Ch.8 Gradient Methods

8.1 Introduction

- We consider a class of search methods for real-valued functions on $\Re^n$.
- The gradient of f at x_0, denoted $\nabla f(x_0)$ is orthogonal to the tangent vector to an arbitrary smooth curve passing through x_0 on the level set $f(x) = c$.
- The direction in which $\nabla f(x)$ points is the direction of maximum rate of increase of f at x.
- The direction of negative gradient is a good direction to search if we want to find a function minimizer.
- Suppose that we are given a point $x^{(k)}$. To find the next point $x^{(k+1)}$, we start at $x^{(k)}$ and move by an amount $-\alpha_k \nabla f(x^{(k)})$, where α_k is a positive scalar called the step size.

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)})$$

- We refer to the above as a gradient descent algorithm (or gradient algorithm).

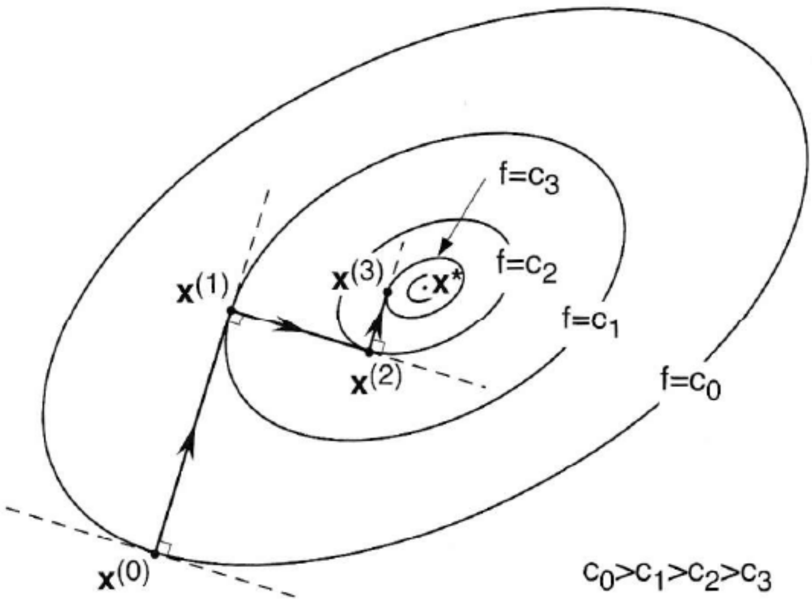
8.2 Method of Steepest Descent

- The method of steepest descent is a gradient algorithm where the step size α_k is chosen to achieve the maximum amount of decrease of the object function at each individual step.

- α_k is chosen to minimize $\phi_k(\alpha) \equiv f(x^{(k)} - \alpha \nabla f(x^{(k)}))$.

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

♣ Fig.8.2 on page 115.



- **Proposition 8.1** *If* $\{x^{(k)}\}_{k=0}^{\infty}$ *is a steepest descent sequence for a given function* $f : \Re^n \to \Re$, *then for each* k *the vector* $x^{(k+1)} - x^{(k)}$ *is orthogonal to the vector* $x^{(k+2)} - x^{(k+1)}$.

- **Proposition 8.2** *If* $\{x^{(k)}\}_{k=0}^{\infty}$ *is the steepest descent sequence for* $f : \Re^n \to \Re$ *and if* $\nabla f(x^{(k)}) \neq 0$, *then* $f(x^{(k+1)}) < f(x^{(k)})$. $\to$ *descent property.*

- Stopping condition for this algorithm is that we stop when

$$|f(x^{(k+1)}) - f(x^{(k)})| < \epsilon$$

where ϵ is a prespecified threshold.

- Another alternative is to compute the norm $\|x^{(k+1)} - x^{(k)}\|$ of the difference, and we stop if the norm is less than a prespecified threshold:

$$\|x^{(k+1)} - x^{(k)}\| < \epsilon$$

- Example 8.1 We use the method of steepest descent to find the minimizer of

$$f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$$

The initial point is $x^{(0)} = [4, 2, -1]^T$.

$$\begin{aligned} \nabla f(x) &= [4(x_1 - 4)^3,\ 2(x_2 - 3),\ 16(x_3 + 5)^3]^T \\ \nabla f(x^{(0)}) &= [0, -2, 1024]^T \end{aligned}$$

To compute $x^{(1)}$,

$$\begin{aligned} \alpha_0 &= \arg\min_{\alpha \geq 0} f(x^{(0)} - \alpha \nabla f(x^{(0)})) \\ &= \arg\min_{\alpha \geq 0} ((2 + 2\alpha - 3)^2 + 4(-1 - 1024\alpha + 5)^4) \\ &= \arg\min_{\alpha \geq 0} \phi_0(\alpha) \\ \alpha_0 &= 3.967 \times 10^{-3} \\ x^{(1)} &= x^{(0)} - \alpha_0 \nabla f(x^{(0)}) = [4.000, 2.008, -5.062]^T \end{aligned}$$

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

$$
\begin{aligned}
\nabla f(x^{(1)}) &= [0.000, -1.984, -0.003875]^T \\
\alpha_1 &= \arg\min_{\alpha \geq 0}(0 + (2.008 + 1.984\alpha - 3)^2 \\
&\quad +4(-5.062 + 0.003875\alpha + 5)^4) \\
&= \arg\min_{\alpha \geq 0}\phi_1(\alpha) \\
\alpha_1 &= 0.5000 \\
x^{(2)} &= x^{(1)} - \alpha_1 \nabla f(x^{(1)}) = [4.000, 3.000, -5.060]^T
\end{aligned}
$$

