On finite termination of quasi-Newton methods on quadratic problems

Aban Ansari-Önnestam, Linköping University
Dept. of Mathematics, Applied Mathematics division
Supervisors: Elina Rönnberg (LiU), Anders Forsgren (KTH)

LINKÖPINGS UNIVERSITET

Motivation & Research Goals

Quasi-Newton methods form an important class of methods for solving nonlinear optimization problems. In the best case, quasi-Newton methods will far outperform steepest descent and other first order methods, without the computational cost of calculating the exact second derivative. These convergence guarantees hold locally, which follows closely from the fact that, if the objective function is strongly convex, it can be approximated well by a quadratic function close to the solution. Understanding the performance of quasi-Newton methods on quadratic problems with a symmetric positive definite Hessian is therefore of vital importance. We show that it suffices to create a memoryless quasi-Newton matrix based on two vectors to give ability to compute a Newton direction within a finite number of iterations, independent of step lengths.

A quasi-Newton method for solving QP

We focus on a quasi-Newton method on the unconstrained quadratic problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T H x + c^T x \tag{QP}$$

for a symmetric positive definite matrix $H \in \mathbb{R}^{n \times n}$.
At a point x_k, using $P_{k-1} = \begin{bmatrix} p_{k-1}^N & q_{k-1} \end{bmatrix}$ with columns of a subspace Newton step and the most recent conjugate vector we may build the Hessian approximation

$$B_k = \sigma_k (I - P_{k-1}(P_{k-1}^T P_{k-1})^{-1} P_{k-1}^T) + H P_{k-1} (P_{k-1}^T H P_{k-1})^{-1} P_{k-1}^T H.$$

Let x_0 be a fixed initial point and let $x_i, i = 0, \dots k$ be generated by Algorithm 1 where the step sizes α_i are arbitrarily chosen for $i < k$ and $\sigma_i > 0$ for $i \leq k$.

When $k \leq r$, the search direction p_k generated by the algorithm increases the dimension of the search space by exactly one conjugate basis vector.

Additionally, for $k \geq r$, the search direction p_k generated by Algorithm 1 is precisely the Newton step so the algorithm will terminate if $\alpha_k = 1$.

Using exact line search means the best possible step is taken at each iteration and the algorithm always leads to descent. However any step size, including a zero step, is sufficient to gather curvature information on quadratic problems.

For a unique value of $\hat{\sigma}_{r-1}$, Algorithm 1 will generate the Newton step $p_{r-1} = p_{r-1}^N(x_{r-1})$, and the algorithm will terminate at iteration r if $\alpha_{r-1} = 1$.

For all other value of σ_{r-1} the algorithm will terminate only when $\alpha_k = 1$ for some $k \geq r$.

An exact algorithm without line search

Algorithm 1 An exact quasi-Newton method for solving QP

Require: $x_0 \in \mathbb{R}^N, p_{-1}^N = 0, \sigma_0 > 0$
  $k \leftarrow 0$;
  $g_k \leftarrow H x_k + c$;
  while $\|g_k\| \neq 0$ **do**
    if $k = 0$ **then**
      $B_k = \sigma_k I$;
    else
      Choose some $\sigma_k > 0$;
      Update B_k using P_{k-1};
    end if
    Solve $B_k p_k = -g_k$;
    Choose step size α_k;
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
    $g_{k+1} \leftarrow g_k + \alpha_k H p_k$;
    if $\|g_{k+1}\| = 0$ **then**
      break
    end if
    $q_k \leftarrow p_k - p_{k-1}^N$;
    if $\|q_k\| \neq 0$ **then**
      $p_k^N \leftarrow \left(-\frac{g_k^T q_k}{q_k^T H q_k} - 1 \right) q_k + (1 - \alpha_k) p_k$;
      if rank($\begin{bmatrix} p_k^N & q_k \end{bmatrix}$) $= 2$ **then**
        $P_k = \begin{bmatrix} p_k^N & q_k \end{bmatrix}$;
      else
        $P_k = \begin{bmatrix} q_k \end{bmatrix}$;
      end if
    else
      $p_k^N = (1 - \alpha_k) p_{k-1}^N$;
      $P_k = \begin{bmatrix} p_k^N \end{bmatrix}$;
    end if
    $k \leftarrow k + 1$;
  end while

Future work

- Adapting the exact quasi-Newton algorithm to general nonlinear functions.
- Analyzing the differences between performance on quadratic and almost quadratic functions.
- Improving performance on limited memory methods to achieve the same convergence rates as full memory methods.

References

[1] On finite termination of quasi-Newton methods on quadratic problems
Aban Ansari-Önnestam and Anders Forsgren
https://arxiv.org/abs/2407.03072