

1 Strong convexity and smoothness constants

Let $f : \Omega \rightarrow \mathbb{R}$ be smooth on $\Omega \subseteq \mathbb{R}^n$. Recall the relationship between minimization and the infimum; the analogous concept for maximization is the supremum $\sup_{x \in \Omega} h(x) = -\inf_{x \in \Omega} (-h(x))$. Define

$$\mu^* = \inf_{x \in \Omega} \lambda_{\min}(\nabla^2 f(x)), \quad L^* = \sup_{x \in \Omega} \|\nabla^2 f(x)\|_{\text{op}}, \quad L_2^* = \sup_{u \neq w} \frac{\|\nabla^2 f(u) - \nabla^2 f(w)\|_{\text{op}}}{\|u - w\|}.$$

Then f is μ -strongly convex if and only if $0 < \mu \leq \mu^*$, ∇f is L -Lipschitz if and only if $L \geq L^*$, and $\nabla^2 f$ is L_2 -Lipschitz if and only if $L_2 \geq L_2^*$. When $n = 1$, $L_2^* = \sup_{x \in \Omega} |f'''(x)|$. For each function below:

- (a) compute μ^* , L^* , and L_2^* ;
- (b) determine whether f is strongly convex, whether ∇f is Lipschitz, and whether $\nabla^2 f$ is Lipschitz.

Rigorously justify each conclusion.

Worked example 1.1

Consider $f(x) = \frac{1}{2}x^2 + \cos x$ for $x \in \mathbb{R}$.

Solution. (a) Here $f''(x) = 1 - \cos x$ and $f'''(x) = \sin x$. Then $\mu^* = \inf_x (1 - \cos x) = 0$, attained at $x = 0$. Next, $L^* = \sup_x (1 - \cos x) = 2$, attained at $x = \pi$. Finally, $L_2^* = \sup_x |\sin x| = 1$. (b) Since $\mu^* = 0$, f is not strongly convex. Since L^* and L_2^* are finite, ∇f and $\nabla^2 f$ are Lipschitz.

1. $f(x) = 2x_1^2 + 2x_1x_2 + 2x_2^2$ for $x \in \mathbb{R}^2$.
2. $f(x) = \frac{1}{2}(x_1 - x_2)^2$ for $x \in \mathbb{R}^2$.
3. $f(x) = x^4$ for $x \in \mathbb{R}$.
4. $f(x) = e^x$ for $x \in (-1, 1)$.
5. $f(x) = \sin x$ for $x \in \mathbb{R}$.
6. $f(x) = x^2 + x^4$ for $x \in \mathbb{R}$.
7. $f(x) = -\log x$ for $x \in (0, 1)$.
8. **(G)** $f(x) = -\log(1 - \|x\|^2)$ for $x \in \mathbb{R}^2$ with $\|x\| < 1$.

2 Minimizing quadratics

Let $H = H^T \in \mathbb{R}^{n \times n}$ and $g \in \mathbb{R}^n$, and consider

$$q(x) = g^T x + \frac{1}{2} x^T H x.$$

For each pair (g, H) below, either find all global minimizers of q , or prove that q is unbounded below by stating a path $x(t)$ such that $q(x(t)) \rightarrow -\infty$ as $t \rightarrow \infty$. Rigorously justify each conclusion.

1. $g = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $H = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

2. $g = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, $H = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$.

3. $g = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $H = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$.

4. $g = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}$, $H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

5. $g = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$.

6. **(G)** General $g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ and $H = \text{diag}(\lambda_1, \lambda_2)$ with $\lambda_1 \geq 0$ and $\lambda_2 < 0$.

7. **(G)** General $g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ and $H = \text{diag}(\lambda_1, 0)$ with $\lambda_1 > 0$.

3 Exact solution of the cubic-regularization subproblem

Let $g \in \mathbb{R}^n$, $M > 0$, and $d \in \mathbb{R}^n$ with $d_1 \leq \dots \leq d_n$, and write $D = \text{diag}(d)$. The cubic-regularization subproblem minimizes

$$c_x(v) = g^T v + \frac{1}{2} v^T D v + \frac{M}{6} \|v\|^3. \quad (\text{CR})$$

A vector v is globally optimal if and only if some $\lambda \geq 0$ satisfies

$$(D + \lambda I)v = -g, \quad D + \lambda I \succeq 0, \quad \lambda = \frac{M}{2} \|v\|. \quad (\text{C})$$

The subproblem is solved by bisection on the secular equation $\psi(\lambda) = 0$, where

$$\lambda_0 = \max\{0, -d_1\}, \quad \psi(\lambda) = \sum_{i=1}^n \frac{g_i^2}{(d_i + \lambda)^2} - \frac{4\lambda^2}{M^2} \quad \text{for } \lambda > \lambda_0.$$

A root $\lambda_\star > \lambda_0$ gives the global minimizer $v = -(D + \lambda_\star I)^{-1}g$.

