Running Notes

Daniel Henderson

May 24, 2025

1 Preliminaries

1.1 Notation

We use the following notation

- $\|\cdot\|$ denotes the usual ℓ_2 norm for vectors $\boldsymbol{x}$ in $\mathbb{R}^n$ and $p = 2$ norm for matrices in $\mathbb{R}^{n\times m}$. i.e.,

$$\|x\| := \left(\sum_i x_i^2\right)^{1/2}$$
$$\|A\| := (\lambda_{\max}(A^\top A))^{1/2} = \max(\sigma(A))$$

- $\sigma(A) := \{\text{singular values of } A\}$.
- $A \in \mathbb{R}^{n\times n} \implies \sigma(A) = \{\text{eigenvalues of } A \text{ (i.e. spectrum)}\}$
- $\sigma_{\max}(A) := \max(\sigma(A))$ and $\sigma_{\min}(A) := \min(\sigma(A))$.
- $A \in \mathbb{R}^{n\times n} \implies \lambda_{\max}(A) := \sigma_{\max}(A)$ and $\lambda_{\min}(A) = \sigma_{\min}(A)$
- $\langle\cdot,\cdot\rangle$ denotes the usual inner product on $\mathbb{R}^n$, i.e.,

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i = \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos(\theta)$$

  where θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

- $\mathcal{B}_r(\boldsymbol{x}) := \{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < r\}$ is the open ball of radius r centered at $\boldsymbol{x} \in \mathbb{R}^n$.

1.2 Assumptions

We assume the following assumptions

- Let $\Omega \subset \mathbb{R}^n$ be a bounded open subset with a smooth boundary $\partial\Omega$, which we assume is a C^2-manifold. Additionally, we often consider Ω and $\partial\Omega$ to be the $n-1$-dimensional hypersurface of a n-dimensional ball.
- Let be a function $f \in C^\infty(\Omega; \mathbb{R})$ is a continous and infinitely-differentiable mapping from Ω to $\mathbb{R}$. In general, f is nonlinear and nonconvex on Ω.
- We assume C^2 regularity of the boundary $\partial\Omega$ ao that the derivatives of f may be extended to $\partial\Omega$.
- The typefont $\boldsymbol{x}$ will denote a vector in $\mathbb{R}^n$.

1.3 Definitions

Using the above notation, we assume our assumptions hold, and we define the following

Definition 1. f is **L-Lipschitz** if $\forall\ \boldsymbol{x_1}, \boldsymbol{x_2}$

$$\exists\ L \geq 0\ :\ \|f(\boldsymbol{x_1}) - f(\boldsymbol{x_2})\| \leq L\|\boldsymbol{x_1} - \boldsymbol{x_2}\|$$

Definition 2. f has **ℓ-Lipschitz gradient**, or, f is **ℓ-smooth** if $\forall\ \boldsymbol{x_1}, \boldsymbol{x_2}$

$$\exists\ \ell \geq 0\ :\ \|\nabla f(\boldsymbol{x_1}) - \nabla f(\boldsymbol{x_2})\| \leq \ell\|\boldsymbol{x_1} - \boldsymbol{x_2}\|$$

Definition 3. f has **ρ-Lipschitz Hessian** if $\forall\ \boldsymbol{x_1}, \boldsymbol{x_2}$

$$\exists\ \rho \geq 0\ :\ \|\nabla^2 f(\boldsymbol{x_1}) - \nabla^2 f(\boldsymbol{x_2})\| \leq \rho\|\boldsymbol{x_1} - \boldsymbol{x_2}\|$$

Definition 4. f is **convex** if $\forall\ \boldsymbol{x_1}, \boldsymbol{x_2}$

$$\begin{aligned} f(\boldsymbol{x_2}) &\geq f(\boldsymbol{x_1}) + \langle \boldsymbol{x_2} - \boldsymbol{x_1}, \nabla f(\boldsymbol{x_1})\rangle \\ &= f(\boldsymbol{x_1}) + \nabla f(\boldsymbol{x_1})^T(\boldsymbol{x_2} - \boldsymbol{x_1}) \end{aligned}$$

Definition 5. f is **strictly convex** if

$$\begin{aligned} \exists\, \mu > 0\ :\ \nabla^2 f \succeq \mu I \\ \iff \lambda_{\min}(\nabla^2 f) \geq \mu > 0 \end{aligned}$$

Definition 6. f is α**-strongly convex** if $\forall\ \boldsymbol{x_1}, \boldsymbol{x_2} \exists\ \alpha > 0$ s.t.

$$\begin{aligned} & f(\boldsymbol{x_2}) \geq f(\boldsymbol{x_1}) + \langle \nabla f(\boldsymbol{x_1}), \boldsymbol{x_2} - \boldsymbol{x_1} \rangle + \frac{\alpha}{2} \|\boldsymbol{x_2} - \boldsymbol{x_1}\|^2 \\ & \iff \lambda_{\min}(\nabla^2 f(\boldsymbol{x})) \geq -\alpha. \end{aligned}$$

Definition 7. $\boldsymbol{x^*}$ is a **first-order stationary point** if $\|\nabla f(x^*)\| = 0$.

