A "Miscellaneous" Introduction to the Theory and Applications of Gradient Flows

$$\boxed{\frac{dX}{dt} = -\nabla F(X)}$$

Aaron Nung Kwan Yip
Purdue University

Central Michigan University
AMS Graduate Student Chapter
April, 22nd, 2022

Outline

1. Concept of Gradient Flows
   Examples from Statistics & physics

2. Long time behavior, convergence rate to minimum

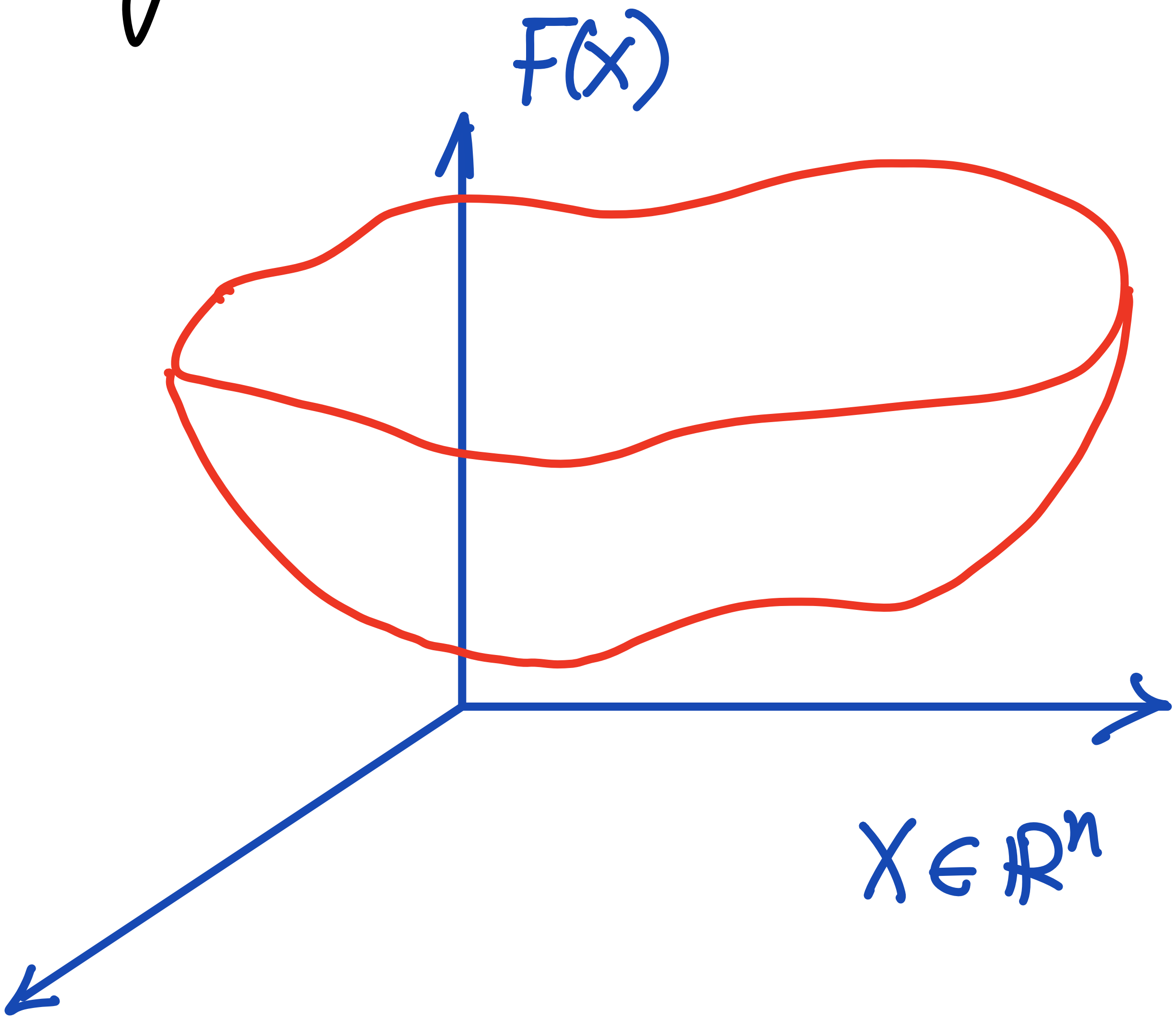
3. Examples of PDE: heat equation

What is a gradient?

Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dx}{dt} = -\nabla F(x)}$$



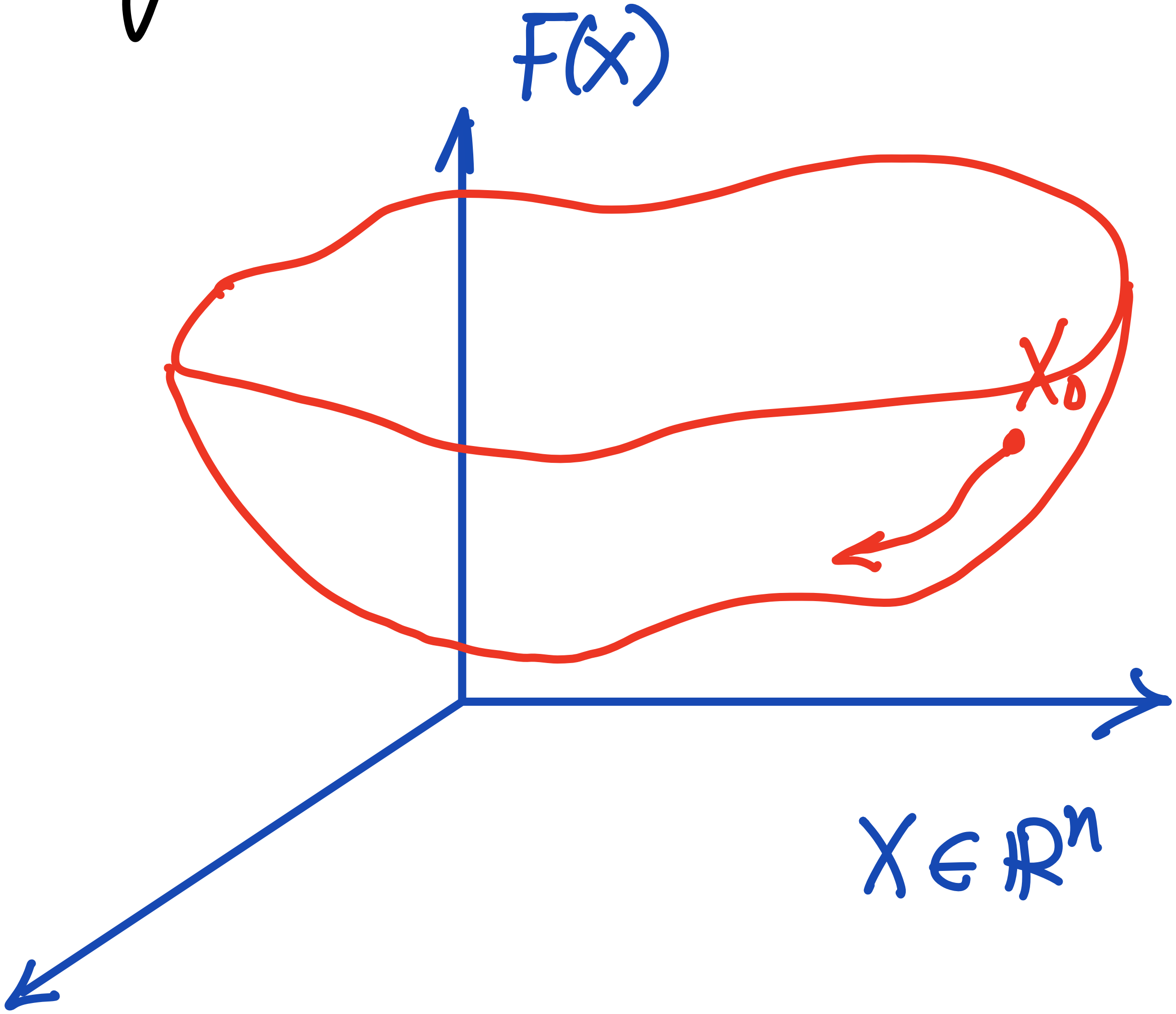
$$\min_{x \in \mathbb{R}^n} F(x)$$

Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dx}{dt} = -\nabla F(x)}$$

$$\min_{x \in \mathbb{R}^n} F(x)$$



Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dX}{dt} = -\nabla F(X)}$$

$$\frac{d}{dt} F(X(t)) = \left\langle \nabla F(X(t)), \frac{d}{dt} X \right\rangle$$

$$= \left\langle \nabla F(X(t)), -\nabla F(X(t)) \right\rangle$$

$$= -\|\nabla F(X(t))\|^2 \leq 0$$

Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dX}{dt} = -\nabla F(X)}$$

$$\frac{d}{dt} F(X(t)) = \left\langle \nabla F(X(t)), \frac{d}{dt} X \right\rangle$$

$$= \left\langle \nabla F(X(t)), -\nabla F(X(t)) \right\rangle$$

$$= -\left\| \nabla F(X(t)) \right\|^2 \leq 0$$

$F(X(t)) \downarrow$ in t

$F(X(t)) < F(X(0))$

Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dX}{dt} = -\nabla F(X)}$$

$$F(X(t)) + \int_0^t \|\nabla F(X(s))\|^2 ds = F(X(0))$$

Gradient Descent

(Negative) Gradient Flow

$$\boxed{\frac{dX}{dt} = -\nabla F(X)}$$

$$F(X(t)) + \int_0^t \underbrace{\|\nabla F(X(s))\|^2}_{\text{Dissipation}} ds = F(X(0))$$

$$\int_0^\infty \|\nabla F(X(s))\|^2 ds = \underbrace{F(X(0)) - \lim_{t \to +\infty} F(X(t))}_{>0}$$

Examples (from Statistics, ML)

(1) Least Square

$$\min_{X} \|AX - b\|^2 \quad (\text{Solve } AX = b)$$

Examples (from Statistics, ML)

(1) Least Square

$$\min_{X} \|AX - b\|^2 \quad (\text{Solve } AX = b)$$

$$\min_{X} \|AX-b\|^2 + \delta \underbrace{\|X\|^2}_{\text{regularization}}$$

$$\min_{X \in \mathcal{S}} \|AX-b\|^2$$

constraint set

Examples (from Statistics, ML)

(2) Maximum Likelihood Estimation

$$\{X_i = x_i\}_{i=1}^{N}, \quad X_i \sim f_\theta(x), \text{ iid}$$

Data

parameter

$$P(X_i = x_i,\ i=1,2,\dots N) = \prod_{i=1}^{N} f_\theta(x_i)$$

Examples (from Statistics, ML)

(2) Maximum Likelihood Estimation

$$\{X_i = x_i\}_{i=1}^{N}, \quad X_i \sim f_\theta(x), \text{ iid}$$

Data

parameter

$$P(X_i = x_i,\ i = 1, 2, \dots N) = \prod_{i=1}^{N} f_\theta(x_i)$$

Likelihood fct $L(\theta)$

$\max_\theta L(\theta)$, $\hat{\theta}$ = maximum likelihood est.

Examples (from Statistics, ML)

(3) Kullback-Leibler Divergence

$$D_{KL}(f\|g) = \int f \log \frac{f}{g}\, dx$$

$f, g \geq 0$

$\int f = \int g = 1$

relative entropy

Examples (from Statistics, ML)

(3) Kullback-Leibler Divergence

$$D_{KL}(f\|g) = \int f \log \frac{f}{g}\, dx$$

$f, g \geq 0$

$\int f = \int g = 1$

$$= \int \left(\frac{f}{g}\right) \log\left(\frac{f}{g}\right) (g\, dx)$$

Examples (from Statistics, ML)

(3) Kullback-Leibler Divergence

$$D_{KL}(f\|g) = \int f \log \frac{f}{g}\, dx$$

$f, g \geq 0$

$\int f = \int g = 1$

$$= \int \left(\frac{f}{g}\right) \log\left(\frac{f}{g}\right) (g\, dx)$$

$$\boxed{D_{KL}(f\|g) \geq 0, \quad = 0 \text{ iff } f = g}$$

(i.e. given g, $D_{KL}(\cdot\|g)$ minimizes at g.)

