Gradient Flows in Metric Spaces

Lecture 1:
Recap of gradient flows in $\mathbb{R}^d$

0. Recap on convexity in $\mathbb{R}^d$

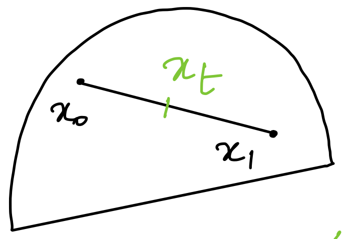
0.1 Convex sets

$C \subset \mathbb{R}^d$ convex if $\forall\, x_0, x_1 \in C$, $\forall\, t \in [0,1]$
$x_t := (1-t)x_0 + t\,x_1 \in C$.

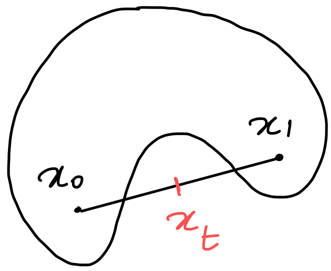
Thm (Supporting hyperplane)

Let $C \subset \mathbb{R}^d$ convex
and $x_0 \in \partial C$.
Then there exists $u \in \mathbb{R}^d \setminus \{0\}$,
such that $\forall x \in C$

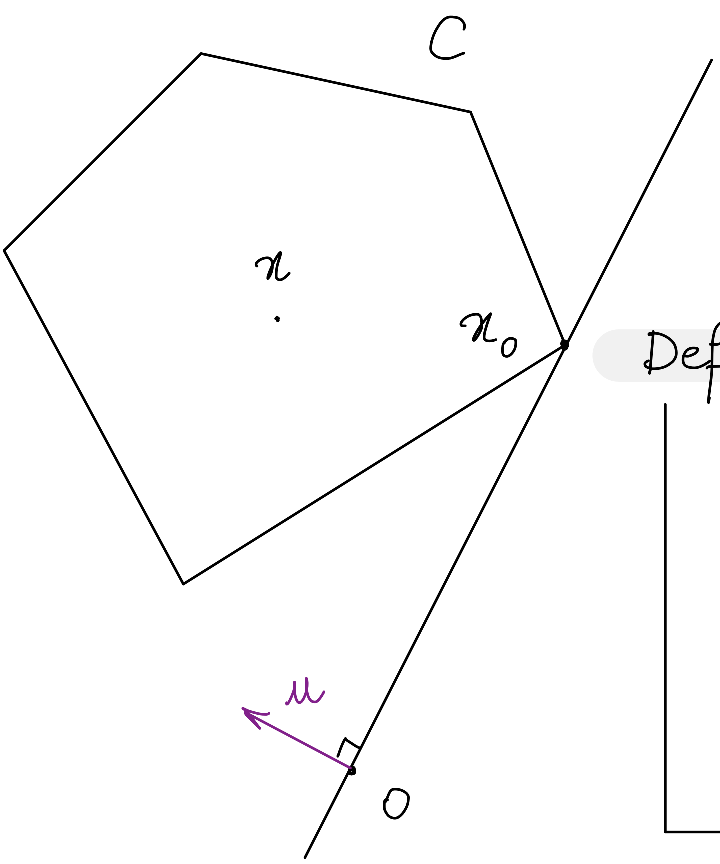
$$\langle u, x - x_0 \rangle \geqslant 0$$



convex ✓



non convex ✗



0.2 Convex functions.

Def. $C \subset \mathbb{R}^d$ convex set.
$f : C \to \mathbb{R}$ convex if
$\forall\, x_0, x_1 \in C$

$$f(x_t) \leqslant (1-t) f(x_0) + t\, f(x_1)$$

$$x_t := (1-t)x_0 + t\,x_1$$

One checks that $f : C \to \mathbb{R}$ convex iff its epigraph

$$\mathrm{epi}(f) := \{ (x,t) \in C \times \mathbb{R} : f(x) \leqslant t \}$$

is convex in $\mathbb{R}^d \times \mathbb{R}$.

