4.4. Some stochastic aspects in gradient algorithm. (not required for Exam).

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

1. Probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- Ω a set
- $\mathcal{F}$: σ-field
  - $\{\phi, \Omega\} \subseteq \mathcal{F}$
  - $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
  - $\{A_n\}_{n\geq 1} \subseteq \mathcal{F} \Rightarrow \bigcup_{n\geq 1} A_n \in \mathcal{F}$.
- $\mathbb{P}: \mathcal{F} \to [0,1]$
  - $\mathbb{P}[\Omega] = 1$. $\quad \mathbb{P}[\phi \cup \Omega] = \mathbb{P}[\phi] + \mathbb{P}[\Omega] \Rightarrow \mathbb{P}[\phi] = 0$.
  - $\{A_n\}_{n\geq 1} \subseteq \mathcal{F}$. s.t $A_i \cap A_j = \phi$. $\forall i \neq j$, then $\mathbb{P}(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mathbb{P}[A_n]$

2. Stochastic process. is a family of random variables, indexed by time. $t \in [0, \infty)$.

$X : \Omega \to \mathbb{R}$

i.e. $(X_t)_{t\in[0,\infty)}$.

$\mathbb{E}[|X|] := \int_\Omega |X(\omega)| \, \mathbb{P}(d\omega)$

— A stochastic process $X = (X_t)_{t\in[0,\infty)}$ is called a $\boxed{\text{martingale}}$.

if $\mathbb{E}[|X_t|] < +\infty$, $\forall t \in [0,\infty)$ and $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$, $\forall s \le t$.

$\{\omega : X_r(\omega) \le a\} \in \mathcal{F}_s$.

① $\mathcal{F}_s := \sigma(X_r : r \le s)$ is the smallest σ-field w.r.t. which X_r is measurable $\forall r \le s$.

② $\mathbb{E}[X_t | \mathcal{F}_s]$ is a random variable Z s.t. → Z is $\mathcal{F}_s$-measurable
→ $\mathbb{E}[Z \mathbb{1}_A] = \mathbb{E}[X_t \mathbb{1}_A]$, $\forall A \in \mathcal{F}_s$.

— $X = (X_t)_{t \in [0,\infty)}$ is called a super martingale (resp. sub martingale)
if $\mathbb{E}[|X_t|] < \infty$ $\forall t$, and $\mathbb{E}[X_t | \mathcal{F}_s] \le X_s$. (resp. $\mathbb{E}[X_t | \mathcal{F}_s] \ge X_s$.)

Doob's martingale convergence theorem. Assume that $X = (X_t)_{t \in [0,\infty)}$ is a super martingale s.t. $t \mapsto X_t(\omega)$ is right continuous for a.e. $\omega \in \Omega$. and $\sup_{t \in [0,\infty)} \mathbb{E}[X_t^-] < +\infty$. $(X_t^- := \max(-X_t, 0))$

Then, there exists a r.v. X_∞, s.t. $X_t(\omega) \longrightarrow X_\infty(\omega)$. a.e. $\omega \in \Omega$.

Remark: Given a discrete time process $(X_{t_k})_{k=0,1,\dots}$, one can consider it as a continuous time process by setting $X_t = X_{t_k}$, $\forall t \in [t_k, t_{k+1})$.

↳ which has right continuous paths

3. Example: Brownian motion. B.

A process $(B_t)_{t\geq 0}$ is calle a standard B.M. if
- $B_0 = 0$, $t \mapsto B_t(\omega)$ is continuous $\forall$ a.e. ω.
- $B_t - B_s \perp\!\!\!\perp \sigma(B_r : r \leq s)$. $\forall s \leq t$.
- $B_t - B_s \sim N(0, t-s)$.

Intuition: $(\xi_k)_{k\geq 1}$ i.i.d. r.v. $\sim N(0,1)$

$t_k := \frac{k}{n}$, $B^n_{t_k} := \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \xi_i = \sum_{i=1}^{k} \underbrace{\frac{1}{\sqrt{n}} \xi_i}_{N(0,\frac{1}{n})} \sim N(0, t_k) \xrightarrow{n \to +\infty} B_t$ in some sense.

