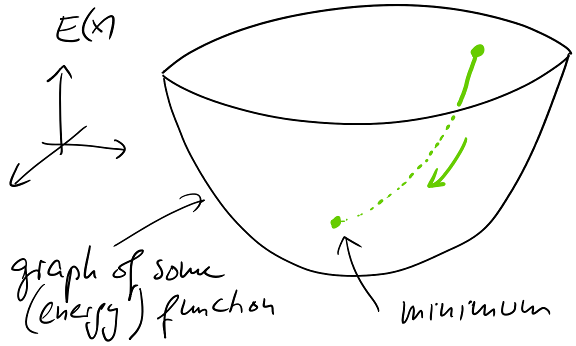
Gradient flows:

- special class of ordinary differential equations
- steepest descent

$t \mapsto x(t)$ curve

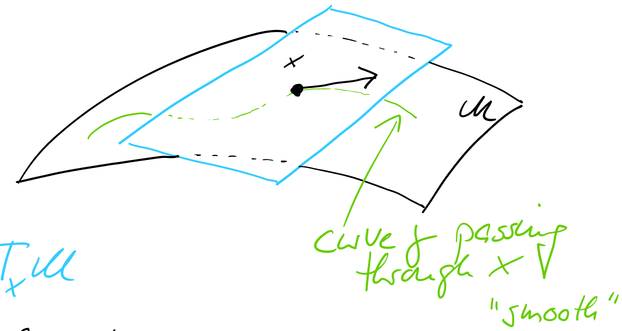
s.t. $\dot{x}(t)$ points in the direction where E decreases at maximal rate



Ingredients

(1.) manifold $\mathcal{M}$

$\hookrightarrow$ tangent space $T_x\mathcal{M}$



vector space of velocity vectors $\dot{\gamma}(0)$ for curves $t \mapsto \gamma(t) \in \mathcal{M}$ with $\gamma(0) = x$

(one possible definition, there are others ...)

(2) function $E : \mathcal{M} \longrightarrow \mathbb{R}$ "smooth"

(energy, entropy ...)

$\hookrightarrow$ differential

$$dE_x : T_x\mathcal{M} \longrightarrow \mathbb{R} \quad , \; x \in \mathcal{M}$$

continuous linear map from tangent space $T_x\mathcal{M}$ to $\mathbb{R}$

$(T_x\mathcal{M})^*$ dual space

Problem $\quad \dot{x}(t) \in T_{x(t)} \mathcal{M} \quad , \quad dE_{x(t)} \in \left(T_{x(t)} \mathcal{M}\right)^*$

different spaces!

(3.) Riemannian metric $\langle \cdot, \cdot \rangle_x$

inner product on $T_x\mathcal{M}$, $x \in \mathcal{M}$

Riesz representation theorem:

There exists a **unique** $\nabla E(x) \in T_x\mathcal{M}$ (tangent vector) with the property that

$$dE_x(v) = \langle \nabla E(x), v \rangle_x \quad \text{for all } v \in T_x\mathcal{M},\ x \in \mathcal{M}$$

this is the gradient of E at x

(One tends to forget about the significance of $\langle \cdot, \cdot \rangle_x$ because one is used to think of Euclidean norm in $\mathbb{R}^n$...)

Note The metric (= local geometry) plays a crucial role.

$\hookrightarrow$ gradient flow: a curve $t \mapsto \gamma(t)$ s.t.

$$\boxed{\dot{\gamma}(t) = -\nabla E(\gamma(t)), \quad t \in \text{Interval}}$$

Some properties:

- $t \mapsto E(\gamma(t))$ decreasing

Γ
$$\frac{d}{dt}E(\gamma(t)) = dE_{\gamma(t)}\left(\dot{\gamma}(t)\right)$$

γ gradient flow

$$= \left\langle \nabla E(\gamma(t)), -\nabla E(\gamma(t)) \right\rangle_{\gamma(t)}$$

definition of gradient

$$= - \underbrace{\left| \nabla E(\gamma(t)) \right|^2}_{\geq 0} \leq 0$$

non degeneracy of scalar product $\lrcorner$

- Convergence to local minimum:

  we have $\dot{\gamma}(t) = 0$ iff $\nabla E(\gamma(t)) = 0$

- for convex E:

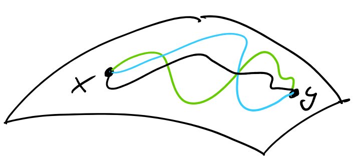
  the local minimum is the (unique) global minimum

$\boxed{\text{There is more}}$

E (uniformly) convex $\Longrightarrow$ contraction / convergence rates

Γ How to probe convexity?

