Nikola Janjušević
Fall 2025

Sparsity Based Image Processing

① Regularized Image Recovery

1.1

ex: Image Deblur $\quad Hx = y$

Noisy: $\quad \min_x \frac{1}{2}\|Hx - y\|_2^2 + \frac{\alpha}{2}\|x\|_2^2$

ex: Subsample MRI (single coil)

$$y = Ax = MFx \quad . \quad \text{rank}(A) < n$$

A not invertible.

Need prior information to fill in the gaps...

1.2 MAP formulation ⇒ Energy Min formulation

↳ $y \sim \mathcal{N}(Ax, \sigma^2 I)$, $\quad x \sim p(x)$

↳ General $y \sim p(y|x)$

↳ $p(x|y) \propto p(y|x)\,p(x)$.

↳ Analytical solve? → Yes → great.

↳ No? Need Numerical solver. ex. GD.

→ When do we have a unique solution?

↳ $\boxed{\text{convex}}$ loss function:

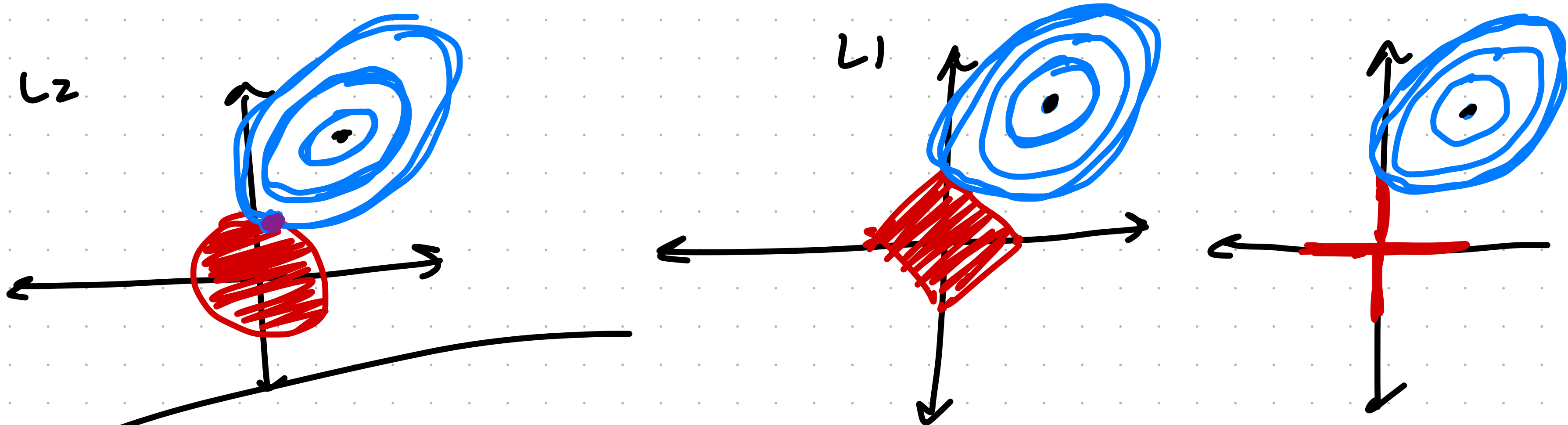
1.3 Sparsity Priors

↳ 1 Wavelet sparsity 2 $\phi = \|Wx\|_1$

↳ Total Variation $\phi = \|\nabla x\|_1$, $\phi = \|\nabla x\|_{2,1}$

↳ TGV

1 $\boxed{\text{How do we encode sparsity}}$?



→ convex relaxation of L0 pseudo norm.

So solve $x^* = \arg\min \frac{1}{2} \| y - Hx \|_2^2 + \phi(x)$

where $\phi = \| Wx \|_1$.

Convex ? ✓

Differentiable ?

1) Sub GD
2) Prox GD (ISTA)
3) ADMM.
4) PDS

① Signal / Image Recovery

We want to recover x from some measurement y.

Linear Measurements: $y = Hx$.

ex. Blurring: $Hx = h * x$ (conv)

ex. Down Sampling: $Hx = (\downarrow 2)(h * x)$

ex. In-painting: $Hx = m \odot x$ m mask.

ex MRI: $Hx = Fx$ Fourier Transform

F: DFT matrix.

ex. X-Ray CT: $Hx = R\{x\}$ ← Radon Transform.