Lemma: The standard B.M. is a martingale.

Proof: ① $B_t = B_t - B_0 \sim N(0,t) \implies \mathbb{E}[|B_t|] < +\infty$.

② $\mathbb{E}[B_t | \mathcal{F}_s] = \underbrace{\mathbb{E}[B_t - B_s | \mathcal{F}_s]}_{= \mathbb{E}[B_t - B_s] = 0} + \underbrace{\mathbb{E}[B_s | \mathcal{F}_s]}_{= B_s} = B_s$. #

4. Stochastic gradient algorithm.

Z is a r.v.

(P). $\min_{x\in\mathbb{R}^n} f(x)$, where. $f(x) = \mathbb{E}[F(x,Z)]$

$\boxed{\text{Algo 1: } x_{k+1} = x_k - \rho_k \nabla f(x_k).}$

$\nabla f(x) = \nabla \cdot \mathbb{E}[F(x,Z)] = \mathbb{E}[\nabla_x F(x,Z)]$

Stochastic gradient algo: $\boxed{X_{k+1} = X_k - \rho_k \cancel{\mathbb{E}[}\nabla_x F(x, Z_k)\cancel{]}}$

where $(Z_k)_{k\geq 1}$ is i.i.d sequence $\sim Z$.

Assumption: ① $\nabla_x F(x,Z)$ and $\nabla_x f(x) = \mathbb{E}[\nabla_x F(x,Z)]$ is uniformly bounded.

② $\boxed{\exists\, x^* \text{ s.t. } \langle x-x^*, \nabla f(x)\rangle > 0.}$ $\forall\ x \neq x^*$.

③ $\rho_k > 0$, $\sum_{k=1}^{\infty} \rho_k = +\infty$, $\sum_{k=1}^{\infty} \rho_k^2 < +\infty$.

Example $\rho_k = \frac{1}{k}$.

Theorm. Under the above Assumption: $X_k(\omega) \xrightarrow{k\to+\infty} x^*$. for a.e. $\omega\in\Omega$.

Proof. Let $\xi_{k+1} := \nabla_x F(X_k, Z_{k+1}) - \nabla f(X_k)$.

$\mathbb{E}[\xi_{k+1} \mid \mathcal{F}_k] = \mathbb{E}[\nabla_x F(X_k, Z_{k+1}) \mid \mathcal{F}_k]$

$- \mathbb{E}[\nabla f(X_k) \mid \mathcal{F}_k]$

Then. $X_{k+1} = X_k - \rho_k \nabla f(X_k) - \rho_k \xi_{k+1}$

$$= \nabla f(X_k) - \nabla f(X_k)$$
$$= 0$$

① Let. $S_n := \sum_{k=0}^{n-1} \rho_{k+1}^2 \, \mathbb{E}\left[|\nabla_x F(X_k, Z_{k+1})|^2 \,\middle|\, \mathcal{F}_k \right]$

$$\le C \cdot \sum_{k=0}^{n-1} \rho_{k+1}^2 \le C \cdot \sum_{k=0}^{\infty} \rho_{k+1}^2 < +\infty \quad a.s.$$

- $S_n \nearrow$ in n.
- $(S_n)_{n\ge1}$ is unif. bounded.

$\Rightarrow S_n(\omega) \nearrow S_\infty(\omega)$ for a.e. ω.

$$S_\infty(\omega) = \sum_{k=0}^{\infty} \rho_{k+1}^2 \, \mathbb{E}\left[|\nabla_x F(X_k, Z_{k+1})|^2 \,\middle|\, \mathcal{F}_k \right]$$

Next, Let. $Y_n := |X_n - x^*|^2 - S_n$.