$\boxed{I}$ along geodesics ← shortest curves between points:



Distance between x, y:

$$d(x,y)^2 = \inf \left\{ \int_0^1 \left| \dot{\gamma}(t) \right|^2_{\gamma(t)} dt : \begin{array}{l} t \mapsto \gamma(t) \in \mathcal{M} \\ \gamma(0) = x \\ \gamma(1) = y \end{array} \right\}$$

Riemannian metric

Under suitable assumption
there exists a minimizing curve (= geodesic).
It is characterized by the condition

$$\boxed{\frac{D}{dt}\dot{\gamma}(t) = 0}$$

Covariant derivative

(takes into account that the tangent spaces $T_{\gamma(t)}\mathcal{M}$ change with t!)

$\boxed{\text{Then}}$ $E : \mathcal{M} \longrightarrow \mathbb{R}$ is (geodesically) convex

if $t \longmapsto E(\gamma(t))$ convex

along $\underline{\underline{\text{all}}}$ geodesics $t \mapsto \gamma(t)$.

Recall that this is equivalent (for smooth setting) to

$$\frac{d^2}{dt^2} E(\gamma(t)) \geq 0$$

or to

Hess $E_{\gamma(t)}$ positive semidefinite

(i.e., $\langle v, (\text{Hess}\, E_{\gamma(t)}) v \rangle_{\gamma(t)} \geq 0$ for all $v \in T_{\gamma(t)}\mathcal{M}$

linear map: $T_{\gamma(t)}\mathcal{M}$ to itself.

[Proof: By chain rule

$$\frac{d^2}{dt^2} E(\gamma(t)) = \frac{d}{dt} \left\langle \nabla E(\gamma(t)), \dot\gamma(t) \right\rangle_{\gamma(t)}$$

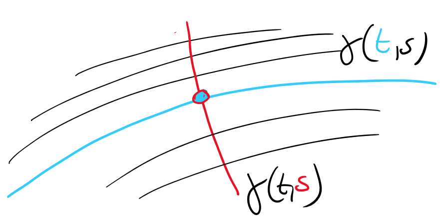
$$= \left\langle \underbrace{\frac{D}{dt} \nabla E(\gamma(t))}_{= (\mathrm{Hess}\, E_{\gamma(t)}) \dot\gamma(t)}, \dot\gamma(t) \right\rangle_{\gamma(t)}$$

$$+ \cancel{\left\langle \nabla E(\gamma(t)), \underbrace{\frac{D}{dt} \dot\gamma(t)}_{=0 \text{ for geodesics!}} \right\rangle_{\gamma(t)}}$$

need to know the geodesics!

$$= \left\langle (\mathrm{Hess}\, E_{\gamma(t)}) \dot\gamma(t), \dot\gamma(t) \right\rangle_{\gamma(t)}$$

$\boxed{\text{II}}$ Alternative way to probe convexity:



consider one-parameter family of gradient flow trajectories

$$\dot\gamma(t,s) = -\nabla E(\gamma(t,s))$$

parameter

$\Longrightarrow$ The perturbation

$$\delta\gamma_t := \frac{d}{ds}\Big|_{s=0} \gamma(t,s)$$

satisfies also a differential equation:

$$\frac{D}{dt} \frac{d}{ds} \gamma(t,s) = \frac{D}{ds} \underbrace{\frac{d}{dt} \gamma(t,s)}_{= \dot\gamma(t,s)}$$

torsion-free metric

$$= \frac{D}{ds} \left(-\nabla E(\gamma(t,s)) \right)$$

gradient flow

$$= -\left(\mathrm{Hess}\, E_{\gamma(t,s)}\right) \underbrace{\frac{d}{ds} \gamma(t,s)}_{\text{derivative wrt. } s!}$$

