

Motivation: TV-norm minimization

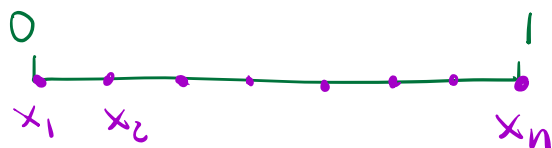
① The continuum model

$$\Omega = [0, 1] \times [0, 1]$$

$$L^2(\Omega) = \{ u(x, y) : \iint_{\Omega} |u(x, y)|^2 dx dy < +\infty \}$$

$$\|u\|_{TV} = \iint_{\Omega} |\nabla u| dx dy = \iint_{\Omega} \sqrt{u_x^2 + u_y^2} dx dy$$

② 1D Discrete Model



$$\Delta x = \frac{1}{n}$$

$$u_j = u(x_j)$$

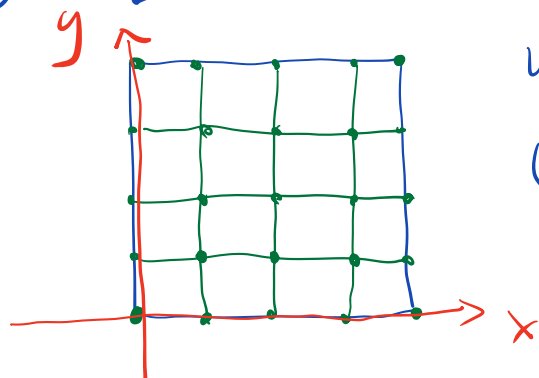
$$D = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \\ & & & & & 0 \end{pmatrix}_{n \times n}$$

$$D^T = \begin{pmatrix} -1 & & & & \\ 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ & & & & 1 & 0 \end{pmatrix}_{n \times n}$$

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$$\frac{1}{\Delta x} D U = \begin{pmatrix} \frac{u_2 - u_1}{\Delta x} \\ \vdots \\ \frac{u_n - u_{n-1}}{\Delta x} \\ 0 \end{pmatrix} \approx \begin{pmatrix} u'(x_1) \\ \vdots \\ u'(x_{n-1}) \\ 0 \end{pmatrix}$$

③ 2D Discrete Model



uniform grid (x_i, y_j) $\Delta x = \Delta y = h$

U is a $n \times n$ 2D array with

$$U(j, i) = u(x_i, y_j)$$

$$\begin{cases} U_x = \frac{1}{h} U D^T \approx U_x \\ U_y = \frac{1}{h} D U \approx U_y \end{cases}$$

Discrete TV Norm

$$\|U\|_{TV} = \sum_i \sum_j h^2 \sqrt{\underbrace{U_x^2(j,i) + U_y^2(j,i)}_{|U(j,i+1) - U(j,i)|^2}}$$

$U \in \mathbb{R}^{n \times n}$

Given a noisy image $B \in \mathbb{R}^{n \times n}$, want to solve

$$\min_{U \in \mathbb{R}^{n \times n}} \|U\|_{TV} + \frac{\lambda}{2h} \|U - B\|_{L^2}^2$$

$$\|U - B\|_{L^2}^2 = \sum_i \sum_j h^2 \cdot (U_{ij} - B_{ij})^2$$

$\lambda = 10 \sim 15$ is usually good for images

$$\Leftrightarrow \min_U \sum_i \sum_j \left[\sqrt{(DU)_{ij}^2 + (UD^T)_{ij}^2} + \frac{\lambda}{2} |U_{ij} - B_{ij}|^2 \right] h$$

[D] ROF (Rudin, Osher, Fatemi 1992) model

1D signal

$$\min_{u \in \mathbb{R}^n} \|u\|_{TV} + \frac{\alpha}{2} \|u - d\|^2 \quad \left(\frac{1}{2} \|x\|^2\right)^* = \frac{1}{2} \|x\|^2$$

$$\|Du\|_1 + \frac{\alpha}{2} \|u - d\|^2$$

$$\|u\|_{TV} = \sum_i |u_{i+1} - u_i|$$

$$= \|Du\|_1$$

$$Du = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$(n-1) \times n$



(a) Noisy Image



(b) $\lambda = 4$



(c) $\lambda = 8$



(d) $\lambda = 12$

Figure 4.2: ROF solutions using isotropic TV-norm with different λ .

$$\min_u \|Du\|_1 + \frac{\alpha}{2} \|u - d\|^2$$

we have prox for $f(x) = \|x\|_1$, but not $\|Du\|_1$
 Need to go to dual problems!

