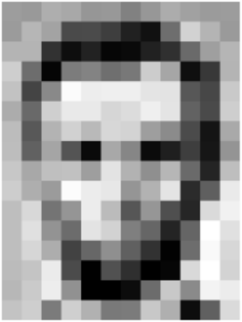
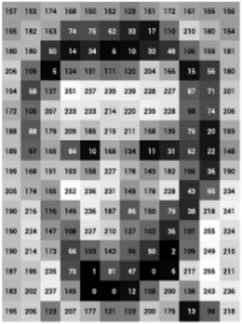
Convex optimization in image/signal processing

L. M. Briceño-Arias
Universidad Técnica Federico Santa María

FGV 2023

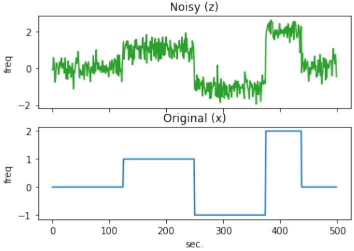
Signals and images

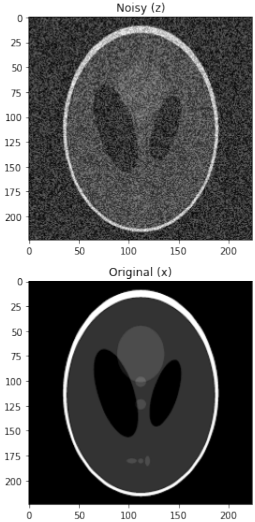




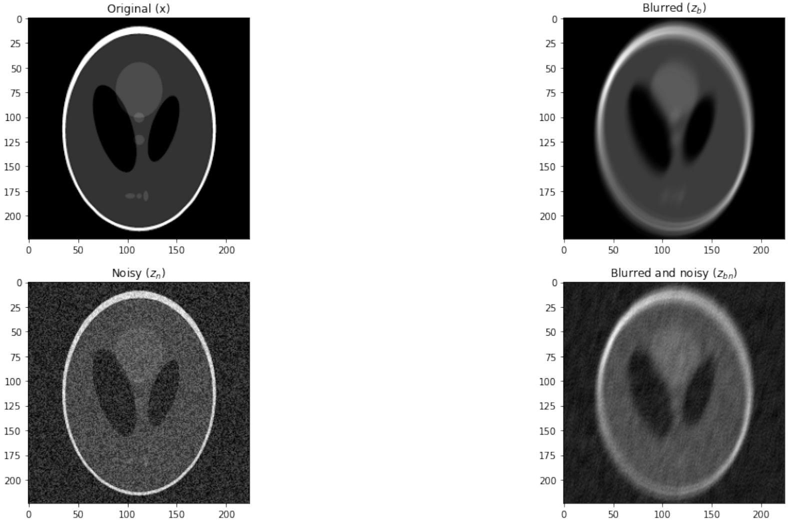
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Denoising


Noisy (z)
freq
2
0
−2
Original (x)
freq
2
1
0
−1
0
100
200
300
400
500
sec.


Noisy (z)
0
25
50
75
100
125
150
175
200
0
50
100
150
200
Original (x)
0
25
50
75
100
125
150
175
200
0
50
100
150
200

Denoising and/or Deblurring



Denoising models

- N: dimension (signal) or pixels (image $N = m \times n$).
- $z \in \mathbb{R}^N$: observed image/signal (noisy).
- $\overline{x} \in \mathbb{R}^N$: image/signal to recover.
- $\varepsilon \in \mathbb{R}^N$: additive noise (random gaussian variable).

$$z = \overline{x} + \varepsilon$$

Denoising optimization problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - z\|_2^2 + \mathcal{R}(x)$$

- $\|x\|_2^2 = \sum_{i=1}^N |x_i|^2$.
- $\mathcal{R} \colon \mathbb{R}^N \to \,]-\infty, +\infty]$: regularization term depending on z.
- If $\mathcal{R} = 0$, the unique solution is $x = z$.

Deblurring/denoising models

- N: dimension (signal) or pixels (image $N = m \times n$).
- Φ: $N \times N$ blur matrix.
- $z \in \mathbb{R}^N$: observed image/signal (noisy).
- $\overline{x} \in \mathbb{R}^N$: image/signal to recover.
- $\varepsilon \in \mathbb{R}^N$: additive noise (random Gaussian variable).

$$z = \Phi\overline{x} + \varepsilon$$

Deblurring/denoising optimization problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|\Phi x - z\|_2^2 + \mathcal{R}(x)$$

- $\mathcal{R}\colon \mathbb{R}^N \to \,]-\infty, +\infty]$: regularization term depending on z.
- If $\mathcal{R} = 0$ and Φ is invertible, the unique solution is $x = \Phi^{-1}z$.

Discrete gradient in regularizations

- In the case of signals $x \in \mathbb{R}^N$, the discrete gradient is

$$D = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \quad Dx = \begin{pmatrix} 0 \\ x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_N - x_{N-1} \end{pmatrix} \in \mathbb{R}^N.$$

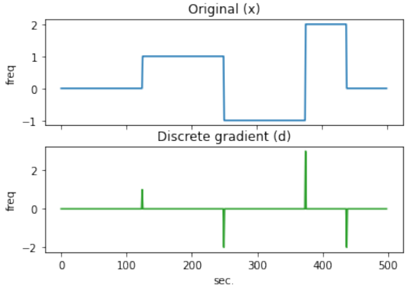
- The discrete gradient operator in the case of images is

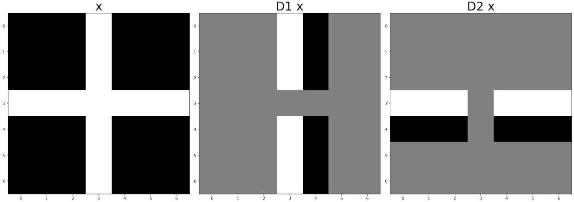
$$D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix},$$

where D_1 and D_2 are $N \times N$ real matrices considering horizontal and vertical differences at each pixel, respectively.

Discrete gradient in regularizations

Python. Dgrad.





Regularizations and main problem

Given $\lambda > 0$, in this lecture we will study images/signals whose Dx is small (piecewise constant).[1]

- $\mathcal{R} = \lambda\|D\cdot\|_1$: TV-$\ell_1$ regularization. $\Rightarrow$ nonsmooth convex
- $\mathcal{R} = \lambda\|D\cdot\|_2^2$: TV-$\ell_2$ regularization. $\Rightarrow$ smooth convex

$$\mathcal{R} = g \circ D, \qquad \text{where} \quad g \in \{\lambda\|\cdot\|_1, \lambda\|\cdot\|_2^2\}$$

Main problem

$$\min_{x\in\mathbb{R}^N} F(x) = \frac{1}{2}\|Ax - z\|_2^2 + g(Dx)$$

- $A = \mathrm{Id}$ if denoising. $\Rightarrow$ strongly convex
- $A = \Phi$ if deblurring. $\Rightarrow$ convex

[1] $\|x\|_1 = \sum_{i=1}^N |x_i|$.

1. Motivation
2. Part I: Convex functions in image/signal processing
3. Part II: TV-ℓ_2 regularization: smooth convex functions
4. Part III: TV-ℓ_1 regularization: nonsmooth convex functions

Part I: Convex functions in image/signal processing

1. Existence and uniqueness
2. F smooth: gradient
   - F convex
   - F strongly convex
3. F nonsmooth: subdifferential
   - Proximity operator
   - Examples

Convex optimization problems

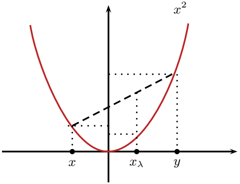
Problem (P)

$$\min_{x\in\mathbb{R}^N} F(x).$$

- $F\colon \mathbb{R}^N \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is
  - **convex :** $(\forall x, y \in \mathbb{R}^N)(\forall \lambda \in [0,1])$ $F(x+\lambda(y-x)) \leq F(x) + \lambda(F(y) - F(x))$.
  - **lower semicontinuous (l.s.c):** $(\forall \gamma > \inf F) \quad \{x \in \mathbb{R}^N \mid F(x) \leq \gamma\}$ is closed.
  - **proper:** F is not always $+\infty$ and never $-\infty$.
- $\Gamma_0(\mathbb{R}^N)$: Class of functions satisfying above conditions.

