Optimization

$$
\begin{aligned}
& x^\star \in \operatorname*{argmin}_{x \in X} f(x) \\
\iff \quad & \delta_{x^\star} \in \operatorname*{argmin}_{\mu \text{ prob. distrib.}} \int_X f(x) \mathrm{d}\mu(x)
\end{aligned}
$$

$$f^*(\omega) \overset{\text{def.}}{=} \sup_x \langle x,\, \omega \rangle - f(x)$$

$$(f \diamond g)(x) \overset{\text{def.}}{=} \inf_y f(y) + g(x-y)$$

$$\textit{Theorem:}\ (f \diamond g)^* = f^* + g*$$

$$\hat{f}(\omega) \overset{\text{def.}}{=} \int f(x) e^{-\mathrm{i}\langle \omega,\, x \rangle} \mathrm{d}x$$

$$(f \star g)(x) \overset{\text{def.}}{=} \int f(y) g(x-y) \mathrm{d}y$$

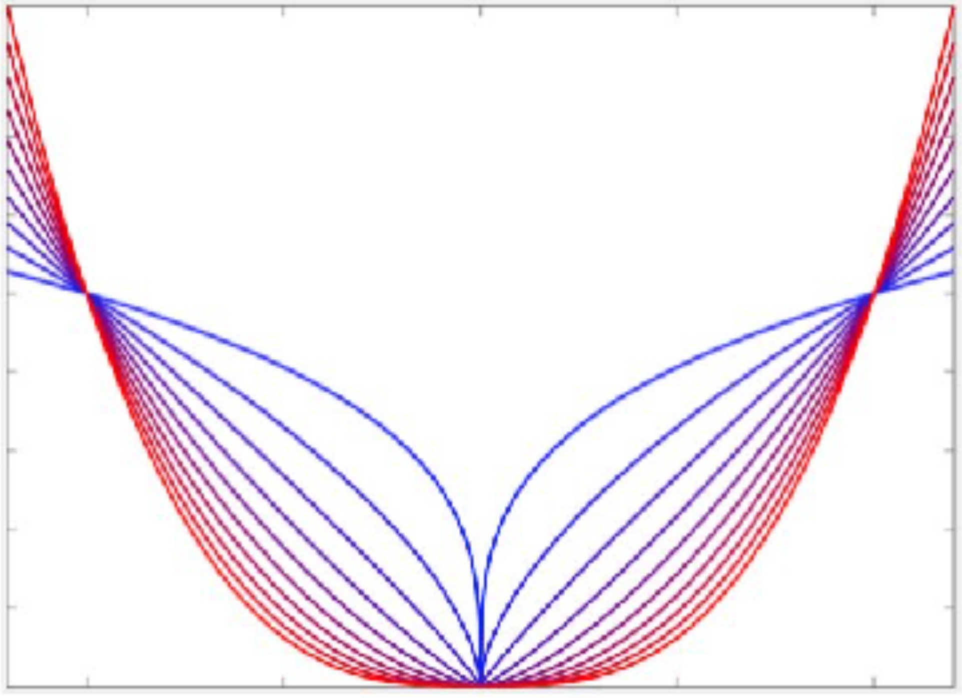
$$\textit{Theorem:}\ \widehat{f \star g} = \hat{f} \cdot \hat{g}$$

Kurdyka-Łojasiewicz:

(at minimum $f(0) = 0$)

$$\exists \varphi, \|\nabla(\varphi \circ f)\| \geqslant 1$$

$$\iff \quad \exists (K, \tau), \|\nabla f(x)\| \geqslant K |f(x)|^{\tau}$$



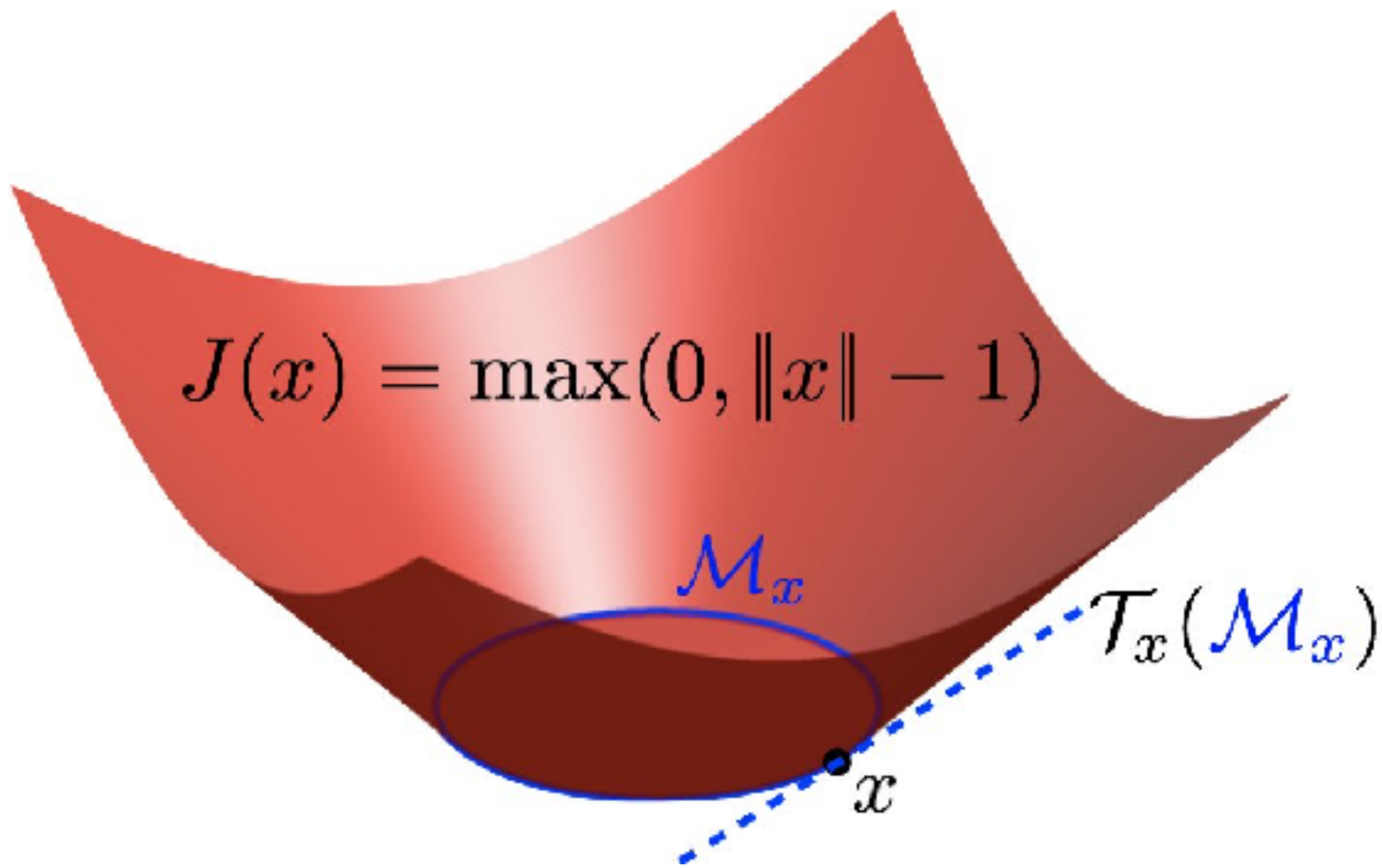
$J : \mathbb{R}^N \to \mathbb{R}$ is partly smooth at x for a manifold $\mathcal{M}_x$

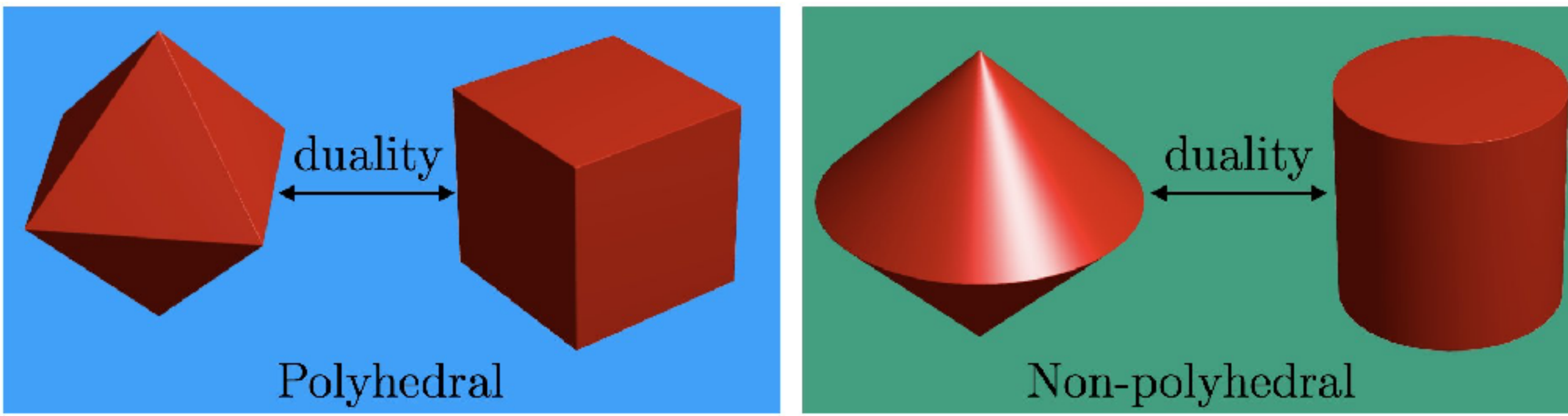
(i) J is C^2 along $\mathcal{M}_x$ around x ;

(ii) $\forall\, h \in \mathcal{T}_x(\mathcal{M}_x)^\perp$, $t \mapsto J(x + th)$ non-smooth at $t = 0$.

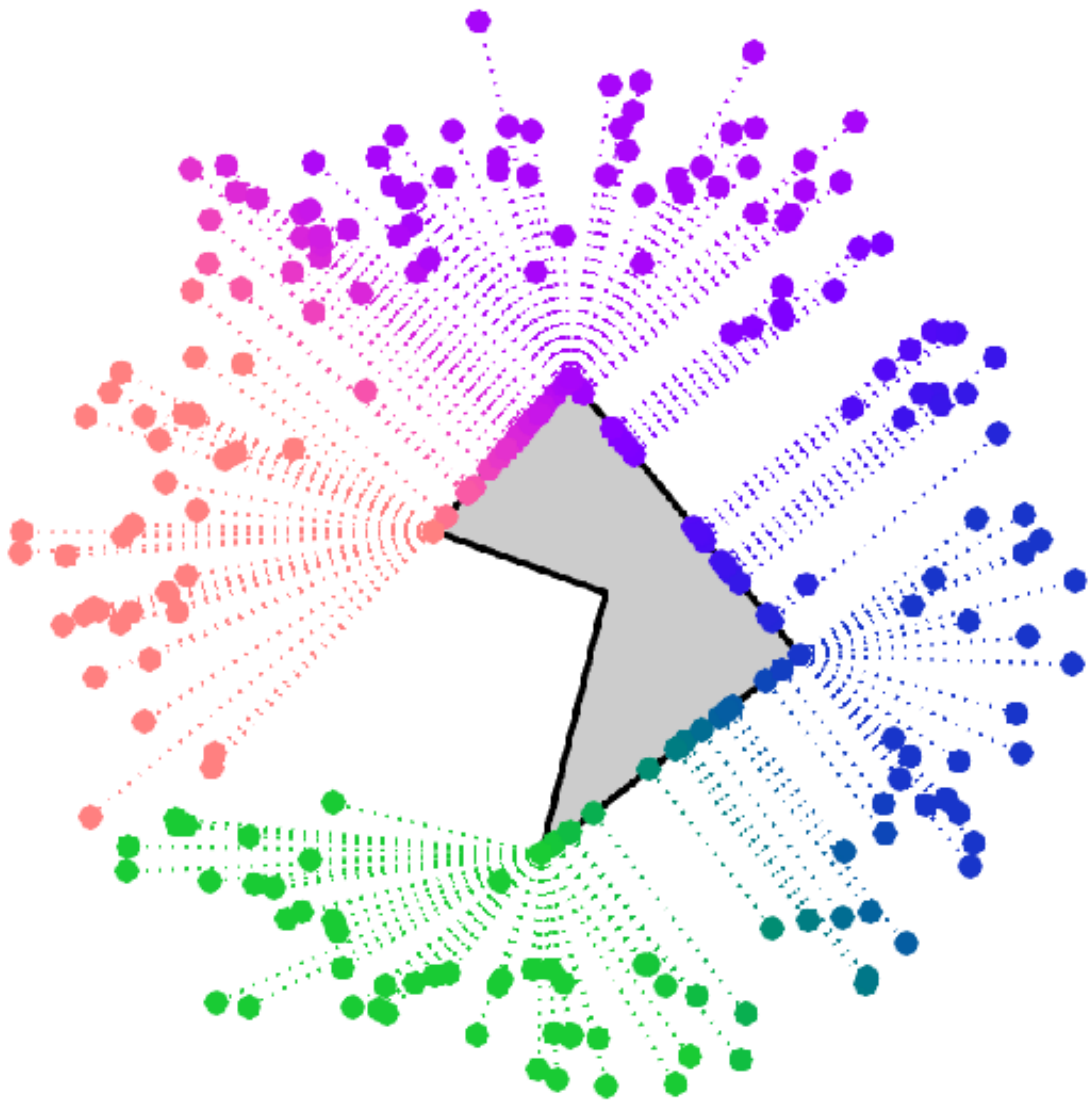
(iii) ∂J is continuous on $\mathcal{M}_x$ around x.

[Lewis 2003]

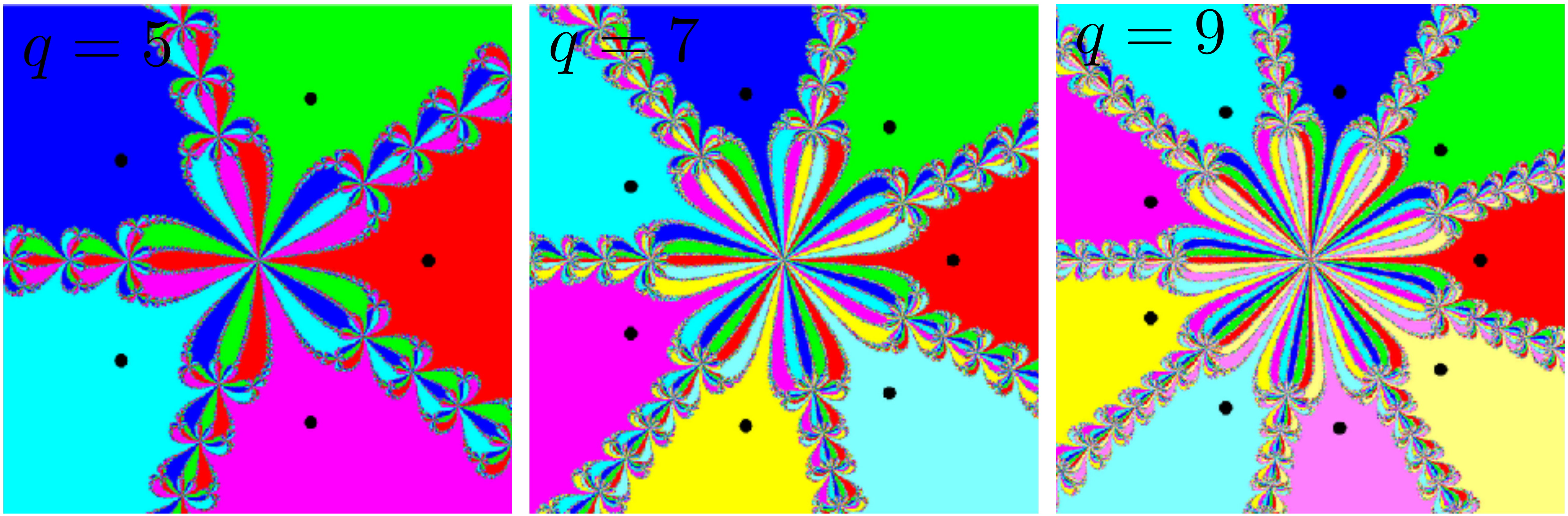




duality
Polyhedral
duality
Non-polyhedral

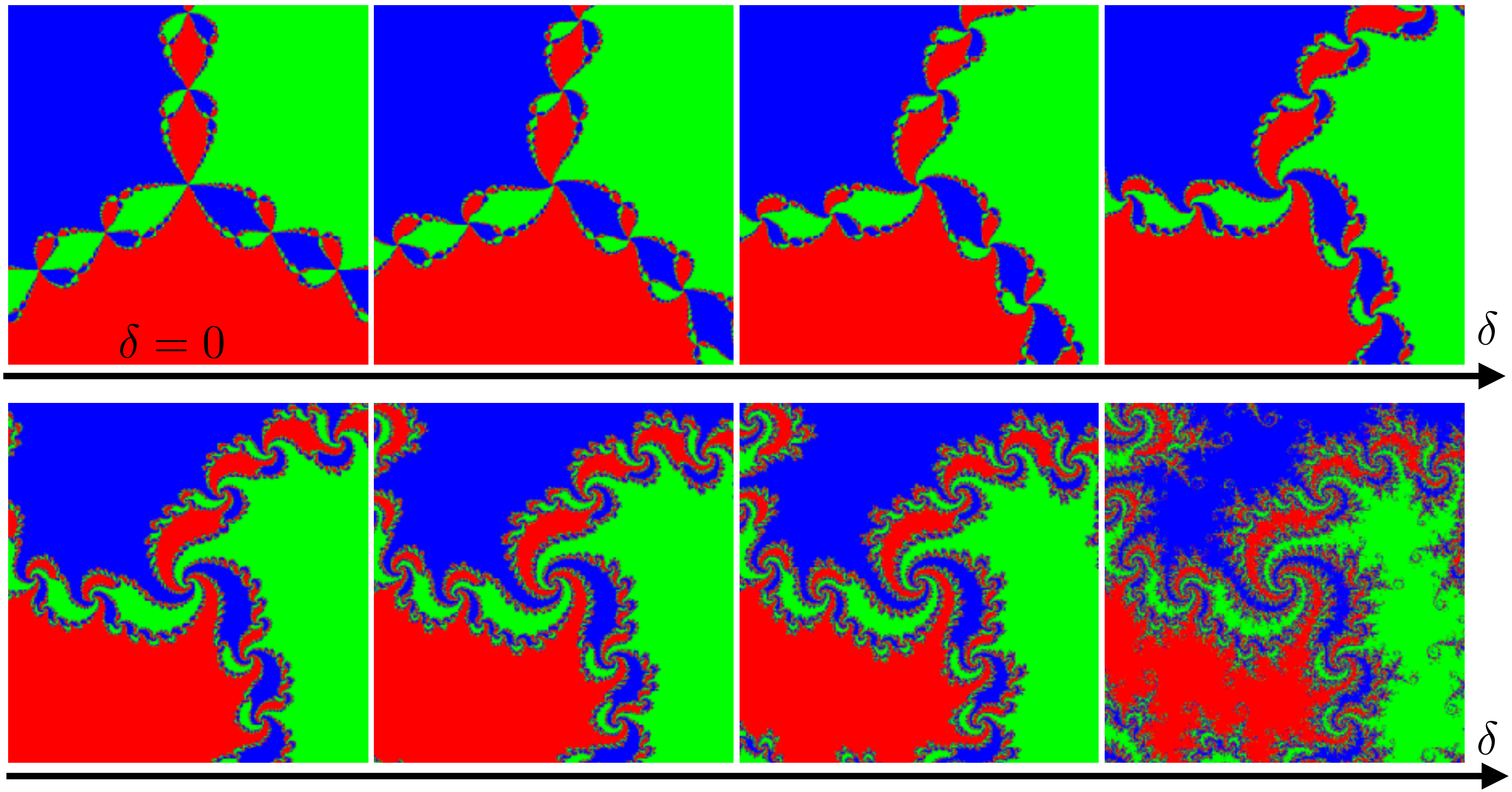


Newton method: $z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}$



Attraction bassins for $f(z) = z^q - 1$

“Twisted” Newton: $z_{k+1} = z_k - (1 + \delta e^{\mathrm{i}\theta}) \frac{f(z_k)}{f'(z_k)}$ $\quad f(z) = z^3 - 1$