So if $H \in \mathbb{R}^{n \times n}$ square.

Then $\underline{x^* = H^{-1}y} = H^{-1}Hx = x$.

ex. $Hx = h * x$. What is H^{-1}?

<u>Conv theorem</u>: $h * x \iff \hat{h} \circ \hat{x} = \mathcal{F}\{h\} \circ \mathcal{F}\{x\}$

$$\longrightarrow \quad F(h * x) = \hat{h} \circ \hat{x} = \mathrm{diag}(\hat{h})\hat{x} = \mathrm{diag}(\hat{h})Fx$$

$$FHx = \mathrm{diag}(\hat{h})Fx \quad \forall x.$$

$$FH = \mathrm{diag}(\hat{h})F$$

$$H = F^{-1}\mathrm{diag}(\hat{h})F$$

$$H = F^{H}\mathrm{diag}(\hat{h})F$$

↳ apply H in fourier domain

↳ Eigen Decomp of H.

So. What is H^{-1}?

$$H^{-1} = \left(F^{H} \operatorname{diag}(\hat{h}) F \right)^{-1}$$

$$= F^{-1} \operatorname{diag}(\hat{h})^{-1} F^{-H}$$

$$\underline{H^{-1} = F^{H} \operatorname{diag}\left(\frac{1}{\hat{h}}\right) F}$$ ← Assumes $\hat{h} > 0$.

So $x^* = H^{-1} y$ ✓

Blur kernel is normalized s.t. $\|h\|_1 = 1$.

so that DC response is 1.

$$\mathcal{F}_2\{h\}(0) = \sum_{-\infty}^{\infty} h[n] e^{-j\omega n} = \sum_{-\infty}^{\infty} h[n] = 1.$$

What if our measurements aren't exact?
i.e. we have some measurement noise?

$$y = Hx + v. \quad \text{Let } v \sim N(0, \sigma^2 I)$$

then $x^* = H^{-1} y = H^{-1}(Hx + v)$

$$= x + H^{-1} v$$

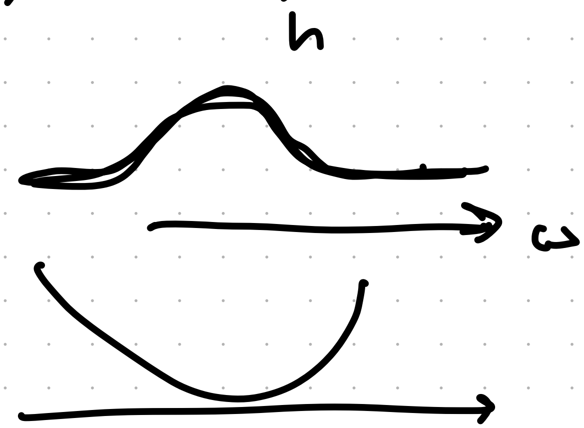
$$x^* = x + v^*, \quad v^* \sim N(0, \sigma^2 H^{-1} H^{-T})$$

$$H^{-1}H^{-H} = F^H \mathrm{diag}\left(\frac{1}{\hat h}\right) F \left(F^H \mathrm{diag}\left(\frac{1}{\hat h^*}\right) F\right)$$

$$= F^H \mathrm{diag}\left(\frac{1}{|\hat h|^2}\right) F$$

i.e. if $\hat h$ looks like

$|\hat h|^{-2} \longrightarrow$



High freq noise amplified!!!

$\Longrightarrow$ $x^\diamond$ explodes in magnitude.!

Solution: encode prior information

↳ image magnitude reasonable

↳ penalize large images.

$$x^* = \underset{x}{\operatorname{argmin}} \; \underbrace{\tfrac{1}{2}\|y - Hx\|_2^2}_{f \;\; \text{Data Fid.}} + \underbrace{\tfrac{\lambda}{2}\|x\|_2^2}_{g \;\; \text{Prior}}$$

Solve? Unique Min? Where?

$$\nabla_x \left(f(x^*) + g(x^*) \right) = 0$$

$$\nabla_x f(x^*) + \nabla_x g(x^*) = 0$$

$$H^T(Hx - y) + \lambda x = 0$$

$$(\lambda I + H^T H)x = H^T y$$

$$x = (\lambda I + H^T H)^{-1} H^T y$$

$$= \left(\lambda F^H F + F^H \operatorname{diag}(|\hat{h}|^2) F\right)^{-1} H^T y$$

$$= \left(F^H(\lambda I + \operatorname{diag}(|\hat{h}|^2)F\right)^{-1} F^H \operatorname{diag}(\hat{h}^*) F y$$

$$= F^H \operatorname{diag}\left(\frac{1}{\lambda + |\hat{h}|^2}\right) \operatorname{diag}(\hat{h}^*) F y$$

$$\boxed{x^\lambda = F^H \operatorname{diag}\left(\frac{\hat{h}^*}{\lambda + |\hat{h}|^2}\right) F y}$$

(Weiner Filter)

↳ Choose λ reasonable s.t. high freq noise doesn't blow up.

