Implicit differentiation for fast hyperparameter selection in non-smooth convex learning

Quentin Bertrand
Mila, Université de Montral
quentin.bertrand@mila.quebec

Finding the optimal hyperparameters of a model can be cast as a bilevel optimization problem, typically solved using zero-order techniques such as grid search or random search. In this work we study first-order methods when the inner optimization problem is convex but non-smooth. We show that the forward-mode differentiation of proximal gradient descent and proximal coordinate descent yield sequences of Jacobians converging toward the exact Jacobian. Using implicit differentiation, we show it is possible to leverage the non-smoothness of the inner problem to speed up the computation. Finally, we provide a bound on the error made on the hypergradient when the inner optimization problem is solved approximately. Results on regression and classification problems reveal computational benefits for hyperparameter optimization, especially when multiple hyperparameters are required. This work is illustrated in Figure 1 and is based on the following publications [1, 2].

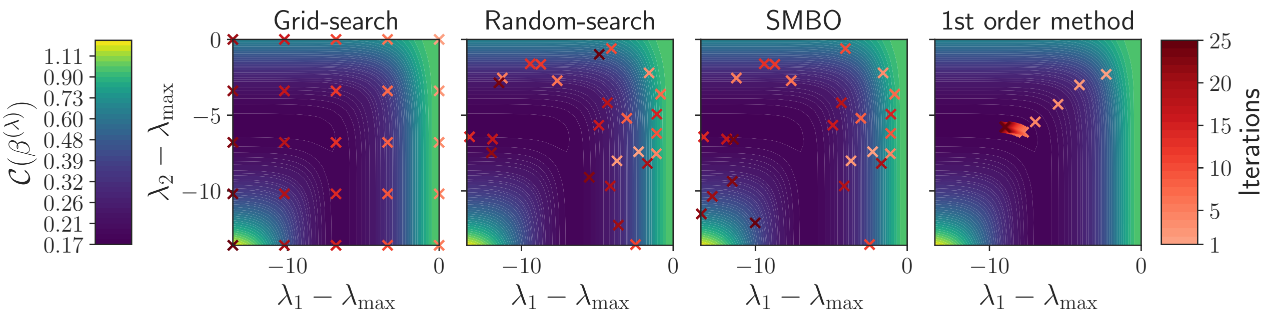

Figure 1: **5-fold cross-validation error** $\mathcal{C}(\beta^{(\lambda)})$: elastic net CV error with respect to λ_1 and λ_2 for multiple hyperparameter optimization methods on the *rcv1* dataset. Crosses represent the 25 first error evaluations for each method.

References

[1] Q. Bertrand, Q. Klopfenstein, S. Vaiter, A. Gramfort, and J. Salmon. Implicit differentiation for Lasso-type problems *ICML 2020*.

[2] Q. Bertrand, Q. Klopfenstein, M. Massias, S. Vaiter, A. Gramfort, and J. Salmon. Implicit differentiation for fast hyperparameter selection in non-smooth convex learning *JMLR 2022*.