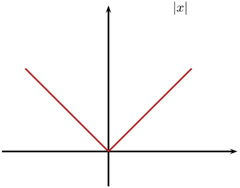
Examples of functions in $\Gamma_0(\mathbb{R}^N)$

- **Differentiable convex functions:** $x \mapsto e^x$, $x \mapsto \|x\|_2^2$,...



$$F(\underbrace{x + \lambda(y - x)}_{x_\lambda}) \leq F(x) + \lambda(F(y) - F(x))$$

- **Non-smooth convex functions:** $x \mapsto |x|$, $x \mapsto \max\{0, x\}$, $x \mapsto \|x\|_1$...

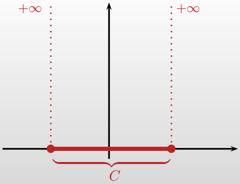


Examples of functions in $\Gamma_0(\mathbb{R}^N)$

- **Discontinuous convex functions:** $C \subset \mathbb{R}^N$ is closed and convex.

Indicator function

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases}$$



- **Constrained convex functions:** Let $f \in \Gamma_0(\mathbb{R}^N)$ and C be closed and convex:

$$F(x) = \begin{cases} f(x), & \text{if } x \in C \\ +\infty, & \text{otherwise.} \end{cases} = f(x) + \iota_C(x)$$

Examples of functions in $\Gamma_0(\mathbb{R}^N)$

- If f and g are in $\Gamma_0(\mathbb{R}^N)$, we have $f+g \in \Gamma_0(\mathbb{R}^N)$.
- If $g \in \Gamma_0(\mathbb{R}^M)$, and $L\colon \mathbb{R}^N \to \mathbb{R}^M$ is linear ($M \times N$ matrix), then $g \circ L \in \Gamma_0(\mathbb{R}^N)$. Dem.
- In particular,

$$\|x\|_1 = |x_1|+|x_2|+\cdots+|x_N| \quad \text{and} \quad \|x\|_2^2 = x_1^2+x_2^2+\cdots+x_N^2$$

are in $\Gamma_0(\mathbb{R}^N)$ Dem. $N=1$.

Exercise

$$(\forall \lambda \in [0,1]) \quad \|(1-\lambda)x+\lambda y\|_2^2 = (1-\lambda)\|x\|_2^2+\lambda\|y\|_2^2-\lambda(1-\lambda)\|x-y\|_2^2.$$

- $\mathcal{R} = g \circ L$ and $f\colon x \mapsto \frac{1}{2}\|Ax-z\|_2^2$ are in $\Gamma_0(\mathbb{R}^N)$.
- $F\colon x \mapsto f(x)+g(Lx)$ is also in $\Gamma_0(\mathbb{R}^N)$.

Existence of solutions

Let $F \in \Gamma_0(\mathbb{R}^N)$ be coercive, i.e.,

$$\lim_{\|x\|\to+\infty} F(x) = +\infty.$$

Then $\arg\min F = \big\{x^* \in \mathbb{R}^N \mid (\forall x \in \mathbb{R}^N)\ F(x^*) \leq F(x)\big\} \neq \varnothing.$

Dem. We have

$$\arg\min F = \bigcap_{\gamma > \inf F} \big\{x \in \mathbb{R}^N \mid F(x) \leq \gamma\big\}$$

and coercivity implies that $\big\{x \in \mathbb{R}^N \mid F(x) \leq \gamma\big\}$ is bounded (and closed). Intersection of nested nonempty compact sets is nonempty. $\square$

Strong convexity and uniqueness of solutions

- F is β-strongly convex for some $\beta > 0$ if $F - \frac{\beta}{2}\|\cdot\|_2^2$ is convex.
- Strongly convex functions are coercive.

Exercise

F is β-strongly convex $\Leftrightarrow$ $(\forall x, y \in \mathbb{R}^N)(\forall \lambda \in [0,1])$
$$F(x + \lambda(y - x)) \leq F(x) + \lambda(F(y) - F(x)) - \tfrac{\beta}{2}\lambda(1-\lambda)\|x - y\|_2^2.$$

Uniqueness of solutions

Suppose that $F \in \Gamma_0(\mathbb{R}^N)$ is β-strongly convex $(\beta > 0)$. Then $\arg\min F$ is a singleton.

Dem. Existence is ok. Suppose that $\{x^*, y^*\} \subset \arg\min F$, $x^* \neq y^*$.

$$F(x^* + \lambda(y^* - x^*)) \leq F(x^*) + \lambda(F(y^*) - F(x^*)) - \frac{\beta}{2}\lambda(1-\lambda)\|x^* - y^*\|_2^2$$
$$< F(x^*) = F(y^*) \Rightarrow\Leftarrow \quad \square.$$

F differentiable

- For all $i \in \{1, \ldots, N\}$, let $e^i = (0, \ldots, \underbrace{1}_{i}, \ldots, 0)^\top \in \mathbb{R}^N$.
- For all $x \in \mathbb{R}^N$ and $i \in \{1, \ldots, N\}$,

$$\frac{\partial F(x)}{\partial x_i} = \lim_{t \to 0} \frac{F(x + te^i) - F(x)}{t} \text{ when the limit exists.}$$

- $\nabla F(x) = \left(\frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial F(x)}{\partial x_N}\right)^\top \in \mathbb{R}^N$: gradient of F at x.
- F is differentiable: $\frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial F(x)}{\partial x_N}$ are continuous.

Exercises

- Prove that $F\colon x \mapsto \|x\|_2^2/2$ is differentiable and $\nabla F(x) = x$.
- Suppose that F is differentiable. Prove that

$$(\forall h \in \mathbb{R}^N) \quad \lim_{t\to 0^+} \frac{F(x+th) - F(x)}{t} = \nabla F(x)^\top h.$$

- Suppose that F is differentiable and let $x^* \in \arg\min F$. Prove that $\nabla F(x^*) = 0$.
- (Chain's rule) Suppose that
  - $f\colon \mathbb{R}^N \to \mathbb{R}$ and $g\colon \mathbb{R}^M \to \mathbb{R}$ are differentiable.
  - L is a $M \times N$ real matrix.
  - $F = f + g \circ L$.

  Prove that F is differentiable and

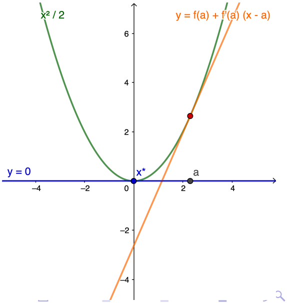
$$(\forall x \in \mathbb{R}^N) \quad \nabla F(x) = \nabla f(x) + L^\top \nabla g(Lx).$$

F differentiable and convex

$$(\forall x \in \mathbb{R}^N)(\forall y \in \mathbb{R}^N) \quad \begin{cases} F(x) + \nabla F(x)^\top (y-x) \leq F(y) \\ \big(\nabla F(x) - \nabla F(y)\big)^\top (x-y) \geq 0. \end{cases}$$

Dem. Let x and y in $\mathbb{R}^N$. From convexity, we have, for every $\lambda \in [0,1]$,

$$\underbrace{\frac{F(x+\lambda(y-x)) - F(x)}{\lambda}}_{\to \nabla F(x)^\top (y-x)} \leq F(y) - F(x). \quad \square$$