1. **Reduction to diagonal form.** Suppose an oracle $v = \text{SolveCR}(g, d, M)$ returns a global minimizer of (CR). Explain how a single call to $\text{SolveCR}(\cdot, \cdot, \cdot)$ globally solves

$$\min_{v \in \mathbb{R}^n} f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v + \frac{M}{6} \|v\|^3.$$

2. **Monotonicity.** Prove that ψ is strictly decreasing on (λ_0, ∞) .
3. **Easy case.** Prove that if $g_1 \neq 0$, then $\psi(\lambda_\star) = 0$ for some $\lambda_\star > \lambda_0$.
4. **Hard case.** Explain why the bisection method can fail when $\lambda_0 > 0$ and $g_1 = 0$.
5. **(G) Bisection.** It can be shown that ψ is strictly decreasing on (λ_0, ∞) . Suppose that $\psi(\lambda_\star) = 0$ for some $\lambda_\star > \lambda_0$. Find a choice of λ_1 and λ_2 for which $\psi(\lambda_1) < 0 < \psi(\lambda_2)$, and explain how to use bisection to find $\lambda_\star$.
6. **(G) Hard-case minimizers.** Suppose that

$$d_1 < \min\{0, d_2\}, \quad g_1 = 0, \quad \bar{v} = - \left(0, \frac{g_2}{d_2 - d_1}, \dots, \frac{g_n}{d_n - d_1} \right)^T, \quad \|\bar{v}\| \leq -\frac{2d_1}{M}.$$

Prove that $\bar{v} \pm \tau e_1$ with $\tau = \sqrt{4d_1^2/M^2 - \|\bar{v}\|^2}$ are global minimizers of (CR). The boundary case $\|\bar{v}\| = 2|d_1|/M$ gives $\tau = 0$.

4 Newton convergence analysis

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable, and assume that for all $x, u, w \in \mathbb{R}^n$,

$$\mu I \preceq \nabla^2 f(x) \preceq L_1 I, \quad \|\nabla^2 f(u) - \nabla^2 f(w)\|_{\text{op}} \leq L_2 \|u - w\|, \quad (\text{A})$$

with $0 < \mu \leq L_1$ and $L_2 > 0$. Then f has a unique minimizer x^* , and $f^* = f(x^*)$. At an iterate x , define the Newton direction and the Newton decrement

$$v = -\nabla^2 f(x)^{-1} \nabla f(x), \quad \delta^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) = -\nabla f(x)^T v = v^T \nabla^2 f(x) v. \quad (\text{N})$$

Backtracking Newton sets $x_+ = x + \alpha v$, where α is the first element of $\{1, c_2, c_2^2, \dots\}$ satisfying

$$f(x + \alpha v) \leq f(x) + c_1 \alpha \nabla f(x)^T v, \quad (\text{Armijo})$$

with $0 < c_1 < 1/2$ and $0 < c_2 < 1$.

1. **Gradient remainder.** Starting from

$$\nabla f(x + d) - \nabla f(x) = \int_0^1 \nabla^2 f(x + td) d \, dt,$$

prove that for all $x, d \in \mathbb{R}^n$,

$$\|\nabla f(x + d) - \nabla f(x) - \nabla^2 f(x) d\| \leq \frac{L_2}{2} \|d\|^2.$$

2. **Function remainder.** Derive the identity

$$f(x + d) = f(x) + \nabla f(x)^T d + \int_0^1 (1 - t) d^T \nabla^2 f(x + td) d \, dt,$$

and use it to prove that for all $x, d \in \mathbb{R}^n$,

$$\left| f(x + d) - f(x) - \nabla f(x)^T d - \frac{1}{2} d^T \nabla^2 f(x) d \right| \leq \frac{L_2}{6} \|d\|^3.$$

3. **Local quadratic convergence.** Starting from

$$e = x - x^*, \quad e_+ = x + v - x^*, \quad \nabla f(x) = \int_0^1 \nabla^2 f(x^* + te) e \, dt,$$

prove that $\|e_+\| \leq \frac{L_2}{2\mu} \|e\|^2$. Explain why one full Newton step minimizes a strictly convex quadratic exactly.

4. **Iteration complexity.** Let δ_k denote the Newton decrement at x_k . Suppose that constants $\eta, \gamma > 0$ satisfy, for every k ,

- if $\delta_k > \eta$, then $f(x_{k+1}) \leq f(x_k) - \gamma$;
- if $\delta_k \leq \eta$, then $\delta_{k+1} \leq \delta_k^2 / (2\eta)$.

