Fallacies about Newton Method

Konstantin Burlachenko (burlachenkok@gmail.com, bruziuz@stanford.edu), 07JUNE 2018

1. Method always needs big amount of time per iteration

http://cs229.stanford.edu/notes/cs229-notes1.pdf
"Newton's method typically enjoys faster convergence than (batch) gradient descent, and requires many fewer iterations to get very close to the minimum. One iteration of Newton's can, however, be more expensive than one iteration of gradient descent, since it requires finding and inverting an n-by-n Hessian."
Of course it's true but sometimes we can exploit structure in this case .

2. Subgradient sound fancy so it's better then old Newton Method

S.Boyd asked in video EE364B https://youtu.be/B51GgGCHBRk?t=1981 that if it is possible for the problem then avoid subgradients methods. Theory is beautifull. There is interconnection between subgradients and directional derivative, both of which are well-defined for non-differentiable, but convex functions. But method are very slow. Even various modification like heavy-ball or momentum method proposed by B.Polyak, et al. is fundamentally don't change situation. Also I have found the information that Newton method is fast way to sovle convex optimization problem here

http://stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf, p.17

"However, the slow convergence of ADMM also distinguishes it from algorithms such as Newton's method (or, for constrained problems, interior-point methods), where high accuracy can be attained in a reasonable amount of time"

3. After select direction method exact line search can give good next point.

Cvxbook, p.492: "An exact line search usually gives a very small improvement of convergence of Newton's method."

4. Issue is compute time for Newton Method.

Usually no. If change level of linear algebra from usual to advanced. But it need carefully exploit structure of Hessian.

1. Newton method is invariant of the affine change of coordinates[p.495]. While choice of coordinates (or condition number of sublevel sets) is a first-order issue for gradient and steepest descent methods.
2. Newton's method scales well with problem size.

The main disadvantage of Newton's method is the cost of forming and storing the Hessian, and the cost of computing the Newton step, which requires solving a set of linear equations. Classical analysis of Newton method which has been done by Kantarovich (cvxbook, p.504,eq.9.53) and modern analysis for self-concordant functions done by Nesterov and Nemervoskiy (cvxbook, p.505,eq.9.57) shows that upper bound of number of iterations depend lineary of initial gap.