Def (subgradients and subdifferential)

$C \subset \mathbb{R}^d$ convex and $f : C \to \mathbb{R}$.
A subgradient of f at $x \in C$ is a vector $g_x \in \mathbb{R}^d$ such that:

$$\forall y \in C, \quad f(y) \geqslant f(x) + \langle g_x , y - x \rangle$$

The subdifferential of f at $x \in C$, denoted $\partial f(x)$ is the set of all subgradients of f at x.

Thm. $C \subset \mathbb{R}^d$ convex and $f : C \to \mathbb{R}$.

i) $\left(\forall x \in C, \ \partial f(x) \neq \emptyset \right) \Rightarrow f$ convex

ii) f convex $\Rightarrow \left(\forall x \in \mathring{C}, \ \partial f(x) \neq \emptyset \right)$

iii) (Monotonicity of subgradients)
f convex $\Rightarrow \begin{cases} \forall x, y \in \mathring{C}, \ \forall g_x \in \partial f(x), \ \forall g_y \in \partial f(y) : \\ \langle g_x - g_y , x - y \rangle \geqslant 0 \end{cases}$

Proof. i) Take $x_0, x_1 \in C$ and $x_t := (1-t)x_0 + t x_1$.
By assumption there exists $g \in \partial f(x_t)$.
Hence we have

$$f(x_0) \geq f(x_t) + \langle g, x_0 - x_t \rangle$$
$$f(x_1) \geq f(x_t) + \langle g, x_1 - x_t \rangle$$

Multiplying the first inequality by $(1-t)$ the second one by t and summing them, we obtain $f(x_t) \leq (1-t) f(x_0) + t f(x_1)$.

ii) Consider $x \in C$. The pair $(x, f(x))$ belongs to the frontier $\partial \operatorname{epi}(f)$ of the epigraph of f.
By the supporting hyperplane theorem, there exists $(a, b) \in \mathbb{R}^d \times \mathbb{R}$, $\neq (0,0)$, such that $\forall (y, t) \in \operatorname{epi}(f)$

$$\langle a, y \rangle + bt \geq \langle a, x \rangle + b f(x). \qquad (*_1)$$

Note that $(y,t) \in \operatorname{epi}(f)$ and $t' \geq t$ implies $(y, t') \in \operatorname{epi}(f)$. In particular inequality $(*_1)$ should hold for $t \to +\infty$ which imposes

$$b \geq 0.$$

Now suppose $x \in \mathring{C}$. Then for $\varepsilon < 0$ small enough $z := x + \varepsilon a \in C$. Then $(*_1)$ implies that $\forall t \geq f(z)$:

$$\begin{aligned} \langle a, x \rangle + b f(x) &\leq \langle a, z \rangle + bt \\ &= \langle a, x \rangle + \varepsilon \|a\|^2 + bt \end{aligned}$$

$$\Rightarrow \quad b f(x) \leq \varepsilon \|a\|^2 + bt.$$

If $b=0$, $\varepsilon<0$ implies $a=0$ which is a contradiction. Hence :

$$b>0.$$

Finally, $\forall y \in C$, applying $(*_1)$ for $t=f(y)$ leads to

$$f(y) \geq f(x) + \langle \frac{a}{b}, x-y \rangle$$

$$\Rightarrow \; -a/b \in \partial f(x).$$

iii) Sum inequalities
$f(y) \geq f(x) + \langle g_x, y-x \rangle$ and $f(x) \geq f(y) + \langle g_y, x-y \rangle$.

$\square$

Thm (Convexity and differentiability)

$C \subset \mathbb{R}^d$ convex and $f : C \to \mathbb{R}$ convex.

i) f differentiable $\Rightarrow$ $\forall x \in \mathring{C}$, $\partial f(x) = \{\nabla f(x)\}$

ii) f twice differentiable $\Rightarrow$ $\forall x \in \mathring{C}$, $\forall h \in \mathbb{R}^d$
$\langle \nabla^2 f(x) h, h \rangle \geq 0$