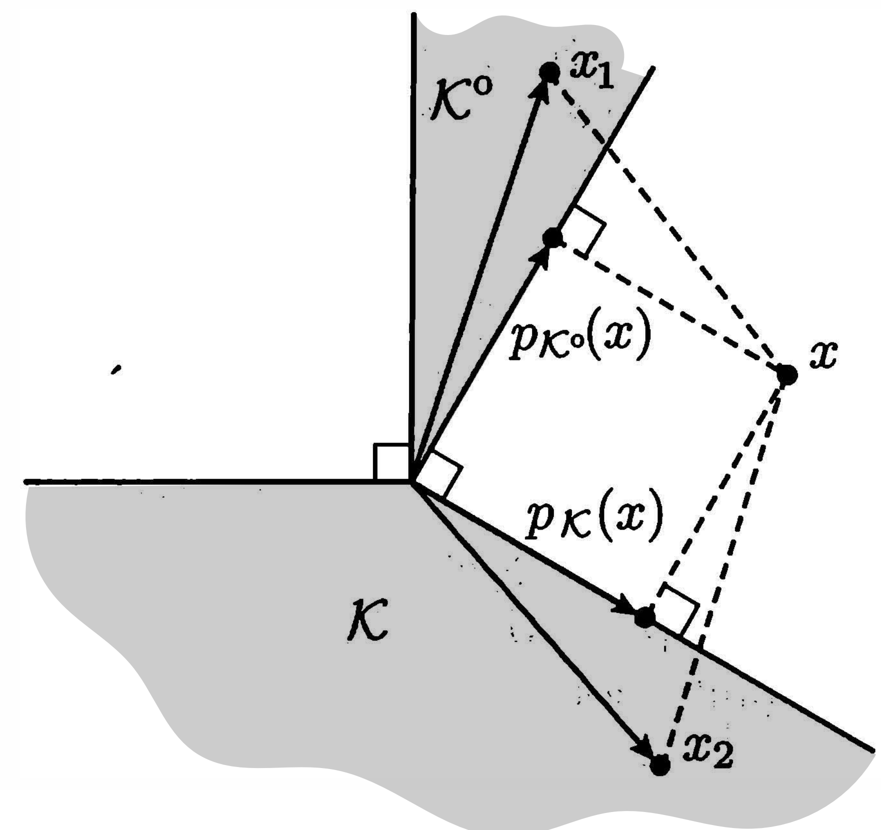


Polar cone:

$$\mathcal{K}^\circ \overset{\text{def.}}{=} \{x \,;\, \forall y \in \mathcal{K}, \langle x,\, y\rangle \leqslant 0\}$$

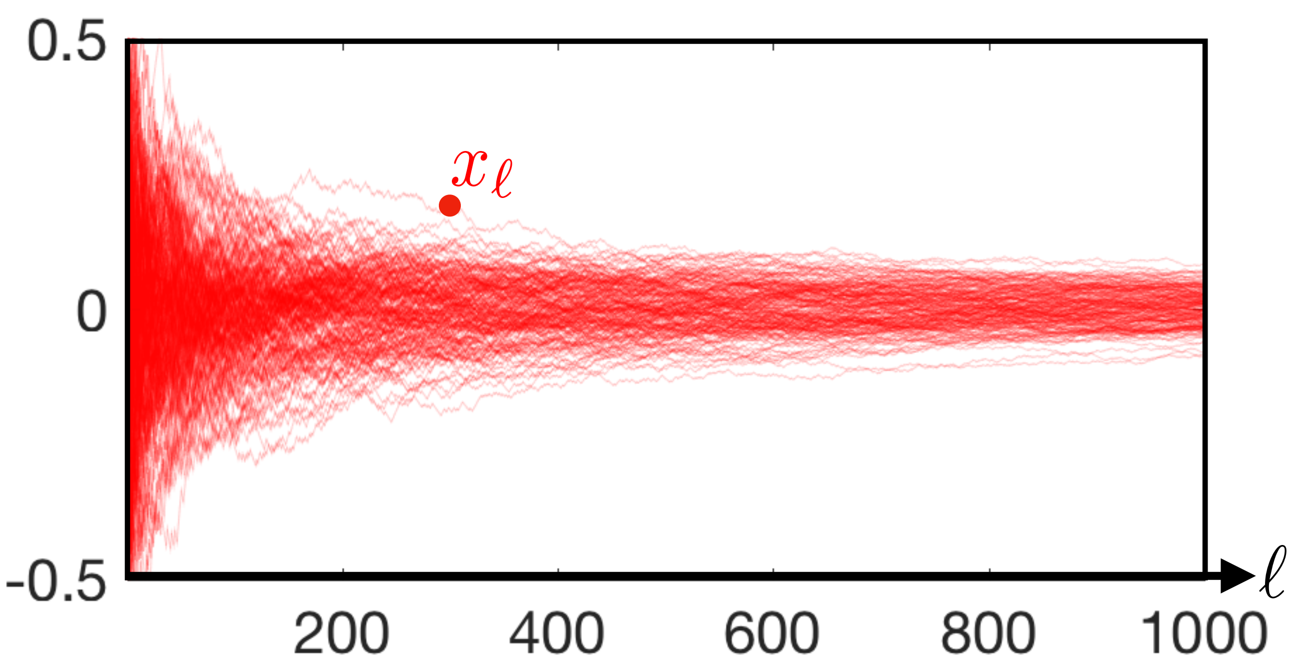
Theorem: (Moreau's decomposition)

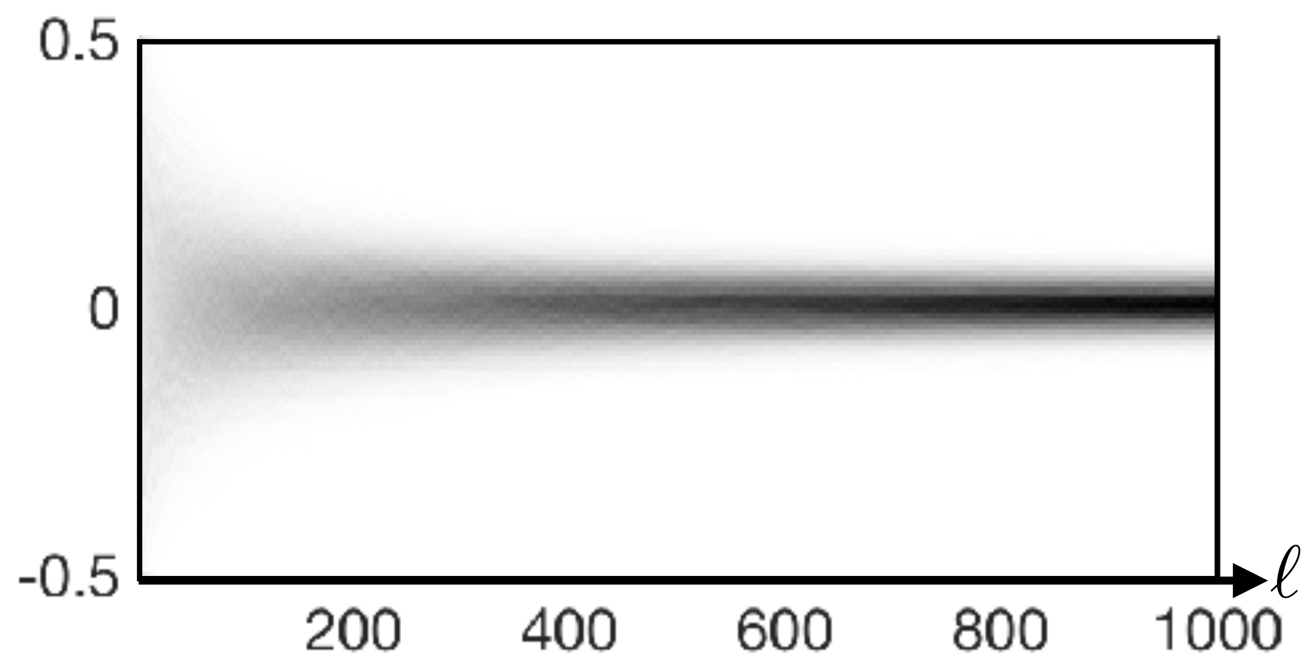
$$x = \mathrm{Proj}_{\mathcal{K}}(x) +^{\perp} \mathrm{Proj}_{\mathcal{K}^\circ}(x)$$



$$\min_{x\in\mathbb{R}} \underset{=f_1(x)}{(x+1)^2} + \underset{=f_2(x)}{(x-1)^2}$$

$$x_{\ell+1} \overset{\text{def.}}{=} \begin{cases} x_\ell - \frac{1}{\ell}\nabla f_1(x_\ell) \text{ with proba } \frac{1}{2} \\ x_\ell - \frac{1}{\ell}\nabla f_2(x_\ell) \text{ with proba } \frac{1}{2} \end{cases}$$

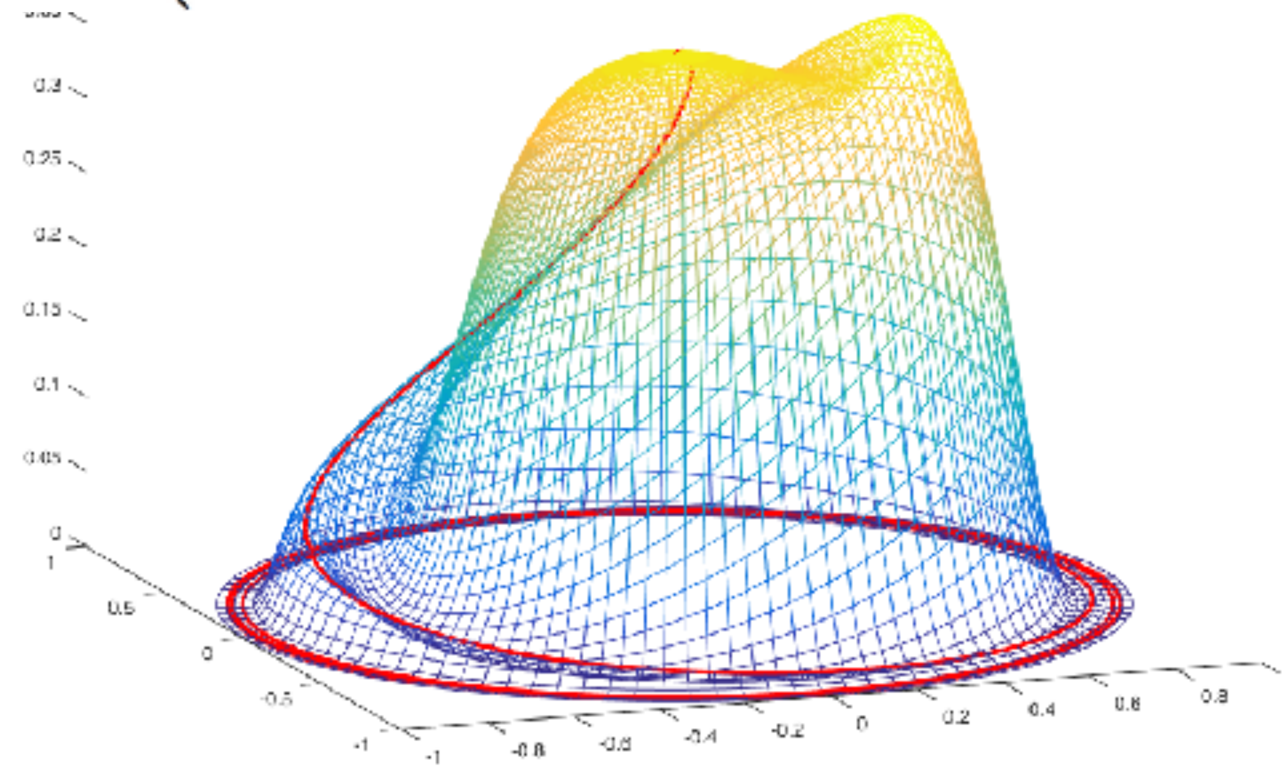




Gradient descent dynamic:

$$x(t) = -\nabla f(x(t)))$$

$$f(r,\theta) := \begin{cases} e^{-\frac{1}{1-r^2}} \left[1 - \frac{4r^4}{4r^4+(1-r^2)^4} \sin\left(\theta - \frac{1}{1-r^2}\right)\right] & \text{if } r < 1, \\ 0 & \text{if } r \geq 1, \end{cases}$$



On the "mexican hat":

$$x(t) = r(t)(\cos(\theta(t)), \sin(\theta(t))$$

$$\theta(t) = (1 - r(t))^{-2}$$

$$\rightarrow \text{Length}(x) = +\infty$$

$$f : z \stackrel{\text{def.}}{=} x + \mathrm{i}y \in \mathbb{C} \longmapsto f(z) \in \mathbb{C}$$

$$\frac{\partial}{\partial z} \stackrel{\text{def.}}{=} \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right) \qquad \frac{\partial}{\partial \bar{z}} \stackrel{\text{def.}}{=} \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)$$

Proposition: f holomorphic $\Leftrightarrow \frac{\partial f}{\partial \bar{z}} = 0.$

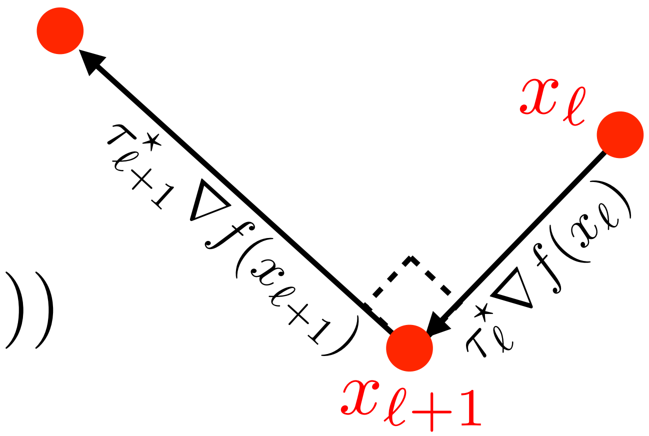
Examples:

$$\frac{\partial z}{\partial z} = 1 \qquad \frac{\partial \bar{z}}{\partial z} = 0 \qquad \frac{\partial z}{\partial \bar{z}} = 0 \qquad \frac{\partial \bar{z}}{\partial \bar{z}} = 1$$

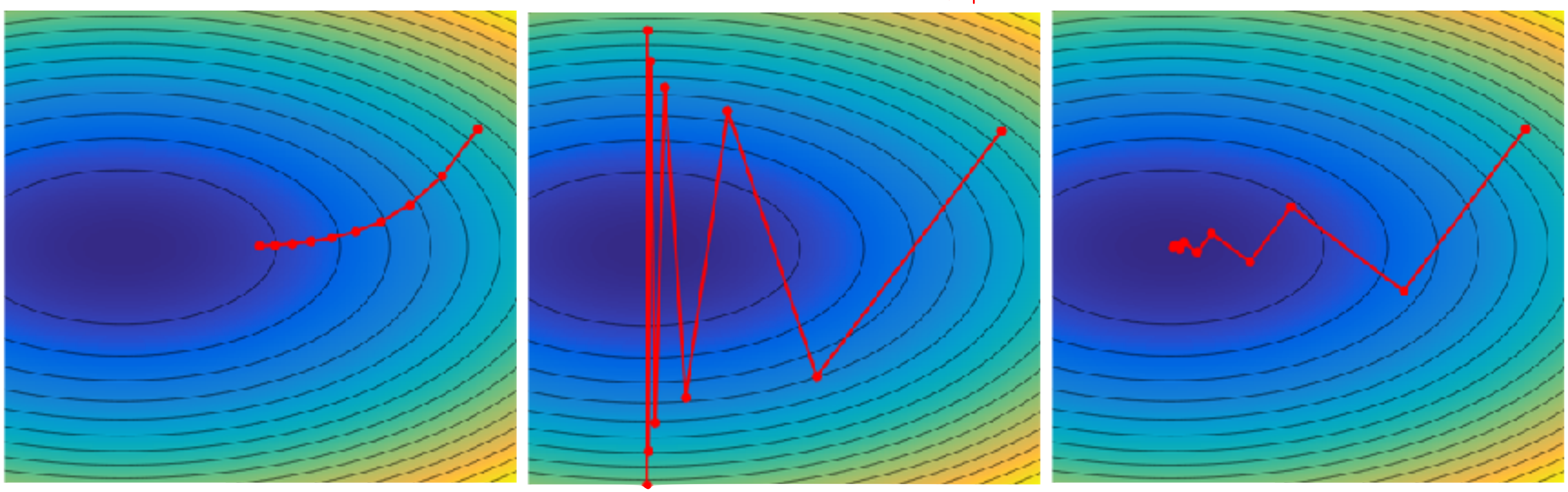
$$\frac{\partial c z}{\partial z} = c \qquad \frac{\partial |z|^2}{\partial z} = \frac{\partial (z\bar{z})}{\partial z} = \bar{z}$$

$$x_{\ell+1} = x_\ell - \tau_\ell \nabla f(x_\ell)$$

$$\tau_\ell^\star = \underset{\tau}{\text{argmin}}\ f(x_\ell - \tau \nabla f(x_\ell))$$

