Proximal Splitting and Optimal Transport

Gabriel Peyré

www.numerical-tours.com

Overview

- *Optimal Transport and Imaging*

- Convex Analysis and Proximal Calculus

- Forward Backward

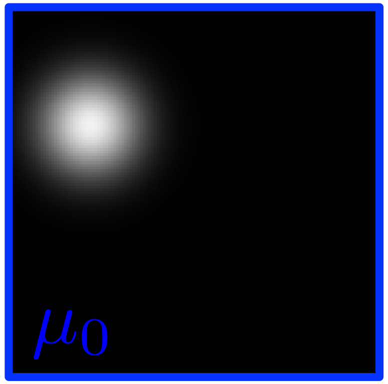
- Douglas Rachford and ADMM

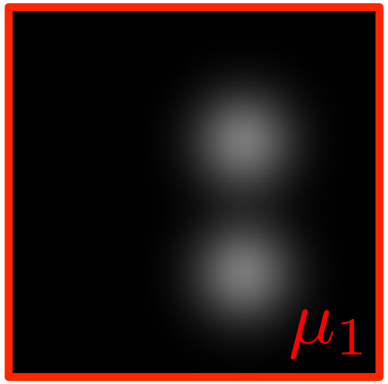
- Generalized Forward-Backward

- Primal-Dual Schemes

Measure Preserving Maps

Distributions μ_0, μ_1 on $\mathbb{R}^k$.



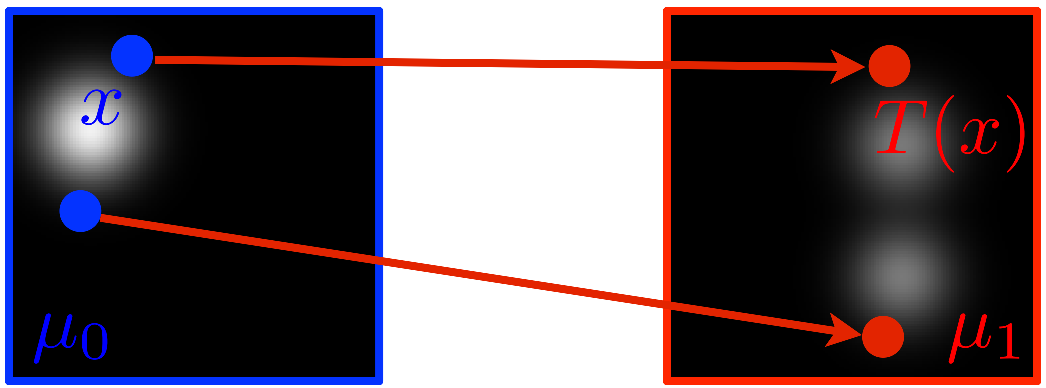


Measure Preserving Maps

Distributions μ_0, μ_1 on $\mathbb{R}^k$.

Mass preserving map $T : \mathbb{R}^k \to \mathbb{R}^k$.

$$\mu_1 = T\sharp\mu_0 \quad \text{where} \quad (T\sharp\mu_0)(A) = \mu_0(T^{-1}(A))$$

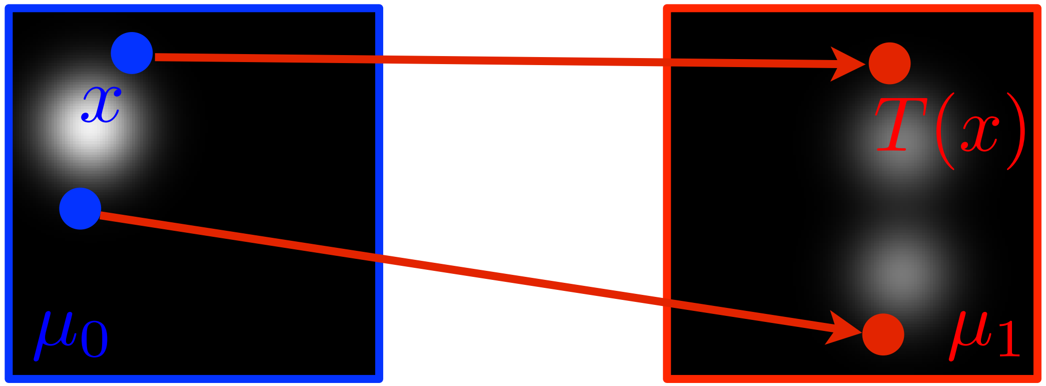


Measure Preserving Maps

Distributions μ_0, μ_1 on $\mathbb{R}^k$.

Mass preserving map $T : \mathbb{R}^k \to \mathbb{R}^k$.

$$\mu_1 = T\sharp\mu_0 \quad \text{where} \quad (T\sharp\mu_0)(A) = \mu_0(T^{-1}(A))$$



Distributions with densities: $\quad \mu_i = \rho_i(x)\mathrm{d}x$

$$T\sharp\mu_0 = \mu_1 \quad \Longleftrightarrow \quad \rho_1(T(x))|\det \partial T(x)| = \rho_0(x)$$

Optimal Transport

L^p optimal transport:

$$W_2(\mu_0, \mu_1)^p = \min_{T \sharp \mu_0 = \mu_1} \int \|T(x) - x\|^p \mu_0(\mathrm{d}x)$$

Optimal Transport

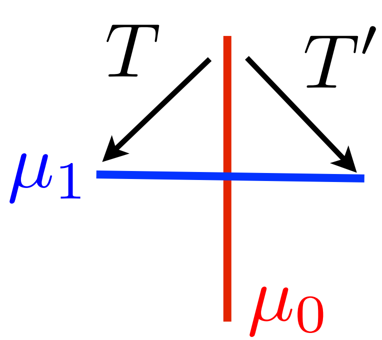
L^p optimal transport:

$$W_2(\mu_0, \mu_1)^p = \min_{T\sharp\mu_0=\mu_1} \int \|T(x) - x\|^p \mu_0(\mathrm{d}x)$$

Regularity condition:

μ_0 or μ_1 does not give mass to "small sets".

Theorem $(p > 1)$*:* there exists a unique optimal T.



Optimal Transport

L^p optimal transport:

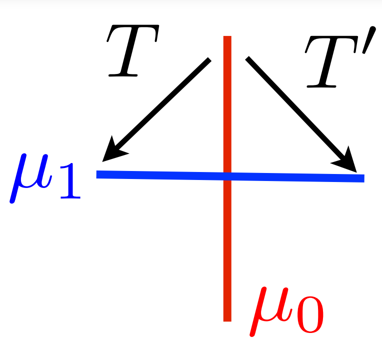
$$W_2(\mu_0, \mu_1)^p = \min_{T\sharp\mu_0=\mu_1} \int \|T(x) - x\|^p \mu_0(\mathrm{d}x)$$

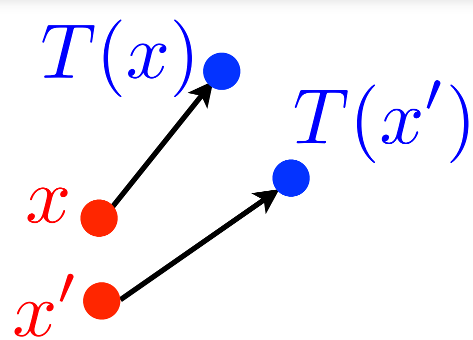
Regularity condition:

μ_0 or μ_1 does not give mass to "small sets".

Theorem $(p > 1)$*:* there exists a unique optimal T.

Theorem $(p = 2)$*:* T is defined as $T = \nabla\psi$ with ψ convex.





T is monotone:

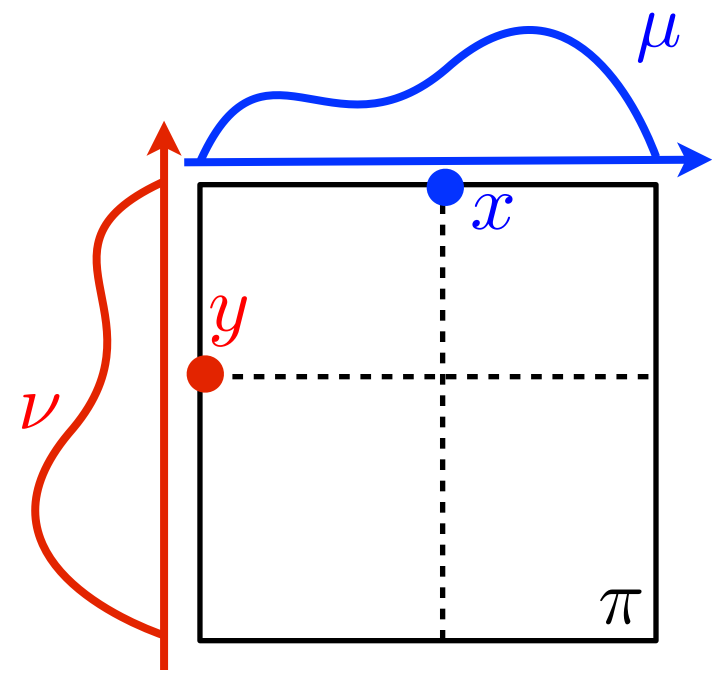
$$\langle T(x) - T(x'),\, x - x' \rangle \geqslant 0$$

Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.

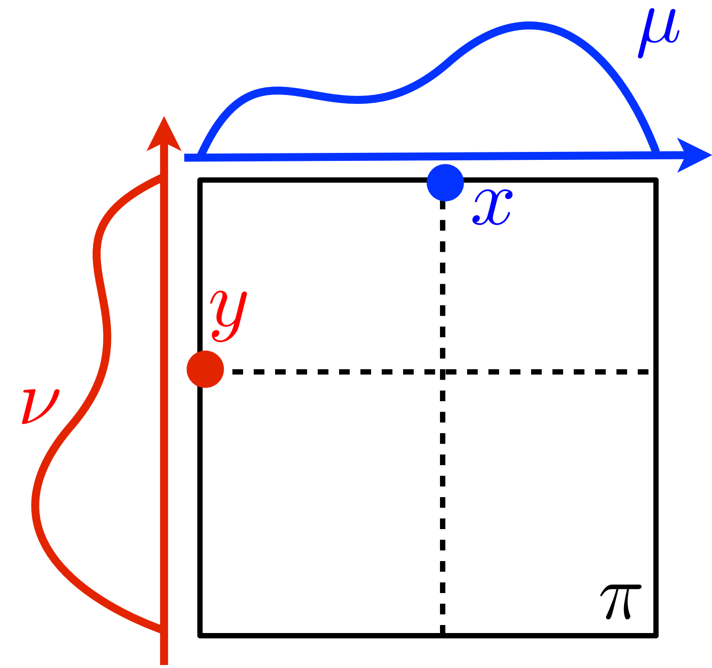
Couplings: $\pi \in \Pi_{\mu,\nu}$

$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$



Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.



Couplings: $\pi \in \Pi_{\mu,\nu}$

$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$

Transportation cost:

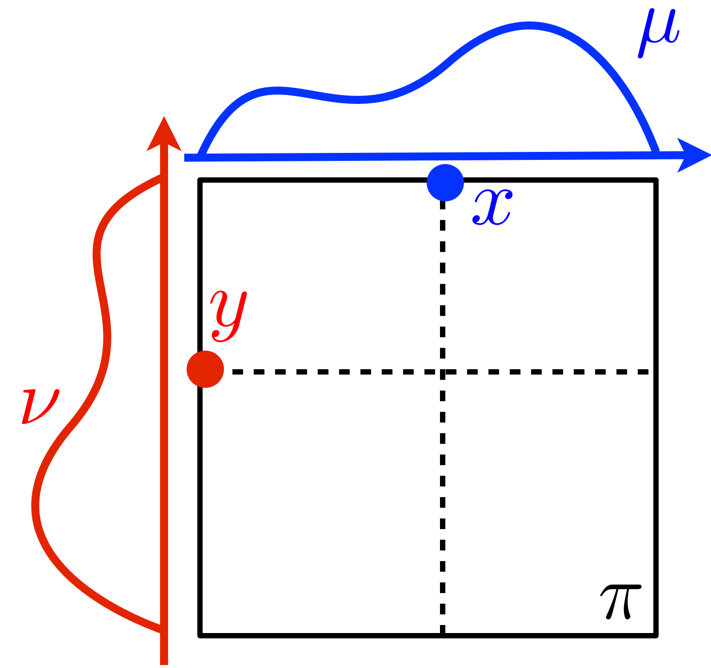
$$c(x,y) = \|x - y\|^p$$

L^p Wasserstein distance:

$$W_p(\mu,\nu)^p = \min_{\pi \in \Pi_{\mu,\nu}} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \mathrm{d}\pi(x,y)$$

Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.



Couplings: $\pi \in \Pi_{\mu,\nu}$

$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$

Transportation cost:

$$c(x, y) = \|x - y\|^p$$

L^p Wasserstein distance:

$$W_p(\mu, \nu)^p = \min_{\pi \in \Pi_{\mu,\nu}} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \mathrm{d}\pi(x, y)$$

Uniqueness: (μ does not vanish on small sets)

If $p > 1$, $\exists !\, \pi$ optimal.

Optimal Transport

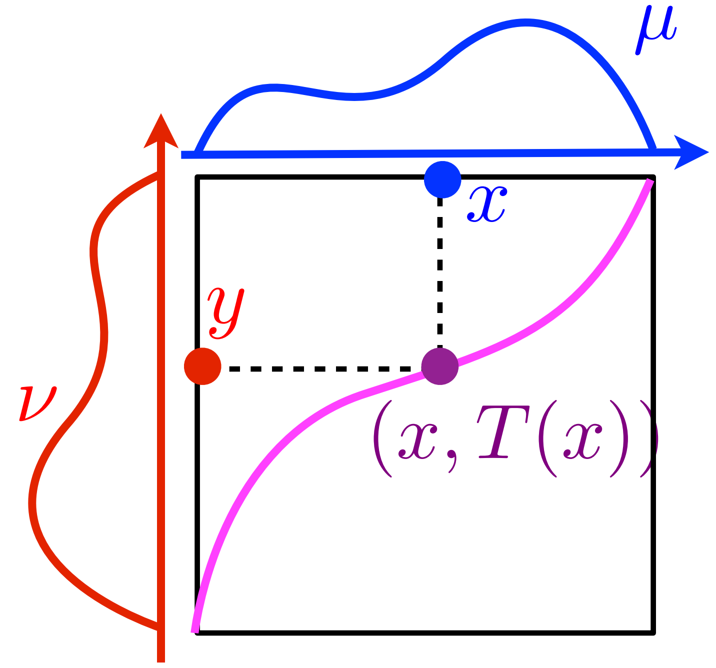
Let $p > 1$ and μ does not vanish on small sets.

$$\text{Unique } \pi \in \Pi_{\mu,\nu} \text{ s.t.} \quad W_p(\mu,\nu)^p = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\mathrm{d}\pi(x,y)$$

Optimal transport $T : \mathbb{R}^d \to \mathbb{R}^d$*:*

π is supported on the graph of $x \mapsto y = T(x)$.

$$\mathrm{d}\pi(x,y) = \mathrm{d}\mu(x)\delta(y = T(x))$$



Optimal Transport

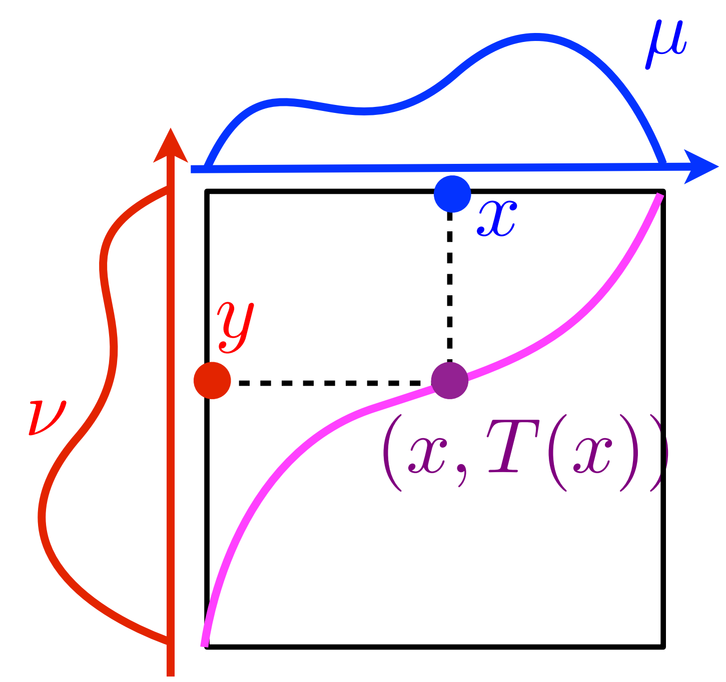
Let $p > 1$ and μ does not vanish on small sets.