Proof. i) If $x \in \mathring{C}$, then $\forall h \in \mathbb{R}^d$ and $t \in \mathbb{R}$ small enough, $x \pm th \in C$. Taylor expansion around x :

$$f(x \pm th) = f(x) \pm t \langle \nabla f(x), h \rangle + o(t).$$

Convexity of f implies, $\forall g_x \in \partial f(x)$,

$$f(x \pm th) \geq f(x) \pm t \langle g_x, h \rangle$$

$$\Rightarrow \forall h \in \mathbb{R}^d : \pm t \langle \nabla f(x), h \rangle + o(t) \geq \pm t \langle g_x, h \rangle$$
$$\Rightarrow (t \downarrow 0) \ \forall h \in \mathbb{R}^d : \quad \langle \nabla f(x) - g_x, h \rangle \geq 0$$

$$\Rightarrow \quad \nabla f(x) = g_x$$

ii) Similar idea:
- Taylor expansion around $x \in \mathring{C}$:

$$f(x \pm th) = f(x) \pm t \langle \nabla f(x), h \rangle + \frac{t^2}{2} \langle \nabla^2 f(x) h, h \rangle + o(t^2)$$

- Convexity: $f(x \pm th) \geq f(x) \pm t \langle \nabla f(x), h \rangle$

$$\Rightarrow \quad \frac{t^2}{2} \langle \nabla^2 f(x) h, h \rangle + o(t^2) \geq 0 \qquad \square$$

0.3 λ- convexity

Def Let $C \subset \mathbb{R}^d$ convex and $\lambda \in \mathbb{R}$.
$f : C \to \mathbb{R}$ is called λ - convex if

$$x \in C \mapsto f_\lambda(x) := f(x) - \frac{\lambda}{2} \|x\|_2^2$$

is convex

Thm. $C \subset \mathbb{R}^d$, $\lambda \in \mathbb{R}$, $f : C \to \mathbb{R}$ λ-convex
Then the following holds

i) $\forall x_0, x_1 \in C$, $\forall t \in [0,1]$:

$$f(x_t) \leqslant (1-t) f(x_0) + t f(x_1) - \frac{\lambda}{2} t(1-t) \|x_0 - x_1\|^2$$

ii) If f differentiable, then $\forall x \in \mathring{C}$, $\forall y \in C$:

$$f(y) \geqslant f(x) + \langle \nabla f(x), y-x \rangle + \frac{\lambda}{2} \|x-y\|^2$$

iii) If f differentiable, $\forall x, y \in \mathring{C}$:

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geqslant \lambda \|x-y\|^2$$

iv) If f twice differentiable, $\forall x \in \mathring{C}$, $\forall h \in \mathbb{R}^d$:

$$\langle \nabla^2 f(x) h, h \rangle \geqslant \lambda \|h\|^2.$$

Proof. Exercise.
In particular, show conversely that these properties characterize λ-convexity.

Thm (Case $\lambda > 0$)

Suppose $f : C \to \mathbb{R}$ is λ-convex, $\lambda > 0$.
Then:

i) If C closed and f lower semi continuous, f has a unique minimizer x^*.

ii) (Quadratic growth): $\forall x \in C$

$$\frac{\lambda}{2} \|x - x^*\|^2 \leq f(x) - f(x^*).$$

iii) (Polyak-Łojasiewicz (PL) inequality) If f is differentiable, $\forall x \in \mathring{C}$:

$$f(x) - f(x^*) \leq \frac{1}{2\lambda} \|\nabla f(x)\|^2$$

Proof

i) Let $x_n \in C$ be a minimizing sequence ($\lim_n f(x_n) = \inf f$). By point i) of the previous thm C, we obtain

$$\frac{\lambda}{2} \|x_n - x_m\|^2 \leq \frac{1}{2} f(x_n) + \frac{1}{2} f(x_m) - f\left(\frac{x_n + x_m}{2}\right)$$

Letting $n, m \to +\infty \Rightarrow \lim_{n,m \to +\infty} \|x_n - x_m\| = 0$
$\Rightarrow (x_n)$ Cauchy and hence converges in C (closed).
If $x^* := \lim x_n$. We get by lower semi-continuity

$$f(x^*) \leq \lim f(x_n)$$

$\Rightarrow x^*$ minimum.
Uniqueness follows from ii).

