Convex Optimization for Imaging

Gabriel Peyré

www.numerical-tours.com

Convex Optimization

Setting: $G : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$

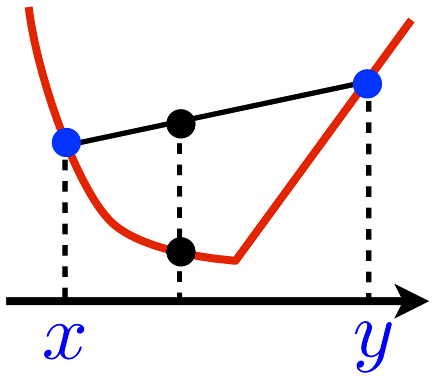
$\mathcal{H}$: Hilbert space. Here: $\mathcal{H} = \mathbb{R}^N$.

Problem: $\min_{x \in \mathcal{H}} G(x)$

Convex Optimization

Setting: $G : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$

$\mathcal{H}$: Hilbert space. Here: $\mathcal{H} = \mathbb{R}^N$.



$$\textit{Problem:} \quad \min_{x \in \mathcal{H}} G(x)$$

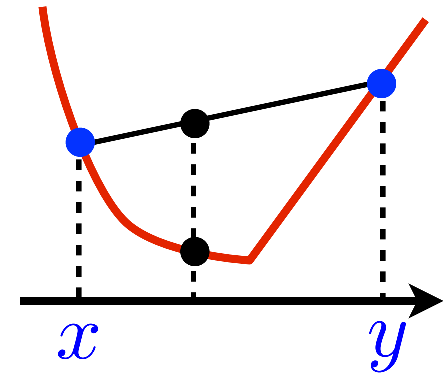
Class of functions:

Convex: $G(tx + (1-t)y) \leqslant tG(x) + (1-t)G(y) \quad t \in [0,1]$

Convex Optimization

Setting: $G : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$

$\mathcal{H}$: Hilbert space. Here: $\mathcal{H} = \mathbb{R}^N$.

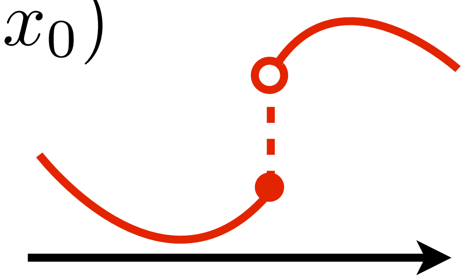


Problem: $\min_{x \in \mathcal{H}} G(x)$

Class of functions:

Convex: $G(tx + (1-t)y) \leqslant tG(x) + (1-t)G(y) \quad t \in [0,1]$

Lower semi-continuous: $\liminf_{x \rightarrow x_0} G(x) \geqslant G(x_0)$

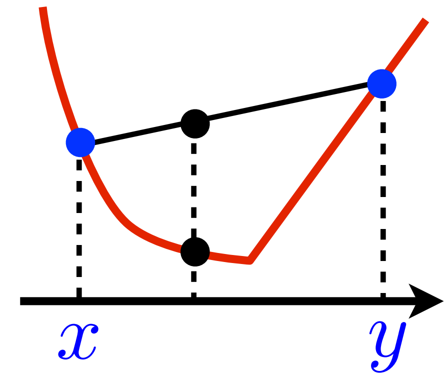


Proper: $\{x \in \mathcal{H} \setminus G(x) \neq +\infty\} \neq \emptyset$

Convex Optimization

Setting: $G : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$

$\mathcal{H}$: Hilbert space. Here: $\mathcal{H} = \mathbb{R}^N$.

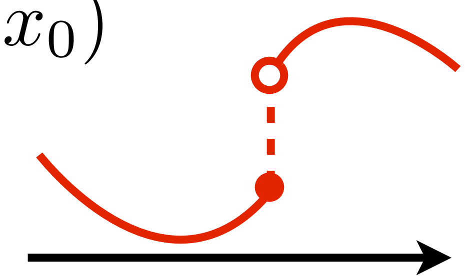


$$\textit{Problem:} \quad \min_{x \in \mathcal{H}} G(x)$$

Class of functions:

Convex: $G(tx + (1-t)y) \leqslant tG(x) + (1-t)G(y) \quad t \in [0,1]$

Lower semi-continuous: $\liminf_{x \rightarrow x_0} G(x) \geqslant G(x_0)$



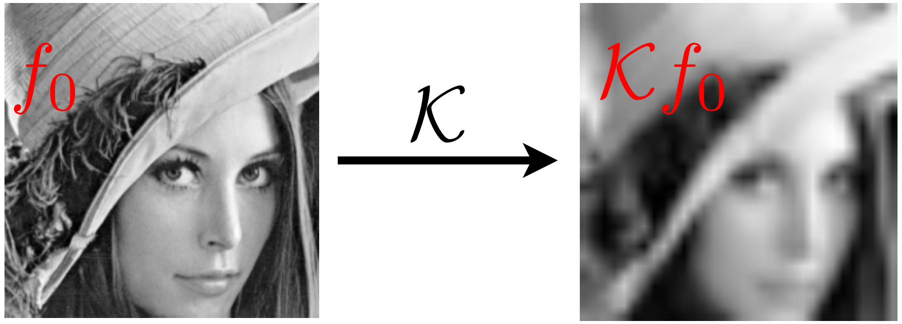
Proper: $\{x \in \mathcal{H} \setminus G(x) \neq +\infty\} \neq \emptyset$

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

($\mathcal{C}$ closed and convex)

Example: ℓ^1 Regularization

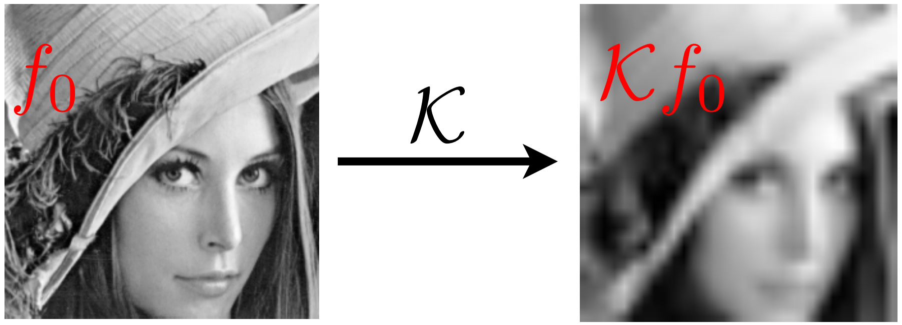
Inverse problem: measurements $y = \mathcal{K} f_0 + w$



$$\mathcal{K} : \mathbb{R}^N \rightarrow \mathbb{R}^P, \quad P \leqslant N$$

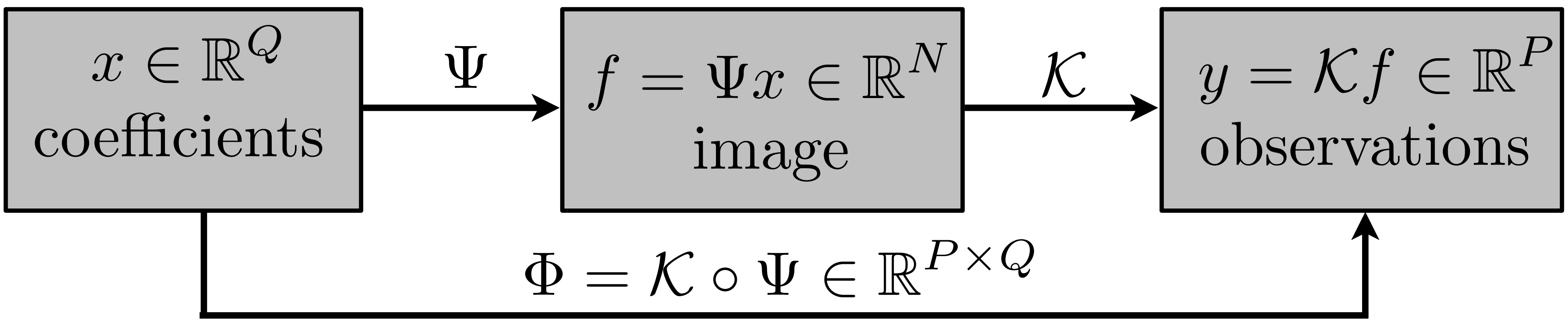
Example: ℓ^1 Regularization

Inverse problem: measurements $y = \mathcal{K} f_0 + w$



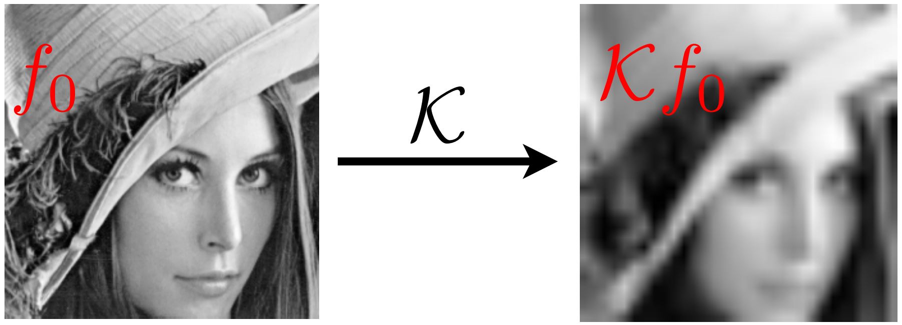
$\mathcal{K} : \mathbb{R}^N \to \mathbb{R}^P, \quad P \leqslant N$

Model: $f_0 = \Psi x_0$ sparse in dictionary $\Psi \in \mathbb{R}^{N \times Q}, Q \geqslant N$.



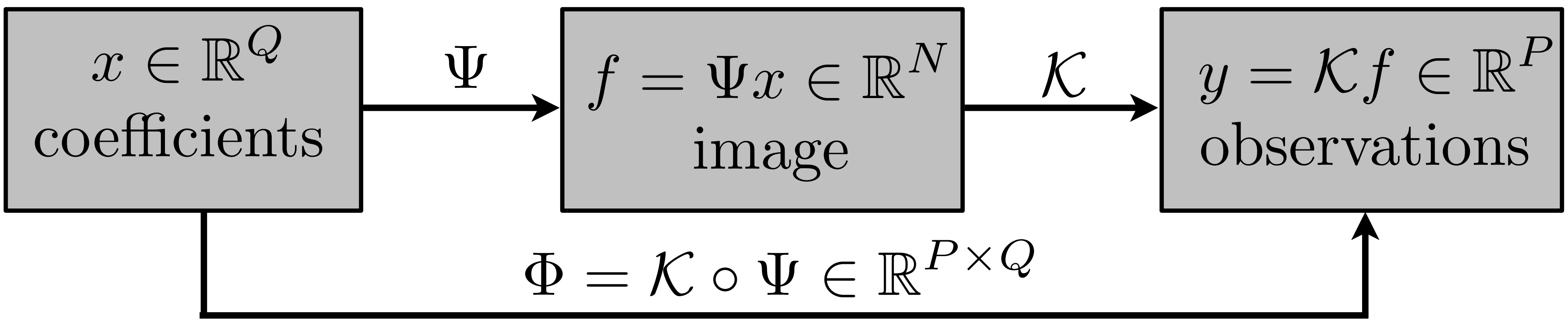
Example: ℓ^1 Regularization

Inverse problem: measurements $y = \mathcal{K} f_0 + w$



$\mathcal{K} : \mathbb{R}^N \to \mathbb{R}^P, \quad P \leqslant N$

Model: $f_0 = \Psi x_0$ sparse in dictionary $\Psi \in \mathbb{R}^{N \times Q}, Q \geqslant N$.



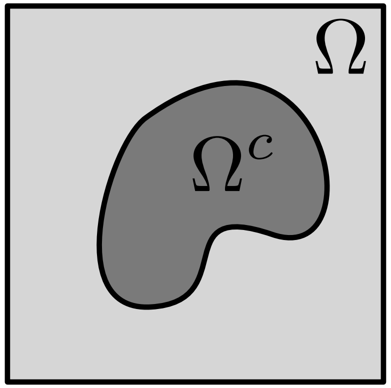
Sparse recovery: $f^\star = \Psi x^\star$ where $x^\star$ solves

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|^2 + \lambda \|x\|_1$$

Fidelity   Regularization

Example: ℓ^1 Regularization

Inpainting: masking operator $\mathcal{K}$

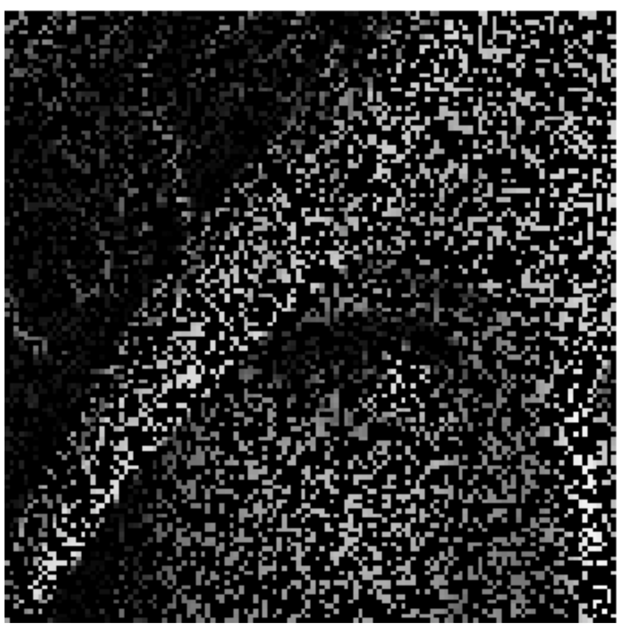


$$(\mathcal{K}f)_i = \begin{cases} f_i & \text{if} \quad i \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

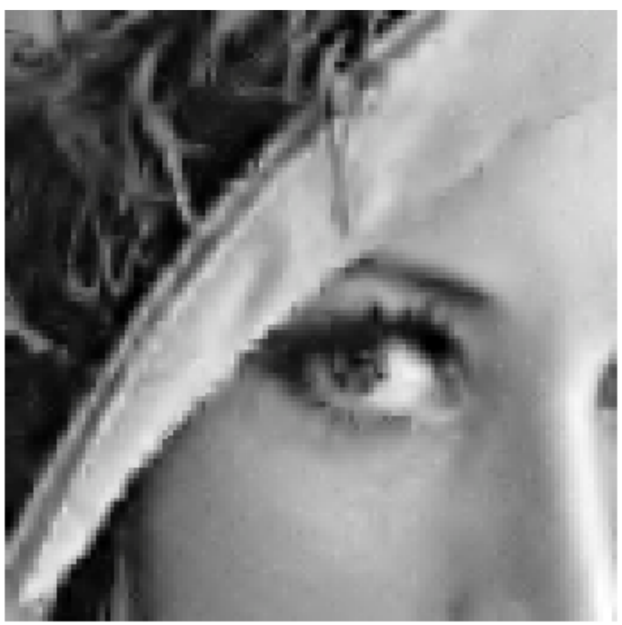
$$\mathcal{K} : \mathbb{R}^N \rightarrow \mathbb{R}^P \qquad P = |\Omega|$$

$\Psi \in \mathbb{R}^{N \times Q}$ translation invariant wavelet frame.

Orignal $f_0 = \Psi x_0$



$y = \Phi x_0 + w$



Recovery $\Psi x^\star$

Overview

- Smooth optimization

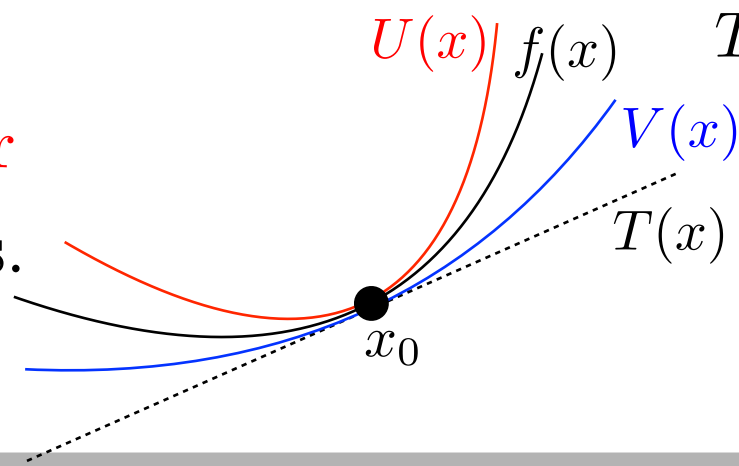
Hypotheses: $\mu \mathrm{Id}_n \preceq \partial^2 f(x) \preceq L\, \mathrm{Id}_n$

strong convexity — smoothness

Conditionning:

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

Quadratic lower / upper approximants.



