Newton's method

Exact relation:

$$f(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(\xi)(x - x_k),$$

where $\xi = \alpha x_k + (1 - \alpha)x$ for some $\alpha \in [0, 1]$.

Quadratic model for Newton:

$$\phi_k(x) = f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

- Recall: The solution to $Hs_* = -g$ is the unique minimizer of $\phi(s) = \alpha + g^\mathsf{T} s + \frac{1}{2} s^\mathsf{T} H s$, for H symmetric, positive definite.
- Thus, if $\nabla^2 f(x_k)$ is positive definite, ϕ_k has a unique minimum for

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$$

Algorithm (Basic Newton)

1. Choose initial guess x_0 & tolerance ϵ
2. For $k = 0, 1, \ldots$
   - 2.1 If $\|\nabla f(x_k)\| \le \epsilon$ **stop**
   - 2.2 Solve $\nabla^2 f(x_k) s_k = -\nabla f(x_k)$
   - 2.3 Set $x_{k+1} = x_k + s_k$

Newton's method converges quadratically to a local minimum x_*

- if the Hessian is positive definite at x_*, and
- if started close enough to the solution

Problems with basic Newton

(i) Needs derivatives: may be expensive or impossible to compute

(ii) Needs solution of a linear system each iteration. Demanding for large problems

(iii) Converges only when starting close enough

(iv) The Hessian may be singular. Yields problems to generate the step (to solve the linear system)

Guaranteeing descent (problem (iv))

- p is a **descent direction** for f at x_k if $p^\mathsf{T} \nabla f(x_k) < 0$
- If p is a descent direction for f at x_k, then $f(x_k + \alpha p) < f(x_k)$ for a sufficiently small $\alpha > 0$
- The Newton direction $s_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$ is a descent direction **if the Hessian $\nabla^2 f(x_k)$ is positive definite**

The Cholesky Decomposition

A version of the LU-factorization for symmetric positive definite matrices, where $A = LL^\mathsf{T}$, with L lower triangular. Partition:

$$A = \begin{pmatrix} \alpha & a^\mathsf{T} \\ a & \hat{A} \end{pmatrix}$$

where $\alpha > 0$ (since A is PD), a is n-by-1, $\hat{A}$ is $n-1$-by-$n-1$

Steps of algorithm:

- $L_1 = \begin{pmatrix} \sqrt{\alpha} & 0^\mathsf{T} \\ \frac{a}{\sqrt{\alpha}} & \hat{A}_1 \end{pmatrix}$, where $\hat{A}_1 = \hat{A} - \dfrac{aa^\mathsf{T}}{\alpha}$

  (Note that $L_1\hat{L}_1^\mathsf{T} = \begin{pmatrix} \sqrt{\alpha} & 0^\mathsf{T} \\ \frac{a}{\sqrt{\alpha}} & \hat{A} - \frac{aa^\mathsf{T}}{\alpha} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & \frac{a^\mathsf{T}}{\sqrt{\alpha}} \\ 0 & \hat{I} \end{pmatrix} = \begin{pmatrix} \alpha & a^\mathsf{T} \\ a & \hat{A} \end{pmatrix} = A$, so first row and column done.)
- $\hat{A}$ is symmetric $n-1$-by-$n-1$. Can show: $\hat{A}_1$ is positive definite if and only if A is positive definite
- Repeat the above recursively for $\hat{A}$

Review: Newton's method/Cholesky

Modified Cholesky

- Cholesky needs no pivoting
- Half the computational effort and half the storage compared to standard LU
- The α occurring in current submatrix is ≤ 0 only if the matrix is not positive definite
- If $\alpha \leq 0$ happens, replace with a small positive number
- Can show: equivalent to compute $LL^\mathsf{T} = E + A$ with $E \geq 0$ diagonal such that $E + A$ is positive definite