Definition 8. $\boldsymbol{x^*}$ is an ϵ**-first-order stationary point** if $\|\nabla f(x^*)\| \leq \epsilon$.

Definition 9. $\boldsymbol{x}^* \in \mathbb{R}^n$ is a **second-order stationary point** if $\|\nabla f(x^*)\| = 0$ and $\nabla^2 f(x^*) \succeq 0$.

Definition 10. if f has ρ-Lipschitz Hessian, $\boldsymbol{x}^* \in \mathbb{R}^n$ is a ϵ**-second-order stationary point** if

$$\|\nabla f(x^*)\| \leq \epsilon \text{ and } \nabla^2 f(x^*) \succeq -\sqrt{\rho\epsilon}$$

Remark. Note that the Hessian is not required to be positive definite, but it is required to have a small eigenvalue.

Definition 11. We consider the general form of our **unconstrained optimization problem** to be

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{x} \in K \subseteq \mathbb{R}^n \tag{1}$$

where K is a compact set in $\mathbb{R}^n$. We denote the **optimal solution** and **optimal value** of the optimization problem

$$\begin{aligned} \boldsymbol{x}^* &= \arg\min_{\boldsymbol{x} \in K} f(\boldsymbol{x}) \\ \boldsymbol{f}^* &= \min_{\boldsymbol{x} \in K} f(\boldsymbol{x}) \end{aligned}$$

where $\boldsymbol{x}^*$ satisfies the first-order optimality condition, i.e., $\nabla f(\boldsymbol{x}^*) = 0$.

Definition 12. Let the **gradient flow** of f be a solution to the dynamical system defined as

$$\boldsymbol{\gamma}'(t) = -\nabla f(\boldsymbol{\gamma}(t))$$

where the evolution of our phase space is driven by the negative gradient of f. A **gradient flow line** of f is an integral curve $\boldsymbol{\gamma} : [0, t_f] \to \Omega$ satisfying the above evolutionary system (ordinary-differential equation) subject to $\boldsymbol{\gamma}(0) = \boldsymbol{x_0}$.

We aim to classify the phase space of the *gradient flow* of f on Ω. First we notice that for and any critical point $\boldsymbol{x}^*$,

$$\begin{aligned}
&\boldsymbol{\gamma}(t) = \boldsymbol{x}^* \ \forall\ t \in [0, t_f] \\
&\implies \quad \boldsymbol{\gamma}'(t) = \mathbf{0} \ \text{ and } \ -\nabla f(\boldsymbol{\gamma}(t)) = -\nabla f(\boldsymbol{x}^*) = \mathbf{0} \quad \because\ \boldsymbol{x}^* \ \text{ is a critical point} \\
&\quad \therefore \quad \boldsymbol{\gamma}(t) = -\nabla f(\boldsymbol{\gamma}(t)) \ \forall\ t \in [0, t_f]
\end{aligned}$$

Consequently, by the uniqueness of solutions for ordinary differential equations, if any flow line contains a *first-order* critical point $\boldsymbol{x}^*$, it must be a constant flow line.

Lemma A. *The function $f : \omega \to \mathbb{R}$ is nonincreasing along any flow-line $\boldsymbol{\gamma}(t)$ and strictly decreasing along flow lines not containing a critical point x^*.*

Proof. Let $\boldsymbol{\gamma} : [0, t_f] \to \Omega$ be a flow line. Consider the composition $f \circ \boldsymbol{\gamma} : [0, t_f] \to \mathbb{R}$, its derivative is

$$\begin{aligned}
\frac{d}{dt}\big(f(\boldsymbol{\gamma}(t))\big) &= \langle \nabla_{\boldsymbol{\gamma(t)}}(f), \frac{d\boldsymbol{\gamma}}{dt} \rangle \\
&= \langle \nabla_{\boldsymbol{\gamma(t)}}(f), -\nabla_{\gamma(t)}(f) \rangle \\
&= -\langle \nabla_{\boldsymbol{\gamma(t)}}(f), \nabla_{\gamma(t)}(f) \rangle \\
&= -|\nabla_{\boldsymbol{\gamma(t)}}(f)|^2 \\
&\leq 0
\end{aligned}$$

