Unconstrained Optimization

Anders Poirel

June 5, 2019

University of California, Santa Cruz

Multivariable Calculus

- Gradient of f: $\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$
- $\mathbf{x}^*$ is a stationary point of f: $\nabla f(\mathbf{x}^*) = \mathbf{0}$
- Directional derivative of f at $\mathbf{x}$ along direction $\mathbf{v}$:
  $\frac{d}{dt} f(\mathbf{x} + t\mathbf{v})\Big|_{t=0} = \nabla f(\mathbf{x}) \cdot \mathbf{v}$
- Hessian of f: $\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$
- $n \times n$ matrix A is positive definite: $\forall \mathbf{z} \neq \mathbf{0}, \mathbf{z}^T A \mathbf{z} > 0$

Sufficient Conditions for Local Minima
Suppose ∇^{f} is continous in an open neighborhood of $\mathbf{x}^*$,and $\nabla f(\mathbf{x}^*) = \mathbf{0}$, and $\nabla^2 f(\mathbf{x}^*)$ is positive definite.
Then $\mathbf{x}^*$ is a strict local minimum.

Problem

Let $f : \mathbb{R}^n \to \mathbb{R}$.

- We assume f is C^1, that is, for all $\mathbf{x}$, $\nabla f(\mathbf{x})$ exist and is continuous.
- We want to find $\text{argmin}_{\mathbf{x}} f(\mathbf{x})$
- However, this problem does not always have a closed solution (or finding one is impractical), so we want a numerical method to solve this.

Descent Directions (1)

Steepest descent
At $\mathbf{x}$, the direction along which f is decreasing the most rapidly is $\mathbf{p} = -\nabla f(\mathbf{x})$

Proof
Let θ be the angle between ∇f and $\mathbf{p}$.

The directional derivative of f at $\mathbf{x}$ is given by $\nabla f(\mathbf{x}) \cdot \mathbf{p} = \|\nabla f(\mathbf{x})\| \|\mathbf{p}\| \cos(\theta)$. This is minimized when $\cos(\theta) = \pm\pi$, that is, $\mathbf{p} = -c\nabla f(\mathbf{x})$. $\square$

Descent Directions (2)

- In general, we are "going down" as long as the direction is not orthorgonal to ∇f:
- $\mathbf{p}$ is a descent direction at $\mathbf{x}$: $-\frac{\pi}{2} < (\nabla f(\mathbf{x}), \mathbf{p}) < \frac{\pi}{2}$
- So to tend towards a minimum from a point $\mathbf{x_k}$, perform an update of the form $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha \mathbf{p_k}$

Wolfe Conditions (1)

Sufficient Decrease

- α_k should yield a large enough decrease in f:

  $$f(\mathbf{x_k}+\alpha_k \mathbf{p_k}) \leq f(\mathbf{x_k})+c_1\alpha_k \nabla f(\mathbf{x_k})^{\top}$$

  Where $0 < c_1 < 1$

- i.e reduction in f should be proportional to both the step length and the directional derivative.

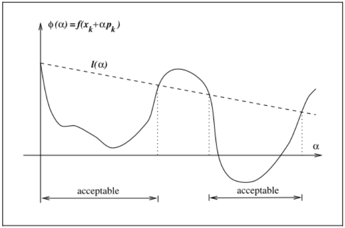

Figure 1: Sufficient decrease condition

Curvature Condition

- Sufficient decrease condition is satisfied for all sufficiently small a, so we want to prevent steps lengths α_k that are too small: $\nabla f(\mathbf{x_k} + \alpha_k \mathbf{p_k})^\top \mathbf{p_k} \geq c_2 \nabla f(\mathbf{x_k})^\top \mathbf{p_k}$ Where $0 < c_1 < c_2 < 1$.
- i.e. the slope be decreased by a sufficient amount.

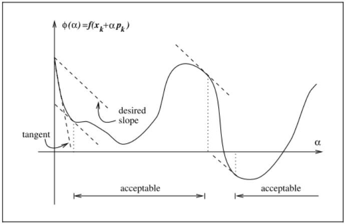

Figure 2: Curvature condition

Line Search Methods

Putting all of this together we get an optimization algorithm:
Set a starting point $\mathbf{x_0}$. Then for $k = 1, 2, \ldots$,

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_k \mathbf{p_k}$$

Where $\mathbf{p_k}$ is a descent direction and α_k satisfies the Wolfe conditions. As we get closer to a stationary point, $\mathbf{p_k}$ approaches $\mathbf{0}$.

