Extreme values: Iterative methods (II)

Discrete Mathematics and Optimization
Bioinformatics

1. Recall

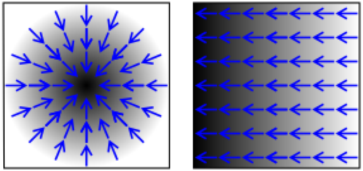
Definition

An iterative method to find an extreme value of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is a procedure to find a sequence

$$\mathbf{x}_k = h(\mathbf{x}_{k-1}),\ k \geq 1,$$

from some initial value $\mathbf{x}_0$ such that $\mathbf{x}_k$ converges to the optimal value.

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. At each point $\mathbf{x}_0 \in D$ the gradient $\nabla f(\mathbf{x}_0)$ points at the direction of most rapid increase of the function at x_0. The rate of increase is $\|\nabla f(\mathbf{x}_0)\|$.

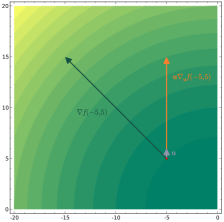


1. Recall

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. At each point $\mathbf{x}_0 \in D$ the gradient $\nabla f(\mathbf{x}_0)$ points at the direction of most rapid increase of the function at x_0. The rate of increase is $\|\nabla f(\mathbf{x}_0)\|$.

Fix $\mathbf{u}$ a unit vector $\|\mathbf{u}\| = 1$.

- $\phi_{\mathbf{u}}(t) = f(\mathbf{x}_0 + t\mathbf{u})$.
- $\phi'(t) = \nabla f(\mathbf{x}_0 + t\mathbf{u}) \cdot \mathbf{u}$.
- $\phi'_{\mathbf{u}}(0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{u}$.
- The scalar product is maximum when $\mathbf{u}$ is parallel to $\nabla f(\mathbf{x}_0)$.
- In that case , $|\phi'_{\mathbf{u}}(0)| = \|\nabla f(\mathbf{x}_0)\|$.



2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k), \ k \geq 0,$

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

Choose t_k as to locally optimize the decrease (greedy strategy)

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

Example: $f(x, y) = 4x^2 - 4xy + 2y^2$; $\nabla f(x, y) = (8x - 4y, -4x + 4y)$

- Initial point $\mathbf{x}_0 = (2, 3)$; $\nabla f(2, 3) = (4, 4)$.

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k), \; k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)), \; t \geq 0.$$

Example: $f(x, y) = 4x^2 - 4xy + 2y^2$; $\nabla f(x, y) = (8x - 4y, -4x + 4y)$

- Initial point $\mathbf{x}_0 = (2, 3)$; $\nabla f(2, 3) = (4, 4)$.
- $\phi_{\mathbf{x}_0}(t) = f((2, 3) - t(4, 4))$;
- $\phi'_{\mathbf{x}_0}(t) = \nabla f(2 - 4t, 3 - 4t) \cdot (4, 4) = -16(2 - 4t) \longrightarrow t_0 = 1/2$.
- $\mathbf{x}_1 = \mathbf{x}_0 - (1/2)\nabla f(\mathbf{x}_0) = (0, 1)$.

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

Example: $f(x, y) = 4x^2 - 4xy + 2y^2$; $\nabla f(x, y) = (8x - 4y, -4x + 4y)$

- First point $\mathbf{x}_1 = (0, 1)$; $\nabla f(0, 1) = (-4, 4)$.

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

Example: $f(x, y) = 4x^2 - 4xy + 2y^2$; $\nabla f(x, y) = (8x - 4y, -4x + 4y)$

- First point $\mathbf{x}_1 = (0, 1)$; $\nabla f(0, 1) = (-4, 4)$.
- $\phi_{\mathbf{x}_1}(t) = f((0, 1) - t(-4, 4))$;
- $\phi'_{\mathbf{x}_1}(t) = \nabla f(4t, 1 - 4t) \cdot (-4, 4) = -16(2 - 20t) \longrightarrow t_1 = 1/10.$
- $\mathbf{x}_2 = \mathbf{x}_1 - (1/10)\nabla f(\mathbf{x}_1) = (4/10, 6/10)$.

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

Example: $f(x, y) = 4x^2 - 4xy + 2y^2$; $\nabla f(x, y) = (8x - 4y, -4x + 4y)$

- Second point $\mathbf{x}_2 = (4/10, 6, 10))$; $\nabla f(4/10, 6, 10) = (8/10, 8/10)$.
- $\phi_{\mathbf{x}_2}(t) = f((4/10, 6/10) - t(8/10, 8/10))$;
- $\phi'_{\mathbf{x}_2}(t) = \nabla f(4/10 - 8t/10, 6/10 - 8/10t) \cdot (8/10, 8/10) = -16(2 - 20t) \longrightarrow t_2 = 1/10$.
- $\mathbf{x}_3 = \mathbf{x}_2 - (1/10)\nabla f(\mathbf{x}_2) = (0, 2/10)$.