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

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

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

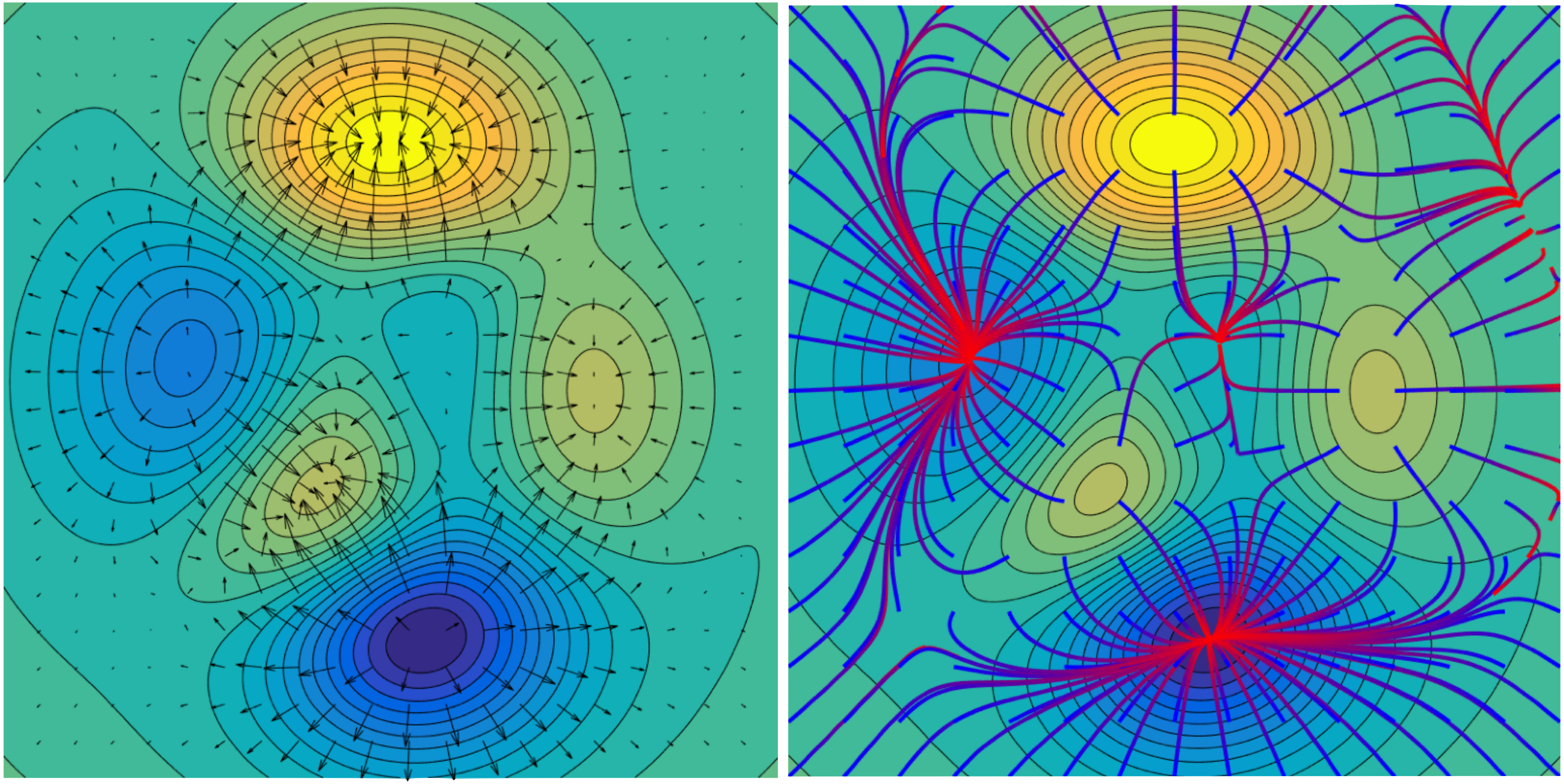
Gradient descent: $x_{k+1} = x_k - \tau_k \nabla f(x_k)$

Theorem:

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

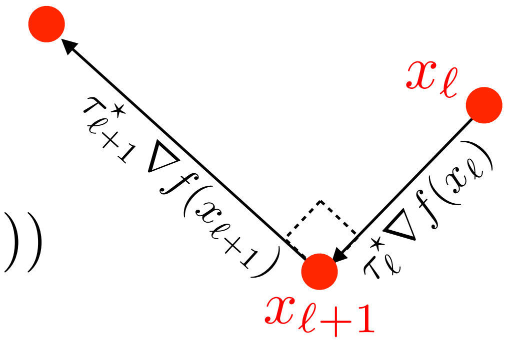
If $\mu > 0$, $L < +\infty$, $0 < \tau < \frac{2}{L}$ $\quad \|x_k - x^\star\| \leqslant \rho^\ell \|x_0 - x^\star\|$

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

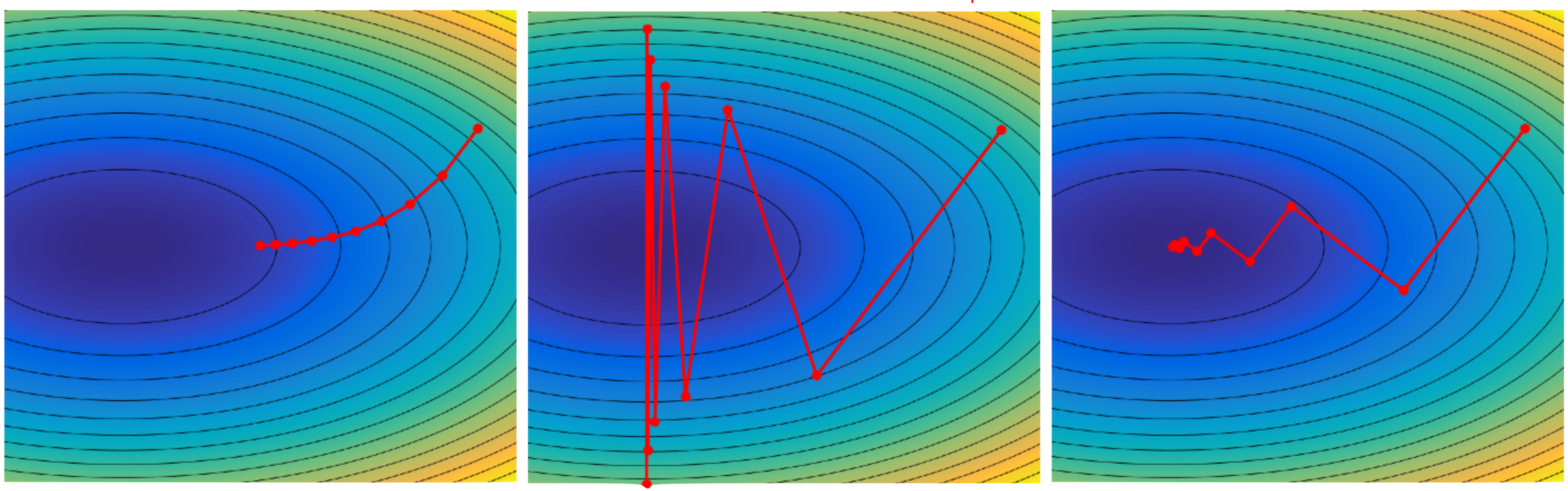


$$x_{\ell+1} = x_\ell - \tau_\ell \nabla f(x_\ell)$$

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

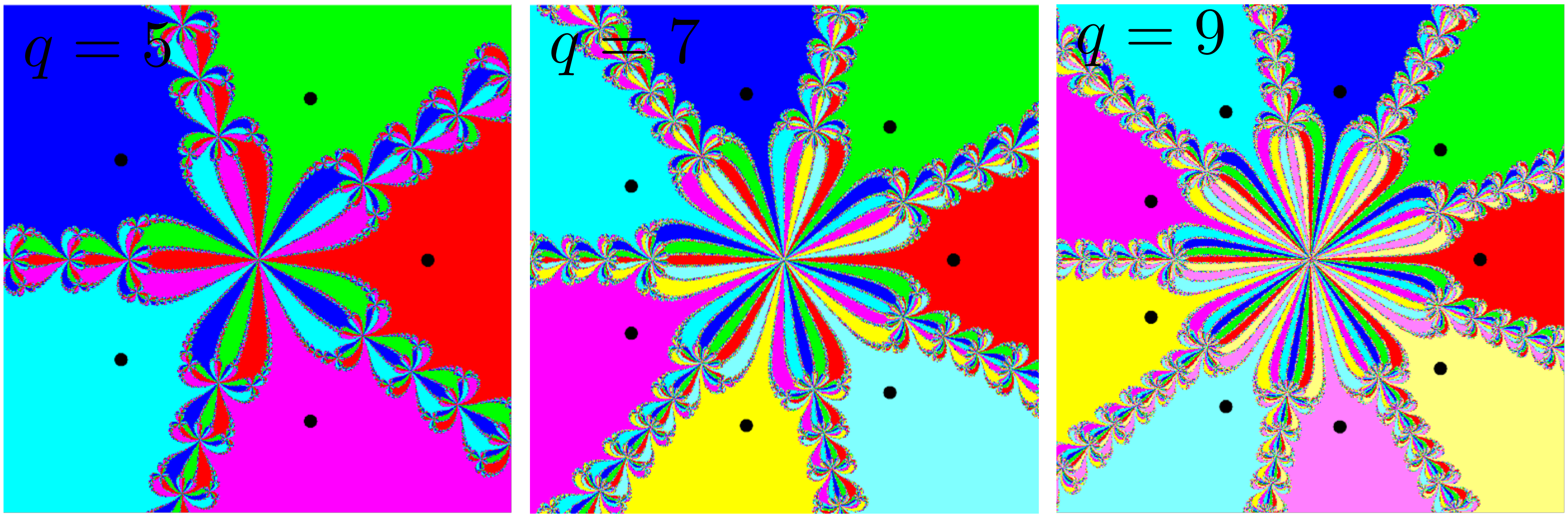


$$\nabla f(x_\ell) \perp \nabla f(x_{\ell+1})$$



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

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



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

Overview

- Motivation for Non-smooth Optimization

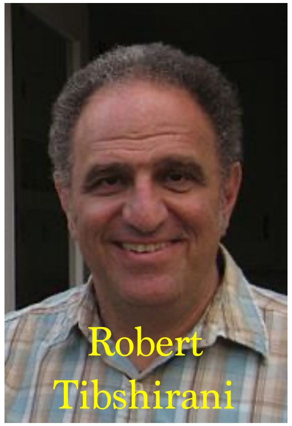
ℓ^p "norms" $\quad \|x\|_p^p \overset{\text{def.}}{=} \sum_i |x_i|^p$

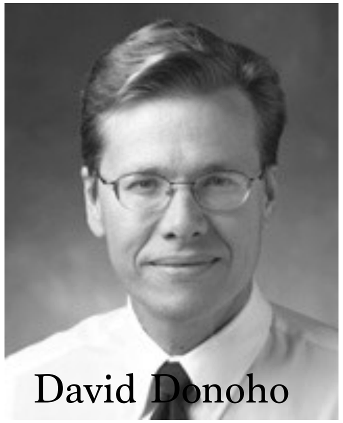
Lasso / Basis-Pursuit:

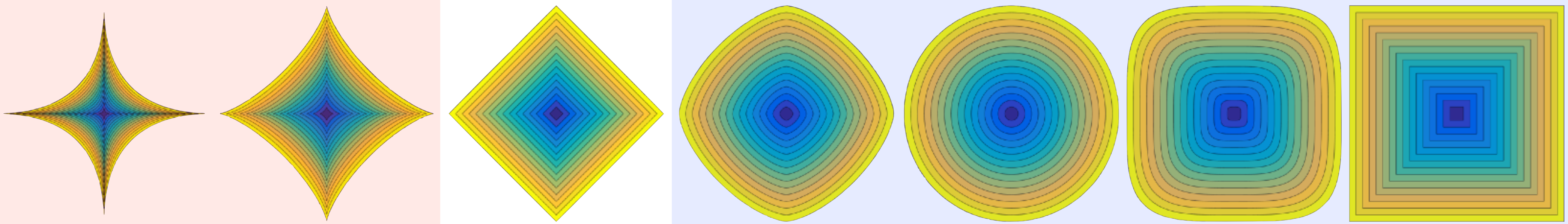
$$\min_x \|x\|_1 + \frac{1}{2\lambda}\|Ax - y\|^2$$

$$\min_{Ax=y} \|x\|_1$$

$\lambda \to 0$





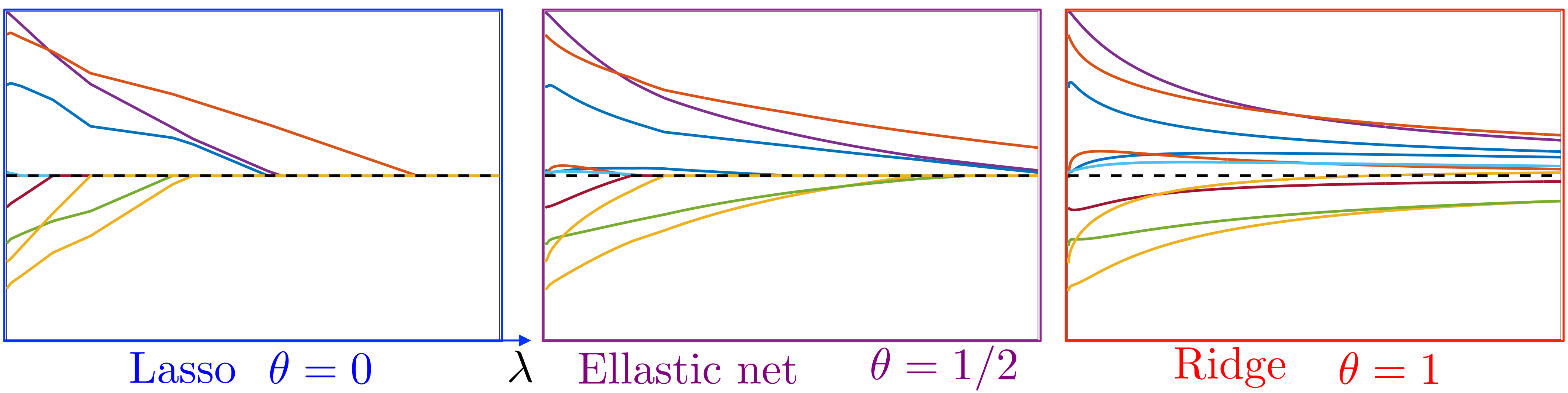


$p = 0.5$ $\quad p = 0.75$ $\quad p = 1$ $\quad p = 1.5$ $\quad p = 2$ $\quad p = 4$ $\quad p = \infty$

Non-convex $\qquad$ Non-sparse

Ellastic net: $x_\lambda \in \underset{x}{\operatorname{argmin}} \frac{1}{2\lambda}\|Ax - y\|^2 + (1-\theta)\|x\|_1 + \frac{\theta}{2}\|x\|_2^2$

Regularization path: $\lambda \longmapsto x_\lambda$



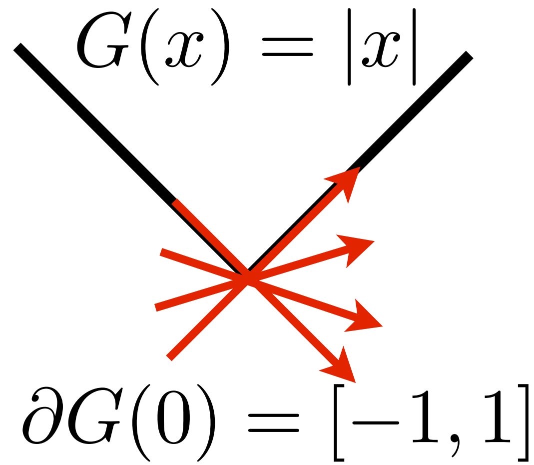
Overview

- **Subdifferential Calculus**
- Proximal Calculus
- Forward Backward
- Douglas Rachford
- Generalized Forward-Backward
- Duality

Sub-differential

Sub-differential:

$$\partial G(x) = \{u \in \mathcal{H} \setminus \forall\, z,\ G(z) \geqslant G(x) + \langle u,\ z - x\rangle\}$$



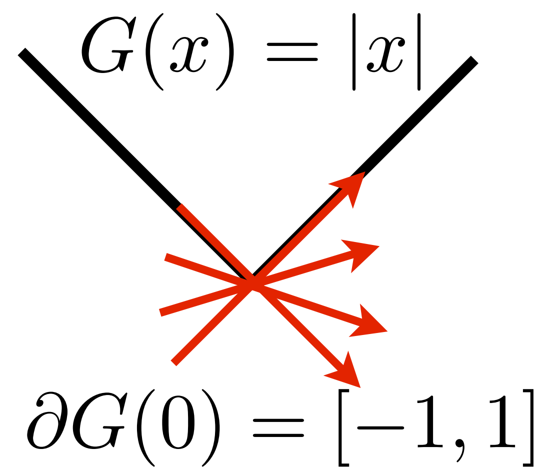
Sub-differential

Sub-differential:

$$\partial G(x) = \{u \in \mathcal{H} \setminus \forall\, z,\; G(z) \geqslant G(x) + \langle u,\; z - x\rangle\}$$

Smooth functions:

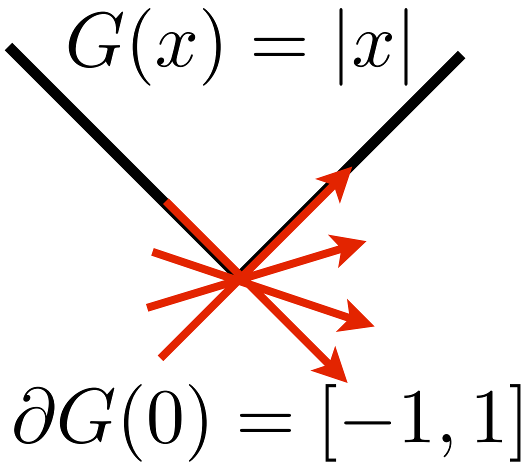
If F is C^1, $\partial F(x) = \{\nabla F(x)\}$



Sub-differential

Sub-differential:

$$\partial G(x) = \{u \in \mathcal{H} \setminus \forall\, z,\ G(z) \geqslant G(x) + \langle u,\ z - x\rangle\}$$



Smooth functions:

$$\text{If } F \text{ is } C^1,\ \partial F(x) = \{\nabla F(x)\}$$

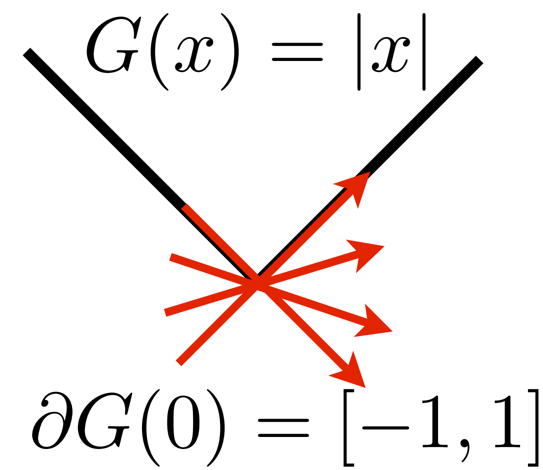
First-order conditions:

$$x^\star \in \underset{x \in \mathcal{H}}{\text{argmin}}\ G(x) \iff 0 \in \partial G(x^\star)$$

Sub-differential

Sub-differential:

$$\partial G(x) = \{u \in \mathcal{H} \setminus \forall\, z,\ G(z) \geqslant G(x) + \langle u,\ z - x\rangle\}$$

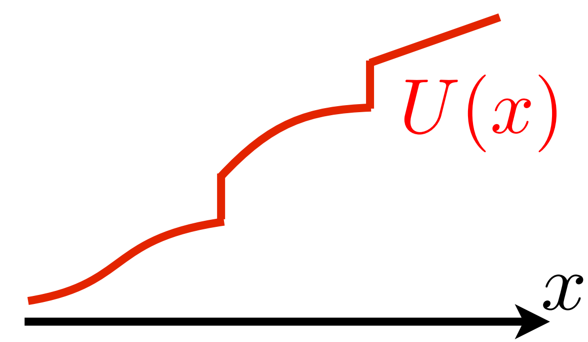


Smooth functions:

$$\text{If } F \text{ is } C^1,\ \partial F(x) = \{\nabla F(x)\}$$

First-order conditions:

$$x^\star \in \underset{x \in \mathcal{H}}{\text{argmin}}\ G(x) \iff 0 \in \partial G(x^\star)$$



Monotone operator: $\quad U(x) = \partial G(x)$

$$\forall\,(u, v) \in U(x) \times U(y),\quad \langle y - x,\ v - u\rangle \geqslant 0$$

Example: ℓ^1 Regularization

$$x^\star \in \underset{x \in \mathbb{R}^Q}{\text{argmin}}\ G(x) = \frac{1}{2}\|y - \Phi x\|^2 + \lambda \|x\|_1$$

$$\partial G(x) = \Phi^*(\Phi x - y) + \lambda \partial \|\cdot\|_1(x)$$

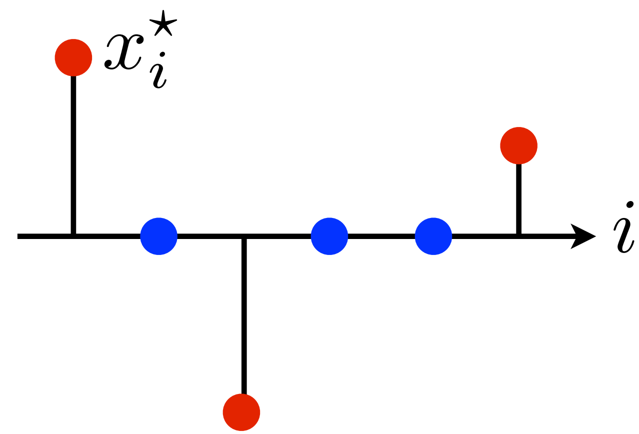
$$\partial \|\cdot\|_1(x)_i = \begin{cases} \text{sign}(x_i) & \text{if} \quad x_i \neq 0, \\ [-1, 1] & \text{if} \quad x_i = 0. \end{cases}$$

