MAT 3007 Optimization: Tutorial 13

Kangran ZHAO
kangranzhao@link.cuhk.edu.cn

School of Data Science,
Chinese University of Hong Kong, Shen Zhen

July 16, 2025

Recap: General Method

Algorithm 1 (General Scheme).

Choose initial point x^0 and tolerance ϵ>0, set hyperparameters (like σ, γ in Armijo line search) , let k=0

(1). *Check if $\|\nabla f(x^k)\| \le \epsilon$, stop and output x^k; otherwise, continue to Step 2.*

(2). *Calculate d^k.*

(3). *Select a proper step-size α_k.*

(4). *Update $x^{k+1} = x^k + \alpha_k d^k$, let $k = k + 1$. Go back to Step 1.*

Recap: Parameter Stategies

Choice for d^k:

- Gradient Descent: $d^k = -\nabla f(x^k)$;
- Newton Method: $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$.

Choice for α_k:

- Constant;
- Exact line search: $\alpha_k = \underset{\alpha}{\operatorname{argmin}} f(x^k + \alpha d^k)$;
- Armijo line search (backtracking): Choose σ, $\gamma \in (0, 1)$. If

$$f(x^k + \alpha d^k) \leq f(x^k) + \gamma \cdot \alpha \nabla f(x)^\top d^k,$$

  then choose $\alpha_k = \alpha$; otherwise, set $\alpha = \sigma \cdot \alpha$ and repeat the step;
- Diminishing;

Recap: Convergence Analysis

- When f is continuously differentiable and let $\{x^k\}_k$ be generated by the gradient method for solving

$$\min f(x), \text{ s.t. } x \in \mathbb{R}^n$$

  with one of the following step size:
  - exact line search
  - Armijo line search with $\sigma, \gamma > 0$.

  Then $\{f(x^k)\}_k$ is nonincreasing and every accumulation point of $\{x^k\}_k$ is a stationary point of f.
- Let $f \in C_L^{1,1}$ and be strongly convex. Let $\{x^k\}_k$ be generated by gradient method with the following stepsize:
  - constant step size $\alpha < \frac{2}{L}$.
  - exact linear search
  - Armijo linear search

  x^* be the solution of $\min f(x)$. Then $\{x^k\}_k$ converges linearly to x^*.

Exercise 1: Algorithms with Quadratic function

Consider the following quadratic problem

$$\min_{x} \quad f(x) := \frac{1}{2}x^\top Gx + b^\top x$$

where $G > 0$.

(a) We now use **gradient descent method** with **exact line search**, show that if $x^k - x^*$ is the eigenvector of G, then $x^{k+1} = x^*$.

(b) We now use **Newton method** with **exact line search**, show that wherever initial point x^0 is, $x^1 = x^*$.
i.e. Convex Quadratic function with Newton method and exact line search, one step to optimality!

Solution to Exercise 1 (a)

By optimality condition $\nabla f(x^*) = 0$, i.e., $b = -Gx^*$, and if there exist some λ such that $G(x^k - x^*) = \lambda(x^k - x^*)$, then

$$d^k \overset{\text{alg.}}{=} -\nabla f(x^k) = -(Gx^k + b) = -G(x^k - x^*) = -\lambda(x^k - x^*).$$

For exact line search, we have

$$\begin{aligned}\alpha_k &= \arg_\alpha \min \quad f(x^k + \alpha d^k) \\ &= \arg_\alpha \min \quad \frac{1}{2}(x^k + \alpha d^k)^\top G(x^k + \alpha d^k) + b^\top (x^k + \alpha d^k) \\ &= \frac{-(Gx^k + b)^\top d^k}{{d^k}^\top G d^k} = \frac{{d^k}^\top d^k}{{d^k}^\top G d^k} = 1/\lambda.\end{aligned}$$

Finally, we have $x^{k+1} = x^k + \alpha_k d^k = x^k - \frac{1}{\lambda}\lambda(x^k - x^*) = x^*$.

Solution to Exercise 1 (b)

By optimality condition $\nabla f(x^*) = 0$, i.e., $b = -Gx^*$, then

$$d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k) = -G^{-1} G(x^k - x^*) = -(x^k - x^*).$$

For exact line search, we have

$$\begin{aligned}
\alpha_k &= \arg_\alpha \min \quad f(x^k + \alpha d^k) \\
&= \arg_\alpha \min \quad \frac{1}{2}(x^k + \alpha d^k)^\top G(x^k + \alpha d^k) + b^\top (x^k + \alpha d^k) \\
&= \frac{-(Gx^k + b)^\top d^k}{{d^k}^\top G d^k} = \frac{{d^k}^\top G d^k}{{d^k}^\top G d^k} = 1.
\end{aligned}$$

Finally, we have $x^{k+1} = x^k + \alpha_k d^k = x^k - 1 * (x^k - x^*) = x^*$.

