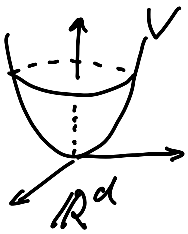
Lecture 21, Friday, 5/13/2022

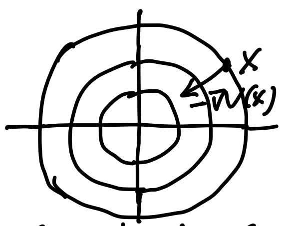
- Finish the proof of convergence theorem
- The JKO scheme

The steepest descent / gradient descent

$x = x(t) \in \mathbb{R}^d$: $\quad \dot{x} = -\nabla V(x)$. i.e. $\frac{dx(t)}{dt} = -\nabla V(x(t))$

$V: \mathbb{R}^d \to [0,\infty)$ given, smooth, convex, ...





level sets of V

$$\frac{d}{dt}\left[V(x(t))\right] = \nabla V(x(t)) \cdot \dot{x}(t) = -|\nabla V(x(t))|^2 \le 0.$$

$\forall x \in \mathbb{R}^d$: $\nabla V(x) \perp \{ y \in \mathbb{R}^d : V(y) = V(x) \}$.

$|\nabla V(x) \cdot \vec{v}|$ ($\vec{v}$: unit vector) is maximized at $\pm \nabla V(x)/|\nabla V(x)|$.

$\Rightarrow$ steepest descent of $V(x(t))$.

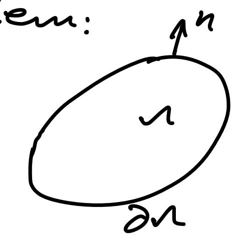
The heat equation $\quad \partial_t u = \Delta u \quad u = u(x,t)$

Given: $\Omega \subseteq \mathbb{R}^d$, bounded and smooth domain

$u_0 = u_0(x)$, bounded and smooth function.

The initial-boundary-value problem:

$$\begin{cases} \partial_t u = \Delta u & \Omega \times (0,\infty) \\ \partial_n u = 0 & \partial\Omega \times (0,\infty) \\ u(\cdot, 0) = u_0 \end{cases}$$



$\partial_n u = (\nabla \cdot n) u$

- Conservation of mass:

$$\frac{d}{dt}\int_\Omega u\,dx = \int_\Omega \partial_t u\,dx = \int_\Omega \Delta u\,dx = \int_{\partial\Omega} \partial_n u\,dS = 0$$

$$\Longrightarrow \int_\Omega u\,dx = \text{const.} = \int_\Omega u_0\,dx \qquad \forall t \ge 0.$$

- This is the gradient flow. $\partial_t u = -\delta I[u]$.

$I[v] := \int_\Omega \frac{1}{2}|\nabla v|^2 dx. \qquad \delta I[v] = -\Delta v.$

$$\delta I[v][\varphi] = \frac{d}{dt}\Big|_{t=0} I[v + t\varphi] = \frac{d}{dt}\Big|_{t=0} \int_\Omega \frac{1}{2}|\nabla v + t\nabla\varphi|^2 dx$$

$$= \int_\Omega \nabla v \cdot \nabla \varphi \, dx = -\int_\Omega \Delta v \, \varphi \, dx = \langle -\Delta v, \varphi \rangle_{L^2(\Omega)} \quad \forall \varphi \in C^1(\bar\Omega).$$

Identify $\delta I[v] = -\Delta v$.

Note: This is the gradient flow w.r.t. $L^2(\Omega)$.

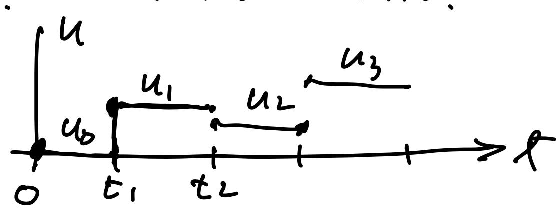
<u>The implicit Euler scheme</u> Let $\tau > 0$ be small (time step). Let $t_k = k\tau$ ($k = 0, 1, 2, \cdots$). Approximate $u(x, t_k)$ by $u_k(x) = u_k^\tau(x)$ ($x \in \Omega$, $k = 1, 2, \cdots$).

$$\begin{cases} u_0 : \text{given}, \\ \dfrac{u_{k+1} - u_k}{\tau} = \Delta u_{k+1} \quad (k = 0, 1, \cdots). \end{cases}$$

Ideas: $\partial_t u(x, t+\tau) \approx \dfrac{u(x, t+\tau) - u(x, t)}{\tau}$.

