Ch.10 Conjugate Direction Methods

10.1 Introduction

- Intermediate bw. the method of steepest descent and Newton's method.
- This performs better than the method of steepest descent, but not as well as Newton's method.
- For a quadratic function of n variables $f(x) = \frac{1}{2}x^TQx - x^Tb,\ x \in \Re^n,\ Q = Q^T > 0$, the best direction of search is in the Q-conjugate direction.
- **Def. 10.1** let Q be a real symmetric $n \times n$ matrix. The directions $d^{(0)},\ d^{(1)},\ d^{(2)},\ \cdots d^{(m)}$ are Q-conjugate if, for all $i \neq j$, we have $d^{(i)}Qd^{(j)} = 0$.
- **Lemma 10.1** Let Q be a symmetric positive definite $n \times n$ matrix. If the directions $d^{(0)}, d^{(1)}, \cdots, d^{(k)} \in \Re^n, k \leq n-1$, are nonzero and Q-conjugate, then they are linearly independent.
- Ex. 10.1 Consider

$$Q = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$Q = Q^T > 0$. Q is positive definite because all its leading principle minors are positive.

Our goal is to construct a set of Q-conjugate vectors $d^{(0)}, d^{(1)}, d^{(2)}$. Let $d^{(0)} = [1, 0, 0]^T$,$d^{(1)} =$ $[d_1^{(1)}, d_2^{(1)}, d_3^{(1)}]^T$,$d^{(2)} = [d_1^{(2)}, d_2^{(2)}, d_3^{(2)}]^T$.

For $d^{(0)}Qd^{(1)} = 0$,

$$d^{(0)T}Qd^{(1)} = 3d_1^{(1)} + d_3^{(1)}$$

Then, $d^{(1)} = [1, 0, -3]^T$, and thus $d^{(0)T}Qd^{(1)} = 0$.

To find $d^{(2)}$, $d^{(0)T}Qd^{(2)} = 0$ and $d^{(1)}Qd^{(2)} = 0$.

Therefore $d^{(2)} = [1, 4, -3]^T$.

10.2 Conjugate Direction Algorithm

- Let's minimize the quadratic function of n variables

$$f(x) = \frac{1}{2}x^TQx - x^Tb$$

where $Q = Q^T, x \in \Re^n$. Because $Q > 0$, the function f has a global minimizer that can be found by solving $Qx = b$.

- Basic Conjugate Direction Algorithm: given a starting point $x^{(0)}$ and Q-conjugate directions $d^{(0)}, d^{(1)}, \cdots, d^{(n-1)}$, for $k \geq 0$,

$$\begin{aligned} g^{(k)} &= \nabla f(x^{(k)}) = Qx^{(k)} - b \\ \alpha_k &= -\frac{g^{(k)T}d^{(k)}}{d^{(k)T}Qd^{(k)}}, \\ x^{(k+1)} &= x^{(k)} + \alpha_k d^{(k)} \end{aligned}$$

- **Theorem 10.1** For any starting point $x^{(0)}$, the basic conjugate direction algorithm converges to the unique x^*(that solves $Qx = b$) in n steps; that is, $x^{(n)} = x^*$.
- Ex. 10.2 find the minimizer of

$$f(x_1, x_2) = \frac{1}{2} x^T \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} x - x^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \ x \in \Re^2$$

using the conjugate direction method with the initial point $x^{(0)} = [0, 0]^T$, and Q conjugate direction $d^{(0)} = [1, 0]^T$ and $d^{(1)} = [-3/8, 3/4]^T$.

$$\begin{aligned} g^{(0)} &= -b = [1, -1]^T \\ \alpha_0 &= -\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} = -\frac{1}{4} \\ x^{(1)} &= x^{(0)} + \alpha_0 d^{(0)} = \begin{bmatrix} -1/4 \\ 0 \end{bmatrix} \end{aligned}$$

To find $x^{(2)}$,

$$\begin{aligned} g^{(1)} &= Qx^{(1)} - b = \begin{bmatrix} 0 \\ -3/2 \end{bmatrix} \\ \alpha_1 &= -\frac{g^{(1)T} d^{(1)}}{d^{(1)T} Q d^{(1)}} = 2 \\ x^{(2)} &= x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix} -1 \\ 3/2 \end{bmatrix} \end{aligned}$$

Because f is quadratic function in two variables, $x^{(2)} = x^*$.