Examples (from Statistics, ML)

(4) Posterior Distribution

u - parameter(s), y - data

$p(u)$ - prior dist. of u

$p(u|y)$ - post. dist. of u, given y

$$\frac{p(u,y)}{p(y)} = \frac{p(y|u)\,p(u)}{p(y)}$$

Examples (from Statistics, ML)

(4) Posteria Distribution

prior of u

$$\boxed{p(u/y) = \frac{p(y/u)\, p(u)}{p(y)}}$$

Posterior of u, given y

Likelihood func $L(u/y)$

Examples (from Statistics, ML)

(4) Posteria Distribution

$$\boxed{J_{KL}(q(\cdot)) = D_{KL}(q\|p) - \int (\log L(u|y))\, q(u)\, du}$$

Examples (from Statistics, ML)

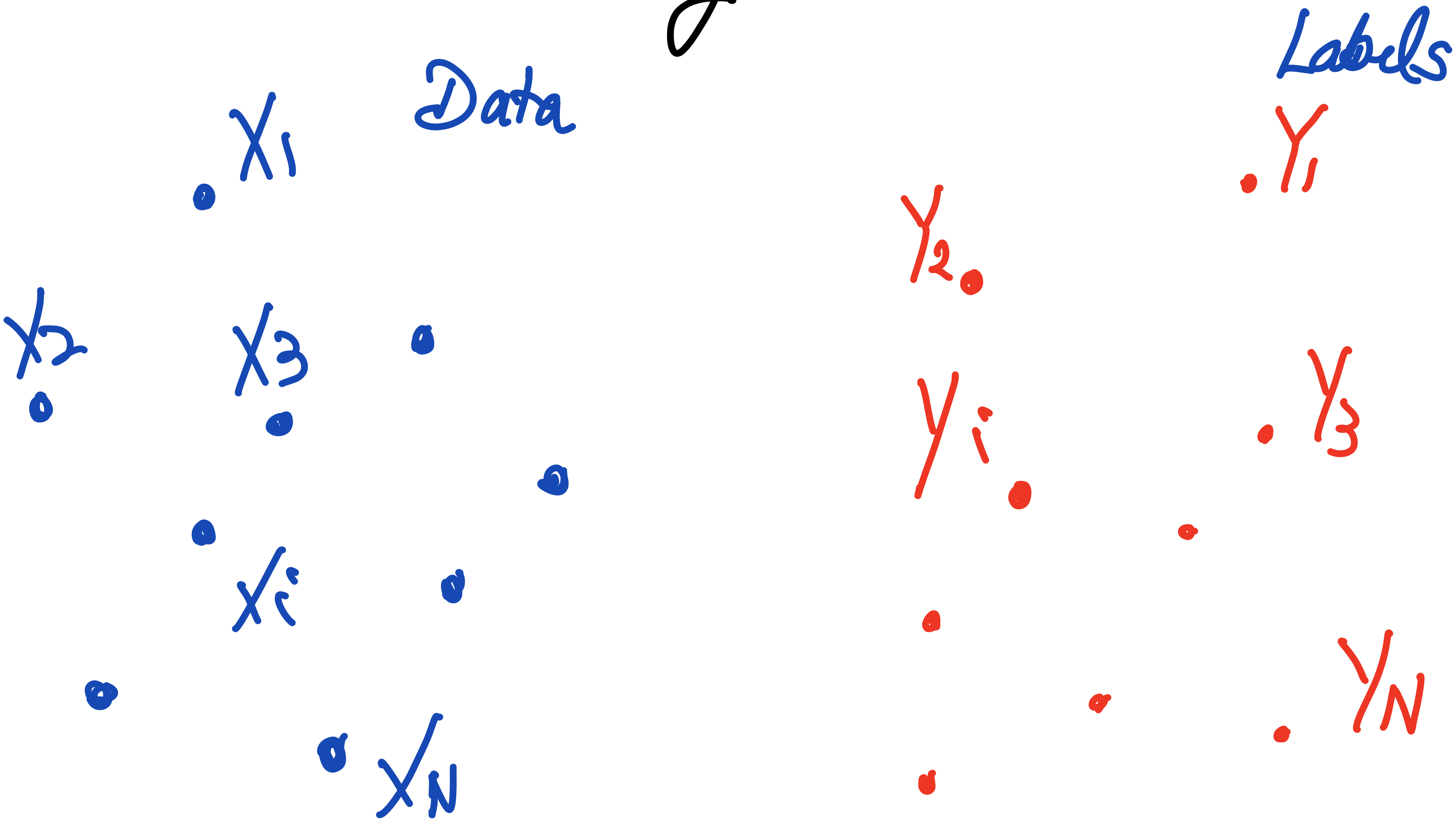
(4) Posterior Distribution

$$\boxed{J_{KL}(q(\cdot)) = D_{KL}(q\|p) - \int \big(\log L(u|y)\big)\, q(u)\, du}$$

$$\boxed{p(u|y) \text{ minimizes } J_{KL}(\cdot)}$$

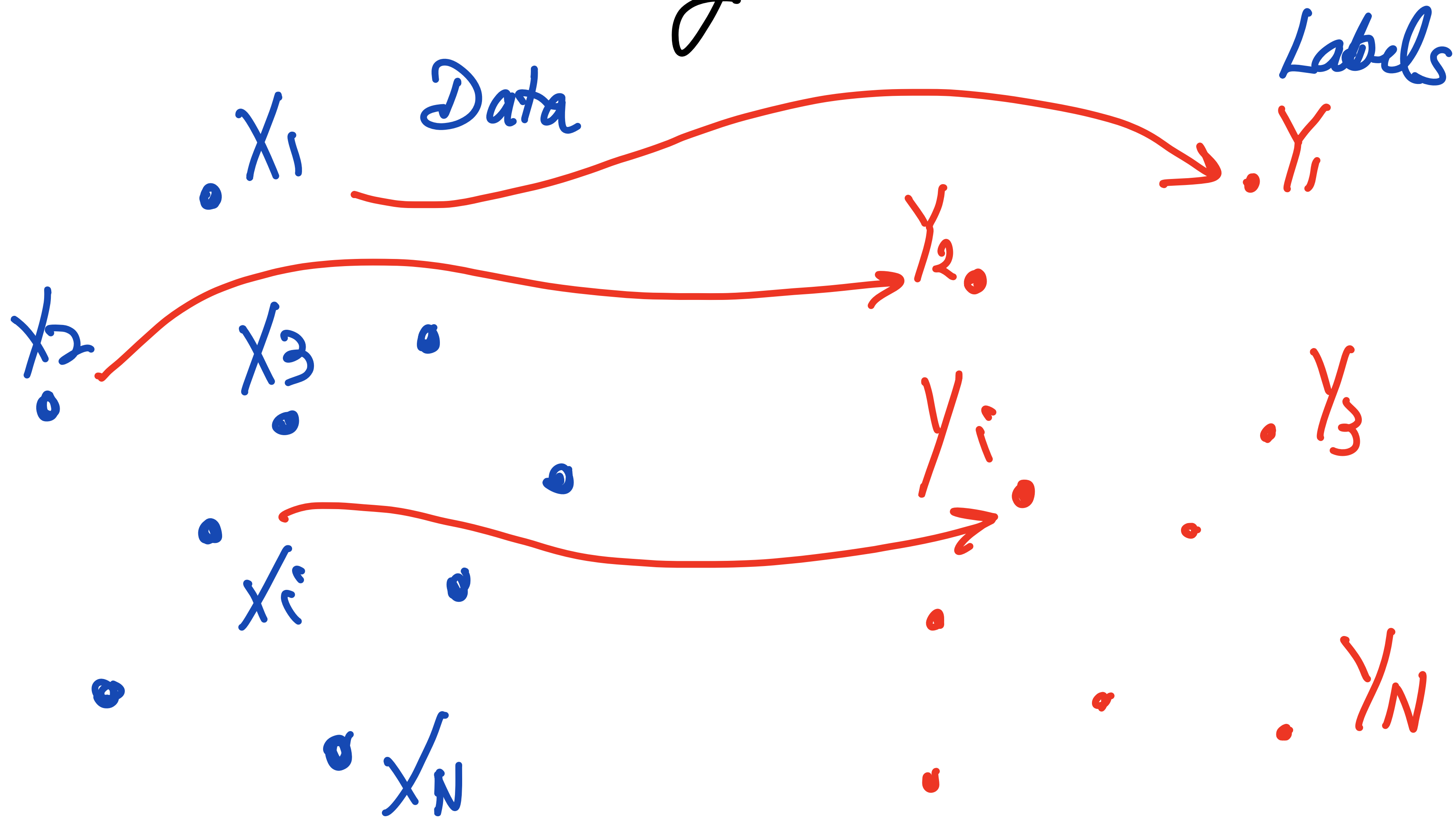
Examples (from Statistics, ML)

(5) Machine Learning



Examples (from Statistics, ML)

(5) Machine Learning



Examples (from Statistics, ML)

(5) Machine Learning

$$\mathcal{X} = \{X\text{'s}\} \xrightarrow{\quad f \quad} \mathcal{Y} = \{Y\text{'s}\}$$

$$\min_{f \in \mathcal{F}} \int_{X \times Y} \| f(X) - Y \|^2 \, d\mathbb{P}(X, Y)$$

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \| f(X_i) - Y_i \|^2$$

Examples (from Statistics, ML)

(5) Machine Learning

$$\mathcal{X} = \{X's\} \xrightarrow{f} \mathcal{Y} = \{Y's\}$$

$$\min_{f \in \mathcal{F}} \int_{X \times Y} \| f(X) - Y \|^2 \, d\mathbb{P}(X, Y)$$

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \| f_{\theta}(X_i) - Y_i \|^2$$

parameter

Examples (from Physics)

$V: \mathbb{R}^n \longrightarrow \mathbb{R}$, potential fct

$F: \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $F(x) = -\nabla V(x)$

$F = ma$ (Newton's 2nd Law)

$m\ddot{X} = -\nabla V(x)$

$m_i \ddot{X}_i = -\partial_{x_i} V(x)$

Examples (from Physics)

$$\frac{d}{dt}\Big(\underbrace{\tfrac{1}{2}m\|\dot{x}\|^2}_{K.E} + \underbrace{V(x)}_{P.E.}\Big)$$

(E)

K.E + P.E. = Total Energy

$$= m\langle \dot{x}, \ddot{x}\rangle + \langle \nabla V(x), \dot{x}\rangle$$

$m\ddot{x} = -\nabla V$

$$= 0 \, !$$

Examples (from Physics)

$$\frac{d}{dt}\Big(\underbrace{\tfrac{1}{2} m \|\dot{x}\|^2}_{K.E} + \underbrace{V(x)}_{P.E.}\Big) \qquad (E)$$

$K.E + P.E. =$ Total Energy

$$= m\langle \dot{x}, \ddot{x} \rangle + \langle \nabla V(x), \dot{x} \rangle$$

$m\ddot{x} = -\nabla V$

$$= 0!$$

(Conservation of energy)

Examples (from Physics)

$V : \mathbb{R}^n \longrightarrow \mathbb{R}$, potential fct

$F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $F(x) = -\nabla V(x)$

$F = ma$ (Newton's 2^{nd} Law)

$m\ddot{x} = -\nabla V(x) - \gamma \dot{x}$ (friction)

$m_i \ddot{x}_i = -\partial_{x_i} V(x) - \gamma \dot{x}_i$

Examples (from Physics)

$$\frac{d}{dt}\Big(\underbrace{\tfrac{1}{2} m \|\dot{x}\|^2}_{K.E} + \underbrace{V(x)}_{P.E.}\Big) \qquad (E)$$

K.E + P.E. = Total Energy

$$= m\langle \dot{x}, \ddot{x} \rangle + \langle \nabla V(x), \dot{x} \rangle$$

$m\ddot{x} = -\nabla V - \gamma\dot{x}$

$$= \langle \dot{x}, -\gamma\dot{x} \rangle$$

$$= -\gamma \|\dot{x}\|^2 < 0 \quad (E \downarrow \text{ in time})$$

Examples (from Physics)