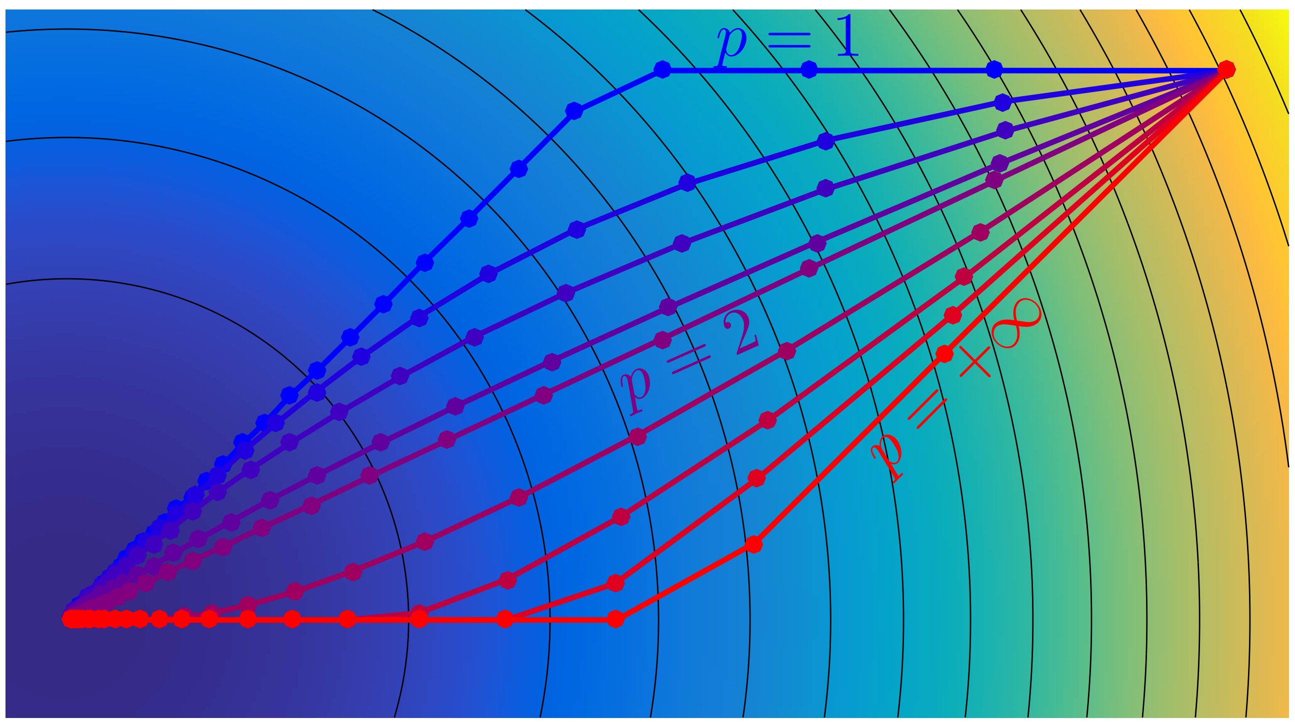


$$\nabla f(x_\ell) \perp \nabla f(x_{\ell+1})$$



Small τ_ℓ | Large τ_ℓ | Optimal $\tau_\ell = \tau_\ell^\star$

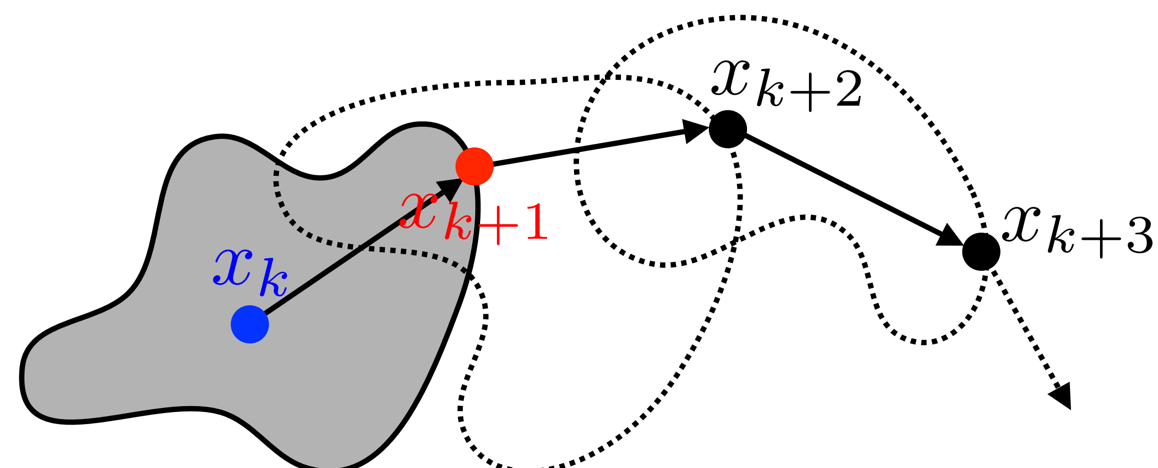
Metric space $(\mathcal{X}, d)$, minimize $F(x)$ on $\mathcal{X}$.



$F(x) = \|x\|^2$ on $(\mathcal{X} = \mathbb{R}^2, \|\cdot\|_p)$

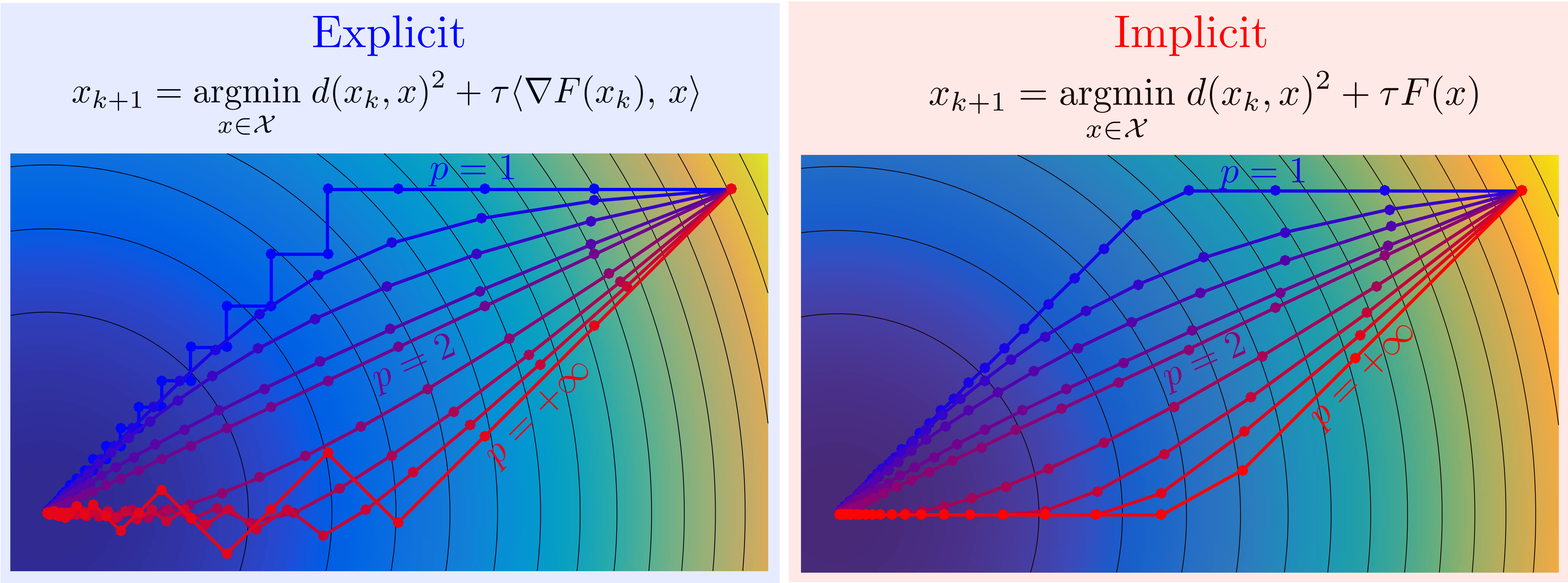
Implicit Euler step:

$$x_{k+1} \overset{\text{def.}}{=} \underset{x \in \mathcal{X}}{\operatorname{argmin}}\; d(x_k, x)^2 + \tau F(x)$$



$\{x \;;\; d(x_k, x) \sim \tau\}$

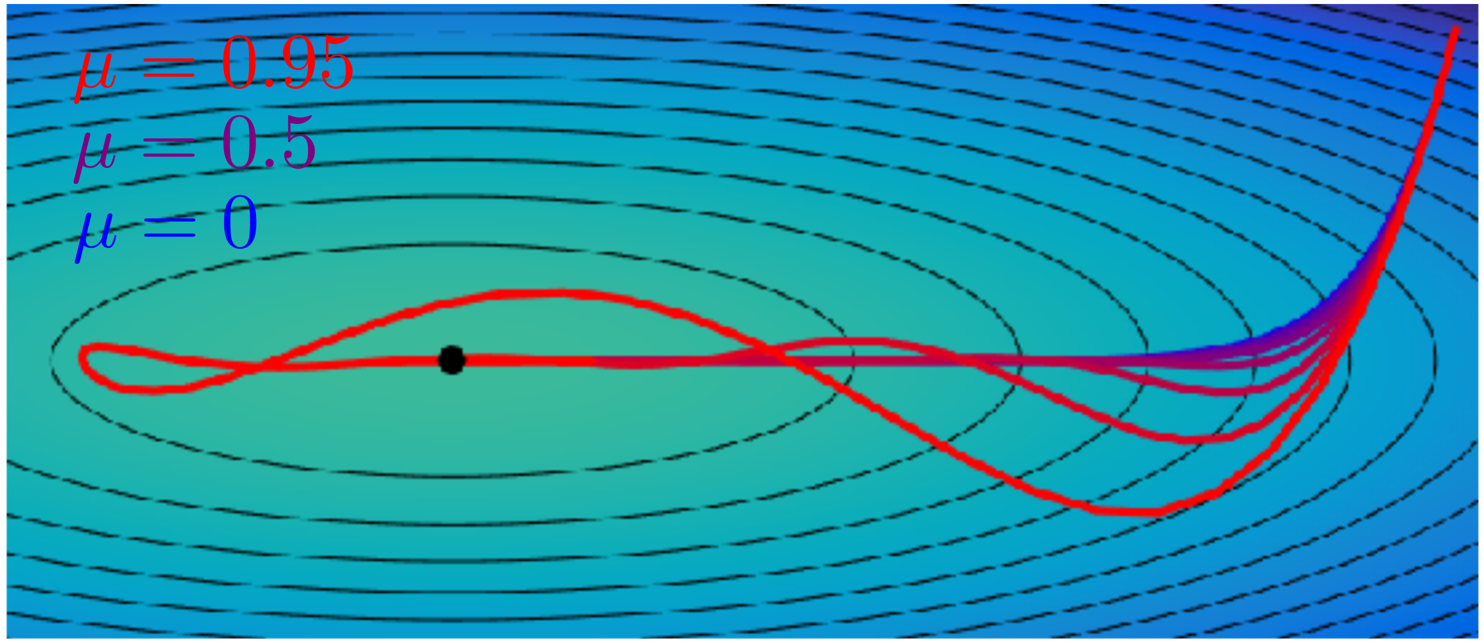
Metric space $(\mathcal{X}, d)$, minimize $F(x)$ on $\mathcal{X}$.



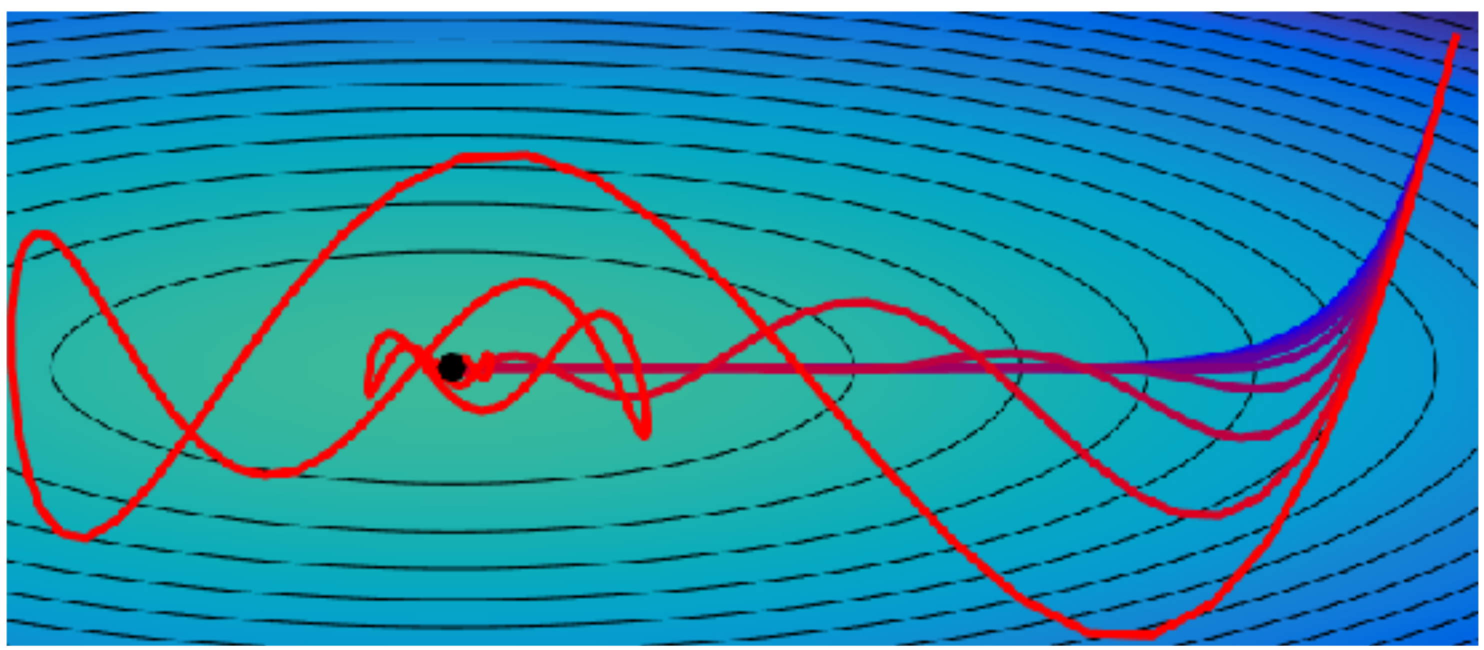
$$F(x) = \|x\|^2 \text{ on } (\mathcal{X} = \mathbb{R}^2, \|\cdot\|_p)$$

$$
\begin{aligned}
x_{k+1} &= x_k + p_k \\
p_{k+1} &= \mu p_k - \tau \begin{cases} \nabla f(x_k) & \text{Polyak} \\ \nabla f(x_k + \mu p_k) & \text{Nesterov} \end{cases}
\end{aligned}
$$

Yurii Nesterov



Boris Polyak



Gradient descent

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

$\tau \to 0$ $\quad$ $k\tau \to t$

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\nabla f(x(t))$$

Nesterov's acceleration

$$x_{k+1} = y_k - \tau \nabla f(y_k)$$

$$y_{k+1} = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k)$$

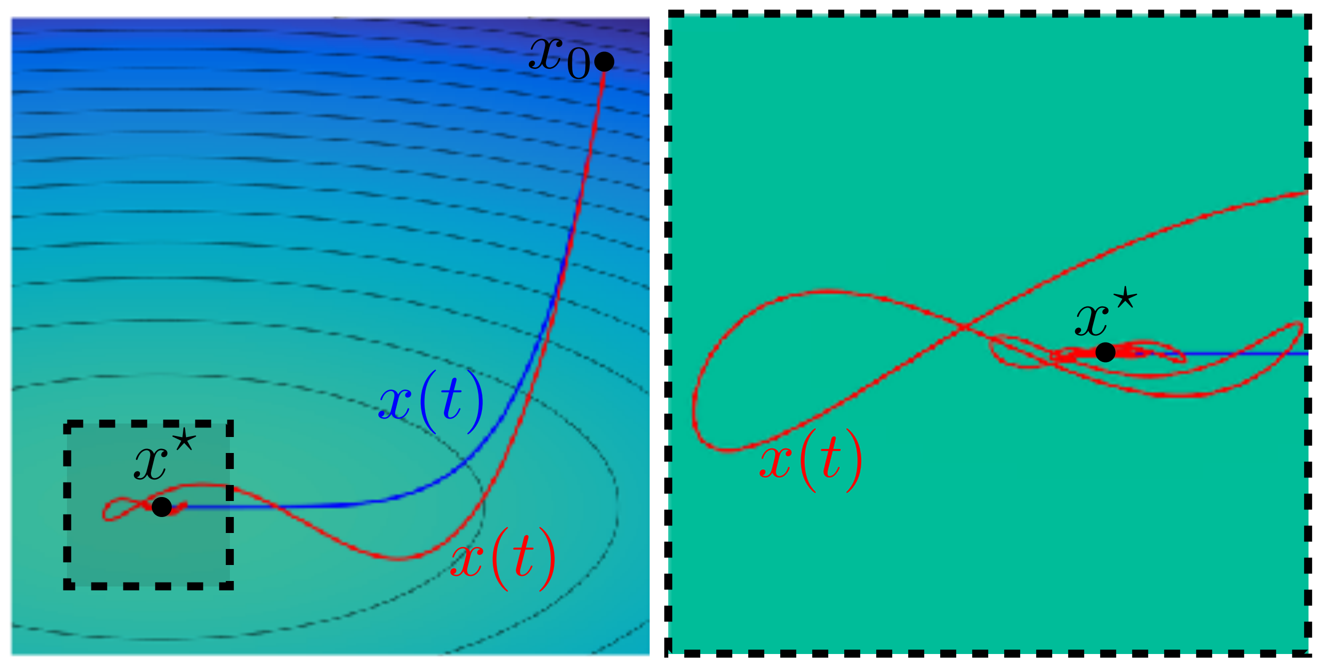
$\tau \to 0$ $\quad$ $k\sqrt{\tau} \to t$

$$\frac{\mathrm{d}^2 x(t)}{\mathrm{d}t^2} + \frac{3}{t}\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\nabla f(x(t))$$

Theorem:

