Conjugate Direction Methods

- First, consider a quadratic problem:

$$\min f(x) = \tfrac{1}{2} x'Qx - b'x$$

where Q is positive definite

- Is it convex?
- What is the equivalent linear program?

How to solve it?

- Start gently
  - Consider the simplest case
  - When $Q = I \cancel{= 0}$
- Orthogonality of searching directions a must?
- Can we extend it to general Q?

Conjugacy

- Given a positive definite matrix Q, a set of nonzero vectors $d^1, \ldots, d^k$ are *Q-conjugate* if

$$d^{i\prime} Q d^j = 0, \quad \text{for all } i \text{ and } j \text{ such that } i \neq j$$

- If $d^1, \ldots, d^k$ are Q-conjugate, they are linearly independent

Conjugate Direction Method for Quadratic Optimization

1. Start with arbitrary x^0
2. Get $d^0, \ldots, d^{n-1}$ which are *Q-conjugate*
3. Repeat

$$x^{k+1} = x^k + \alpha^k d^k$$

$$\text{Where } \alpha^k = argmin_{\alpha} f(x^k + \alpha\, d^k)$$

for $k = 0, \ldots, n-1$

- Intuition: dimensional decomposition
  - "orthogonal" w.r.t. Q

GENERATING Q-CONJUGATE DIRECTIONS

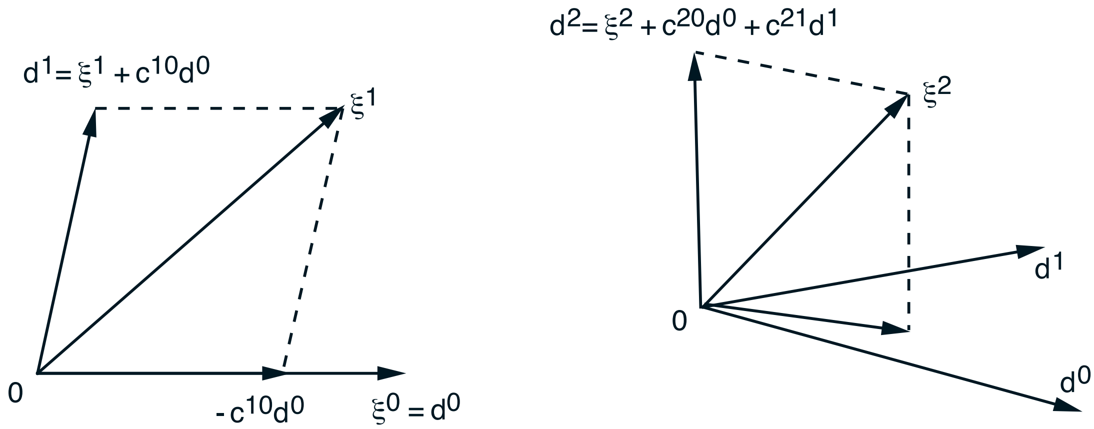
- Given set of linearly independent vectors $\xi^0, \ldots, \xi^k$, we can construct a set of Q-conjugate directions $d^0, \ldots, d^k$ s.t. $Span(d^0, \ldots, d^i) = Span(\xi^0, \ldots, \xi^i)$

- $Gram\text{-}Schmidt\ procedure$. Start with $d^0 = \xi^0$. If for some $i < k$, $d^0, \ldots, d^i$ are Q-conjugate and the above property holds, take

$$d^{i+1} = \xi^{i+1} + \sum_{m=0}^{i} c^{(i+1)m} d^m;$$

choose $c^{(i+1)m}$ so d^{i+1} is Q-conjugate to $d^0, \ldots, d^i$,

$$d^{i+1'} Q d^j = \xi^{i+1'} Q d^j + \left(\sum_{m=0}^{i} c^{(i+1)m} d^m \right)' Q d^j = 0.$$



CONJUGATE GRADIENT METHOD

- Apply Gram-Schmidt to the vectors $\xi^k = -g^k = -\nabla f(x^k)$, $k = 0, 1, \ldots, n-1$. Then

$$d^k = -g^k + \sum_{j=0}^{k-1} \frac{g^{k\prime} Q d^j}{d^{j\prime} Q d^j} d^j$$

- **Key fact:** Direction formula can be simplified.