Thus:

$$\frac{D}{dt} \delta\gamma(t) = -\left(\mathrm{Hess}\, E_{\gamma(t)}\right) \delta\gamma(t)$$

$\gamma(t,0)$

In particular, this implies

$$\frac{d}{dt}\,\frac{1}{2}\left|\delta\gamma(t)\right|^2_{\gamma(t)} = -\left\langle \delta\gamma(t), \left(\mathrm{Hess}\,E_{\gamma(t)}\right)\delta\gamma(t)\right\rangle_{\gamma(t)}$$

≤ 0 if $\mathrm{Hess}\,E_{\gamma(t)} \geq 0$
(i.e., E convex)

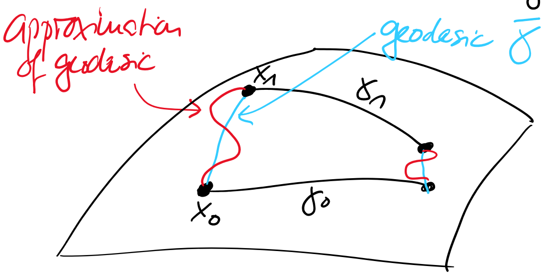
$\Longrightarrow$ Perturbations along trajectories are damped.

potentially much easier to check

(no need for geodesics)

This implies a $\boxed{\text{contraction estimate}}$

consider two gradient flow trajectories



$$\dot\gamma_i(t) = -\nabla E(\gamma_i(t)), \quad i=1..2$$

$$\begin{cases} \gamma_1(0) = x_1 \\ \gamma_2(0) = x_2 \end{cases}$$

The distance

Perturbation along trajectories

$$d(x_1,x_2)^2 = \inf\left\{ \int_0^1 \left|\frac{d}{ds}\gamma(t=0,s)\right|^2 ds : \quad s \mapsto \gamma(t=0,s) \text{ curve connecting } \gamma(0,s=0)=x_1,\ \gamma(0,s=1)=x_2 \right\}$$

$$= \int_0^1 \left|\frac{d}{ds}\bar\gamma(s)\right|^2 ds$$

geodesic connecting x_1, x_2

Therefore we can estimate

$$d(\gamma(t,s=0),\gamma(t,s=1))^2$$

distance = inf ...

$$\leq \int_0^1 \left|\frac{d}{ds}\gamma(t,s)\right|^2 ds$$

nonpositive

$$= \int_0^t\int_0^1 \underbrace{\frac{d}{dt}\left|\frac{d}{ds}\gamma(t,s)\right|^2}_{\leq 0 \text{ if } E \text{ convex}} ds\, dt$$

$$+ \int_0^1 \left|\frac{d}{ds}\gamma(t=0,s)\right|^2 ds$$

approximates $d(x_1,x_2)^2$ for suitable γ

This implies:

$$\boxed{d(\gamma_0(t),\gamma_1(t)) \leq d(x_0,x_1)}$$

contraction

Note: If E is uniformly convex in the sense that

$$\langle v,(\mathrm{Hess}\, E_x)v\rangle_x \geq \alpha\, |v|_x^2 \quad \text{for } x\in M,\ v\in T_xM$$

and some number $\alpha\in\mathbb{R}$, then by Gronwall's Lemma

$$\boxed{d(\gamma_0(t),\gamma_1(t)) \leq e^{-\frac{\alpha}{2}t}\, d(x_0,x_1)}$$

Importance of local geometry:

Riemannian metric $\langle\cdot,\cdot\rangle$

→ Distance on M (geodesics)

$$d(x_0,x_1) = \inf\left\{\int_0^1 |\dot\gamma(t)|^2 dt : \gamma:[0,1]\to M,\ \gamma(0)=x_0,\ \gamma(1)=x_1\right\}$$

→ Gradient from Differential

$$dE_x(v) =: \langle\nabla E(x),v\rangle_x, \quad v\in T_xM$$

Application: Porous Medium Equation

$$\boxed{\partial_t \rho - \Delta P(\rho) = 0 \quad \text{in } [0,\infty) \times \mathbb{R}^d}, \quad \rho(t=0,\cdot) = \bar{\rho}$$

suitable initial data

Density: $\rho = \rho(t,x)$

$$\rho(t,\cdot) \in \mathcal{P}_2(\mathbb{R}^d)$$

$$= \Big\{ \mu \text{ probability measures on } \mathbb{R}^d, \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty \Big\}$$

finite second moments

with $P: [0,\infty) \to \mathbb{R}$ some function (see below).