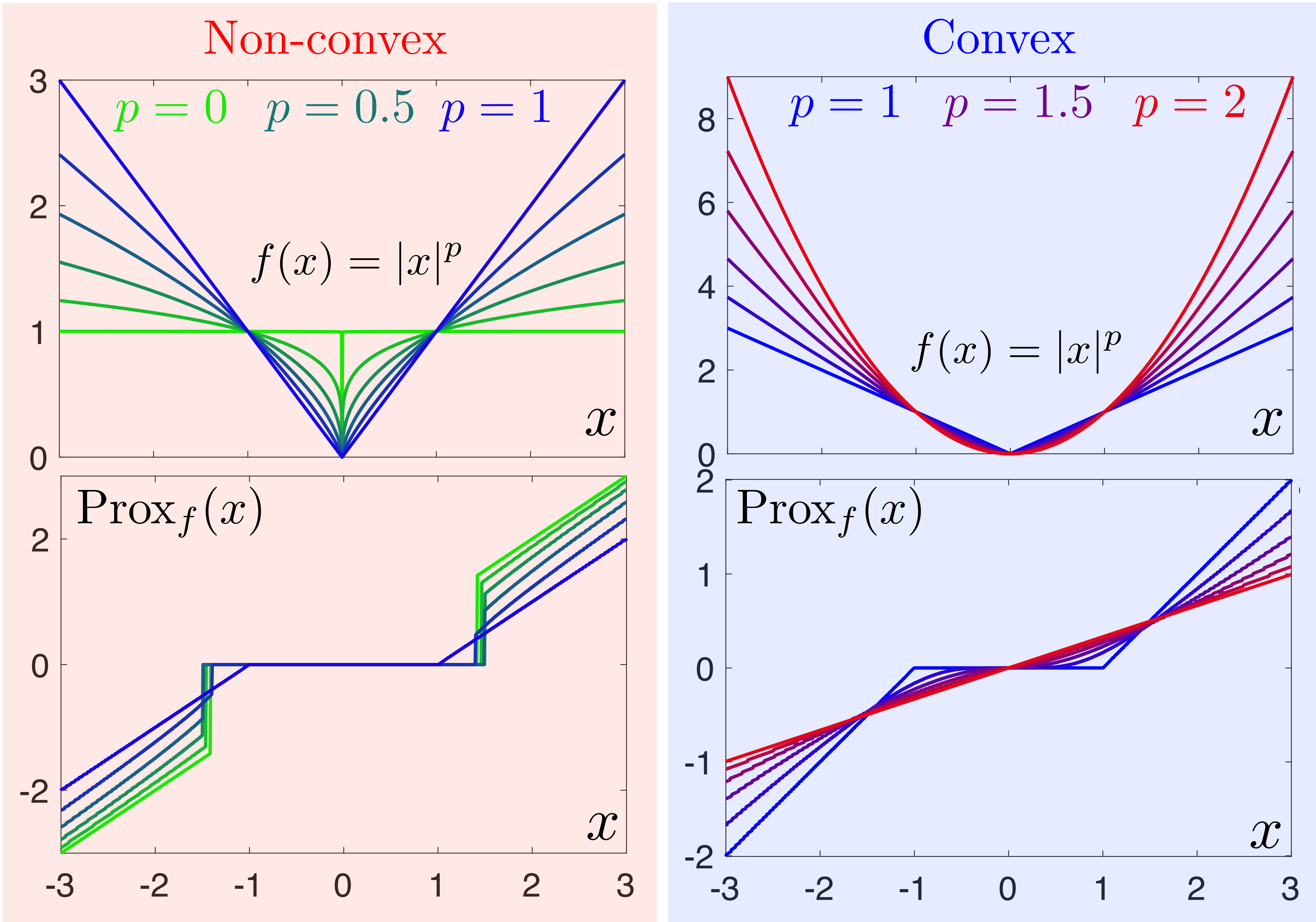
$$f(x_k) - f(x^\star) = O(1/k)$$

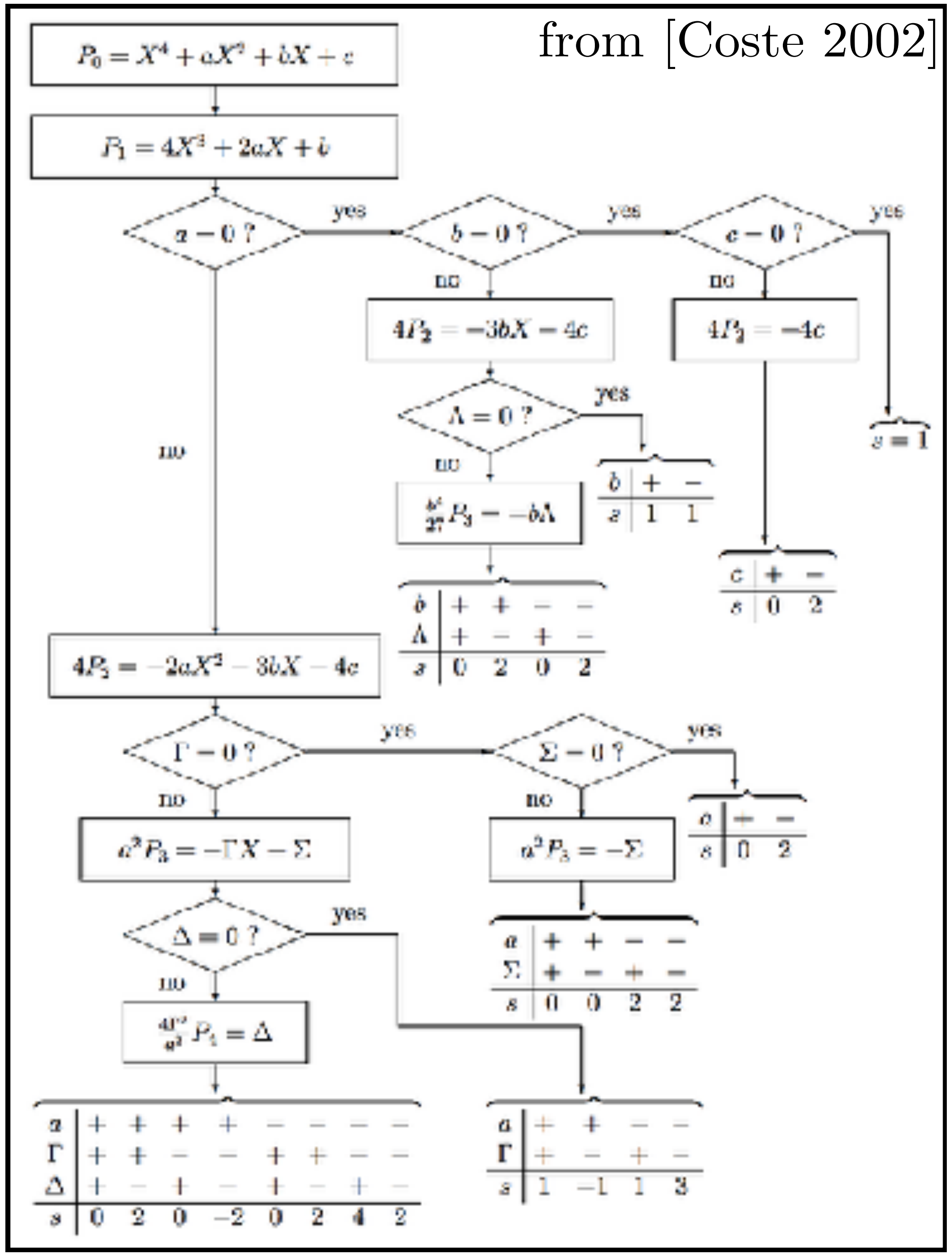
$$f(x_k) - f(x^\star) = O(1/k^2)$$



Yurii Nesterov

$$\mathrm{Prox}_f(x) = \underset{x'}{\mathrm{argmin}}\ \frac{1}{2}\|x - x'\|^2 + f(x')$$





algebraic set

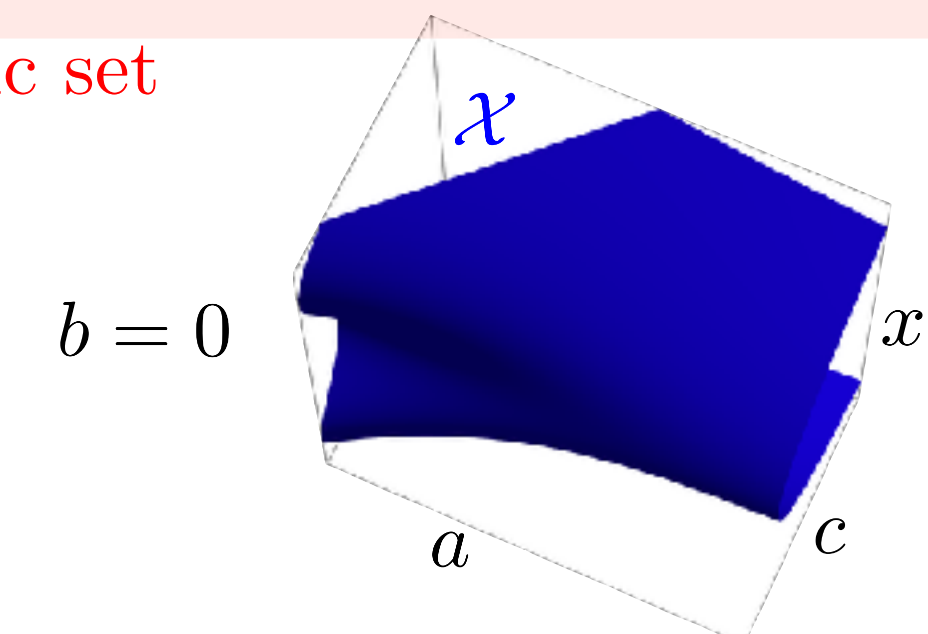
$$\mathcal{X} \stackrel{\text{def.}}{=} \{(a,b,c,X)\ ;\ X^4 + aX^2 + bX + c = 0\}$$

projection

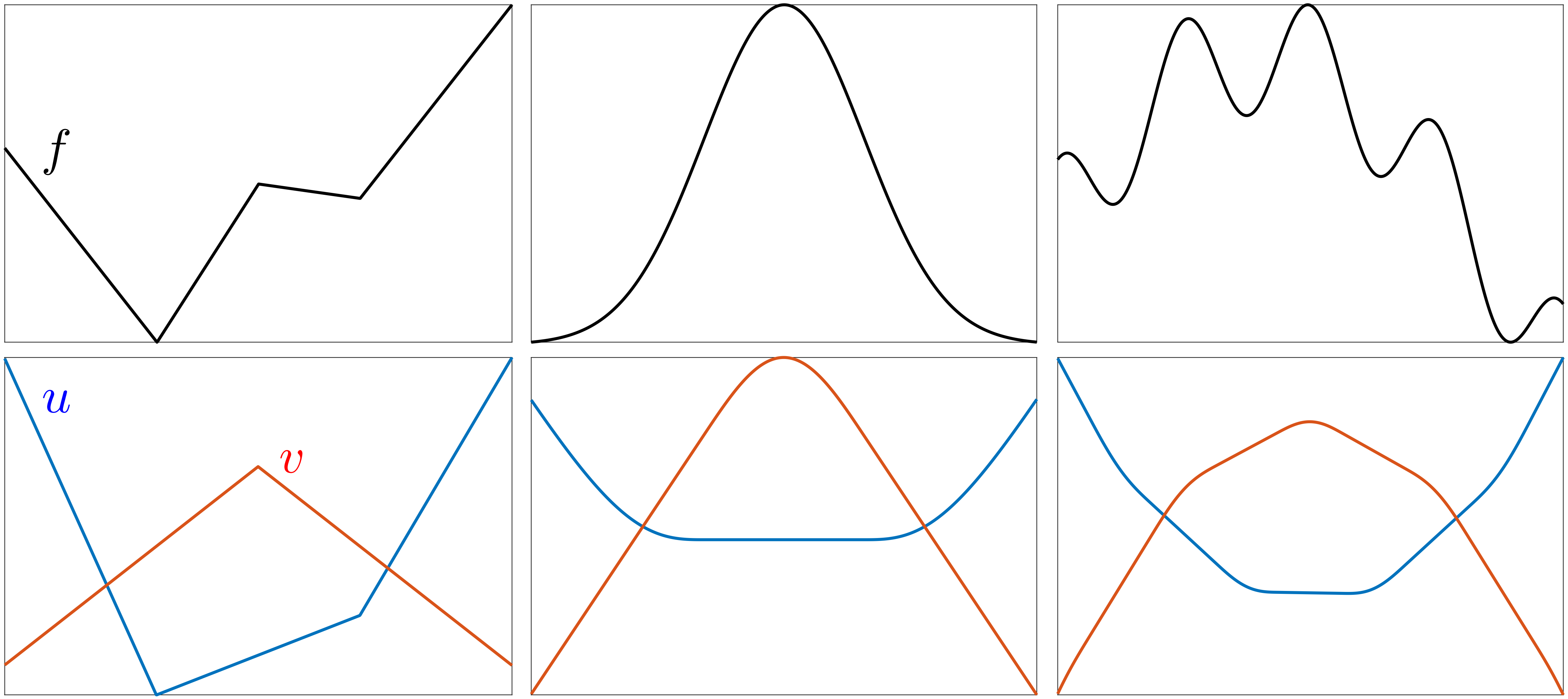
$$(a,b,c,X) \mapsto (a,b,c)$$

$$\{(a,b,c)\ ;\ \exists X \in \mathbb{R},\ X^4 + aX^2 + bX + c = 0\}$$

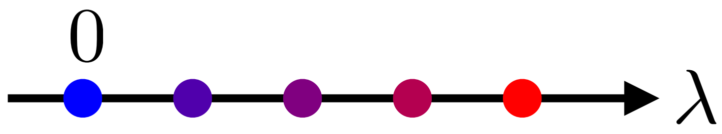
semi-algebraic set

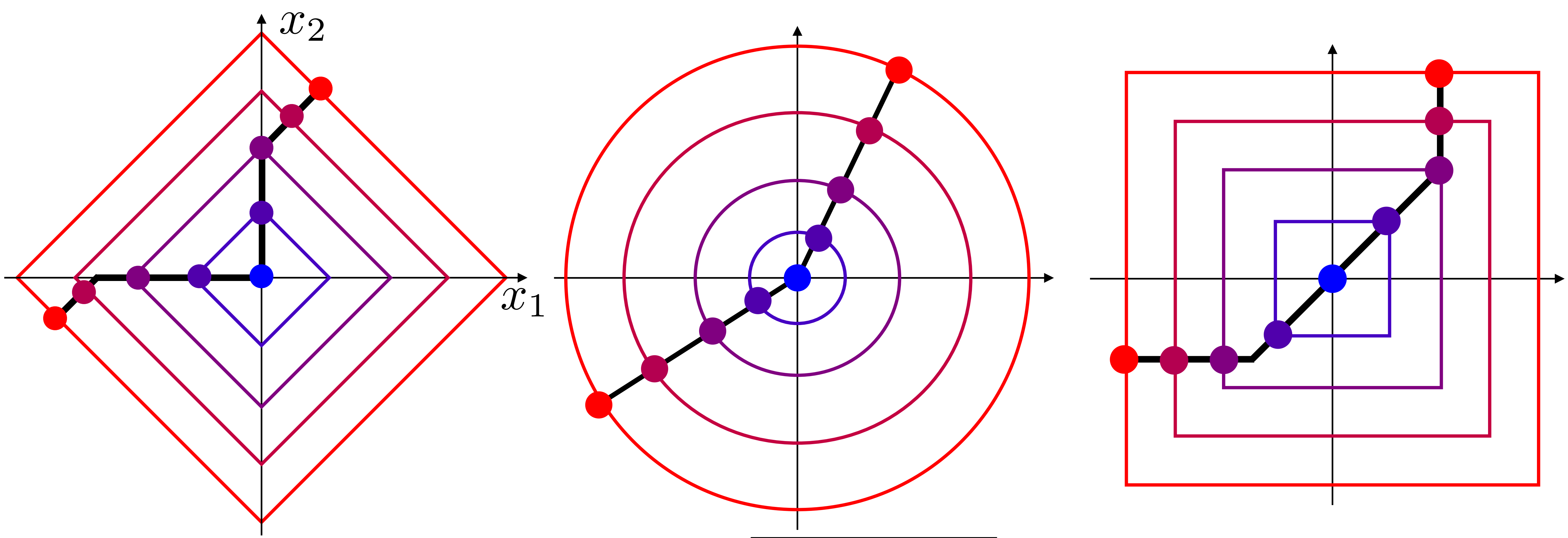


$$f = u + v = \text{convex} + \text{concave}$$



$$\mathrm{Prox}_{\lambda f}(x) = \underset{x'}{\mathrm{argmin}}\ \frac{1}{2}\|x - x'\|^2 + \lambda f(x')$$





$f(x) = |x_1| + |x_2|$ $\quad$ $f(x) = \sqrt{|x_1|^2 + |x_2|^2}$ $\quad$ $f(x) = \max(|x_1|, |x_2|)$

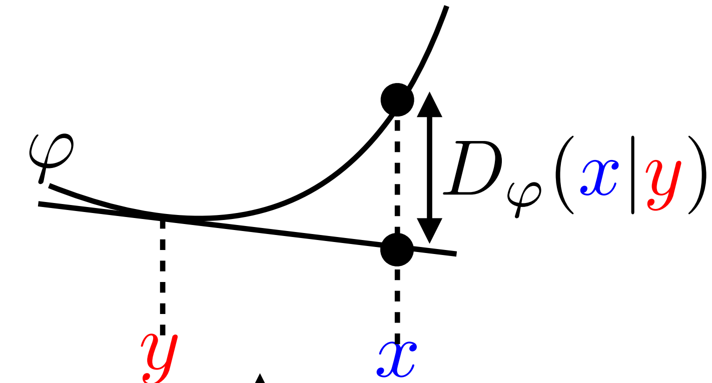
John Toland

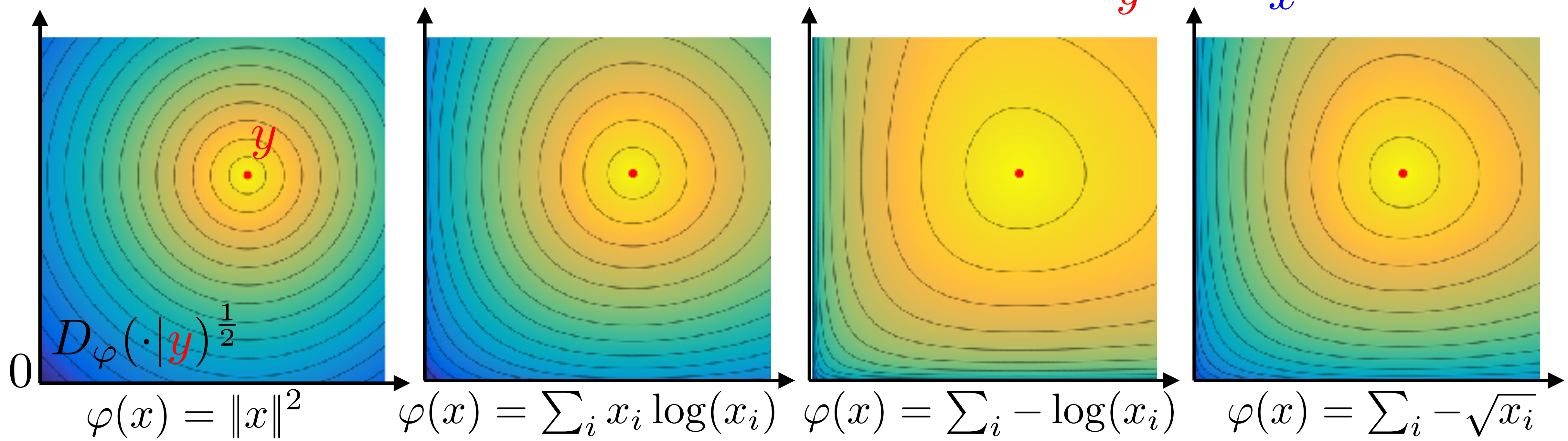
(f, g) convex functions. $$f^*(p) \stackrel{\text{def.}}{=} \sup_x \langle p,\, x\rangle - f(x)$$

Toland's duality: $$\inf f - g = \inf g^* - f^*$$

Bregman divergence:

$$D_\varphi(x|y) \stackrel{\text{def.}}{=} \varphi(x) - \varphi(y) - \langle x - y,\, \nabla\varphi(y)\rangle$$