Def $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a function

$f^*(x) = \sup_y \langle x, y \rangle - f(y)$ is called

the convex conjugate of $f(x)$.

a.k.a. Legendre Transform
Fenchel Transform
Fenchel dual

If max exists
then $\sup = \max$

Theorem $f^*(x)$ is convex on its domain $\{x: f^*(x) < +\infty\}$

even if $f(x)$ is not convex

Example: $f(x) = e^x$

$$f^*(x) = \sup_y (xy - e^y)$$

sup attains at critical point $\Rightarrow x - e^{y^*} = 0$

$$\Rightarrow y_* = \log x$$

$$\Rightarrow f^*(x) = \begin{cases} x \log x - x & , x > 0 \\ 0 & , x = 0 \\ +\infty & , x < 0 \end{cases}$$

Example: $f(x) = ax + b$

$$f^*(x) = \sup_y (xy - f(y)) = \sup_y (xy - ay - b)$$

$$= \begin{cases} -b & , \quad x=a \\ +\infty & , \quad x \neq a \end{cases}$$

Theorem $f^*(x) + f(y) \geq \langle x, y \rangle$ if $f^*(x) \in \mathbb{R}$

Proof: $f^*(x) = \sup_y \langle x, y \rangle - f(y) \geq \langle x, y \rangle - f(y)$

Theorem ① $f^{**}(x) \leq f(x)$

② $f(x)$ is convex $\Rightarrow f^{**}(x) = f(x)$

Example: ① $f(x) = \|x\|$ for some norm $\|\cdot\|$

$$\Rightarrow f^*(x) = \begin{cases} 0 & , \quad \|x\|_* \leq 1 \\ +\infty & , \quad \text{otherwise} \end{cases}$$

indicator function of
unit ball

$\|x\|_*$ is the dual norm

② $f(x) = \frac{1}{2} \|x\|^2 \Rightarrow f^*(x) = \frac{1}{2} \|x\|_*^2$

Examples of dual norms for $x \in \mathbb{R}^n$

1) The dual norm of $\|x\|$ is $\|x\|$
 $\hookrightarrow$ vector 2-norm

2) The dual norm of $\|x\|_1$ is $\|x\|_\infty = \max_i |x_i|$

3) The dual norm of $\|x\|_\infty$ is $\|x\|_1$

4) The dual norm of $\|x\|_p$ is $\|x\|_q$ $\frac{1}{p} + \frac{1}{q} = 1$

$$\|x\|_p = \left[\sum_{i=1}^n |x_i|^p \right]^{\frac{1}{p}}$$

Example: $f(x) = \|x\|_1$

$$f^*(x) = \begin{cases} 0 & , \|x\|_\infty \leq 1 \\ +\infty & , \|x\|_\infty > 1 \end{cases}$$

Example:

$$f(x) = \begin{cases} 0 & , \|x\|_\infty \leq 1 \\ +\infty & , \|x\|_\infty > 1 \end{cases}$$

$$f^*(x) = \|x\|_1 \quad f \text{ is convex} \Rightarrow f^{**} = f$$

Moreau-Decomposition

$f(x)$ is convex

$$\text{Prox}_f^\eta(x) + \eta \text{Prox}_{f^*}^{\frac{1}{\eta}}(x/\eta) = x$$

Example: Find $\text{Prox}_f^\eta(x)$ for $f(x) = \begin{cases} 0 & , \|x\|_\infty \leq 1 \\ +\infty & , \|x\|_\infty > 1 \end{cases}$

Solution I: $f^*(x) = \|x\|_1$ $f^*(x) = \|x\|_1$

$$\begin{aligned} \text{Prox}_f^\eta(x) &= x - \eta \text{Prox}_{f^*}^{\frac{1}{\eta}}\left(\frac{x}{\eta}\right) \\ &= x - \eta S_{\frac{1}{\eta}}\left(\frac{x}{\eta}\right) \end{aligned}$$

$$S_\gamma(x)_i = \begin{cases} x_i - \gamma & , x_i > \gamma \\ x_i + \gamma & , x_i < -\gamma \\ 0 & , x_i \in [-\gamma, \gamma] \end{cases}$$

$$S_{\frac{1}{\eta}}\left(\frac{x}{\eta}\right)_i = \begin{cases} x_i/\eta - 1/\eta & , x_i > 1 \\ x_i/\eta + 1/\eta & , x_i < -1 \\ 0 & , x_i \in [-1, 1] \end{cases}$$

$$\text{Prox}_f^\eta(x)_i = x_i - \eta S_{\frac{1}{\eta}}\left(\frac{x}{\eta}\right)_i = \begin{cases} 1 & , x_i > 1 \\ -1 & , x_i < -1 \\ x_i & , x_i \in [-1, 1] \end{cases}$$

$$\text{Solution II: } f(x) = \begin{cases} 0 & , \|x\|_\infty \leq 1 \\ +\infty & , \|x\|_\infty > 1 \end{cases}$$

is the indicator function of $\|\cdot\|_\infty$ -ball

So $\text{Prox}_f^\eta(x)$ should be projection to this ball.