ii) $\forall x \in C$, denote $x_t := (1-t)x^* + tx$.
Point i) of previous thm $\Rightarrow$
$$f(x_t) \leq (1-t)f(x^*) + t f(x) - \frac{\lambda}{2} t(1-t) \|x - x^*\|^2.$$

$\Rightarrow \forall t \in (0,1)$:

$$\frac{\lambda}{2} \|x - x^*\|^2 \leq \frac{f(x) - f(x^*)}{1-t} + \underbrace{\frac{f(x^*) - f(x_t)}{t(1-t)}}_{\leq 0}$$
$$\leq \frac{f(x) - f(x^*)}{1-t}.$$

$t \downarrow 0 \Rightarrow$ Conclusion.

iii) Fix $x \in \mathring{C}$. Suppose $x^* \in \mathring{C}$. Then optimizing both sides of inequality (in $y \in \mathring{C}$)

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\lambda}{2} \|x - y\|^2$$

implies
$$f(x^*) \geq f(x) - \frac{1}{2\lambda} \|\nabla f(x)\|^2,$$
as desired. If $x^* \in \partial C$, we conclude by an approximation argument.

$\square$

1. Gradient flows in $\mathbb{R}^d$.

1.1 Definition

Consider $f : \mathbb{R}^d \to \mathbb{R}$, $C^2(\mathbb{R}^d)$.

Def. The gradient flow of f is the family of maps $(S_t)_{t \geq 0}$, $S_t : \mathbb{R}^d \to \mathbb{R}^d$ such that

- $S_0 = id$,
- $\forall x_0 \in \mathbb{R}^d$, the map
  $$t \in (0, +\infty) \mapsto S_t(x_0)$$
  is the unique solution of the Cauchy problem
  $$\begin{cases} \dfrac{d}{dt} x_t = -\nabla f(x_t) \\ \lim\limits_{t \to 0} x_t = x_0 \end{cases} \quad : \quad GF(f, x_0)$$

Notation: For $x : (0, +\infty) \to \mathbb{R}^d$, we use indifferently notation

$$x_t \text{ or } x(t)$$

as well as

$$\dot{x}_t \text{ or } \frac{d}{dt} x_t$$

1.2 Basic properties

Lemma (Gronwall)

Let $g : [0, +\infty) \to \mathbb{R}$ be differentiable and satisfying $g'(t) \leq \beta(t) g(t)$ for some $\beta : [0, +\infty) \to \mathbb{R}$.
Then
$$g(t) \leq g(0) \exp\left(\int_0^t \beta(s)\, ds \right).$$

Theorem. Suppose $x : [0, \infty) \to \mathbb{R}^d$ solves $GF(f, x_0)$. Then the following holds.

i) ("Energy identity") $\forall t > 0$:
$$\frac{d}{dt} f(x_t) = - \|\dot{x}_t\|^2 = - \|\nabla f(x_t)\|^2 \leq 0. \quad (EI)$$
If in addition f is λ-convex, for some $\lambda > 0$, and achieves its minimum at x^*, Then:
$$f(x_t) - f(x^*) \leq (f(x_0) - f(x^*)) e^{-2\lambda t}$$

ii) ("Slope inequality") If f is λ-convex, $\lambda \in \mathbb{R}$, then $\forall t > 0$:
$$\|\nabla f(x_t)\| \leq \|\nabla f(x_0)\| e^{-\lambda t}. \quad (SI)$$

Comments

- Point i) States that f decreases along the trajectory of x_t. Note that identity (EI) does not require f to be convex.

- Point ii) states that the "slope" of f, i.e. $\|\nabla f\|$, decreases along the trajectory of x_t.

Proof. i) By the chain rule,

$$\frac{d}{dt} f(x_t) = \langle \nabla f(x_t), \dot{x}_t \rangle$$

and (EI) follows from the fact $\dot{x}_t = -\nabla f(x_t)$. Now if f is λ-convex (for $\lambda > 0$), we have seen earlier that

$$f(x_t) - f(x^*) \leq \frac{1}{2\lambda} \|\nabla f(x_t)\|^2.$$

Hence, we deduce from (EI) that

$$\frac{d}{dt}\left(f(x_t) - f(x^*)\right) = \frac{d}{dt} f(x_t) = -\|\nabla f(x_t)\|^2$$
$$\leq -2\lambda\left(f(x_t) - f(x^*)\right).$$