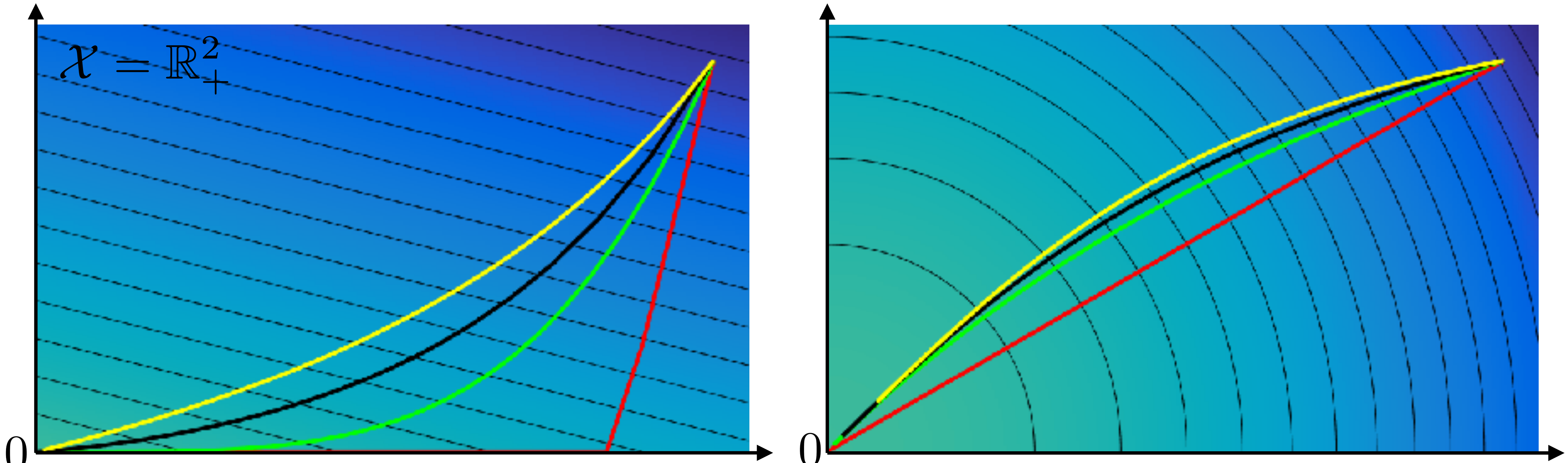




$$\left.\begin{array}{l} D_\varphi(x|x+\varepsilon) \\ D_\varphi(x+\varepsilon|x) \end{array}\right\} = \frac{1}{2}\langle \partial^2\varphi(x)\varepsilon,\, \varepsilon\rangle + o(\|\varepsilon\|^2)$$

Bregman divergence: $$D_\varphi(x|y) \overset{\text{def.}}{=} \varphi(x) - \varphi(y) - \langle x - y, \nabla\varphi(y) \rangle$$

Mirror descent: $$x_{k+1} = \underset{x \in \mathcal{X}}{\operatorname{argmin}} \; D_\varphi(x|x_k) + \tau \langle \nabla f(x_k), x \rangle$$
$$= (\nabla\varphi)^{-1} \left(\nabla\varphi(x_k) - \tau \nabla f(x_k) \right)$$



$\varphi(x) = \|x\|^2$ $\varphi(x) = \sum_i -\log(x_i)$ $\varphi(x) = \sum_i x_i \log(x_i)$ $\varphi(x) = \sum_i -\sqrt{x_i}$

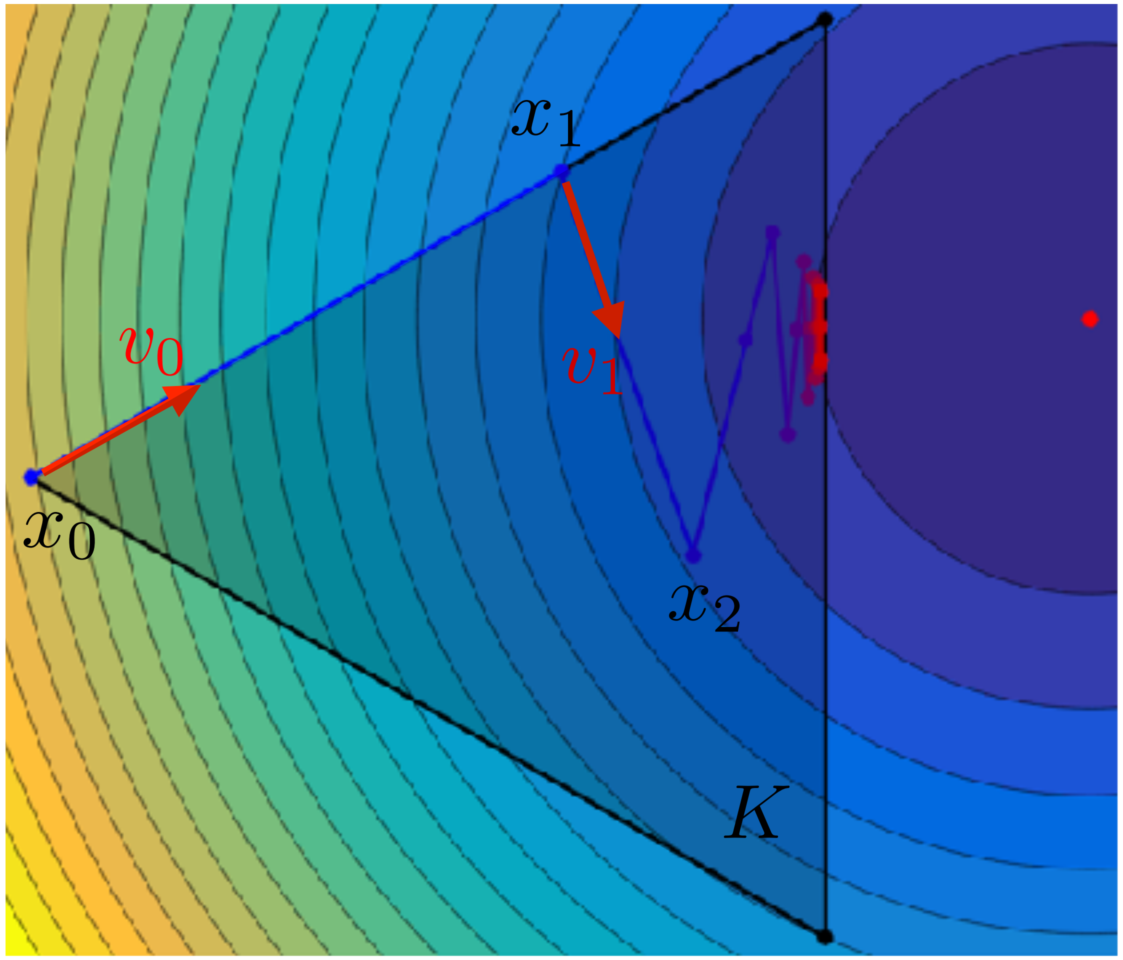
Directional derivative: $D_v f(x) \overset{\text{def.}}{=} \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$

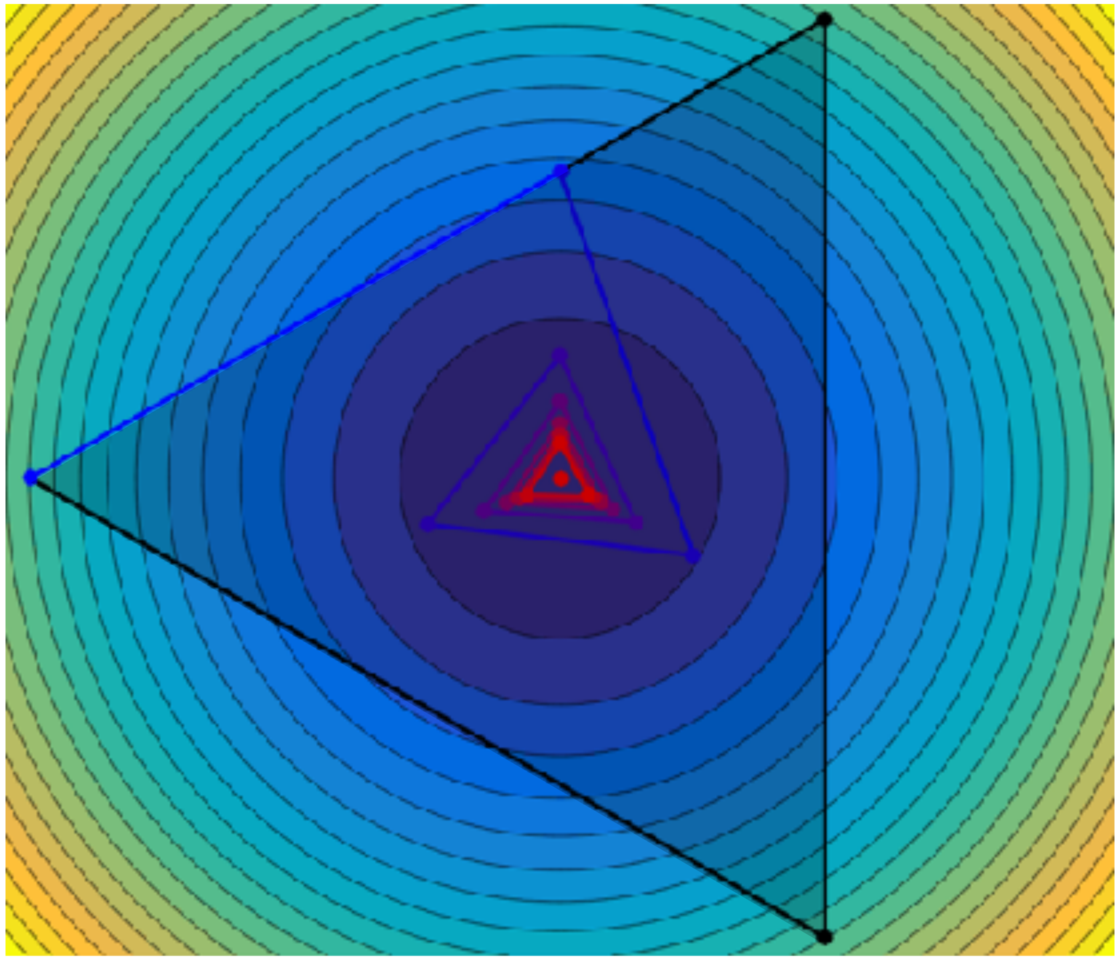
$$\min_{x \in K} f(x)$$

Frank-Wolfe

$$v_\ell \overset{\text{def.}}{=} \underset{v \in K}{\operatorname{argmin}} \, D_v f(x_\ell)$$

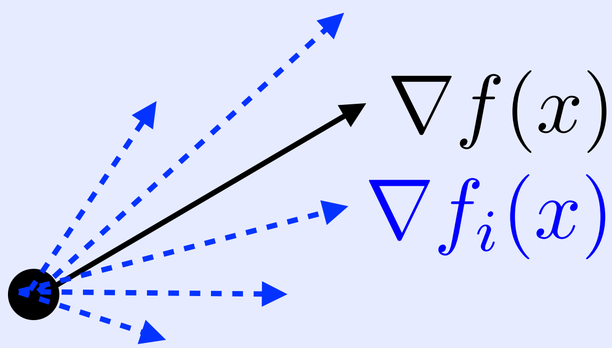
$$x_{\ell+1} \overset{\text{def.}}{=} x_\ell + \frac{2}{2+\ell}(v_\ell - x_\ell)$$





Finite sums

$$f(x) \overset{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^n f_i(x)$$
$$\nabla f(x) = \frac{1}{n}\sum_i \nabla f_i(x)$$

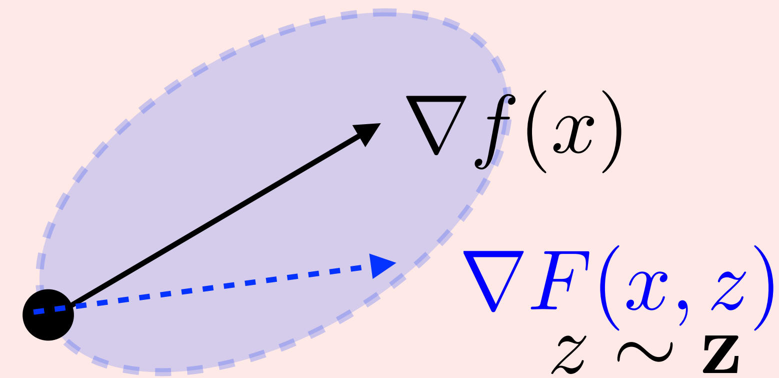


Draw $i \in \{1,\ldots,n\}$ uniformly.

$$x_{k+1} = x_k - \tau_k \nabla f_i(x_k)$$

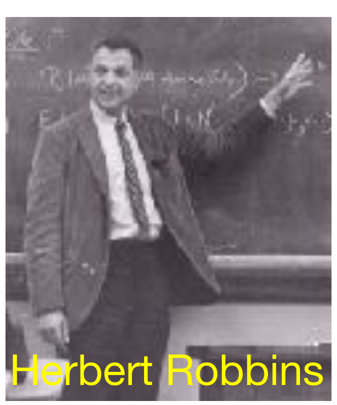
Expectation

$$f(x) \overset{\text{def.}}{=} \mathbb{E}_{\mathbf{z}}(f(x,\mathbf{z}))$$
$$\nabla f(x) = \mathbb{E}_{\mathbf{z}}(\nabla F(x,\mathbf{z}))$$



Draw $z \sim \mathbf{z}$

$$x_{k+1} = x_k - \tau_k \nabla F(x,z)$$

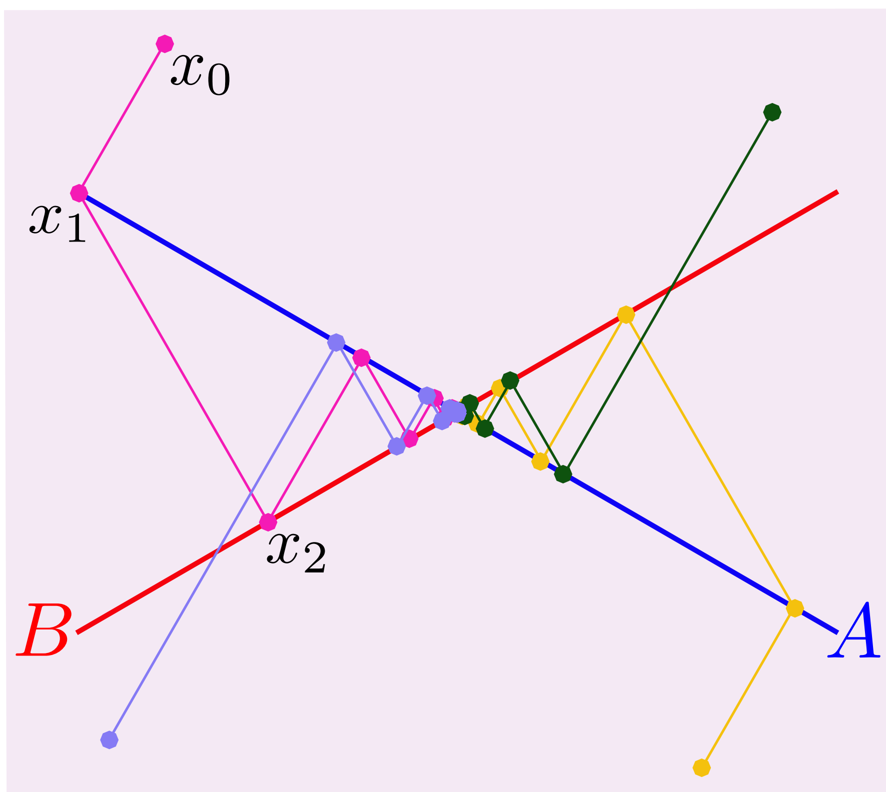


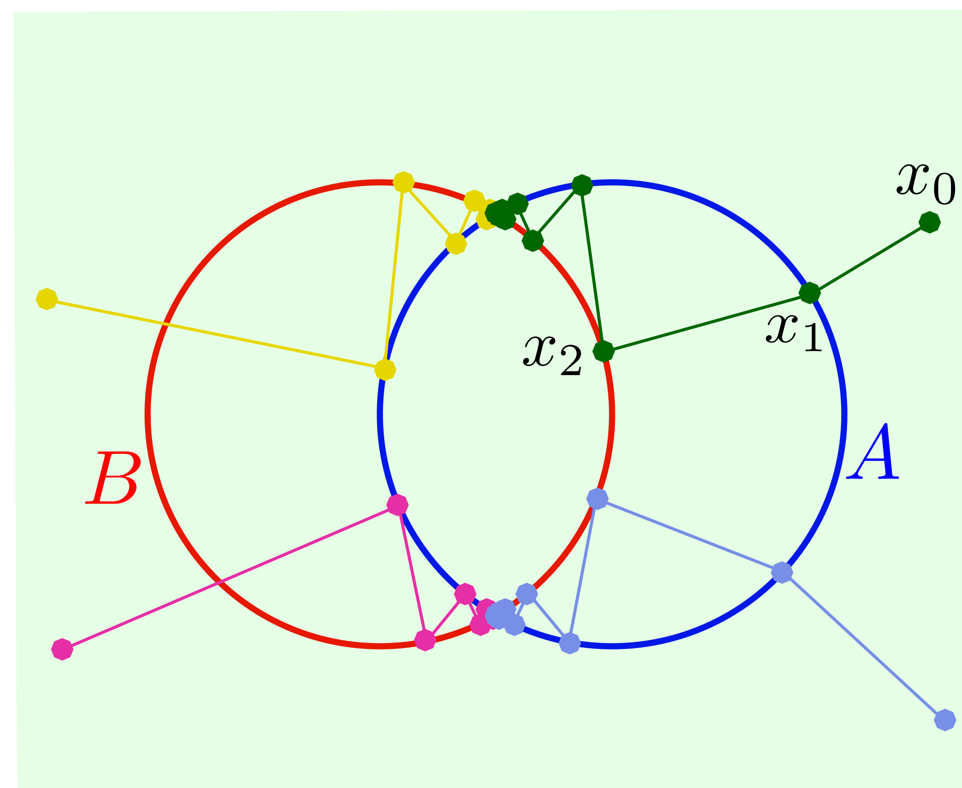
Theorem: If f is strongly convex and $\tau_k \sim 1/k$,

$$\mathbb{E}(\|x_k - x^\star\|^2) = O(1/k)$$

Iterative projections: $\begin{cases} x_{2k+1} = \mathrm{Proj}_A(x_{2k}) \\ x_{2k+2} = \mathrm{Proj}_B(x_{2k+1}) \end{cases}$