Then: $\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = \mathbb{E}[|X_{n+1} - x^*|^2 | \mathcal{F}_n] - \mathbb{E}[S_{n+1} | \mathcal{F}_n]$

$$= \mathbb{E}[|X_n - x^*|^2 | \mathcal{F}_n] + \mathbb{E}[\langle X_n - x^*, \nabla_x F(X_n, Z_{n+1})\rangle | \mathcal{F}_n] \, \rho_n$$

$$+ \rho_n^2 \, \mathbb{E}[|\nabla_x F(X_n, Z_{n+1})|^2 | \mathcal{F}_n] - \mathbb{E}[S_{n+1} | \mathcal{F}_n]$$

$= \langle X_n - x^*, \nabla f(X_n) \rangle \le 0.$

$\le -S_n$

$$|X_{n+1} - x^*|^2 = |(X_n - x^*) + \rho_n \nabla_x F(X_n, Z_{n+1})|^2$$

$$\leq |X_n - x^*|^2 - S_n = Y_n.$$

$$\mathbb{E}[Y_{n+1} | \mathcal{F}_n] \leq Y_n.$$

$\Rightarrow$ (Y_n) is a super martingale.

and $\sup_n \mathbb{E}[Y_n^-] \leq \sup_n \mathbb{E}[S_n] < +\infty$

$\Rightarrow$ $Y_n \xrightarrow{a.s} Y_\infty$ for some r.v. Y_∞.

Besides: $S_n \xrightarrow{a.s} S_\infty$.

Then: $\underbrace{|X_n - x^*|^2}_{\geq 0} \xrightarrow{a.s.} Y_\infty - S_\infty =: L \geq 0$ a.s.

(2) We will prove that $L = 0$ a.s.

Assume that $\mathbb{P}[L > 0] > 0$. then for some $\delta > 0$, $A_\delta := \{\omega : L(\omega) \geq \delta\}$

$\mathbb{P}[A_\delta] > 0$.

Let $\eta := \inf_{\delta \leq |x - x^*|^2 \leq 2L} \langle \nabla f(x), x - x^* \rangle > 0. \quad \forall \delta > 0$

$$\Rightarrow \sum_{k=0} \rho_{k+1} \langle \nabla f(X_k), X_k - x^* \rangle \geq \sum_{k=0} \rho_{k+1} \cdot \eta = +\infty \quad \text{on } A_\delta.$$

$$\left(|X_k - x^*|^2 \to L \text{ so that } \delta \leq |X_k - x^*|^2 \leq 2L \text{ for } k \text{ large enough on } A_\delta. \right)$$

$$\Rightarrow \mathbb{E}\left[\sum_{k=0}^{\infty} \rho_{k+1} \langle \nabla f(X_k), X_k - x^* \rangle \right] \geq \mathbb{E}\left[+\infty \, \mathbb{1}_{A_\delta} \right] = \infty$$

However, $\mathbb{E}\left[\sum_{k=0}^{\infty} \rho_{k+1} \langle \nabla f(X_k), X_k - x^* \rangle \right]$

$$= \sum_{k=0}^{\infty} \rho_{k+1} \cdot \mathbb{E}\left[\cancel{\mathbb{E}}\left[\underbrace{\langle \nabla_x F(X_k, Z_{k+1}), X_k - x^* \rangle}_{\langle \nabla f(X_k), X_k - x^* \rangle} \cancel{\mid \mathcal{F}_k} \right] \right]$$

$$= \sum_{k=0}^{\infty} \cancel{\rho} \, \mathbb{E}\left[\langle X_{k+1} - X_k , X_k - x^* \rangle \right]$$

$$= \frac{1}{2} \sum_{k=0}^{\infty} \mathbb{E}\left[\underbrace{|X_{k+1} - x^*|^2}_{|X_{k+1} - X_k + X_k - x^*|^2} - |X_k - x^*|^2 - |\rho_{k+1} \nabla_x F(X_k, Z_{k+1})|^2 \right]$$

$$= \frac{1}{2} \lim_{n\to+\infty} \left(\sum_{k=0}^{n} \mathbb{E}\left[|X_{k+1} - x^*|^2 - |X_k - x^*|^2 \right] - \sum_{k=0}^{n} \mathbb{E}\left[|\rho_{k+1} \nabla_x F(X_k, Z_{k+1})|^2 \right] \right)$$