Previous property implies

Fermat's Theorem (diff)

$$x^* \in \arg\min F \quad \Leftrightarrow \quad \nabla F(x^*) = 0$$

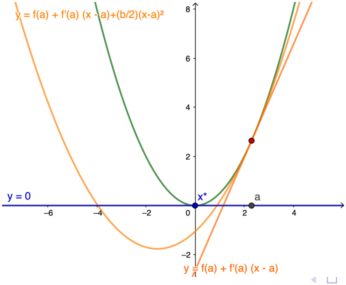
F differentiable and strongly convex

Since $F - \frac{\beta}{2}\|\cdot\|_2^2$ is convex and differentiable:

Exercise

Suppose that F is differentiable. Prove that F is β-strongly convex $\Leftrightarrow$

$$(\forall x \in \mathbb{R}^N)(\forall y \in \mathbb{R}^N) \quad \begin{cases} F(x) + \nabla F(x)^\top (y - x) + \frac{\beta}{2}\|x - y\|_2^2 \leq F(y) \\ \big(\nabla F(x) - \nabla F(y)\big)^\top (x - y) \geq \beta \|x - y\|_2^2 \end{cases}$$



Example

- Suppose that $\overline{x} = 0 \in \mathbb{R}^N$ is the signal to be recovered.
- $z = \overline{x} + \varepsilon = \varepsilon \in \mathbb{R}^N$ is a Gaussian noise (known).
- $\lambda \geq 0$

Denoising with ℓ_2 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \frac{1}{2}\|x - z\|_2^2 + \lambda\|x\|_2^2$$

$F \in \Gamma_0(\mathbb{R}^N)$ and is $(\lambda + 1/2)$-strongly convex. Therefore, there exists a unique solution x^*. By Fermat's Theorem (diff)

$$\begin{aligned} \{x^*\} = \arg\min F \ &\Leftrightarrow \ 0 = \nabla F(x^*) = x^* - z + \lambda x^* \\ &\Leftrightarrow \ x^* = \frac{z}{1+\lambda}. \end{aligned}$$

Python. Signal denoising.

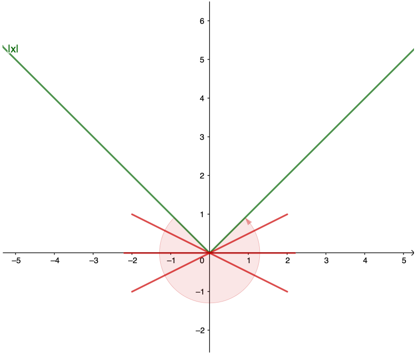
F convex nonsmooth

Subdifferential of F

$$\partial F\colon \mathbb{R}^N \to 2^{\mathbb{R}^N} = \mathcal{P}(\mathbb{R}^N)$$
$$x \mapsto \left\{u \in \mathbb{R}^N \mid (\forall y \in \mathbb{R}^N)\ F(x) + u^\top (y-x) \le F(y)\right\}$$

Example: $|x| = \begin{cases} x, & \text{if } x \ge 0; \\ -x & \text{if } x < 0. \end{cases}$

$$\partial|\cdot|\colon x \mapsto \begin{cases} \{1\}, & \text{if } x > 0; \\ [-1,1], & \text{if } x = 0; \\ \{-1\}, & \text{if } x < 0. \end{cases}$$



Subdifferential properties

Exercise

If F is differentiable, then, for every $x \in \mathbb{R}^N$, $\partial F(x) = \{\nabla F(x)\}$.

Dem. Let $u \in \partial F(x)$, let $h \in \mathbb{R}^N$, and let $t > 0$. Set $y = x + th$

$$u^\top h \leq \frac{F(x+th) - F(x)}{t} \to \nabla F(x)^\top h \Rightarrow (u - \nabla F(x))^\top h \leq 0. \quad \square$$

Monotonicity of ∂F

For every x and y in $\mathbb{R}^N$, $u \in \partial F(x)$, and $v \in \partial F(y)$,

$$(u - v)^\top (x - y) \geq 0.$$

When F is differentiable, it reduces to monotonicity of the gradient.

F convex

- If F is convex and non necessarily differentiable, we have

Fermat's Theorem

$$x \in \arg\min F \quad \Leftrightarrow \quad 0 \in \partial F(x)$$

Dem. $0 \in \partial F(x) \;\Leftrightarrow\; (\forall y \in \mathbb{R}^N) \quad F(x) + 0^\top (y-x) \leq F(y)$ $\square$.

Moreau-Rockafellar's Theorem

Suppose that $g\colon \mathbb{R}^M \to \mathbb{R}$ is continuous, L is a $M \times N$ real matrix, and set $F = f + g \circ L$. Then[a]

$$(\forall x \in \mathbb{R}^N) \quad \partial F(x) = \partial f(x) + L^\top \partial g(Lx).$$

[a] $L(C) = \{Lx \mid x \in C\}$.

Dem. Hahn-Banach's Theorem.

F convex nonsmooth

If $F = f + g \circ L$ and f and g are differentiable, then F is differentiable and Moreau-Rockafellar's Theorem becomes

$$(\forall x \in \mathbb{R}^N) \quad \{\nabla F(x)\} = \{\nabla f(x)\} + L^\top \{\nabla g(Lx)\},$$

which is equivalent to chain's rule.

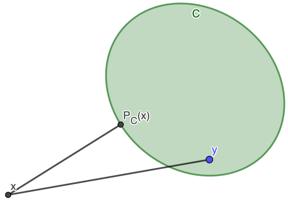
Proximity operator

Suppose that $F \in \Gamma_0(\mathbb{R}^N)$.

Proximity operator of F

$$\operatorname{prox}_F \colon x \mapsto \underset{y\in\mathbb{R}^N}{\operatorname{argmin}} \; F(y) + \frac{1}{2}\|y-x\|_2^2$$

Example: $F = \iota_C$



Projection

$$\operatorname{prox}_{\iota_C} = P_C x = \underset{y\in C}{\operatorname{argmin}} \; \frac{1}{2}\|y-x\|_2^2.$$

Proximity operator

- Since $F + \|\cdot - x\|_2^2/2 \in \Gamma_0(\mathbb{R}^N)$ is strongly convex, $\arg\min(F + \|\cdot - x\|_2^2/2) = \{p^*\}$ and prox_F is well defined.
- By Fermat's and Moreau-Rockafellar's Theorems:

$$0 \in \partial(F + \|\cdot - x\|_2^2/2)(p^*) = \partial F(p^*) + \{p^* - x\}$$

- Then $p^* = \mathrm{prox}_F(x)$ is the unique solution to the inclusion

$$x \in p^* + \partial F(p^*) = (\mathrm{Id} + \partial F)(p^*)$$

or, equivalently,

$$\boxed{\mathrm{prox}_F(x) = (\mathrm{Id} + \partial F)^{-1}(x).}$$

Example: smooth thresholder

Set $\lambda > 0$ and $F = \lambda|\cdot|$.

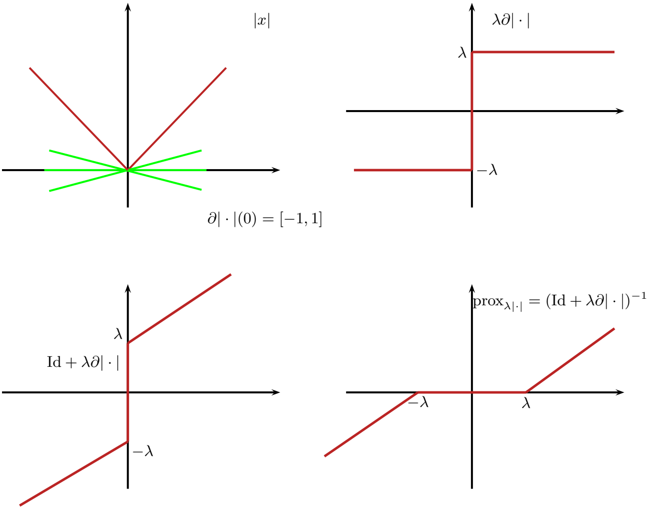
Proximity operator of $\lambda|\cdot|$

$$p = \mathrm{prox}_{\lambda|\cdot|}x \quad \Leftrightarrow \quad x - p \in \lambda\partial|\cdot|(p) = \begin{cases} \{\lambda\}, & \text{if } p > 0; \\ [-\lambda, \lambda], & \text{if } p = 0; \\ \{-\lambda\}, & \text{if } p < 0. \end{cases}$$