In practice, if we have Prox for $f(x)$
then we also have Prox for $f^*(x)$

$$f(x) \text{ is convex} \Rightarrow f = (f^*)^*$$

$\Rightarrow$

$$\begin{aligned} & \min_x f(x) + g(x) \\ &= \min_x \left(\max_y [\langle x, y \rangle - f^*(y)] + g(x) \right) \\ &= \min_x \max_y (\langle x, y \rangle - f^*(y) + g(x)) \end{aligned}$$

min & max
can be switched
under some assumptions

$$\begin{aligned}
 & \leftarrow \textcircled{=} \max_y \min_x (\langle x, y \rangle - f^*(y) + g(x)) \\
 &= \max_y \left[\min_x (\langle x, y \rangle + g(x)) - f^*(y) \right] \\
 &= \max_y \left[-\max_x (\langle x, -y \rangle - g(x)) - f^*(y) \right] \\
 &= \max_y \left[-g^*(-y) - f^*(y) \right] \\
 &= -\min_y \left[f^*(y) + g^*(-y) \right]
 \end{aligned}$$

Fenchel's Duality Theorem $\left. \begin{array}{l} f(x): \mathbb{R}^n \rightarrow \mathbb{R} \\ g(x): \mathbb{R}^n \rightarrow \mathbb{R} \end{array} \right\}$ are convex

$$\min_x [f(x) + g(x)] = -\min_y [f^*(y) + g^*(-y)]$$

Primal-Dual Relation $f^*(x) = \sup_y \langle x, y \rangle - f(y)$

If y_x is maximizer to $\langle x, y \rangle - f(y)$

then $\left\{ \begin{array}{l} f^*(x) = \langle x, y_x \rangle - f(y_x) \end{array} \right.$

$$0 \in x - \partial f(y_x) \Leftrightarrow x \in \partial f(y_x)$$

$$\Leftrightarrow f^*(x) + f(y_x) = \langle x, y_x \rangle$$

Strong Convexity - Smoothness Correspondence

$$\textcircled{1} \quad \left. \begin{array}{l} f(x) \text{ is convex} \\ \nabla f(x) \text{ is } L\text{-cont. with } L = \sigma > 0 \end{array} \right\} \Leftrightarrow \left. \begin{array}{l} f^*(x) \text{ is strongly} \\ \text{convex with } \mu = \frac{1}{\sigma} \end{array} \right.$$

② $f(x)$ is strongly convex with $\mu = \sigma \iff \begin{cases} f^*(x) \text{ is convex} \\ \nabla f^*(x) \text{ is } L\text{-cont. with } L = \frac{1}{\sigma}. \end{cases}$

Remark: Strong convexity in $f(x) \iff$ Smoothness in dual

Fact If $f(x)$ is strongly convex,

then $\nabla f^*(y) = x \iff x \in \partial f(y)$

∇f^* is inverse to ∂f

Example: $\min_x \underbrace{\|x\|_1}_{f(x)} + \underbrace{\mathcal{I}_{\{x: Ax=b\}}(x) + \frac{1}{2\alpha} \|x\|_2^2}_{g(x)}$

We can use

① Douglas-Rachford for $\min_x f(x) + g(x)$

② Douglas-Rachford for $\min_y f^*(y) + g^*(-y)$

③ Fast Proximal Gradient for $\min_y f^*(y) + g^*(-y)$
if gradient can be computed

g is $\frac{1}{2}$ -strongly convex $\iff \nabla g^*$ is L -continuous
with $L = 2$.

$$\nabla g^*(y) = 2y - A^T(AA^T)^{-1}(2Ay - b)$$

Consider $\min_x \|x\|_1 \text{ s.t. } Ax = b$



$$\min_x \|x\|_1 + \mathbb{I}_{\{x: Ax=b\}}(x)$$

$$f(x) + g(x)$$



$$\min_y f^*(y) + g^*(-y)$$

we have

$$\left\{ \begin{array}{ll} \text{Prox}_f^\eta(x) = S_\eta(x) & \text{shrinkage} \\ \text{Prox}_g^\eta(x) = P(x) & \text{projection} \\ & = x + A^T(AA^T)^{-1}(b - Ax) \end{array} \right.$$

By Moreau-Decomposition, we get

$$\text{Prox}_{f^*}^\eta(x) = x - \eta \text{Prox}_f\left(\frac{x}{\eta}\right) = P_1(x) = \begin{cases} 1 & , x_i > 1 \\ -1 & , x_i < -1 \\ x_i & , x_i \in [-1, 1] \end{cases}$$

$$\begin{aligned} \text{Prox}_{g^*}^\eta(x) &= x - \eta \text{Prox}_g^\eta\left(\frac{x}{\eta}\right) \\ &= x - \eta \left[\frac{x}{\eta} + A^T(AA^T)^{-1}(b - A \frac{x}{\eta}) \right] \\ &= -A^T(AA^T)^{-1}(\eta b - Ax) \\ &= A^T(AA^T)^{-1}(Ax - \eta b) \end{aligned}$$