Lipschitz Continuity

- $F : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continous on S: $\exists L > 0$ such that $\forall \mathbf{x}, \mathbf{y} \in S$, $\|F(\mathbf{x}) - F(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$

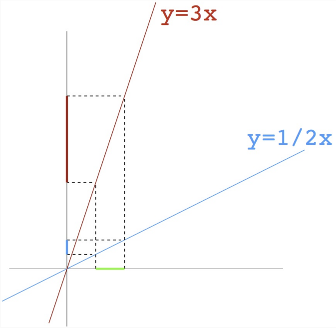

Figure 3: Lipschitz continuous functions

Convergence of Line Search (1)

Zoutendjik's Theorem
Consider an iteration of the form $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_k \mathbf{p_k}$, where:

- $\mathbf{p_k}$ is a descent direction and α_k satisfies the Wolfe conditions.
- f is bounded below in $\mathbb{R}^n$
- f continuously differentiable in open set $\mathcal{N}$ containing level set $\mathcal{L} = \{x : f(\mathbf{x}) \leq f(\mathbf{x_0})\}$, where $\mathbf{x_0}$ is the starting point of the iteration.
- ∇f is Lipschitz continous on $\mathcal{N}$.

Then, $\exists M \in \mathbb{R}$ such that

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|\nabla f\|^2 \leq M$$

Convergence of Line Search (2)

Proof
From second Wolfe condition, and $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_k \mathbf{p_k}$,

$$(\nabla f(\mathbf{x_{k+1}}) - \nabla f_k)^\top \mathbf{p_k} \geq (c_2 - 1)\nabla f_k^\top \mathbf{p_k}$$

And the Lipschitz condition implies that $\exists L$ such that

$$(\nabla f(\mathbf{x_{k+1}}) - \nabla f(\mathbf{x_k}))^\top \mathbf{p_k} \leq \alpha_k L \|\mathbf{p_k}\|^2$$

Convergence of Line Search (3)

Proof (continued)
Combining the two previous results yields

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{(\nabla f(\mathbf{x_k})^\top \mathbf{p_k})^2}{\|\mathbf{p_k}\|^2}$$

Substituting this into the first Wolfe condition gives

$$f(\mathbf{x_{k+1}}) \leq f(\mathbf{x_k}) - c_1 \frac{1 - c_2}{L} \frac{(\nabla f(\mathbf{x_k})^\top \mathbf{p_k})^2}{\|\mathbf{p_k}\|^2}$$

Convergence of Line Search (4)

Proof (continued)

Using $cos(\theta_k) = \frac{\nabla f(\mathbf{x_k})^\top \mathbf{p_k}}{\|\nabla f(\mathbf{x_k})\| \|\mathbf{p_k}\|}$ yields

$$f(\mathbf{x_{k+1}}) \leq f(\mathbf{x_k} - c\cos^2\theta_k)\|\nabla f(\mathbf{x_k})\|^2$$

where $c = \frac{c_1(1-c_2)}{L}$.

Sum this expression over all indices up to k to obtain

$$f(\mathbf{x_{k+1}}) \leq f(\mathbf{x_0}) - c\sum_{j=0}^{k}\cos^2\theta_j)\|\nabla f(\mathbf{x_j}).$$

Convergence of Line Search (5)

We know that f is bounded below, hence for some C, $f(\mathbf{x_0}) - f(x\mathbf{x_{k+1}}) < C$. By taking limits in the expression in the previous slide, we get

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|\nabla f(\mathbf{x_k})|^2 < M$$

For some $M > 0$. $\square$

Convergence of Line Search (6)

Line Search Converges
The sequence of points $\{\mathbf{x_k}\}$ given by $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha_k \mathbf{p_k}$ (where $\mathbf{p_k}$ is a descent direction and α_k satisfies the Wolfe conditions) converges to a stationary point.

Proof
As p_k is a descent direction, $\frac{\pi}{2} < \theta_k < \frac{\pi}{2}$. Thus $\exists \delta > 0$ such that for all k, $\cos(\theta_k) > \delta$. We have $\sum_{k=0}^{\infty} \cos^2(\theta_k) \|\nabla f(\mathbf{x_k})\|^2 < M$ hence we must have

$$\lim_{k \to \infty} \|\nabla f(\mathbf{x_k})\| = 0$$

$\square$

Visualization

https://www.benfrederickson.com/numerical-optimization/

References

- W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1976.
- J. Nocedal, S. Wright, *Numerical Optimization*, Springer, 2006.