Computational Optimization and Applications (2022) 83:723–726
https://doi.org/10.1007/s10589-022-00426-3

COAP 2021 Best Paper Prize

Each year, the editorial board of Computational Optimization and Applications selects a paper from the preceding year's publications for the Best Paper Award. In 2021, 88 papers were published in the journal. This article highlights the research related to the award winning work of Christian Kanzow and Theresa Lechner (University of Würzburg) in their paper "Globalized inexact proximal Newton-type methods for nonconvex composite functions," published in volume 78, pages 377–410.

The paper [6] considers the composite optimization problem

$$\min f(x) + \varphi(x), \quad x \in \mathbb{R}^n$$

with $f : \mathbb{R}^n \to \mathbb{R}$ being (twice) continuously differentiable (not necessarily convex) and $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ convex (possibly nonsmooth and extended-valued.) Proximal methods compute a sequence $\{x^k\}$ such that x^{k+1} is a solution or a stationary point (at least inexactly) of the subproblem

$$\min f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T H_k(x - x^k) + \varphi(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where only f is linearized around the current iterate x^k, and H_k is a suitable approximation of the (not necessarily existing) Hessian $\nabla^2 f(x^k)$. The classical proximal gradient method corresponds to the choice $H_k = \gamma_k I$ for all k with some (penalty or line search parameter) $\gamma_k > 0$, whereas $H_k = \nabla^2 f(x^k)$ leads to the proximal Newton method, and $H_k \approx \nabla^2 f(x^k)$ with a (limited memory) quasi-Newton approximation of the Hessian is naturally called a (limited memory) proximal quasi-Newton method.

In principle, the proximal gradient method has the major advantage that the subproblems (1) can be solved very efficiently (even analytically) for some practically important proximal functions φ. In general, this is not true for the proximal Newton subproblem, so that iterative methods have to be used in order to solve the subproblems themselves, which causes more work in each step than for the proximal gradient method. On the other hand, the proximal Newton method typically requires less many (outer) iterations. Moreover, in those situations where the function φ belongs to a class where, in any case, the solution of the proximal gradient subproblem cannot be computed analytically or efficiently, the overhead for computing an inexact solution of the proximal Newton subproblems seems to be minor.

For a convergence theory of the proximal gradient method, we refer to the excellent book [1] by Beck for more details, where f is assumed to satisfy the standard assumptions for proving nice global convergence results, in particular, f is supposed to be convex with ∇f satisfying a global Lipschitz condition. (The global Lipschitz assumption on the gradient is rather annoying and there is currently quite some research in order to overcome this problem.) Since the proximal gradient method reduces to the standard steepest descent method for the particular case where $\varphi \equiv 0$, the rate-of-convergence is typically slow (sublinear). On the other hand, proximal Newton and quasi-Newton methods are much faster (locally) convergent, see, e.g., [9, 13, 15] and references therein. Typically, the underlying approaches either assume that f is (strongly) convex or that a positive definite Hessian approximation is used, so that the resulting subproblems (1) still have a (unique) solution.

The motivation for the work in [6] is to deal with functions f which are, in general, neither convex nor do their gradients satisfy a global Lipschitz condition. The interest in this situation partially results from the fact that these proximal-type methods are used to solve certain subproblems in an augmented Lagrangian setting for some classes of highly difficult optimization problems, and where these subproblems are typically nonconvex, and the corresponding objective functions (the augmented Lagrangians) do not satisfy a (global) Lipschitz condition, see [4, 5].

The method presented in [6] combines the (inexact) proximal Newton step with a proximal gradient step in a suitable way such that it inherits automatically the nice global convergence properties of the former (without using a Lipschitz assumption on the gradient) and the strong local convergence of the latter one. Note, however, that the proximal Newton subproblems do not need to have solutions (or inexact solutions), so some care has to be taken by introducing a suitable and computable criterion which decides when the method has to switch from a proximal Newton to a proximal gradient step. Numerical results indicate, however, that the method accepts the (inexact) proximal Newton step quite frequently, which is an important observation since otherwise the overall approach would be rather inefficient.

There exist already some subsequent works which may be viewed as modifications or improvements of the paper [6]. First of all, the PhD thesis [8] contains improved convergence results based on the theory of Kurdyka-Łojasiewicz functions. The report [7] presents a regularized quasi-Newton method for composite optimization which incorporates second-order information in a very efficient way into the proximal term by using ideas from [2] and the compact representation of limited memory quasi-Newton matrices [3]. The basic idea behind this approach is the fact that solutions of a limited memory proximal quasi-Newton subproblem can be obtained from the solution of the corresponding proximal gradient subproblem using only some algebraic manipulations together with the solution of a small-dimensional and strongly monotone nonlinear (though nonsmooth) system of equations.