Example: ℓ^1 Regularization

$$x^\star \in \underset{x \in \mathbb{R}^Q}{\text{argmin}}\; G(x) = \frac{1}{2}\|y - \Phi x\|^2 + \lambda \|x\|_1$$

$$\partial G(x) = \Phi^*(\Phi x - y) + \lambda \partial \|\cdot\|_1(x)$$

$$\partial \|\cdot\|_1(x)_i = \begin{cases} \text{sign}(x_i) & \text{if} \quad x_i \neq 0, \\ [-1, 1] & \text{if} \quad x_i = 0. \end{cases}$$



Support of the solution:

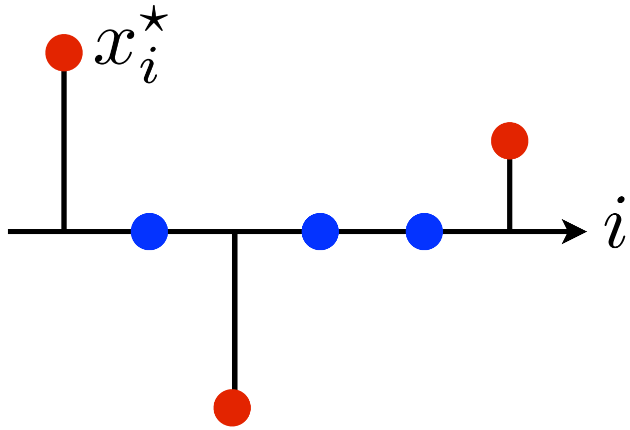
$$I = \{ i \in \{0, \ldots, N-1\} \setminus x_i^\star \neq 0 \}$$

Example: ℓ^1 Regularization

$$x^\star \in \underset{x \in \mathbb{R}^Q}{\text{argmin}}\ G(x) = \frac{1}{2}\|y - \Phi x\|^2 + \lambda \|x\|_1$$

$$\partial G(x) = \Phi^*(\Phi x - y) + \lambda \partial \| \cdot \|_1(x)$$

$$\partial \| \cdot \|_1(x)_i = \begin{cases} \text{sign}(x_i) & \text{if} \quad x_i \neq 0, \\ [-1, 1] & \text{if} \quad x_i = 0. \end{cases}$$



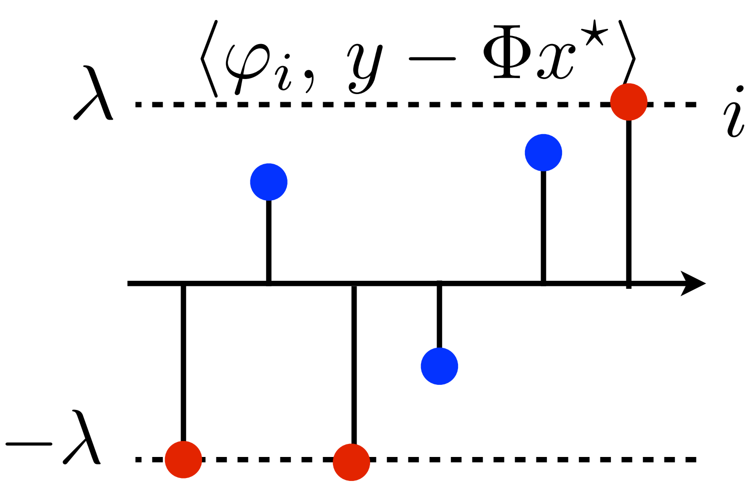
Support of the solution: ●

$$I = \{ i \in \{0, \ldots, N-1\} \setminus x_i^\star \neq 0 \}$$

First-order conditions:

$$\exists s \in \mathbb{R}^N, \quad \Phi^*(\Phi x^\star - y) + \lambda s = 0$$

$$\begin{cases} s_I = \text{sign}(x_I), \\ \|s_{I^c}\|_\infty \leqslant 1. \end{cases}$$



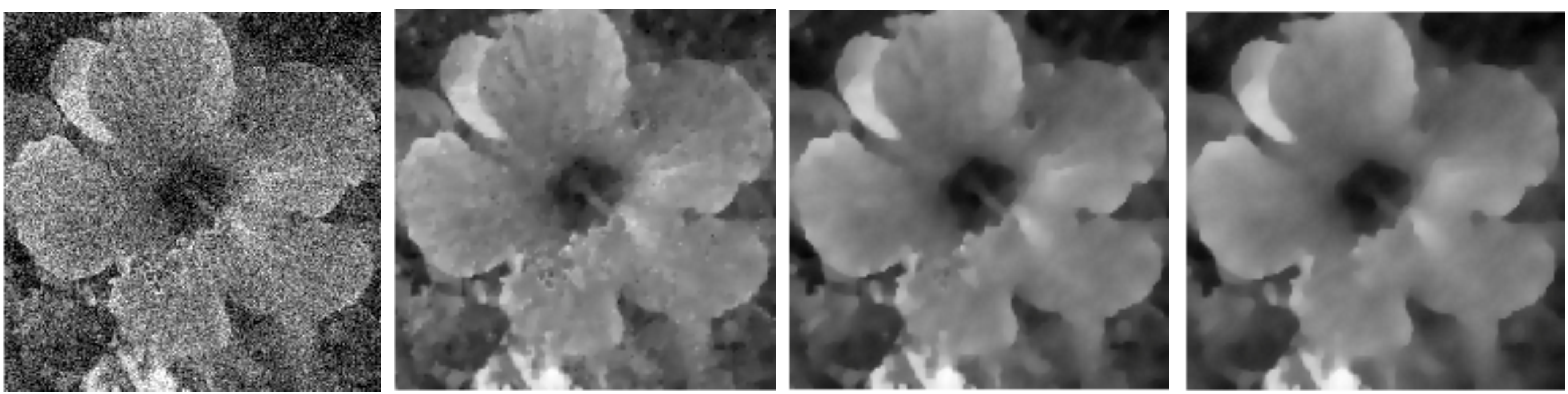
Example: Total Variation Denoising

Important: the optimization variable is f.

$$f^\star \in \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - f\|^2 + \lambda J(f)$$

Finite difference gradient: $\nabla : \mathbb{R}^N \to \mathbb{R}^{N \times 2}$ $\quad (\nabla f)_i \in \mathbb{R}^2$

Discrete TV norm: $J(f) = \sum_i \|(\nabla f)_i\|$



$\lambda = 0$ (noisy) $\longrightarrow$ λ

Example: Total Variation Denoising

$$f^\star \in \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - f\|^2 + \lambda J(f)$$

$$J(f) = G(\nabla f) \qquad G(u) = \sum_i \|u_i\|$$

Composition by linear maps: $\partial(J \circ A) = A^* \circ (\partial J) \circ A$

$$\partial J(f) = -\mathrm{div}\left(\partial G(\nabla f)\right)$$

$$\partial G(u)_i = \begin{cases} \frac{u_i}{\|u_i\|} & \text{if} \quad u_i \neq 0, \\ \left\{\eta \in \mathbb{R}^2 \setminus \|\eta\| \leqslant 1\right\} & \text{if} \quad u_i = 0. \end{cases}$$

Example: Total Variation Denoising

$$f^\star \in \underset{f \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|y - f\|^2 + \lambda J(f)$$

$$J(f) = G(\nabla f) \qquad G(u) = \sum_i \|u_i\|$$

Composition by linear maps: $\partial(J \circ A) = A^* \circ (\partial J) \circ A$

$$\partial J(f) = -\mathrm{div}\left(\partial G(\nabla f)\right)$$

$$\partial G(u)_i = \begin{cases} \frac{u_i}{\|u_i\|} & \text{if} \quad u_i \neq 0, \\ \{\eta \in \mathbb{R}^2 \setminus \|\eta\| \leqslant 1\} & \text{if} \quad u_i = 0. \end{cases}$$

First-order conditions: $\exists\, v \in \mathbb{R}^{N \times 2},\ f^\star = y + \lambda \mathrm{div}(v)$

$$\begin{cases} \forall\, i \in I,\ v_i = \frac{\nabla f_i^\star}{\|\nabla f_i^\star\|}, \\ \forall\, i \in I^c,\ \|v_i\| \leqslant 1 \end{cases} \qquad I = \{i \setminus (\nabla f^\star)_i \neq 0\}$$

Overview

- Subdifferential Calculus
- **Proximal Calculus**
- Forward Backward
- Douglas Rachford
- Generalized Forward-Backward
- Duality

Proximal Operators

Proximal operator of G:

$$\mathrm{Prox}_{\gamma G}(x) = \underset{z}{\mathrm{argmin}}\ \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

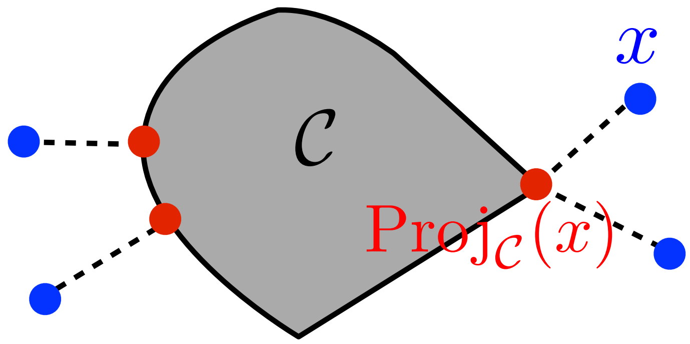
Proximal Operators

Proximal operator of G:

$$\text{Prox}_{\gamma G}(x) = \underset{z}{\text{argmin}} \ \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

Indicators: $G(x) = \iota_{\mathcal{C}}(x)$

$$\text{Prox}_{\gamma G}(x) = \text{Proj}_{\mathcal{C}}(x) = \underset{z \in \mathcal{C}}{\text{argmin}} \ \|x - z\|$$



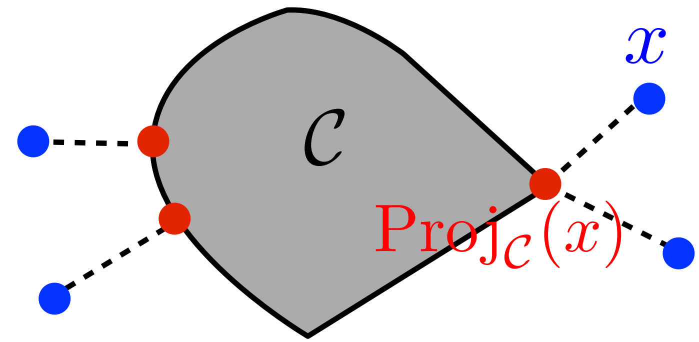
Proximal Operators

Proximal operator of G:

$$\mathrm{Prox}_{\gamma G}(x) = \underset{z}{\mathrm{argmin}}\ \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

Indicators: $G(x) = \iota_{\mathcal{C}}(x)$

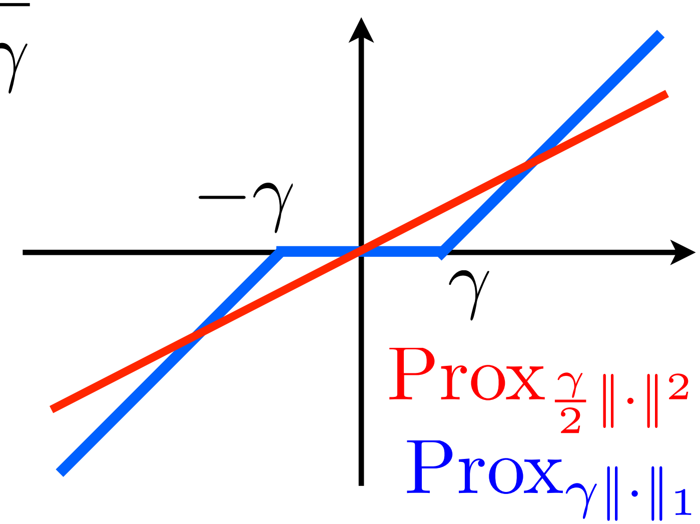
$$\mathrm{Prox}_{\gamma G}(x) = \mathrm{Proj}_{\mathcal{C}}(x) = \underset{z \in \mathcal{C}}{\mathrm{argmin}}\ \|x - z\|$$



ℓ^2 norm squared: $\mathrm{Prox}_{\frac{\gamma}{2}\|\cdot\|^2}(x) = \dfrac{x}{1+\gamma}$

ℓ^1 norm: $G(x) = \|x\|_1 = \sum_i |x_i|$

$$\mathrm{Prox}_{\gamma G}(x)_i = \max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i$$



Proximal Calculus

Separability: $G(x) = G_1(x_1) + \ldots + G_n(x_n)$

$$\mathrm{Prox}_G(x) = (\mathrm{Prox}_{G_1}(x_1), \ldots, \mathrm{Prox}_{G_n}(x_n))$$

Proximal Calculus

Separability: $G(x) = G_1(x_1) + \ldots + G_n(x_n)$

$$\mathrm{Prox}_G(x) = (\mathrm{Prox}_{G_1}(x_1), \ldots, \mathrm{Prox}_{G_n}(x_n))$$

Quadratic functionals: $G(x) = \frac{1}{2}\|\Phi x - y\|^2$

$$\mathrm{Prox}_{\gamma G} = (\mathrm{Id} + \gamma \Phi^* \Phi)^{-1} \Phi^*$$

$$= \Phi^* (\mathrm{Id} + \gamma \Phi \Phi^*)^{-1}$$

Proximal Calculus

Separability: $G(x) = G_1(x_1) + \ldots + G_n(x_n)$

$$\text{Prox}_G(x) = (\text{Prox}_{G_1}(x_1), \ldots, \text{Prox}_{G_n}(x_n))$$

Quadratic functionals: $G(x) = \frac{1}{2}\|\Phi x - y\|^2$

$$\text{Prox}_{\gamma G} = (\text{Id} + \gamma \Phi^* \Phi)^{-1} \Phi^*$$

$$= \Phi^* (\text{Id} + \gamma \Phi \Phi^*)^{-1}$$

Composition by tight frame: $A \circ A^* = \text{Id}$

$$\text{Prox}_{G \circ A}(x) = A^* \circ \text{Prox}_G \circ A + \text{Id} - A^* \circ A$$

Ortho-basis A: $\text{Prox}_{G \circ A} = A^* \circ \text{Prox}_G \circ A$

Non-convex Proximal Operators

Proximal operator of G:

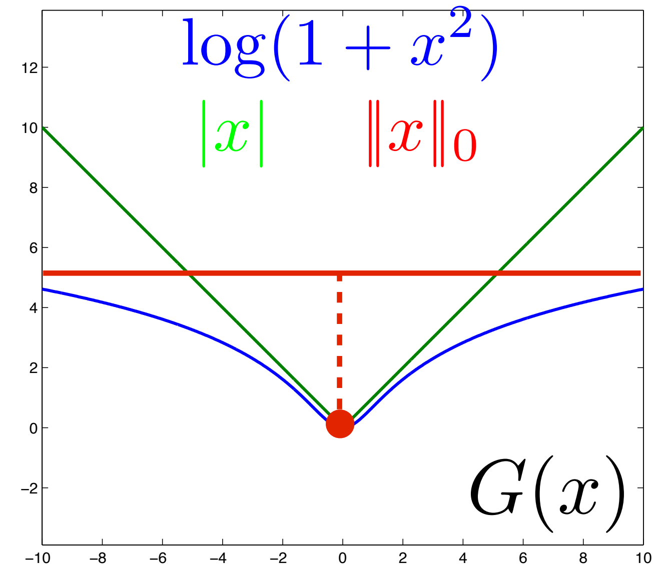
$$\mathrm{Prox}_{\gamma G}(x) = \underset{z}{\mathrm{argmin}} \ \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

Non-convex Proximal Operators

Proximal operator of G:

$$\text{Prox}_{\gamma G}(x) = \underset{z}{\text{argmin}} \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

$$G(x) = \|x\|_1 = \sum_i |x_i|$$



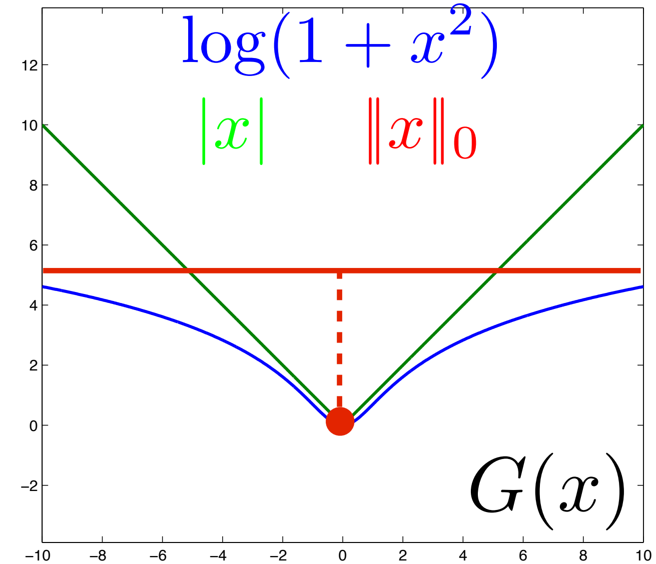
$$G(x) = \|x\|_0 = |\,\{i \setminus x_i \neq 0\}\,|$$

$$G(x) = \sum_i \log(1 + |x_i|^2)$$

Non-convex Proximal Operators

Proximal operator of G:

$$\mathrm{Prox}_{\gamma G}(x) = \underset{z}{\mathrm{argmin}}\ \frac{1}{2}\|x - z\|^2 + \gamma G(z)$$