Therefore, $f'(\boldsymbol{\gamma(t)}) = 0$ iff $\boldsymbol{\gamma}(t)$ is on a critical point of f. In particular, if $\boldsymbol{\gamma}(t)$ does not contain in its image a critical point of f, then the above inequality implies that f is strictly decreasing along the integral curve $\boldsymbol{\gamma(t)}$. $\square$

theorem. For all $\boldsymbol{x}$ in the closed manifold $\overline{\Omega}$, there exists uniquely $\boldsymbol{\gamma_x}(t) : \mathbb{R} \to \overline{\Omega}$ such that $\boldsymbol{\gamma_x}(0) = \boldsymbol{x}$ and the limits

$$\lim_{t \to -\infty} \boldsymbol{\gamma_x}(t) \ \text{ and } \ \lim_{t \to \infty} \boldsymbol{\gamma_x}(t)$$

exist and converge to *critical-points* of f.

Now it may be shown that the flow map operator $T : \overline{\Omega} \times \mathbb{R} \to \overline{\Omega}$ defined as $T(\boldsymbol{x}, t) := \boldsymbol{\gamma_x}(t)$. This implies theoretically that we may perform analysis of our phase space as constructed by a smooth union of integral curves. Consequently, by our previous lemma, if the flow map T contains a critical point $\boldsymbol{x}^*$ then it ought to be that $T(\boldsymbol{x}^*, t) \equiv \text{ const } \forall t$, otherwise, T is descending for all of time.

Definition 13. A **scheme** for solving a general form *unconstrained optimization problem* is a one-parameter family of iteration operators:

$$T_h : \mathbb{R}^n \to \mathbb{R}^n \text{ where } \boldsymbol{x}_{k+1} = T_h(\boldsymbol{x}_k),\ h \in (0, h_0]$$

where h_0 is a constant and h is the step size. The scheme is well-defined such that the triplet $(\boldsymbol{x}_0, h, T_h)$ satisfy

1. *Consistency*: $\forall \boldsymbol{x} \in K$
$$(T_h(\boldsymbol{x}) - \boldsymbol{x})\, h^{-1} + \nabla f(\boldsymbol{x}) \to 0 \text{ as } h \to 0.$$
implying that a single step approximates the continous gradient flow w/ local error $\mathcal{O}(h^{p+1})$ where p is the global order of the scheme.

2. *Stability*: $\exists\ c > 0, h_0 > 0\ :\ \forall h \in (0, h_0]$, and for all $\boldsymbol{x_1}, \boldsymbol{x_2}$ in a *neighborhood* $N \subset K$ around an *optimal solution* $\boldsymbol{x}^*$
$$\|T_h(\boldsymbol{x_1}) - T_h(\boldsymbol{x_2})\| \leq (1 - ch)\|\boldsymbol{x_1} - \boldsymbol{x_2}\|$$
where c is a constant that depends on the scheme and h is the step size. Or, equivalently, the scheme is *stable* if $\exists\ c > 0, h_0 > 0\ :\ \forall h \in (0, h_0]$, and for all $\boldsymbol{x}$ in a *neighborhood* about $\boldsymbol{x}^*$, each step results in a strict decrease of by atleast a factor of $1 - ch$, i.e.,
$$\|J(T_h(\boldsymbol{x}))\| \leq (1 - ch)$$
where $J(T_h(\boldsymbol{x}))$ is the Jacobian of the scheme.

3. *Convergence*: $\forall\ \boldsymbol{x_0} \in N \subset K$, s.t. N is some neighborhood around a strict minimizer x^*. and $\forall \epsilon > 0$
$$\exists\ K \in \mathbb{N}\ :\ \forall k\ > K, \boldsymbol{x}_k \in N \text{ and } d(\boldsymbol{x}_k, \boldsymbol{x}^*) \leq \epsilon.$$

Definition 14. $\boldsymbol{x}^*$ is **non-degenerate** if $\nabla^2 f(x^*)$ is non-singular.

Definition 15. The **level set** of f at c is the set of points $\boldsymbol{x} \in \Omega$ such that $f(\boldsymbol{x}) = c$, i.e.,

$$L_c = \{\boldsymbol{x} \in \Omega : f(\boldsymbol{x}) = c\}$$

The level set L_c is a smooth manifold with boundary ∂L_c. The **sublevel** and **superlevel** sets of

f at c are the sets of points $\boldsymbol{x} \in \Omega$ such that $f(\boldsymbol{x}) \leq c$ and $f(\boldsymbol{x}) \geq c$, respectively, i.e.,

$$L_c^- = \{\boldsymbol{x} \in \Omega : f(\boldsymbol{x}) \leq c\}$$
$$L_c^+ = \{\boldsymbol{x} \in \Omega : f(\boldsymbol{x}) \geq c\}.$$

The sublevel set L_c^- is a smooth manifold with boundary ∂L_c^- and the superlevel set L_c^+ is a smooth manifold with boundary ∂L_c^+.

theorem. For a Morse function f on Ω, the gradient of f is either zero or orthogonal to the tangent space of the level set L_c at $\boldsymbol{x} \in L_c$.