$$= \frac{1}{2} \lim_{n\to+\infty} \left(\mathbb{E}\left[|X_{n+1} - x^*|^2 \right] - \mathbb{E}\left[|X_0 - x^*|^2 \right] - \sum_{k=0}^{n} \rho_{k+1}^2 \, \mathbb{E}\left[|\nabla_x F(X_k, Z_{k+1})|^2 \right] \right)$$

$$= \frac{1}{2} \lim_{n\to+\infty} \left(\mathbb{E}[Y_{n+1}] - \mathbb{E}\left[|X_0 - x^*|^2 \right] \right) < +\infty.$$

Since $(Y_n)_{n\geq 1}$ is a super martingale and $\sup_n \mathbb{E}[Y_n^-] < +\infty$

This is a contradiction.!

Then $L = 0$ a.s. #

(P) $\min_{x\in\mathbb{R}^n} f(x)$.

$$\boxed{x_{k+1} = x_k - \gamma \cdot \nabla f(x_k).}$$

When $f: \mathbb{R}^n \to \mathbb{R}$ is not convex | variation:

$$\boxed{X_{k+1} = X_k - \Delta t \cdot \nabla f(X_k) + \boxed{\sigma \sqrt{\Delta t}\, \xi_{k+1}}}$$

where $(\xi_k)_{k\geq 1}$ i.i.d. r.v. $N(0,1)$

$\Longrightarrow$ ① $(X_k)_{k=0,1,\dots}$ is a Markov chain. invariant measure

② limit $\Delta t \downarrow 0$ $\quad X_{k+1} = X_0 - \sum_{i=0}^{k} \nabla f(X_i)\Delta t$

$$+ \sigma \underbrace{\sum_{i=0}^{k} \sqrt{\Delta t}\, \xi_{i+1}}$$

$\to$ B.M.

$\to N\left(0, \sum_{i=0}^{k} \Delta t\right) = N(0, k\cdot\Delta t)$

$$\boxed{X_t = X_0 - \int_0^t \nabla f(X_s)\, ds + \sigma \cdot B_t.}$$

Markov diffusion process (Langevin Equation)

$\boxed{\mathcal{L}(X_t)} \longrightarrow$ $\boxed{\text{invariant measure of } X.}$

with density $\frac{1}{C(\sigma)} \exp\left(-\frac{1}{\sigma^2} f(x)\right) \xrightarrow{\sigma \to 0} \delta_{x^*}$

Exercise.

— Fokker – Planck Equation on $\mathcal{L}(X_t)$

Fokker Planck equation on $\mathcal{L}(X_t)$.

Let $\varphi \in C_c^\infty(\mathbb{R}^n)$ $\qquad 0 = t_0 < t_1 < \cdots < t_n = t.$

$$\varphi(X_t) = \varphi(X_0) + \sum_{k=0}^{n-1}\left(\varphi(X_{t_{k+1}}) - \varphi(X_{t_k})\right)$$

$$\approx \varphi(X_0) + \sum_{k=0}^{n-1}\varphi'(X_{t_k})\left(X_{t_{k+1}} - X_{t_k}\right) + \sum_{k=0}^{n-1}\frac{1}{2}\varphi''(X_{t_k})\left(X_{t_{k+1}} - X_{t_k}\right)^2.$$

$$\approx \varphi(X_0) - \sum_{k=0}^{n-1}\varphi'(X_{t_k})\,\nabla f(X_{t_k})\,\Delta t$$

$$+\mathbb{E}\left[\sum_{k=0}^{n-1}\varphi'(X_{t_k})\,\sigma\sqrt{\Delta t}\,\underbrace{\xi_{k+1}}_{\approx (B_{t_{k+1}} - B_{t_k})}\right] = \mathbb{E}[\cdot] = 0$$

$$+\sum_{k=0}^{n-1}\frac{1}{2}\varphi''(X_{t_k})\cdot\underbrace{\left(X_{t_{k+1}} - X_{t_k}\right)^2}_{\approx \Delta t}.$$