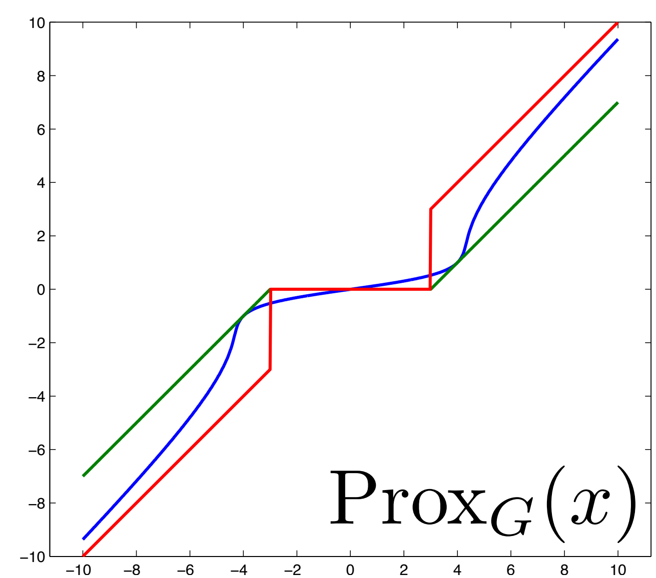


$$G(x) = \|x\|_1 = \sum_i |x_i|$$

$$\mathrm{Prox}_{\gamma G}(x)_i = \max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i$$

$$G(x) = \|x\|_0 = |\,\{i \setminus x_i \neq 0\}\,|$$

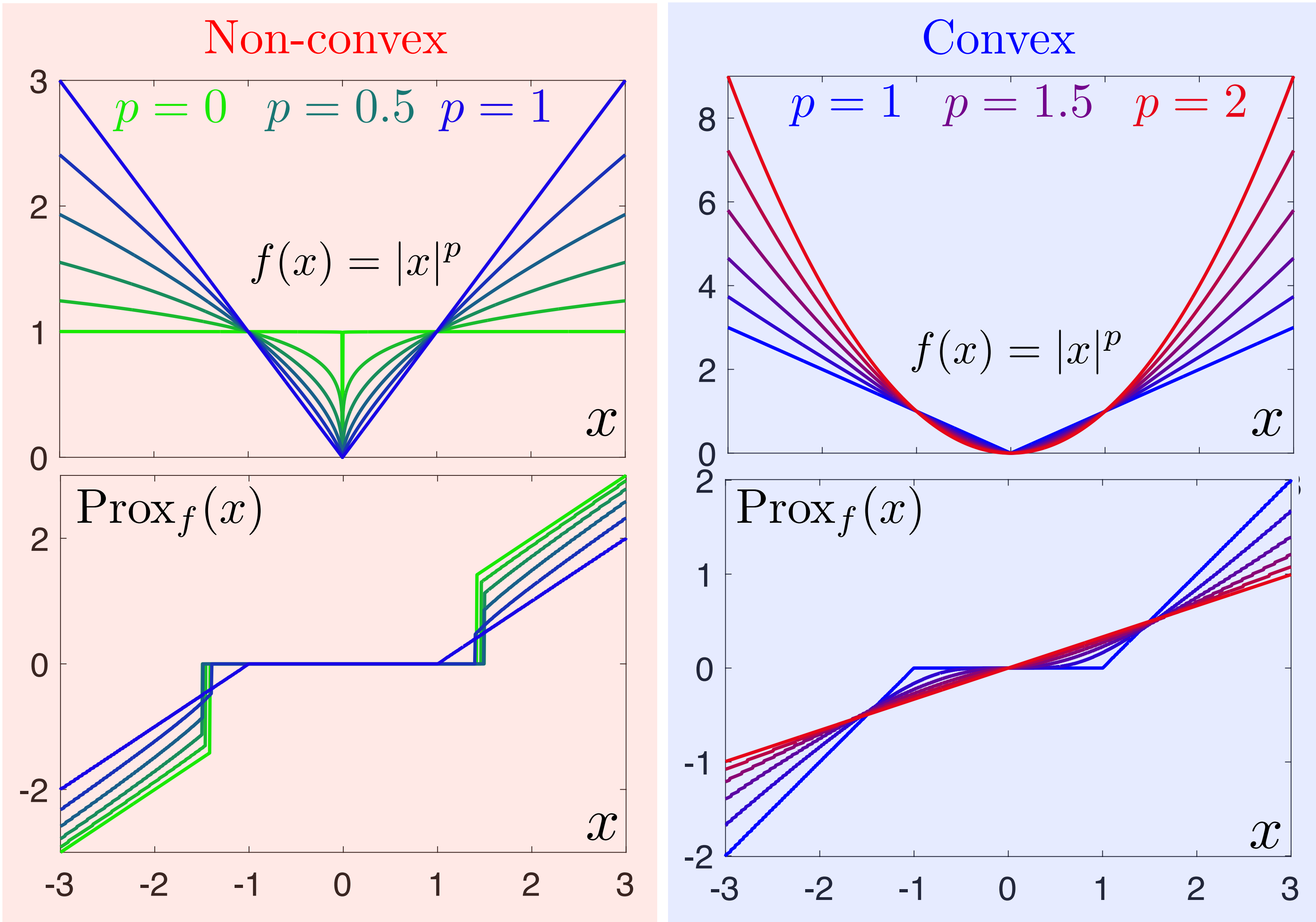
$$\mathrm{Prox}_{\gamma G}(x)_i = \begin{cases} x_i & \text{if} \quad |x_i| \geqslant \sqrt{2\gamma}, \\ 0 & \text{otherwise.} \end{cases}$$



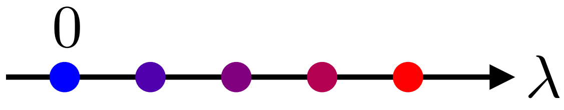
$$G(x) = \sum_i \log(1 + |x_i|^2)$$

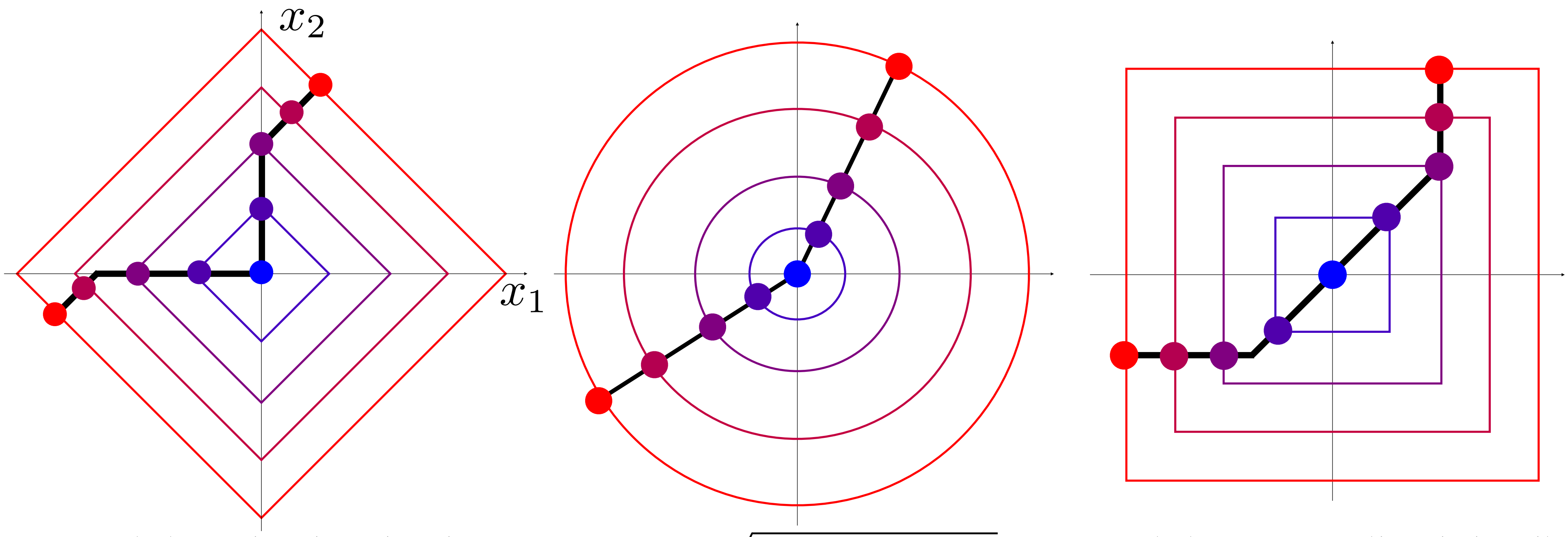
$\longrightarrow$ 3rd order polynomial root.

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



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





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

Prox and Subdifferential

Resolvant of ∂G:

$$z = \text{Prox}_{\gamma G}(x) \iff 0 \in z - x + \gamma \partial G(z)$$

$$\iff x \in (\text{Id} + \gamma \partial G)(z) \iff z = (\text{Id} + \gamma \partial G)^{-1}(x)$$

Inverse of a set-valued mapping:

$$\text{where } x \in U(y) \iff y \in U^{-1}(x)$$

$$\text{Prox}_{\gamma G} = (\text{Id} + \gamma \partial G)^{-1} \quad \text{is a single-valued mapping}$$

Prox and Subdifferential

Resolvant of ∂G:

$$z = \text{Prox}_{\gamma G}(x) \iff 0 \in z - x + \gamma \partial G(z)$$
$$\iff x \in (\text{Id} + \gamma \partial G)(z) \iff z = (\text{Id} + \gamma \partial G)^{-1}(x)$$

Inverse of a set-valued mapping:

$$\text{where } x \in U(y) \iff y \in U^{-1}(x)$$

$$\text{Prox}_{\gamma G} = (\text{Id} + \gamma \partial G)^{-1} \quad \text{is a single-valued mapping}$$

Fix point: $x^\star \in \underset{x}{\text{argmin}}\, G(x)$

$$\iff 0 \in \partial G(x^\star) \iff x^\star \in (\text{Id} + \gamma \partial G)(x^\star)$$
$$\iff x^\star = (\text{Id} + \gamma \partial G)^{-1}(x^\star) = \text{Prox}_{\gamma G}(x^\star)$$

Gradient and Proximal Descents

Gradient descent: $x^{(\ell+1)} = x^{(\ell)} - \gamma_\ell \nabla G(x^{(\ell)})$ **[explicit]**

G is C^1 and ∇G is L-Lipschitz

Theorem: If $0 < \gamma_\ell < 2/L$, $x^{(\ell)} \to x^\star$ a solution.

Gradient and Proximal Descents

Gradient descent: $x^{(\ell+1)} = x^{(\ell)} - \gamma_\ell \nabla G(x^{(\ell)})$ **[explicit]**

G is C^1 and ∇G is L-Lipschitz

Theorem: If $0 < \gamma_\ell < 2/L$, $x^{(\ell)} \rightarrow x^\star$ a solution.

Sub-gradient descent: $x^{(\ell+1)} = x^{(\ell)} - \gamma_\ell v^{(\ell)}, \quad v^{(\ell)} \in \partial G(x^{(\ell)})$

Theorem: If $\gamma_\ell \sim 1/\ell$, $x^{(\ell)} \rightarrow x^\star$ a solution.

$\longrightarrow$ Problem: slow.

Gradient and Proximal Descents

Gradient descent: $x^{(\ell+1)} = x^{(\ell)} - \gamma_\ell \nabla G(x^{(\ell)})$ **[explicit]**

G is C^1 and ∇G is L-Lipschitz

Theorem: If $0 < \gamma_\ell < 2/L$, $x^{(\ell)} \to x^\star$ a solution.

Sub-gradient descent: $x^{(\ell+1)} = x^{(\ell)} - \gamma_\ell v^{(\ell)}, \quad v^{(\ell)} \in \partial G(x^{(\ell)})$

Theorem: If $\gamma_\ell \sim 1/\ell$, $x^{(\ell)} \to x^\star$ a solution.

$\longrightarrow$ Problem: slow.

Proximal-point algorithm: $x^{(\ell+1)} = \mathrm{Prox}_{\gamma_\ell G}(x^{(\ell)})$ **[implicit]**

Theorem: If $\gamma_\ell \geqslant c > 0$, $x^{(\ell)} \to x^\star$ a solution.

$\longrightarrow$ $\mathrm{Prox}_{\gamma G}$ hard to compute.

Overview

- Subdifferential Calculus
- Proximal Calculus
- Forward Backward
- Douglas Rachford
- Generalized Forward-Backward
- Duality

Proximal Splitting Methods

Solve $\min_{x \in \mathcal{H}} E(x)$

Problem: $\text{Prox}_{\gamma E}$ is not available.

Proximal Splitting Methods

Solve $\min_{x \in \mathcal{H}} E(x)$

Problem: $\text{Prox}_{\gamma E}$ is not available.

Splitting: $E(x) = F(x) + \sum_i G_i(x)$

$F(x)$: Smooth

$G_i(x)$: Simple

Proximal Splitting Methods

Solve $\min_{x \in \mathcal{H}} E(x)$

Problem: $\text{Prox}_{\gamma E}$ is not available.

Splitting: $E(x) = F(x) + \sum_i G_i(x)$

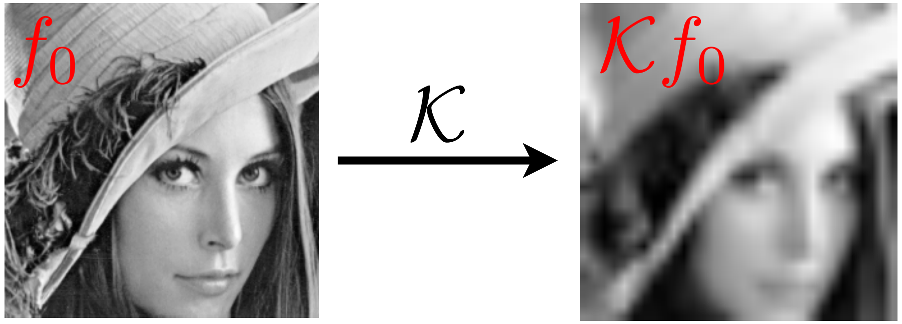
Smooth ($F(x)$) — Simple ($G_i(x)$)

Iterative algorithms using: $\nabla F(x)$ and $\text{Prox}_{\gamma G_i}(x)$

	solves	
Forward-Backward:	$\longrightarrow$	$F + G$
Douglas-Rachford:	$\longrightarrow$	$\sum G_i$
Primal-Dual:	$\longrightarrow$	$\sum G_i \circ A$
Generalized FB:	$\longrightarrow$	$F + \sum G_i$

Smooth + Simple Splitting

Inverse problem: measurements $y = \mathcal{K} f_0 + w$



$\mathcal{K} : \mathbb{R}^N \to \mathbb{R}^P, \quad P \leqslant N$

Model: $f_0 = \Psi x_0$ sparse in dictionary Ψ.

Sparse recovery: $f^\star = \Psi x^\star$ where $x^\star$ solves

$$\min_{x \in \mathbb{R}^N} \underbrace{F(x)}_{\text{Smooth}} + \underbrace{G(x)}_{\text{Simple}}$$

Data fidelity: $F(x) = \frac{1}{2} \|y - \Phi x\|^2 \qquad \Phi = \mathcal{K} \circ \Psi$

Regularization: $G(x) = \|x\|_1 = \sum_i |x_i|$

Forward-Backward

Fix point equation:

$$x^\star \in \underset{x}{\operatorname{argmin}}\, F(x) + G(x) \iff 0 \in \nabla F(x^\star) + \partial G(x^\star)$$

$$\iff (x^\star - \gamma \nabla F(x^\star)) \in x^\star + \gamma \partial G(x^\star)$$

$$\iff x^\star = \operatorname{Prox}_{\gamma G}(x^\star - \gamma \nabla F(x^\star))$$

Forward-Backward

Fix point equation:

$$x^\star \in \underset{x}{\operatorname{argmin}}\ F(x) + G(x) \iff 0 \in \nabla F(x^\star) + \partial G(x^\star)$$

$$\iff (x^\star - \gamma \nabla F(x^\star)) \in x^\star + \gamma \partial G(x^\star)$$

$$\iff x^\star = \operatorname{Prox}_{\gamma G}(x^\star - \gamma \nabla F(x^\star))$$

Forward-backward:

$$x^{(\ell+1)} = \operatorname{Prox}_{\gamma G}\left(x^{(\ell)} - \gamma \nabla F(x^{(\ell)}) \right)$$

Forward-Backward

Fix point equation:

$$x^\star \in \underset{x}{\text{argmin}}\; F(x) + G(x) \iff 0 \in \nabla F(x^\star) + \partial G(x^\star)$$

$$\iff (x^\star - \gamma \nabla F(x^\star)) \in x^\star + \gamma \partial G(x^\star)$$

$$\iff x^\star = \text{Prox}_{\gamma G}(x^\star - \gamma \nabla F(x^\star))$$

Forward-backward:

$$x^{(\ell+1)} = \text{Prox}_{\gamma G}\left(x^{(\ell)} - \gamma \nabla F(x^{(\ell)})\right)$$

Projected gradient descent: $G = \iota_{\mathcal{C}}$

Forward-Backward

Fix point equation:

$$x^\star \in \underset{x}{\text{argmin}}\; F(x) + G(x) \iff 0 \in \nabla F(x^\star) + \partial G(x^\star)$$

$$\iff (x^\star - \gamma \nabla F(x^\star)) \in x^\star + \gamma \partial G(x^\star)$$

$$\iff x^\star = \text{Prox}_{\gamma G}(x^\star - \gamma \nabla F(x^\star))$$

Forward-backward:

$$x^{(\ell+1)} = \text{Prox}_{\gamma G}\left(x^{(\ell)} - \gamma \nabla F(x^{(\ell)})\right)$$

Projected gradient descent: $G = \iota_{\mathcal{C}}$

Theorem: Let ∇F be L-Lipschitz.

If $\gamma < 2/L$, $x^{(\ell)} \to x^\star$ a solution of $(\star)$

Example: L1 Regularization

$$\min_x \frac{1}{2}\|\Phi x - y\|^2 + \lambda\|x\|_1 \quad \Longleftrightarrow \quad \min_x F(x) + G(x)$$

$$F(x) = \frac{1}{2}\|\Phi x - y\|^2$$

$$\nabla F(x) = \Phi^*(\Phi x - y) \qquad L = \|\Phi^*\Phi\|$$

$$G(x) = \lambda\|x\|_1$$

$$\mathrm{Prox}_{\gamma G}(x)_i = \max\left(0, 1 - \frac{\gamma\lambda}{|x_i|}\right) x_i$$

Forward-backward $\Longleftrightarrow$ Iterative soft thresholding

Convergence Speed

$$\min_x E(x) = F(x) + G(x)$$

∇F is L-Lipschitz.

G is simple.

Theorem: If $L > 0$, FB iterates $x^{(\ell)}$ satisfies

$$E(x^{(\ell)}) - E(x^\star) = O(1/\ell)$$

C degrades with $L \to 0$.

Multi-steps Accelerations

Beck-Teboule accelerated FB: $t^{(0)} = 1$

$$x^{(\ell+1)} = \mathrm{Prox}_{1/L}\left(y^{(\ell)} - \frac{1}{L}\nabla F(y^{(\ell)})\right)$$

$$t^{(\ell+1)} = \frac{1+\sqrt{1+4(t^{(\ell)})^2}}{2}$$

$$y^{(\ell+1)} = x^{(\ell+1)} + \frac{t^{(\ell)}-1}{t^{(\ell+1)}}(x^{(\ell+1)} - x^{(\ell)})$$

(see also Nesterov method)

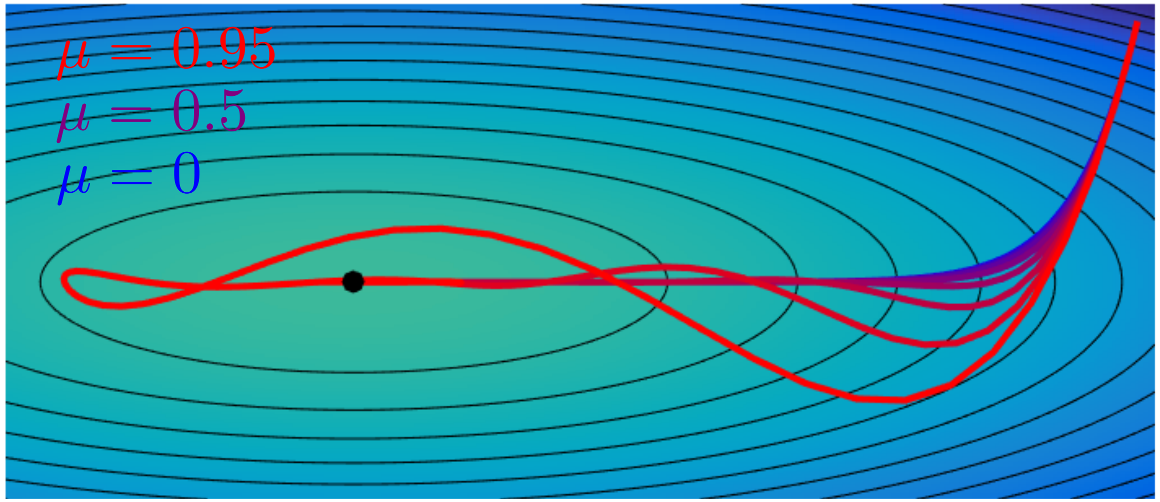
Theorem: If $L > 0$, $E(x^{(\ell)}) - E(x^\star) = O(1/\ell^2)$

Complexity theory: optimal in a worse-case sense.

