Smooth unconstrained minimization

Pramesh Kumar

IIT Delhi

September 4, 2025

Unconstrained problems

To $\min_{\mathbf{x}} f(\mathbf{x})$, where f is convex and differentiable, we generally adopt an iterative procedure.

- Start with some initial point $\mathbf{x}^{(0)}$ and then generate a sequence of points $\{\mathbf{x}_k\}$.
- We want improving objective values in each iteration i.e., $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$.
- Hopefully, our sequence of points $\{\mathbf{x}^{(k)}\}$ will converge to a local minimizer $\mathbf{x}^*$ (or global minimizer).

Descent methods

General procedure

From $\mathbf{x}^{(0)}$, we generate a sequence of points using the following procedure:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)}$$

where, $\mathbf{d}^{(k)}$ is called the search direction (must be descent direction[1] i.e., $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0$) and t_k is called the step size. Continue until a stopping criterion is satisfied.

Line search types

- Exact line search $t_k = \text{argmin}_{t>0}\{f(\mathbf{x}^{(k)} + t\mathbf{d}^{(k)})\}$ (can be done using Bisection method)
- Backtracking line search For $0 < \alpha < 0.5$ and $0 < \beta < 1$
  - Start with $t = 1$, update $t := \beta t$
  - until $f(\mathbf{x} + t\mathbf{d}) \leq f(\mathbf{x}) + t\alpha \nabla f(\mathbf{x})^T \mathbf{d}$

[1] By Taylor expansion $f(\mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)}) \approx f(\mathbf{x}^{(k)}) + t_k \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)}$. We need to choose $\mathbf{d}^{(k)}$ such that $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0$

Gradient descent method

initialize $\mathbf{x}^{(0)} \in \mathbf{dom}(f)$. Set $k = 0$ and tolerance $\epsilon > 0$
repeat

1. Evaluate $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ (check if this is a decent direction)
2. Use a line search method to evaluate t_k
3. Update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)}$; Set $k := k + 1$.
4. Go back to Step 1 until $\|\nabla f(\mathbf{x}^{(k)})\| \leq \epsilon$ in which case output $\mathbf{x}^{(k)}$ as the solution.

Remark. Gradient method is guaranteed to converge to a local minimizer. As we know if $f(\mathbf{x})$ is convex, local is also global.

Newton's method

In this method, at any point $\mathbf{x}^{(k)}$, we approximate the objective function $f(\mathbf{x})$ by its second order Taylor series expansion:

$$f(\mathbf{x}) \approx f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^T(\mathbf{x}-\mathbf{x}^{(k)}) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^{(k)})^T \nabla^2 f(\mathbf{x}^{(k)})(\mathbf{x}-\mathbf{x}^{(k)}) \quad (1)$$

We minimize (1) to get

$$\mathbf{x}^* = \mathbf{x}^{(k)} - [\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (2)$$

In Newton's method, the search direction (also called Newton's step) is:

$$\mathbf{d}^{(k)} = -[\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (3)$$

Question: Is this a descent direction? Answer: Yes, because $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})^T [\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)}) < 0$ since $\nabla^2 f(\mathbf{x}^{(k)})$ is positive definite for strictly convex $f(\mathbf{x})$

Newton's method

initialize $\mathbf{x}^{(0)} \in \mathbf{dom}(f)$. Set $k = 0$ and tolerance $\epsilon > 0$
repeat

1. Evaluate $\mathbf{d}^{(k)} = -[\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)})$
2. Use a line search method to evaluate t_k
3. Update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)}$; Set $k := k + 1$.
4. Go back to Step 1 until $\|\nabla f(\mathbf{x}^{(k)})\| \leq \epsilon$ in which case output $\mathbf{x}^{(k)}$ as the solution.

A few remarks

Gradient method

- No matter where it starts, it will always converge to a local minimizer.
- It is easy to implement.
- Only need to know the first-order (gradient) information
- Linear convergence rate

Newton's method

- Sensitive to the initial point
- Require second-order information (second-order derivative)
- Quadratic convergence rate (much faster than the gradient)

Suggested reading

Boyd, Stephen P., and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

Thank you!