- If $p > 0$, then $x - p = \lambda$ and $p = x - \lambda > 0$.
- If $p < 0$, then $x - p = -\lambda$ and $p = x + \lambda < 0$.
- If $p = 0$, then $x - p = x \in [-\lambda, \lambda]$.

$$\boxed{\mathrm{prox}_{\lambda|\cdot|}x = \begin{cases} x - \lambda, & \text{if } x > \lambda; \\ 0, & \text{if } x \in [-\lambda, \lambda]; \\ x + \lambda, & \text{if } x < -\lambda \end{cases} = \mathrm{sign}(x)\max\{0, |x| - \lambda\}.}$$

Example: smooth thresholder



Exercise

For every $i \in \{1, \dots, N\}$, let $f_i \in \Gamma_0(\mathbb{R})$. Define $F \colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ as

$$(\forall x = (x_i)_{1 \leq i \leq N} \in \mathbb{R}^N) \quad F(x) = \sum_{i=1}^{N} f_i(x_i).$$

Prove that

$$(\forall x = (x_i)_{1 \leq i \leq N} \in \mathbb{R}^N) \quad \mathrm{prox}_F x = (\mathrm{prox}_{f_i} x_i)_{1 \leq i \leq N}.$$

Example

- Suppose that $\overline{x} = 0 \in \mathbb{R}^N$ is the signal to be recovered (sparse).
- $z = \overline{x} + \varepsilon = \varepsilon \in \mathbb{R}^N$ is a Gaussian noise (known).
- $\lambda \geq 0$

Denoising with ℓ_1 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \frac{1}{2}\|x - z\|_2^2 + \lambda\|x\|_1$$

Since F is strongly convex, there exists a unique solution x^*. By Fermat's Theorem and $\lambda\|\cdot\|_1 : x \mapsto \sum_{i=1}^N \lambda|x_i|$, we have

$$\{x^*\} = \arg\min F \quad \Leftrightarrow \quad x^* = \mathrm{prox}_{\lambda\|\cdot\|_1} z = (\mathrm{prox}_{\lambda|\cdot|} z_i)_{1 \leq i \leq N}.$$

Python. Signal denoising.

1. Motivation
2. Part I: Convex functions in image/signal processing
3. Part II: TV-ℓ_2 regularization: smooth convex functions
4. Part III: TV-ℓ_1 regularization: nonsmooth convex functions

Part II: TV-ℓ_2 regularization: smooth convex functions.

1. TV-ℓ_2 denoising
    - Problem
    - Fixed point theory: Banach-Picard's theorem and strict contractions
    - Application: Gradient algorithm
2. TV-ℓ_2 deblurring
    - Problem
    - Fixed point theory: Opial's Lemma and averaged nonexpansive ops.
    - Application: Gradient algorithm
3. Appendix

TV-ℓ_2 denoising

- Suppose that $\overline{x} \in \mathbb{R}^N$ is the signal/image to be recovered (piecewise constant).
- $z = \overline{x} + \varepsilon$, where ε is a Gaussian noise, and $\lambda > 0$.

Denoising with TV-ℓ_2 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \frac{1}{2}\|x - z\|_2^2 + \lambda \|Dx\|_2^2$$

Since F is differentiable and 1-strongly convex, Fermat's theorem and chain's rule yield

$$\begin{aligned}
\{x^*\} = \arg\min F \ &\Leftrightarrow\ 0 = x^* - z + \lambda D^\top D x^* = \nabla F(x^*) \\
&\Leftrightarrow\ z = (\mathrm{Id} + \lambda D^\top D) x^* \\
&\Leftrightarrow\ x^* = (\mathrm{Id} + \lambda D^\top D)^{-1} z
\end{aligned}$$

$(\mathrm{Id} + \lambda D^\top D)^{-1}$ costly if N large (Python. Signal denoising.)
$\Rightarrow$ Algorithms !

From Fermat to fixed points

- F differentiable and convex and $\gamma > 0$.

$$\begin{aligned} x^* \in \arg\min F \quad &\Leftrightarrow \quad 0 = \nabla F(x^*) \\ &\Leftrightarrow \quad x^* = x^* - \gamma \nabla F(x^*) \\ &\Leftrightarrow \quad x^* = G_{\gamma F} x^* \end{aligned}$$

Gradient operator

$$G_{\gamma F} x = x - \gamma \nabla F(x).$$

Banach-Picard's theorem

Definition

Let $L > 0$ and let $T\colon \mathbb{R}^N \to \mathbb{R}^N$.

- T is L-Lipschitz continuous if $(\forall x, y \in \mathbb{R}^N)\ \|Tx - Ty\|_2 \leq L\|x - y\|_2$.
- T is a strict contraction if it is L-Lipschitz with $L \in \,]0, 1[$.
- $\operatorname{Fix} T = \big\{x \in \mathbb{R}^N \ \big|\ x = Tx\big\}$: fixed points of T.

Banach-Picard's theorem

- Suppose that $T\colon \mathbb{R}^N \to \mathbb{R}^N$ is a strict contraction with constant $L \in [0, 1[$.
- Let $x_0 \in \mathbb{R}^N$ and $(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n$.

Then $\operatorname{Fix} T = \{x^*\}$ and $(\forall n \in \mathbb{N})\ \|x_n - x^*\|_2 \leq L^n\|x_0 - x^*\|_2$.
Hence, $x_n \to x^*$ with linear convergence rate L.

Proof of Banach-Picard's Theorem

Dem. For every $m > n$,

$$\begin{aligned}\|x_m - x_n\|_2 &\leq \|x_m - x_{m-1}\|_2 + \cdots + \|x_{n+1} - x_n\|_2\\ &\leq \|Tx_{m-1} - Tx_{m-2}\|_2 + \cdots + \|Tx_n - Tx_{n-1}\|_2\\ &\leq (L^{m-2} + \cdots + L^{n-1})\|x_1 - x_0\|_2\\ &= (L^{n-1} - L^{m-1})/(1-L)\|x_1 - x_0\|_2 \to 0, m, n \to +\infty\end{aligned}$$

- $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, hence it converges to x^* and

$$\begin{aligned}\|x^* - Tx^*\|_2 &\leq \|x^* - x_n\|_2 + \|Tx_{n-1} - Tx^*\|_2\\ &\leq \|x^* - x_n\|_2 + \|x_{n-1} - x^*\|_2 \to 0.\end{aligned}$$

- Then $x^* \in \operatorname{Fix} T$. Uniqueness (exercise).
- $\|x_n - x^*\|_2 = \|Tx_{n-1} - Tx^*\|_2 \leq L\|x_{n-1} - x^*\|_2 \leq \cdots \leq L^n\|x_0 - x^*\|_2$. $\square$

$G_{\gamma F}$ strict contraction

Theorem

Suppose that

- F is differentiable, ρ-strongly convex,
- ∇F is L-Lipschitz continuous,
- $0 < \gamma < \frac{2}{L}$.

Then, $G_{\gamma F}$ is $r_G(\gamma)$-strict contraction, where

$$r_G(\gamma) = \max\{|1-\gamma\rho|, |1-\gamma L|\} \in \left]0,1\right[.$$

Moreover, $r_G(\frac{2}{L+\rho}) = \min_{\gamma>0} r_G(\gamma) = \frac{L-\rho}{L+\rho}$.

Dem. See Appendix.