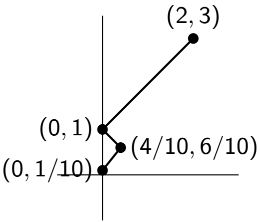
2. The method of steepest descent

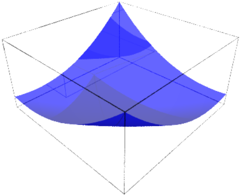
Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$





2. The method of steepest descent

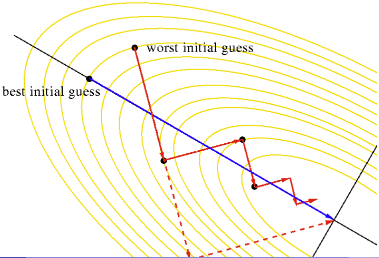
Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

- At each step the direction becomes orthogonal to the previous one.



2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

- At each step the direction becomes orthogonal to the previous one.

$$(\mathbf{x}_{k+1} - \mathbf{x}_k) \cdot (\mathbf{x}_{k+2} - \mathbf{x}_{k+1}) = t_{k+1} t_k \nabla f(\mathbf{x}_k) \cdot \nabla f(\mathbf{x}_{k+1}) = 0$$

because

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)) \to 0 = \phi_k'(t_k) = -\nabla f(\mathbf{x}_{k+1}) \cdot \nabla f(\mathbf{x}_k).$$

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

- It is a descending method:

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$$

because

$$f(\mathbf{x}_{k+1}) = \phi_k(t_k) \leq \phi_k(\tilde{t}) < \phi_k(0) = f(\mathbf{x}_k).$$

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

- It converges for every initial point $\mathbf{x}_0$ under general conditions.
  - if $f \in \mathcal{C}^1(\mathbb{R}^n)$ and

    $$C(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$$

    is bounded.
- If $Hf(\mathbf{x}_{min})$ is positive definite, it converges linearly.

2. The method of steepest descent

Definition

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{C}^1$. The Steepest Descent Sequence with initial point $\mathbf{x}_0$ is defined by

- $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k),\ k \geq 0,$
- t_k is chosen as to minimize the function

$$\phi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)),\ t \geq 0.$$

- It converges linearly under some general conditions.
- It is a descending method.
- It runs by zig-zag, it may be not efficient.
- It only requires computation of ∇f
- In practical applications there are several stopping rules: if $f|(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)| < \epsilon$, if $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \epsilon$, if $\|\nabla f(\mathbf{x}_{k+1})\| < \epsilon$,...

3. Iterative methods

General framework of iterative methods

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$$

At point $\mathbf{x}_k$,

- Choose a descent direction $\mathbf{d}_k$: $\mathbf{d}_k \cdot \nabla f(\mathbf{x}_k) < 0$.
- Determine t_k such that $f(\mathbf{x}_k + t_k \mathbf{d}_k) < f(\mathbf{x}_k)$ (line search).
- Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$

3. Iterative methods

General framework of iterative methods

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$$

At point $\mathbf{x}_k$,

- Choose a descent direction $\mathbf{d}_k$: $\mathbf{d}_k \cdot \nabla f(\mathbf{x}_k) < 0$.
- Determine t_k such that $f(\mathbf{x}_k + t_k \mathbf{d}_k) < f(\mathbf{x}_k)$ (line search).
- Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$

Goals to achieve:

- Convergence: The sequence must converge to a minimization point.
- Efficiency
  - Each step must be computationally simple.
  - The number of steps must be small.
- Robustness: Apply to a broad class of functions.
- Stability: No dramatic changes on choices of initial point $\mathbf{x}_0$.

3. Iterative methods

General framework of iterative methods

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$$

At point $\mathbf{x}_k$,

- Choose a descent direction $\mathbf{d}_k$: $\mathbf{d}_k \cdot \nabla f(\mathbf{x}_k) < 0$.
- Determine t_k such that $f(\mathbf{x}_k + t_k \mathbf{d}_k) < f(\mathbf{x}_k)$ (line search).
- Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$

	Newton	Stepest descent
Convergence	Not always	Reasonably wide
Complexity of steps	Gradient and Hessian	Gradient and minimization
Number of steps	Quadratic	Can be inefficient
Robustness	Quadratic functions	Reasonably wide
Stability	Closedness	Reasonable

3. Iterative methods

General framework of iterative methods

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$$

At point $\mathbf{x}_k$,

- Choose a descent direction $\mathbf{d}_k$: $\mathbf{d}_k \cdot \nabla f(\mathbf{x}_k) < 0$.
- Determine t_k such that $f(\mathbf{x}_k + t_k \mathbf{d}_k) < f(\mathbf{x}_k)$ (line search).
- Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$

Some alternatives:

- Conjugate gradient method
- QuasiNewton method

5. Summary

What we did

- We provided tools from caluclus for the search an analysis of extremal values of functions of several variables.
- We have discussed convexity and its role in optimization.
- We have described the two basic iterative methods for optimization: Newton method and Gradient method, discussing some of its performances.

What we mentioned

- The existence of convex optimization problems.
- Brief description of refinements of iterative methods: conjugate gradient descent and quasi-Newton methods.

What we did not do

- Discussing specific optimization problems (Least Squares, Curve fitting,...)
- Secant methods
- Constrained optimization: adding constraints to the optimization problem.
- Heuristic methods