Unique $\pi \in \Pi_{\mu,\nu}$ s.t. $\quad W_p(\mu,\nu)^p = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \mathrm{d}\pi(x,y)$

Optimal transport $T : \mathbb{R}^d \to \mathbb{R}^d$*:*

π is supported on the graph of $x \mapsto y = T(x)$.

$\mathrm{d}\pi(x,y) = \mathrm{d}\mu(x)\delta(y = T(x))$



$p = 2$: $T = \nabla\varphi$ unique solution of

$$\begin{cases} \varphi \text{ is convex l.s.c.} \\ (\nabla\varphi)\sharp\mu = \nu \end{cases}$$

1-D Continuous Wasserstein

Distributions μ, ν on $\mathbb{R}$.

Cumulative functions: $$C_\mu(t) = \int_{-\infty}^{t} \mathrm{d}\mu(x)$$

For all $p > 1$: $T = C_\nu^{-1} \circ C_\mu$

T is non-decreasing ("change of contrast")

1-D Continuous Wasserstein

Distributions μ, ν on $\mathbb{R}$.

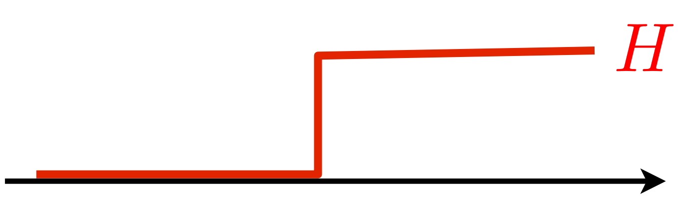
Cumulative functions: $$C_\mu(t) = \int_{-\infty}^{t} \mathrm{d}\mu(x)$$

For all $p > 1$: $T = C_\nu^{-1} \circ C_\mu$

T is non-decreasing ("change of contrast")

Explicit formulas:

$$W_p(\mu, \nu)^p = \int_0^1 |C_\mu^{-1} - C_\nu^{-1}|^p$$



$$W_1(\mu, \nu) = \int_{\mathbb{R}} |C_\mu - C_\nu| = \|(C_\mu - C_\nu) \star H\|_1$$

Grayscale Histogram Transfer

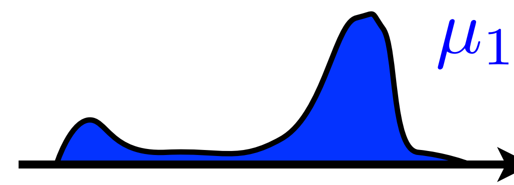
Input images: $f_i : [0,1]^2 \rightarrow [0,1], i = 0, 1.$

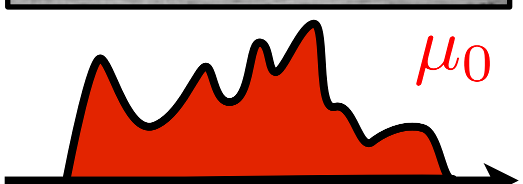
Grayscale Histogram Transfer

Input images: $f_i : [0,1]^2 \rightarrow [0,1], i = 0, 1.$

Gray-value distributions: μ_i defined on $[0,1]$.

$$\mu_i([a,b]) = \int_{[0,1]^2} 1_{\{a \leqslant f \leqslant b\}}(x) \mathrm{d}x$$





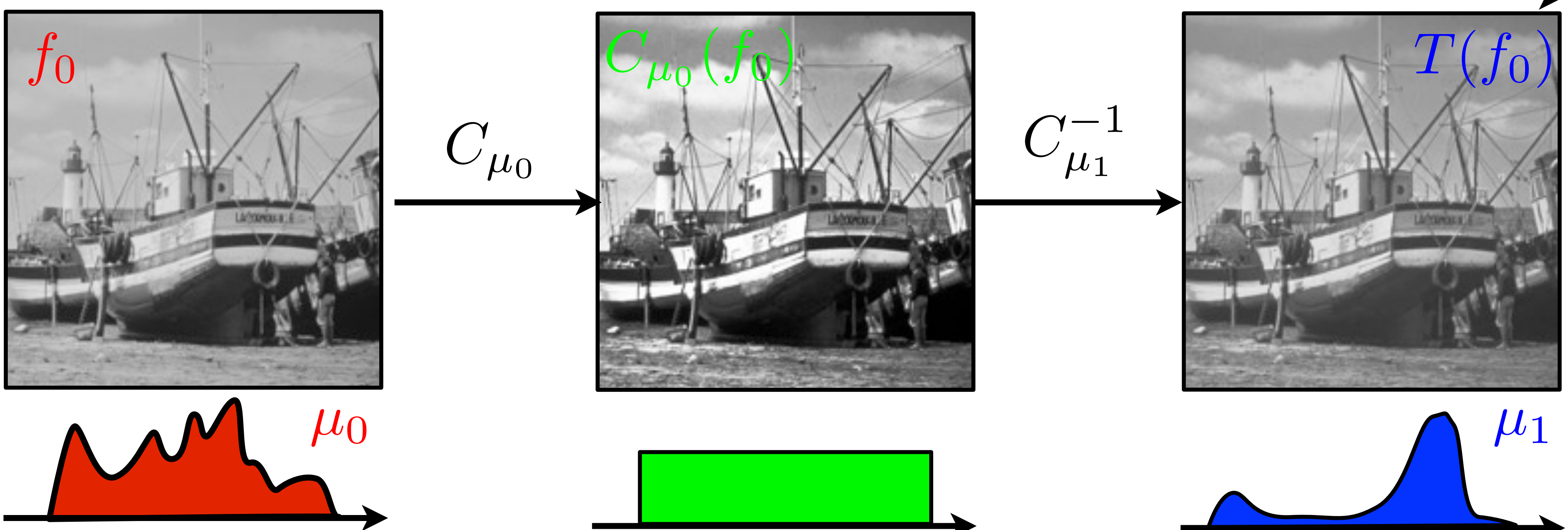
Grayscale Histogram Transfer

Input images: $f_i : [0,1]^2 \rightarrow [0,1], i = 0, 1.$

Gray-value distributions: μ_i defined on $[0,1]$.

$$\mu_i([a,b]) = \int_{[0,1]^2} 1_{\{a \leqslant f \leqslant b\}}(x) \mathrm{d}x$$

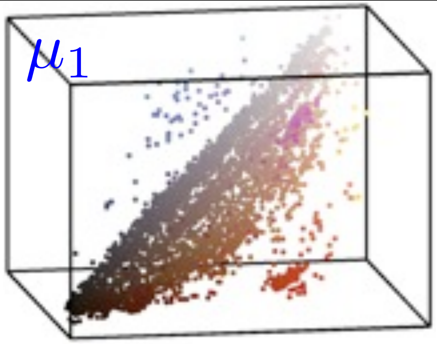
Optimal transport: $T = C_{\mu_1}^{-1} \circ C_{\mu_0}$.

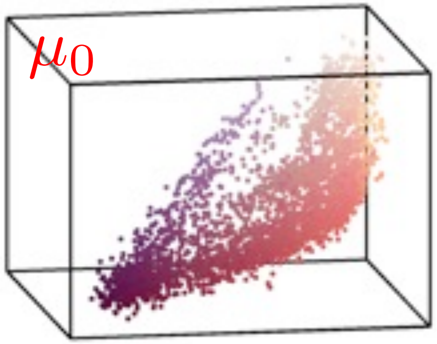


Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$.

$$\nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$$





Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$. $\qquad \nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$

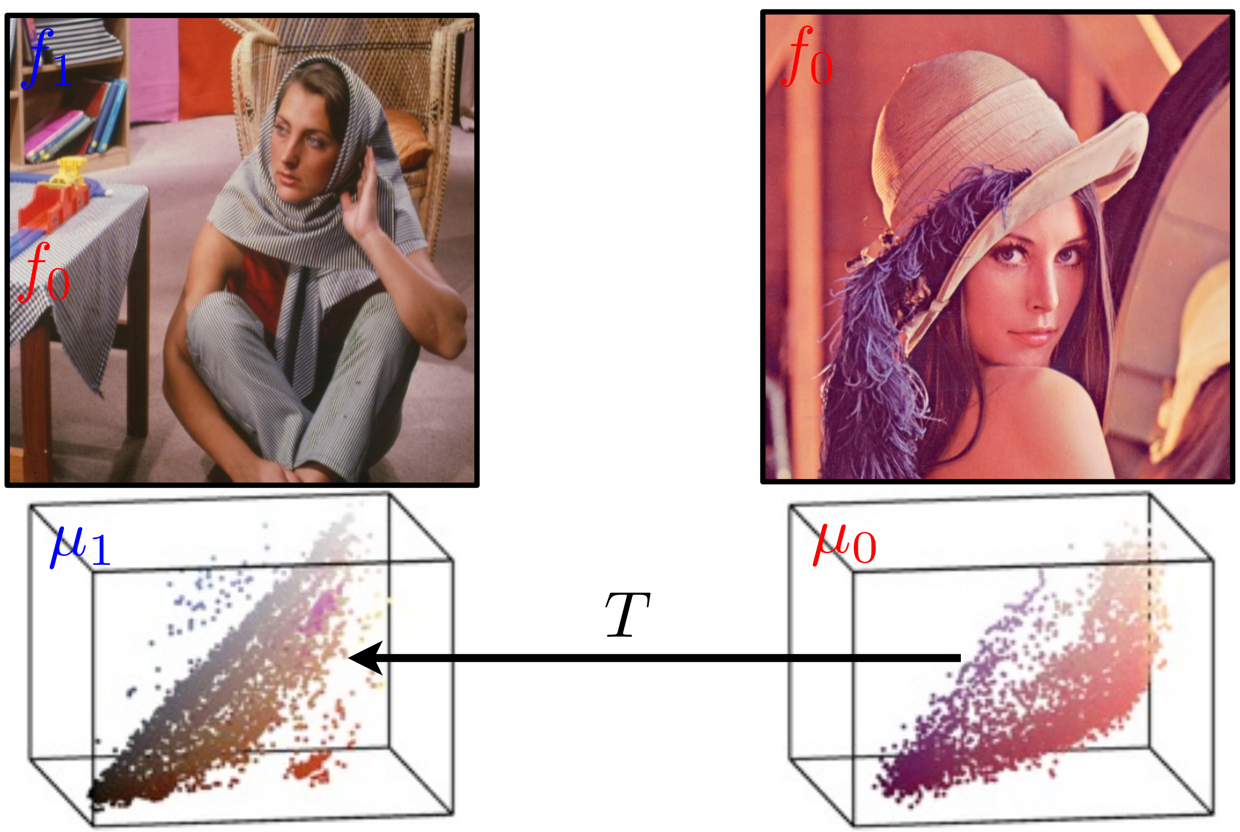
Optimal assignement: $\qquad \min_{\sigma \in \Sigma_N} \| f_0 - f_1 \circ \sigma \|$

Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$. $\qquad \nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$

Optimal assignement: $\min_{\sigma \in \Sigma_N} \|f_0 - f_1 \circ \sigma\|$

Transport: $T : f_0(x) \in \mathbb{R}^3 \mapsto f_1(\sigma(i)) \in \mathbb{R}^3$



Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N\times 3}$. $\nu_i = \frac{1}{N}\sum_x \delta_{f_i(x)}$