Application: TV-ℓ_2 denoising

Denoising with TV-ℓ_2 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \frac{1}{2}\|x-z\|_2^2 + \lambda\|Dx\|_2^2$$

- $\nabla F\colon x \mapsto x - z + \lambda D^\top Dx$ is $(1+\lambda\|D\|_2^2)$-Lipschitz cont.[2]
- F is 1-strongly convex.
- If $\gamma \in \left]0, 2/(2+\lambda\|D\|_2^2)\right[$, $G_{\gamma F} = \mathrm{Id} - \gamma\nabla F$ is a $\underbrace{\max\{|1-\gamma|, |1-\gamma(1+\lambda\|D\|_2^2)|\}}_{r_G(\gamma)}$- strict contraction.

Gradient algorithm

$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma(x_n - z + \lambda D^\top Dx_n).$
Then $\|x_n - x^*\| \leq r_G(\gamma)^n\|x_0 - x^*\|$.

Python. Signal-TV-l2-grad-theo-vs-num & Image#

[2] $\|D\|_2 = \max_{\|x\|=1}\|Dx\|_2 = \sqrt{\lambda_{\max}(D^\top D)}$.

TV-ℓ_2 deblurring

- Suppose that $\overline{x} \in \mathbb{R}^N$ is the signal/image to be recovered (piecewise constant).
- $z = \Phi\overline{x} + \varepsilon$, where $\varepsilon \in \mathbb{R}^N$ is a Gaussian noise and Φ is a blur operator.

Deblurring with TV-ℓ_2 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \frac{1}{2}\|\Phi x - z\|_2^2 + \lambda\|Dx\|_2^2$$

Since F is differentiable and convex (not strongly convex unless Φ injective), Fermat's theorem and chain's rule yield

$$\begin{aligned} x^* \in \arg\min F \;&\Leftrightarrow\; 0 = \Phi^\top(\Phi x^* - z) + \lambda D^\top D x^* = \nabla F(x^*) \\ &\Leftrightarrow\; \Phi^\top z = (\Phi^\top\Phi + \lambda D^\top D)x^* \\ &\Leftrightarrow\; x^* = (\Phi^\top\Phi + \lambda D^\top D)^{-1}\Phi^\top z \end{aligned}$$

$(\Phi^\top\Phi + \lambda D^\top D)^{-1}$ difficult/costly $\Rightarrow$ Algorithms !

Gradient operator without strong conv.

- Since F is differentiable and convex, we already know that

$$x^* \in \arg\min F \quad \Leftrightarrow \quad x^* \in \operatorname{Fix} G_{\gamma F} = (\mathrm{Id} - \gamma \nabla F)$$

- Since F is not strongly convex, $G_{\gamma F}$ is no longer a strict contraction and Banach-Picard's theorem does not guarantee the convergence of the gradient method.
- Still, we have (see Appendix)

Baillon-Haddad (1977)

- $F\colon \mathbb{R}^N \to \mathbb{R}$ be convex differentiable.
- ∇F is L-Lipschitz continuous ($L > 0$).

Then, for all x and y in $\mathbb{R}^N$,
$(\nabla F(x) - \nabla F(y))^\top (x - y) \geq \frac{1}{L} \|\nabla F(x) - \nabla F(y)\|^2$.

Gradient operator without strong conv.

$$\begin{aligned}
\|G_{\gamma F}x - G_{\gamma F}y\|_2^2 &= \|x-y\|_2^2 + \gamma^2\|\nabla F(x) - \nabla F(y)\|_2^2 \\
&\qquad - 2\gamma(\nabla F(x) - \nabla F(y))^\top (x-y) \\
&\leq \|x-y\|_2^2 - \gamma\left(\frac{2}{L} - \gamma\right)\|\nabla F(x) - \nabla F(y)\|_2^2 \\
&= \|x-y\|_2^2 - \left(\frac{1-\frac{L\gamma}{2}}{\frac{L\gamma}{2}}\right)\|\underbrace{\gamma\nabla F}_{\mathrm{Id}-G_{\gamma F}}(x) - \underbrace{\gamma\nabla F}_{\mathrm{Id}-G_{\gamma F}}(y)\|_2^2
\end{aligned}$$

Definition

T is α-averaged nonexpansive ($\alpha \in \,]0,1[$) if for every $x, y \in \mathbb{R}^N$,

$$\|Tx - Ty\|_2^2 \leq \|x-y\|_2^2 - \frac{1-\alpha}{\alpha}\|(\mathrm{Id}-T)x - (\mathrm{Id}-T)y\|_2^2.$$

In particular, T is nonexpansive (i.e., 1-Lipschitz cont.)

- Then $G_{\gamma F}$ is $L\gamma/2$-averaged nonexpansive if $0 < \gamma < 2/L$.

Fixed point convergence: averaged nonexpansive

Theorem

- T is α-averaged nonexpansive with $\operatorname{Fix} T \neq \varnothing$.
- Let $x_0 \in \mathbb{R}^N$ and $(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n$.

Then, there exists $x^* \in \operatorname{Fix} T$ such that $x_n \to x^*$.

Dem.

- For every $x^* \in \operatorname{Fix} T$ and $n \in \mathbb{N}$, the av. nonexpansive property implies

$$\|x_{n+1}-x^*\|_2^2 = \|Tx_n - Tx^*\|_2^2 \leq \|x_n - x^*\|_2^2 - \frac{1-\alpha}{\alpha}\|Tx_n - x_n\|_2^2.$$

- We obtain that $(\|x_n - x^*\|_2)_{n\in\mathbb{N}}$ is decreasing and bounded from below (by 0). Thus,

(1)

$$(\forall x^* \in \operatorname{Fix} T)\ (\|x_n - x^*\|_2)_{n\in\mathbb{N}} \text{ converges.}$$

Fixed point convergence: averaged nonexpansive

- As a by-product we obtain $(1-\alpha)/\alpha\|Tx_n - x_n\|_2^2 \leq \|x_n - x^*\|_2^2 - \|x_{n+1} - x^*\|_2^2$ and

$$\frac{1-\alpha}{\alpha}\sum_{n=0}^{N}\|Tx_n - x_n\|_2^2 \leq \|x_0 - x^*\|_2^2 - \|x_{N+1} - x^*\|_2^2$$

- Hence, the series $\sum_{n\geq 0}\|Tx_n - x_n\|_2^2$ converges implying $x_n - Tx_n \to 0$.
- Since $(x_{n_k})_{n\in\mathbb{N}}$ is bounded, let y^* be any accumulation point, i.e., $x_{n_k} \to y^*$. Since T is 1-Lipschitz,

$$\begin{aligned}
&\|x_{n_k} - y^*\|_2^2 + \|y^* - Ty^*\|_2^2 + 2(y^* - Ty^*)^\top(x_{n_k} - y^*) \\
&\quad = \|x_{n_k} - Ty^*\|_2^2 \\
&\quad = \|x_{n_k} - Tx_{n_k}\|_2^2 + \|Tx_{n_k} - Ty^*\|_2^2 + 2(x_{n_k} - Tx_{n_k})^\top(Tx_{n_k} - Ty^*) \\
&\quad \leq \underbrace{\|x_{n_k} - Tx_{n_k}\|_2^2}_{\to 0} + \|x_{n_k} - y^*\|_2^2 + 2\underbrace{(x_{n_k} - Tx_{n_k})}_{\to 0}{}^\top\underbrace{(Tx_{n_k} - Ty^*)}_{\text{bounded}}
\end{aligned}$$

Fixed point convergence: averaged nonexpansive

(2)

Any accumulation point of $(x_n)_{n\in\mathbb{N}}$ is in $\operatorname{Fix} T$.