Theorem: if (A, B) convex, $x_k \xrightarrow{k \to +\infty} A \cap B$





Iterative Projections

$$x_{k+1} = P_B(P_A x_k)$$

$$P_A \overset{\text{def.}}{=} \mathrm{Proj}_A$$

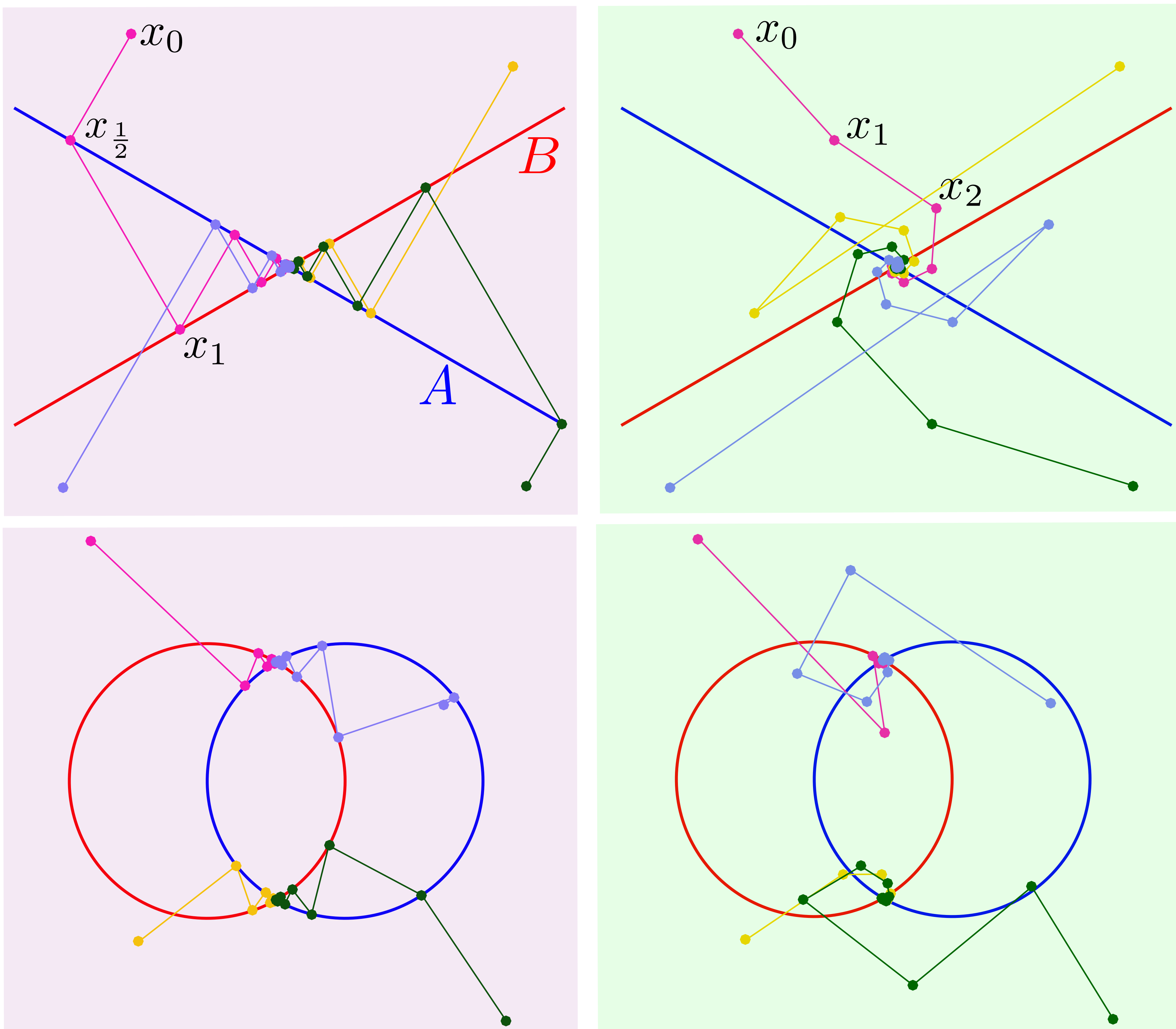
Douglas-Rachford

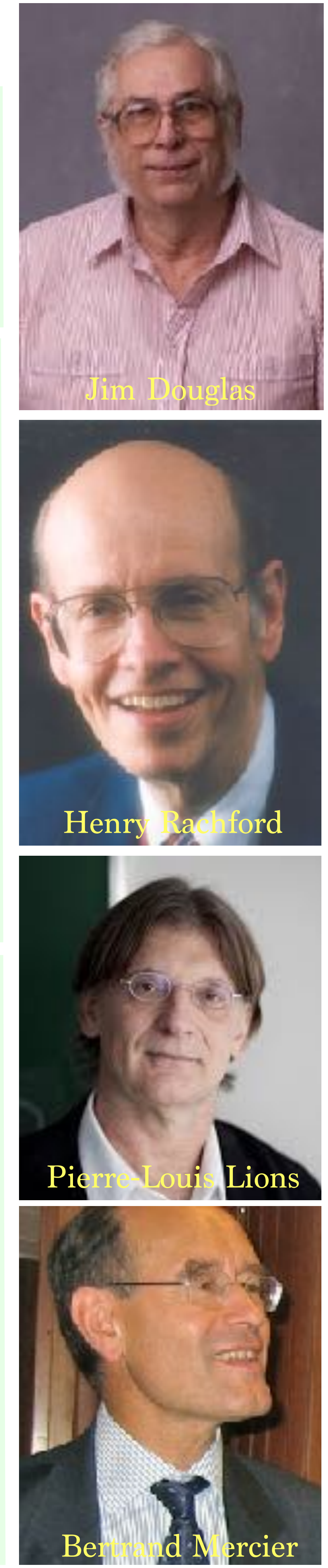
$$x_k = \bar{P}_A(y_k) \overset{\text{def.}}{=} 2P_A(y_k) - y_k$$

$$y_{k+1} = \frac{1}{2} y_k + \frac{1}{2} \bar{P}_B(x_k)$$

Convex

Non-convex

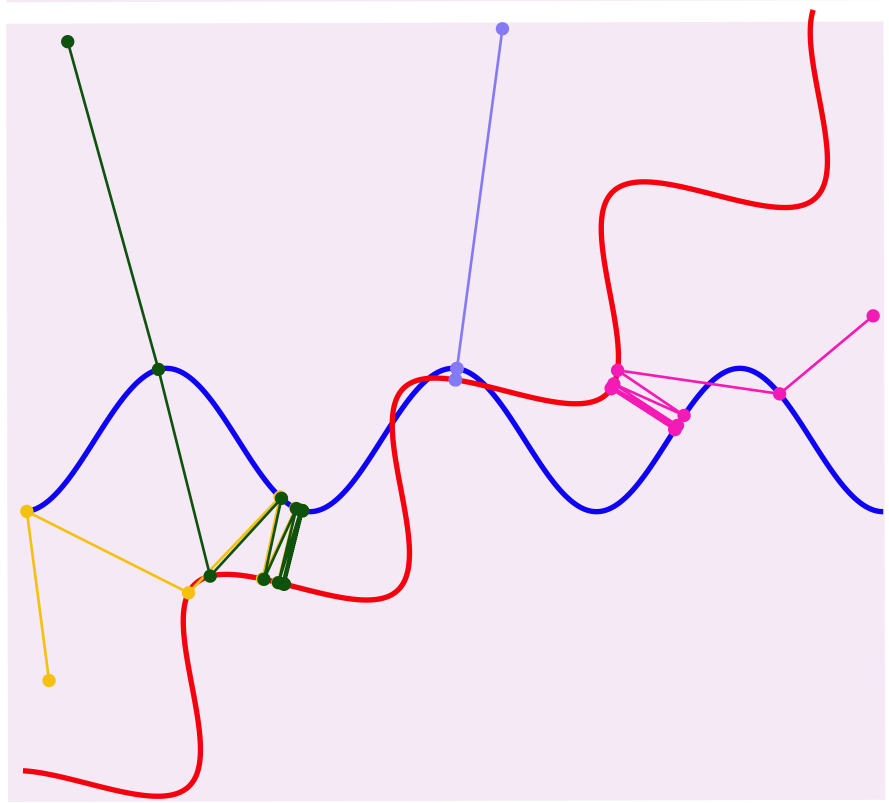




Iterative Projections

$$x_{k+1} = P_B(P_A x_k)$$

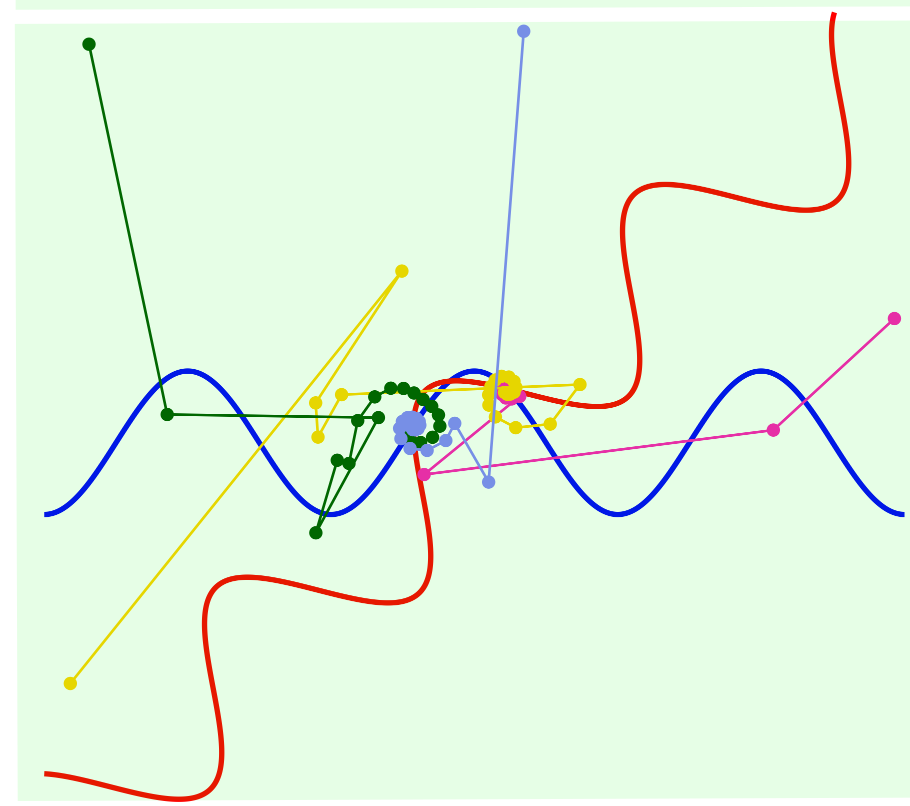
$$P_A \overset{\text{def.}}{=} \text{Proj}_A$$

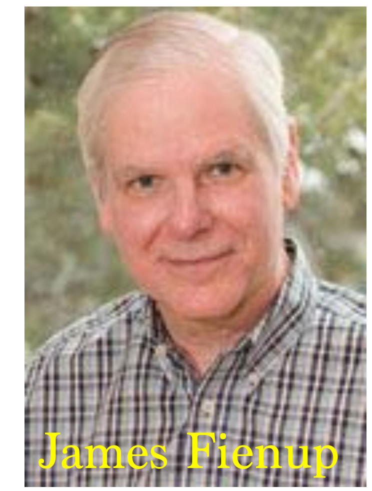


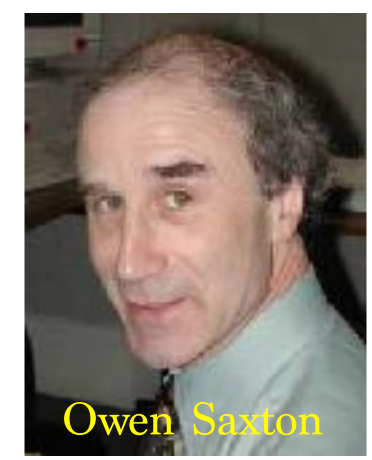
Douglas-Rachford

$$x_k = \bar{P}_A(y_k) \overset{\text{def.}}{=} 2P_A(y_k) - y_k$$

$$y_{k+1} = \frac{1}{2} y_k + \frac{1}{2} \bar{P}_B(x_k)$$



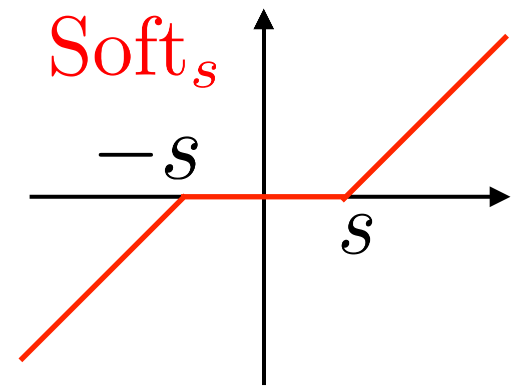


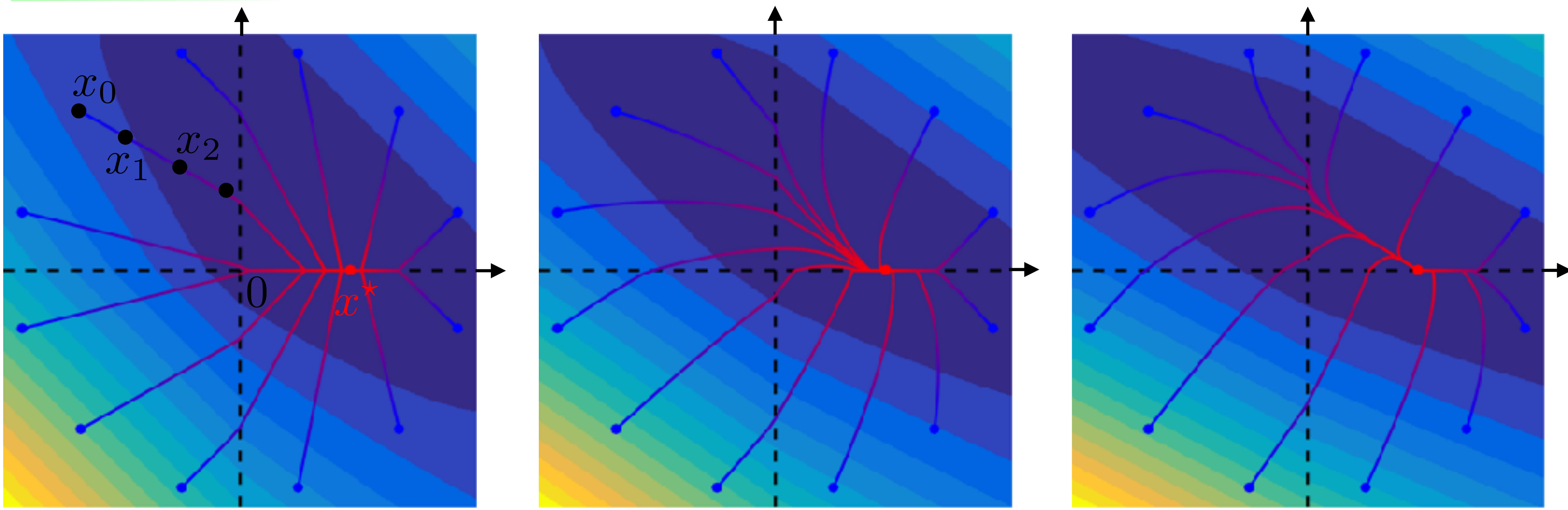


Lasso: $\min_x \frac{1}{2}\|y - Ax\|^2 + \lambda\|x\|_1$

Fwd-Bwd (ISTA): $x_{k+1} \stackrel{\text{def.}}{=} \text{Soft}_{\tau\lambda}(x_k - \tau A^\top(Ax_k - y))$

Theorem: if $0 < \tau < 2/\|A\|^2$, $x_k \to x^\star$ solution of Lasso.





$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $y = A\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\lambda = 0.3$

$A = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}$

$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$

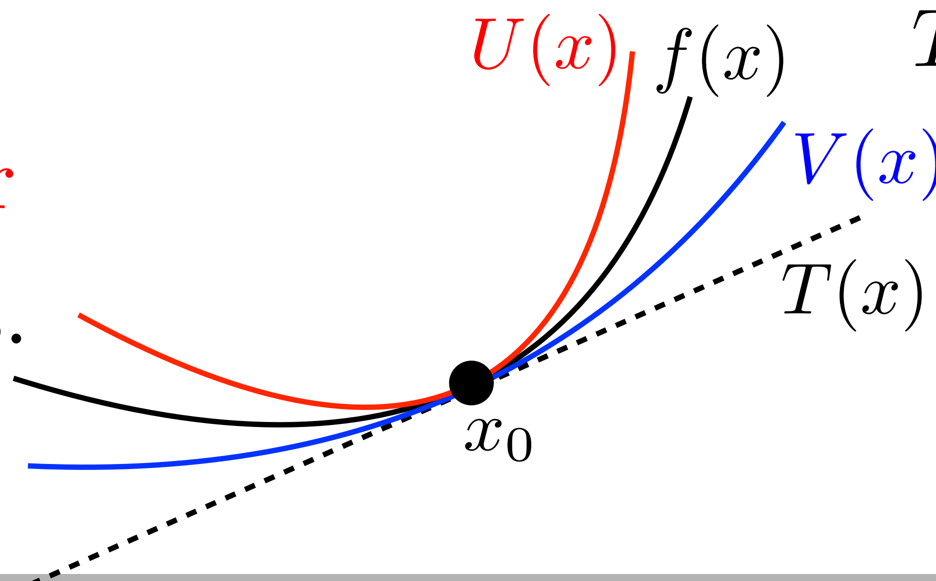
Hypotheses: $\color{blue}{\mu \mathrm{Id}_n} \preceq \partial^2 f(x) \preceq \color{red}{L\, \mathrm{Id}_n}$

strong convexity — smoothness

Conditionning:

$$\varepsilon \overset{\text{def.}}{=} \frac{\mu}{L} \leqslant 1$$

Quadratic lower / upper approximants.



$$T(x) \overset{\text{def.}}{=} f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$$

$$U(x) \overset{\text{def.}}{=} T(x) + \frac{L}{2}\|x - x_0\|^2$$

$$V(x) \overset{\text{def.}}{=} T(x) + \frac{\mu}{2}\|x - x_0\|^2$$

Gradient descent: $x_{k+1} = x_k - \tau_k \nabla f(x_k)$

Theorem:

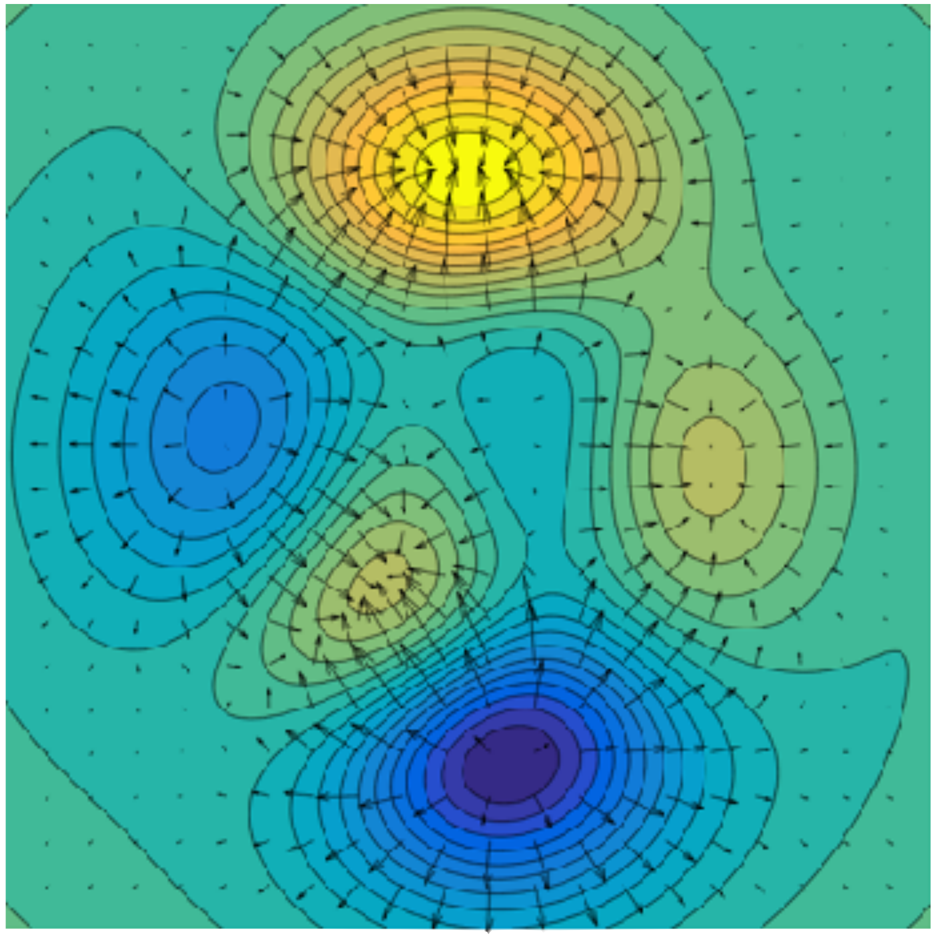
If $L < +\infty$, $0 < \tau < \frac{2}{L}$ $\qquad f(x_k) - f(x^\star) \leqslant \frac{C}{\ell + 1}$