Proposition : The directions of the CGM are generated by $d^0 = -g^0$, and

$$d^k = -g^k + \beta^k d^{k-1}, \qquad k = 1, \ldots, n-1,$$

where β^k is given by

$$\beta^k = \frac{g^{k\prime} g^k}{g^{k-1\prime} g^{k-1}} \quad \text{or} \quad \beta^k = \frac{(g^k - g^{k-1})' g^k}{g^{k-1\prime} g^{k-1}}$$

Furthermore, the method terminates with an optimal solution after at most n steps.

- Extension to nonquadratic problems.

Conjugate Gradient Methods applied to Nonquadratic Problems

- min f(x), where f(x) is a general function
  - use the same algorithm
- $x^{k+1} = x^k + \alpha^k d^k$

  Where $\alpha^k = argmin_{\alpha} f(x^k + \alpha\, d^k)$

  $d^k = - g^k + \beta^k d^{k-1}$
- Approximation: *d0, d1,* ..., gradually lose conjugacy. Remedies:
  - Restart every *n* iterations (with a steepest descent)
  - Restart when conjugacy is lost
- Line search may be expensive

Quasi-Newton Method

- Newton's method:
  - $d^k = -(H^k)^{-1} g^k$
  - But $(H^k)^{-1}$ is expensive to evaluate
- Can we approximate $(H^k)^{-1}$ given
  - $x^0, x^1, \ldots, x^{k-1}$
  - $f^0, f^1, \ldots, f^{k-1}$
  - $g^0, g^1, \ldots, g^{k-1}$

QUASI-NEWTON METHODS

- $x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$, where D^k is an inverse Hessian approximation.
- Key idea: Successive iterates x^k, x^{k+1} and gradients $\nabla f(x^k)$, $\nabla f(x^{k+1})$, yield curvature info

$$q^k \approx \nabla^2 f(x^{k+1}) p^k,$$

$$p^k = x^{k+1} - x^k, \quad q^k = \nabla f(x^{k+1}) - \nabla f(x^k),$$

$$\nabla^2 f(x^n) \approx \begin{bmatrix} q^0 & \cdots & q^{n-1} \end{bmatrix} \begin{bmatrix} p^0 & \cdots & p^{n-1} \end{bmatrix}^{-1}$$

- Most popular Quasi-Newton method is a clever way to implement this idea

$$D^{k+1} = D^k + \frac{p^k p^{k\prime}}{p^{k\prime} q^k} - \frac{D^k q^k q^{k\prime} D^k}{q^{k\prime} D^k q^k} + \xi^k \tau^k v^k v^{k\prime},$$

$$v^k = \frac{p^k}{p^{k\prime} q^k} - \frac{D^k q^k}{\tau^k}, \quad \tau^k = q^{k\prime} D^k q^k, \quad 0 \le \xi^k \le 1$$

and $D^0 > 0$ is arbitrary, α^k by line minimization, and $D^n = Q^{-1}$ for a quadratic.

Summary of Gradient Methods

- Steepest Descent
  - 1 iteration for linear problems
- Newton's Method
  - 1 iteration for quadratic problems
- Conjugate gradient and quasi-Newton
  - n iterations for quadratic problems
- All solve non-quadratic problems heuristically
  - needs infinite number of iterations to converge

More comparisons

- **Advantages of Quasi-Newton over conjugate gradient for nonquadratic problems:**
  - Better local convergence since QN approximates Newton
  - Does not needs to periodically restart
    - CGM loses conjugacy, while QN improves approximation
  - Experimentally, QN is less sensitive to linear search quality
- **Complexity per iteration:**
  - Newton: $O(n^3)$, f, gradient, Hessian, to generate H^{k-1}
  - QN: $O(n^2)$, f, gradient, to generate D^k
  - CGM: $O(n)$, f, gradient, to generate d^k
  - CGM is preferable when n is large and computing gradient is efficient