Optimal assignement: $\min_{\sigma\in\Sigma_N} \|f_0 - f_1 \circ \sigma\|$

Transport: $T : f_0(x) \in \mathbb{R}^3 \mapsto f_1(\sigma(i)) \in \mathbb{R}^3$

Equalization: $\tilde{f}_0 = T(f_0) \iff \tilde{f}_0 = f_1 \circ \sigma$

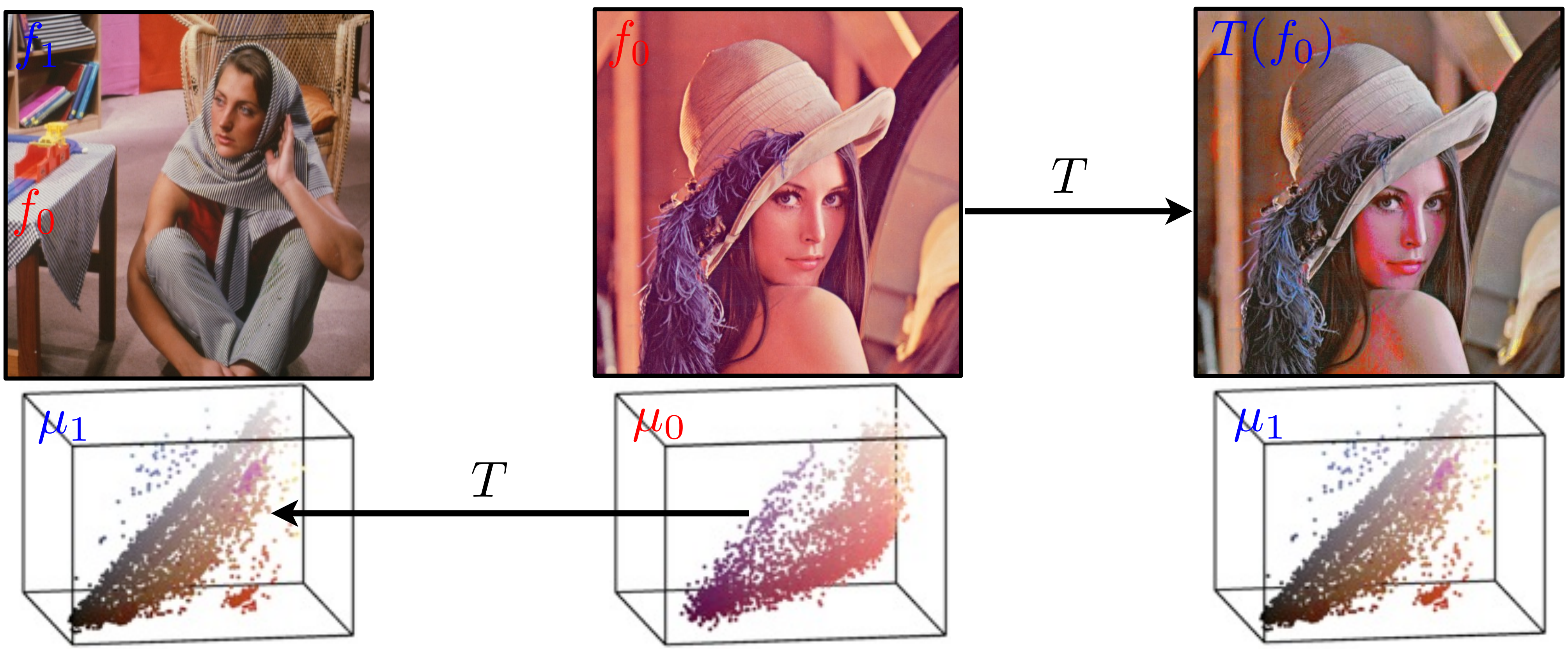
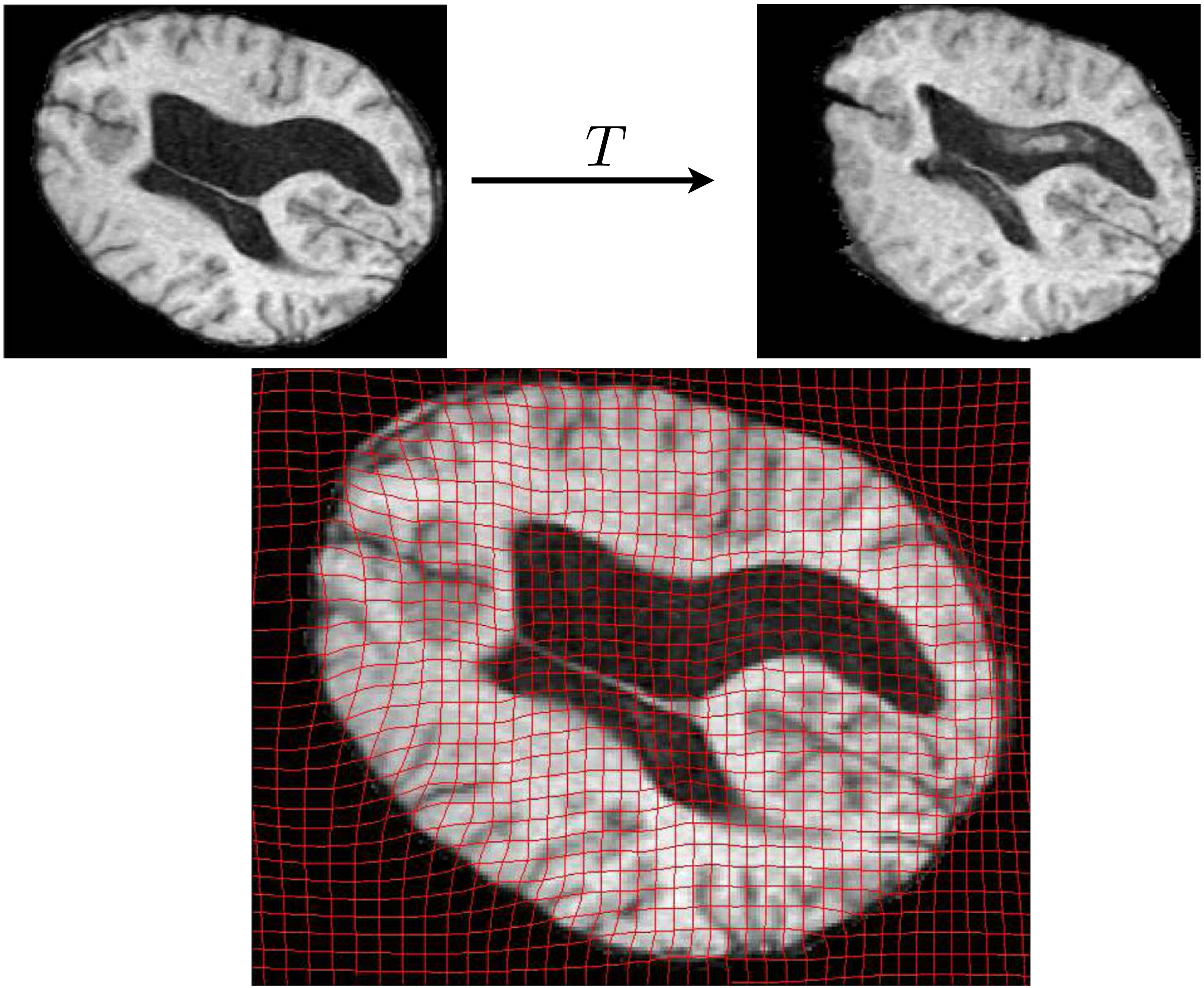


Image Registration



[ur Rehman et al, 2009]

Convex Formulation (Benamou-Brenier)

Find $\begin{cases} \rho \ : \mathbb{R}^d \times [0,1] \to \mathbb{R}^+ \\ m : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d \end{cases}$ solving:

$$\boxed{W(\mu_0, \mu_1)^2 = \min_{x=(m,\rho)} J(x) + \iota_{\mathcal{C}}(x)}$$

Convex Formulation (Benamou-Brenier)

Find $\begin{cases} \rho \;: \mathbb{R}^d \times [0,1] \to \mathbb{R}^+ \\ m : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d \end{cases}$ solving:

$$\boxed{W(\mu_0,\mu_1)^2 = \min_{x=(m,\rho)} J(x) + \iota_{\mathcal{C}}(x)}$$

$$J(x) = \int_{s\in\mathbb{R}^d} \int_{t=0}^{1} j(x(s,t))\mathrm{d}t\mathrm{d}s$$

$$j(\tilde m, \tilde\rho) = \begin{cases} \frac{\|\tilde m\|^2}{\tilde\rho} & \text{if} \quad \tilde\rho > 0, \\ 0 \quad \text{if} \quad \tilde\rho = 0 \quad \text{and} \quad \tilde m = 0, \\ +\infty \quad \text{otherwise.} \end{cases}$$

$$\in \mathbb{R} \qquad \in \mathbb{R}^2$$

Convex Formulation (Benamou-Brenier)

Find $\begin{cases} \rho \ : \mathbb{R}^d \times [0,1] \to \mathbb{R}^+ \\ m : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d \end{cases}$ solving:

$$\boxed{W(\mu_0,\mu_1)^2 = \min_{x=(m,\rho)} J(x) + \iota_{\mathcal{C}}(x)}$$

$$J(x) = \int_{s\in\mathbb{R}^d} \int_{t=0}^{1} j(x(s,t))\mathrm{d}t\mathrm{d}s$$

$$\underbrace{j(\tilde{m},\tilde{\rho})}_{\tilde{m}\in\mathbb{R}^2,\ \tilde{\rho}\in\mathbb{R}} = \begin{cases} \frac{\|\tilde{m}\|^2}{\tilde{\rho}} & \text{if} \quad \tilde{\rho} > 0, \\ 0 \quad \text{if} \quad \tilde{\rho} = 0 \quad \text{and} \quad \tilde{m} = 0, \\ +\infty \quad \text{otherwise.} \end{cases}$$

$$\mathcal{C} = \{x = (m,\rho) \setminus \mathrm{div}(x) = 0,\ \mathcal{B}(\rho) = (\rho_0,\rho_1)\}$$

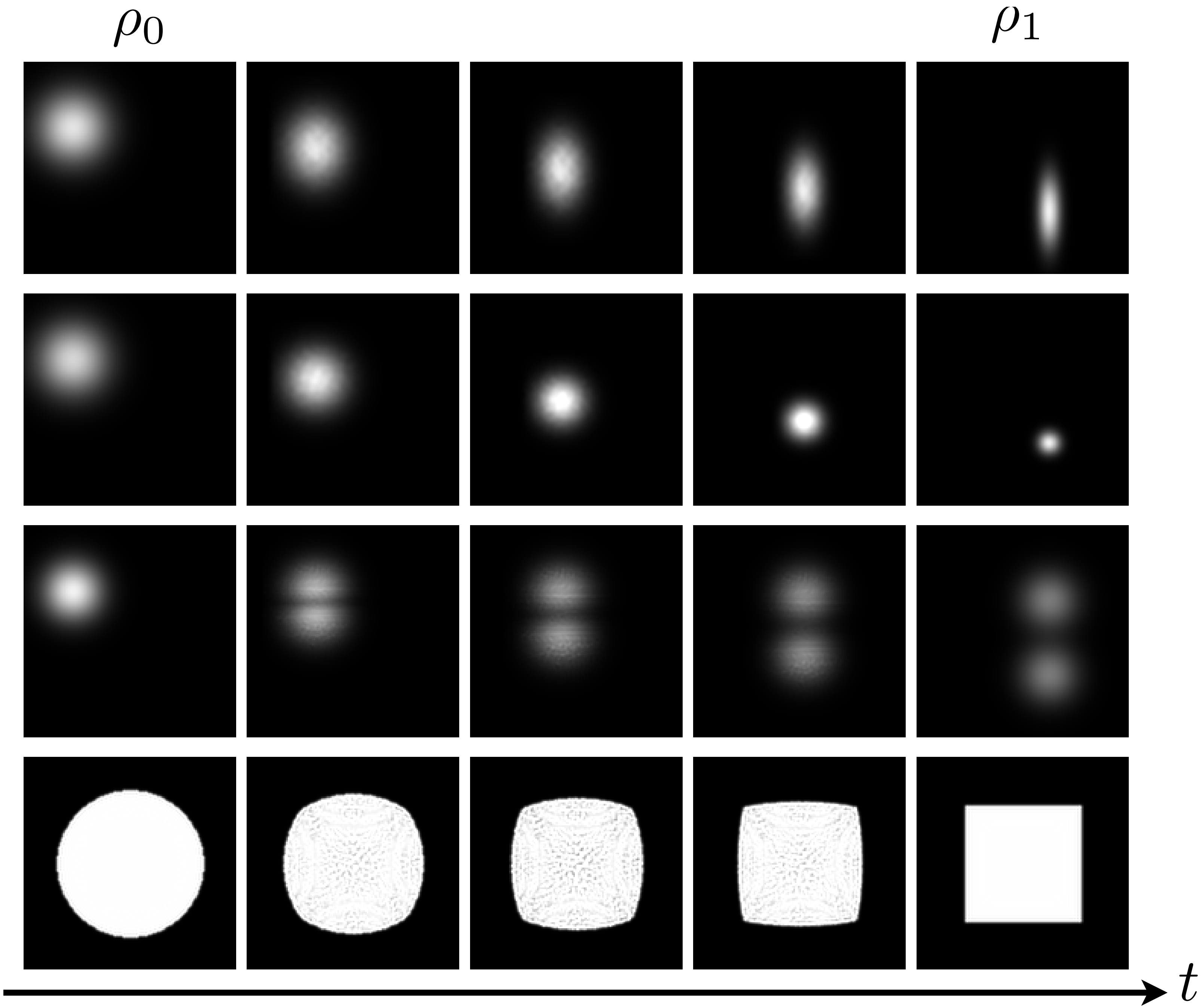
$$\mathcal{B}(\rho) = (\rho(0,\cdot),\rho(1,\cdot))$$

Numerical Examples

ρ_0 ρ_1

t

Numerical Examples

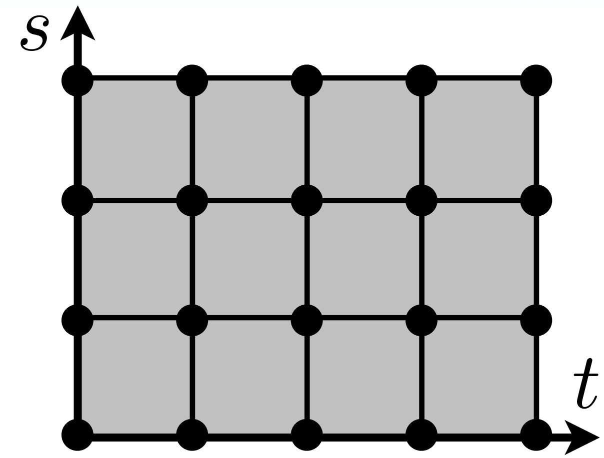


Discrete Formulation

Centered grid formulation ($d = 1$):

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$J(x) = \sum_{i \in \mathcal{G}_c} j(x_i)$$



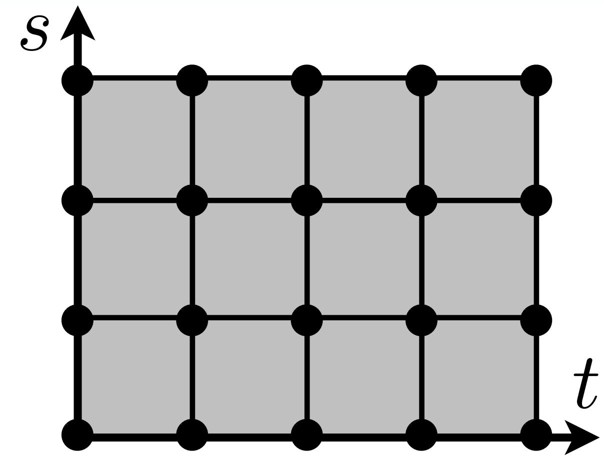
Centered grid $\mathcal{G}_c$

Discrete Formulation

Centered grid formulation ($d = 1$):

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

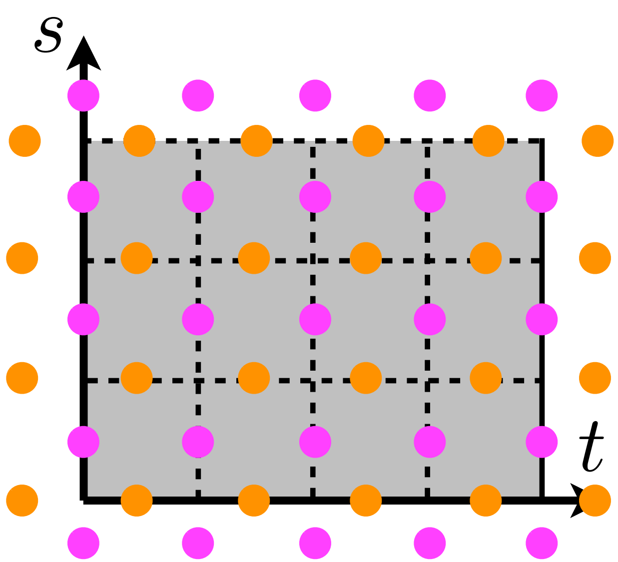
$$J(x) = \textstyle\sum_{i \in \mathcal{G}_c} j(x_i)$$



Centered grid $\mathcal{G}_c$

Staggered grid formulation :

$$\min_{x \in \mathbb{R}^{\mathcal{G}^1_{\text{st}}} \times \mathbb{R}^{\mathcal{G}^2_{\text{st}}}} J(\mathcal{I}(x)) + \iota_{\mathcal{C}}(x)$$



Staggered grid

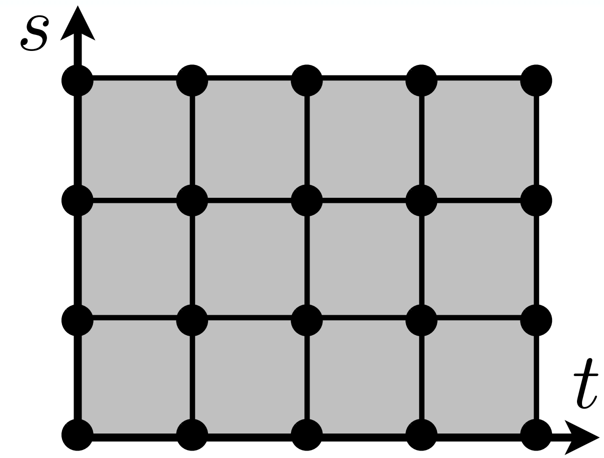
$\mathcal{G}^1_{\text{st}}$ $\mathcal{G}^2_{\text{st}}$