Gronwall $\Rightarrow$

$$f(x_t) - f(x^*) \leq \left(f(x_0) - f(x^*)\right) e^{-2\lambda t}.$$

ii) By the chain rule again

$$\frac{d}{dt}\|\nabla f(x_t)\|^2 = 2\langle \nabla^2 f(x_t)\cdot \dot{x}_t , \nabla f(x_t)\rangle$$
$$= -2\langle \nabla^2 f(x_t)\cdot \nabla f(x_t), \nabla f(x_t)\rangle$$
$$\leq -2\lambda \|\nabla f(x_t)\|^2 .$$

$\langle \nabla^2 f(x)h, h\rangle \geq \lambda \|h\|^2$

Gronwall $\Rightarrow \|\nabla f(x_t)\|^2 \leq \|\nabla f(x_0)\|^2 e^{-2\lambda t}$.

$\square$

Exercise. Show that if f is λ-convex, for $\lambda \in \mathbb{R}$, and if x solves $GF(f, x_0)$ and y solves $GF(f, y_0)$, then $\forall t > 0$

$$\|x_t - y_t\| \leq \|x_0 - y_0\| e^{-\lambda t}$$

1.3 Metric characterization of Gflows in $\mathbb{R}^d$: The Energy Dissipation Inequality (EDI)

Thm. A C^1 curve $x : [0, \infty) \to \mathbb{R}^d$ is a solution of $GF(f, x_0)$ if and only if $\forall t \in (0, +\infty)$

$$f(x_t) + \frac{1}{2}\int_0^t \left(\|\dot{x}_s\|^2 + \|\nabla f(x_s)\|^2\right) ds \leq f(x_0)$$

(EDI)

Proof. $\Rightarrow$) Suppose x solves $GF(f, x_0)$. Then the Energy identity (already proved) shows that

$$\frac{d}{ds} f(x_s) = -\|\dot{x}_s\|^2 = -\|\nabla f(x_s)\|^2$$

$$\Rightarrow \quad \frac{d}{ds} f(x_s) = -\frac{1}{2}\left(\|\dot{x}_s\|^2 + \|\nabla f(x_s)\|^2\right)$$

Integrating from 0 to t implies that

$$f(x_t) - f(x_0) = -\frac{1}{2}\int_0^t \left(\|\dot{x}_s\|^2 + \|\nabla f(x_s)\|^2\right) ds$$

so that the EDI actually holds as an identity.

$\Leftarrow$) Conversely suppose that the EDI holds. Then, observing that (using only the differentiability of f and x)

$$f(x_t) = f(x_0) + \int_0^t \langle \nabla f(x_s), \dot{x}_s \rangle \, ds$$

we deduce that

$$\frac{1}{2}\int_0^t \|\dot{x}_s + \nabla f(x_s)\|^2 ds$$

$$= \frac{1}{2}\int_0^t \left(\|\dot{x}_s\|^2 + \|\nabla f(x_s)\|^2\right) ds + \int_0^t \langle \nabla f(x_s), \dot{x}_s \rangle ds$$

$$= \frac{1}{2}\int_0^t \left(\|\dot{x}_s\|^2 + \|\nabla f(x_s)\|^2\right) ds + f(x_t) - f(x_0)$$

$$\overset{(EDI)}{\leq} 0 .$$

Hence, $\forall t>0$, $\|\dot{x}_s + \nabla f(x_s)\| = 0$ for Lebesgue almost all $s \in [0,t]$. Since this holds $\forall t>0$ and since x is C^1, this implies : $\forall t>0$

$$\dot{x}_t = -\nabla f(x_t).$$

$\square$

Comment. The above result provides a "metric" characterization of the solution of $GF(f, x_0)$ since quantities $\|\dot{x}_s\|$ and $\|\nabla f(x_s)\|$ can be defined only in terms of the Euclidean metric $(x,y) \mapsto \|x-y\|$ and therefore can be defined analogously in arbitrary metric spaces. Indeed, denoting $d(x,y) := \|x-y\|$,

$$\|\dot{x}_t\| = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon \neq 0}} \frac{d(x_t, x_{t+\varepsilon})}{|\varepsilon|}$$

and

$$\|\nabla f(x)\| = \limsup_{\substack{y \to x \\ y \neq x}} \frac{|f(x)-f(y)|}{d(x,y)}$$