$\partial_t u(x, t+\tau) = \Delta u(x, t+\tau)$. $\quad$ r.h.s = r.h.s.

Set $u^\tau(x,t) = u_k^\tau(x)$ if $t_k \le t < t_{k+1}$.



Convergence:

$u^\tau \to u$ as $\tau \to 0$.

<u>Claim</u>

$$u_{k+1} = \operatorname*{argmin}_u \Big(\underbrace{\frac{1}{2\tau} \|u - u_k\|^2_{L^2(\Omega)} + I[u]}_{I_k[u]} \Big), \text{ unique.}$$

<u>Proof</u> $I_k[u] := \frac{1}{2\tau} \|u - u_k\|^2_{L^2(\Omega)} + I[u]$

$$= \int_\Omega \left(\frac{1}{2\tau} (u - u_k)^2 + \frac{1}{2} |\nabla u|^2 \right) dx$$

$$\delta I_k[u][\varphi] := \frac{d}{dt} \left(I_k[u + t\varphi] \right) \Big|_{t=0}$$

$$= \int_\Omega \left[\nabla u \cdot \nabla \varphi + \frac{1}{\tau}(u - u_k)\varphi \right] dx$$

$$= \int_\Omega \left[-\Delta u + \frac{1}{\tau}(u - u_k) \right] \varphi \, dx.$$

So, $\delta I_k[u] = -\Delta u + \frac{1}{\tau}(u - u_k)$. Now, u_{k+1} solves the the Euler-Lagrange eq. $\delta I_k[u] = 0$. Since I_k is convex, u_{k+1} is the unique minimizer. QED

<u>The JKO scheme</u> (Jordan-Kinderlehrer-Otto, 1998)

Define

$$F(\rho) = \int_\Omega \rho \log \rho \, dx, \quad \rho = \rho(x) \geq 0, \; x \in \Omega.$$

$$\mathcal{P}(\Omega) = \{ \rho \in L^1(\Omega) : \rho \geq 0, \int_\Omega \rho \, dx = 1 \}. \Rightarrow \int_\Omega |x|^2 \rho(x) dx < \infty.$$

$\rho \in \mathcal{P}(\Omega)$

Given: $\rho_0 \in \mathcal{P}(\Omega)$, $\tau > 0$, $t_k = k\tau$, $k = 0, 1, 2, \dots$

$$\rho_{k+1} = \arg\min_\rho \left(\frac{1}{2\tau} W_2^2(\rho, \rho_k) + F[\rho] \right).$$

Note: for $\xi, \eta \in \mathcal{P}(\Omega)$, $W_2(\xi, \eta) = W_2(\xi dx, \eta dx)$.

$X = \Omega$: compact, Euclidean distance.

<u>Convergence</u> $\rho^\tau \to \rho$ weakly and $\partial_t \rho = \Delta \rho$.

<u>The Fokker-Planck eq. (FPE):</u>

$$\partial_t \rho = \beta^{-1} \Delta \rho + \nabla \cdot (\rho \nabla V)$$

$V : \mathbb{R}^d \to \mathbb{R}$ given, bounded below.

Known: ρdx is the probability distribution for the stochastic process $X_t \in \mathbb{R}^d$:

$$dX_t = -\nabla V(X_t) dt + \sqrt{2\beta^{-1}} \, dW_t$$

Here: W_t is the standard d-dimensional Brownian motion.

Lemma Let $\rho_0 \in \mathcal{P}(\Omega)$ be such that $F[\rho_0] < \infty$. There exists a unique $\hat{\rho} \in \mathcal{P}(\Omega)$ s.t.

$$\hat{\rho} = \underset{\rho \in \mathcal{P}(\Omega)}{\operatorname{argmin}} J[\rho]$$

where $J[\rho] = W_2^2(\rho, \rho_0) + F[\rho]$,

$$F[\rho] = \int_\Omega \rho \log \rho \, dx.$$

Proof Let $\alpha := \inf_{\rho \in \mathcal{P}(\Omega)} J_0[\rho] < \infty$. ($F[\rho_0] < \infty$)

Let $\rho_n \in \mathcal{P}(\Omega)$ $(n = 1, 2, \cdots)$ be such that $J[\rho_n] \to \alpha$.

Since $\{F[\rho_n]\}$ is bounded and $s \mapsto s \log s$ is superlinear, there exists a subseq. of $\{\rho_n\}$, not relabelded, such that $\rho_n \to \hat{\rho}$ weakly in $L^1(\Omega)$. (by de la Vallée Poussin's criterior of weakly sequential compactness of $L^1(\Omega)$.) Since all $\rho_n \in \mathcal{P}(\Omega)$, we have $\hat{\rho} \in \mathcal{P}(\Omega)$.