↳ <u>Extend this idea:</u>

- Problem: ~~Loss of information~~ approx loss yields unstable solution.

Extreme Example: $y = MFx$.

mask decimates information,
we cannot recover w/out assumptions.

~~Vexample~~: if x is band limited to
$|\omega| < \frac{\pi}{N}$ $\Longrightarrow$ $(\downarrow N)x$ has same
freq content.

↳ Prior knowledge allows us to
successfully recover certain signals.

1.2 Probability Perspective :

ex. let $x \sim p(x)$.

let $y \sim N(Hx, \sigma^2 I)$

then MAP estimation :

$$x^* = \underset{x}{\mathrm{argmax}}\ p(x|y)$$

$$= \underset{x}{\mathrm{argmax}}\ \frac{p(y|x)p(x)}{p(y)} \quad \text{Bayes.}$$

$$= \underset{x}{\mathrm{argmax}}\ p(y|x)\,p(x)$$

$$= \underset{x}{\operatorname{amax}} \ \log p(y|x) + \log p(x)$$

$$= \underset{x}{\operatorname{amin}} \ -\log p(y|x) - \log p(x)$$

$$x^* = \underset{x}{\operatorname{amin}} \ \frac{1}{2\sigma^2} \|y - Hx\|_2^2 - \log p(x)$$

let $x \sim N(0, \rho^2 I)$

then $x^* = \underset{x}{\operatorname{amin}} \ \frac{1}{2\sigma^2} \|y - Hx\|^2 \ \frac{1}{2\rho^2} \|x\|^2$

$$x^* = \underset{x}{\operatorname{amin}} \ \frac{1}{2} \|y - Hx\|_2^2 \ \frac{\lambda}{2} \|x\|^2$$

Weiner filter.

where $\lambda = \frac{\sigma^2}{\rho^2}$

if I choose a function $\phi(x)$ as my regularization, can interpret it as a prior

$$\phi(x) = -\log p(x)$$

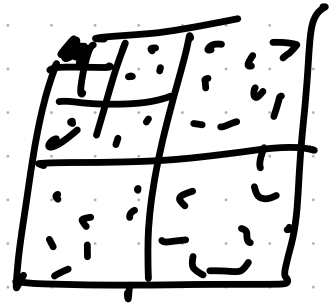
$$\Longleftrightarrow$$

$$p(x) \propto e^{-\phi(x)}$$

Also, my data fidelity implicitly encodes a noise model

($L2 \Longleftarrow$ Gaussian)

1.3 Sparsity Priors

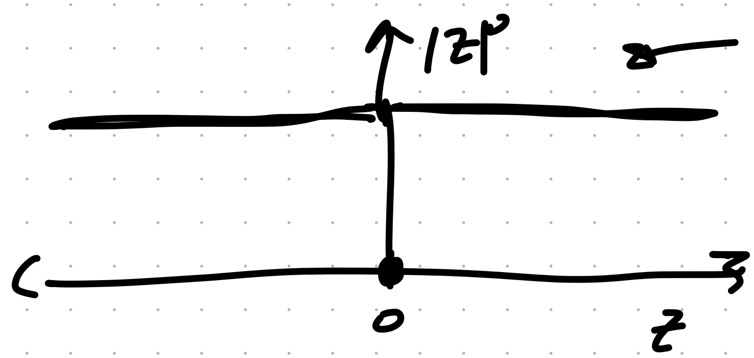
example: Wavelet(x) = 

few non-zero coeffs.

