Motivation: TV-norm minimization

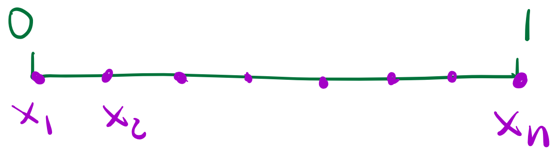
① The continuum model

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

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

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

② 1D Discrete Model



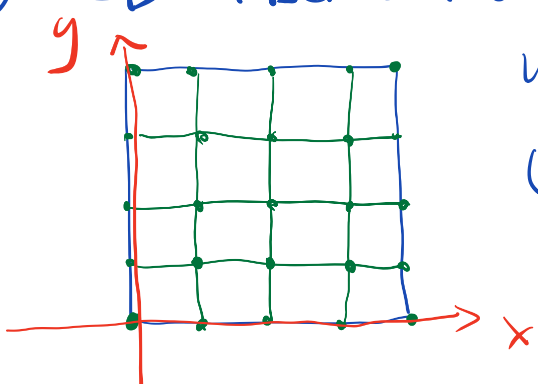
$\Delta x = \frac{1}{n}$

$u_j = u(x_j)$

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

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

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

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

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

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

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

1D

ROF (Rudin, Osher, Fatemi 1992) model

1D signal

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

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

$$\boxed{\begin{aligned} \|u\|_{TV} &= \sum_i |u_{i+1} - u_i| \\ &= \|Du\|_1 \end{aligned} \qquad Du = \underbrace{\begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}}_{(n-1)\times n} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}}$$

(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 $f: \mathbb{R}^n \to \mathbb{R}$ is a function

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

the convex conjugate of $f(x)$.

a.k.a. Legendre Transform

Fenchel Transform

Fenchel dual

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: $f(x) = ax + b$

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

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

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 $\|\cdot\|$

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

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 $\|x\|_q$ $\quad \frac{1}{p}+\frac{1}{q}=1$

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

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

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

Example:

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

$$f^*(x) = \|x\|_1$$

$\boxed{\text{Morean-Decomposition}}$ $f(x)$ is convex

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

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

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

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

$$= x - \eta S_{\frac{1}{\eta}}\left(\frac{x}{\eta}\right)$$

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

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

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

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

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 $f^*(x)$

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

$\Rightarrow$

$$\min_x f(x) + g(x)$$

$$= \min_x \left(\max_y \left[\langle x, y \rangle - f^*(y) \right] + g(x) \right)$$

min & max can be switched under some assumptions $\quad = \min_x \max_y \left(\langle x, y \rangle - f^*(y) + g(x) \right)$

$$\textcircled{=} \max_y \min_x \left(\langle x, y\rangle - f^*(y) + g(x) \right)$$

$$= \max_y \left[\min_x \left(\langle x, y\rangle + g(x) \right) - f^*(y) \right]$$

$$= \max_y \left[-\max_x \left(\langle x, -y\rangle - g(x) \right) - f^*(y) \right]$$

$$= \max_y \left[-g^*(-y) - f^*(y) \right]$$

$$= -\min_y \left\{ f^*(y) + g^*(-y) \right\}$$

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

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

Primal-Dual Relation $\min \underbrace{\|x\|_1 + \alpha \|x\|_2^2}_{f(x)} + \underbrace{i_{\{Ax=b\}}}_{g(x)}$

$$x^* = \arg\min_x \langle x, y^*\rangle + g(x) \Leftrightarrow 0 \in y^* + \partial g(x^*) \Leftrightarrow -y^* \in \partial g(x^*)$$

Similarly $y^* = \arg\min_y \left[f^*(y) - \langle x^*, y\rangle \right] \Leftrightarrow x^* \in \partial f^*(y^*)$

Strong Convexity – Smoothness Correspondance

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

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

Remark: strong convexity in $f(x) \Leftrightarrow$ Smoothness in dual

Example: $\min_x \|x\|_1$ s.t. $Ax=b$

$$\Updownarrow$$

$$\min_x \|x\|_1 + i_{\{x:Ax=b\}}(x)$$

$$\Updownarrow$$ ← proven true if $\alpha>0$ is large

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

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

g is $\frac{1}{\alpha}$-strongly convex $\Leftrightarrow$ ∇g^* is L-continuous with $L=\alpha$.

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

$$\Updownarrow$$

$$\min_{x} \|x\|_1 + i_{\{x:Ax=b\}}(x)$$

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

$$\Updownarrow$$

$$\min_{y} f^*(y) + g^*(-y)$$

We have $\begin{cases} \text{Prox}_f^{\eta}(x) = S_\eta(x) & \text{shrinkage} \\ \text{Prox}_g^{\eta}(x) = P(x) & \text{projection} \\ \qquad\qquad = x + A^T(AA^T)^{-1}(b-Ax) & \end{cases}$