Example: $P(r) = r^m$, for suitable $m \in \mathbb{R}$.

$\boxed{\text{Claim}}$ The porous medium equation can be understood as a (generalized) gradient flow [Otto]

① **Manifold** $\mathcal{M} = \mathcal{P}_2(\mathbb{R}^d)$

Q: what is the tangent space at some point $x \in \mathcal{M}$ $\iff$ $\rho \in \mathcal{P}_2(\mathbb{R}^d)$?

$\rightsquigarrow$ consider curve $t \mapsto \rho_t$ in $\mathcal{P}_2(\mathbb{R}^d)$ and <u>differentiate it</u>:

$$\partial_t \rho_t = \lim_{h \to 0} \frac{1}{h} (\rho_{t+h} - \rho_t)$$

distribution in $\mathbb{R}^d$

$$= -\nabla \cdot (\rho_t u_t), \quad u_t \in L^2(\mathbb{R}^d, \rho_t)$$

"integral" zero by Gauss-Green

Eulerian <u>velocity</u> field

$\iff$ total mass of ρ_t is constant in time

Therefore: identify tangent vectors to $\mathcal{P}_2(\mathbb{R}^d)$ in ρ_t with vector fields $u_t \in L^2(\mathbb{R}^d, \rho_t)$

square integrable

square integrable w.r.t. ρ_t

Problem: Not unique!

If $w \in L^2(\mathbb{R}^d, \rho_t)$ satisfies $\nabla \cdot (\rho_t w) = 0$ in distributional sense, then

$$\nabla \cdot (\rho(u_t + w)) = \nabla \cdot (\rho u_t)$$

Solution:

- Note first that $L^2(\mathbb{R}^d, \rho_t)$ induces an <u>inner product</u>:

$$\langle v, w \rangle_x \quad , \quad v, w \in T_x \mathcal{M}$$

$$\Updownarrow$$

$$\int_{\mathbb{R}^d} \rho \, v \cdot w \quad , \quad v, w \in L^2(\mathbb{R}^d, \rho)$$

$\rightsquigarrow$ Induced norm $\int_{\mathbb{R}^d} \rho |v|^2$

- Consider $u_t \in L^2(\mathbb{R}^d, \rho_t)$ with minimal length

$$\int_{\mathbb{R}^d} \rho_t |u_t|^2 = \underset{\text{attained}}{\inf} \left\{ \int_{\mathbb{R}^d} \rho_t |u_t + w|^2 : w \in L^2(\mathbb{R}^d, \rho_t), \ \nabla \cdot (\rho_t w) = 0 \right\}$$

$\rightsquigarrow$ The optimal u_t must be orthogonal to all w as above

$$\int_{\mathbb{R}^d} \rho_t u_t \cdot w = 0 \quad \text{if} \quad \nabla \cdot (\rho_t w) = 0$$

gradients!

closure

$$\Longrightarrow \quad u_t \in \overline{\{ \nabla \phi : \phi \in \mathcal{D}(\mathbb{R}^d) \}}^{L^2(\mathbb{R}^d, \rho_t)}$$

$$=: \mathrm{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d)$$ <u>tangent space</u>

Distance: $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$

Riemannian metric on manifold

$$W(\rho_0,\rho_1)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} \rho_t |u_t|^2 \, dt \; : \; t \longmapsto \rho_t \in \mathcal{P}_2(\mathbb{R}^d), \right.$$

$$\partial_t \rho_t + \nabla \cdot (\rho_t u_t) = 0,$$

Wasserstein distance

justifies

$$\left. u_t \in \mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d), \quad \rho_{t=0} = \rho_0, \; \rho_{t=1} = \rho_1 \right\}$$

Benamou Brenier formula