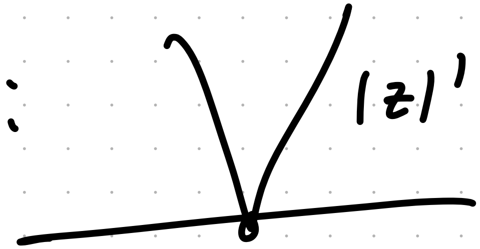
Prior: $\phi(x) = (\#\text{ non-zeros})(Wx)$

$$\#\text{ non-zeros}(z) = \sum_n |z_n|^0$$

$\Longrightarrow$



Not convex.

↳ Closest convex analogue: 

So let $\phi(z) = \sum_i |z_i| = \|z\|_1$

then $x^* = \underset{x}{\arg\min}\ \frac{1}{2}\|y - Hx\|_2^2 + \lambda \|Wx\|_1$

Remember W is unitary, so if $z = Wx$,

$x = W^T z$.

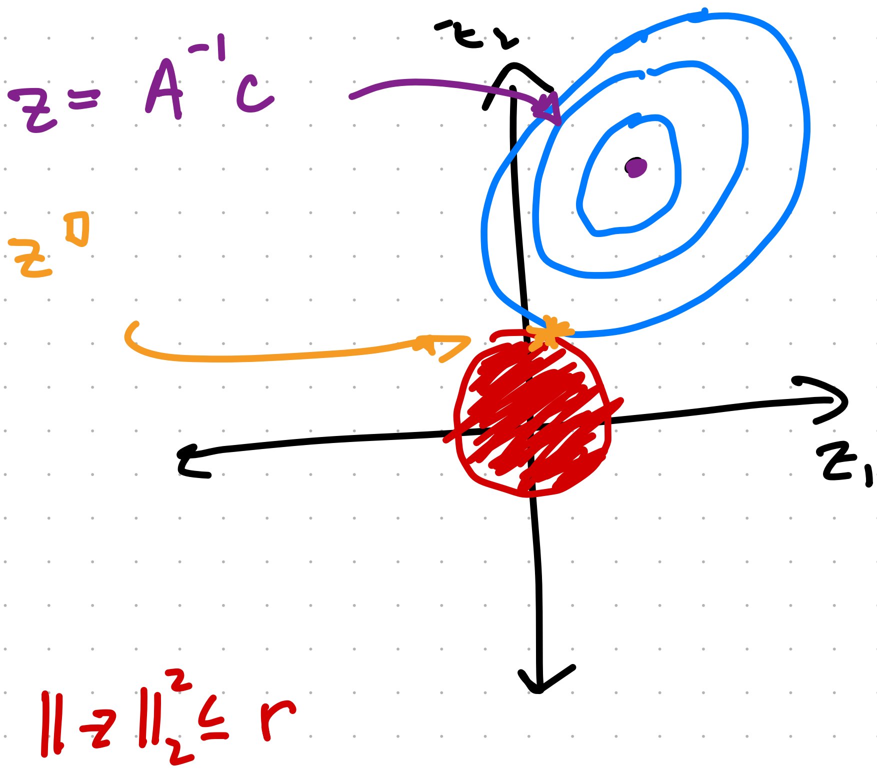
$$z^* = \underset{z}{\arg\min}\ \frac{1}{2}\|W(y - HW^T z)\|_2^2 + \lambda\|z\|_1$$

