ACHIEVING LINEAR OR QUADRATIC CONVERGENCE ON PIECEWISE SMOOTH OPTIMIZATION PROBLEMS

ANDREAS GRIEWANK[1] AND ANDREA WALTHER[2]

Abstract. Many problems in machine learning involve objective functions that are piecewise smooth [7] due to the occurrence of absolute values mins and maxes in their evaluation procedures. See e.g. [8]. For such function we derived in [3] first order (KKT) and second order (SSC) optimality conditions, which can be checked on the basis of a local piecewise linearization [2] that can be computed in an AD like fashion, e.g. using ADOL-C or Tapenade.

In that analysis, a key assumption on the local piecewise linearization was the Linear Independence Kink Qualification (LIKQ), a generalization of the Linear Independence Constraint Qualification (LICQ) known from smooth Nonlinear Optimization. A rather surprising consequence is that checking the optimality conditions is not at all combinatorial but can be done with a cubic effort like in the classical smooth case. Moreover, as we show here first under LIKQ with SSC the natural algorithm of successive piecewise linear optimization with a proximal term (SPLOP) achieves a linear rate of convergence. A version of SPLOP has already been implemented and tested in [4, 1].

Secondly, we observe that, even without any kink qualifications, local optimality of the nonlinear objective always requires local optimality of its piecewise linearization, and strict minimality of the latter is in fact equivalent to sharp minimality of the former. Moreover, we show that SPLOP will converge quadratically to such sharp minimizers, where the function exhibits linear growth. These results are independent of the particular function representation, and allow in particular duplications of switching variables and other intermediates.

We note that the classical theory for subgradient [9] , proximal [6] and bundle [5] methods usually only yields convergence rates like $1/sqrt(k)$ or $\log(k)/k$, where k is the iteration counter. Only for strongly convex functions a linear convergence rate can sometimes be established. Our assumptions LIKQ and SSC are certainly quite strong, but they do not require convexity, even locally, near a minimizer. In case of the Lasso problem $\min ||x||_1 + \rho||Ax - b||$ our method coincides with ISTA as described in [6].

Our current implementation of SPLOPT allows the verification of the theoretical results mentioned above on the usual set of academic test problems. The number of outer iterations is usually extremely low compared to more established approaches. However, the setting up and solving the local piecewise linear problem is not yet adapted to large structured problems.

In effect we have to solve a sequence of closely related, convex Quadratic Optimization Problems (QOP), while marching through a polyhedral decomposition of the variable domain. For several aspects, like the selection of the next polyhedron, the handling of the many locally redundant constraints, and the exploitation of sparsity there are obvious improvements, which we are currently exploring and implementing. We expect to present results at least on the Lasso problem [6] and fuzzy pattern trees as described in [8].

[1]School of Mathematical Sciences and Information Technology, Yachaytech, Urcuquí, Imbabura, Ecuador

[2]Institut für Mathematik, Universität Paderborn, Paderborn, Germany

REFERENCES

[1] S. Fiege, A. Walther, and A. Griewank. An algorithm for nonsmooth optimization by successive piecewise linearization. Technical report, Universität Paderborn, 2016. available at optimization-online.

[2] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013.

[3] A. Griewank and A. Walther. First and second order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5):904–930, 2016.

[4] A. Griewank, A. Walther, S. Fiege, and T. Bosse. On Lipschitz optimization based on gray-box piecewise linearization. *Mathematical Programming Series A*, 158(1-2):383–415, 2016.

[5] N. Karmitsa and M. Mäkelä. Limited memory bundle method for large bound constrained nonsmooth optimization: convergence analysis. *Optimization Methods and Software*, 25(6):895–916, 2010.

[6] Juan Peypouquet. *Convex Optimization in Normed Spaces*. Number ISBN 978-3-319-13710-0. Springer, 2015.

[7] S. Scholtes. *Introduction to Piecewise Differentiable Functions*. Springer, 2012.

[8] R. Senge and E. Hüllermeier. Top-down induction of fuzzy pattern trees. *IEEE Transactions on Fuzzy Systems,*, 19(2):241–252, 2011.

[9] N.Z. Shor. *Nondifferentiable optimization and polynomial problems*. Kluwer, 1998.