Discrete Formulation

Centered grid formulation ($d = 1$):

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} \; J(x) + \iota_{\mathcal{C}}(x)$$

$$J(x) = \sum_{i \in \mathcal{G}_c} j(x_i)$$



Centered grid $\mathcal{G}_c$

Staggered grid formulation :

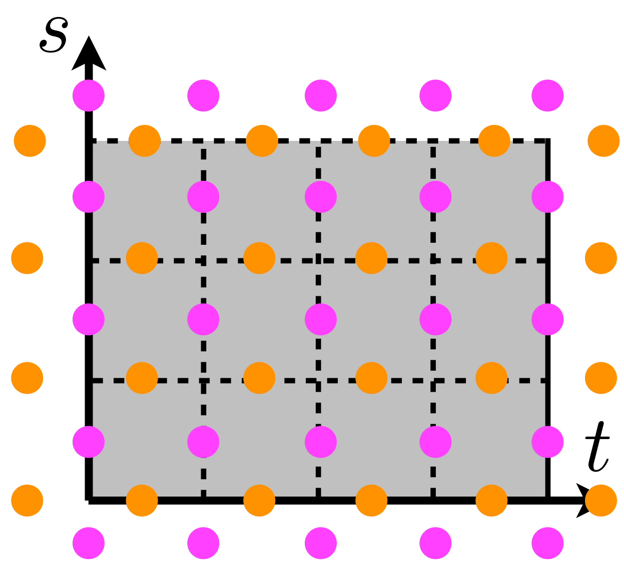
$$\min_{x \in \mathbb{R}^{\mathcal{G}^1_{\mathrm{st}}} \times \mathbb{R}^{\mathcal{G}^2_{\mathrm{st}}}} \; J(\mathcal{I}(x)) + \iota_{\mathcal{C}}(x)$$

Interpolation operator:

$$\mathcal{I} = (\mathcal{I}^1, \mathcal{I}^2) : \mathbb{R}^{\mathcal{G}^1_{\mathrm{st}}} \times \mathbb{R}^{\mathcal{G}^2_{\mathrm{st}}} \longrightarrow \mathbb{R}^{\mathcal{G}_c}$$

$$2\mathcal{I}_1(m)_{i,j} = m_{i+\frac{1}{2},j} + m_{i-\frac{1}{2},j}$$

$\longrightarrow$ Projection on $\mathrm{div}(x) = 0$ using FFTs.



Staggered grid

$\mathcal{G}^1_{\mathrm{st}}$ $\mathcal{G}^2_{\mathrm{st}}$

SOCP Formulation

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times d}} \; J(x) + \iota_{\mathcal{C}}(x) \qquad J(x) = \sum_{i \in \mathcal{G}_c} j(x_i)$$

$$\Longleftrightarrow \quad \min_{x \in \mathbb{R}^{\mathcal{G}_c \times d}, r \in \mathbb{R}^{\mathcal{G}_c}} \; \sum_i r_i \qquad \text{s.t.} \qquad \forall\, i \in \mathcal{G}_c, \; (m_i, \rho_i, r_i) \in \mathcal{K}$$

(Rotated) Lorentz cone: $\mathcal{K} = \left\{ (\tilde{m}, \tilde{\rho}, \tilde{r}) \in \mathbb{R}^{d+2} \setminus \|\tilde{m}\|^2 \leqslant \tilde{\rho}\tilde{r} \right\}$

SOCP Formulation

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times d}} J(x) + \iota_{\mathcal{C}}(x) \qquad J(x) = \sum_{i \in \mathcal{G}_c} j(x_i)$$

$$\Longleftrightarrow \min_{x \in \mathbb{R}^{\mathcal{G}_c \times d}, r \in \mathbb{R}^{\mathcal{G}_c}} \sum_i r_i \quad \text{s.t.} \quad \forall\, i \in \mathcal{G}_c,\ (m_i, \rho_i, r_i) \in \mathcal{K}$$

(Rotated) Lorentz cone: $\mathcal{K} = \left\{ (\tilde{m}, \tilde{\rho}, \tilde{r}) \in \mathbb{R}^{d+2} \setminus \|\tilde{m}\|^2 \leqslant \tilde{\rho}\tilde{r} \right\}$

Second order cone program:

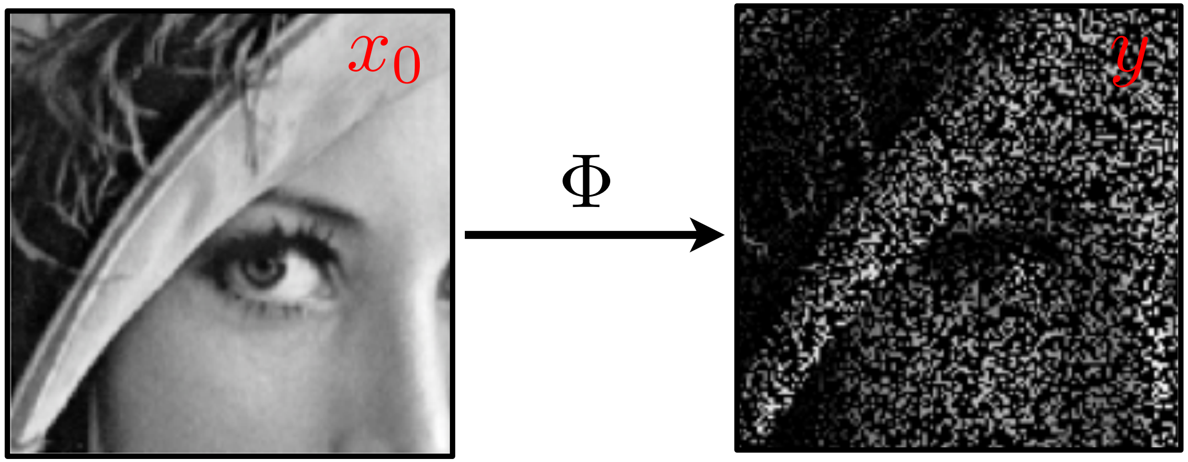
$\longrightarrow$ Use interior point methods (e.g. MOSEK software).

- Linear convergence with iteration #.
- Poor scaling with dimension $|\mathcal{G}_c|$.

Efficient for medium scale problems ($N \sim 10^4$).

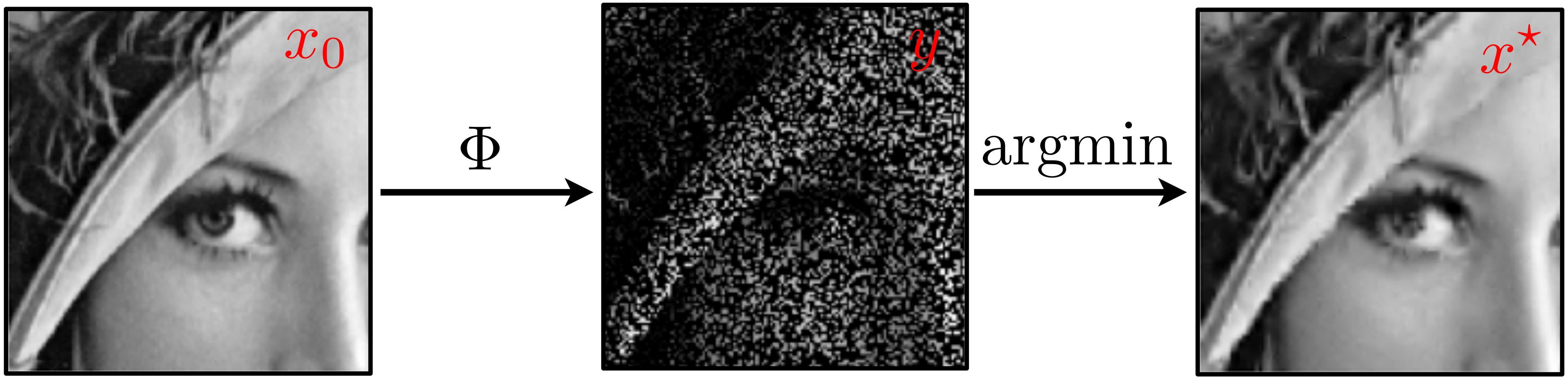
Example: ℓ^1 Regularization

Inverse problem: measurements $y = \Phi x_0 + w$



Example: ℓ^1 Regularization

Inverse problem: measurements $y = \Phi x_0 + w$

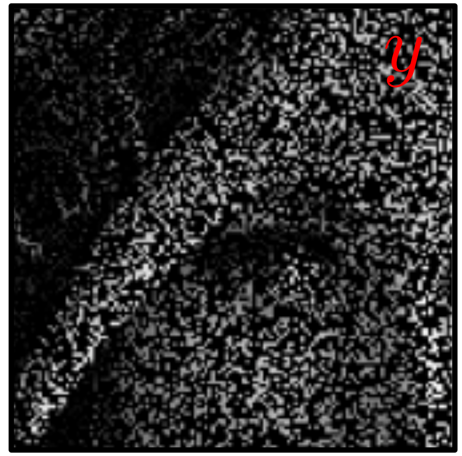


Regularized inversion: $$x^\star \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \underbrace{\frac{1}{2}\|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{R(x)}_{\text{Regularity}}$$

Example: ℓ^1 Regularization

Inverse problem: measurements $y = \Phi x_0 + w$

$\xrightarrow{\Phi}$



$\xrightarrow{\text{argmin}}$

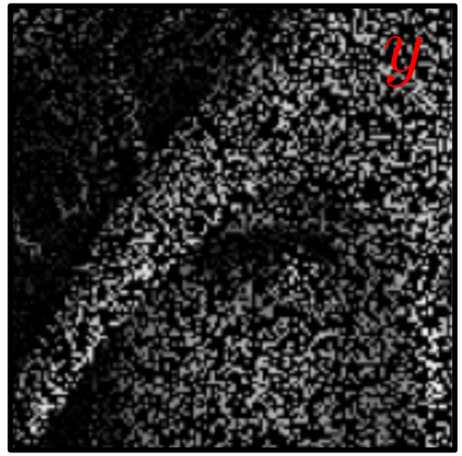
Regularized inversion: $$x^\star \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \ \underbrace{\frac{1}{2}\|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{R(x)}_{\text{Regularity}}$$

Total Variation: $R(x) = \sum_i \|(\nabla x)_i\|$

Example: ℓ^1 Regularization

Inverse problem: measurements $y = \Phi x_0 + w$

$\xrightarrow{\Phi}$



$\xrightarrow{\text{argmin}}$

Regularized inversion: $$x^\star \in \underset{x \in \mathbb{R}^N}{\text{argmin}} \; \underbrace{\frac{1}{2}\|y - \Phi x\|^2}_{\text{Data fidelity}} + \lambda \underbrace{R(x)}_{\text{Regularity}}$$

Total Variation: $R(x) = \sum_i \|(\nabla x)_i\|$

ℓ^1 sparsity: $R(x) = \sum_i |x_i|$

Images are sparse in wavelet bases.

$\xrightarrow{\Psi^*}$ $\xleftarrow{\Psi}$

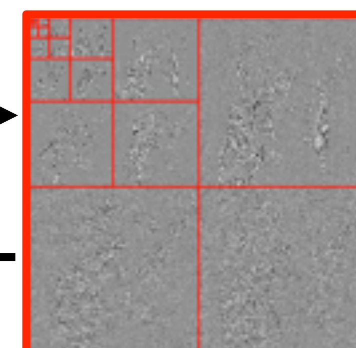


Image $f = \Psi x$ Coeff. $x = \Psi^* f$

Overview

- Optimal Transport and Imaging
- *Convex Analysis and Proximal Calculus*
- Forward Backward
- Douglas Rachford and ADMM
- Generalized Forward-Backward
- Primal-Dual Schemes

Convex Optimization

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

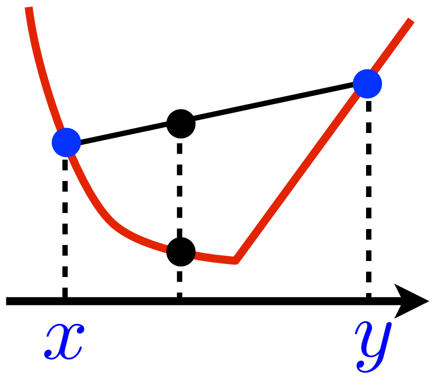
$\mathcal{H}$: Hilbert space. Here: $\mathcal{H} = \mathbb{R}^N$.

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

Convex Optimization

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

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



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

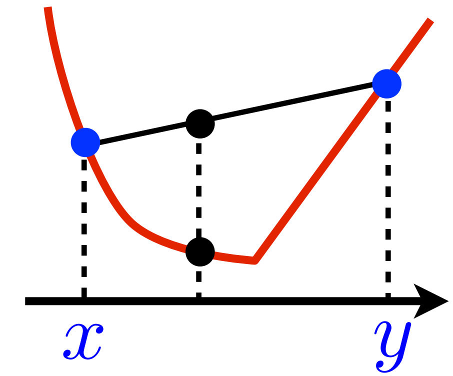
Class of functions:

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

Convex Optimization

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

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

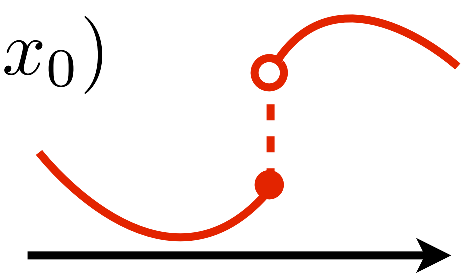


$$\boxed{\textit{Problem:} \quad \min_{x \in \mathcal{H}} G(x)}$$

Class of functions:

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

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



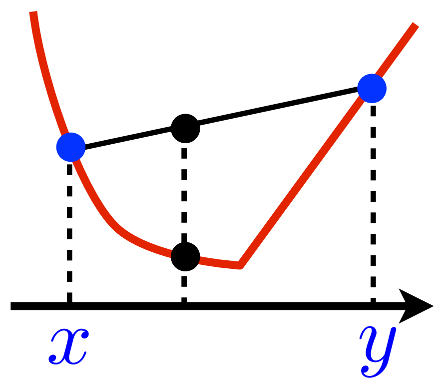
Proper: $\{x \in \mathcal{H} \setminus G(x) \neq +\infty\} \neq \emptyset$

Convex Optimization

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

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

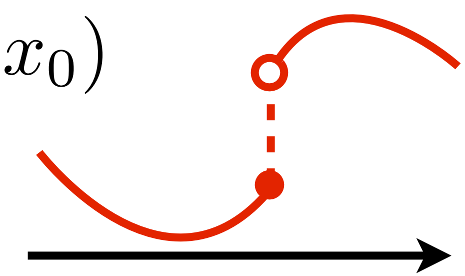
$$\textit{Problem:} \quad \min_{x \in \mathcal{H}} G(x)$$



Class of functions:

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

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



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

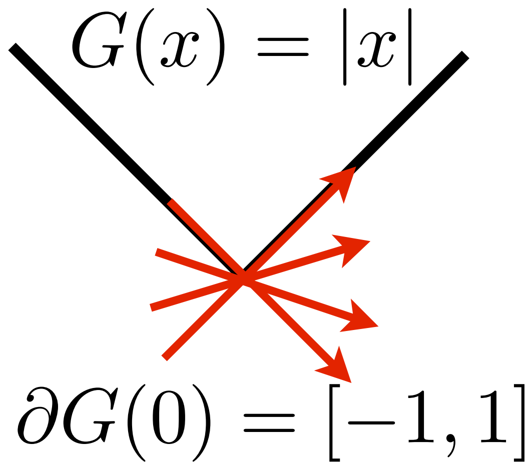
Indicator: $$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if} \quad x \in \mathcal{C}, \\ +\infty & \text{otherwise.} \end{cases}$$

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

Sub-differential

Sub-differential:

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



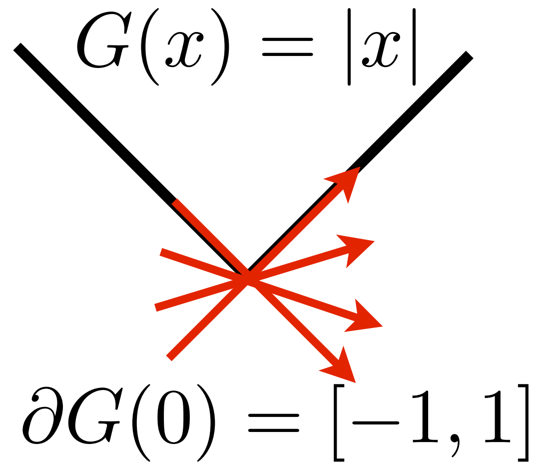
Sub-differential

Sub-differential:

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

Smooth functions:

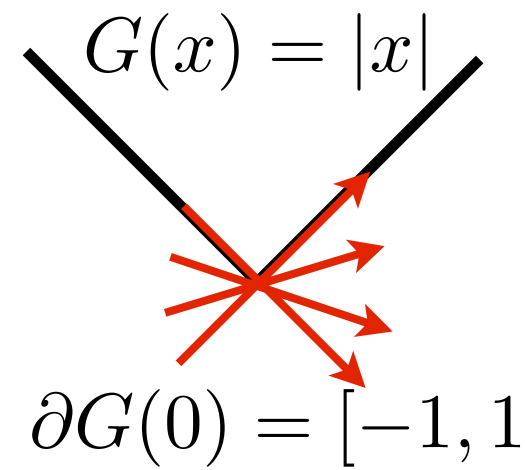
If F is C^1, $\partial F(x) = \{\nabla F(x)\}$



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:

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

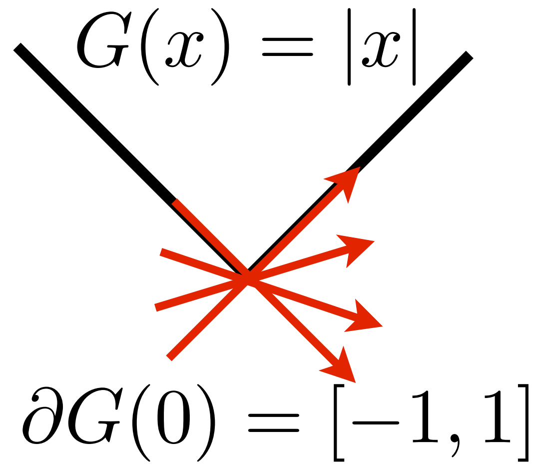
Sub-differential

Sub-differential:

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

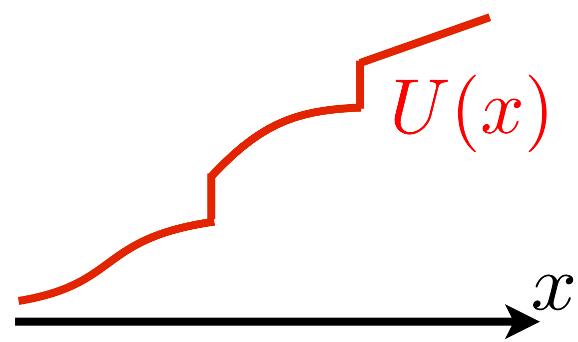
Smooth functions:

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



First-order conditions:

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



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

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

Prox and Subdifferential

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

Prox and Subdifferential

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

Resolvant of ∂G*:*

$$z = \mathrm{Prox}_{\gamma G}(x) \quad \Longleftrightarrow \quad 0 \in z - x + \gamma \partial G(z)$$

$$\Longleftrightarrow \quad x \in (\mathrm{Id} + \gamma \partial G)(z)$$

Prox and Subdifferential

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

Resolvant of ∂G:

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

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

Inverse of a set-valued mapping:

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

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

Prox and Subdifferential

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

Resolvant of ∂G:

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

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

Inverse of a set-valued mapping:

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

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

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

$$\iff 0 \in \partial G(x^\star) \iff x^\star \in (\mathrm{Id} + \gamma \partial G)(x^\star)$$

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

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

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

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

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

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

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

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

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

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

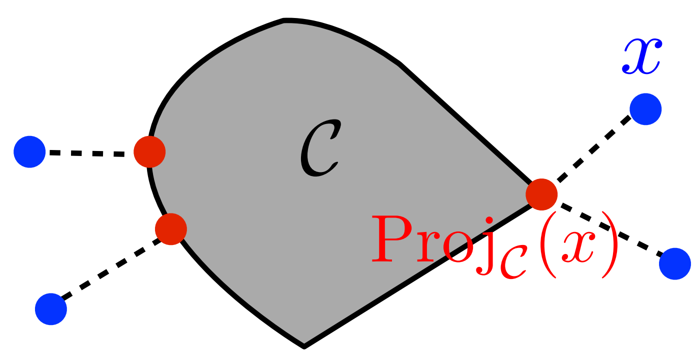
Composition by tight frame: $A \circ A^* = \mathrm{Id}$

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

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

$$\mathrm{Prox}_{\gamma G}(x) = \mathrm{Proj}_{\mathcal{C}}(x)$$

$$= \underset{z \in \mathcal{C}}{\mathrm{argmin}} \|x - z\|$$



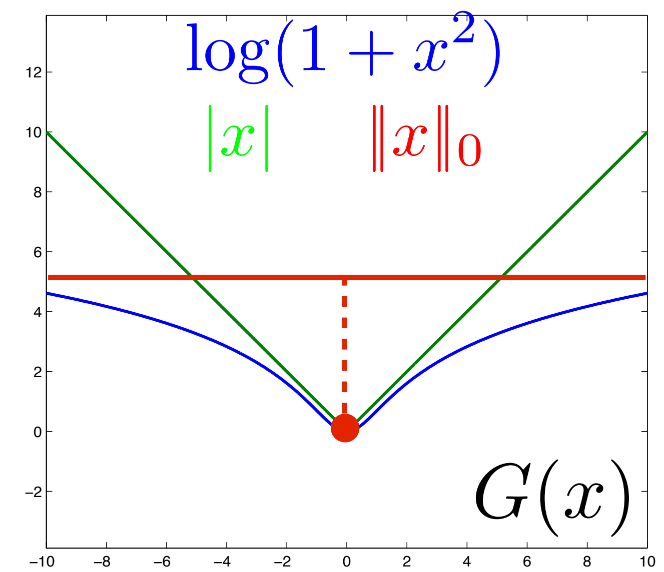
Prox of Sparse Regularizers

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

Prox of Sparse Regularizers

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

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



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

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

Prox of Sparse Regularizers

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

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

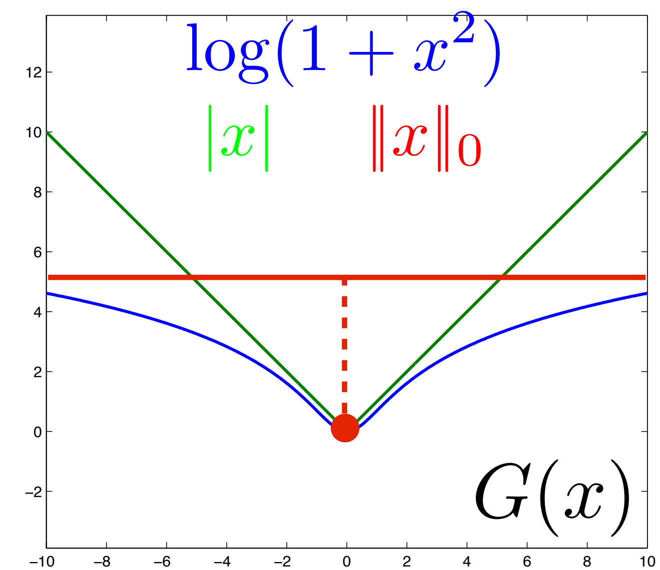
$$\mathrm{Prox}_{\gamma G}(x)_i = \max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i$$

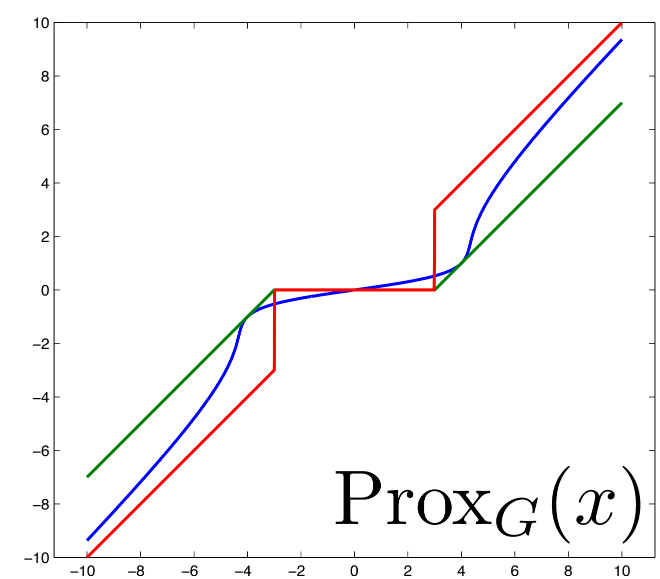
$$G(x) = \|x\|_0 = |\,\{i \setminus x_i \neq 0\}\,|$$

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

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

$\longrightarrow$ 3rd order polynomial root.

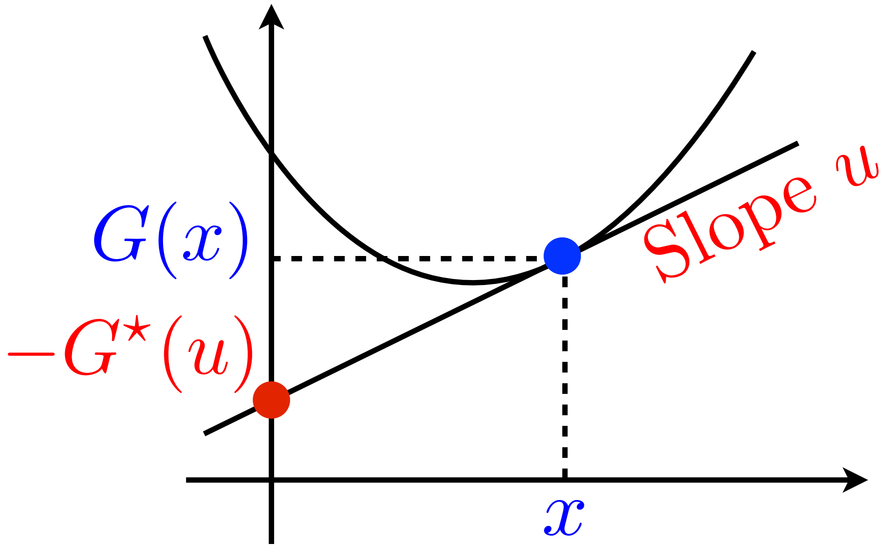




Legendre-Fenchel Duality

Legendre-Fenchel transform:

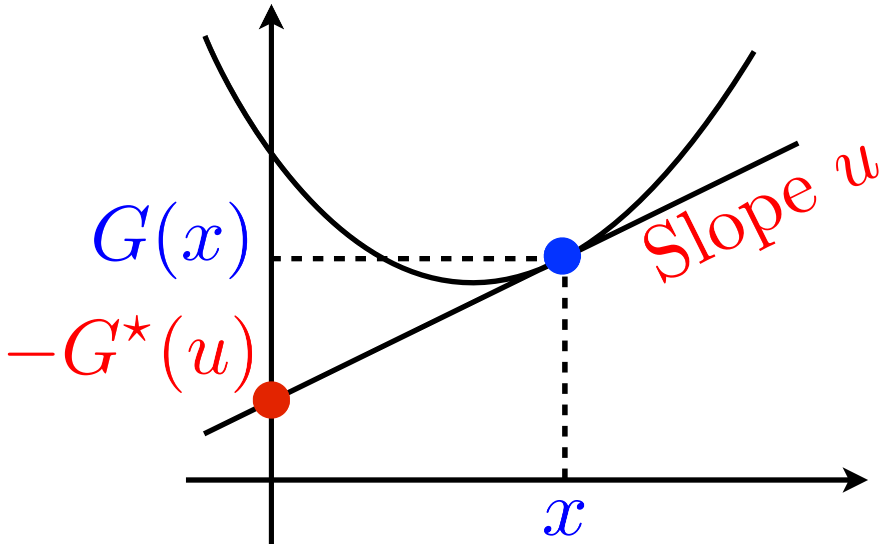
$$G^\star(u) = \sup_{x \in \mathrm{dom}(G)} \langle u, x \rangle - G(x)$$



Legendre-Fenchel Duality

Legendre-Fenchel transform:

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



Example: quadratic functional

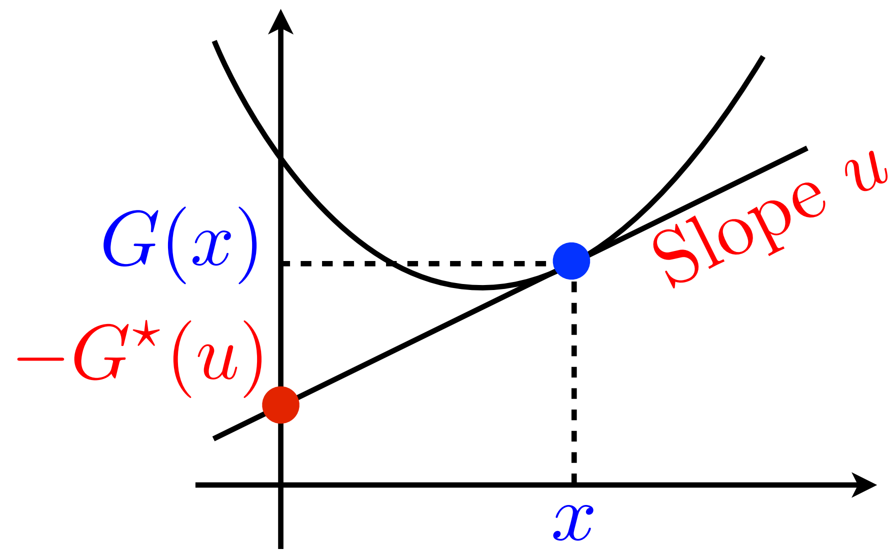
$$G(x) = \frac{1}{2}\langle Ax, x \rangle + \langle x, b \rangle$$