The above theorem implies that at a stationary point $\boldsymbol{x}^*$, a level set $L_{\boldsymbol{x}^*}$ is reduced to a single point when $\boldsymbol{x}^*$ is a local minimum or maximum. Otherwise, the level set may have a singularity such as a self-intersection or a cusp.

2 Overview of Optimization Schemes

2.1 Gradient Descent (GD)

Continous Model

TODO: Explain the gradient descent phase space and the level set equations for the continous model.

Scheme

The **gradient descent** line search scheme is

$$T_h(\boldsymbol{x}) = \boldsymbol{x} - h\nabla f(\boldsymbol{x})$$

first-order ($p = 1$) and contractive when $\nabla^2 f \succeq \mu I \succeq 0$.

2.2 Newton's Method (NM)

Continous Model

TODO: Explain the Newton's method phase space and the level set equations for the continous model.

Scheme

The **Newton's method** line search scheme is

$$T_h(\boldsymbol{x}) = \boldsymbol{x} - h\nabla^2 f(\boldsymbol{x})^{-1}\nabla f(\boldsymbol{x})$$

second-order ($p = 2$) and contractive when $\nabla^2 f \succeq \mu I \succeq 0$.

2.3 Trust Region Methods (TR)

Continous Model

TODO: Explain the trust region method phase space and the level set equations for the continous model.

Scheme

The **trust region method** line search scheme is

$$T_h(\boldsymbol{x}) = \boldsymbol{x} + \arg\min_{\boldsymbol{\tau}} m_{\boldsymbol{x}}(\boldsymbol{\tau})$$

where $m_{\boldsymbol{x}}(\boldsymbol{\tau}) = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{\tau}\rangle + \frac{1}{2}\langle \boldsymbol{\tau}, \nabla^2 f(\boldsymbol{x})\boldsymbol{\tau}\rangle$ is the quadratic approximation of f at $\boldsymbol{x}$ and $\|\boldsymbol{\tau}\| \leq \Delta$ is the trust region constraint.

2.4 Quasi-Newton Methods (QN)

Continous Model

TODO : Explain the quasi-newton method phase space and the level set equations for the continous model.

Scheme

The **quasi-newton method** line search scheme is

$$T_h(\boldsymbol{x}) = \boldsymbol{x} - hB\nabla f(\boldsymbol{x})$$

where $B \approx \nabla^2 f^{-1}(\boldsymbol{x})$ is a positive-definite approximation of the Hessian. The quasi-newton method is a first-order ($p = 1$) scheme and contractive when $\nabla^2 f \succeq \mu I \succeq 0$.

3 Theory

3.1 Morse Theory

Note that Ω is a bounded subset of $\mathbb{R}^n$, so its closure $\overline{\Omega} = \Omega \cup \partial\Omega$ is a compact subset of $\mathbb{R}^n$, by the Heine-Borel Theorem. Also, the boundary $\partial\Omega$ is sufficiently smooth, so we can apply the theory of smooth manifolds. The closure $\overline{\Omega}$ is a compact subset of $\mathbb{R}^n$ and is a smooth manifold with boundary $\partial\Omega$. The interior Ω is an open subset of $\mathbb{R}^n$ and is a smooth manifold.

Definition 16. A point $\boldsymbol{x}^* \in \Omega$ is a **critical point** of f if the differentiable map $df_p : T_p\Omega \to \mathbb{R}$ is zero. (Here $T_p\Omega$ is a tangent space of the Manifold M at p.) The set of critical points of f is denoted by $\mathrm{crit}(f)$.

Definition 17. A point $\boldsymbol{x}^* \in \Omega$ is a **non-degenerate critical point** of f if the Hessian $H_p f$ is non-singular.

Definition 18. The **index** of a *non-degenerate critical point* $\boldsymbol{x}^*$ is defined to be the dimension of the negative eigenspace of the Hessian $H_p f$.

- local minima at $\boldsymbol{x}^*$ have index 0.
- local maxima at $\boldsymbol{x}^*$ have index n.
- saddle points at $\boldsymbol{x}^*$ have index k where $0 < k < n$.