(1) Gravitational motions:

Gravitational Potential

$$V(x_1, \cdots, x_N) = -\sum_{i \neq j} \frac{G m_i m_j}{\|x_i - x_j\|}$$

$$\partial_{x_i} V(x_1, \cdots, x_N) = -\sum_{j \neq i} \frac{G m_i m_j (x_j - x_i)}{\|x_i - x_j\|^3}$$

Examples (from Physics)

(1) Gravitational motions:

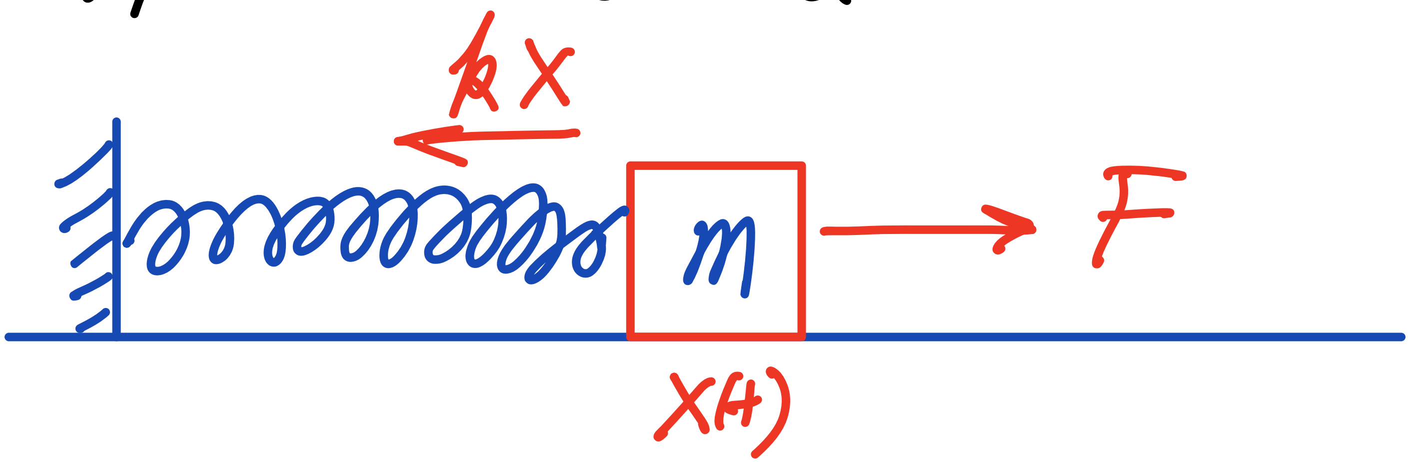
Gravitational Potential

$$V(x_1, \cdots, x_N) = - \sum_{i \neq j} \frac{G m_i m_j}{\| x_i - x_j \|}$$

$$m_i \ddot{x}_i = \sum_{j \neq i} \frac{G m_i m_j (x_j - x_i)}{\| x_i - x_j \|^3}$$

Examples (from Physics)

(2) Harmonic Oscillator



$$m\ddot{x} = F - kx - \gamma\dot{x}$$

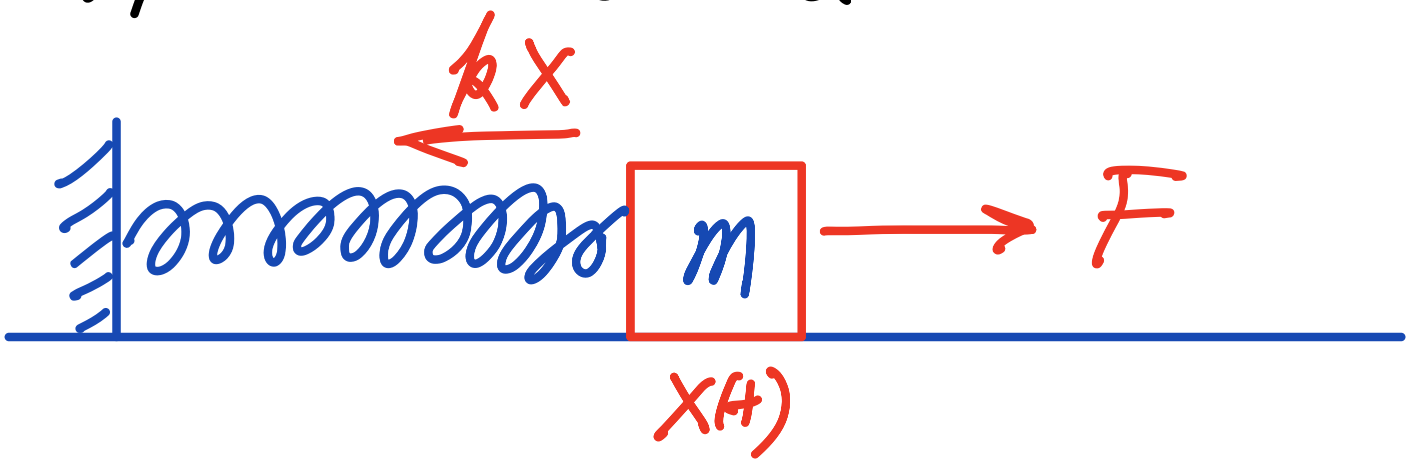
ma

Hooke's Law

friction

Examples (from Physics)

(2) Harmonic Oscillator



$$E = \frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2$$

K.E.

P.E. energy stored in spring

Interpolation Between Hamiltonian and Gradient Flows

$$m\ddot{x} = -\nabla V(x) - \gamma \dot{x}$$

$\gamma \to 0$

$$m\ddot{x} = -\nabla V(x)$$

$$\frac{d}{dt}\left(\frac{1}{2}m|\dot{x}|^2 + V(x)\right) = 0$$

Interpolation Between Hamiltonian and Gradient Flows

$$m\ddot{x} = -\nabla V(x) - \gamma\dot{x}$$

$m \to 0$

$$\gamma\dot{x} = -\nabla V(x)$$

$$\frac{d}{dt}V(x) = -\frac{1}{\gamma}\|\nabla V(x)\|^2 < 0$$

Long Time Behavior of Grad. Flow

$$\begin{cases} \dfrac{dX}{dt} = -\nabla F(X) \\ X(0) = X_0 \end{cases}, \qquad F(X(t)) \downarrow \text{ in } t.$$

Q1: $F(X(t)) \xrightarrow{t \to \infty} \min F$?

Q2: $X(t) \xrightarrow{t \to \infty} X_*$, X_* minimizes F ?

Long Time Behavior of Grad. Flow

$$\begin{cases} \dfrac{dX}{dt} = -\nabla F(X) \\ X(0) = X_0 \end{cases}, \qquad F(X(t)) \downarrow \text{ in } t.$$

Q1: $F(X(t)) \xrightarrow{t \to \infty} \min F$? No!

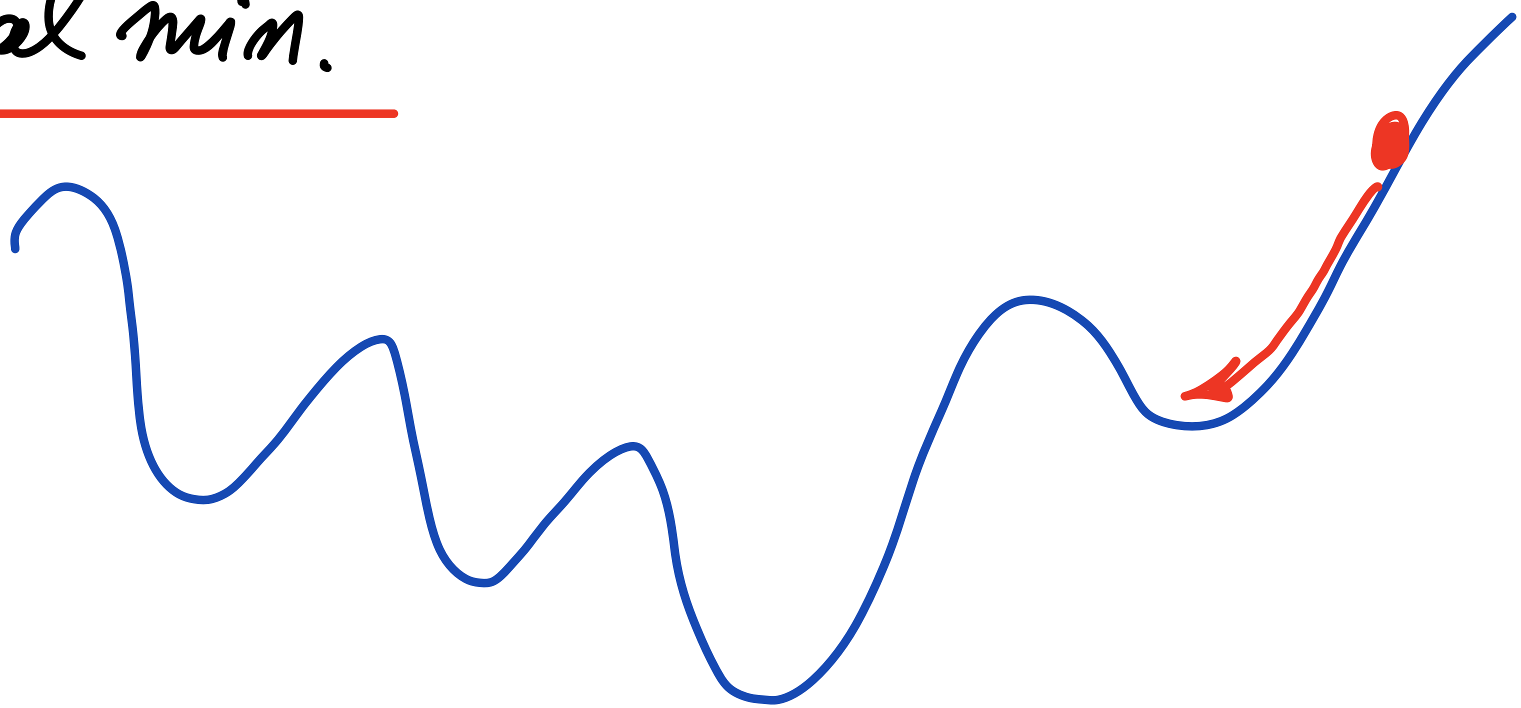
Q2: $X(t) \xrightarrow{t \to \infty} X_*$, X_* minimizes F ? No!

Long Time Behavier of Grad. Flow

Q1: $F(X(t)) \xrightarrow{t\to\infty} \min F$? No!

There might be other critical pts.

Local min.

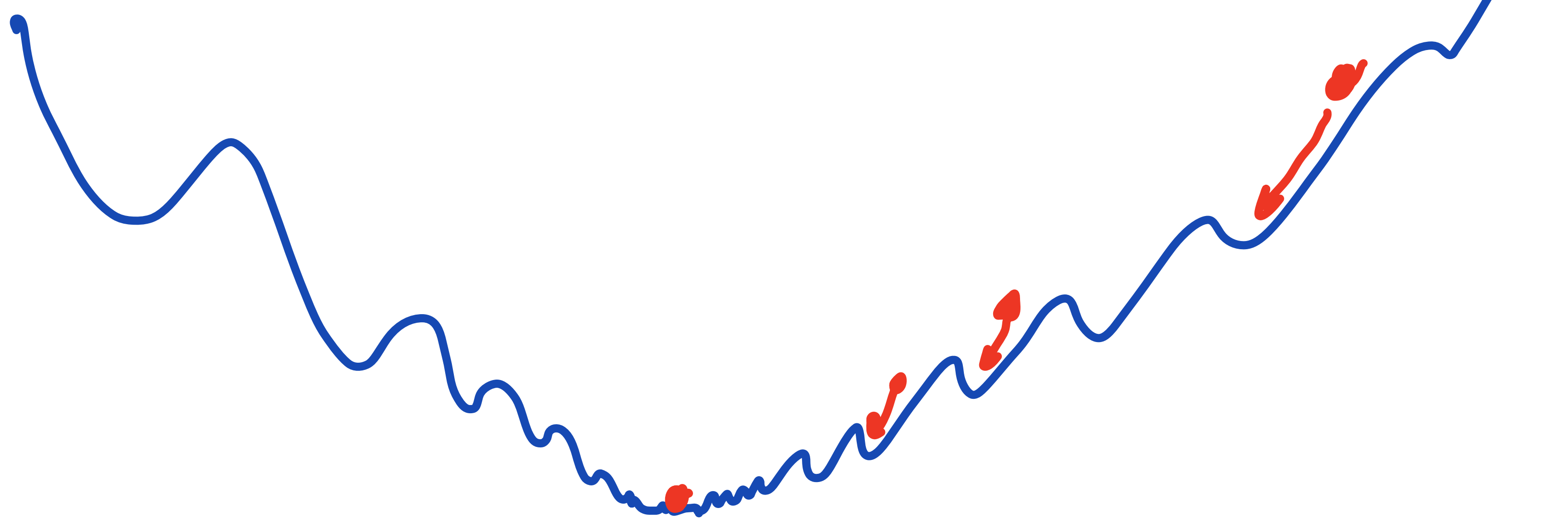


Long Time Behavior of Grad. Flow

Q1: $F(X(t)) \xrightarrow{t \to \infty} \min F$? No!

