Plan

① Steepest Descent

$$x_{k+1} = x_k - \eta_k \nabla f(x_k)$$

$$\eta_k = \underset{\eta>0}{\operatorname{argmin}}\, f(x_k - \eta \nabla f(x_k))$$

② Numerics for quadratic examples

$$f(x) = \frac{1}{2} x^T K x - x^T b + c$$

③ Lemma $f(x)$ is convex and $\nabla f(x)$ is L-cont. with L

$$\Longleftrightarrow \quad 0 \le \langle \nabla f(x) - \nabla f(y), x-y \rangle \le L\|x-y\|^2$$

Theorem 2.13. *For a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, assume $\mu I \le \nabla^2 f(x) \le LI$ where $L > \mu > 0$ are constants (eigenvalues of Hessian have uniform positive bounds), thus f is strongly convex has a unique minimizer $\mathbf{x}_*$. Then the steepest descent method (2.9) satisfies*

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}_*) \le \left(1 - \frac{\mu}{L}\right)^k [f(\mathbf{x}_0) - f(\mathbf{x}_*)].$$

Remark: With strong convexity, and L-cont. ∇f,

$$\|x_k - x^*\|^2 \le \left(1 - \frac{2\eta\mu L}{\mu+L}\right)^k \|x_0 - x^*\|^2$$

$$f(x_k) \le f(x_*) + \nabla f(x_*)^T (x_k - x_*) + \frac{L}{2}\|x_k - x_*\|^2$$

$$\Rightarrow \quad f(x_k) - f(x_*) \le \frac{L}{2}\|x_k - x_*\|^2$$

$$\eta = \frac{2}{L+\mu} \Rightarrow \left(\frac{L-\mu}{L+\mu}\right)^2 < 1 - \frac{\mu}{L}$$

Proof. For convenience, let $\mathbf{h}_k = \nabla f(\mathbf{x}_k)$. By Multivariate Quadratic Taylor's Theorem (Theorem 1.4), for any $\alpha > 0$, there exists $\theta \in (0, 1)$ and $\mathbf{z}_k = \mathbf{x}_k + \theta(\mathbf{x}_k - \alpha\mathbf{h}_k)$ such that

$x_{k+1} = x_k - \eta_k h_k$

$$f(\mathbf{x}_k - \alpha\mathbf{h}_k) = f(\mathbf{x}_k) - \alpha\mathbf{h}_k^T \nabla f(\mathbf{x}_k) + \frac{1}{2}\alpha^2 \mathbf{h}_k^T \nabla^2 f(\mathbf{z}_k)\mathbf{h}_k.$$

The assumption $\nabla^2 f(\mathbf{x}) \le LI, \forall \mathbf{x}$ implies

$$f(\mathbf{x}_k - \alpha\mathbf{h}_k) \le f(\mathbf{x}_k) - \alpha\mathbf{h}_k^T \nabla f(\mathbf{x}_k) + \frac{1}{2}L\alpha^2\|\mathbf{h}_k\|^2.$$

$f(x_{k+1}) \le f(x_k) - \alpha h^T \nabla f(x_k) + \frac{1}{2} L\alpha^2 \|h\|^2$

The minimum of the left hand side with respect to α is $f(\mathbf{x}_{k+1})$. The right hand side is a quadratic function of α. The inequality above still holds if minimizing both sides with respect to α:

$$f(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|^2,$$

thus

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}_*) \le f(\mathbf{x}_k) - f(\mathbf{x}_*) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|^2. \tag{2.10}$$

Similarly, by Multivariate Quadratic Taylor's Theorem and lower bound assumption $\mu I \le \nabla^2 f(\mathbf{x})$, we get

$\frac{1}{2}(x - x_k)^T \nabla^2 f(z)(x - x_k)$

$$f(\mathbf{x}) \ge f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T(\mathbf{x} - \mathbf{x}_k) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_k\|^2.$$

Minimizing both sides w.r.t. $\mathbf{x}$, we get $f(x) \ge f(x_k) - \frac{1}{2\mu}\|\nabla f\|^2$

$$f(\mathbf{x}_*) \ge f(\mathbf{x}_k) - \frac{1}{2\mu}\|\nabla f(\mathbf{x}_k)\|^2,$$

thus $-\|\nabla f(\mathbf{x}_k)\|^2 \le 2\mu[f(\mathbf{x}_*) - f(\mathbf{x}_k)]$. Plugging it into (2.10), we get the convergence rate. $\square$

$f(x_{k+1}) - f(x_*) \le \left(1 - \frac{\mu}{L}\right)[f(x_k) - f(x_*)]$

Theorem 2.10. *The assumptions that a function $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex and its gradient $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L are equivalent to the following for any $\mathbf{x}, \mathbf{y}$:*

1.

$$0 \leq f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2. \tag{2.5}$$

2.

$$f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq f(\mathbf{y}). \tag{2.6}$$

3.

$$\frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle. \tag{2.7}$$

assumptions $\Leftarrow$

4.

$$0 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq L\|\mathbf{x} - \mathbf{y}\|^2. \tag{2.8}$$

Proof. First of all, assume a function $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex and its gradient $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L, then (2.5) holds because of the first order condition of convexity (Lemma 1.1) and descent lemma (Lemma 2.1).