For using generalized Douglas-Rachford, we get four families of algorithms:

$$\begin{aligned} x_k &= \text{Prox}_g(y_k) & R_g(y_k) &= 2\text{Prox}_g(y_k) - y_k = 2x_k - y_k \\ R_f[R_g(y_k)] &= 2\text{Prox}_f(2x_k - y_k) - (2x_k - y_k) \\ \frac{I + R_f R_g}{2}(y_k) &= \text{Prox}_f(2x_k - y_k) - x_k + y_k \end{aligned}$$

General DR: $y_{k+1} = (1-\lambda)y_k + \lambda \frac{I + R_f R_g}{2}(y_k)$

① $\begin{cases} x_k = \text{Prox}_g^\eta(y_k) & x_k \rightarrow x_* \\ y_{k+1} = (1-\lambda)y_k + \lambda [\text{Prox}_f^\eta(2x_k - y_k) + y_k - x_k] \end{cases} \quad \lambda \in (0, 2)$

② $\begin{cases} x_k = \text{Prox}_f^\eta(y_k) & x_k \rightarrow x_* \\ y_{k+1} = (1-\lambda)y_k + \lambda [\text{Prox}_g^\eta(2x_k - y_k) + y_k - x_k] \end{cases} \quad \lambda \in (0, 2)$

Theorem 7.8. The sequence produced by

$$\begin{cases} z_{k+1} &= R_g^\eta R_f^\eta(z_k) \\ x_k &= R_f^\eta(z_k) \end{cases}, \quad z_0 = R_g^\eta(y_0),$$

is the same as the sequence produced by

$$\begin{cases} y_{k+1} &= R_f^\eta R_g^\eta(y_k) \\ x_k &= R_g^\eta(y_k) \end{cases}, \quad \forall y_0.$$

Remark 7.1. For general Douglas-Rachford splitting, though the same result cannot be shown, in practice the difference in numerical performance between two different versions caused by switching f and g is marginal and minimal.

So switching f and g doesn't really make a difference

$$\min_z f^*(z) + g^*(-z) \quad F(z) = f^*(z) \\ \Leftrightarrow \min_z F(z) + G(z) \quad G(z) = g^*(-z)$$

$$\text{Prox}_F^\eta(x) = \text{Prox}_{f^*}^\eta(x) = x - \eta \text{Prox}_f^{\frac{1}{\eta}}\left(\frac{x}{\eta}\right)$$

$$\begin{aligned} \text{Prox}_G^\eta(x) &= \arg\min_u [g^*(u) + \frac{1}{2\eta} \|u - x\|^2] \\ &= \arg\min_v [g^*(v) + \frac{1}{2\eta} \|-v - x\|^2] \\ &= \arg\min_v [g^*(v) + \frac{1}{2\eta} \|v - (-x)\|^2] \\ &= \text{Prox}_{g^*}^\eta(-x) \\ &= -x - \eta \text{Prox}_g^{\frac{1}{\eta}}\left(-\frac{x}{\eta}\right) \end{aligned}$$

with $f(\mathbf{x}) = \|\mathbf{x}\|_1$ and $g(\mathbf{x}) = \iota_{\{\mathbf{x}: A\mathbf{x}=b\}}(\mathbf{x})$. To apply the Douglas-Rachford splitting, it seems that there are at least four choices to do fixed point iteration $\mathbf{y}_{k+1} = T(\mathbf{y}_k)$:

1. $T = \frac{1}{2}[\mathbb{I} + R_{f(\mathbf{x})} R_{g(\mathbf{x})}]$. (P) $\min_x f(x) + g(x)$
2. $T = \frac{1}{2}[\mathbb{I} + R_{g(\mathbf{x})} R_{f(\mathbf{x})}]$. $\Downarrow$
3. $T = \frac{1}{2}[\mathbb{I} + R_{f^*(\mathbf{y})} R_{g^*(-\mathbf{y})}]$. (D) $-\min_y f^*(y) + g^*(-y)$
4. $T = \frac{1}{2}[\mathbb{I} + R_{g^*(-\mathbf{y})} R_{f^*(\mathbf{y})}]$.

② with step size $\eta > 0 \Leftrightarrow$ ④ with step size $\frac{1}{\eta}$
 $\downarrow$ for any convex f & g

Theorem 7.15.

Remark: In practice, difference between ① and ② is marginal