$$\mathbb{E}[\varphi(X_t)] = \mathbb{E}[\varphi(X_0)] - \int_0^t \mathbb{E}[\nabla\varphi(X_s)\cdot\nabla f(X_s)]\,ds$$

$$+\mathbb{E}\left[\int_0^t \nabla\varphi(X_s)\cdot\sigma\,dB_s\right] = 0$$

$$+\frac{1}{2}\int_0^t \mathbb{E}[\nabla^2\varphi(X_s)]\,\sigma^2\,ds$$

$$\boxed{m_t \triangleq \mathcal{L}(X_t)}$$

$$\Rightarrow \int \varphi(x)\, m_t(x)\, dx = \langle \varphi, m_t \rangle$$

$$= \langle \varphi, m_0 \rangle - \int_0^t \langle \nabla\varphi \cdot \nabla f, m_s \rangle\, ds + \frac{1}{2}\sigma^2 \int_0^t \langle \nabla^2 \varphi, m_s \rangle\, ds.$$

$$\Rightarrow \int_0^t \langle \varphi, \partial_t m_s \rangle\, ds + \int_0^t \langle \nabla\varphi, \nabla f\, m_s \rangle\, ds - \frac{1}{2}\sigma^2 \int_0^t \langle \nabla^2\varphi, m_s \rangle\, ds$$

integration by parts.

$$\xRightarrow[\varphi \in C_c^\infty(\mathbb{R}^n)]{} \int_0^t \langle \varphi, \partial_t m_s \rangle\, ds - \int_0^t \langle \varphi, \nabla(\nabla f\, m_s) \rangle\, ds$$

$$- \frac{1}{2}\sigma^2 \int_0^t \langle \varphi, \nabla^2 m_s \rangle\, ds, \quad \forall \varphi \in C_c^\infty$$

$$\Longrightarrow \left\langle \boxed{\cancel{\partial_t m_s}} - \nabla(\nabla f \cdot m_s) + \frac{1}{2}\sigma^2 \cdot \nabla^2 m_s ,\ \varphi \right\rangle = 0 \quad \forall \varphi \in C_c^\infty$$

$$\nabla(\nabla f \cdot m_s) + \frac{1}{2}\sigma^2 \nabla^2 m_s = 0$$

Fokker-Planck Equation.

Invariant measure m_∞ should satisfy

$$\nabla(\nabla f \cdot m_\infty) + \frac{1}{2}\sigma^2 \nabla^2 m_\infty = 0.$$

$$\Updownarrow$$

$$\nabla^2 f \cdot m + \nabla f \cdot \nabla m + \frac{1}{2}\sigma^2 \nabla^2 m = 0$$

$$\boxed{\text{Let } m_\infty(x) = \frac{1}{C(\sigma)} \exp\left(-\frac{2}{\sigma^2} f(x)\right)}$$

$$\nabla m_\infty(x) = m_\infty(x) \left(-\frac{2}{\sigma^2} \nabla f(x)\right).$$

$$\nabla^2 m_\infty(x) = m_\infty(x) \left(-\frac{2}{\sigma^2} \nabla^2 f(x)\right) + m_\infty(x) \cdot \frac{4}{\sigma^4} \langle \nabla f, \nabla f \rangle$$

$$\Longrightarrow \nabla^2 f \cdot m + \left\langle \nabla f,\ m \left(-\frac{2}{\sigma^2} \nabla f\right)\right\rangle$$

$$+\cancel{\frac{1}{2}\sigma^2 \cdot m \cdot \left(-\frac{2}{\sigma^2}\right) \cdot \nabla^2 f} + \frac{1}{2}\sigma^2 m \frac{4}{\sigma^4} \langle \nabla f, \nabla f \rangle$$

$$= 0$$

$$m_\sigma(x) = \frac{1}{\sigma(\sigma)} \exp\left(-\frac{2}{\sigma^2} f(x)\right) dx \xrightarrow{\text{as } \sigma \downarrow 0} \delta_{x^*}(dx) \quad \#$$