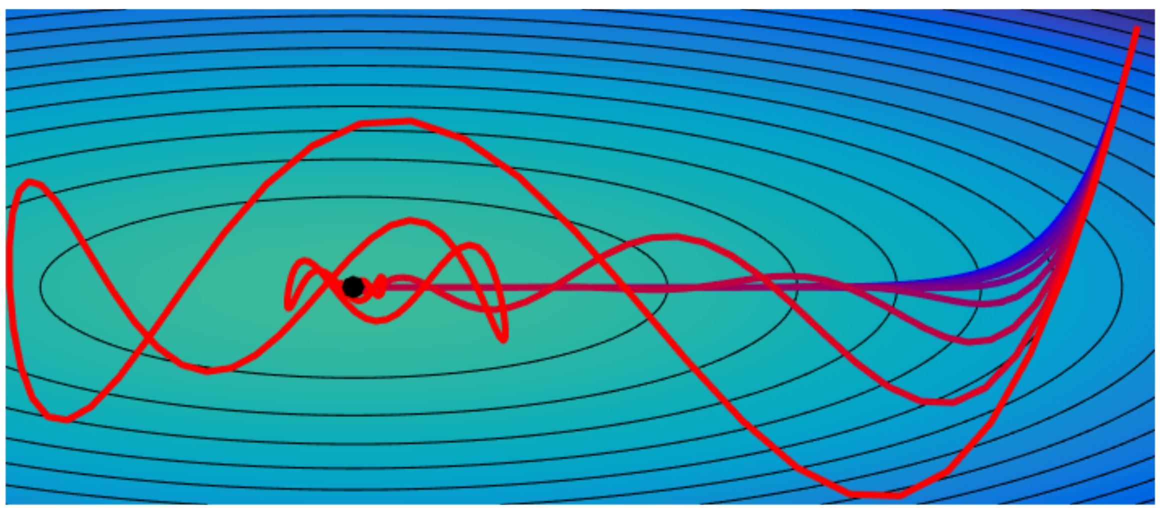
$$x_{k+1} = x_k + p_k$$

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

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

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

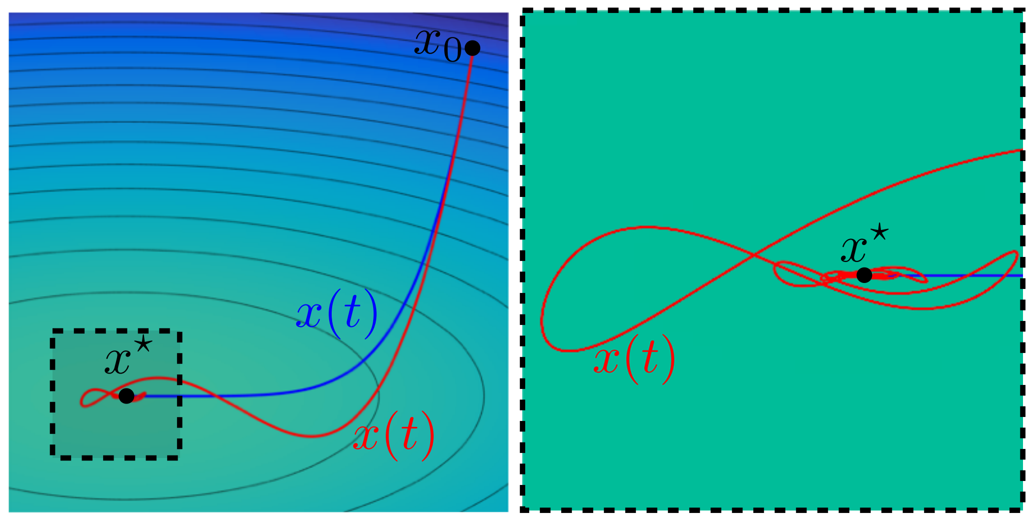
$\tau \to 0 \quad k\sqrt{\tau} \to t$

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

Theorem:

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

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



Yurii Nesterov

Overview

- Subdifferential Calculus
- Proximal Calculus
- Forward Backward
- **Douglas Rachford**
- Generalized Forward-Backward
- Duality

Douglas Rachford Scheme

$$\min_x \boxed{G_1(x)} + \boxed{G_2(x)} \qquad (\star)$$

Simple Simple

Douglas-Rachford iterations:

$$z^{(\ell+1)} = \left(1 - \frac{\alpha}{2}\right) z^{(\ell)} + \frac{\alpha}{2} \mathrm{RProx}_{\gamma G_2} \circ \mathrm{RProx}_{\gamma G_1}(z^{(\ell)})$$

$$x^{(\ell+1)} = \mathrm{Prox}_{\gamma G_1}(z^{(\ell+1)})$$

Reflexive prox:

$$\mathrm{RProx}_{\gamma G}(x) = 2\mathrm{Prox}_{\gamma G}(x) - x$$

Douglas Rachford Scheme

$$\min_x \boxed{G_1(x)} + \boxed{G_2(x)} \qquad (\star)$$

Simple Simple

Douglas-Rachford iterations:

$$z^{(\ell+1)} = \left(1 - \frac{\alpha}{2}\right) z^{(\ell)} + \frac{\alpha}{2} \text{RProx}_{\gamma G_2} \circ \text{RProx}_{\gamma G_1}(z^{(\ell)})$$

$$x^{(\ell+1)} = \text{Prox}_{\gamma G_1}(z^{(\ell+1)})$$

Reflexive prox:

$$\text{RProx}_{\gamma G}(x) = 2\text{Prox}_{\gamma G}(x) - x$$

Theorem: If $0 < \alpha < 2$ and $\gamma > 0$,

$x^{(\ell)} \rightarrow x^\star$ a solution of $(\star)$

DR Fix Point Equation

$$\min_x G_1(x) + G_2(x) \quad \Longleftrightarrow \quad 0 \in \partial(G_1 + G_2)(x)$$

$$\Longleftrightarrow \quad \exists z,\ z - x \in \partial(\gamma G_1)(x) \quad \text{and} \quad x - z \in \partial(\gamma G_2)(x)$$

$$\Longleftrightarrow \quad x = \mathrm{Prox}_{\gamma G_1}(z) \quad \text{and} \quad \boxed{(2x - z) - x \in \partial(\gamma G_2)(x)}$$

DR Fix Point Equation

$$\min_x G_1(x) + G_2(x) \quad \Longleftrightarrow \quad 0 \in \partial(G_1 + G_2)(x)$$

$$\Longleftrightarrow \quad \exists z,\; z - x \in \partial(\gamma G_1)(x) \quad \text{and} \quad x - z \in \partial(\gamma G_2)(x)$$

$$\Longleftrightarrow \quad x = \operatorname{Prox}_{\gamma G_1}(z) \quad \text{and} \quad \boxed{(2x - z) - x \in \partial(\gamma G_2)(x)}$$

$$\Longleftrightarrow \quad x = \operatorname{Prox}_{\gamma G_2}(2x - z) = \operatorname{Prox}_{\gamma G_2} \circ \operatorname{RProx}_{\gamma G_1}(z)$$

$$\Longleftrightarrow \quad z = 2\operatorname{Prox}_{\gamma G_2} \circ \operatorname{RProx}_{\gamma G_1}(y) - (2x - z)$$

$$\Longleftrightarrow \quad z = 2\operatorname{Prox}_{\gamma G_2} \circ \operatorname{RProx}_{\gamma G_1}(z) - \operatorname{RProx}_{\gamma G_1}(z)$$

$$\Longleftrightarrow \quad z = \operatorname{RProx}_{\gamma G_2} \circ \operatorname{RProx}_{\gamma G_1}(z)$$

$$\Longleftrightarrow \quad z = \left(1 - \frac{\alpha}{2}\right) z + \frac{\alpha}{2} \operatorname{RProx}_{\gamma G_2} \circ \operatorname{RProx}_{\gamma G_1}(z)$$

Iterative Projections

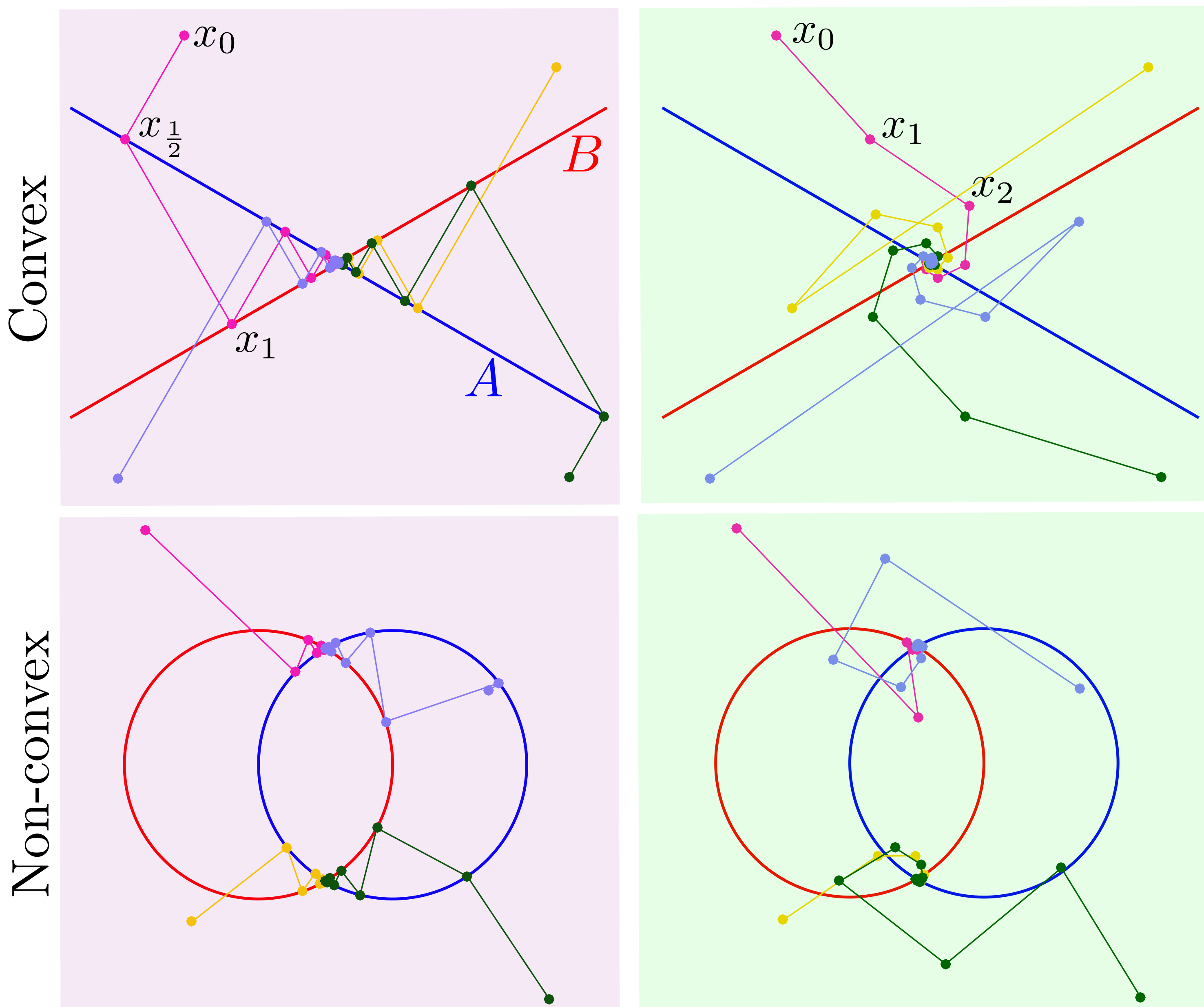
$$x_{k+1} = P_B(P_A x_k)$$

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

Douglas-Rachford

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

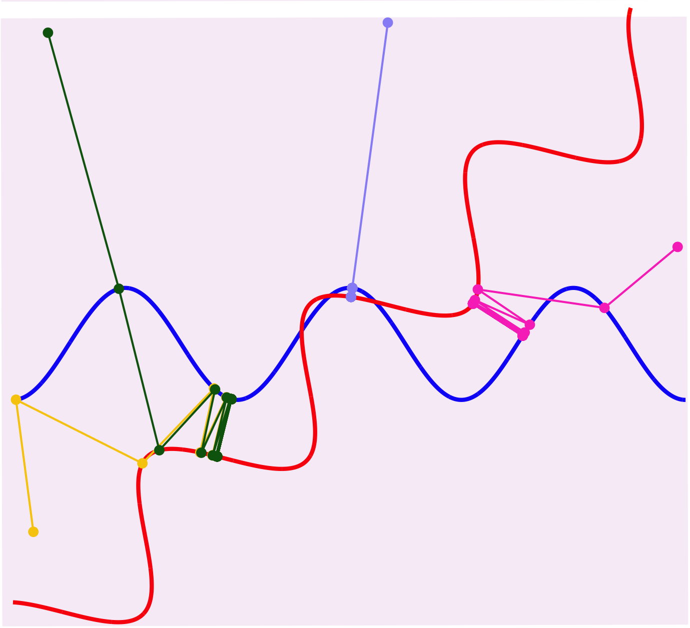
$$y_{k+1} = \frac{1}{2} y_k + \frac{1}{2} \bar{P}_B(x_k)$$



Iterative Projections

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

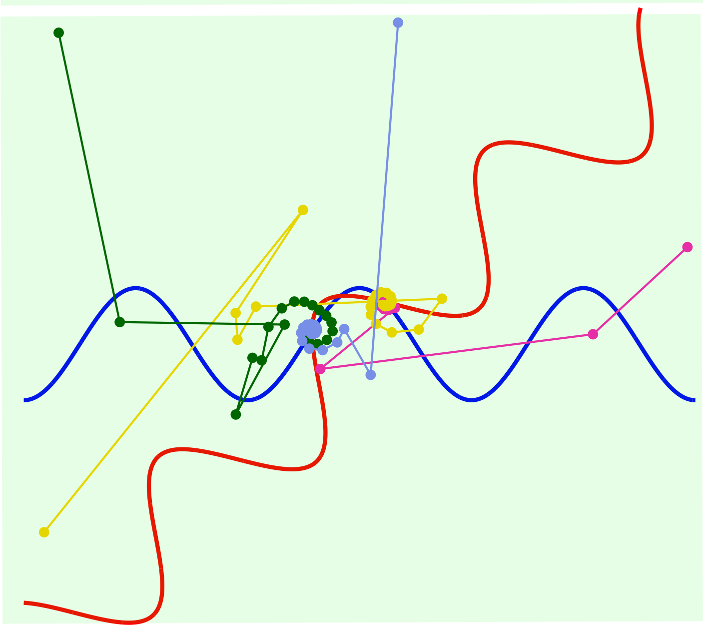
$$P_A \overset{\text{def.}}{=} \text{Proj}_A$$

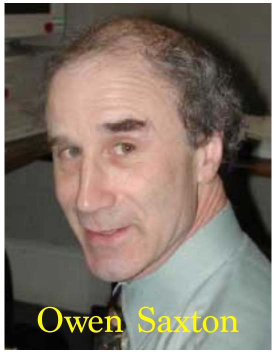


Douglas-Rachford

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

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





Example: Constrainted L1

$$\min_{\Phi x = y} \|x\|_1 \quad \Longleftrightarrow \quad \min_x G_1(x) + G_2(x)$$

$$G_1(x) = i_{\mathcal{C}}(x), \qquad \mathcal{C} = \{x \setminus \Phi x = y\}$$

$$\text{Prox}_{\gamma G_1}(x) = \text{Proj}_{\mathcal{C}}(x) = x + \Phi^*(\Phi\Phi^*)^{-1}(y - \Phi x)$$

$$G_2(x) = \|x\|_1 \qquad \text{Prox}_{\gamma G_2}(x) = \left(\max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i \right)_i$$

$\longrightarrow$ efficient if $\Phi\Phi^*$ easy to invert.

Example: Constrainted L1

$$\min_{\Phi x = y} \|x\|_1 \quad \Longleftrightarrow \quad \min_x G_1(x) + G_2(x)$$

$$G_1(x) = i_{\mathcal{C}}(x), \qquad \mathcal{C} = \{x \setminus \Phi x = y\}$$

$$\text{Prox}_{\gamma G_1}(x) = \text{Proj}_{\mathcal{C}}(x) = x + \Phi^*(\Phi\Phi^*)^{-1}(y - \Phi x)$$

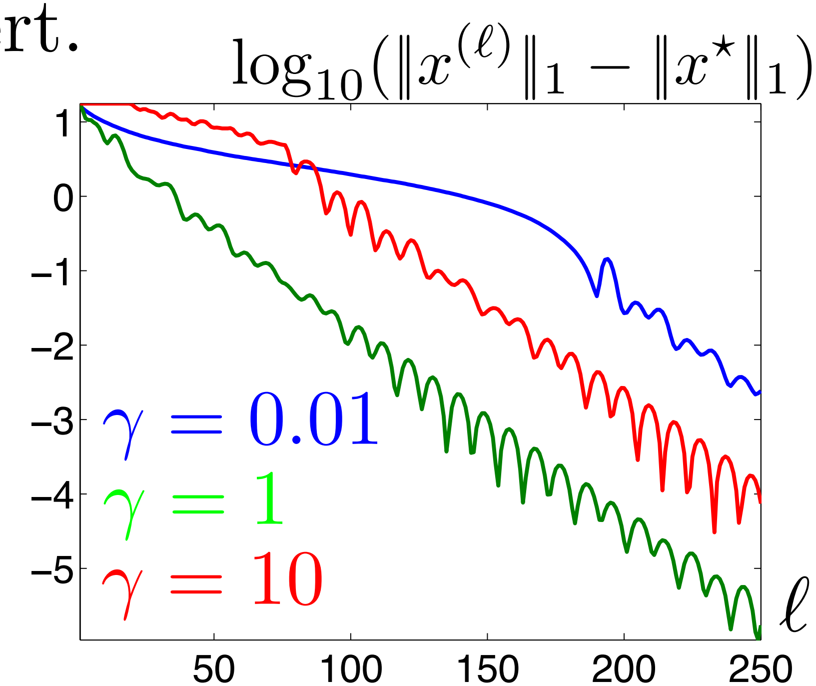
$$G_2(x) = \|x\|_1 \qquad \text{Prox}_{\gamma G_2}(x) = \left(\max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i \right)_i$$

$\longrightarrow$ efficient if $\Phi\Phi^*$ easy to invert.