To see why this second identity holds, observe that $\forall y \in \mathbb{R}^d$ (by Taylor expansion):

$$f(y) - f(x) = \langle \nabla f(x), y-x \rangle + o(\|y-x\|)$$

Hence $y \neq x \Rightarrow$

$$\frac{f(y)-f(x)}{d(y,x)} = \left\langle \nabla f(x), \frac{y-x}{\|y-x\|} \right\rangle + \varepsilon(\|y-x\|)$$

where $\varepsilon(u) \to 0$ as $u \to 0$. Hence,

$$\limsup_{\substack{y\to x \\ y\neq x}} \frac{|f(x)-f(y)|}{d(x,y)} = \limsup_{\substack{y\to x \\ y\neq x}} \left|\left\langle \nabla f(x), \frac{y-x}{\|y-x\|} \right\rangle\right|.$$

Finally, note that:

$$\limsup_{\substack{y\to x \\ y\neq x}} \left|\left\langle \nabla f(x), \frac{y-x}{\|y-x\|} \right\rangle\right| = \sup_{\substack{u\in\mathbb{R}^d \\ \|u\|=1}} |\langle \nabla f(x), u\rangle|$$

$$= \|\nabla f(x)\|,$$

which proves the claim.

1.4 Metric characterization: Evolution Variational Inequality (EVI)

Thm. Suppose $f \in C^2(\mathbb{R}^d)$ is λ- convex. Then a C^1 curve $x : [0, \infty) \to \mathbb{R}^d$ solves $GF(f, x_0)$ if and only if $\forall v \in \mathbb{R}^d$, $\forall t > 0$:

$$\frac{d}{dt}\left(\frac{1}{2} d(x_t, v)^2\right) + \frac{\lambda}{2} d(x_t, v)^2 \leq f(v) - f(x_t)$$

where $d(\cdot - \cdot) = \| \cdot - \cdot \|$ is the Euclidean distance.

Proof. $\Rightarrow$) Suppose x solve $GF(f, x_0)$. Then

$$\frac{d}{dt}\left(\frac{1}{2} d(x_t, v)^2\right) = \langle \dot{x}_t, x_t - v \rangle$$

$$= - \langle \nabla f(x_t), x_t - v \rangle.$$

f λ convex $\Rightarrow$

$$f(v) \geq f(x_t) + \langle \nabla f(x_t), v - x_t \rangle + \frac{\lambda}{2} \| x_t - v \|^2.$$

Combining these two observations we get exactly:

$$\frac{d}{dt}\left(\frac{1}{2} d(x_t, v)^2\right) + \frac{\lambda}{2} d(x_t, v)^2 \leq f(v) - f(x_t).$$

$\Leftarrow$) Conversely, if the EVI holds $\forall v \in \mathbb{R}^d$, $\forall t > 0$, then

Since $\dfrac{d}{dt}\left(\dfrac{1}{2} d(x_t, v)^2\right) = \langle \dot{x}_t, x_t - v \rangle$,

the EVI reads :

$$\langle \dot{x}_t, x_t - v \rangle \leqslant f(v) - f(x_t) - \frac{\lambda}{2} \| x_t - v \|^2 .$$

Choosing $v = x_t + \varepsilon \xi$, $\varepsilon > 0$, $\xi \in \mathbb{R}^d$, and dividing by ε, we obtain

$$- \langle \dot{x}_t, \xi \rangle \leqslant \frac{f(x_t + \varepsilon \xi) - f(x_t)}{\varepsilon} - \frac{\lambda \varepsilon}{2} \| \xi \|^2 .$$

Taking $\varepsilon \downarrow 0$, we deduce that :

$$\forall \xi \in \mathbb{R}^d : \quad - \langle \dot{x}_t, \xi \rangle \leqslant \langle \nabla f(x_t), \xi \rangle .$$

This implies $\dot{x}_t = - \nabla f(x_t)$. $\square$