$$= \underset{z}{\arg\min}\ \frac{1}{2}\|c - Az\|_2^2 + \lambda\|z\|_1$$

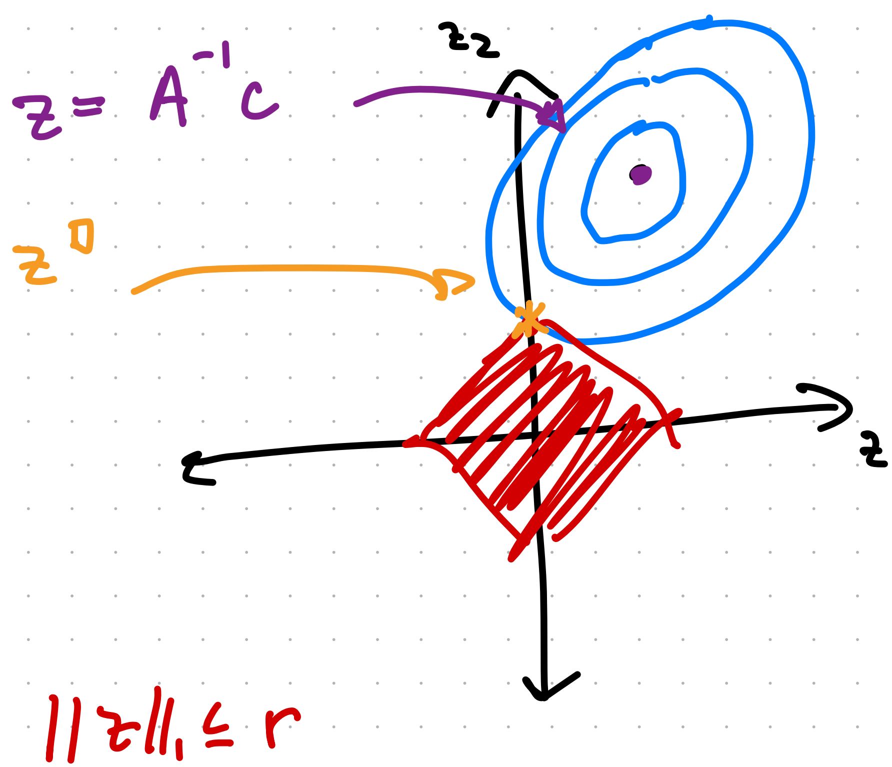
$$= \underset{z}{\arg\min}\ \mathcal{L}(z).$$

How / Why Do we expect L1 to encourage sparsity?

L2



L1



$\Longrightarrow$ L1 is pointy!

Simple case: $H = I \Rightarrow y = x + v$

$Wy = Wx + Wv$

$c = z + v_w$

$\Rightarrow$ Denoise in the wavelet Domain. $\Leftarrow A = WHW^T = WW^T = I$

$$z^* = \underset{z}{\mathrm{amin}} \underbrace{\frac{1}{2}\|c - z\|_2^2 + \lambda \|z\|_1}_{\mathcal{L}(z)}$$

$$\mathcal{L}(z) = \frac{1}{2}\sum_i (c_i - z_i)^2 + \lambda |z_i|$$

$$= \mathcal{L}_i(z_i)$$

Can optimize for each z_i individually:

$$z^* = \underset{z}{\operatorname{argmin}} \ \frac{1}{2}(c - z)^2 + \lambda |z|$$

(1) <u>assume $z^* \neq 0$:</u>

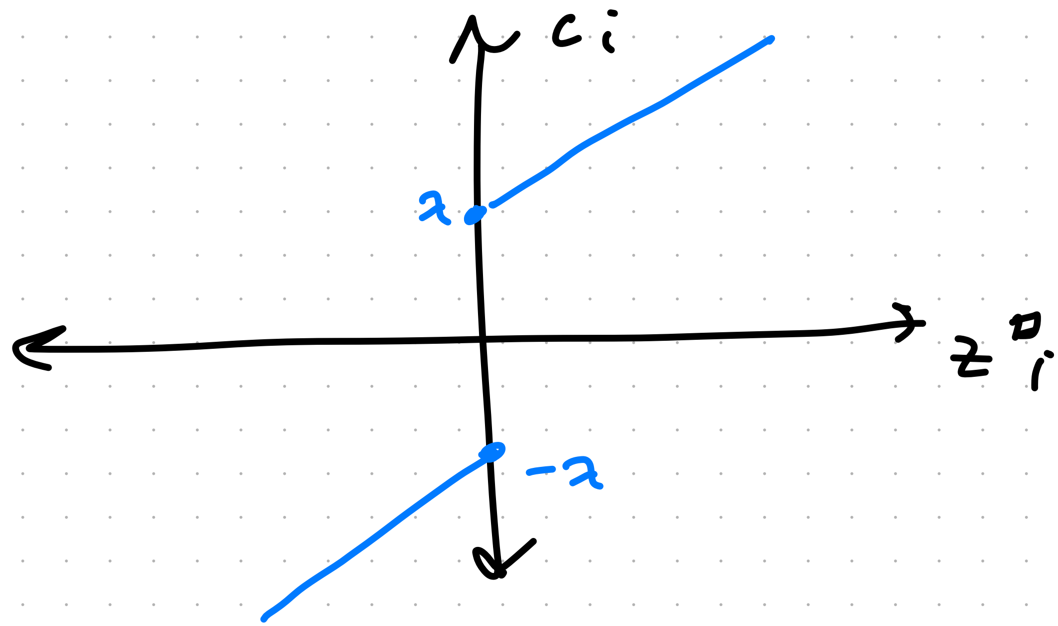
$$\frac{\partial}{\partial z}\left(\frac{1}{2}(c - z)^2 + \lambda |z|\right) = 0$$