There might be other critical pts.

Local min.

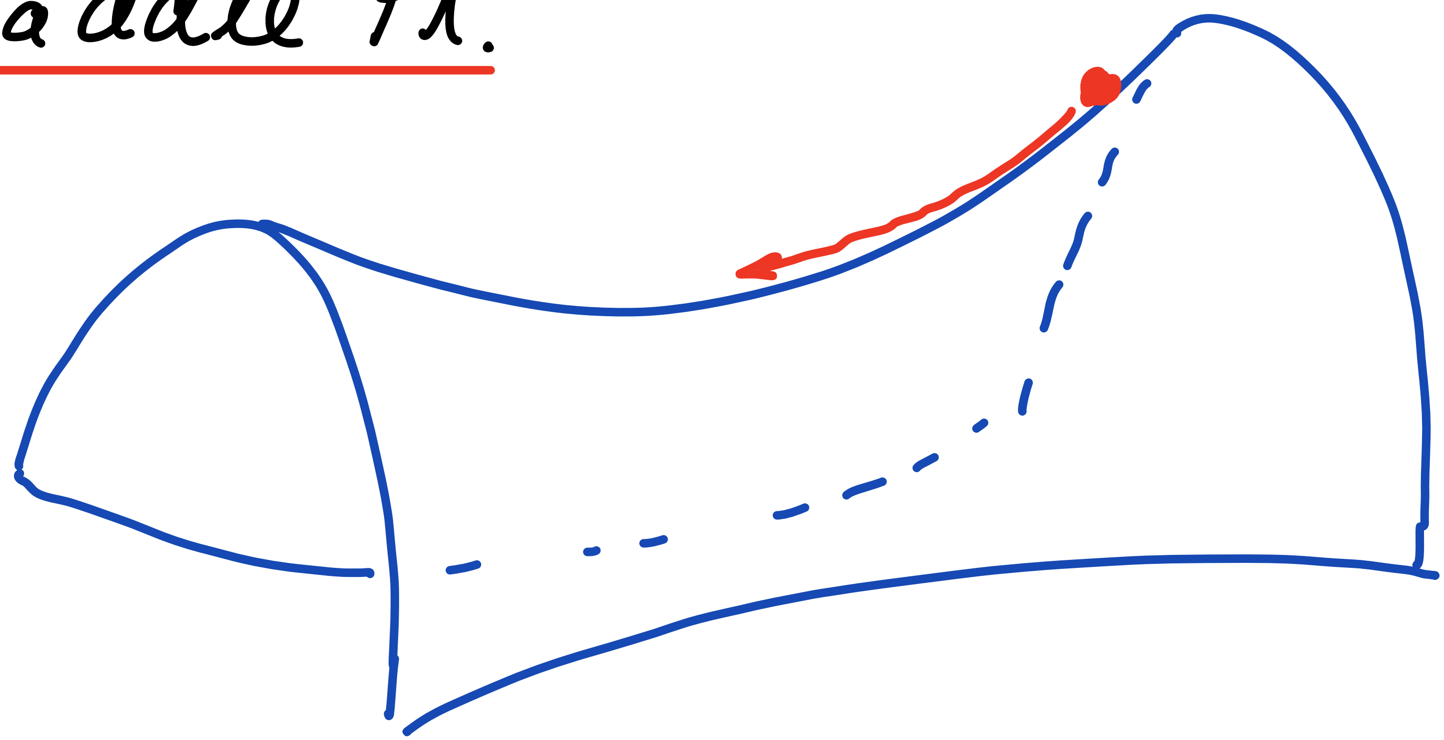


Long Time Behavior of Grad. Flow

Q1: $F(X(t)) \xrightarrow{t\to\infty} \min F$? No!

There might be other critical pts.

Saddle Pt.

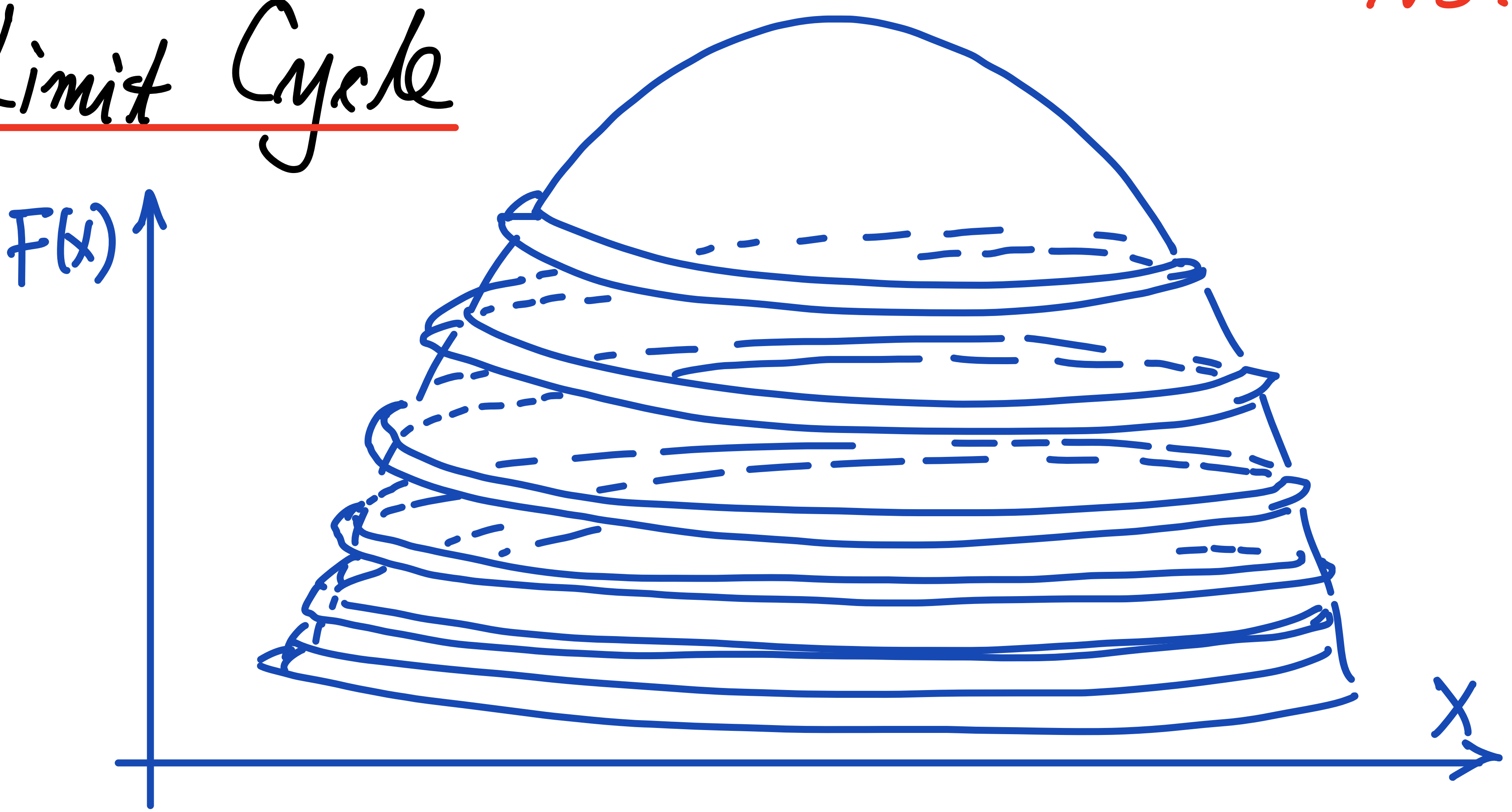


Long Time Behavior of Grad. Flow

Q2: $X(t) \xrightarrow{t\to\infty} X_*$, X_* minimizes F?

No!

Limit Cycle

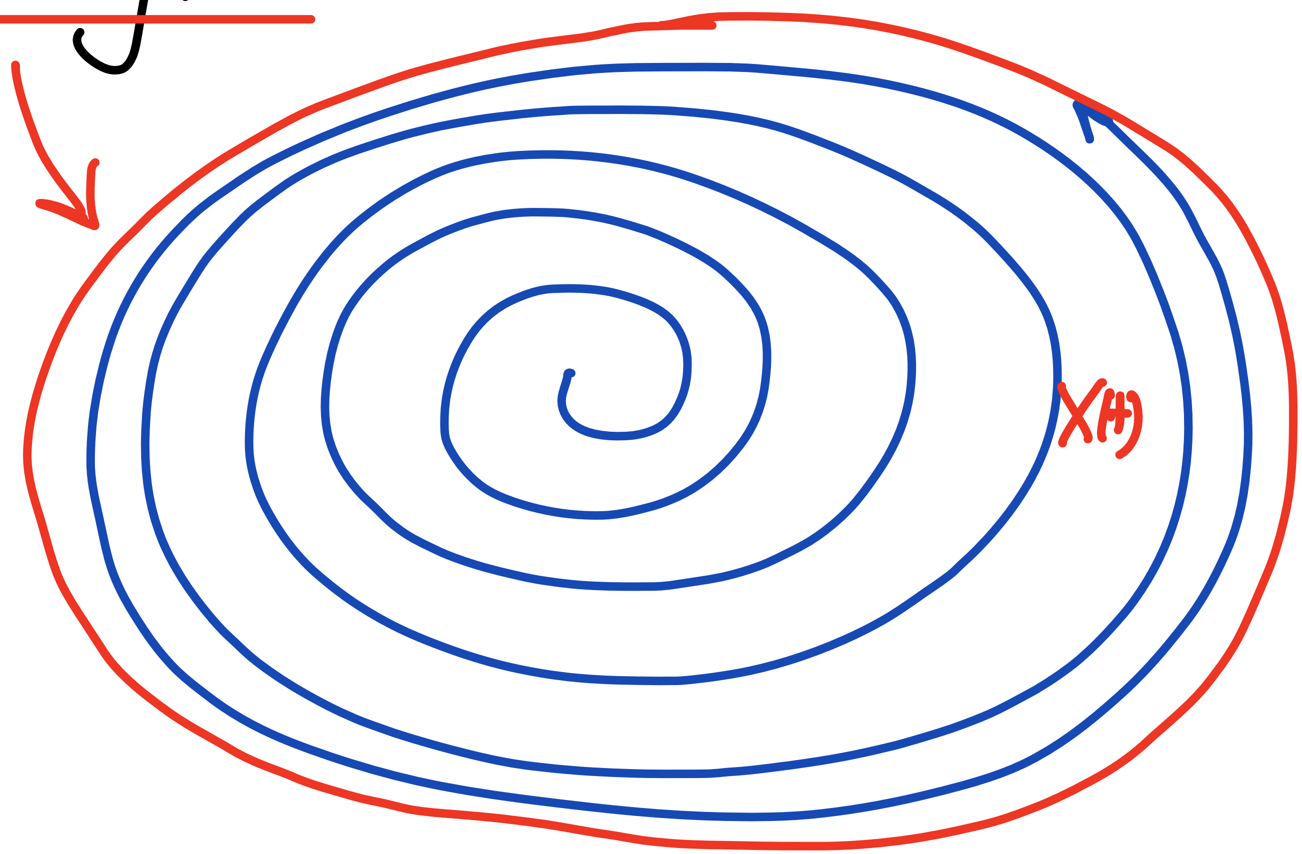


Long Time Behavior of Grad. Flow

Q2: $X(t) \xrightarrow{t \to \infty} X_*$, X_* minimizes F ?