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

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 Functionals

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

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

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

Indicator and Homogeneous Functionals

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

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

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

ℓ^p *norms:* $G(x) = \|x\|_p$

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

Indicator and Homogeneous Functionals

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

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

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

ℓ^p *norms:* $G(x) = \|x\|_p$

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

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 $\tau = \tau(x, \gamma)$

Prox of the J Functional

$$J(m,\rho) = \sum_i j(m_i,\rho_i) \qquad j(\tilde{m},\tilde{\rho}) = \frac{\|\tilde{m}\|^2}{\tilde{\rho}} \quad \text{for } \tilde{\rho} > 0$$

Prox of the J Functional

$$J(m,\rho) = \sum_i j(m_i,\rho_i) \qquad j(\tilde m,\tilde\rho) = \frac{\|\tilde m\|^2}{\tilde\rho} \quad \text{for } \tilde\rho > 0$$

$$\mathrm{Prox}_{\gamma J}(m,\rho) = (\mathrm{Prox}_{\gamma j}(m_i,\rho_i))_i$$

Prox of the J Functional

$$J(m,\rho) = \sum_i j(m_i,\rho_i) \qquad j(\tilde{m},\tilde{\rho}) = \frac{\|\tilde{m}\|^2}{\tilde{\rho}} \quad \text{for } \tilde{\rho} > 0$$

$$\text{Prox}_{\gamma J}(m,\rho) = (\text{Prox}_{\gamma j}(m_i,\rho_i))_i$$

$$j^* = \iota_{\mathcal{C}} \quad \text{where} \quad \mathcal{C} = \left\{(a,b) \in \mathbb{R}^2 \times \mathbb{R} \setminus 2\|a\|^2 + b \leqslant 0\right\}$$

$$\text{Prox}_{\gamma j}(\tilde{x}) = \tilde{x} - \gamma \text{Proj}_{\mathcal{C}}(\tilde{x}/\gamma) \qquad \text{where} \quad \tilde{x} = (\tilde{m},\tilde{\rho})$$

Prox of the J Functional

$$J(m,\rho) = \sum_i j(m_i,\rho_i) \qquad j(\tilde m,\tilde\rho) = \frac{\|\tilde m\|^2}{\tilde\rho} \quad \text{for } \tilde\rho > 0$$

$$\mathrm{Prox}_{\gamma J}(m,\rho) = (\mathrm{Prox}_{\gamma j}(m_i,\rho_i))_i$$

$$j^* = \iota_{\mathcal{C}} \quad \text{where} \quad \mathcal{C} = \left\{(a,b) \in \mathbb{R}^2 \times \mathbb{R} \setminus 2\|a\|^2 + b \leqslant 0\right\}$$

$$\mathrm{Prox}_{\gamma j}(\tilde x) = \tilde x - \gamma \mathrm{Proj}_{\mathcal{C}}(\tilde x/\gamma) \qquad \text{where} \quad \tilde x = (\tilde m, \tilde\rho)$$

Proposition: $\mathrm{Prox}_{\gamma}(\tilde m,\tilde\rho) = \begin{cases} (m^\star,\rho^\star) & \text{if} \quad \rho^\star > 0 \\ (0,0) & \text{otherwise.} \end{cases}$

where $m^\star = \dfrac{\rho^\star \tilde m}{\rho^\star + 2\gamma}$ and $\rho^\star$ is the largest root of

$$X^3 + (4\gamma - \tilde\rho)X^2 + 4\gamma(\gamma - \tilde\rho)X - \gamma\|\tilde m\|^2 - 4\gamma^2\tilde\rho = 0$$

Overview

- Optimal Transport and Imaging
- Convex Analysis and Proximal Calculus
- ***Forward Backward***
- Douglas Rachford and ADMM
- Generalized Forward-Backward
- Primal-Dual Schemes

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. [Rockafellar, 70]

Proximal Splitting Methods

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

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

Proximal Splitting Methods

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

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

Splitting: $E(x) = \underbrace{\boxed{F(x)}}_{\text{Smooth}} + \sum_i \underbrace{\boxed{G_i(x)}}_{\text{Simple}}$

Proximal Splitting Methods

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

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

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

Smooth Simple

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

Forward-Backward: $\xrightarrow{\text{solves}}$ $F + G$

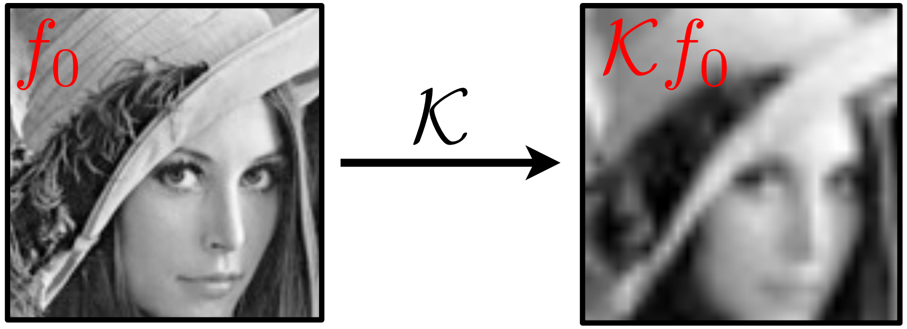
Douglas-Rachford: $\longrightarrow$ $\sum G_i$

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

$$\min_{x \in \mathbb{R}^N} \boxed{F(x)} + \boxed{G(x)}$$

Smooth 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:

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

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

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

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:

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

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

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

Forward-Backward

Fix point equation:

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

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

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

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

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 $(\star)$

[Passty 79, Gabay, 83]

Example: L1 Regularization

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

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

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

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

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

$$\boxed{\text{Forward-backward} \quad \Longleftrightarrow \quad \text{Iterative soft thresholding}}$$

Convergence Speed

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

∇F is L-Lipschitz.

G is simple.

$$\boxed{\textit{Theorem:} \quad \text{If } L > 0, \text{ FB iterates } x^{(\ell)} \text{ satisfies} \quad E(x^{(\ell)}) - E(x^\star) \leqslant C/\sqrt{\ell}}$$

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 $L > 0$, $\quad E(x^{(\ell)}) - E(x^\star) \leqslant \frac{C}{\ell}$

Complexity theory: optimal in a worse-case sense.

Overview

- Optimal Transport and Imaging
- Convex Analysis and Proximal Calculus
- Forward Backward
- *Douglas Rachford and ADMM*
- Generalized Forward-Backward
- Primal-Dual Schemes

Douglas Rachford Scheme

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

Simple Simple

Douglas-Rachford iterations:

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

$$x^{(\ell+1)} = \mathrm{Prox}_{\gamma G_2}(z^{(\ell+1)})$$

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:

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

$$x^{(\ell+1)} = \mathrm{Prox}_{\gamma G_2}(z^{(\ell+1)})$$

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

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

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

[Lions, Mercier, 79]

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

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

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

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

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

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

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

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

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

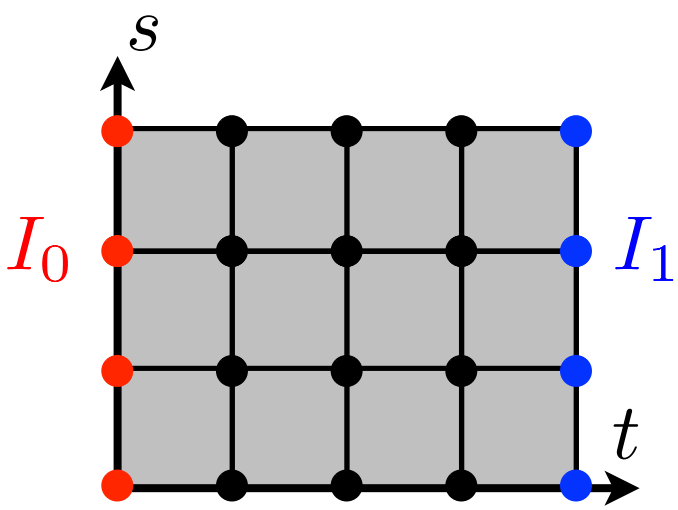
Example: Optimal Transport on Centered Grid

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{C} = \{x = (m, \rho) \setminus Ax = b\}$$

$$b = (0, \rho_0, \rho_1)$$

$$A(x) = (\mathrm{div}(x), \rho_{I_0}, \rho_{I_1})$$



Centered grid $\mathcal{G}_c$

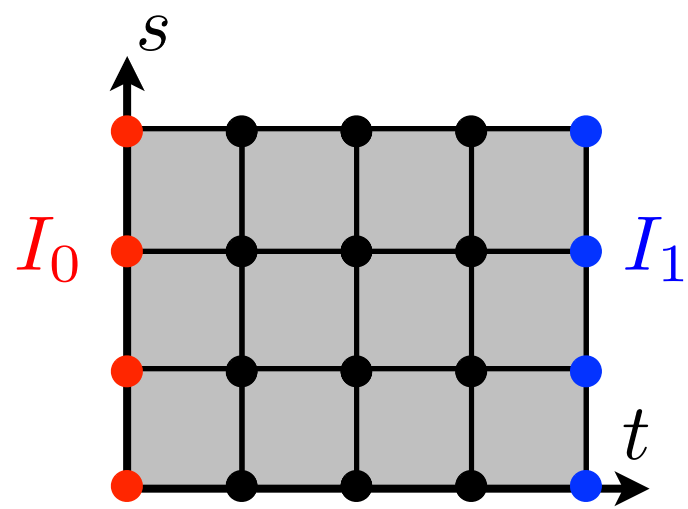
Example: Optimal Transport on Centered Grid

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{C} = \{x = (m, \rho) \setminus Ax = b\}$$

$$b = (0, \rho_0, \rho_1)$$

$$A(x) = (\mathrm{div}(x), \rho_{I_0}, \rho_{I_1})$$



Centered grid $\mathcal{G}_c$

$\mathrm{Prox}_{\gamma J}$: cubic root (closed form).

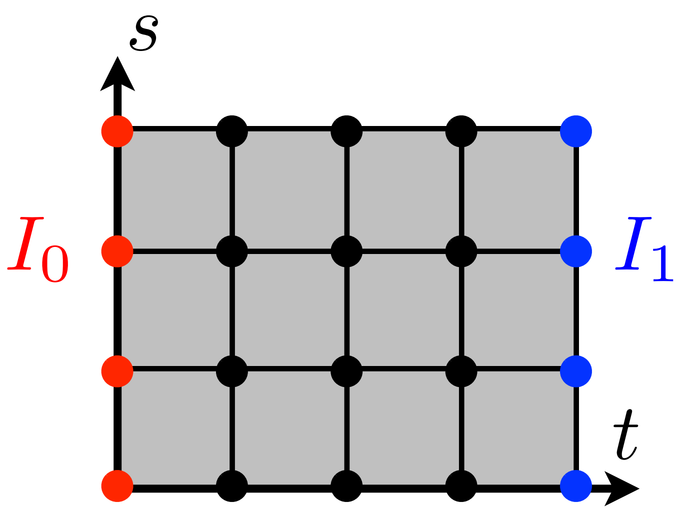
Example: Optimal Transport on Centered Grid

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{C} = \{x = (m, \rho) \setminus Ax = b\}$$

$$b = (0, \rho_0, \rho_1)$$

$$A(x) = (\mathrm{div}(x), \rho_{I_0}, \rho_{I_1})$$



Centered grid $\mathcal{G}_c$

$\mathrm{Prox}_{\gamma J}$: cubic root (closed form).

$$\mathrm{Prox}_{\iota_{\mathcal{C}}} = \mathrm{Proj}_{\mathcal{C}} = (\mathrm{Id} - A^* \Delta^{-1} A) + A^* \Delta^{-1} y$$

$\Delta^{-1} = (AA^*)^{-1}$: solving a Poisson equation with b.c.

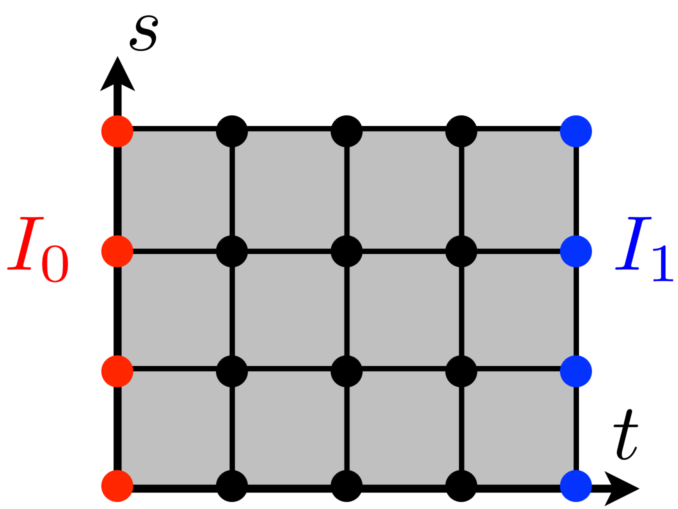
Example: Optimal Transport on Centered Grid

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{C} = \{x = (m, \rho) \setminus Ax = b\}$$

$$b = (0, \rho_0, \rho_1)$$

$$A(x) = (\mathrm{div}(x), \rho_{I_0}, \rho_{I_1})$$



Centered grid $\mathcal{G}_c$

$\mathrm{Prox}_{\gamma J}$: cubic root (closed form).

$$\mathrm{Prox}_{\iota_{\mathcal{C}}} = \mathrm{Proj}_{\mathcal{C}} = (\mathrm{Id} - A^*\Delta^{-1}A) + A^*\Delta^{-1}y$$

$\Delta^{-1} = (AA^*)^{-1}$: solving a Poisson equation with b.c.

Proposition: DR($\alpha = 1$) is ALG2 of [Benamou, Brenier 2000]

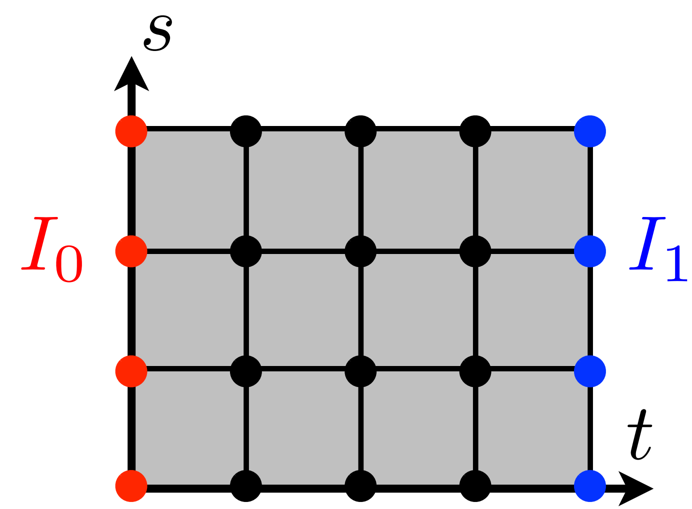
Example: Optimal Transport on Centered Grid

$$\min_{x \in \mathbb{R}^{\mathcal{G}_c \times 2}} J(x) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{C} = \{x = (m, \rho) \setminus Ax = b\}$$

$$b = (0, \rho_0, \rho_1)$$

$$A(x) = (\mathrm{div}(x), \rho_{I_0}, \rho_{I_1})$$



Centered grid $\mathcal{G}_c$

$\mathrm{Prox}_{\gamma J}$: cubic root (closed form).

$\mathrm{Prox}_{\iota_{\mathcal{C}}} = \mathrm{Proj}_{\mathcal{C}} = (\mathrm{Id} - A^* \Delta^{-1} A) + A^* \Delta^{-1} y$

$\Delta^{-1} = (AA^*)^{-1}$: solving a Poisson equation with b.c.