If $\mu > 0$, $L < +\infty$, $0 < \tau < \frac{2}{L}$ $\qquad \|x_k - x^\star\| \leqslant \rho^\ell \|x_0 - x^\star\|$

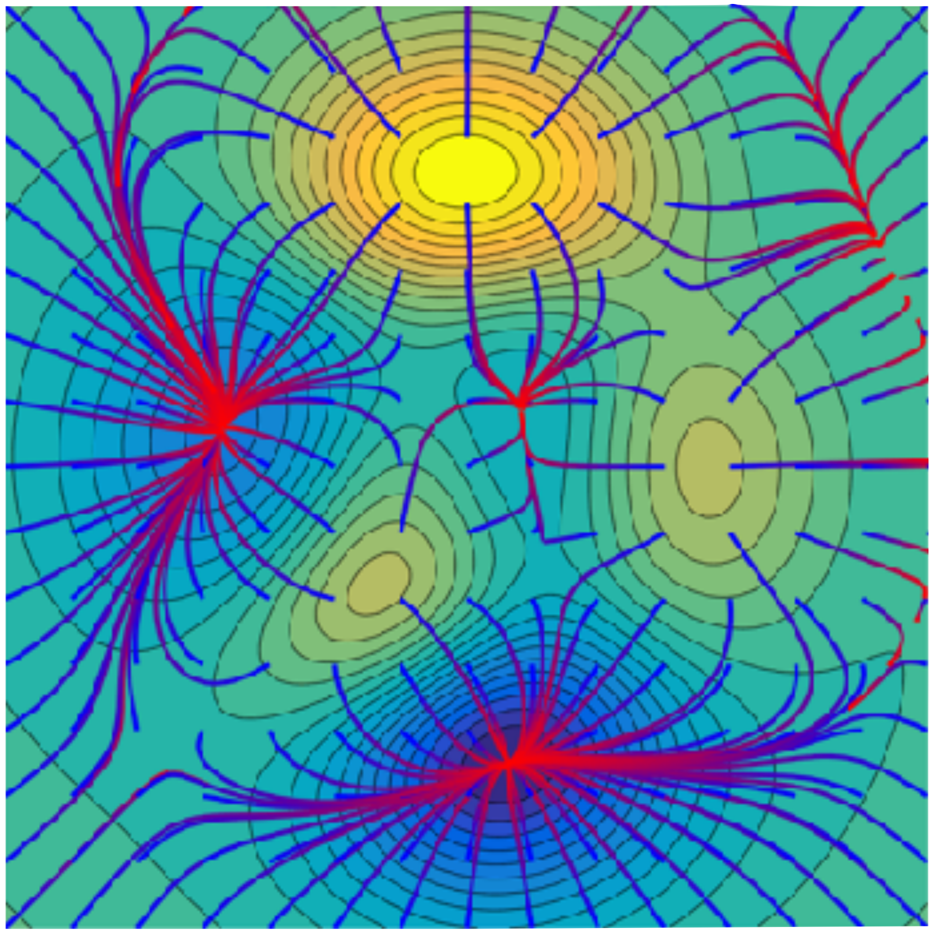
$$\rho = (1 + \varepsilon)^{-\frac{1}{2}} < 1$$

Gradient field: $f(x+\varepsilon) = f(x) + \varepsilon\langle\varepsilon, \nabla f(x)\rangle + o(\varepsilon)$

Gradient flow: $x'(t) = -\nabla f(x(t))$



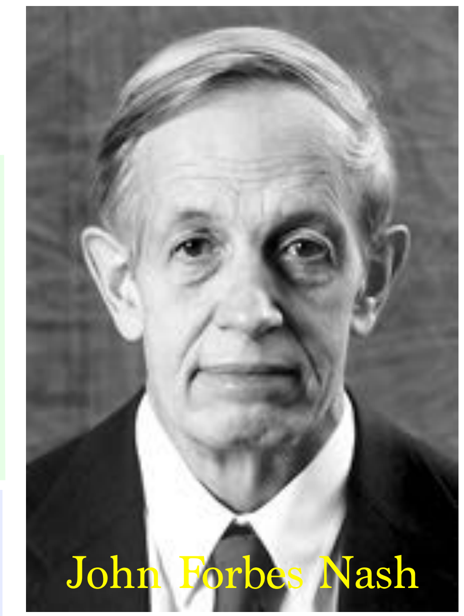
Gradient field ∇f



Gradient flows $x(t)$

$t = 0$ t medium t large

Min-max game: $$\min_{x} \max_{y} f(\overset{\text{convex}}{x}, \overset{\text{concave}}{y}) \geqslant \max_{y} \min_{x} f(x, y)$$

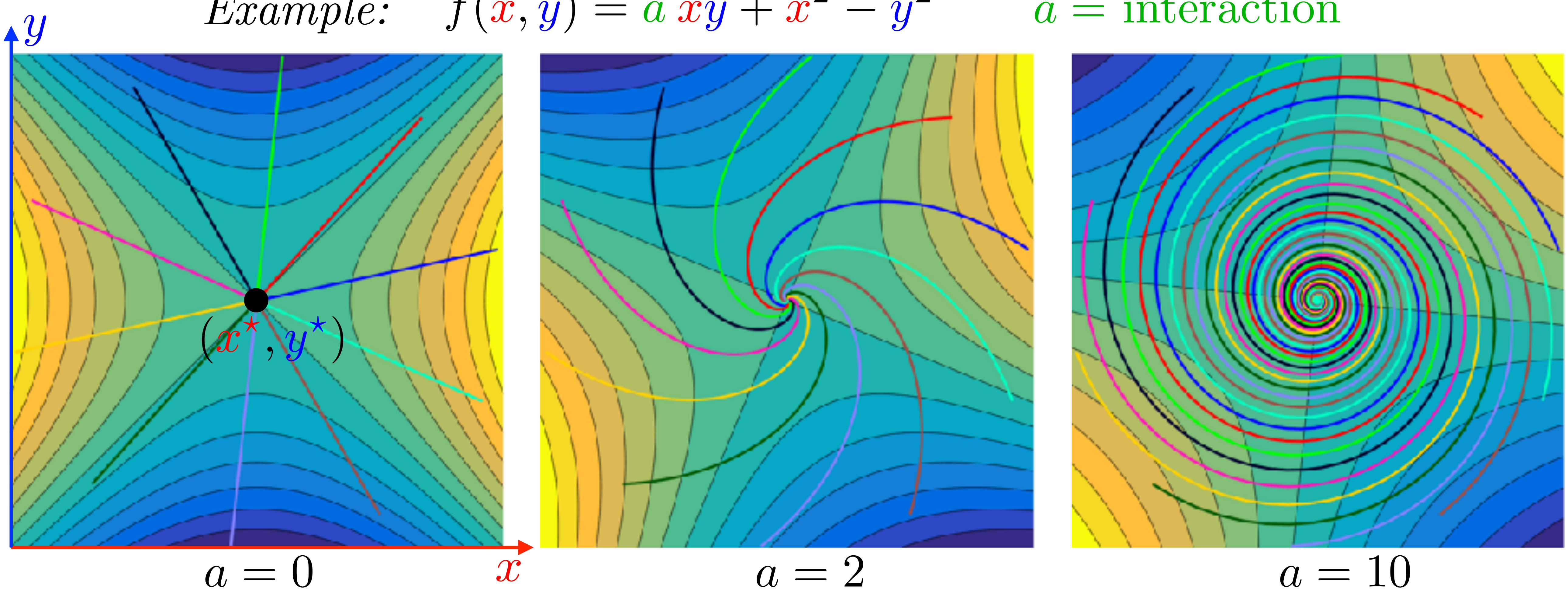


Saddle point $(x^\star, y^\star)$: $$f(x^\star, y) \leqslant f(x^\star, y^\star) \leqslant f(x, y^\star)$$

$\longrightarrow$ Strong duality: $$\min_{x} \max_{y} f(x, y) = \max_{y} \min_{x} f(x, y)$$

Gradient descent: $$\begin{cases} x_{k+1} = x_k - \tau \nabla_x f(x_k, y_k) \\ y_{k+1} = y_k + \tau \nabla_y f(x_k, y_k) \end{cases}$$

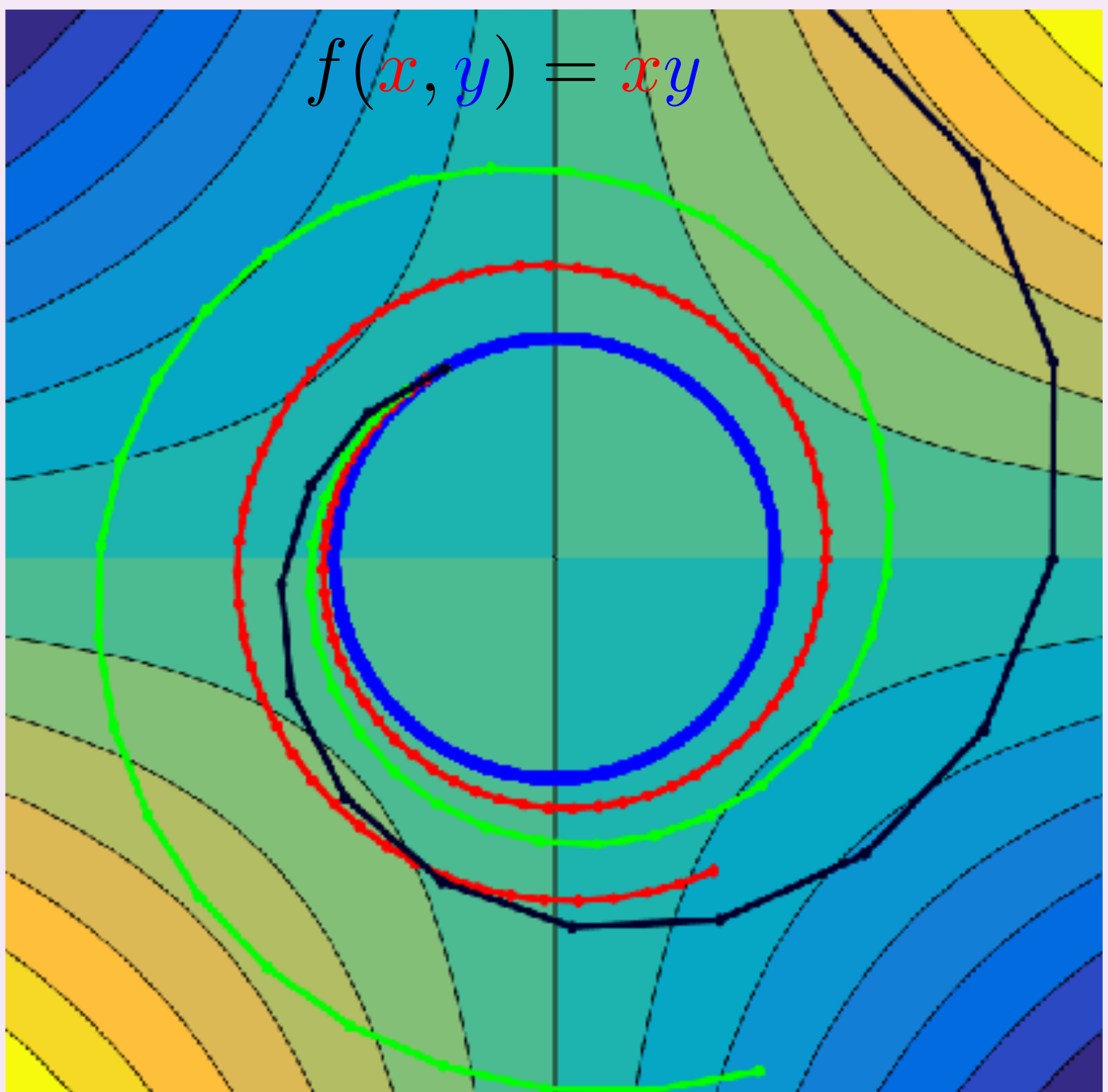
Example: $f(x, y) = a\, xy + x^2 - y^2$ $\quad a =$ interaction



Min-max game: $\min_{x} \max_{y} f(x,y) \geqslant \max_{y} \min_{x} f(x,y)$

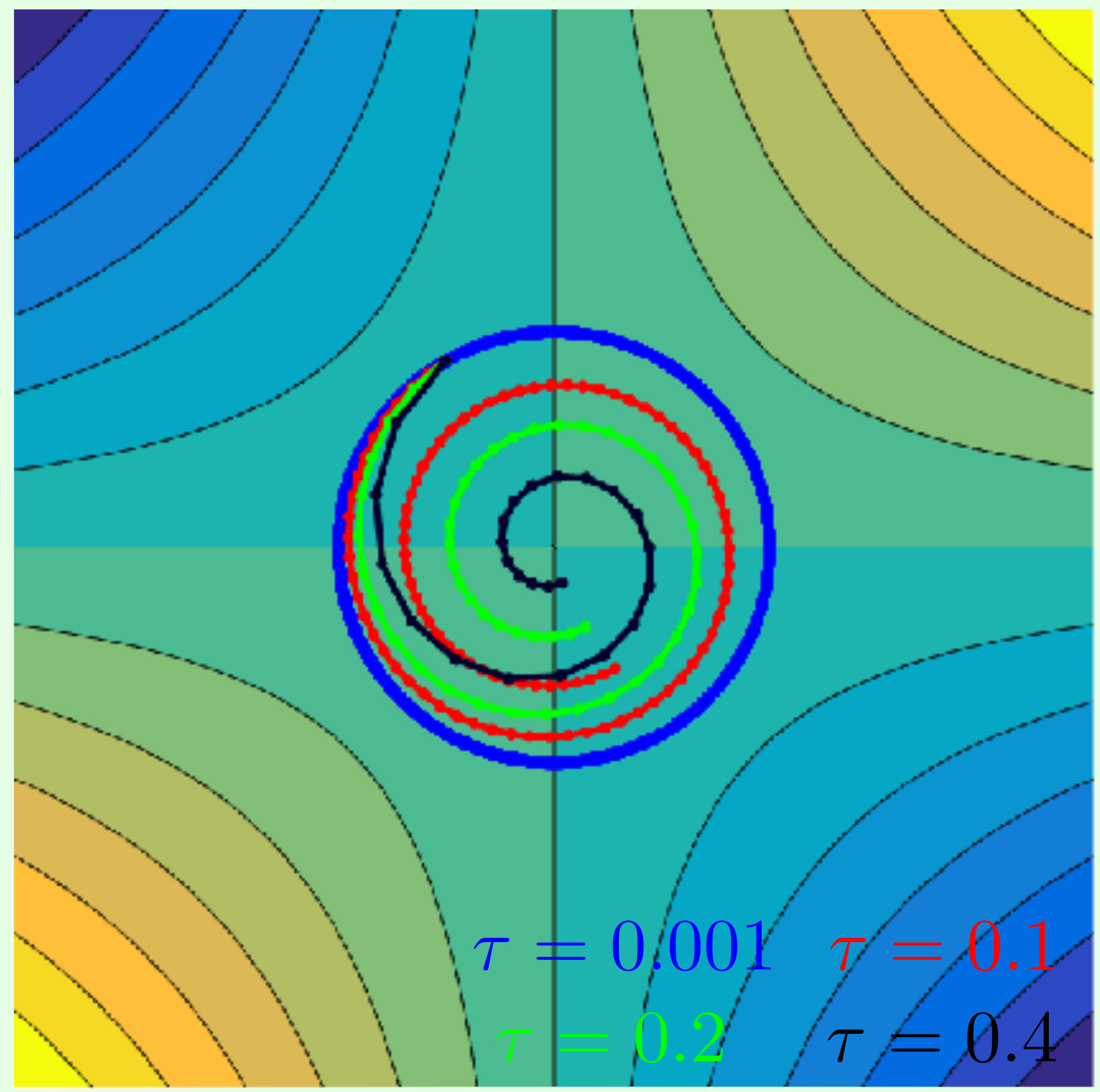
Explicit

$$\begin{cases} x_{k+1} = x_k - \tau \nabla_x f(x_k, y_k) \\ y_{k+1} = y_k + \tau \nabla_y f(x_k, y_k) \end{cases}$$



Implicit

$$\begin{cases} x_{k+1} = x_k - \tau \nabla_x f(x_{k+1}, y_{k+1}) \\ y_{k+1} = y_k + \tau \nabla_y f(x_{k+1}, y_{k+1}) \end{cases}$$



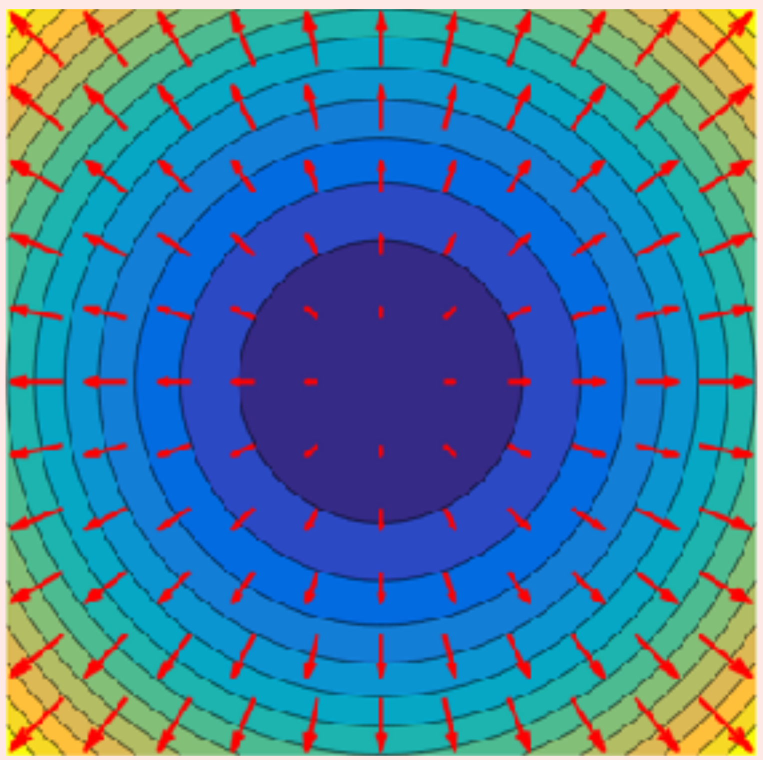
Monotone operator $V : \mathbb{R}^d \to \mathbb{R}^d$: $\quad \forall (x,y) \in (\mathbb{R}^d)^2,\ \langle V(x) - V(y),\, x - y \rangle \geqslant 0$

Convex optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

Theorem: ∇f is monotone.

f *quadratic:* ∇f symmetric.