No!

Limit Cycle



Long Time Behavior of Grad. Flow

Q2: $X(t) \xrightarrow{t\to\infty} X_*$, X_* minimizes F?

No!

What is true:

$$\bigcap_{n=1}^{\infty} \overline{\{X(t) : t \geq n\}} \subseteq \{X : \nabla F(X) = 0\}$$

Long Time Behavior of Grad. Flow

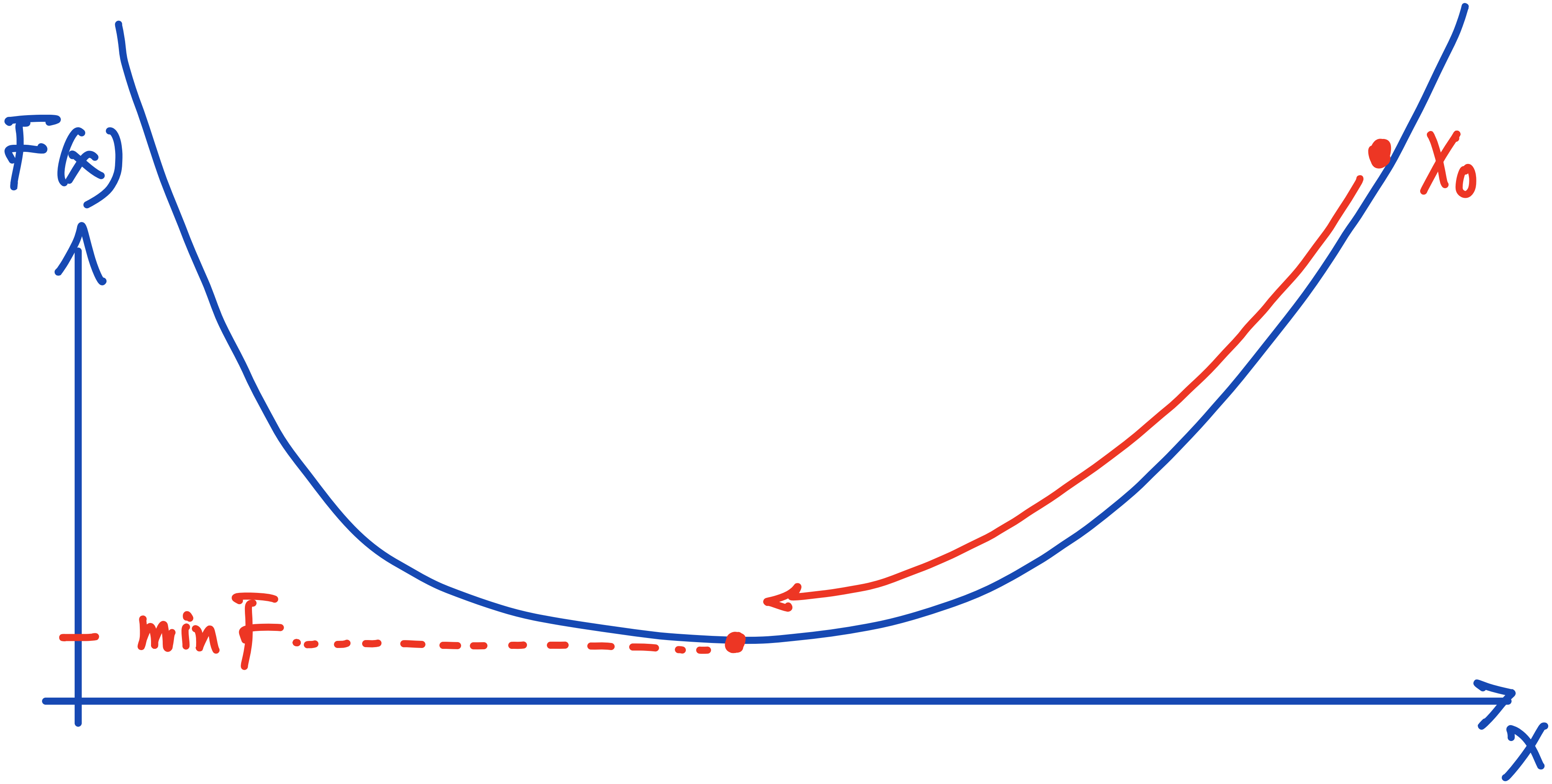
Q2: $X(t) \xrightarrow{t\to\infty} X_*$, X_* minimizes F?

No!

What is true:

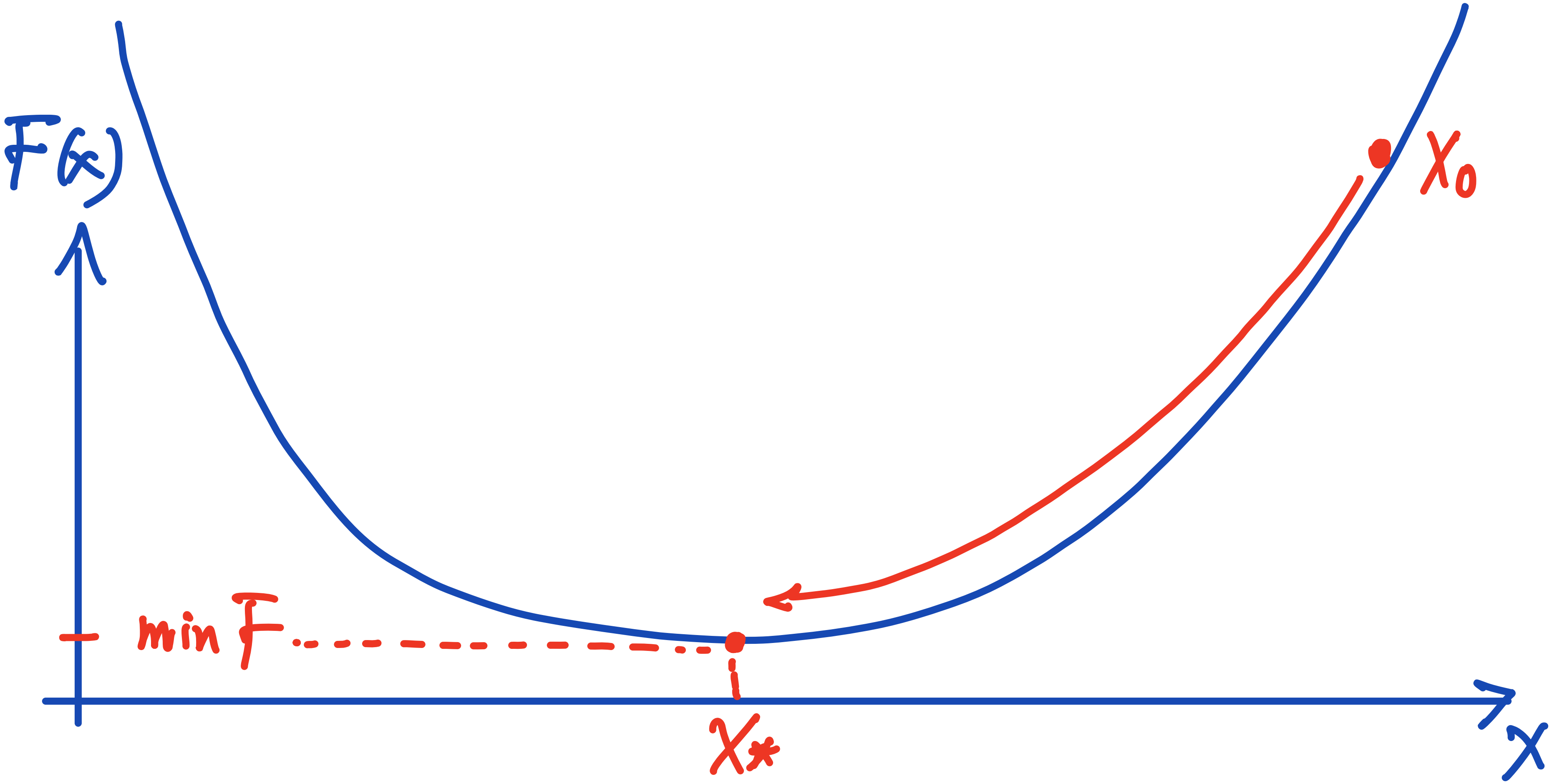
$$\underbrace{\bigcap_{n=1}^{\infty} \overline{\{X(t) : t \geq n\}}}_{\omega\text{-limit set}} \subseteq \{X : \nabla F(X) = 0\}$$

First Useful Simplification: Convex F



$F(X(t)) \longrightarrow \min F$

First Useful Simplification: Convex F



$F(X(t)) \longrightarrow \min F$ Q: How fast?

First Useful Simplification: Convex F

Suppose further, $D^2F(x_*) \geq \lambda I$

(uniformly convex)

Then there is unique x_* that minimizes F and $x(t) \longrightarrow x_*$ exponentially fast.

[$F(x) = \frac{1}{2}\lambda x^2$, $\nabla F(x) = \lambda x$.

$\dot{x} = -\lambda x$, $x(t) = x_0 e^{-\lambda t} \longrightarrow 0$
exp. fast.]

First Useful Simplification: Convex F

For any convex F, $\quad \dot{x} = -\nabla F(x)$,

then
$$\boxed{F(x(t)) - \min F \leq \frac{\|x_0 - x_*\|^2}{2t}}$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

then

$$\boxed{F(x(t)) - \min F \leq \frac{\|x_0 - x_*\|^2}{2t}}$$

simple and dimensional independent.

So

$$\boxed{F(x(t)) - \min F \lesssim O\left(\frac{1}{t}\right)}$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

then
$$\boxed{F(x(t)) - \min F \leq \frac{\|x_0 - x_*\|^2}{2t}}$$

Pf Introduce,

$$E(x(t)) = \frac{\|x(t) - x_*\|^2}{2} + t\left[F(x(t)) - \overset{\min F}{F_*}\right]$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

$$E(X(t)) = \frac{\|X(t) - X_*\|^2}{2} + t\left[F(X(t)) - F_*\right]$$

$$\frac{dE}{dt} = \langle X(t) - X_*, \dot{x} \rangle + \left(F(X(t)) - F_*\right) + t\langle \nabla F(X), \dot{x} \rangle$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