- With (1) and (2), we can conclude the uniqueness of the accumulation point.[3]
- Suppose that $x_{n_k} \to x^*$ and $x_{n_m} \to y^*$.
- (2) $\Rightarrow$ x^* and y^* are in $\operatorname{Fix} T$.
- (1) $\Rightarrow$ $\|x_n - x^*\|_2 \to L_1$ and $\|x_n - y^*\|_2 \to L_2$.
- $\underbrace{\|x_n - x^*\|_2^2}_{\to L_1^2} = \underbrace{\|x_n - y^*\|_2^2}_{\to L_2^2} + \|y^* - x^*\|_2^2 + 2(x_n - y^*)^\top (y^* - x^*)$
- $(y^* - x^*)^\top x_n \to L = \frac{1}{2}(L_1^2 - L_2^2 - \|y^* - x^*\|_2^2) + (y^* - x^*)^\top y^*$.
- $(y^* - x^*)^\top x^* \underset{n=n_k}{\leftarrow} (y^* - x^*)^\top x_n \underset{n=n_m}{\rightarrow} (y^* - x^*)^\top y^*$
- $\|y^* - x^*\|_2^2 = 0 \quad \Rightarrow \quad x^* = y^*$. $\square$

[3] The argument is known as Opial's Lemma

Application: TV-ℓ_2 deblurring

Deblurring with TV-ℓ_2 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \frac{1}{2}\|\Phi x - z\|_2^2 + \lambda\|Dx\|_2^2$$

- F convex and differentiable.
- $\nabla F\colon x \mapsto \Phi^\top(\Phi x - z) + \lambda D^\top D x$ is L-Lipschitz continuous, where $L = \|\Phi^\top\Phi + \lambda D^\top D\|_2 \leq \|\Phi\|_2^2 + \lambda\|D\|_2^2$.
- For $\gamma \in \left]0, \frac{2}{L}\right[$, $G_{\gamma F}$ is $\frac{L\gamma}{2}$-averaged nonexpansive.
- $\operatorname{Fix} G_{\gamma F} = \arg\min F \neq \varnothing$. Why ?

Gradient algorithm

$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma\big(\Phi^\top(\Phi x_n - z) + \lambda D^\top D x_n\big)$.
There exists $x^* \in \arg\min F$ such that $x_n \to x^*$.

Python. TV-l2-image-deblurring-l1

Appendix: Enhanced Baillon-Haddad[4]

Theorem

Let $F\colon \mathbb{R}^N \to \mathbb{R}$ be convex and differentiable. Then, the following are equivalent:

1. ∇F is L-Lipschitz continuous.
2. $(\forall x, y \in \mathbb{R}^N) \quad F(x) \leq F(y) + \nabla F(y)^\top (x-y) + \frac{L}{2}\|x-y\|_2^2.$
3. $(\forall x, y \in \mathbb{R}^N) \quad (\nabla F(x) - \nabla F(y))^\top (x-y) \leq L\|x-y\|_2^2.$
4. $(\forall x, y \in \mathbb{R}^N) \quad (\nabla F(x) - \nabla F(y))^\top (x-y) \geq \frac{1}{L}\|\nabla F(x) - \nabla F(y)\|_2^2.$

[4] see, e.g. Bauschke–Combettes (2017)

Appendix: Enhanced Baillon–Haddad ($1 \Rightarrow 2$ & $2 \Leftrightarrow 3$)

Dem. $1 \Rightarrow 2$: Given x and y in $\mathbb{R}^N$, define $\phi \colon t \mapsto F(y+t(x-y))$, which is differentiable, $\phi'(t) = \nabla F(y+t(x-y))^\top(x-y)$ (exercise), $\phi(0) = F(y)$ and $\phi(1) = F(x)$. Hence, from C-S, Lipschitz cont., and FTC we have

$$
\begin{aligned}
F(x) - F(y) &= \int_0^1 \nabla F(y+t(x-y))^\top (x-y)dt \\
&= \int_0^1 \nabla F(y+t(x-y) - \nabla F(y))^\top (x-y)dt + \nabla F(y)^\top (x-y) \\
&\leq \int_0^1 \|\nabla F(y+t(x-y) - \nabla F(y)\|_2 \|x-y\|_2 dt + \nabla F(y)^\top (x-y) \\
&\leq L\|x-y\|_2^2 \int_0^1 tdt + \nabla F(y)^\top (x-y) \\
&= \frac{L}{2}\|x-y\|_2^2 + \nabla F(y)^\top (x-y).
\end{aligned}
$$

Dem. $2 \Leftrightarrow 3$: Change the roles of x and y and sum. exercise: $\Leftarrow$

Appendix: Enhanced Baillon–Haddad ($3 \Rightarrow 4$ & $4 \Rightarrow 1$)

$3 \Rightarrow 4$: Using 2 and convexity we have, for every x, y, z in $\mathbb{R}^N$

$$\begin{aligned} F(x) + \nabla F(x)^\top (z - x) &\leq F(z) \\ F(z) &\leq F(y) + \nabla F(y)^\top (z - y) + \frac{L}{2}\|z - y\|_2^2 \end{aligned}$$

which leads to

$$F(x) + \nabla F(x)^\top (y - x) \leq F(y) + \underbrace{(\nabla F(y) - \nabla F(x))^\top (z - y) + \frac{L}{2}\|z - y\|_2^2}_{\varphi(z)}.$$

Since $\varphi\colon \mathbb{R}^N \to \mathbb{R}$ is strongly convex and differentiable, admits a unique minimizer satisfying (Fermat)

$$0 = \nabla\varphi(z^*) = \nabla F(y) - \nabla F(x) + L(z^* - y) \;\Leftrightarrow\; z^* - y = \frac{1}{L}(\nabla F(x) - \nabla F(y))$$

obtaining $F(x) + \nabla F(x)^\top (y - x) \leq F(y) - \frac{1}{2L}\|\nabla F(x) - \nabla F(y)\|_2^2$. Changing the roles of x and y, the result follows.

$4 \Rightarrow 1$: C-S. $\square$.

$G_{\gamma F}$ strict contraction

Theorem

Suppose that

- F is differentiable, ρ-strongly convex,
- ∇F is L-Lipschitz continuous,
- $0 < \gamma < \frac{2}{L}$.

Then, $G_{\gamma F}$ is $r_G(\gamma)$-strict contraction, where

$$r_G(\gamma) = \max\{|1-\gamma\rho|, |1-\gamma L|\} \in \,]0,1[\,.$$

Moreover, $r_G(\frac{2}{L+\rho}) = \min_{\gamma>0} r_G(\gamma) = \frac{L-\rho}{L+\rho}$.

Appendix: Proof

- F is ρ-strongly convex $\Leftrightarrow$ $h = F - \frac{\rho}{2}\|\cdot\|_2^2$ is convex.
- F differentiable $\Leftrightarrow$ h differentiable and $\nabla h = \nabla F - \rho \mathrm{Id}$.
- ∇F is L-Lipschitz continuous (B-H)

$$\begin{aligned}
&\Leftrightarrow (\nabla F(x) - \nabla F(y))^\top (x-y) \leq L\|x-y\|_2^2 \\
&\Leftrightarrow (\nabla h(x) - \nabla h(y)))^\top (x-y) \leq (L-\rho)\|x-y\|_2^2 \\
&\Leftrightarrow \nabla h \text{ is } (L-\rho)\text{-Lipschitz continuous.}
\end{aligned}$$

For every x and y in $\mathbb{R}^N$,

$$\begin{aligned}
\|G_{\gamma F}x - G_{\gamma F}y\|_2^2 &= \|x - y - \gamma(\nabla F(x) - \nabla F(y))\|_2^2 \\
&= \|(1-\gamma\rho)(x-y) - \gamma(\nabla h(x) - \nabla h(y))\|_2^2 \\
&= (1-\gamma\rho)^2\|x-y\|_2^2 + \gamma^2\|\nabla h(x) - \nabla h(y)\|_2^2 \\
&\qquad - 2\gamma(1-\gamma\rho)(\nabla h(x) - \nabla h(y))^\top (x-y) \\
&\leq (1-\gamma\rho)^2\|x-y\|_2^2 \\
&\qquad + \gamma(\gamma(L+\rho) - 2)(\nabla h(x) - \nabla h(y))^\top (x-y)
\end{aligned}$$

Appendix: Proof

Two cases:

- If $\gamma < \frac{2}{L+\rho}$: Since h is convex and differentiable, ∇h is monotone, i.e., $(\nabla h(x) - \nabla h(y))^\top (x-y) \geq 0$, obtaining

$$\|G_{\gamma F}x - G_{\gamma F}y\|_2^2 \leq \underbrace{(1-\gamma\rho)^2}_{\leq r_G(\gamma)^2} \|x-y\|_2^2.$$

- If $\gamma \geq \frac{2}{L+\rho}$: h is convex and ∇h is $(L-\rho)$-Lipschitz, B-H implies $(\nabla h(x) - \nabla h(y))^\top (x-y) \leq (L-\rho)\|x-y\|^2$, which yields

$$\begin{aligned}\|G_{\gamma F}x - G_{\gamma F}y\|_2^2 &\leq \big((1-\gamma\rho)^2 + \gamma(L-\rho)(\gamma(L+\rho)-2)\big)\|x-y\|^2 \\ &= \underbrace{(1-\gamma L)^2}_{\leq r_G(\gamma)^2} \|x-y\|_2^2. \quad \square\end{aligned}$$

1. Motivation
2. Part I: Convex functions in image/signal processing
3. Part II: TV-ℓ_2 regularization: smooth convex functions
4. Part III: TV-ℓ_1 regularization: nonsmooth convex functions

Part III: TV-ℓ_1 regularization: nonsmooth convex functions.

1. TV-ℓ_1 denoising
   - Problem
   - Douglas-Rachford splitting
   - Application to signals
   - Dual Forward-Backward-Splitting
   - Application to images
2. TV-ℓ_1 deblurring
   - Problem
   - Primal-dual algorithms
   - Application to images
3. Concluding remarks

TV-ℓ_1 denoising

- Suppose that $\overline{x} \in \mathbb{R}^N$ is the signal/image to be recovered.
- $z = \overline{x} + \varepsilon \in \mathbb{R}^N$, where ε is a Gaussian noise.
- $\lambda > 0$

Denoising with TV-ℓ_1 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \frac{1}{2}\|x - z\|_2^2 + \lambda \|Dx\|_1$$

- Non-smooth convex function. Gradient method not available.
- $F \in \Gamma_0(\mathbb{R}^N)$ is 1-strongly convex. Then $\arg\min F = \{x^*\}$.
- Fermat and Moreau-Rockafellar theorems yields

$$0 \in x^* - z + \lambda \partial(\|D \cdot\|_1)(x^*) \quad \Leftrightarrow \quad x^* = \mathrm{prox}_{\lambda\|D\cdot\|_1}(z).$$

- $\mathrm{prox}_{\lambda\|D\cdot\|_1}$ not easy in general !

Recall: proximity operator

Let $F \in \Gamma_0(\mathbb{R}^N)$ and $\gamma > 0$

Proximity operator

$$p = \mathrm{prox}_{\gamma F} x \;\Leftrightarrow\; \frac{x-p}{\gamma} \in \partial F(p)$$

Property

$\mathrm{prox}_{\gamma F}$ is 1/2-averaged nonexpansive

Dem. Let $x, y \in \mathbb{R}^N$, set $p = \mathrm{prox}_{\gamma F} x$, and $q = \mathrm{prox}_{\gamma F} y$. The monotonicity of ∂F implies

$$\begin{aligned} 0 &\leq \left(\frac{x-p}{\gamma} - \frac{y-q}{\gamma}\right)^{\top} (p-q) \\ &= \frac{1}{\gamma}\big((x-y)^{\top}(p-q) - \|p-q\|_2^2\big). \end{aligned}$$

$\mathrm{prox}_{\gamma\|D\cdot\|_1}$ if $DD^\top = \mu\mathrm{Id}$?

- Suppose that $DD^\top = \mu\mathrm{Id}$ for some $\mu > 0$.
- Let $x \in \mathbb{R}^N$. We have (Moreau-Rockafellar)

$$p = \mathrm{prox}_{\gamma\|D\cdot\|_1} x \;\Leftrightarrow\; x - p \in \gamma\partial(\|D\cdot\|_1)(p) = \gamma D^\top \underbrace{\partial\|\cdot\|_1(Dp)}_{u\in}$$

$$\Leftrightarrow\; (\exists\, u \in \partial\|\cdot\|_1(Dp))\;\; p = x - \gamma D^\top u.$$

- In addition, since $DD^\top = \mu\mathrm{Id}$,

$$Dx - Dp = \gamma \underbrace{DD^\top}_{\mu\mathrm{Id}} u \in \gamma \underbrace{DD^\top}_{\mu\mathrm{Id}} \partial\|\cdot\|_1(Dp) \Rightarrow \begin{cases} Dp = \mathrm{prox}_{\gamma\mu\|\cdot\|_1}(Dx) \\ u = \frac{Dx - Dp}{\gamma\mu}. \end{cases}$$

If $DD^\top = \mu\mathrm{Id}$

$$\mathrm{prox}_{\gamma\|D\cdot\|_1} x = x - \frac{1}{\mu} D^\top\left(Dx - \mathrm{prox}_{\gamma\mu\|\cdot\|_1}(Dx)\right)$$

Reformulation

- However, $DD^\top$ is not diagonal. Python. Signal denoising.

Reformulation

- However, $DD^\top$ is not diagonal. Python. Signal denoising.
- In the case of signals: $\|Dx\|_1 = \sum_{i=2}^{N} |x_i - x_{i-1}| = \sum_{j=1}^{N/2} |x_{2j} - x_{2j-1}| + \sum_{j=1}^{N/2} |x_{2j+1} - x_{2j}| = \|D_e x\|_1 + \|D_o x\|_1$, where

$$D_e = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ & \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, D_o = \begin{bmatrix} 0 & 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ & \vdots & & & \ddots & \vdots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 & 0 \end{bmatrix}$$

- $D_e D_e^\top = 2\mathrm{Id}$ and $D_o D_o^\top = 2\mathrm{Id}$.

Signal denoising with TV-ℓ_1 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \underbrace{\frac{1}{2}\|x - z\|_2^2 + \lambda \|D_e x\|_1}_{f(x)} + \underbrace{\lambda \|D_o x\|_1}_{g(x)}$$

Reflections

Let f and g be functions in $\Gamma_0(\mathbb{R}^N)$.

Problem

$$\min_{x\in\mathbb{R}^N} f(x) + g(x)$$

Definition: Reflection operator

$$R_f = 2\mathrm{prox}_f - \mathrm{Id}$$

Proposition

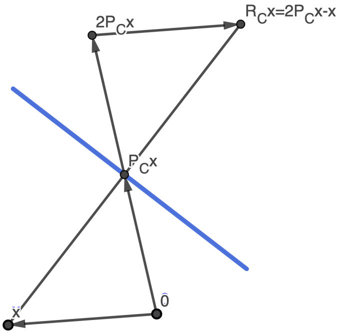
$(\forall\gamma > 0)$ $R_{\gamma f}$ is nonexpansive (1-Lipschitz)

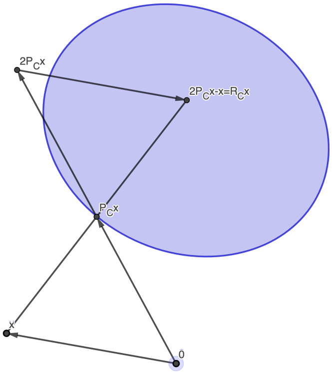
Dem.

$$\begin{aligned}\|R_f x - R_f y\|_2^2 &= 4(\|\mathrm{prox}_f x - \mathrm{prox}_f y\|_2^2 - (\mathrm{prox}_f x - \mathrm{prox}_f y)^\top (x-y)) \\ &\quad + \|x-y\|_2^2 \\ &\leq \|x-y\|_2^2 \quad \square\end{aligned}$$

Reflections

Example: $f = \iota_C$. Then $\mathrm{prox}_f = P_C$ y $R_f = 2P_C - \mathrm{Id}$.