$$z^* - c + \lambda \operatorname{sign}(z^*) = 0$$

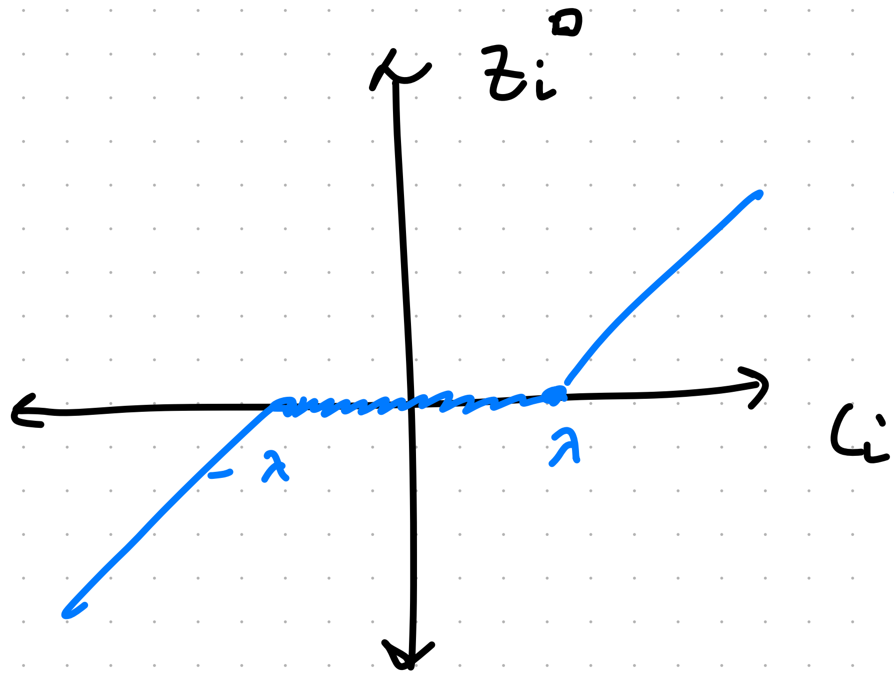
$$c = z^* + \lambda \operatorname{sign}(z^*)$$

(2) if $z^* = 0 \implies z^* = 0$:)

∇



flip



$\equiv ST$

$$z_i^* = ST_\lambda(c_i) = \text{sign}(c_i)\max(0, |c_i| - \lambda)$$

So $\Rightarrow$ $$z^0 = \underset{z}{\operatorname{argmin}} \frac{1}{2}\| c - z\|^2 + \lambda \|z\|_1$$

$$\boxed{z^0 = ST_\lambda(c)}$$ ← element-wise

Lets Generalize:

Define the <u>proximal operator of ϕ</u>

$$\operatorname{prox}_{\lambda\phi}(u) = \underset{x}{\operatorname{argmin}} \lambda\phi(x) + \frac{1}{2}\|u - x\|_2^2$$

for ϕ <u>(strictly) convex</u>. (i.e unique solution)

So ϕ might not be smooth

(no cts. derivative).

$\hookrightarrow$ but as in the case of $\phi = |\cdot|$,

we can still get a solution.

for smooth ϕ, say

x^0 s.t. $\nabla_x \left(\frac{1}{2} \| u - x \| + \lambda \phi(x) \right) = 0$

$$x - u + \lambda \nabla \phi(x) = 0$$

for non-smooth x, generalize to subgradient:

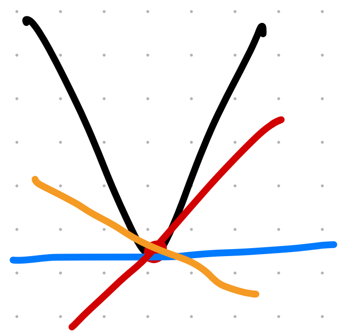
$$x^* \text{ s.t. } \partial_x \left(\frac{1}{2}\|x-u\|^2 + \lambda \partial\phi(x) \right) \ni 0$$

$$0 \in x - u + \lambda \partial\phi(x)$$

what does $\partial\phi$ mean?