Example: compressed sensing

$\Phi \in \mathbb{R}^{100 \times 400}$ Gaussian matrix

$y = \Phi x_0$ $\qquad \|x_0\|_0 = 17$



More than 2 Functionals

$$\min_{x} G_1(x) + \ldots + G_k(x)$$ each F_i is simple

$$\iff \min_{(x_1,\ldots,x_k)} G(x_1,\ldots,x_k) + \iota_{\mathcal{C}}(x_1,\ldots,x_k)$$

$$G(x_1,\ldots,x_k) = G_1(x_1) + \ldots + G_k(x_k)$$

$$\mathcal{C} = \left\{(x_1,\ldots,x_k) \in \mathcal{H}^k \setminus x_1 = \ldots = x_k\right\}$$

More than 2 Functionals

$$\min_x G_1(x) + \ldots + G_k(x) \qquad \text{each } F_i \text{ is simple}$$

$$\iff \quad \min_{(x_1,\ldots,x_k)} G(x_1,\ldots,x_k) + \iota_{\mathcal{C}}(x_1,\ldots,x_k)$$

$$G(x_1,\ldots,x_k) = G_1(x_1) + \ldots + G_k(x_k)$$

$$\mathcal{C} = \left\{ (x_1,\ldots,x_k) \in \mathcal{H}^k \setminus x_1 = \ldots = x_k \right\}$$

G and $\iota_{\mathcal{C}}$ are simple:

$$\mathrm{Prox}_{\gamma G}(x_1,\ldots,x_k) = (\mathrm{Prox}_{\gamma G_i}(x_i))_i$$

$$\mathrm{Prox}_{\gamma \iota_{\mathcal{C}}}(x_1,\ldots,x_k) = (\tilde{x},\ldots,\tilde{x}) \qquad \text{where} \quad \tilde{x} = \frac{1}{k}\sum_i x_i$$

Auxiliary Variables: DR

$$\min_x G_1(x) + G_2 \circ A(x)$$ Linear map $A : \mathcal{E} \to \mathcal{H}$.

$$\iff \min_{z \in \mathcal{H} \times \mathcal{E}} G(z) + \iota_{\mathcal{C}}(z)$$ G_1, G_2 simple.

$$G(x, y) = G_1(x) + G_2(y)$$

$$\mathcal{C} = \{(x, y) \in \mathcal{H} \times \mathcal{E} \setminus Ax = y\}$$

Auxiliary Variables: DR

$$\min_x G_1(x) + G_2 \circ A(x)$$ Linear map $A : \mathcal{E} \to \mathcal{H}$.

$$\iff \min_{z \in \mathcal{H} \times \mathcal{E}} G(z) + \iota_{\mathcal{C}}(z)$$ G_1, G_2 simple.

$$G(x, y) = G_1(x) + G_2(y)$$

$$\mathcal{C} = \{(x, y) \in \mathcal{H} \times \mathcal{E} \setminus Ax = y\}$$

$$\text{Prox}_{\gamma G}(x, y) = (\text{Prox}_{\gamma G_1}(x), \text{Prox}_{\gamma G_2}(y))$$

$$\text{Prox}_{\iota_{\mathcal{C}}}(x, y) = (x + A^* \tilde{y}, y - \tilde{y}) \;= (\tilde{x}, A\tilde{x})$$

where
$$\left| \begin{array}{l} \tilde{y} = (\text{Id} + AA^*)^{-1}(Ax - y) \\ \tilde{x} = (\text{Id} + A^*A)^{-1}(A^*y + x) \end{array} \right.$$

$\longrightarrow$ efficient if $\text{Id} + AA^*$ or $\text{Id} + A^*A$ easy to invert.

Example: TV Regularization

$$\min_f \frac{1}{2}\|\mathcal{K}f - y\|^2 + \lambda\|\nabla f\|_1 \qquad \|u\|_1 = \sum_i \|u_i\|$$

$$\iff \quad \min_x G_1(f) + G_2 \circ \nabla(f)$$

$$G_1(u) = \|u\|_1 \qquad \mathrm{Prox}_{\gamma G_1}(u)_i = \max\left(0, 1 - \frac{\gamma}{\|u_i\|}\right) u_i$$

$$G_2(f) = \frac{1}{2}\|\mathcal{K}f - y\|^2 \qquad \mathrm{Prox}_{\gamma G_2} = (\mathrm{Id} + \gamma\mathcal{K}^*\mathcal{K})^{-1}\mathcal{K}^*$$

$$\mathcal{C} = \left\{(f, u) \in \mathbb{R}^N \times \mathbb{R}^{N\times 2} \setminus u = \nabla f\right\}$$

$$\mathrm{Prox}_{\iota_{\mathcal{C}}}(f, u) = (\tilde{f}, \nabla\tilde{f})$$

Example: TV Regularization

$$\min_f \frac{1}{2}\|\mathcal{K}f - y\|^2 + \lambda\|\nabla f\|_1 \qquad \|u\|_1 = \sum_i \|u_i\|$$

$$\Longleftrightarrow \quad \min_x G_1(f) + G_2 \circ \nabla(f)$$

$$G_1(u) = \|u\|_1 \qquad \mathrm{Prox}_{\gamma G_1}(u)_i = \max\left(0, 1 - \frac{\gamma}{\|u_i\|}\right) u_i$$

$$G_2(f) = \frac{1}{2}\|\mathcal{K}f - y\|^2 \qquad \mathrm{Prox}_{\gamma G_2} = (\mathrm{Id} + \gamma\mathcal{K}^*\mathcal{K})^{-1}\mathcal{K}^*$$

$$\mathcal{C} = \left\{(f, u) \in \mathbb{R}^N \times \mathbb{R}^{N\times 2} \setminus u = \nabla f\right\}$$

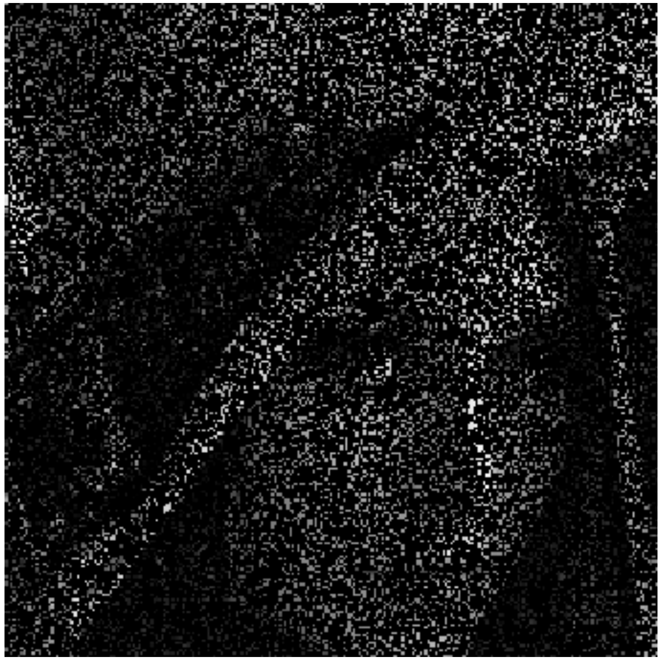
$$\mathrm{Prox}_{\iota_\mathcal{C}}(f, u) = (\tilde{f}, \nabla\tilde{f})$$

Compute the solution of: $\quad (\mathrm{Id} + \Delta)\tilde{f} = -\mathrm{div}(u) + f$

$\longrightarrow O(N\log(N))$ operations using FFT.

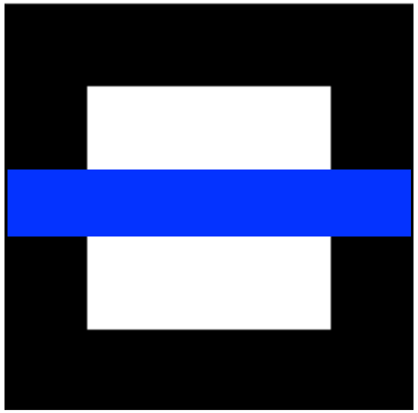
Example: TV Regularization

Orignal f_0

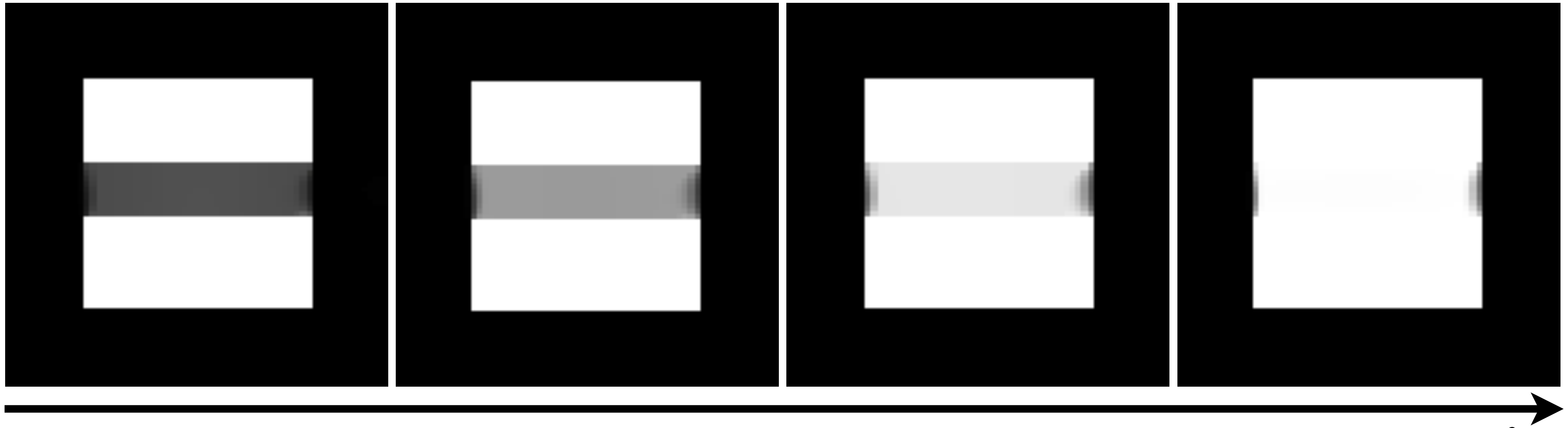


$y = \Phi f_0 + w$

Recovery $f^\star$



$y = \mathcal{K} x_0$



Iteration ℓ

Overview

- Subdifferential Calculus
- Proximal Calculus
- Forward Backward
- Douglas Rachford
- **Generalized Forward-Backward**
- Duality

GFB Splitting

$$\min_{x \in \mathbb{R}^N} F(x) + \sum_{i=1}^{n} G_i(x) \qquad (\star)$$

Smooth Simple

$\forall\, i = 1, \ldots, n,$

$$z_i^{(\ell+1)} = z_i^{(\ell)} + \operatorname{Prox}_{n\gamma G_\ell}(2x^{(\ell)} - z_i^{(\ell)} - \gamma \nabla F(x^{(\ell)})) - x^{(\ell)}$$

$$x^{(\ell+1)} = \frac{1}{n} \sum_{i=1}^{n} z_i^{(\ell+1)}$$

GFB Splitting

$$\min_{x \in \mathbb{R}^N} F(x) + \sum_{i=1}^{n} G_i(x) \qquad (\star)$$

Smooth Simple

$\forall\, i = 1, \ldots, n,$

$$z_i^{(\ell+1)} = z_i^{(\ell)} + \mathrm{Prox}_{n\gamma G_\ell}(2x^{(\ell)} - z_i^{(\ell)} - \gamma \nabla F(x^{(\ell)})) - x^{(\ell)}$$

$$x^{(\ell+1)} = \frac{1}{n} \sum_{i=1}^{n} z_i^{(\ell+1)}$$

Theorem: Let ∇F be L-Lipschitz.

If $\gamma < 2/L$, $x^{(\ell)} \rightarrow x^\star$ a solution of $(\star)$

GFB Splitting

$$\min_{x\in\mathbb{R}^N} F(x) + \sum_{i=1}^{n} G_i(x) \qquad (\star)$$

Smooth Simple

$\forall\, i = 1, \ldots, n,$

$$z_i^{(\ell+1)} = z_i^{(\ell)} + \mathrm{Prox}_{n\gamma G_\ell}(2x^{(\ell)} - z_i^{(\ell)} - \gamma \nabla F(x^{(\ell)})) - x^{(\ell)}$$

$$x^{(\ell+1)} = \frac{1}{n}\sum_{i=1}^{n} z_i^{(\ell+1)}$$

Theorem: Let ∇F be L-Lipschitz.

If $\gamma < 2/L$, $x^{(\ell)} \to x^\star$ a solution of $(\star)$

$n = 1 \longrightarrow$ Forward-backward.

$F = 0 \longrightarrow$ Douglas-Rachford.

GFB Fix Point

$$x \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \; F(x) + \sum_i G_i(x) \iff 0 \in \nabla F(x^\star) + \sum_i \partial G_i(x^\star)$$

$$\iff \quad \exists\, y_i \in \partial G_i(x^\star),\; \nabla F(x^\star) + \sum_i y_i = 0$$

GFB Fix Point

$$x \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \; F(x) + \sum_i G_i(x) \iff 0 \in \nabla F(x^\star) + \sum_i \partial G_i(x^\star)$$

$$\iff \quad \exists\, y_i \in \partial G_i(x^\star), \; \nabla F(x^\star) + \sum_i y_i = 0$$

$$\iff \quad \exists (z_i)_{i=1}^n, \; \forall\, i, \; \frac{1}{n}\big(x^\star - z_i - \gamma \nabla F(x^\star)\big) \in \gamma \partial G_i(x^\star)$$

$$x^\star = \frac{1}{n} \sum_i z_i$$

(use $z_i = x^\star - \gamma \nabla F(x^\star) - N y_i$)

GFB Fix Point

$$x \in \underset{x \in \mathbb{R}^N}{\text{argmin}}\ F(x) + \sum_i G_i(x) \iff 0 \in \nabla F(x^\star) + \sum_i \partial G_i(x^\star)$$

$$\iff \quad \exists\, y_i \in \partial G_i(x^\star),\ \nabla F(x^\star) + \sum_i y_i = 0$$

$$\iff \quad \exists (z_i)_{i=1}^n,\ \forall\, i,\ \frac{1}{n}\left(x^\star - z_i - \gamma \nabla F(x^\star)\right) \in \gamma \partial G_i(x^\star)$$

$$x^\star = \frac{1}{n} \sum_i z_i$$

(use $z_i = x^\star - \gamma \nabla F(x^\star) - N y_i$)

$$\iff \quad (2x^\star - z_i - \gamma \nabla F(x^\star)) - x^\star \in n\gamma \partial G_i(x^\star)$$

$$\iff \quad x^\star = \text{Prox}_{n\gamma G_i}(2x^\star - z_i - \gamma \nabla F(x^\star))$$

$$\iff \quad z_i = z_i + \text{Prox}_{n\gamma G_\ell}(2x^\star - z_i - \gamma \nabla F(x^\star)) - x^\star$$

GFB Fix Point

$$x \in \underset{x \in \mathbb{R}^N}{\text{argmin}}\ F(x) + \textstyle\sum_i G_i(x) \iff 0 \in \nabla F(x^\star) + \textstyle\sum_i \partial G_i(x^\star)$$

$$\iff \quad \exists\, y_i \in \partial G_i(x^\star),\ \nabla F(x^\star) + \textstyle\sum_i y_i = 0$$

$$\iff \quad \exists (z_i)_{i=1}^n,\ \forall\, i,\ \frac{1}{n}\big(x^\star - z_i - \gamma \nabla F(x^\star)\big) \in \gamma \partial G_i(x^\star)$$

$$x^\star = \tfrac{1}{n} \textstyle\sum_i z_i$$

$$(\text{use } z_i = x^\star - \gamma \nabla F(x^\star) - N y_i)$$

$$\iff \quad (2x^\star - z_i - \gamma \nabla F(x^\star)) - x^\star \in n\gamma \partial G_i(x^\star)$$

$$\iff \quad x^\star = \text{Prox}_{n\gamma G_i}(2x^\star - z_i - \gamma \nabla F(x^\star))$$

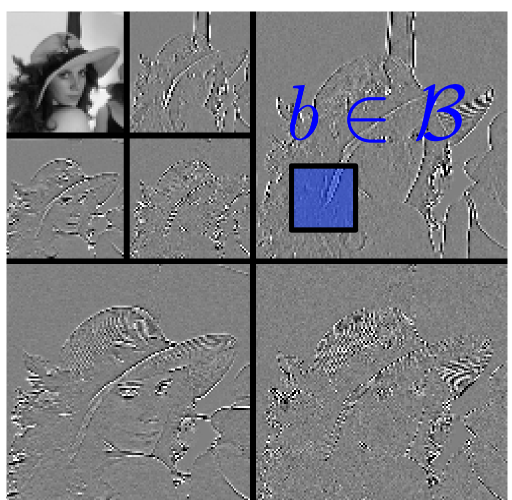
$$\iff \quad z_i = z_i + \text{Prox}_{n\gamma G_\ell}(2x^\star - z_i - \gamma \nabla F(x^\star)) - x^\star$$

[ ] + [ ] $\longrightarrow$ Fix point equation on $(x^\star, z_1, \ldots, z_n)$.

Block Regularization

$\ell^1 - \ell^2$ block sparsity: $G(x) = \sum_{b \in \mathcal{B}} \|x^{[b]}\|, \qquad \|x^{[b]}\|^2 = \sum_{m \in b} x_m^2$

Image $f = \Psi x$



Coefficients x.

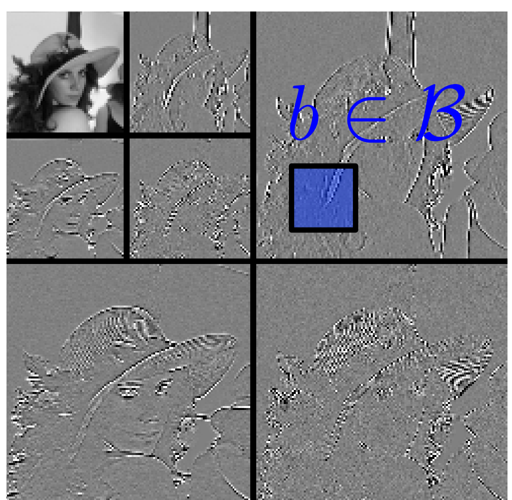
Block Regularization

$\ell^1 - \ell^2$ block sparsity: $G(x) = \sum_{b \in \mathcal{B}} \|x^{[b]}\|, \qquad \|x^{[b]}\|^2 = \sum_{m \in b} x_m^2$

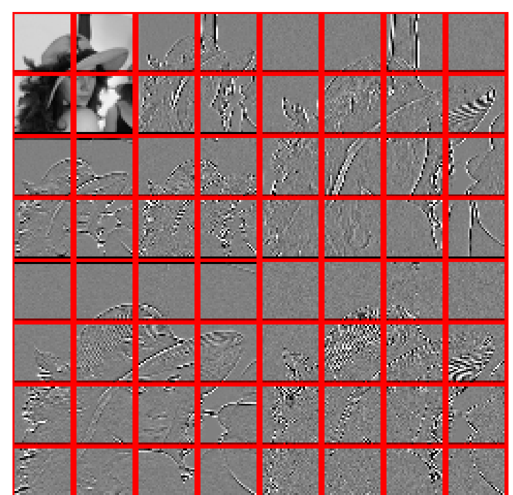
Non-overlapping decomposition: $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_n$

$$G(x) = \sum_{i=1}^{n} G_i(x) \qquad G_i(x) = \sum_{b \in \mathcal{B}_i} \|x^{[b]}\|,$$

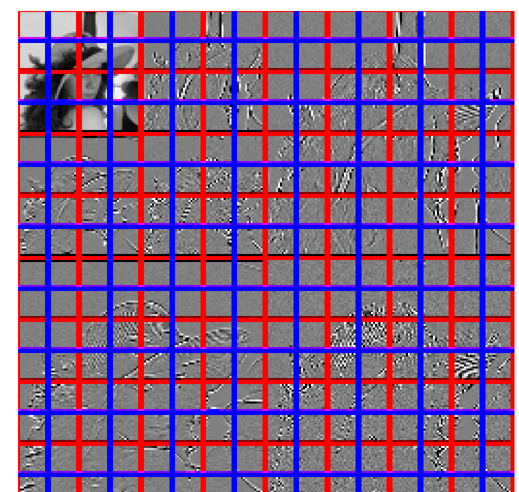
Image $f = \Psi x$



Coefficients x.