Since $s \mapsto s \log s$ is convex, the functional F is sequentially weakly semicontinuous. Thus,

$$\liminf_{n \to \infty} F[\rho_n] \geq F[\hat{\rho}].$$

Let $\gamma_n \in \mathcal{A}(\rho_n, \hat{\rho})$ $(= \mathcal{A}(\rho_n dx, \hat{\rho} dx))$ be such that $W_2^2(\rho_n, \hat{\rho}) = \int_{\Omega \times \Omega} |x - y|^2 d\gamma_n(x, y)$. $(n = 1, 2, \cdots)$.

Since $\rho_n \to \hat{\rho}$ weakly in $L^1(\Omega)$, $\rho_n dx \to \rho dx$

narrowly. Hence, $\{\rho_n dx\}$ is tight by Prokhorov's Theorem. Hence $\{\gamma_n\}$ is also tight, and again by Prokhorov's Theorem, $\{\gamma_n\}$ is sequentially compact. Hence, it has a further subseq., not relabeled, such that $\gamma_n \to \hat{\gamma} \in \mathcal{A}(\hat{\rho}, \rho_0)$ narrowly. Consequently,

$$W_2^2(\rho_n, \rho_0) = \int_{\Omega \times \Omega} |x-y|^2 d\gamma_n(x,y)$$

$$\longrightarrow \int_{\Omega \times \Omega} |x-y|^2 d\hat{\gamma}(x,y) \geq W_2^2(\hat{\rho}, \rho_0).$$

Therefore, $\alpha = \liminf_{n\to\infty} J[\rho_n] \geq J[\hat{\rho}] \geq \alpha$, and $\hat{\rho} \in \mathcal{P}(\Omega)$ is a minimizer of J over $\mathcal{P}(\Omega)$. The uniqueness from the convexity of $F[\rho]$, and $W_2^2(\cdot, \rho_0)$; cf. a result below. QED

Proposition Let (X,d) be a Polish space. Let $\mu_0, \mu_1, \nu \in \mathcal{P}_2(X)$. Then for each $\lambda \in (0,1)$, $\mu_\lambda := (1-\lambda)\mu_0 + \lambda \mu_1 \in \mathcal{P}_2(X)$. Moreover,

$$W_2^2(\mu_\lambda, \nu) \leq (1-\lambda) W_2^2(\mu_0, \nu) + \lambda W_2^2(\mu_1, \nu).$$

Proof clearly $\mu_\lambda \in \mathcal{P}_2(X)$. For $i=1,2$, let $\gamma_i \in \mathcal{A}(\mu_i, \nu)$ be optimal: $W_2^2(\mu_i, \nu) = \int_{X\times X} d^2(x,y)\, d\gamma_i(x,y)$. Then, $\gamma_\lambda = (1-\lambda)\gamma_0 + \lambda \gamma_1 \in \mathcal{A}(\mu_\lambda, \nu)$. Moreover,

$W_2^2(\mu_\lambda, \nu) \leq \int_{X\times X} d^2(x,y)\, d\gamma_\lambda(x,y) = (1-\lambda) W_2^2(\mu_0, \nu) + \lambda W_2^2(\mu_1, \nu)$. QED

Theorem Let $\Omega \subseteq \mathbb{R}^d$ be a bounded and smooth domain. Let $\tau > 0$ and $t_k = k\tau$ $(k = 0, 1, \dots)$. Assume $\rho_0 \in \mathcal{P}(\Omega)$ and $F[\rho_0] = \int_\Omega \rho_0 \log \rho_0 \, dx < \infty$.

Let $\rho_{k+1}^{\tau} \in \mathcal{P}(\Omega)$ be the unique minimizer of

$$J_k[\rho] = \frac{1}{2\tau} W_2^2(\rho, \rho_k) + F[\rho]$$

among all $\rho \in \mathcal{P}(\Omega)$, $k = 0, 1, \cdots$. Let $\rho^{\tau}(t) = \rho_k^{\tau}$ if $t_k \le t < t$. Then there exist $\rho \in L^1_{loc}([0,\infty) \times \Omega)$ and a subsequence of $\{\rho^{\tau}\}_{\tau > 0}$, not relabeled, such that $\rho^{\tau} \to \rho$ weakly in $L_{loc}([0,\infty) \times \Omega)$. Moreover ρ is the distributional solution to the initial-boundary-value problem of the heat equation.

<u>Remark</u> The convergence can be improved to be the strong convergence of the entire original sequence ρ^{τ}.