Proposition: DR($\alpha = 1$) is ALG2 of [Benamou, Brenier 2000]

$\longrightarrow$ *Advantage:* relaxation parameter $\alpha \in]0, 1[$.

Example: Constrained L1

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

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

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

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

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

Example: Constrained L1

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

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

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

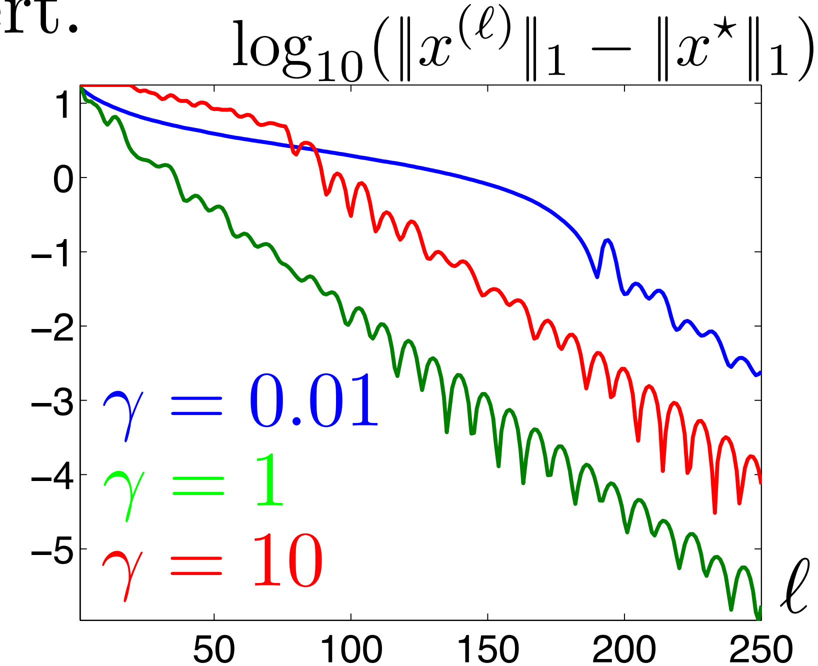
$$G_2(x) = \|x\|_1 \qquad \mathrm{Prox}_{\gamma G_2}(x) = \left(\max\left(0, 1 - \frac{\gamma}{|x_i|}\right) x_i \right)_i$$

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

Example: compressed sensing

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

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



Auxiliary Variables with 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 with DR

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

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

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

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

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

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

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

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

Example: TV Regularization

$$\min_f \frac{1}{2}\|\mathcal{K}f - y\|^2 + \lambda\|\nabla f\|_1$$

$$\|u\|_1 = \sum_i \|u_i\|$$

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

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

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

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

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

Example: TV Regularization

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

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

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

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

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

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

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

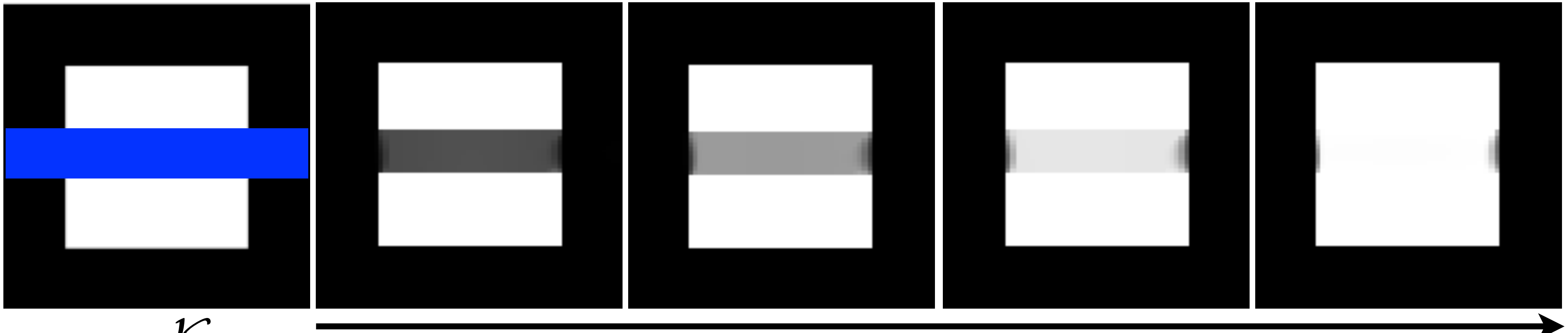
$\longrightarrow O(N\log(N))$ operations using FFT.

Example: TV Regularization

Orignal f_0

$y = \Phi f_0 + w$

Recovery $f^\star$



$y = \mathcal{K} x_0$

Iteration ℓ

Alternating Direction Method of Multipliers

$$\min_x F(x) + G \circ A(x) \quad (\star) \qquad \Longleftrightarrow \qquad \min_{x, y = Ax} F(x) + G(y)$$

$$A : \mathbb{R}^N \to \mathbb{R}^P \text{ injective.}$$

Alternating Direction Method of Multipliers

$$\min_x F(x) + G \circ A(x) \quad (\star) \qquad \Longleftrightarrow \qquad \min_{x, y = Ax} F(x) + G(y)$$

$$A : \mathbb{R}^N \to \mathbb{R}^P \text{ injective.}$$

Lagrangian: $\min_{x,y} \max_u L(x, y, u) = F(x) + G(y) + \langle u,\ y - Ax \rangle$

Alternating Direction Method of Multipliers

$$\min_x F(x) + G \circ A(x) \quad (\star) \qquad \Longleftrightarrow \qquad \min_{x,y=Ax} F(x) + G(y)$$

$$A : \mathbb{R}^N \to \mathbb{R}^P \text{ injective.}$$

Lagrangian: $\min_{x,y} \max_u L(x, y, u) = F(x) + G(y) + \langle u,\ y - Ax\rangle$

Augmented: $\min_{x,y} \max_u L_\gamma(x, y, u) = L(x, y, u) + \frac{\gamma}{2}\|y - Ax\|^2$

Alternating Direction Method of Multipliers

$$\min_x F(x) + G \circ A(x) \quad (\star) \qquad \Longleftrightarrow \qquad \min_{x, y = Ax} F(x) + G(y)$$

$A : \mathbb{R}^N \to \mathbb{R}^P$ injective.

Lagrangian: $\min_{x,y} \max_u L(x, y, u) = F(x) + G(y) + \langle u,\ y - Ax \rangle$

Augmented: $\min_{x,y} \max_u L_\gamma(x, y, u) = L(x, y, u) + \frac{\gamma}{2} \|y - Ax\|^2$

ADMM

$$\left| \begin{aligned} x^{(\ell+1)} &= \text{argmin}_x L_\gamma(x, y^{(\ell)}, u^{(\ell)}) \\ y^{(\ell+1)} &= \text{argmin}_y L_\gamma(x^{(\ell+1)}, y, u^{(\ell)}) \\ u^{(\ell+1)} &= u^{(\ell)} + (y^{(\ell+1)} - Ax^{(\ell+1)}) \end{aligned} \right.$$

Alternating Direction Method of Multipliers

$$\min_x F(x) + G \circ A(x) \quad (\star) \qquad \Longleftrightarrow \qquad \min_{x, y = Ax} F(x) + G(y)$$

$A : \mathbb{R}^N \rightarrow \mathbb{R}^P$ injective.

Lagrangian: $\min_{x,y} \max_u L(x,y,u) = F(x) + G(y) + \langle u,\ y - Ax \rangle$

Augmented: $\min_{x,y} \max_u L_\gamma(x,y,u) = L(x,y,u) + \frac{\gamma}{2}\|y - Ax\|^2$

ADMM

$$\left| \begin{array}{l} x^{(\ell+1)} = \mathrm{argmin}_x L_\gamma(x, y^{(\ell)}, u^{(\ell)}) \\ y^{(\ell+1)} = \mathrm{argmin}_y L_\gamma(x^{(\ell+1)}, y, u^{(\ell)}) \\ u^{(\ell+1)} = u^{(\ell)} + (y^{(\ell+1)} - Ax^{(\ell+1)}) \end{array} \right.$$

Theorem: If $\gamma > 0$, $x^{(\ell)} \rightarrow x^\star$ a solution of $(\star)$

[Gabay, Mercier, Glowinski, Marrocco, 76]

ADMM with Proximal Operators

Proximal mapping for metric A: (A is injective)

$$\mathrm{Prox}_{\gamma F}^{A} = \underset{x}{\operatorname{argmin}} \; \frac{1}{2}\|Ax - z\|^2 + \gamma F(x)$$

ADMM with Proximal Operators

Proximal mapping for metric A: (A is injective)

$$\mathrm{Prox}^{A}_{\gamma F} = \operatorname*{argmin}_{x} \frac{1}{2}\|Ax - z\|^2 + \gamma F(x)$$

Proposition: $\quad \mathrm{Prox}^{A}_{\gamma F} = A^{+}\left(\mathrm{Id} - \gamma \mathrm{Prox}_{F^* \circ A^*/\gamma}(\cdot/\gamma)\right)$

ADMM with Proximal Operators

Proximal mapping for metric A: (A is injective)

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

Proposition: $\mathrm{Prox}^A_{\gamma F} = A^+ \left(\mathrm{Id} - \gamma \mathrm{Prox}_{F^* \circ A^*/\gamma}(\cdot/\gamma)\right)$

ADMM

$$\begin{aligned} x^{(\ell+1)} &= \mathrm{Prox}^A_{F/\gamma}(y^{(\ell)} - u^{(\ell)}) \\ y^{(\ell+1)} &= \mathrm{Prox}_{G/\gamma}(Ax^{(\ell+1)} + u^{(\ell)}) \\ u^{(\ell+1)} &= u^{(\ell)} + (y^{(\ell+1)} - Ax^{(\ell+1)}) \end{aligned}$$

ADMM with Proximal Operators

Proximal mapping for metric A: (A is injective)

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

Proposition: $\mathrm{Prox}^{A}_{\gamma F} = A^{+}\left(\mathrm{Id} - \gamma \mathrm{Prox}_{F^* \circ A^*/\gamma}(\cdot/\gamma)\right)$

ADMM

$$\left|\begin{aligned} x^{(\ell+1)} &= \mathrm{Prox}^{A}_{F/\gamma}(y^{(\ell)} - u^{(\ell)}) \\ y^{(\ell+1)} &= \mathrm{Prox}_{G/\gamma}(Ax^{(\ell+1)} + u^{(\ell)}) \\ u^{(\ell+1)} &= u^{(\ell)} + (y^{(\ell+1)} - Ax^{(\ell+1)}) \end{aligned}\right.$$

$\longrightarrow$ If $G \circ A$ is simple: use DR.

$\longrightarrow$ If $F^* \circ A^*$ is simple: use ADMM.

ADMM vs. DR

Fenchel-Rockafellar duality:

$$\min_x F(x) + G \circ A(x) \quad \longleftrightarrow \quad \min_u F^*(-A^* u) + G^*(u)$$

Important: no bijection between u and x.

ADMM vs. DR

Fenchel-Rockafellar duality:

$$\min_x F(x) + G \circ A(x) \quad \longleftrightarrow \quad \min_u F^*(-A^* u) + G^*(u)$$

Important: no bijection between u and x.

Proposition: DR applied to $F^* \circ -A^* + G^*$ is ADMM.

[Eckstein, Bertsekas, 92]

ADMM vs. DR

Fenchel-Rockafellar duality:

$$\min_x F(x) + G \circ A(x) \quad \longleftrightarrow \quad \min_u F^*(-A^* u) + G^*(u)$$

Important: no bijection between u and x.

> *Proposition:* DR applied to $F^* \circ -A^* + G^*$ is ADMM.

[Eckstein, Bertsekas, 92]

DR iterations (when $\alpha = 1$):

$$z^{(\ell+1)} = \frac{1}{2} z^{(\ell)} + \frac{1}{2} \mathrm{RProx}_{\gamma F^* \circ -A^*} \circ \mathrm{RProx}_{\gamma G^*}(z^{(\ell)})$$

ADMM vs. DR

Fenchel-Rockafellar duality:

$$\min_x F(x) + G \circ A(x) \quad \longleftrightarrow \quad \min_u F^*(-A^* u) + G^*(u)$$

Important: no bijection between u and x.

Proposition: DR applied to $F^* \circ -A^* + G^*$ is ADMM.

[Eckstein, Bertsekas, 92]

DR iterations (when $\alpha = 1$):

$$z^{(\ell+1)} = \frac{1}{2} z^{(\ell)} + \frac{1}{2} \mathrm{RProx}_{\gamma F^* \circ -A^*} \circ \mathrm{RProx}_{\gamma G^*}(z^{(\ell)})$$

The iterates of ADMM are recovered using:

$$x^{(\ell+1)} = \mathrm{Prox}^A_{F/\gamma}(y^{(\ell)} - u^{(\ell)})$$

$$y^{(\ell)} = \frac{1}{\gamma}(z^{(\ell)} - u^{(\ell)})$$

$$u^{(\ell)} = \mathrm{Prox}_{\gamma G^*}(z^{(\ell)})$$

More than 2 Functionals

$$\min_x G_1(x) + \ldots + G_k(x)$$ each F_i is simple

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

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

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

More than 2 Functionals

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

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

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

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

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

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

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

Overview

- Optimal Transport and Imaging
- Convex Analysis and Proximal Calculus
- Forward Backward
- Douglas Rachford and ADMM
- *Generalized Forward-Backward*
- Primal-Dual Schemes

GFB Splitting

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

Smooth Simple

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

$$\boxed{\begin{aligned} z_i^{(\ell+1)} &= z_i^{(\ell)} + \mathrm{Prox}_{n\gamma G_\ell}(2x^{(\ell)} - z_i^{(\ell)} - \gamma\nabla F(x^{(\ell)})) - x^{(\ell)} \\ x^{(\ell+1)} &= \frac{1}{n}\sum_{i=1}^{n} z_i^{(\ell+1)} \end{aligned}}$$

[Raguet, Fadili, Peyré 2012]

GFB Splitting

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

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 $(\star)$

[Raguet, Fadili, Peyré 2012]

GFB Splitting

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

Smooth Simple

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

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

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

Theorem: Let ∇F be L-Lipschitz.

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

[Raguet, Fadili, Peyré 2012]

$n = 1 \longrightarrow$ Forward-backward.

$F = 0 \longrightarrow$ Douglas-Rachford.