Blocks $\mathcal{B}_1$



$\mathcal{B}_1 \cup \mathcal{B}_2$

Block Regularization

$\ell^1 - \ell^2$ block sparsity: $G(x) = \sum_{b \in \mathcal{B}} \|x^{[b]}\|, \qquad \|x^{[b]}\|^2 = \sum_{m \in b} x_m^2$

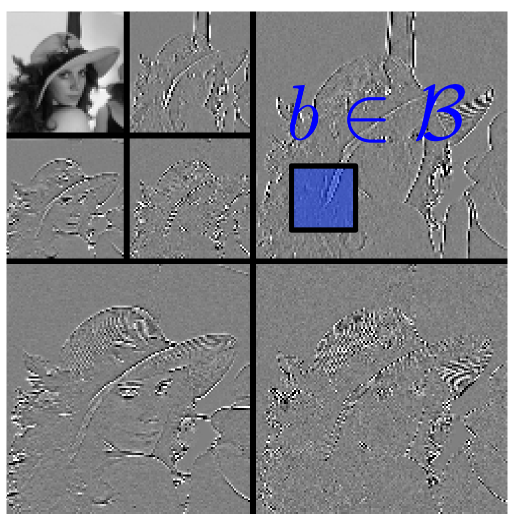
Non-overlapping decomposition: $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_n$

$$G(x) = \sum_{i=1}^{n} G_i(x) \qquad G_i(x) = \sum_{b \in \mathcal{B}_i} \|x^{[b]}\|,$$

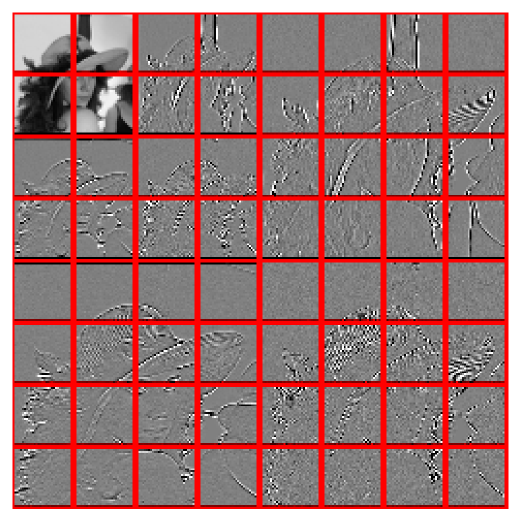
Each G_i is simple:

$$\forall\, m \in b \in \mathcal{B}_i, \quad \mathrm{Prox}_{\gamma G_i}(x)_m = \max\left(0, 1 - \frac{\gamma}{\|x^{[b]}\|}\right) x_m$$

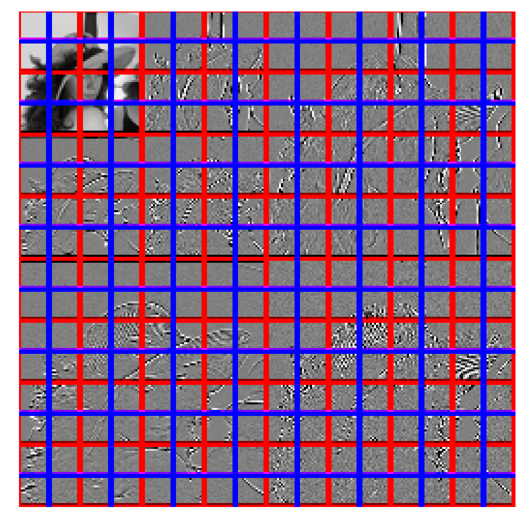
Image $f = \Psi x$



Coefficients x.



Blocks $\mathcal{B}_1$



$\mathcal{B}_1 \cup \mathcal{B}_2$

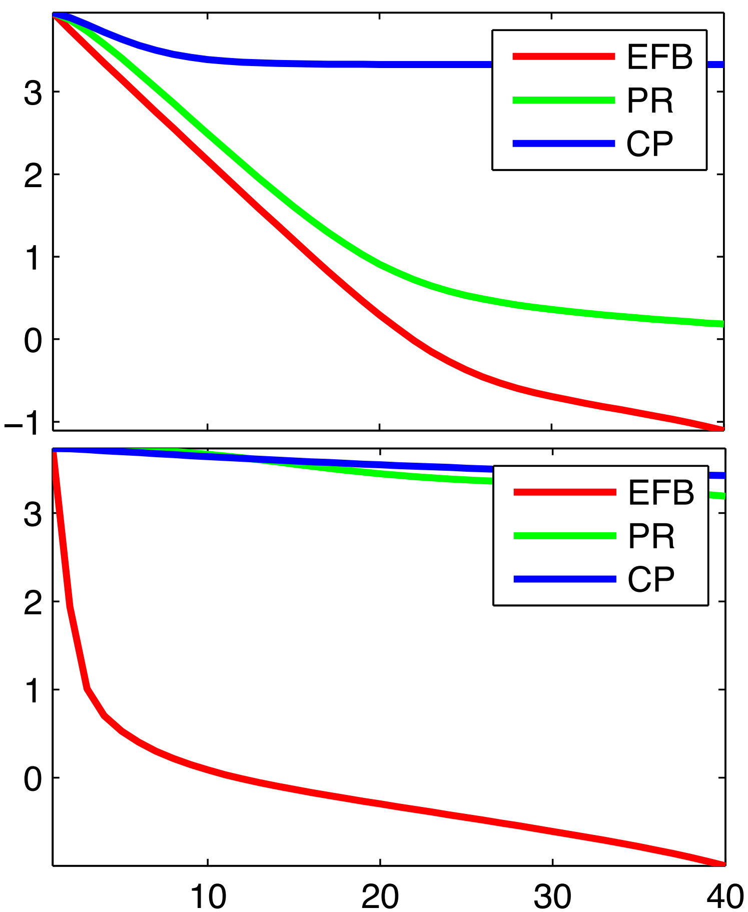
Numerical Illustration

$$\min_x \frac{1}{2}\|y - \Phi\Psi x\|^2 + \lambda \sum_i G_i(x)$$

$\Psi =$ TI wavelets

$\Phi =$ convolution

$\Phi =$ inpainting+convolution



$\log_{10}(E(x^{(\ell)}) - E(x^\star))$

$y = \Phi x_0 + w$ $x^\star$ x_0

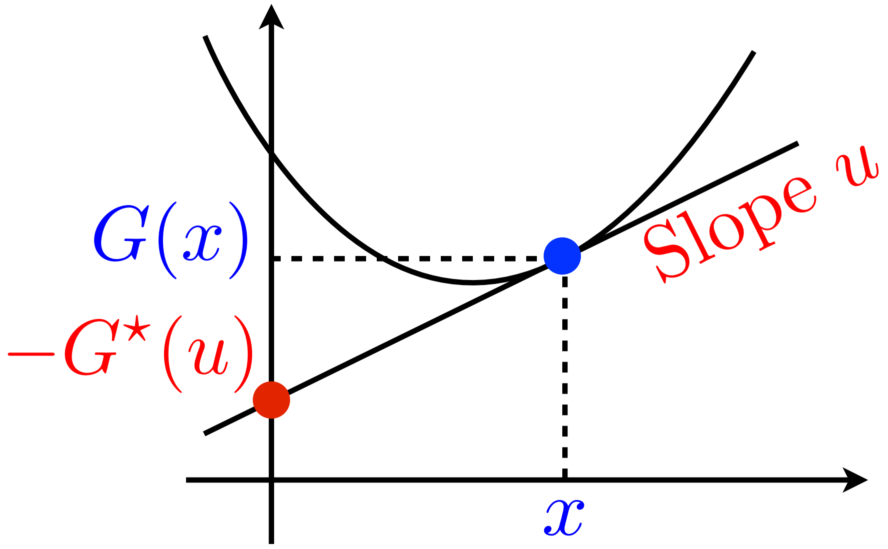
Overview

- Subdifferential Calculus
- Proximal Calculus
- Forward Backward
- Douglas Rachford
- Generalized Forward-Backward
- **Duality**

Legendre-Fenchel Duality

Legendre-Fenchel transform:

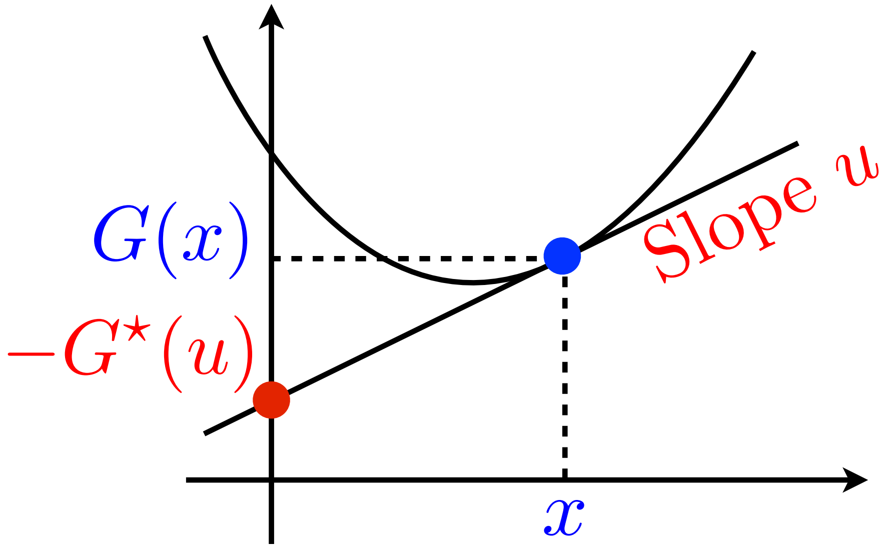
$$G^\star(u) = \sup_{x \in \mathrm{dom}(G)} \langle u,\, x \rangle - G(x)$$



Legendre-Fenchel Duality

Legendre-Fenchel transform:

$$G^\star(u) = \sup_{x \in \mathrm{dom}(G)} \langle u,\, x\rangle - G(x)$$



Example: quadratic functional

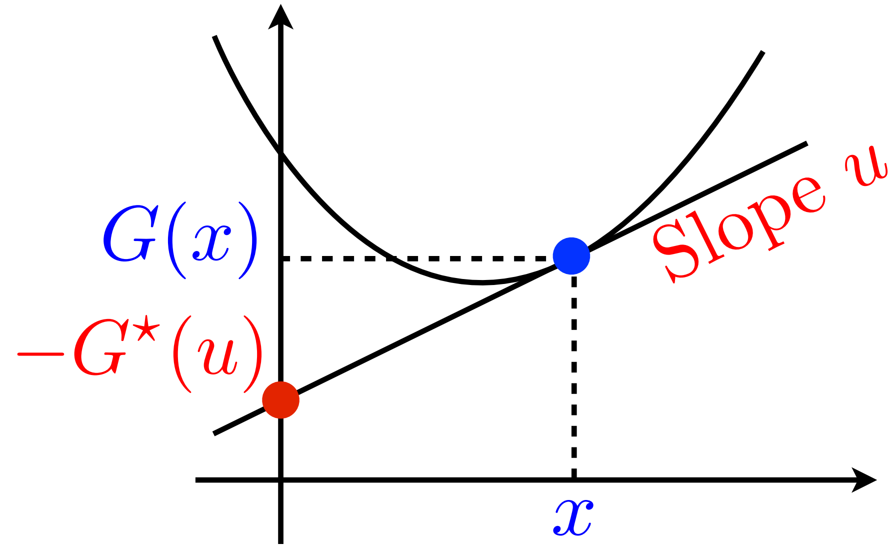
$$G(x) = \frac{1}{2}\langle Ax,\, x\rangle + \langle x,\, b\rangle$$

$$G^\star(u) = \frac{1}{2}\langle u - b,\, A^{-1}(u-b)\rangle$$

Legendre-Fenchel Duality

Legendre-Fenchel transform:

$$G^\star(u) = \sup_{x \in \mathrm{dom}(G)} \langle u,\, x \rangle - G(x)$$



Example: quadratic functional

$$G(x) = \frac{1}{2}\langle Ax,\, x\rangle + \langle x,\, b\rangle$$

$$G^\star(u) = \frac{1}{2}\langle u - b,\, A^{-1}(u - b)\rangle$$

Moreau's identity:

$$\mathrm{Prox}_{\gamma G^\star}(x) = x - \gamma \mathrm{Prox}_{G/\gamma}(x/\gamma)$$

$$G \text{ simple} \quad \Longleftrightarrow \quad G^\star \text{ simple}$$

Indicator and Homogeneous

Positively 1-homogeneous functional: $G(\lambda x) = |x| G(x)$

Example: norm $G(x) = \|x\|$

Duality: $G^\star(x) = \iota_{G_\star(\cdot) \leqslant 1}(x)$ $G_\star(y) = \min_{G(x) \leqslant 1} \langle x, \, y \rangle$

Indicator and Homogeneous

Positively 1-homogeneous functional: $G(\lambda x) = |x| G(x)$

Example: norm $G(x) = \|x\|$

Duality: $G^\star(x) = \iota_{G_\star(\cdot) \leqslant 1}(x)$ $G_\star(y) = \min_{G(x) \leqslant 1} \langle x,\, y \rangle$

ℓ^p *norms:* $G(x) = \|x\|_p$

$G_\star(x) = \|x\|_q$

$$\frac{1}{p} + \frac{1}{q} = 1 \qquad 1 \leqslant p, q \leqslant +\infty$$

Indicator and Homogeneous

Positively 1-homogeneous functional: $G(\lambda x) = |x| G(x)$

Example: norm $G(x) = \|x\|$

Duality: $G^\star(x) = \iota_{G_\star(\cdot) \leqslant 1}(x)$ $G_\star(y) = \min_{G(x) \leqslant 1} \langle x,\, y \rangle$

ℓ^p *norms:* $G(x) = \|x\|_p$

$G_\star(x) = \|x\|_q$ $\frac{1}{p} + \frac{1}{q} = 1$ $1 \leqslant p, q \leqslant +\infty$

Example: Proximal operator of ℓ^∞ norm

$$\mathrm{Prox}_{\gamma \|\cdot\|_\infty} = \mathrm{Id} - \gamma \mathrm{Proj}_{\|\cdot\|_1 \leqslant \gamma}$$

$$\mathrm{Proj}_{\|\cdot\|_1 \leqslant \gamma}(x)_i = \max\left(0, 1 - \frac{\tau}{|x_i|}\right) x_i$$

for a well-chosen $\tau = \tau(x, \gamma)$

Primal-dual Formulation

Fenchel-Rockafellar duality: $A : \mathcal{H} \mapsto \mathcal{L}$ linear

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x) = \min_{x} G_1(x) + \sup_{u \in \mathcal{L}} \langle Ax,\, u \rangle - G_2^*(u)$$

Primal-dual Formulation

Fenchel-Rockafellar duality: $A : \mathcal{H} \mapsto \mathcal{L}$ linear

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x) = \min_{x} G_1(x) + \sup_{u \in \mathcal{L}} \langle Ax,\, u \rangle - G_2^*(u)$$

Strong duality: $0 \in \mathrm{ri}(\mathrm{dom}(G_2)) - A\,\mathrm{ri}(\mathrm{dom}(G_1))$

$(\min \leftrightarrow \max)$

$$= \max_{u} - G_2^*(u) + \min_{x} G_1(x) + \langle x,\, A^* u \rangle$$

$$= \max_{u} - G_2^*(u) - G_1^*(-A^* u)$$

Primal-dual Formulation

Fenchel-Rockafellar duality: $A : \mathcal{H} \mapsto \mathcal{L}$ linear

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x) = \min_x G_1(x) + \sup_{u \in \mathcal{L}} \langle Ax, u \rangle - G_2^*(u)$$

Strong duality: $0 \in \mathrm{ri}(\mathrm{dom}(G_2)) - A\,\mathrm{ri}(\mathrm{dom}(G_1))$

$(\min \leftrightarrow \max)$

$$= \max_u -G_2^*(u) + \boxed{\min_x G_1(x) + \langle x, A^* u \rangle}$$

$$= \max_u -G_2^*(u) - G_1^*(-A^* u)$$

Recovering $x^\star$ from some $u^\star$:

$$\boxed{x^\star = \operatorname*{argmin}_x G_1(x^\star) + \langle x^\star, A^* u^\star \rangle}$$

Primal-dual Formulation

Fenchel-Rockafellar duality: $A : \mathcal{H} \mapsto \mathcal{L}$ linear

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x) = \min_x G_1(x) + \sup_{u \in \mathcal{L}} \langle Ax, u \rangle - G_2^*(u)$$

Strong duality: $0 \in \mathrm{ri}(\mathrm{dom}(G_2)) - A\, \mathrm{ri}(\mathrm{dom}(G_1))$

$(\min \leftrightarrow \max)$

$$= \max_u -G_2^*(u) + \boxed{\min_x G_1(x) + \langle x, A^* u \rangle}$$

$$= \max_u -G_2^*(u) - G_1^*(-A^* u)$$

Recovering $x^\star$ from some $u^\star$:

$$\boxed{x^\star = \operatorname*{argmin}_x G_1(x^\star) + \langle x^\star, A^* u^\star \rangle}$$

$$\iff \quad -A^* u^\star \in \partial G_1(x^\star)$$

$$\iff \quad x^\star \in (\partial G_1)^{-1}(-A^* u^\star) = \partial G_1^*(-A^* u^\star)$$

Forward-Backward on the Dual

If G_1 is strongly convex: $\nabla^2 G_1 \geqslant c\mathrm{Id}$

$$G_1(tx + (1-t)y) \leqslant tG_1(x) + (1-t)G_1(y) - \frac{c}{2}t(1-t)\|x-y\|^2$$

Forward-Backward on the Dual

If G_1 is strongly convex: $\nabla^2 G_1 \geqslant c\mathrm{Id}$

$$G_1(tx + (1-t)y) \leqslant tG_1(x) + (1-t)G_1(y) - \frac{c}{2}t(1-t)\|x-y\|^2$$

- $x^\star$ uniquely defined.
- $G_1^\star$ is of class C^1.

