadratic functions.

Hard hypothesis on commutativity of the Jacobian and Hessian.

No comparison with the implicit gradient.