- For a quadratic function of n variables, the conjugate direction method reaches the solution after n step.
- **Lemma 10.2** In the conjugate direction algorithm,

$$g^{(k+1)T} d^{(i)} = 0$$

for all k, $0 \le k \le n-1$, and $0 \le i \le k$.

- For example, when $k = 0$,

$$\begin{aligned} x^{(1)} &= x^{(0)} - \left(\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} \right) d^{(0)} \\ g^{(1)T} d^{(0)} &= (Qx^{(1)} - b)^T d^{(0)} \\ &= x^{(0)T} Q d^{(0)} - \left(\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} \right) d^{(0)T} Q d^{(0)} - b^T d^{(0)} \\ &= g^{(0)T} d^{(0)} - g^{(0)T} d^{(0)} = 0 \end{aligned}$$

- We can show that for all k,

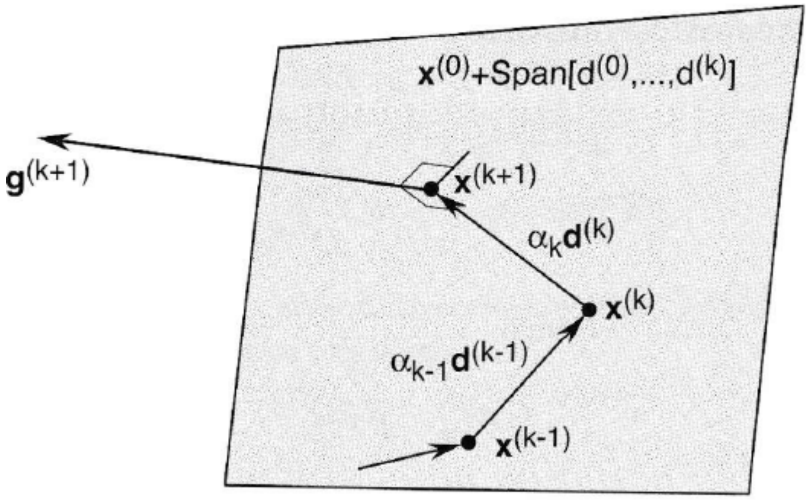
$$g^{(k+1)T} d^{(k)} = 0$$

and hence

$$\alpha_k = \arg\min f(x^{(k)} + \alpha d^{(k)})$$

(Proof):
$\frac{d\Phi_0}{d\alpha}(\alpha_k) = \nabla f(x^{(k)} + \alpha_k d^{(k)})^T d^{(k)} = g^{(k+1)T} d^{(k)}$.

- $g^{(k+1)}$ is orthogonal to any vector from subspace spanned by $d^{(0)}, d^{(1)}, \cdots, d^{(k)}$.
  ♣ Fig. 10.1 on page 157.



10.3 The Conjugate Gradient Algorithm

- The direction is calculated as a linear combination of the previous direction and the current gradient, in such a way that all the directions are mutually Q-conjugate $\Rightarrow$ Conjugate gradient algorithm.
- We find the function minimizer by performing n searches along mutually conjugate directions.
- For the quadratic function

$$f(x) = \frac{1}{2}x^TQx - x^Tb, \quad x \in \Re^n$$

where $Q = Q^T > 0$.

- The conjugate gradient algorithm
  1. Set $k := 0$; select the initial point $x^{(0)}$.
  2. $g^{(0)} = \nabla f(x^{(0)})$. If $g^{(0)} = 0$, stop, else set $d^{(0)} = -g^{(0)}$.
  3. $\alpha_k = -\frac{g^{(k)T} d^{(k)}}{d^{(k)T} Q d^{(k)}}$.
  4. $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$
  5. $g^{(k+1)} = \nabla f(x^{(k+1)})$. if $g^{(k+1)} = 0$, stop.
  6. $\beta_k = \frac{g^{(k+1)T} Q d^{(k)}}{d^{(k)T} Q d^{(k)}}$.
  7. $d^{(k+1)} = -g^{(k+1)} + \beta_k d^{(k)}$
  8. Set $k := k + 1$; go to step 3.
- **Proposition 10.1** In the conjugate gradient algorithm, the directions $d^{(0)}, d^{(1)}, \cdots, d^{(n-1)}$ are Q-conjugate.
- **Ex. 10.3** Consider the quadratic function

  $$f(x_1, x_2, x_3) = \frac{3}{2}x_1^2 + 2x_2^2 + \frac{3}{2}x_3^2 + x_1 x_3 + 2x_2 x_3 - 3x_1 - x_3$$