GFB Fix Point

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

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

GFB Fix Point

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

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

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

$$\boxed{x^\star = \tfrac{1}{n} \textstyle\sum_i z_i} \qquad (\text{use } z_i = x^\star - \gamma \nabla F(x^\star) - N y_i)$$

GFB Fix Point

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

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

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

$$\boxed{x^\star = \tfrac{1}{n} \textstyle\sum_i z_i} \qquad (\text{use } z_i = x^\star - \gamma\nabla F(x^\star) - N y_i)$$

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

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

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

GFB Fix Point

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

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

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

$$x^\star = \tfrac{1}{n} \textstyle\sum_i z_i \qquad (\text{use } z_i = x^\star - \gamma \nabla F(x^\star) - N y_i)$$

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

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

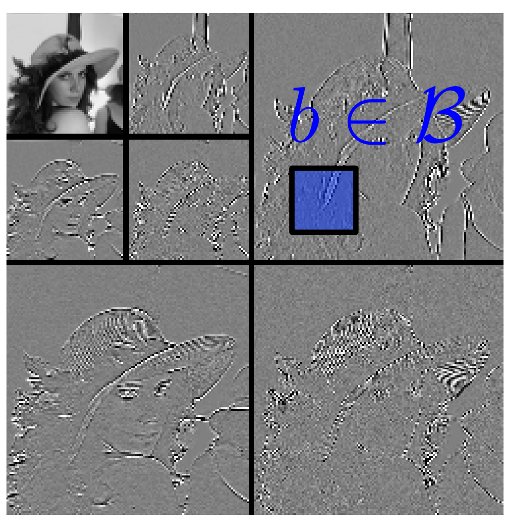
$$\iff \quad z_i = z_i + \operatorname{Prox}_{n\gamma G_\ell}(2x^\star - z_i - \gamma \nabla F(x^\star)) - x^\star$$

$$+ \quad \longrightarrow \text{ Fix point equation on } (x^\star, z_1, \dots, z_n).$$

Block Regularization

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

Image $f = \Psi x$



Coefficients x.

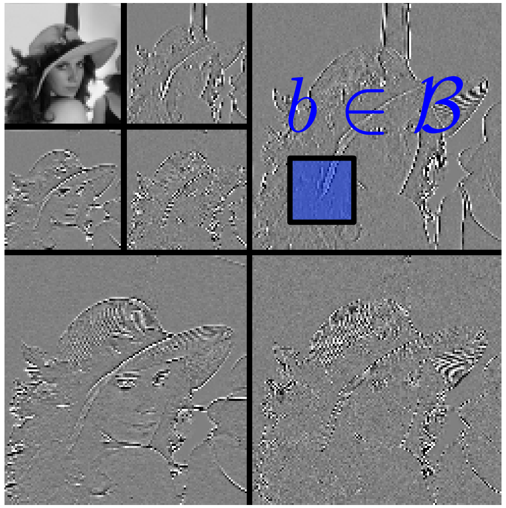
Block Regularization

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

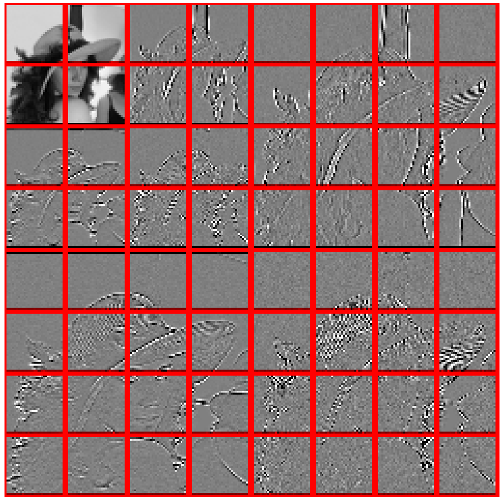
Non-overlapping decomposition: $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_n$

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

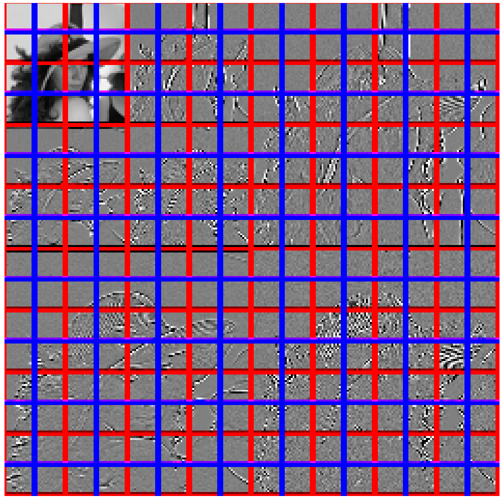
Image $f = \Psi x$



Coefficients x.



Blocks $\mathcal{B}_1$



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

Block Regularization

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

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:

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

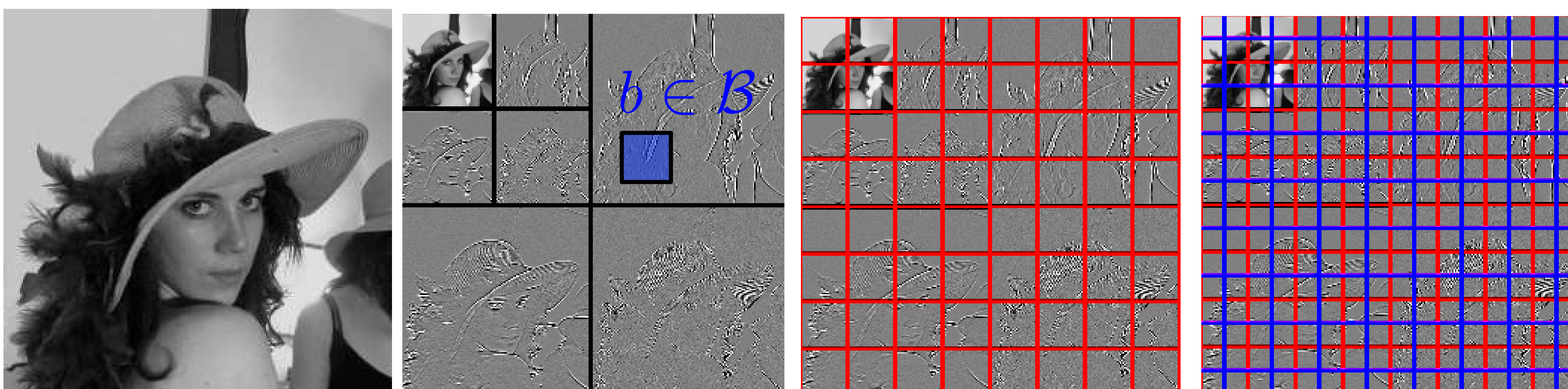


Image $f = \Psi x$ Coefficients x. Blocks $\mathcal{B}_1$ $\mathcal{B}_1 \cup \mathcal{B}_2$

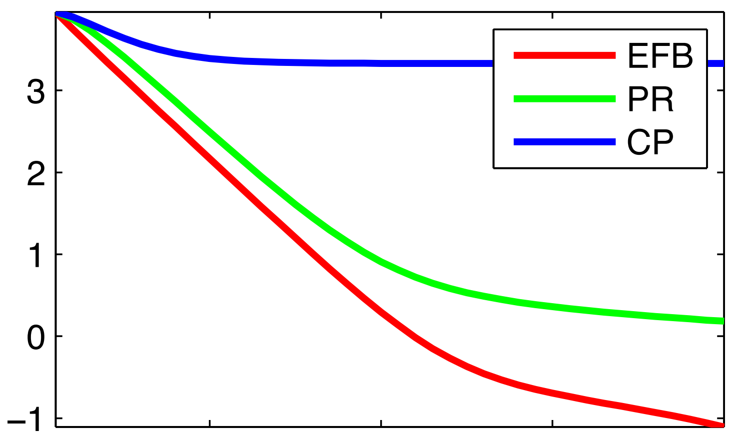
Numerical Illustration

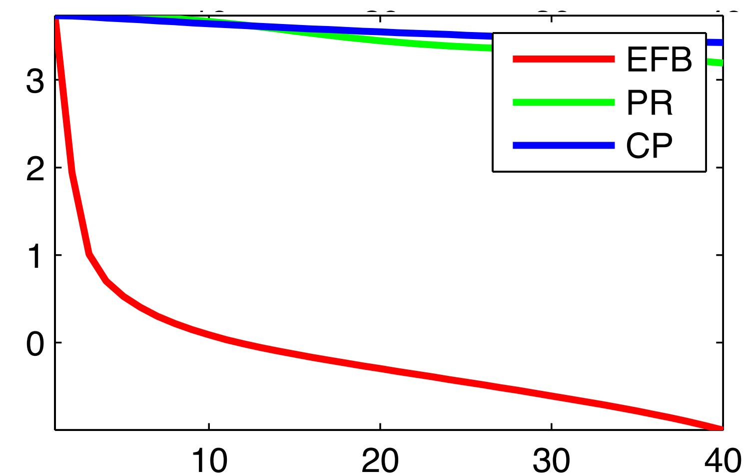
$$\min_x \frac{1}{2}\|y - \Phi\Psi x\|^2 + \lambda \sum_i G_i(x)$$

$\Psi = \text{TI wavelets}$

$\Phi = \text{convolution}$

$\Phi = \text{inpainting+convolution}$





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

x_0

$y = \Phi x_0 + w$

$x^\star$

Overview

- Optimal Transport and Imaging
- Convex Analysis and Proximal Calculus
- Forward Backward
- Douglas Rachford and ADMM
- Generalized Forward-Backward
- *Primal-Dual Schemes*

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 $u^\star$:

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

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 $u^\star$:

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

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

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

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: $\nabla^2 G_1 \geqslant c\mathrm{Id}$

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

$\longrightarrow$ $x^\star$ uniquely defined.

$\longrightarrow$ $G_1^\star$ is of class C^1.

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

Forward-Backward on the Dual

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

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

$\longrightarrow$ $x^\star$ uniquely defined.

$\longrightarrow$ $G_1^\star$ is of class C^1.

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

FB on the dual:

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

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

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

Example: TV Denoising

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

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

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]

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

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

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

Primal-Dual Algorithm

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

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

Primal-Dual Algorithm

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

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

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

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

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

Primal-Dual Algorithm

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

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

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

$\theta = 0$: 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

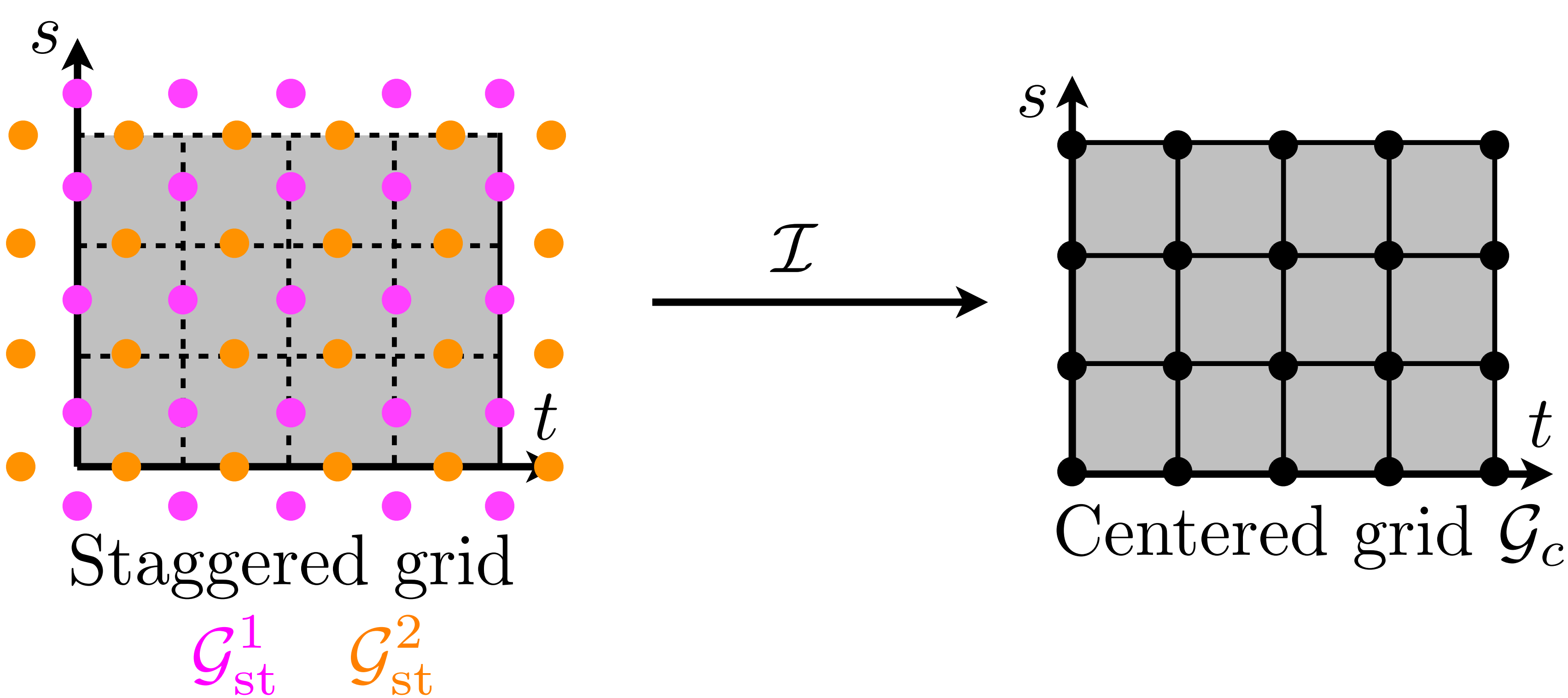
$x^{(\ell)} \rightarrow x^\star$ minimizer of $G_1 + G_2 \circ A$.

Example: Optimal Transport

Staggered grid formulation :

$$\min_{x \in \mathbb{R}^{\mathcal{G}^1_{\mathrm{st}}} \times \mathbb{R}^{\mathcal{G}^2_{\mathrm{st}}}} J(\mathcal{I}(x)) + \iota_{\mathcal{C}}(x)$$

$$\mathcal{I} = (\mathcal{I}^1, \mathcal{I}^2) : \mathbb{R}^{\mathcal{G}^1_{\mathrm{st}}} \times \mathbb{R}^{\mathcal{G}^2_{\mathrm{st}}} \longrightarrow \mathbb{R}^{\mathcal{G}_c}$$



Conclusion

Inverse problems in imaging:

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

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

$\rightarrow$ (Sometimes) convex.

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

Conclusion

Inverse problems in imaging:

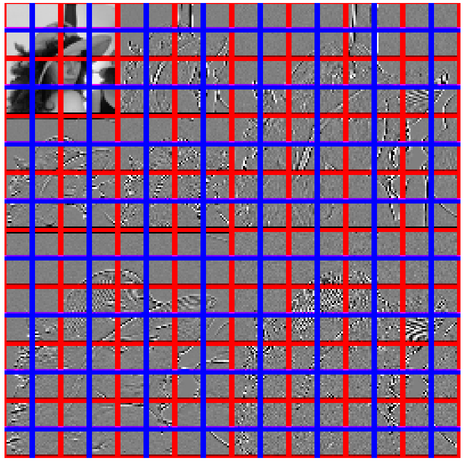
$\longrightarrow$ Large scale, $N \geqslant 10^6$.

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

$\longrightarrow$ (Sometimes) convex.

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

Proximal splitting:



$\longrightarrow$ Unravel the structure of problems.

$\longrightarrow$ Parallelizable.

Conclusion

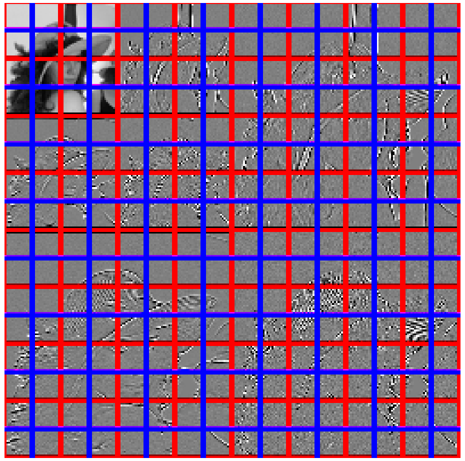
Inverse problems in imaging:

$\longrightarrow$ 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.

Open problems:

$\longrightarrow$ Less structured problems without smoothness.

$\longrightarrow$ Non-convex optimization.