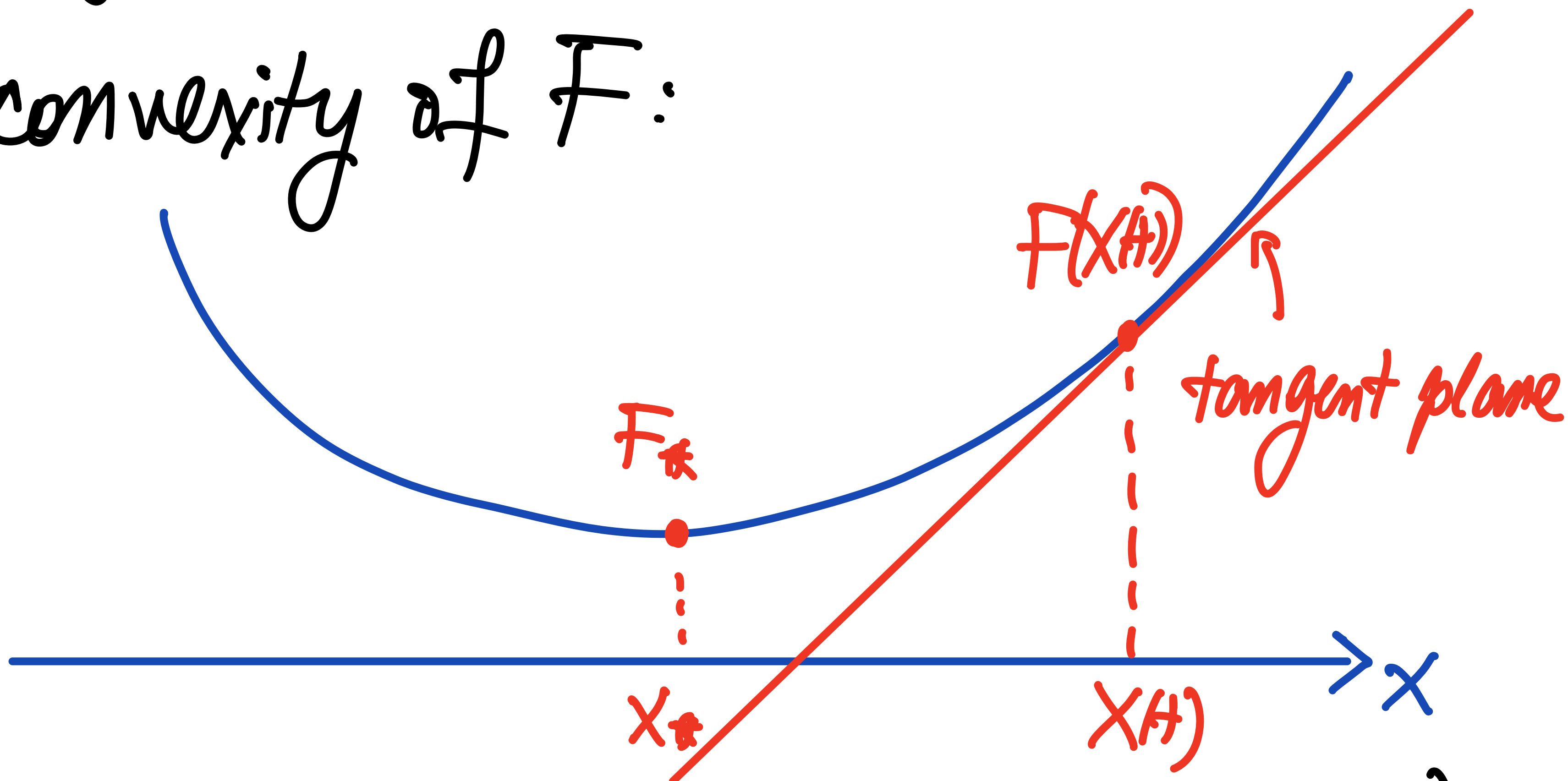
$$\frac{dE}{dt} = \langle X(t) - X_*, \dot{x} \rangle + (F(X(t)) - F_*) + t\langle \nabla F(X), \dot{x} \rangle$$

$$= -\langle X(t) - X_*, \nabla F(X) \rangle + (F(X(t)) - F_*) - t|\nabla F(X)|^2$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

By convexity of F:



$$F_* \geq F(X(t)) + \langle \nabla F(X(t)), X_* - X(t) \rangle$$

First Useful Simplification: Convex F

For any convex F, $\dot{X} = -\nabla F(X)$,

$$F_* \geq F(X(t)) + \langle \nabla F(X(t)), X_* - X(t) \rangle$$

$$\frac{dE}{dt} = -\langle X(t) - X_*, \nabla F(X) \rangle + (F(X(t)) - F_*) - t\,|\nabla F(X)|^2$$

$$\leq 0$$

$$E(X(t)) \leq E(X(0))$$

First Useful Simplification: Convex F

For any convex F, $\dot{x} = -\nabla F(x)$,

$$E(X(t)) \leq E(X(0))$$

$$E(X(t)) = \frac{\|X(t) - X_*\|^2}{2} + t\left[F(X(t)) - F_*\right]$$

$$E(X(0)) = \frac{\|X(0) - X_*\|^2}{2}$$

$$\boxed{F(X(t)) - F_* \leq \frac{\|X(t) - X_*\|^2}{2t}}$$

First Useful Simplification: Convex F

$$\boxed{\ddot{X}(t) = -\frac{3}{t}\dot{X}(t) - \nabla F(X(t))}$$

First Useful Simplification: Convex F

$$\boxed{\ddot{X}(t) = -\frac{3}{t}\dot{X}(t) - \nabla F(X(t))}$$

acceleration

friction (vanishing, $t \to \infty$)

force

$$\boxed{F(X(t)) - F_* \leq \frac{2}{t^2}\|X(0) - X_*\|^2 \lesssim \frac{1}{t^2}}$$

faster convergence

First Useful Simplification: Convex F

$$\boxed{\ddot{X}(t) = -\frac{3}{t}\dot{X}(t) - \nabla F(X(t))}$$

acceleration

friction (vanishing, $t \to \infty$)

force

$$\boxed{F(X(t)) - F_* \leq \frac{2}{t^2}\|X(0) - X_*\|^2 \lesssim \frac{1}{t^2}}$$

accelerated gradient descent

Nesterov Accelerated Grad. Descent

Introduce

$$E(t) = t^2\left(F(X(t)) - F_*\right) + 2\left\| X(t) - X_* + \frac{t\dot{X}}{2} \right\|^2$$

Nesterov Accelerated Grad. Descent

Introduce

$$E(t) = t^2\left(F(X(t)) - F_*\right) + 2\left\| X(t) - X_* + \frac{t\dot{X}}{2} \right\|^2$$

$$\dot{E}(t) = 2t\left(F(X(t)) - F_*\right) + t^2\left\langle \nabla F(X), \dot{X} \right\rangle$$

$$+ 4\left\langle X(t) - X_* + \frac{t\dot{X}}{2},\ \underbrace{\dot{X} + \frac{\dot{X}}{2} + \frac{t\ddot{X}}{2}}_{} \right\rangle$$

$$-\frac{t}{2}\nabla F(X) = \frac{3\dot{X}}{2} + \frac{t\ddot{X}}{2}$$

Nesterov Accelerated Grad. Descent

Introduce

$$E(t) = t^2\left(F(X(t)) - F_*\right) + 2\left\| X(t) - X_* + \frac{t\dot{X}}{2} \right\|^2$$

$$\dot{E}(t) = 2t\left(F(X(t)) - F_*\right) + t^2\left\langle \nabla F(X), \dot{X} \right\rangle$$

$$+ 4\left\langle X(t) - X_* + \frac{t\dot{X}}{2}, \underbrace{\dot{X} + \frac{\dot{X}}{2} + \frac{t\ddot{X}}{2}}_{} \right\rangle$$

$$-\frac{t}{2}\nabla F(X) = \frac{3\dot{X}}{2} + \frac{t\ddot{X}}{2}$$

Nesterov Accelerated Grad. Descent

Introduce

$$E(t) = t^2\left(F(X(t)) - F_*\right) + 2\left\| X(t) - X_* + \frac{t\dot{X}}{2} \right\|^2$$

$$\dot{E}(t) = 2t\left(F(X(t)) - F_*\right) - 2t\langle X(t) - X_*, \nabla F(X)\rangle$$

$$= 2t\Big[\underbrace{F(X(t)) - F_* - \langle X(t) - X_*, \nabla F(X)\rangle}_{\leq 0 \text{ by convexity of } F}\Big]$$

≤ 0

Nesterov Accelerated Grad. Descent

Introduce

$$E(t) = t^2\left(F(X(t)) - F_*\right) + 2\left\| X(t) - X_* + \frac{t\dot{X}}{2} \right\|^2$$

$$E(t) \leq E(0)$$

$$\Rightarrow \boxed{F(X(t)) - F_* \leq \frac{2}{t^2}\left\| X(t) - X_* \right\|^2}$$

Nesterov Accelerated Numerical Scheme

$$\begin{cases} X_{k+1} = Y_k - s\nabla F(Y_k) \\ Y_k = X_k + \left(\frac{k-1}{k+2}\right)(X_k - X_{k-1}) \end{cases}$$

Nesterov Accelerated Numerical Scheme

$$
\begin{cases}
X_{k+1} = Y_k - s\nabla F(Y_k) \\
Y_k = X_k + \left(\dfrac{k-1}{k+2}\right)(X_k - X_{k-1})
\end{cases}
$$

A METHOD OF SOLVING
A CONVEX PROGRAMMING PROBLEM
WITH CONVERGENCE RATE $O(1/k^2)$

YU. E. NESTEROV

0) Select a point $y_0 \in E$. Put

(3) $$k = 0, \quad a_0 = 1, \quad x_{-1} = y_0, \quad \alpha_{-1} = \|y_0 - z\| / \|f'(y_0) - f'(z)\|,$$

where z is an arbitrary point in E, $z \neq y_0$ and $f'(z) \neq f'(y_0)$.

1) *kth iteration*. a) Calculate the smallest index $i \geqslant 0$ for which

(4) $$f(y_k) - f(y_k - 2^{-i}\alpha_{k-1} f'(y_k)) \geqslant 2^{-i-1}\alpha_{k-1}\|f'(y_k)\|^2.$$

b) Put

(5) $$\begin{aligned} \alpha_k &= 2^{-i}\alpha_{k-1}, \quad x_k = y_k - \alpha_k f'(y_k), \\ a_{k+1} &= \left(1 + \sqrt{4a_k^2 + 1}\right)/2, \\ y_{k+1} &= x_k + (a_k - 1)(x_k - x_{k-1})/a_{k+1}. \end{aligned}$$

Nesterov Accelerated Numerical Scheme

$$\begin{cases} X_{k+1} = Y_k - s\nabla F(Y_k) \\ Y_k = X_k + \left(\frac{k-1}{k+2}\right)(X_k - X_{k-1}) \end{cases}$$

$$\frac{\frac{X_{k+1}-X_k}{\sqrt{s}} - \frac{X_k - X_{k-1}}{\sqrt{s}}}{\sqrt{s}} = -\frac{3}{(k+1)\sqrt{s}} \frac{X_k - X_{k-1}}{\sqrt{s}} - \nabla F(Y_k)$$

Nesterov Accelerated Numerical Scheme

$$\begin{cases} X_{k+1} = Y_k - s\nabla F(Y_k) \\ Y_k = X_k + \left(\frac{k-1}{k+2}\right)(X_k - X_{k-1}) \end{cases}$$

$$\frac{\frac{X_{k+1}-X_k}{\sqrt{s}} - \frac{X_k - X_{k-1}}{\sqrt{s}}}{\sqrt{s}} = -\frac{3}{(k+1)\sqrt{s}}\,\frac{X_k - X_{k-1}}{\sqrt{s}} - \nabla F(Y_k)$$