  Use the conjugate gradient algorithm with initial condition as $x^{(0)} = [0, 0, 0]^T$.
  We can represent f as

  $$\begin{aligned} f(x) &= \frac{1}{2}x^T Q x - x^T b, \\ Q &= \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

Then,

$$\begin{aligned} g(x) &= \nabla f(x) = Qx - b \\ &= [3x_1 + x_3 - 3, 4x_2 + 2x_3, x_1 + 2x_2 + 3x_3 - 1]^T. \end{aligned}$$

Following the algorithm procedures,
1-iteration,

$$\begin{aligned} g^{(0)} &= [-3, 0, -1]^T, \\ d^{(0)} &= -g^{(0)} \\ \alpha_0 &= -\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} = 0.2778 \\ x^{(1)} &= x^{(0)} + \alpha_0 d^{(0)} = [0.8333, 0, 0.2778]^T \\ g^{(1)} &= \nabla f(x^{(1)}) = [-0.2, 0.5, 0.6]^T \\ \beta_0 &= \frac{g^{(1)T} Q d^{(0)}}{d^{(0)T} Q d^{(0)}} = 0.08 \\ d^{(1)} &= -g^{(1)} + \beta_0 d^{(0)} = [0.46, -0.55, -0.58]^T \end{aligned}$$

2-iteration

$$\begin{aligned} \alpha_1 &= -\frac{g^{(1)T} d^{(1)}}{d^{(1)T} Q d^{(1)}} = 0.21 \\ x^{(2)} &= x^{(1)} + \alpha_1 d^{(1)} = [0.9, -0.1, 0.1]^T \\ \cdots & \quad \cdots \\ x^{(3)} &= x^{(2)} + \alpha_2 d^{(2)} = [1, 0, 0]^T \end{aligned}$$

Finally

$$g^{(3)} = \nabla f(x^{(3)}) = 0$$

Hence, $x^* = x^{(3)}$.

10.4 Conjugate Gradient Algorithm

for Non-Quadratic Problems

- Conjugate gradient algorithm minimizes a positive definite quadratic function of n variables in n steps.
- For a general nonlinear function, the Hessian is a matrix that has to be reevaluated at each iteration of the algorithm.
- An efficient implementation of the conjugate gradient algorithm that eliminates the Hessian evaluation at each step is desirable.
- Observe that Q appears only in the computation of the scalars α_k and β_k. Because

$$\alpha_k = \arg\min{}_{\alpha \geq 0} f(x^{(k)} + \alpha d^{(k)})$$

  the closed form formula for α_k in the algorithm can be replaced by a numerical line search procedure.
- We only need to concern ourselves with the formula for β_k.
- We manipulate the formula β_k in such a way that Q is eliminated.

- Hestenes-Stiefel formula

$$\beta_k = \frac{g^{(k+1)T} Q d^{(k)}}{d^{(k)T} Q d^{(k)}}$$

Since $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$. Premultiplying both sides by Q, and $g^{(k)} = Qx^{(k)} - b$, we get
$g^{(k+1)} = g^{(k)} + \alpha_k Q d^{(k)}$, which we can rewrite as
$Qd^{(k)} = (g^{(k+1)} - g^{(k)})/\alpha_k$.

$$\beta_k = \frac{g^{(k+1)T}[g^{(k+1)} - g^{(k)}]}{d^{(k)T}[g^{(k+1)} - g^{(k)}]}$$

- The above formula gives us conjugate gradient algorithms that do not require explicit knowledge of the Hessian matrix Q.
- All we need are the objective function and gradient values at each iteration.
- A few more slight modifications to apply the algorithm to general nonlinear functions in practice.
- The termination criterion $\nabla f(x^{(k+1)}) = 0$ is not practical. Instead, a suitable practical stopping criterion needs to be used.