Prove that for $0 < \epsilon < \eta$, backtracking Newton reaches $\delta_k \leq \epsilon$ within

$$\left\lceil \frac{f(x_0) - f^*}{\gamma} + \log_2 \log_2 \left(\frac{2\eta}{\epsilon} \right) \right\rceil$$

iterations.

5. **(G) Full-step acceptance.** Starting from

$$f(x + v) \leq f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v + \frac{L_2}{6} \|v\|^3,$$

prove that

$$f(x + v) \leq f(x) - \frac{1}{2} \delta^2 + \frac{L_2}{6\mu^{3/2}} \delta^3,$$

and conclude that $\alpha = 1$ passes (Armijo) whenever $\delta \leq 3(1 - 2c_1)\mu^{3/2}/L_2$.

6. **(G) Gradient after a full step.** Starting from

$$x_+ = x + v, \quad \|\nabla f(x + d) - \nabla f(x) - \nabla^2 f(x)d\| \leq \frac{L_2}{2} \|d\|^2 \quad \text{for all } d,$$

prove that $\|\nabla f(x_+)\| \leq \frac{L_2}{2\mu} \delta^2$.

7. **(G) Decrement contraction.** Starting from

$$\delta_+^2 = \nabla f(x_+)^T \nabla^2 f(x_+)^{-1} \nabla f(x_+), \quad \|\nabla f(x_+)\| \leq \frac{L_2}{2\mu} \delta^2,$$

prove that $\delta_+ \leq \frac{L_2}{2\mu^{3/2}} \delta^2$, and conclude that $\delta_+ \leq \delta^2/(2\eta)$ for every $0 < \eta \leq \mu^{3/2}/L_2$.

5 Cubic-regularized Newton convergence analysis

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable with $f_{\inf} = \inf_x f(x) > -\infty$, and assume that for all $u, w \in \mathbb{R}^n$,

$$\|\nabla^2 f(u) - \nabla^2 f(w)\|_{\text{op}} \leq L_2 \|u - w\|. \quad (\text{A})$$

At an iterate x , cubic-regularized Newton sets $x_+ = x + v$, where v globally minimizes the model

$$c_x(v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v + \frac{L_2}{6} \|v\|^3. \quad (\text{CN})$$

The minimizer satisfies the first- and second-order necessary conditions

$$\nabla f(x) + \nabla^2 f(x)v + \frac{L_2}{2} \|v\| v = 0, \quad \nabla^2 f(x) + \frac{L_2}{2} \|v\| \left(I + \frac{vv^T}{v^T v} \right) \succeq 0, \quad (\text{O})$$

where $vv^T/v^T v$ is read as 0 when $v = 0$.

1. **Decrease.** Starting from

$$f(x_+) \leq c_x(v), \quad c_x(v) \leq c_x(-v),$$

prove that $\nabla f(x)^T v \leq 0$ and

$$f(x_+) \leq f(x) - \frac{L_2}{12} \|v\|^3.$$

2. **Gradient.** Starting from

$$\|\nabla f(x+v) - \nabla f(x) - \nabla^2 f(x)v\| \leq \frac{L_2}{2} \|v\|^2,$$

prove that $\|\nabla f(x_+)\| \leq L_2 \|v\|^2$.

3. **Curvature.** Using (A) and (O), prove that $\nabla^2 f(x) + L_2 \|v\| I \succeq 0$ and

$$\nabla^2 f(x_+) \succeq -2L_2 \|v\| I.$$

4. **Escaping saddle points.** Starting from

$$\nabla f(x) = 0, \quad \lambda_{\min}(\nabla^2 f(x)) < 0, \quad f(x_+) \leq f(x) - \frac{L_2}{12} \|v\|^3,$$

prove that $v \neq 0$ and $f(x_+) < f(x)$. Explain why backtracking Newton cannot make progress from the same point.

5. **(G) Iteration complexity.** Suppose that the iterates $x_{i+1} = x_i + v_i$ satisfy, for every $i \geq 0$,

$$f(x_{i+1}) \leq f(x_i) - \frac{L_2}{12} \|v_i\|^3, \quad \|\nabla f(x_{i+1})\| \leq L_2 \|v_i\|^2, \quad \nabla^2 f(x_{i+1}) \succeq -2L_2 \|v_i\| I.$$

Let $D = f(x_0) - f_{\inf}$. Prove that for every $k \geq 1$, some $x_{\text{best}} \in \{x_1, \dots, x_k\}$ satisfies

$$\|\nabla f(x_{\text{best}})\| \leq \left(\frac{12L_2^{1/2} D}{k} \right)^{2/3}, \quad \lambda_{\min}(\nabla^2 f(x_{\text{best}})) \geq -2 \left(\frac{12L_2^2 D}{k} \right)^{1/3}.$$