We reserve the integers $c_0, c_1, \ldots, c_i, \ldots, c_n$ to denote the number of critical points of index i.

Remark. For each objective function f we are interested in determining the critical points of f

Remark. The **Morse function** is a smooth function $f : \Omega \to \mathbb{R}$ such that all critical points of f are non-degenerate.

theorem. Let f be a Morse function on Ω, then the Euler characteristic of Ω is given by

$$\chi(\Omega) = \sum_{i=0}^{n} (-1)^i c_i$$

where c_i is the number of critical points of index i.

Remark. The Euler characteristic $\chi(\Omega)$ is a topological invariant of the manifold Ω and is independent of the choice of Morse function f. The Euler characteristic is a measure of the "shape" of the manifold and can be used to distinguish between different topological spaces. The Euler characteristic may be defined by the alternating sum of the Betti numbers b_i of the manifold Ω

$$\chi(\Omega) = \sum_{i=0}^{n} (-1)^i b_i$$

where b_i is the i-th Betti number of the manifold Ω.

theorem. (Sard's theorem) Let f be a Morse function on Ω, then the image $f(\text{crit}(f))$ has Lebesque measure zero in $\mathbb{R}$.

Remark. We state a particular instance of Sard's theorem for continous scalar-valued functions f, which was first proved by Anothony P. Morse in 1939. The theorem asserts that the image of the critical points of a Morse function is a set of measure zero in $\mathbb{R}$. This means that the critical points of a Morse function are "rare" in the sense that they do not form a dense subset of the manifold Ω. Consequently, selecting $\boldsymbol{x} \in \Omega$ at random will almost never yeild a critical point of f.

Remark. The property that $\boldsymbol{x^*} \in \Omega$ being a *critical point* of a Morse function f is not dependent of the metric of $\Omega \subset \mathbb{R}^n$ (and consequently, the norm induced by the metric)

3.2 Analysis of Gradient Descent (GD)

theorem. Assume f is ℓ-smooth and α-strongly convex and that $\epsilon > 0$. If we iterate the gradient descent *scheme* with $h = h_0 = \frac{1}{\ell}$ held fixed, i.e.,

$$T_h(\boldsymbol{x}_k) = \boldsymbol{x}_k - \frac{1}{\ell}\nabla f(\boldsymbol{x}_k),$$

then $d(x_k, x^*) \leq \epsilon$ for all $k > K$ where K is chosen to satisfy

$$\frac{2\ell}{\alpha} \cdot \log\left(\frac{d(x_0, x^*)}{\epsilon}\right) \leq K.$$

Remark. Under *ℓ-smoothness* and *α-strong convexity* assumptions in a neighborhood Ω about $\boldsymbol{x}^*$, it may be shown directly from the above theorem above that the *GD scheme* converges linearly to the optimal solution $\boldsymbol{x}^*$ at a rate of

$$\frac{d(\boldsymbol{x}_k, \boldsymbol{x}^*)}{d(\boldsymbol{x}_{k-1}, \boldsymbol{x}^*)} \leq 1 - \frac{\alpha}{\ell}$$

where $d(\boldsymbol{x}_k, \boldsymbol{x}^*)$ is the distance between the current iterate $\boldsymbol{x}_k$ and the optimal solution $\boldsymbol{x}^*$. The convergence rate is linear in the sense that the distance between the current iterate and the optimal solution decreases by a factor of $1 - \frac{\alpha}{\ell}$ at each iteration. (Ref: TODO)

Remark. Convergence to a first-order stationary point trivally implies convergence to a ϵ-first-order stationary point. Similarly, convergence to a second-order stationary point trivially implies convergence to a ϵ-second-order stationary point.

theorem. Assume f is ℓ-smooth, then for any $\epsilon > 0$, if we iterate the GD scheme with $h = h_0 = \frac{1}{\ell}$ held fixed starting from $\boldsymbol{x}_0 \in \Omega$ where Ω is a neighborhood of x^*, then the number of iterations K required to achieve the stopping condition $\|\nabla f(\boldsymbol{x}_k)\| \leq \epsilon$ is at most

$$\left\lceil \frac{\ell}{\epsilon^2}\left(f(\boldsymbol{x}_0) - f(\boldsymbol{x}^*)\right)\right\rceil$$

Remark. **TODO and Questions**

- State how we use theorems in when performing analysis from the results of our experimewts.
- What is the relationship between ℓ and α?
- In practice do we know how to compute ℓ and α?
- What is the relationship between ℓ and ρ?
- In practice do we know how to compute ℓ and ρ?

Remark. JinSaddle.pdf Section 3.1 - Strict Saddle Property