$s \to 0$

$$\ddot{X} = -\frac{3}{t}\dot{X} - \nabla F(X)$$

Nemirovskii-Yudin Complexity

"First-Order Convex Optimization"

$$\boxed{\frac{1}{k^2} \lesssim F(X_k) - F_*}$$

Nemirovskii–Yudin Complexity

"First-Order Convex Optimization"

$$\boxed{\frac{1}{k^2} \lesssim F(X_k) - F_*}$$

achieved by Nesterov Scheme

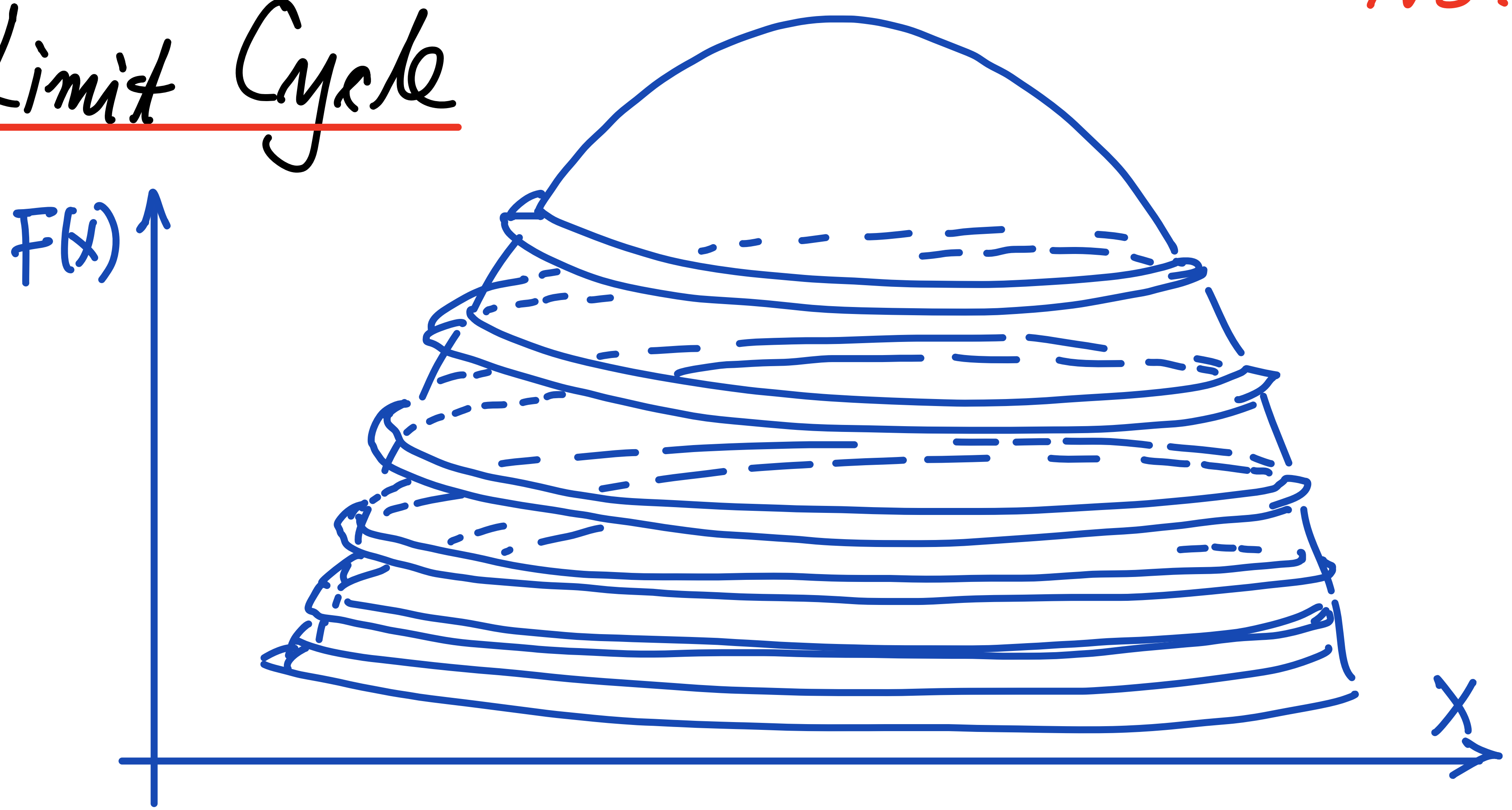
$\lesssim \frac{1}{k}$ (Gradient Flow)

Long Time Behavior of Grad. Flow

Q2: $X(t) \xrightarrow{t\to\infty} X_*$, X_* minimizes F?

No!

Limit Cycle

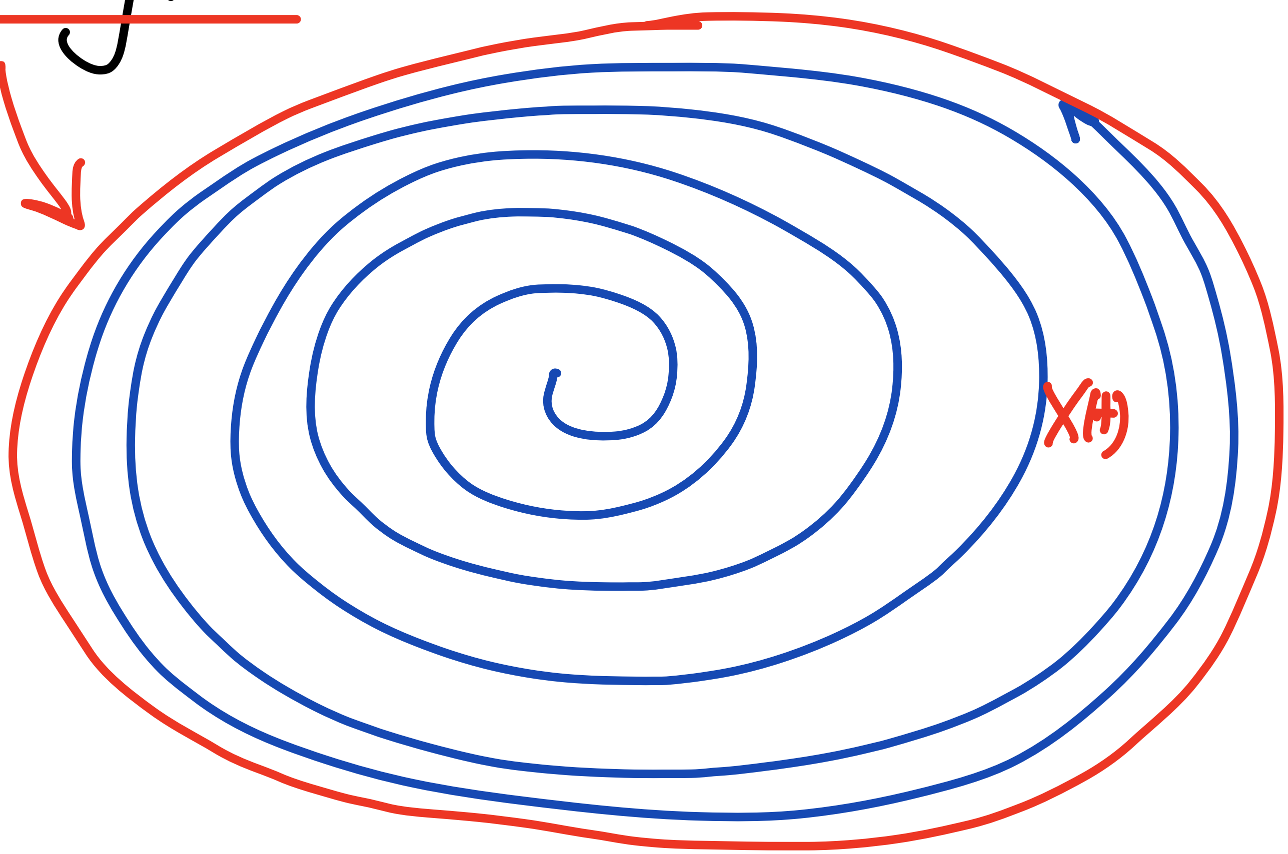


Long Time Behavior of Grad. Flow

Q2: $X(t) \xrightarrow{t\to\infty} X_*$, X_* minimizes F ?

No!

Limit Cycle



Second Useful Simplification

(Lojasiewicz - Simon Inequality)

For any X_* s.t. $\nabla F(X_*) = 0$, X close to X_*,

$$|F(x) - F(X_*)|^{1-\theta} \leq C\|\nabla F(X)\|$$

$\theta \in (0, \frac{1}{2}]$

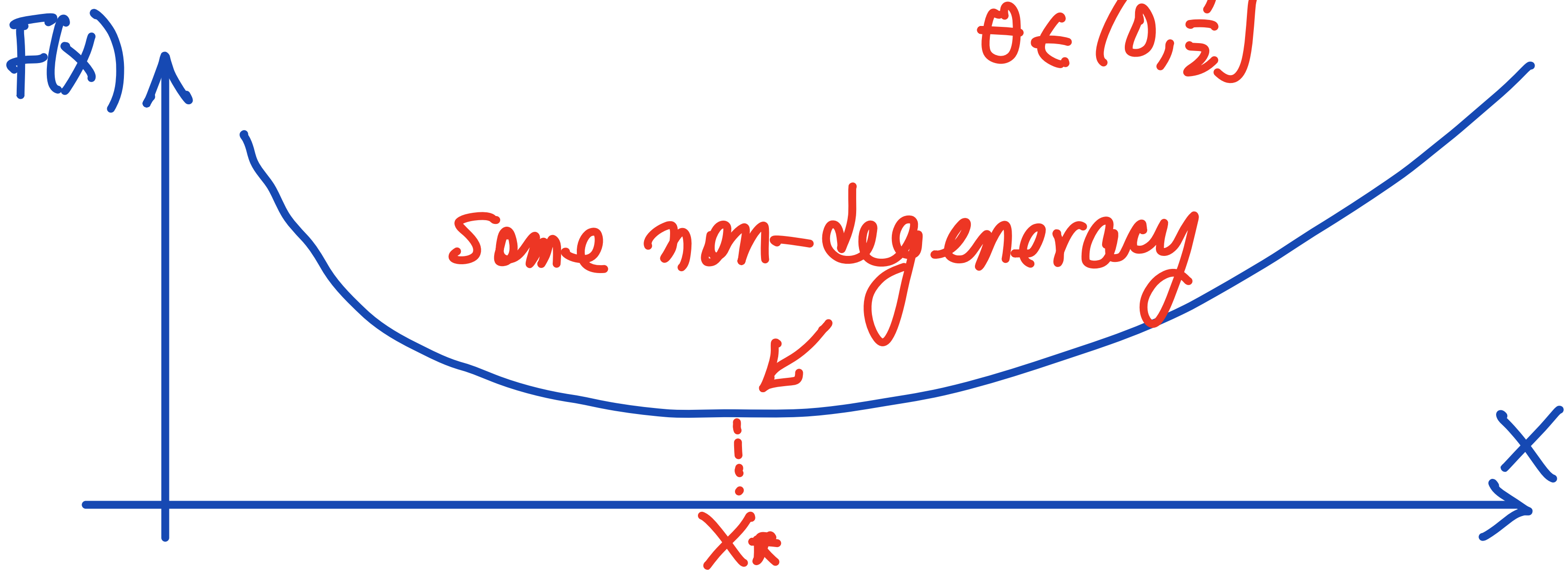
Second Useful Simplification

(Łojasiewicz - Simon Inequality)

For any X_* s.t. $\nabla F(X_*) = 0$, X close to X_*,

$$|F(x) - F(X_*)|^{1-\theta} \leq C\|\nabla F(X)\|$$

$\theta \in (0, \frac{1}{2}]$



Second Useful Simplification

(Lojasiewicz - Simon Inequality)

$$|F(x) - F(x_*)|^{1-\theta} \leq C \|\nabla F(x)\|$$

For $\dot{x} = -\nabla F(x)$,

there is an x_*, $\nabla F(x_*) = 0$ s.t.

$$x(t) \longrightarrow x_* \text{ as } t \to +\infty$$

Second Useful Simplification

(Lojasiewicz - Simon Inequality)

$$|F(x) - F(x_*)|^{1-\theta} \leq C \|\nabla F(x)\|$$

$$\|x(t) - x_*\| = \begin{cases} O(e^{-ct}) & \theta = \frac{1}{2} \\ O\left(t^{-\frac{\theta}{1-2\theta}}\right) & 0 < \theta < \frac{1}{2} \end{cases}$$

Łojasiewicz Inequality

(from "real" algebraic geom.)

Let $f: U \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be a real analytic function.