Exercise 2: Algorithm May Diverge

Consider the function

$$f(x) = |x|^{3/2}, \quad x \in \mathbb{R}.$$

(a). Show the gradient of f is not Lipschitz continuous for any L.

(b). Analyze the behavior of $\{x^k\}$ when implement the gradient descent method with constant step size α.

(c). Show that if the initial point is not 0, then Newton's method will not converge with constant step size $\alpha \geq 1$.

Solution to Exercise 2(a)

We first compute its gradient and Hessian,

$$f'(x) = \begin{cases} \frac{3}{2}x^{1/2}, & x \geq 0 \\ -\frac{3}{2}(-x)^{1/2}, & x \leq 0 \end{cases}$$
$$f''(x) = \begin{cases} \frac{3}{4}x^{-1/2}, & x > 0 \\ \frac{3}{4}(-x)^{-1/2}, & x < 0 \end{cases}$$

For any fixed $L > 0$, since $\lim_{x \to 0^+} f''(x) \to +\infty$, there exists $\delta > 0$, such that $f''(\delta) > L$. Thus, $\nabla f(x)$ will not be L- Lipschitz continuous on $[\delta, +\infty)$. Thus, $\nabla f(x)$ is not Lipschitz continuous for any L.

Solution to Exercise 2(b)

When $x > 0$, the iteraitons of gradient descent method is:

$$x^{k+1} = x^k - \alpha f'(x^k) = \begin{cases} x^k - \frac{3\alpha}{2}\sqrt{|x^k|}, & x^k \geq 0 \\ x^k + \frac{3\alpha}{2}\sqrt{|x^k|}, & x^k \leq 0 \end{cases}$$

Here we suppose $x^k > 0$, the other case is similar.

(1). If $x^0 = 0$, then this is the minimizer.

(2). If $|x^0| > \frac{9\alpha^2}{4}$, x^{k+1} will have the same sign and decrease and when $\exists N$ such that $|x^N| = \frac{9\alpha^2}{4}$, then the algorithm terminates.

(3). If $0 < |x^0| < \frac{9\alpha^2}{4}$, $\{x^{k+1}\}$ does not converge since x^k and x^{k+1} have different sign. However, $|x^k|$ will converge to $\frac{9\alpha^2}{16}$.

- When $\frac{9\alpha^2}{16} < |x^k| < \frac{9\alpha^2}{4}$, $|x^{k+1}| < |x^k|$, since $|x^{k+1}| = \frac{3\alpha}{2}\sqrt{x^k} - x^k \leq x^k$.
- When $|x^k| \leq \frac{9\alpha^2}{16}$, $|x^k| \leq |x^{k+1}| \leq \frac{9\alpha^2}{16}$ since in this case

$$\frac{9\alpha^2}{16} \geq |x^{k+1}| = \frac{3\alpha}{2}\sqrt{x^k} - x^k \geq x^k.$$

Solution to Exercise 2(b)

Thus, when $|x^N| \leq \frac{9\alpha^2}{16}$, $\{|x^k|\}_{k\geq N}$ is nondecreasing and has an upper bound. So the sequence converges. Since

$$|x^{k+1}| = \frac{3\alpha}{2}\sqrt{|x^k|} - |x^k|,$$

if we take the limit on both sides, we can conclude that $|x^k| \to \frac{9\alpha^2}{16}$.
Example: When $x^0 = 10, \alpha = 1$, the sequence will be:

$$10, 5.25658, 1.81750, -0.20472, 0.47397,$$
$$-0.55871, 0.56249, -0.56250, 0.56250, \ldots$$

Solution to Exercise 2(c)

Since $\frac{f'(x)}{f''(x)} = 2x$ for $x > 0$, the iteraitons of Newton's method is:

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

Since $\alpha \geq 1$, $1 - 2\alpha \leq -1$. So if we don't start from 0, Newton's method will generate points:

$$x^k = (1 - 2\alpha)^k x^0$$

which doesn't converge because $|x^k| \geq |x^0|, k = 1, 2, \ldots$

Acknowledgements

Thanks for your coming!

Acknowledgements:
Prof. Andre Milzarek
Prof. Zizhuo WANG,
Prof. Haoxiang Yang
Mr. Wentao Ding
Mr. Hanbin Yang