Advanced Mathematical Programming IE417

Lecture 17

Dr. Ted Ralphs

Reading for This Lecture

- Sections 8.6-8.8

Conjugate Directions

- If $H \in \mathbb{R}^{n\times n}$ is symmetric, the linearly independent vectors $d_1, \ldots, d_n$ are called *H-conjugate* if $d_i^T H d_j = 0$ for $i \neq j$.
- Minimizing the quadratic function $f(x) = c^T x + x^T H x$.
  - Given x_1, any $x \in \mathbb{R}^n$ can be represented as $x_1 + \sum \lambda_j d_j$.
  - $f(x)$ can be rewritten as a function F of λ.

$$F(\lambda) = c^T x_1 + \sum \lambda_j c^T d_j + (x_1 + \sum \lambda_j d_j)^T H (x_1 + \sum \lambda_j d_j)$$
$$= \sum [c^T(x_1 + \lambda_j d_j) + (x_1 + \lambda_j d_j)^T H (x_1 + \lambda_j d_j)]$$

Comments on Conjugate Directions

- The function F is separable so we can minimize over each direction sequentially using line search.
- Hence, we can minimize any quadratic function in n steps.
- At the k^{th} step, we end up at the minimum of f over the subspace spanned by $d_1, \ldots, d_k$.
- Also, $\nabla f(x_k)^T d_j = 0$ for $j = 1, \ldots, k-1$.

Quasi-Newton Methods

Davidon-Fletcher-Powell

- Idea 1: Use a search direction $d_j = -D_j \nabla f(x)$ where D_j is symmetric positive definite and approximates H^{-1}.
- Idea 2: Update D_j at each step so that d_{j+1} is a conjugate direction.
- DFP Update
  - $D_{j+1} = D_j + p_j p_j^T / p_j^T q_j - D_j q_j q_j^T D_j / q_j^T D_j q_j$
  - $p_j = \lambda_j d_j = x_{k+1} - x_k$
  - $q_j = \nabla f(x_k) - \nabla f(x_{k+1})$

Quasi-Newton Methods

- Typical Quasi-Newton Algorithm
  - Start with D_1 symmetric p.d., and initial point $y_1 = x_1, k = 1$.
  - For $j = 1 \; to \; n$
    * If $\|\nabla f(y_j)\| < \varepsilon$, then STOP.
    * Otherwise, perform a line search in direction $d_j = -D_j \nabla f(y_j)$ to find y_{j+1}.
    * Update D_j to D_{j+1}.
  - Set $x_{k+1} = y_n, k = k + 1$.
- If the inner loop exits after only $n' < n$ iterations, we have a *partial quasi-Newton method*.

Comments on the DFP update

- As long as D_1 is symmetric p.d. each D_j is also symmetric p.d.
- This means that each search direction is a descent direction.
- For quadratic functions, the search directions are conjugate to the Hessian.
- Furthermore, for quadratic functions, $D_{n+1} = H^{-1}$.

Basis of Quasi-Newton Methods

- Note that for quadratic functions, the vectors $p_1, \ldots, p_{j-1}$ in the DFP procedure are linearly independent eigenvectors of $D_{j+1}H$ with unit eigenvalues.
- Hence, we are essentially building up H^{-1} as the sum of rank 2 matrices.
- Requirements for quasi-Newton update
  - Maintain symmetry and positive definiteness
  - Maintain the quasi-Newton condition for the correction C_j

$$C_j q_j = p_j - D_j q_j$$

Other Updates

- Broyden Update
  - A parameterized family of quasi-Newton updates
- Broyden-Fletcher-Goldfarb-Shanno
  - One of the Broyden updates, discovered independently
  - Dominant updating scheme computationally
- Updating the Hessian directly
  - Note that the previous updates worked on H^{-1}.
  - It is also possible to approximate H with B_j.
  - In this case, maintain a Cholesky factorization of B_j.

Computational Issues

- Obviously, there are many computational issues to be addressed with these methods.
- We must take care to avoid ill-conditioned, nearly-singular matrices.
- Cholesky factorization is the central tool for solving systems $Bd = -\nabla f(x)$.
- Implementing these techniques efficiently is not easy.