A trust-region modification, based on a nonsmooth reformulaton of the corresponding stationarity conditions, is presented in [12] (note that, in principle, a trust-region strategy avoids the problem that the corresponding subproblems may not have a solution). The paper [10] uses a regularization strategy in order to have strongly convex and hence solvable subproblems (for nonconvex f, for convex functions f, the situation is easier, see [11] for a recent contribution). Another modification, where the Newton direction is computed only on a suitable subspace (and, hence, easier to obtain) is considered in [16] for the particular case of φ being the ℓ_q-quasi norm with $q \in (0, 1)$. Finally, an extension to Hilbert space can be found in [14].

References

1. Beck, A.: First-order methods in optimization. Society for Industrial and Applied Mathematics, MOS-SIAM Series on Optimization (2017)
2. Becker, S., Fadili, J., Ochs, P.: On quasi-Newton forward-backward splitting: proximal calculus and convergence. SIAM J. Optim. **29**(4), 2445–2481 (2019)
3. Byrd, R.H., Nocedal, J., Schnabel, R.B.: Representations of quasi-Newton matrices and their use in limited memory methods. Math. Program. **63**(1), 129–156 (1994)
4. De Marchi, A., Jia, X., Kanzow, C., Mehlitz, P.: Constrained composite optimization and augmented Lagrangian methods (2022). https://doi.org/10.48550/arXiv.2203.05276
5. Jia, X., Kanzow, C., Mehlitz, P., Wachsmuth, G.: An augmented Lagrangian method for optimization problems with structured geometric constraints. Math. Program. (2022). https://doi.org/10.1007/s10107-022-01870-z
6. Kanzow, C., Lechner, T.: Globalized inexact proximal Newton-type methods for nonconvex composite functions. Comput. Optim. Appl. **78**(2), 377–410 (2021)
7. Kanzow, C., Lechner, T.: Efficient regularized proximal quasi-Newton methods for large-scale nonconvex composite optimization problems (2022). https://doi.org/10.48550/arXiv.2210.07644
8. Lechner, T.: Proximal methods for nonconvex composite optimization problems. PhD thesis, University of Würzburg, (2022)
9. Lee, J.D., Sun, Y., Saunders, M.A.: Proximal Newton-type methods for minimizing composite functions. SIAM J. Optim. **24**(3), 1420–1443 (2014)
10. Liu, R., Pan, S., Wu, Y., Yang, X.: An inexact regularized proximal Newton method for nonconvex and nonsmooth optimization (2022). https://doi.org/10.48550/arXiv.2209.09119
11. Mordukhovich, B.S., Yuan, X., Zeng, S., Zhang, J.: A globally convergent proximal Newton-type method in nonsmooth convex optimization. Math. Program. (2022). https://doi.org/10.1007/s10107-022-01797-5
12. Ouyang, W., Milzarek, A.: A trust region-type normal map-based semismooth Newton method for nonsmooth nonconvex composite optimization (2021). https://doi.org/10.48550/arXiv.2106.09340
13. Patrinos, P., Bemporad, A.: Proximal Newton methods for convex composite optimization. In 52nd IEEE Conference on Decision and Control, pp 2358–2363. IEEE (2013)
14. Pötzl, B., Schiela, A., Jaap, P.: Inexact proximal Newton methods in Hilbert spaces(2022). https://doi.org/10.48550/arXiv.2204.12168
15. Scheinberg, K., Tang, X.: Practical inexact proximal quasi-Newton method with global complexity analysis. Math. Program. **160**(1–2), 495–529 (2016)
16. Yuqia, W.U., Shaohua, P., Yang, X.: A globally convergent regularized Newton method for ℓ_q -norm composite optimization problems (2022). arxiv:2203.02957

Christian Kanzow studied at the Universities of Kiel and Hamburg, Germany, and received his diploma and PhD in 1992 and 1995, respectively. He had some longer research stays at the University of Wisconsin in Madison 1997/98 and the University "La Sapienza" in Rome 1998/99, an interim professorship at the University of Trier in 1999/2000, and became full Professor at the University of Würzburg in 2001. His main research interests are in the development of algorithms for several difficult classes of problems in continuous optimization and related areas.

Theresa Lechner studied at the Universities of Würzburg and Kraków. She received her PhD from the University of Würzburg in 2022.