<u>set of all slopes below the fn.</u>

ex. $|z|$



$$(\partial|z|)(z) = \begin{cases} \text{sign}(z), & z \neq 0 \\ [-1, 1], & z = 0 \end{cases}$$

Reduces to regular derivative when regular derivative is defined (i.e. smooth)

So we say we're at the minimizer if $0 \in \partial \mathcal{L}$

Note: $x = \text{prox}_{\lambda\phi}(u)$,

$$0 \in x - u + \lambda \partial \phi(x)$$

$$0 \in (I + \lambda \partial \phi) x - u$$

$$u \in (I - \lambda \partial \phi) x$$

$$x \in (I - \lambda \partial \phi)^{-1} u$$

but x is the unique minimizer

$$\Longrightarrow \quad x = (I - \lambda \partial \phi)^{-1} u$$

$$\Longrightarrow \quad \mathrm{prox}_{\lambda \phi}(u) = (I - \lambda \partial \phi)^{-1} u$$

Now Let's Revisit: $H \neq I$, $\phi = \|\cdot\|_1$

$$z^0 = \operatorname{amin} \underbrace{\tfrac{1}{2}\| c - Az \|_2^2}_{f(z)} + \underbrace{\lambda \phi(z)}_{g(z)}$$

$$0 \in \eta \partial \big(f(z^0) + g(z) \big)$$

$$0 \in \eta \nabla f(z^0) + \eta \partial g(z)$$

$$0 \in \eta \nabla f(z^0) - z^0 + z^0 + \eta \partial g(z^0)$$

$$(I - \eta \nabla f)(z^0) \in (I + \eta \partial g)(z^0)$$

$$z^0 \in (I + \eta\partial g)^{-1}(I - \eta\nabla f)(z^0)$$

$$z^1 = \mathrm{prox}_{\eta g}\big(\underbrace{z^0 - \eta\nabla f(z^0)}_{\text{Gradient Descent on } f}\big)$$

proximal update w/ g.

Proximal Gradient Descent.

When $H = I$, $\phi(x) = \|Wx\|_1$

↳ recast as a wavelet coeff recovery problem:

$$x^{\square} = W^T z^{\square} \leftarrow \text{IWT}$$

$$z^{\square} = \underset{z}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|c - Az\|_2^2}_{f} + \underbrace{\lambda\|z\|_1}_{g}$$

$$\nabla_z f = A^T(Az - c)$$

$$\operatorname{prox}_{\eta g}(u) = \operatorname{prox}_{\eta\lambda\|\cdot\|}(u) = ST_{\lambda}(u)$$

iterative

$$\Rightarrow \quad z^{k+1} = ST_{\eta t}\left(z^k - \eta A^T(Az - c)\right)$$

<u>ISTA!</u>

<u>More Generally</u> we can solve

any $x^* = \underset{x}{\text{amin}}\ f(x) + g(x)$

where f smooth and we have

expression for $\text{prox}_{\eta g}$

(

↳ this becomes difficult when $g(x)$ is not separable,

ex. when our transform domain isn't orthogonal.

ex. $x^{\Pi} = \underset{x}{\mathrm{argmin}} \ \frac{1}{2} \| y - Hx \|^2 + \phi(Dx)$

⟹ ADMM:

t

Proximal Method that Solves

$$x^k, z^k = \underset{x, z}{\operatorname{argmin}} \; f(x) + g(z)$$

$$\text{s.t.} \quad Ax + Bz = c$$

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \; f(x^k) + \frac{\rho}{2} \| Ax^k + Bz^k - c + u^k \|^2$$

$$z^{k+1} = \operatorname{argmin} \; g(z^{k+1}) \; \ldots \; -$$

$$w^{k+1} =$$

Primal Dual Splitting

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

$$\min_x f(x) \max_z z^T Dx - g^*(z)$$

$$\underbrace{\max_z z^T Dx - g^*(z)}_{\text{fenchel conjugat}}$$

$$\min_x \max_z f(x) + z^T Dx - g^*(z)$$

smooth

non smooth?

non smooth?

Do PGD step on x

Do PGA step on z

Repeat

$$x^{k+1} = \text{prox}_{\tau f}\left(x^k - \tau D^T z^k\right)$$

$$z^{k+1} = \text{prox}_{\sigma g^*}\left(z^k + \sigma D x^{k+1}\right)$$

where we use moreau's identity:

$$\left[\, v = \text{prox}_{g^*}(v) + \text{prox}_{g}(v) \,\right]$$

$$\left[g(y) = \max_z \; y^T z - g^*(z) \right)$$