① $\inf_{z:\, f(z)=0} |x - z|^\alpha \leq C\,|f(x)|$

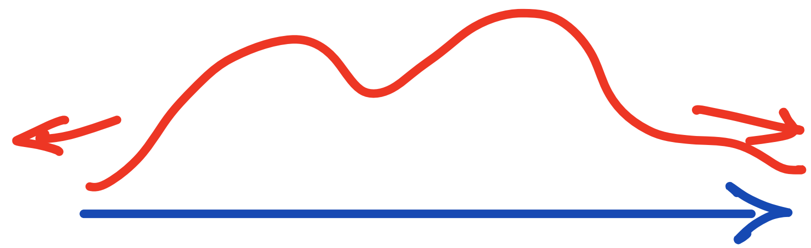
zero set of f

② $|f(x) - f(p)|^\beta \leq C\,|\nabla_x f(x)|$

Examples from Partial Diff. Eqns (PDE)

① Heat Equation

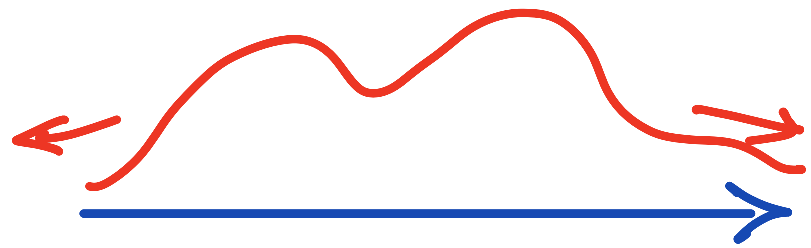
$$u_t = \Delta u$$



Examples from Partial Diff. Eqns (PDE)

① Heat Equation

$$u_t = \Delta u$$



$$\Delta u = \partial_{x_1}^2 u + \partial_{x_2}^2 u + \dots + \partial_{x_n}^2 u$$

$$= \operatorname{div}(\nabla u)$$

$$\boxed{u_t = \operatorname{div}(\nabla u)}$$

Heat Equation $\quad u_t = \operatorname{div}(\nabla u)$

"Energy" of the solution

L^2 norm

$$F(u) = \int \frac{1}{2} u^2 \, dx, \qquad u = u(x,t)$$

$$\frac{d}{dt} F(u) = \frac{d}{dt} \int \frac{1}{2} u^2 dx = \int u \, u_t \, dx$$

$$= \int u \operatorname{div} \nabla u \, dx$$

Heat Equation $\quad u_t = \operatorname{div}(\nabla u)$

"Energy" of the solution

$$F(u) = \int \frac{1}{2} u^2 \, dx, \qquad u = u(x,t)$$

$$\frac{d}{dt} F(u) = \frac{d}{dt} \int \frac{1}{2} u^2 dx = \int u \, u_t \, dx$$

$$= \int u \operatorname{div} \nabla u \, dx = -\int |\nabla u|^2 dx$$

≤ 0

Heat Equation $\quad u_t = \operatorname{div}(\nabla u)$

"Energy" of the solution

Dirichlet norm

$$F(u) = \int \frac{1}{2} |\nabla u|^2 dx, \qquad u = u(x,t)$$

$$\frac{d}{dt} F(u) = \frac{d}{dt} \int \frac{1}{2} \langle \nabla u, \nabla u \rangle = \int \langle \nabla u, \nabla u_t \rangle$$

Heat Equation $\quad u_t = \mathrm{div}(\nabla u)$

"Energy" of the solution

$$F(u) = \int \frac{1}{2} |\nabla u|^2 dx, \quad u = u(x,t)$$

$$\frac{d}{dt} F(u) = \frac{d}{dt} \int \frac{1}{2} \langle \nabla u, \nabla u \rangle = \int \langle \nabla u, \nabla u_t \rangle$$

$$= \int -(\Delta u)\, u_t = -\int (\Delta u)^2 \leq 0$$

Heat Equation $\quad u_t = \operatorname{div}(\nabla u)$

"Energy" of the solution

Entropy $(u \geq 0)$

$$F(u) = \int u \log u \, dx, \qquad u = u(x,t)$$

$$\frac{d}{dt} F(u) = \frac{d}{dt} \int u \log u \, dx = \int (1 + \log u)\, u_t$$

$$= \int (1 + \log u) \operatorname{div} \nabla u$$

Heat Equation $\quad u_t = \mathrm{div}(\nabla u)$

"Energy" of the solution

Entropy ($u \geq 0$)

$$F(u) = \int u \log u \, dx, \qquad u = u(x,t)$$

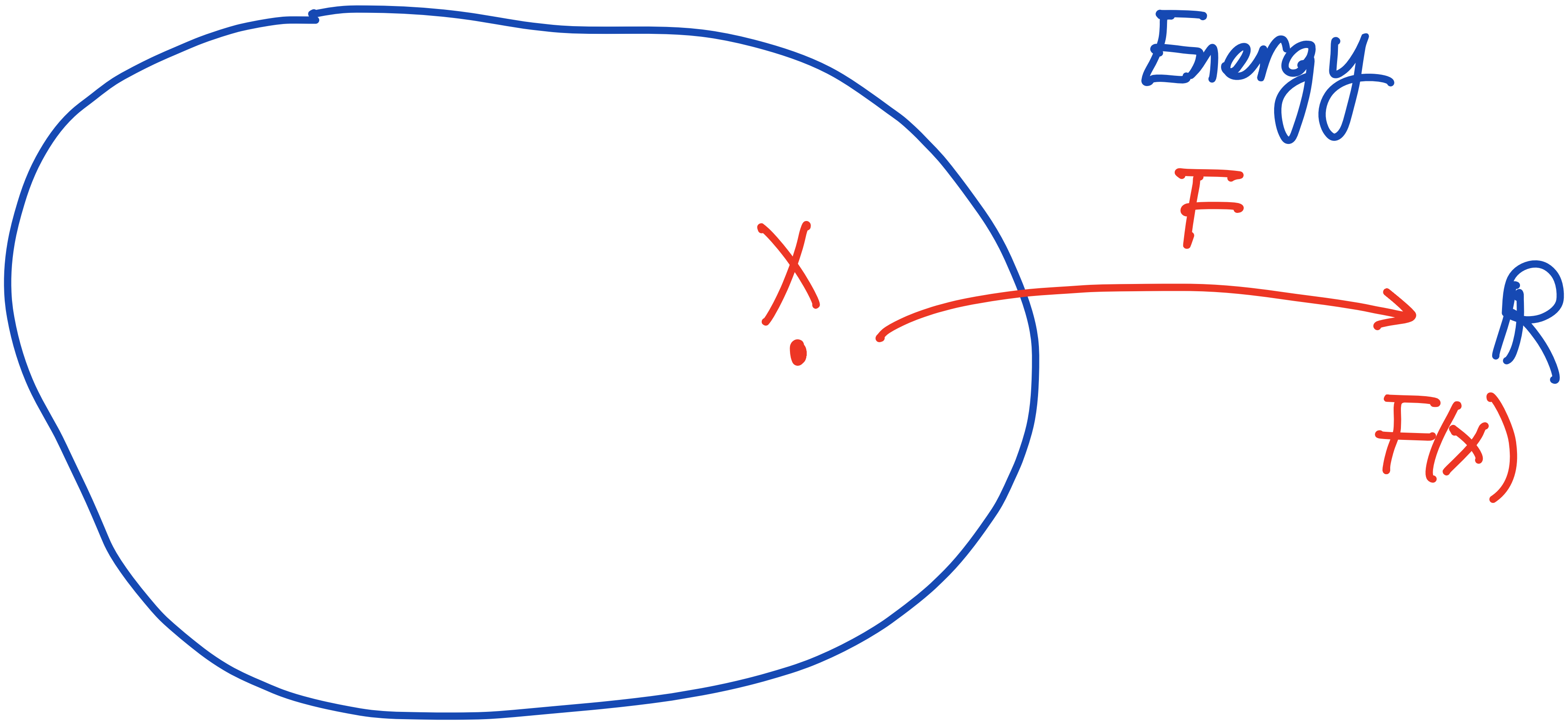
$$\frac{d}{dt} F(u) = \frac{d}{dt}\int u \log u \, dx = \int (1+\log u)\, u_t$$

$$= \int (1+\log u)\, \mathrm{div} \nabla u = -\int \frac{|\nabla u|^2}{u}$$

Fisher Information < 0

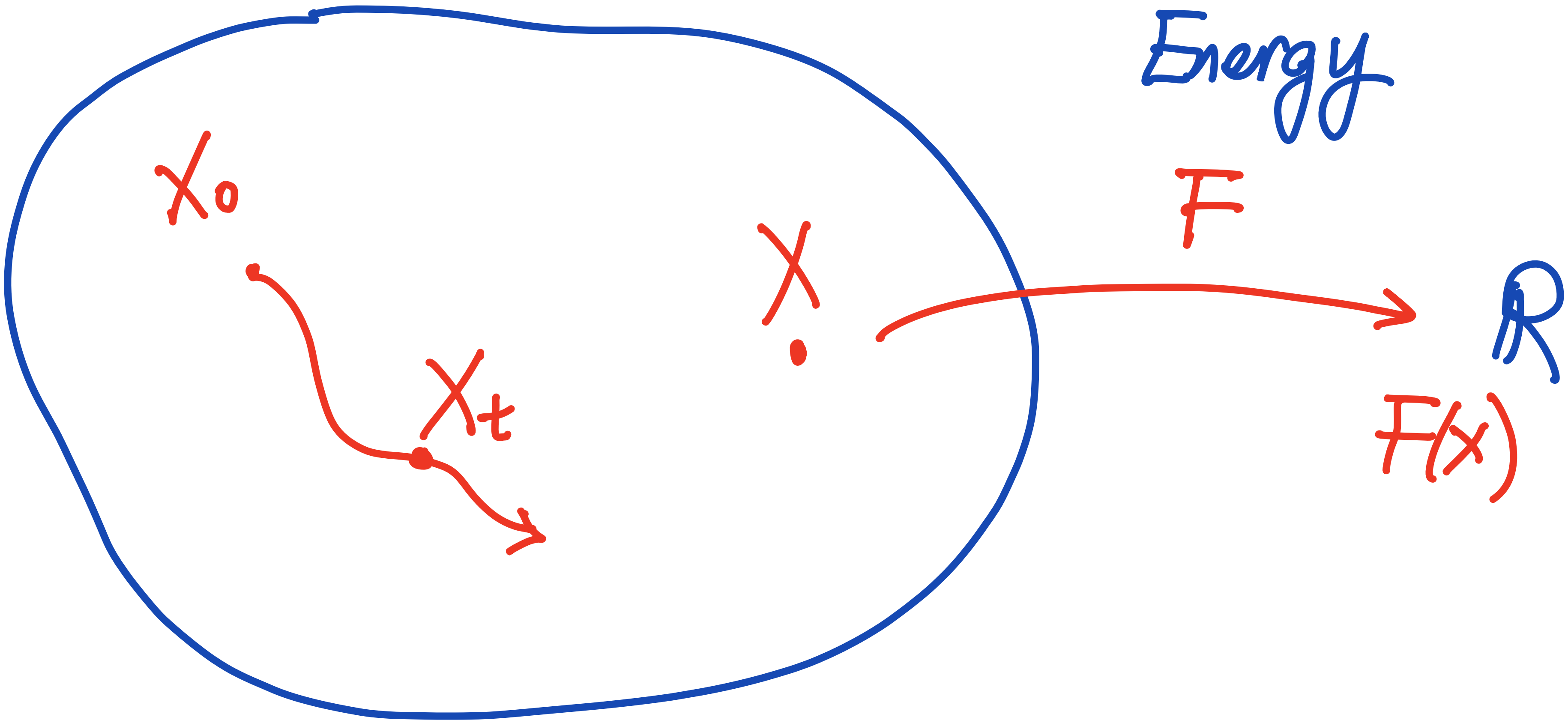
Abstract Formulation of Grad. Flows

X = space of functions



Abstract Formulation of Grad. Flows

X = space of functions



Abstract Formulation of Grad. Flows

$$\frac{d}{dt} F(X_t) = \left(\delta F(X_t), \frac{dX_t}{dt} \right)$$

1st variation of F

Abstract Formulation of Grad. Flows

$$\frac{d}{dt}F(X_t) = \left(\delta F(X_t), \frac{dX_t}{dt}\right)$$

1st variation of F

$$= \left\langle \nabla F(X_t), \frac{dX}{dt} \right\rangle$$

Gradient w.r.t. $\langle \cdot, \cdot \rangle$

inner product on TX

Abstract Formulation of Grad. Flows

$$\frac{d}{dt}F(X_t) = \left\langle \nabla F(X_t), \frac{dX}{dt} \right\rangle$$

$$\frac{dX}{dt} = -\nabla F(X)$$

$$= -|\nabla F(X_t)|^2 < 0$$

By appropriately choosing F, $\langle \cdot, \cdot \rangle$ (and hence ∇F) can model a wide range of physical phenomena.

Thank You !