$$x^\star = \nabla G_1^\star(-A^* u^\star)$$

Forward-Backward on the Dual

If G_1 is strongly convex: $\nabla^2 G_1 \geqslant c\mathrm{Id}$

$$G_1(tx + (1-t)y) \leqslant tG_1(x) + (1-t)G_1(y) - \frac{c}{2}t(1-t)\|x-y\|^2$$

- $x^\star$ uniquely defined.
- $G_1^\star$ is of class C^1.

$$x^\star = \nabla G_1^\star(-A^* u^\star)$$

FB on the dual:

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x)$$

$$= -\min_{u \in \mathcal{L}} G_1^\star(-A^* u) + G_2^\star(u)$$

Smooth Simple

$$u^{(\ell+1)} = \mathrm{Prox}_{\tau G_2^\star}\left(u^{(\ell)} + \tau A^* \nabla G_1^\star(-A^* u^{(\ell)}) \right)$$

Example: TV Denoising

$$\min_{f \in \mathbb{R}^N} \frac{1}{2}\|f - y\|^2 + \lambda \|\nabla f\|_1 \iff \min_{\|u\|_\infty \leqslant \lambda} \|y + \mathrm{div}(u)\|^2$$

$$\|u\|_1 = \sum_i \|u_i\| \qquad \|u\|_\infty = \max_i \|u_i\|$$

Dual solution $u^\star$ $\longrightarrow$ Primal solution $f^\star = y + \mathrm{div}(u^\star)$

[Chambolle 2004]

Example: TV Denoising

$$\min_{f \in \mathbb{R}^N} \frac{1}{2}\|f - y\|^2 + \lambda \|\nabla f\|_1 \iff \min_{\|u\|_\infty \leqslant \lambda} \|y + \mathrm{div}(u)\|^2$$

$$\|u\|_1 = \sum_i \|u_i\| \qquad \|u\|_\infty = \max_i \|u_i\|$$

Dual solution $u^\star$ $\longrightarrow$ Primal solution $f^\star = y + \mathrm{div}(u^\star)$

FB (aka projected gradient descent): [Chambolle 2004]

$$u^{(\ell+1)} = \mathrm{Proj}_{\|\cdot\|_\infty \leqslant \lambda}\left(u^{(\ell)} + \gamma \nabla (y + \mathrm{div}(u^{(\ell)})) \right)$$

$$v = \mathrm{Proj}_{\|\cdot\|_\infty \leqslant \lambda}(u) \qquad v_i = \frac{u_i}{\max(\|u_i\|/\lambda, 1)}$$

Convergence if $\gamma < \dfrac{2}{\|\mathrm{div} \circ \nabla\|} = \dfrac{1}{4}$

Primal-Dual Algorithm

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x)$$

$$\iff \quad \min_x \max_z G_1(x) - G_2^*(z) + \langle A(x),\ z \rangle$$

Primal-Dual Algorithm

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x)$$

$$\iff \quad \min_x \max_z G_1(x) - G_2^*(z) + \langle A(x),\ z \rangle$$

$$z^{(\ell+1)} = \mathrm{Prox}_{\sigma G_2^*}(z^{(\ell)} + \sigma A(\tilde{x}^{(\ell)})$$
$$x^{(\ell+1)} = \mathrm{Prox}_{\tau G_1}(x^{(\ell)} - \tau A^*(z^{(\ell)}))$$
$$\tilde{x}^{(\ell+1)} = x^{(\ell+1)} + \theta(x^{(\ell+1)} - x^{(\ell)})$$

$\theta = 0$: Arrow-Hurwicz algorithm.

$\theta = 1$: convergence speed on duality gap.

Primal-Dual Algorithm

$$\min_{x \in \mathcal{H}} G_1(x) + G_2 \circ A(x)$$

$$\iff \quad \min_x \max_z G_1(x) - G_2^*(z) + \langle A(x),\, z\rangle$$

$$z^{(\ell+1)} = \text{Prox}_{\sigma G_2^*}(z^{(\ell)} + \sigma A(\tilde{x}^{(\ell)})$$
$$x^{(\ell+1)} = \text{Prox}_{\tau G_1}(x^{(\ell)} - \tau A^*(z^{(\ell)}))$$
$$\tilde{x}^{(\ell+1)} = x^{(\ell+1)} + \theta(x^{(\ell+1)} - x^{(\ell)})$$

$\theta = 0$: Arrow-Hurwicz algorithm.
$\theta = 1$: convergence speed on duality gap.

Theorem: [Chambolle-Pock 2011]

If $0 \leqslant \theta \leqslant 1$ and $\sigma\tau\|A\|^2 < 1$ then

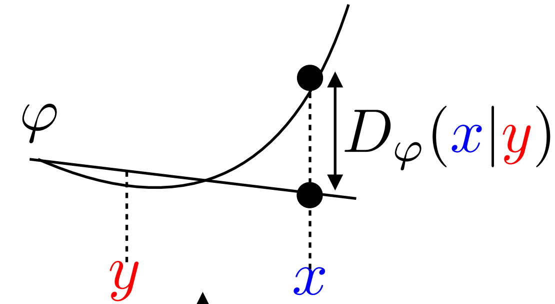
$x^{(\ell)} \rightarrow x^\star$ minimizer of $G_1 + G_2 \circ A$.

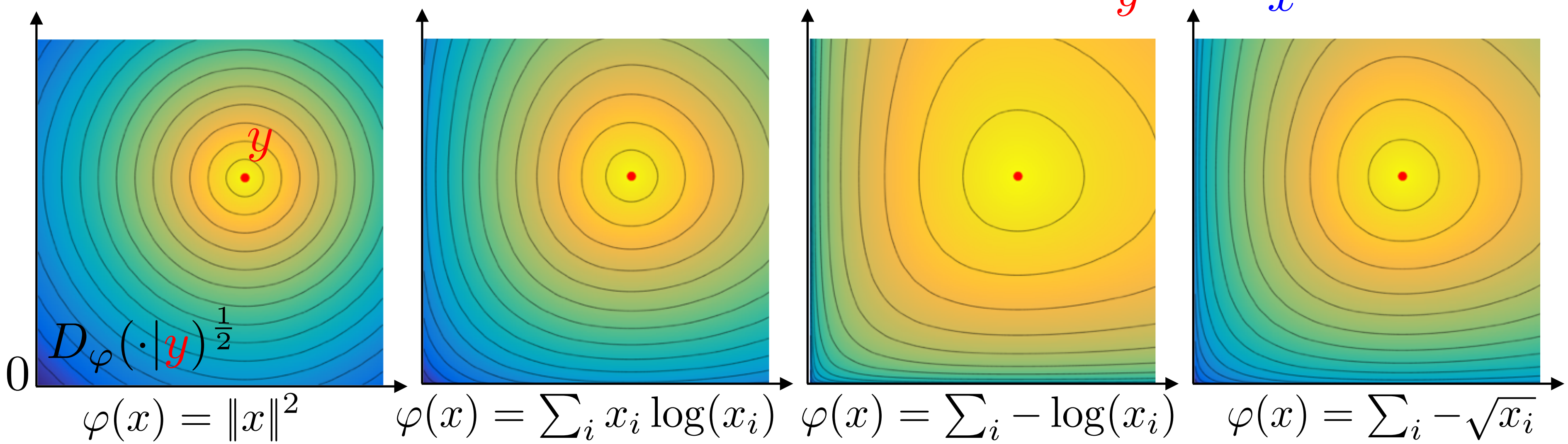
Other algorithms

- Frank Wolfe, Mirror Descent

Bregman divergence:

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

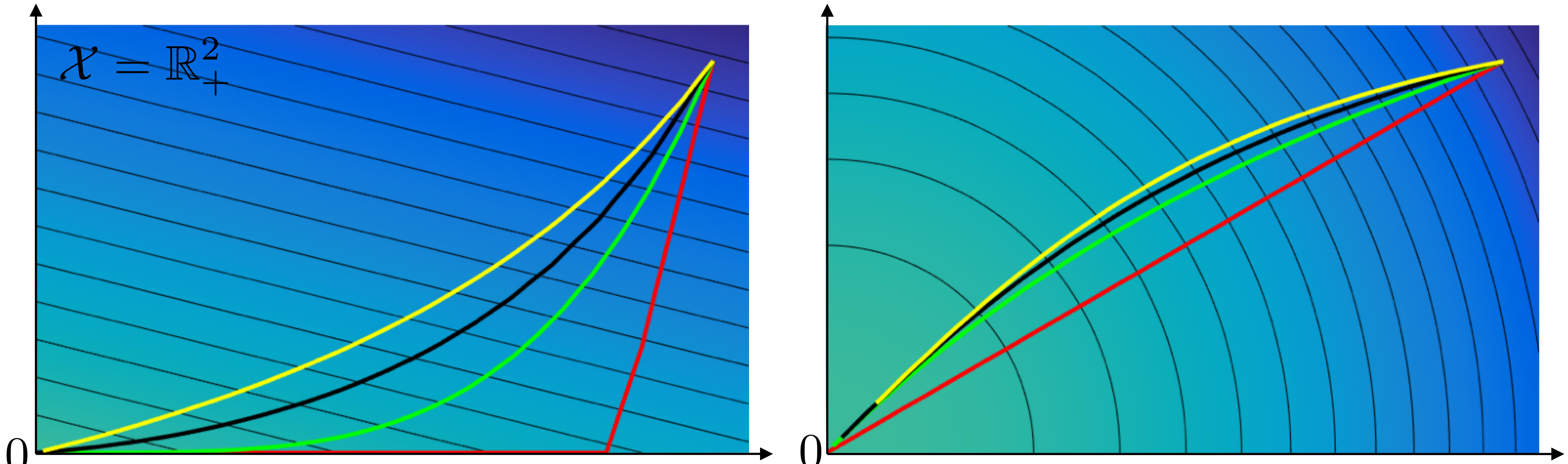




$$\left.\begin{array}{l} D_\varphi(x|x+\varepsilon) \\ D_\varphi(x+\varepsilon|x) \end{array}\right\} = \frac{1}{2}\langle \partial^2\varphi(x)\varepsilon,\, \varepsilon\rangle + o(\|\varepsilon\|^2)$$

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

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



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

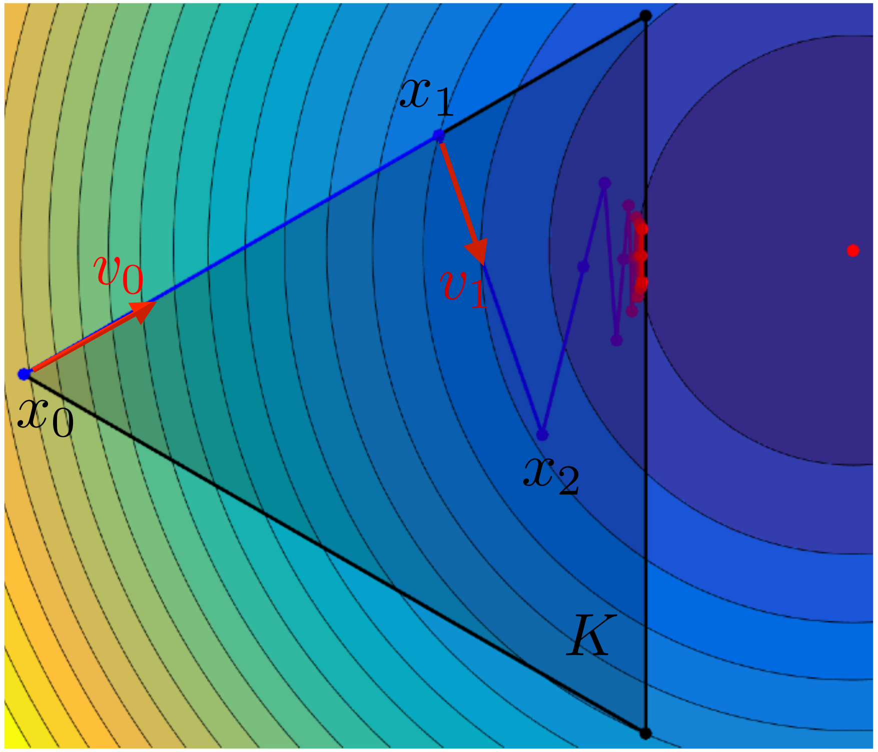
Directional derivative: $D_v f(x) \overset{\text{def.}}{=} \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$

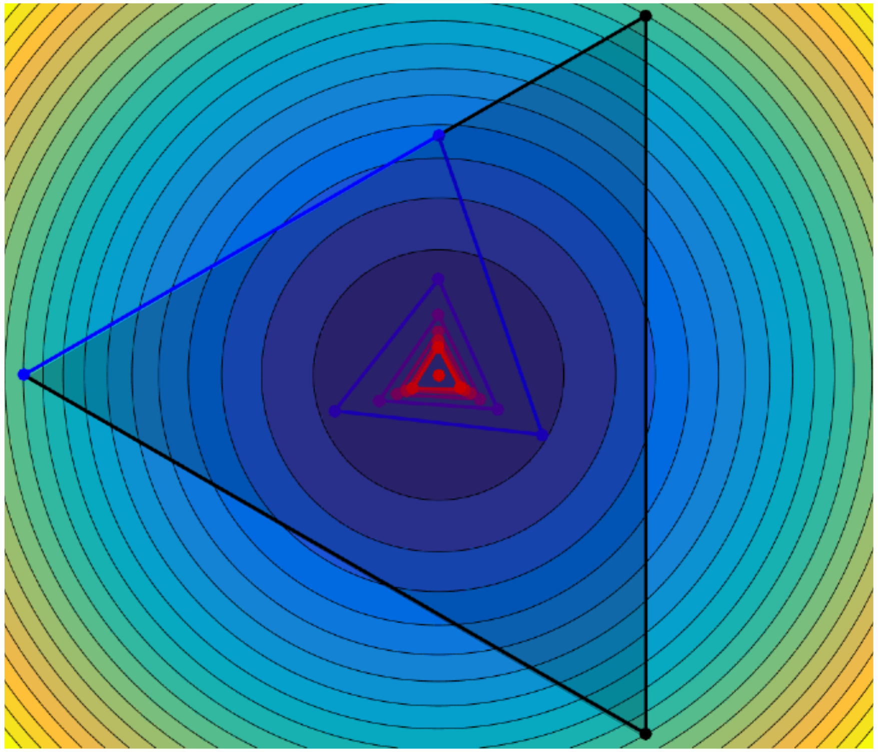
$$\min_{x \in K} f(x)$$

Frank-Wolfe

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

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





Conclusion

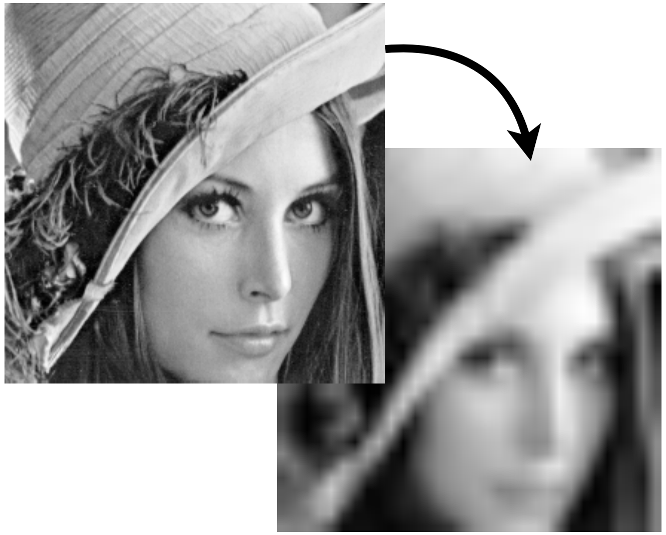
Inverse problems in imaging:

$\longrightarrow$ Large scale, $N \geqslant 10^6$.

$\longrightarrow$ Non-smooth (sparsity, TV, ...)

$\longrightarrow$ (Sometimes) convex.

$\longrightarrow$ Highly structured (separability, ℓ^p norms, ...).



Conclusion

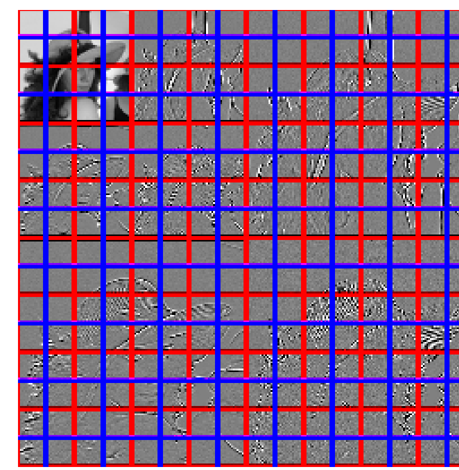
Inverse problems in imaging:

$\longrightarrow$ Large scale, $N \geqslant 10^6$.

$\longrightarrow$ Non-smooth (sparsity, TV, ...)

$\longrightarrow$ (Sometimes) convex.

$\longrightarrow$ Highly structured (separability, ℓ^p norms, ...).



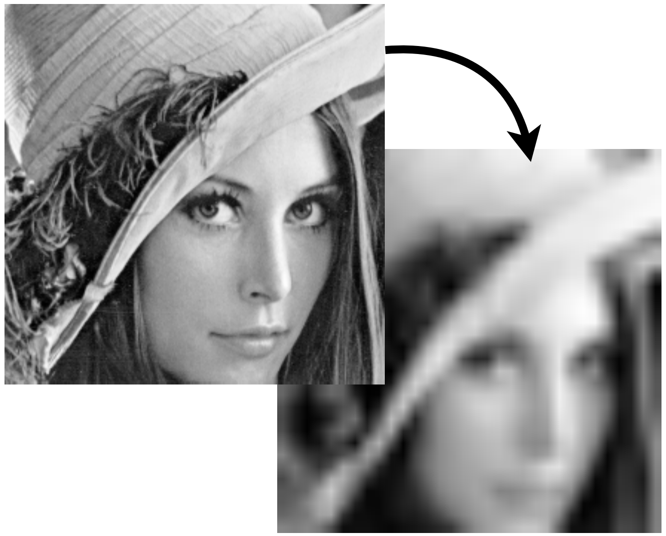
Proximal splitting:

$\longrightarrow$ Unravel the structure of problems.

$\longrightarrow$ Parallelizable.

Conclusion

Inverse problems in imaging:



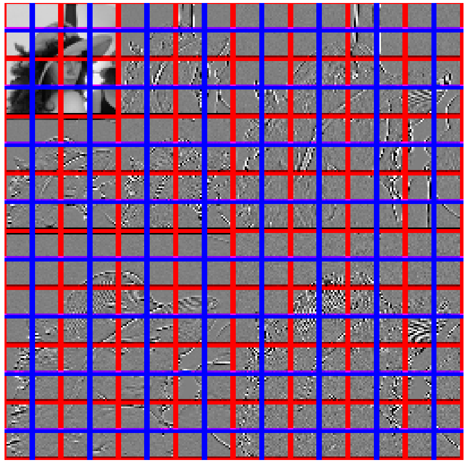
→ Large scale, $N \geqslant 10^6$.

→ Non-smooth (sparsity, TV, . . .)

→ (Sometimes) convex.

→ Highly structured (separability, ℓ^p norms, . . .).

Proximal splitting:



→ Unravel the structure of problems.

→ Parallelizable.

Open problems:

→ Less structured problems without smoothness.

→ Non-convex optimization.