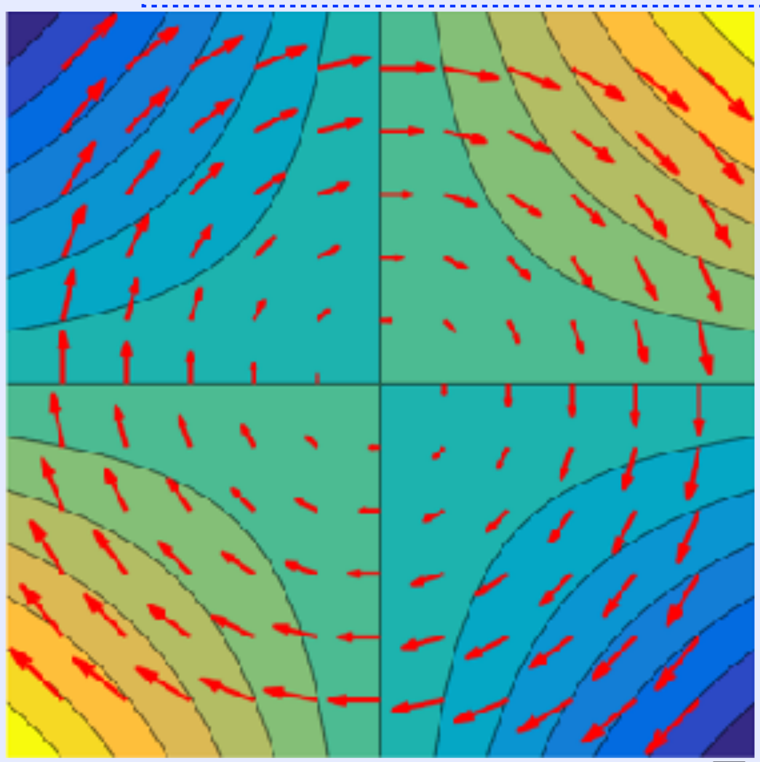
$f(x) = \|x\|^2/2 \qquad V(x) = x$

Saddle point

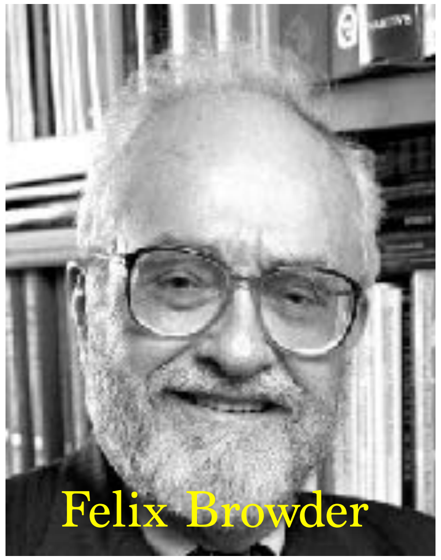
$$\min_{x_1 \in \mathbb{R}^{d_1}} \max_{x_2 \in \mathbb{R}^{d_2}} f(x_1) + \langle A x_1, x_2 \rangle - g(x_2)$$

Theorem: $\begin{pmatrix} \nabla f & A^* \\ -A & \nabla g \end{pmatrix}$ is monotone.

$f = g = 0$: $\begin{pmatrix} 0 & A^* \\ -A & 0 \end{pmatrix}$ skew-symmetric.



$A = 1 \quad V(x) = (x_2, -x_1)^\top$

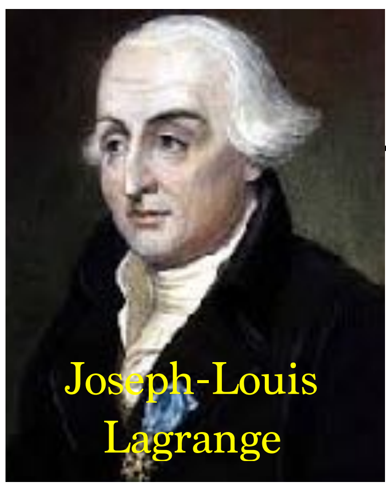


The two "extremal" cases [Edgar Asplund, 1970]

Legendre-Fenchel transform:

$$f^*(u) \overset{\text{def.}}{=} \min_{x} \langle x,\, u\rangle - f(x)$$

$$\min_{x} f(Ax) + g(x) \quad = \min_{Ax=y} f(y) + g(x)$$



$$= \min_{x,y} \max_{u} f(y) + g(x) + \langle Ax - y,\, u\rangle$$

$$= \max_{u} \left[\min_{y} \langle -y,\, u\rangle + f(y)\right] + \left[\min_{x} \langle x,\, A^* u\rangle + g(x)\right]$$

$$= \max_{u} \quad -f^*(u) \quad -g^*(-A^* u)$$

Primal-dual relations:

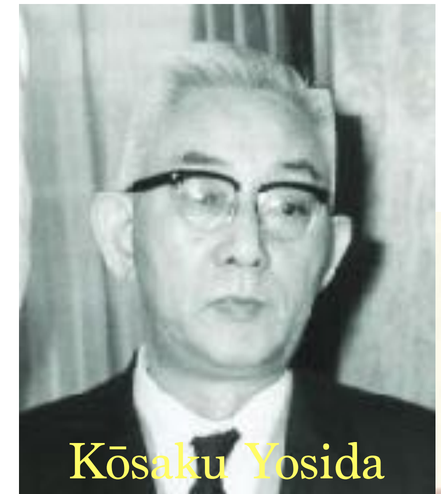
g smooth		g strongly convex
$\nabla g(x) = -A^* u$	$\Longleftrightarrow$	$x = \nabla g^*(-A^* u)$

Moreau-Yosida regularization:

$$f_\mu(x) \overset{\text{def.}}{=} \min_y f(y) + \frac{1}{2\mu}\|x-y\|^2$$

Proximal operator:

$$\text{Prox}_{\mu f} \overset{\text{def.}}{=} \operatorname{argmin}_y f(y) + \frac{1}{2\mu}\|x-y\|^2$$

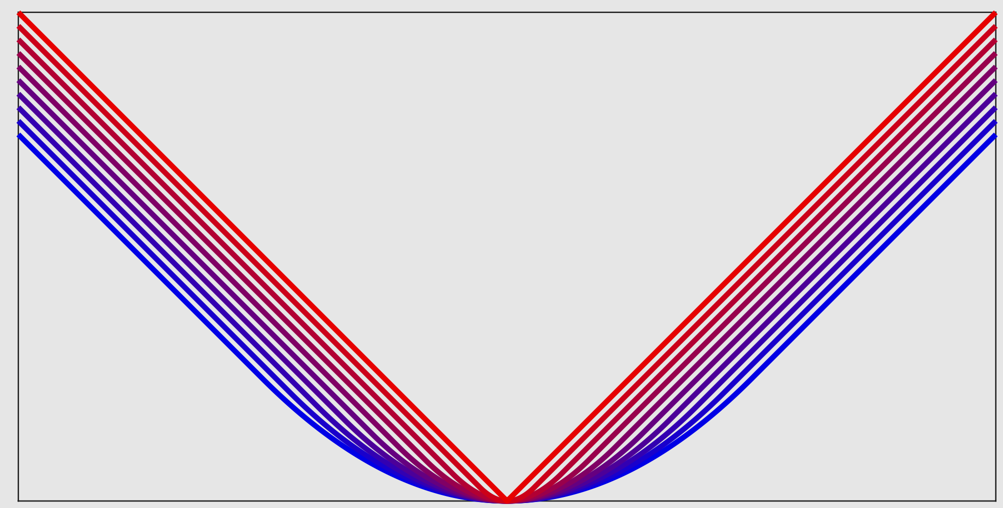


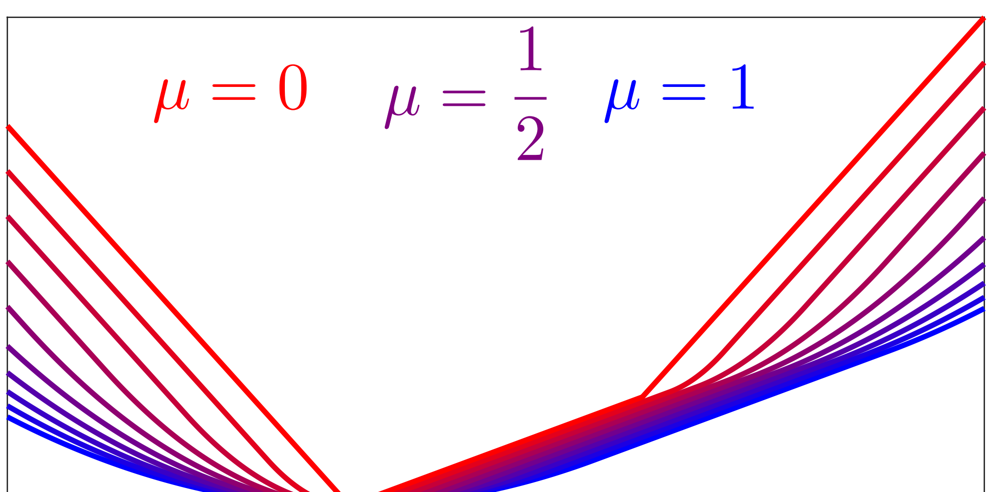
Prop: ∇f_μ is $1/\mu$-Lipscthitz and

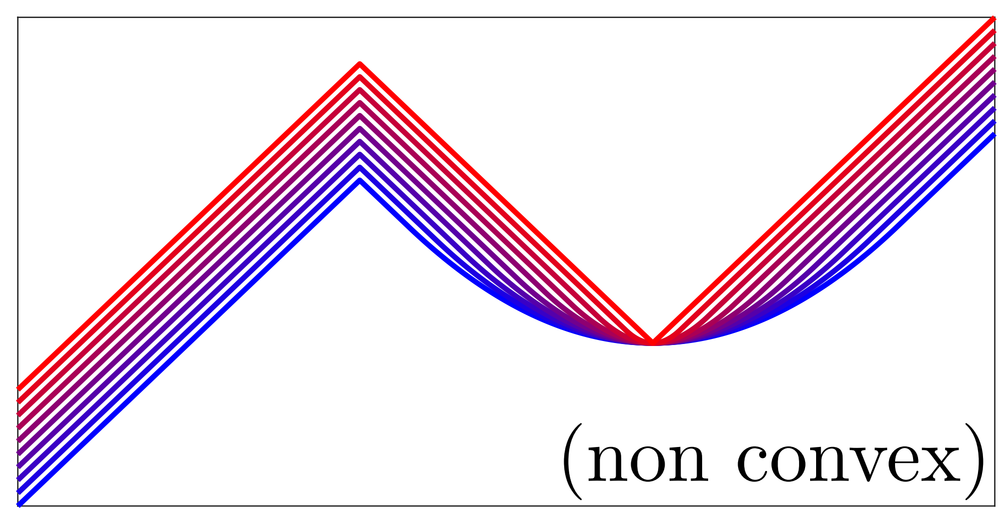
$$\mu \nabla f_\mu(x) = x - \text{Prox}_{\mu f}(x)$$

Huber function: $f(x) = |x|$

$$f_\mu(x) = \begin{cases} x^2/(2\mu) & \text{if} \quad |x| \leqslant \mu \\ |x| - \mu/2 & \text{otherwise.} \end{cases}$$







Fenchel-Legendre transform:

$$f^*(y) = \sup_x \, \langle x,\, y\rangle - f(x)$$

Polar of a set:

$$C^\circ = \{y \; ; \; \forall x \in C, \langle x,\, y\rangle \leqslant 1\}$$

Indicator:

$$\iota_C(x) = \begin{cases} 0 & \text{if} \quad x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Gauge:

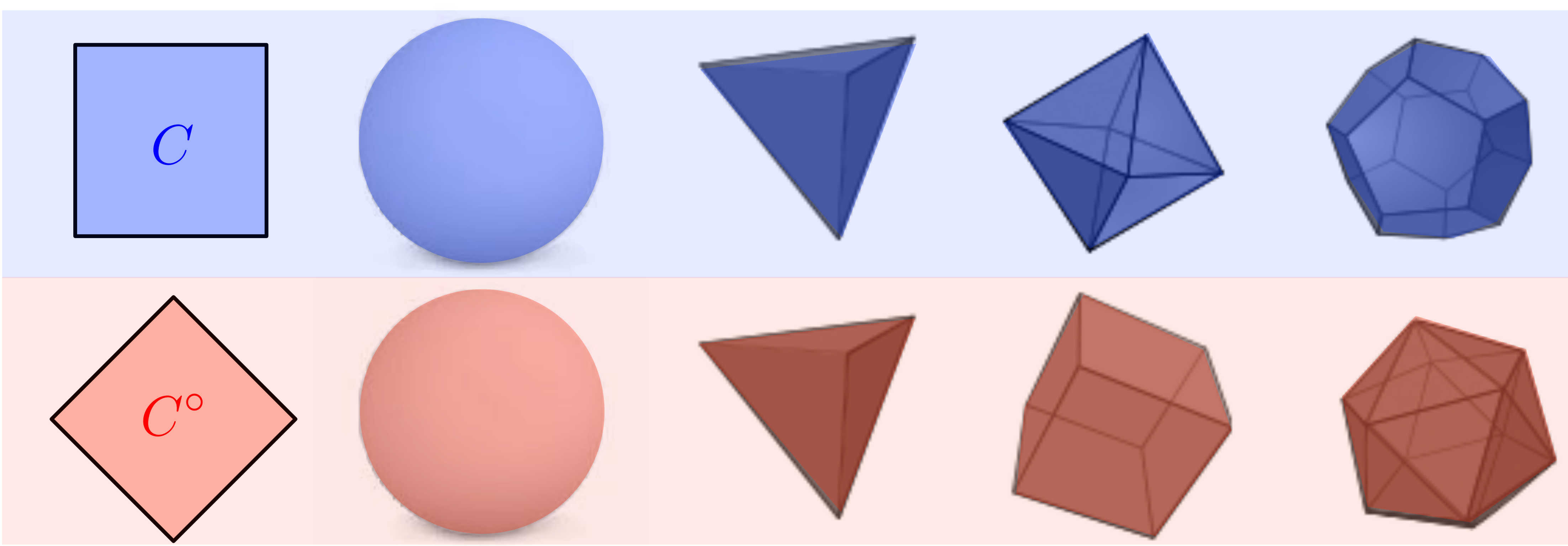
$$\chi_C(x) = \inf\,\{\lambda > 0 \; ; \; x \in \lambda C\}$$

Theorem: If f and C convex, then

$$(C^\circ)^\circ = C \text{ and } (f^*)^* = f.$$

$$(\iota_C)^* = \chi_{C^\circ} \text{ and } (\chi_C)^* = \iota_{C^\circ}.$$

$$\begin{tikzcd}
C = C^{\circ\circ} \arrow[r, shift left, "\boxed{\circ}"] \arrow[d, "\boxed{\iota}"] & C^\circ \arrow[l, shift left] \arrow[d, "\boxed{\chi}"] \\
\iota_C = (\iota_C)^{**} \arrow[r, shift left, "\boxed{*}"] & \chi_{C^\circ} \arrow[l, shift left]
\end{tikzcd}$$

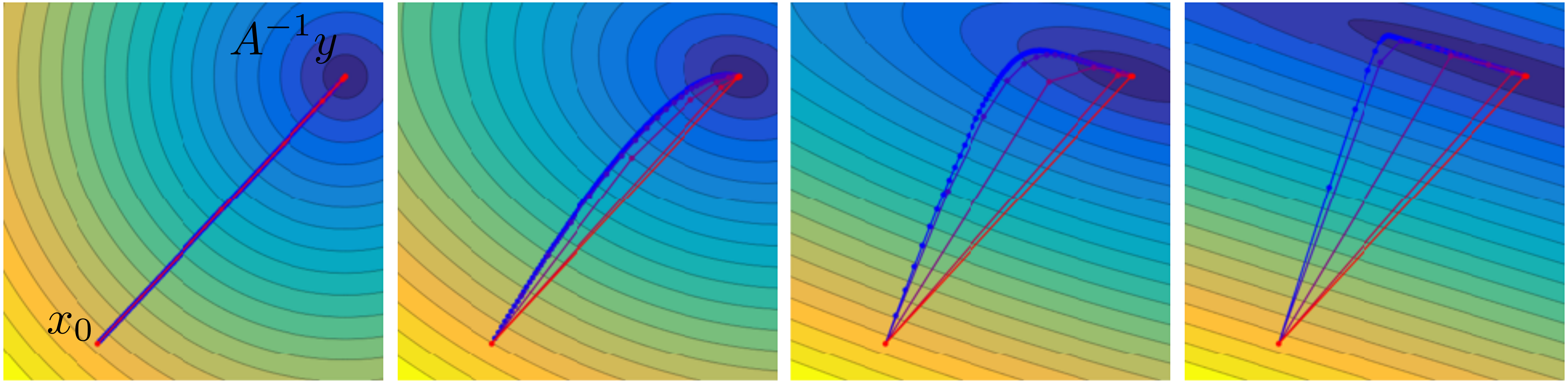


Proximal point: $x_{k+1} \stackrel{\text{def.}}{=} \text{Prox}_{\tau f}(x_k) = \underset{x}{\text{argmin}}\ \frac{1}{2}\|x - x_k\|^2 + \tau f(x)$

Theorem: $\forall\, \tau > 0,\ x_k \xrightarrow{k \to +\infty} x^\star \in \text{argmin}_x\, f(x)$

Example: $f(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle x,\, y\rangle$

$$x_{k+1} = (\text{Id} + \tau A)^{-1}(x_k + \tau y) \xrightarrow{k \to +\infty} A^{-1}y$$



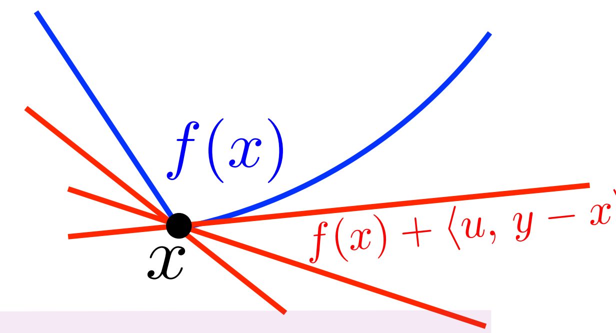
$\tau = 100$ $\tau = 10$ $\tau = 1$ $\tau = 0.1$ $\tau = 0.01$

$f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ convex.

Subdifferential: $\partial f(x) \overset{\text{def.}}{=} \{u \in \mathbb{R}^d \,;\, \forall\, y, f(y) \geqslant f(x) + \langle u,\, y - x\rangle\}$

Theorem: $\operatorname{argmin} f = \{x \,;\, 0 \in \partial f(x)\}$

f is differentiable at $x \Leftrightarrow \partial f(x) = \{\nabla f(x)\}$.



∂f is monotone: $\forall\,(u,v) \in \partial f(x) \times \partial f(y), \langle u - v,\, x - y\rangle \geqslant 0$.