By Moreau-Decomposition, we get

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

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

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

$$x_k = \mathrm{Prox}_g(y_k)$$

$$R_g(y_k) = 2\,\mathrm{Prox}_g(y_k) - y_k = 2x_k - y_k$$

$$R_f[R_g(y_k)] = 2\,\mathrm{Prox}_f(2x_k - y_k) - (2x_k - y_k)$$

$$\frac{I + R_f R_g}{2}(y_k) = \mathrm{Prox}_f(2x_k - y_k) - x_k + y_k$$

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

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

Theorem 7.8. *The sequence produced by*

$$\begin{cases} \mathbf{z}_{k+1} &= \mathrm{R}_g^\eta \mathrm{R}_f^\eta(\mathbf{z}_k) \\ \mathbf{x}_k &= \mathrm{R}_f^\eta(\mathbf{z}_k) \end{cases}, \quad \mathbf{z}_0 = \mathrm{R}_g^\eta(\mathbf{y}_0),$$

is the same as the sequence produced by

$$\begin{cases} \mathbf{y}_{k+1} &= \mathrm{R}_f^\eta \mathrm{R}_g^\eta(\mathbf{y}_k) \\ \mathbf{x}_k &= \mathrm{R}_g^\eta(\mathbf{y}_k) \end{cases}, \quad \forall \mathbf{y}_0.$$

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

$$\min_z f^*(z) + g^*(-z) \qquad F(z) = f^*(z)$$

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

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

$$\begin{aligned}
\mathrm{Prox}_G^\eta(x) &= \operatorname*{argmin}_u \left[g^*(-u) + \frac{1}{2\eta}\|u-x\|^2 \right] \\
&= \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|-v-x\|^2 \right] \\
&= \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|v-(-x)\|^2 \right] \\
&= \mathrm{Prox}_{g^*}^\eta(-x) \\
&= -x - \eta\, \mathrm{Prox}_g^{1/\eta}\left(-\frac{x}{\eta}\right)
\end{aligned}$$

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. $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{f(\mathbf{x})}\,\mathrm{R}_{g(\mathbf{x})}]$.
2. $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{g(\mathbf{x})}\,\mathrm{R}_{f(\mathbf{x})}]$.
3. $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{f^*(\mathbf{y})}\,\mathrm{R}_{g^*(-\mathbf{y})}]$.
4. $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{g^*(-\mathbf{y})}\,\mathrm{R}_{f^*(\mathbf{y})}]$.

2 with step size $\eta > 0$ $\Longleftrightarrow$ 4 with step size $\frac{1}{\eta}$

For solving $\min_{\mathbf{x}}[F(\mathbf{x}) + G(\mathbf{x})]$, with step size $\eta > 0$, it can be written as

$$\text{General Douglas-Rachford}: \mathbf{v}_{k+1} = [(1-\lambda)\mathbb{I} + \lambda \frac{\mathbb{I} + \mathrm{R}_F^\eta \, \mathrm{R}_G^\eta}{2}](\mathbf{v}_k), \quad \lambda \in (0,2).$$

For the primal problem $\min_{\mathbf{x}}[f(\mathbf{x}) + g(\mathbf{x})]$, we take $G(\mathbf{x}) = f(\mathbf{x})$ and $F(\mathbf{x}) = g(\mathbf{x})$, then

$$\text{DR on (P)}: \begin{cases} \mathbf{v}_{k+1} & = [(1-\lambda)\mathbb{I} + \lambda \frac{\mathbb{I} + \mathrm{R}_g^\eta \, \mathrm{R}_f^\eta}{2}](\mathbf{v}_k), \quad \lambda \in (0,2) \\ & = \mathbf{v}_k - \lambda \mathbf{x}_k + \lambda \operatorname{Prox}_f^\eta(2\mathbf{x}_k - \mathbf{v}_k) \\ \mathbf{x}_k & = \operatorname{Prox}_f^\eta(\mathbf{v}_k) \end{cases} . \tag{7.1}$$

For the dual problem $\min_{\mathbf{y}}[f^*(\mathbf{y}) + g^*(-\mathbf{y})]$, we take $F(\mathbf{y}) = g^*(-\mathbf{y})$ and $G(\mathbf{y}) = f^*(\mathbf{y})$, then

$$\operatorname{Prox}_F^\eta = -\operatorname{Prox}_{g^*}^\eta .$$

Using step size $\tau > 0$, we have

$$\text{DR on (D)}: \begin{cases} \mathbf{u}_{k+1} & = [(1-\lambda)\mathbb{I} + \lambda \frac{\mathbb{I} + \mathrm{R}_F^\tau \, \mathrm{R}_G^\tau}{2}](\mathbf{u}_k), \quad \lambda \in (0,2) \\ & = \mathbf{u}_k - \lambda \mathbf{y}_k - \lambda \operatorname{Prox}_{g^*}^\tau(2\mathbf{y}_k - \mathbf{u}_k) \\ \mathbf{y}_k & = \operatorname{Prox}_{f^*}^\tau(\mathbf{u}_k) \end{cases} . \tag{7.2}$$

Theorem 7.9. *The general Douglas-Rachford splitting on the primal problem (7.1) is exactly the same as general Douglas-Rachford splitting on the dual problem (7.2) if $\eta = \frac{1}{\tau}$. In particular, $\mathbf{u}_k = \frac{\mathbf{v}_k}{\eta}$, $\mathbf{x}_k \to \mathbf{x}_*$ and $\mathbf{y}_k \to \mathbf{y}_*$.*