Second, assume (2.5) holds, then (2.5) implies $\phi(\mathbf{x}) = f(\mathbf{x}) - \langle \nabla f(\mathbf{x}_0), \mathbf{x} \rangle$ satisfies

$$① \quad 0 \leq \phi(\mathbf{x}) - \phi(\mathbf{y}) - \langle \nabla \phi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

and

$$
\begin{aligned}
② \qquad & \phi(\mathbf{x}) \leq \phi(\mathbf{y}) + \langle \nabla \phi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2 \\
(|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\|\|\mathbf{b}\|) \qquad & \leq \phi(\mathbf{y}) + \|\nabla \phi(\mathbf{y})\|\|\mathbf{x} - \mathbf{y}\| + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2.
\end{aligned}
$$

$① \Leftrightarrow 0 \leq f(x) - \langle \nabla f(x_0), x \rangle - f(y) + \langle \nabla f(x_0), y \rangle - \langle \nabla f(x) - \nabla f(x_0), x - y \rangle$

$\Leftrightarrow 0 \leq f(x) - f(y) - \langle \nabla f(x), x - y \rangle$

$② \Leftrightarrow f(x) - \langle \nabla f(x_0), x \rangle \leq f(y) - \langle \nabla f(x_0), y \rangle + \langle \nabla f(x) - \nabla f(x_0), x - y \rangle + \frac{L}{2}\|x - y\|^2$

By Theorem 1.5, $\phi(\mathbf{x})$ is also convex. Moreover, $\nabla\phi(\mathbf{x}_0) = \mathbf{0}$, thus by Theorem 2.4, $\mathbf{x}_0$ is a global minimizer of $\nabla\phi(\mathbf{x})$. So we get

$$\begin{aligned}
\phi(\mathbf{x}_0) = \min_{\mathbf{x}} \phi(\mathbf{x}) &\leq \min_{\mathbf{x}} \left[\phi(\mathbf{y}) + \|\nabla\phi(\mathbf{y})\|\|\mathbf{x}-\mathbf{y}\| + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2\right] \\
&\leq \min_{r\geq 0} \left[\phi(\mathbf{y}) + \|\nabla\phi(\mathbf{y})\| r + \frac{L}{2} r^2\right] \\
&= \phi(\mathbf{y}) - \frac{1}{2L}\|\nabla\phi(\mathbf{y})\|^2.
\end{aligned}$$

Thus $\phi(\mathbf{x}_0) \leq \phi(\mathbf{y}) - \frac{1}{2L}\|\nabla\phi(\mathbf{y})\|^2$ implies

$$f(\mathbf{x}_0) - \langle \nabla f(\mathbf{x}_0), \mathbf{x}_0\rangle \leq f(\mathbf{y}) - \langle \nabla f(\mathbf{x}_0), \mathbf{y}\rangle - \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}_0)\|^2.$$

Thus $\phi(\mathbf{x}_0) \leq \phi(\mathbf{y}) - \frac{1}{2L}\|\nabla\phi(\mathbf{y})\|^2$ implies

$$f(\mathbf{x}_0) - \langle \nabla f(\mathbf{x}_0), \mathbf{x}_0\rangle \leq f(\mathbf{y}) - \langle \nabla f(\mathbf{x}_0), \mathbf{y}\rangle - \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}_0)\|^2.$$

Since $\mathbf{x}_0, \mathbf{y}$ are arbitrary, we can also write is as

$$f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{x}\rangle \leq f(\mathbf{y}) - \langle \nabla f(\mathbf{x}), \mathbf{y}\rangle - \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2,$$

which implies

$$f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \leq f(\mathbf{y}).$$

Switching $\mathbf{x}$ and $\mathbf{y}$, we get

$$f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|^2 \leq f(\mathbf{x}),$$

and adding two we get (2.6)

Third, assume (2.7) holds, then $\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0$ implies the convexity by Lemma 1.1, and Cauchy-Schwartz inequality gives Lipschitz continuity by

$$\frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|\|\mathbf{x} - \mathbf{y}\|.$$

Finally, we want to show (2.8) is equivalent to (2.5). Assume (2.5) holds, we get (2.8) by adding the following two:

$$0 \leq f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2,$$

$$0 \leq f(\mathbf{x}) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

Assume (2.8) holds, we get (2.5) by Fundamental Theorem of Calculus on $g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$: $g'(t) = \langle \nabla f[x+t(y-x)], y-x \rangle$

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt$$

$$\begin{aligned}
\Rightarrow f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle &= \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt \\
&= \int_0^1 \frac{1}{t}\langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), t(\mathbf{y} - \mathbf{x}) \rangle dt \\
(2.8) \quad &\leq \int_0^1 Lt\|\mathbf{y} - \mathbf{x}\|^2 dt = \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2.
\end{aligned}$$

□

$0 \leq \langle \nabla f(x) - \nabla f(y), x-y \rangle \leq L\|x-y\|^2$