To find $x^{(3)}$,

$$
\begin{aligned}
\nabla f(x^{(2)}) &= [0.000, 0.000, -0.003525]^T \\
\alpha_2 &= \arg\min_{\alpha \geq 0}(4(-5.060 + 0.003525\alpha + 5)^4) \\
&= \arg\min_{\alpha \geq 0}\phi_2(\alpha) \\
\alpha_2 &= 16.29 \\
x^{(3)} &= x^{(2)} - \alpha_2 \nabla f(x^{(2)}) = [4.000, 3.000, -5.002]^T
\end{aligned}
$$

- Method of steepest descent with a quadratic function of the form

$$
\begin{aligned}
f(x) &= \frac{1}{2}x^T Q x - b^T x \\
Q &= A + A^T \\
x^{(k+1)} &= x^{(k)} - \alpha_k g^{(k)}, \quad g^{(k)} = \nabla f(x^{(k)}) \\
x^{(k+1)} &= x^{(k)} - \left(\frac{g^{(k)T} g^{(k)}}{g^{(k)T} Q g^{(k)}} \right) g^{(k)} \\
g^{(k)} &= \nabla f(x^{(k)}) = Q x^{(k)} - b
\end{aligned}
$$

8.3 Analysis of Gradient Methods

- An iterative algorithm is globally convergent if for any arbitrary starting point the algorithm is guaranteed to generate a sequence of points converging to a point that satisfies the FONC for a minimizer. (vs. locally convergent)
- Rate of convergence: how fast the algorithm converges to a solution point.
- **Theorem 8.2** *In the steepest descent algorithm, we have* $x^{(k)} \to x^*$ *for any* $x^{(0)}$.
- **Theorem 8.3** *For the fixed step size gradient algorithm,* $x^{(k)} \to x^*$ *for any* $x^{(0)}$ *if and only if*

$$0 < \alpha < \frac{2}{\lambda_{max}(Q)}$$

- Example 8.4

$$\begin{aligned} f(x) &= x^T \begin{bmatrix} 4 & 2\sqrt{2} \\ 0 & 5 \end{bmatrix} x + x^T \begin{bmatrix} 3 \\ 6 \end{bmatrix} + 24 \\ x^{(k+1)} &= x^{(k)} - \alpha \nabla f(x^{(k)}) \\ f(x) &= \frac{1}{2} x^T \begin{bmatrix} 8 & 2\sqrt{2} \\ 2\sqrt{2} & 10 \end{bmatrix} x + x^T \begin{bmatrix} 3 \\ 6 \end{bmatrix} + 24 \end{aligned}$$

The eigenvalues of the matrix in the quadratic term are

6 and 12. The above algorithm converges to the minimizer for all $x^{(0)}$ if and only if α lies in the range $0 < \alpha < 2/12$.

- **Definition 8.1** *Given a sequence $\{x^{(k)}\}$ that converges to x^*, that is, $\lim_{k\to\infty} \|x^{(k)} - x^*\| = 0$, we say that the order of convergence is p, if*

$$0 < \lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^p} < \infty$$

If for all $p > 0$,

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^p} = 0$$

then we say that the order of convergence is ∞.

- Example 8.5 1. Suppose that $x^{(k)} = 1/k$, and thus $x^{(k)} \to 0$. Then

$$\frac{|x^{(k+1)}|}{|x^{(k)}|^p} = \frac{1/(k+1)}{1/k^p} = \frac{k^p}{k+1}$$

If $p < 1$, the above sequence converges to 0, whereas if $p > 1$, it grows to ∞. If $p = 1$, the sequence converges to 1.
2. Suppose that $x^{(k)} = \gamma^{(q^k)}$, where $q > 1$ and $0 < \gamma < 1$, and thus $x^{(k)} \to 0$. Then

$$\frac{|x^{(k+1)}|}{|x^{(k)}|^p} = \frac{\gamma^{(q^{k+1})}}{(\gamma^{(q^k)})^p} = \gamma^{(q^{k+1} - pq^k)} = \gamma^{(q-p)q^k}$$

If $p < q$, the above sequence converges to 0, whereas if $p > q$, it grows to ∞. If $p = q$. the sequence converges to 1. Hence, the order of convergence is q.

- The order of convergence can be interpreted using the notion of the order symbol O.
- **Theorem 8.5** *Let $\{x^{(k)}\}$ be a sequence that converges to x^*. If*

$$\|x^{(k+1)} - x^*\| = O(\|x^{(k)} - x^*\|^p)$$

then the order of convergence is at least p.

- Example 8.6 Suppose that we are given a scalar sequence $\{x^{(k)}\}$ that converges with order of convergence p and satisfies

$$\lim_{k\to\infty} \frac{|x^{(k+1)} - 2|}{|x^{(k)} - 2|^3} = 0$$

The limit of $\{x^{(k)}\}$ must be 2 since $|x^{(k+1)} - 2| \to 0$. Since $|x^{(k+1)} - 2| = o(|x^{(k)} - 2|^3)$, hence $p > 3$.

- Example 8.7

$$f(x) = x^2 - \frac{x^3}{3}$$

Suppose we use the algorithm $x^{(k+1)} = x^{(k)} - \alpha f'(x^{(k)})$ with step size $\alpha = 1/2$ and initial condition $x^{(0)} = 1$.

$f'(x) = 2x - x^2$.

$$x^{(k+1)} = x^{(k)} - \alpha f'(x^{(k)}) = \frac{1}{2}(x^{(k)})^2$$

With $x^{(0)} = 1$, we can derive the expression $x^{(k)} = (1/2)^{2^k - 1}$. The order of convergence is

$$\frac{|x^{(k+1)}|}{|x^{(k)}|^2} = \frac{(1/2)^{2^{k+1}-1}}{(1/2)^{2^{k+1}-2}} = \frac{1}{2}$$

Therefore, the order of convergence is 2.