Reflections

Proposition

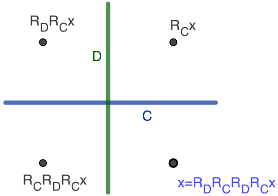
If $z^* \in \operatorname{Fix} R_f R_g$, then $\operatorname{prox}_g z^* \in \arg\min(f+g)$

Dem.

$$
\begin{aligned}
z^* = R_f R_g z^* \quad & \Leftrightarrow \quad z^* = 2\operatorname{prox}_f(2\operatorname{prox}_g z^* - z^*) - 2\operatorname{prox}_g z^* + z^* \\
& \Leftrightarrow \quad \operatorname{prox}_f(2\operatorname{prox}_g z^* - z^*) = \operatorname{prox}_g z^* \\
& \Rightarrow \quad \begin{cases} (2\operatorname{prox}_g z^* - z^*) - \operatorname{prox}_g z^* \in \partial f(\operatorname{prox}_g z^*) \\ z^* - \operatorname{prox}_g z^* \in \partial g(\operatorname{prox}_g z^*) \end{cases} \\
& \Leftrightarrow \quad \begin{cases} \operatorname{prox}_g z^* - z^* \in \partial f(\operatorname{prox}_g z^*) \\ z^* - \operatorname{prox}_g z^* \in \partial g(\operatorname{prox}_g z^*) \end{cases} \\
& \Rightarrow \quad 0 \in \partial f(\operatorname{prox}_g z^*) + \partial g(\operatorname{prox}_g z^*) \\
& \Rightarrow \quad \operatorname{prox}_g z^* \in \arg\min(f+g).
\end{aligned}
$$

Reflections

- Problem: $z_{n+1} = R_f R_g z_n$ does not converge.
- Example: $f = \iota_D$ and $g = \iota_C$.



- $T = R_f R_g$ is (merely) nonexpansive (1-Lipschitz)

T nonexpansive with $\operatorname{Fix} T \neq \varnothing$

- Let $T\colon \mathbb{R}^N \to \mathbb{R}^N$ be nonexpansive.
- $\operatorname{Fix} T \neq \varnothing$.

Then, defining

$$(\forall \alpha \in \,]0,1[) \quad T_\alpha = (1-\alpha)\mathrm{Id} + \alpha T,$$

we have (exercise)

- T_α is α-averaged nonexpansive.
- $\operatorname{Fix} T_\alpha = \operatorname{Fix} T \neq \varnothing$.

Then

Krasnoselskii-Mann (KM)

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = (1-\alpha)z_n + \alpha T z_n,$$

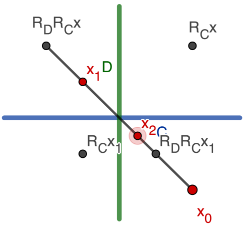
There exists $z^* \in \operatorname{Fix} T$, such that $z_n \to z^*$.

Douglas-Rachford splitting (DRS)

DRS: KM with $T = R_{\gamma f} R_{\gamma g}$ and $\alpha = 1/2$

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = \frac{z_n + R_{\gamma f} R_{\gamma g} z_n}{2}$$
$$= \mathrm{prox}_{\gamma f}(2\mathrm{prox}_{\gamma g} z_n - z_n) + z_n - \mathrm{prox}_{\gamma g} z_n$$

There exists $z^* \in \operatorname{Fix} T$, such that $z_n \to z^*$ and $\mathrm{prox}_{\gamma g} z^* \in \arg\min(f+g)$.



DRS for TV-ℓ_1 signal denoising

Signal denoising with TV-ℓ_1 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \underbrace{\frac{1}{2}\|x-z\|_2^2 + \lambda\|D_e x\|_1}_{f(x)} + \underbrace{\lambda\|D_o x\|_1}_{g(x)}$$

- $f = \frac{1}{2}\|\cdot - z\|_2^2 + \lambda\|D_e\cdot\|_1$ and $g = \lambda\|D_o\cdot\|_1$ are convex nonsmooth.
- Proximity operators simple to compute (Done for g !)

Python. Signal denoising

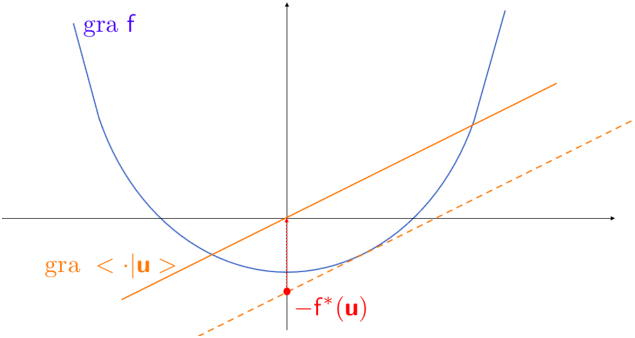
Same argument/formulation is no longer valid for images, since $D = [D1; D2]$...

Fenchel conjugate

Let $f \in \Gamma_0(\mathbb{R}^N)$.

Fenchel conjugate

$$(\forall y \in \mathbb{R}^N) \quad f^*(y) = \sup_{x \in \mathbb{R}^N} \left(x^\top y - f(x)\right)$$



Fenchel-Rockafellar duality

Denoising with TV-ℓ_1 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \underbrace{\frac{1}{2}\|x-z\|_2^2}_{f(x)} + \underbrace{\lambda\|Dx\|_1}_{g(Dx)}$$

- **Primal**: $(P) \quad \min_{x\in\mathbb{R}^N} f(x) + g(Dx)$
- **Dual**: $(D) \quad \min_{u\in\mathbb{R}^N} f^*(-D^\top u) + g^*(u)$
- Both values coincide.
- u^* solves (D), then $x^* = \nabla f^*(-D^\top u^*)$ solves (P).[5]

[5] Proposition 19.4, Bauschke-Combettes (2017)

Dual of TV-ℓ_1 denoising

- $f^*(y) = \frac{1}{2}\|y+z\|_2^2$
- $g^*(u) = \iota_{\lambda B_\infty}$, where $B_\infty = [-1,1]^N$.

Dual TV-ℓ_1 regularization

$$\min_{u\in\mathbb{R}^N} \underbrace{\frac{1}{2}\|-D^\top u + z\|_2^2}_{\varphi(u)} + \underbrace{\iota_{[-\lambda,\lambda]^N}(u)}_{\psi(u)}$$

- $\psi = \iota_{[-\lambda,\lambda]^N}$ convex nonsmooth.
- $\varphi = \frac{1}{2}\|-D^\top\cdot + z\|_2^2$ convex differentiable, and $\nabla\varphi\colon u \mapsto D(D^\top u - z)$ is $\|D\|_2^2$ Lipschitz continuous.
- But no strong convexity... not necessarily unique solution.

Fermat to fixed points

Denoising with TV-ℓ_1 regularization

- φ convex differentiable with L-Lipschitz gradient.
- $\psi \in \Gamma_0(\mathbb{R}^N)$.

$$\min_{u\in\mathbb{R}^N} \varphi(u) + \psi(u)$$

Fermat and M-R imply, for every $\gamma > 0$,

$$\begin{aligned} u^* \in \arg\min(\varphi+\psi) \quad &\Leftrightarrow \quad 0 \in \nabla\varphi(u^*) + \partial\psi(u^*) \\ &\Leftrightarrow \quad u^* - \gamma\nabla\varphi(u^*) \in u^* + \gamma\partial\psi(u^*) \\ &\Leftrightarrow \quad u^* = \operatorname{prox}_{\gamma\psi}\big(u^* - \gamma\nabla\varphi(u^*)\big) \\ &\Leftrightarrow \quad u^* \in \operatorname{Fix} T_{\gamma,\psi,\varphi} \end{aligned}$$

$$\boxed{T_{\gamma,\psi,\varphi} = \operatorname{prox}_{\gamma\psi} \circ G_{\gamma\varphi}}$$

Forward-backward algorithm

- Let $\gamma < 2/L_\varphi$. Then $G_{\gamma\varphi}$ is $\gamma L_\varphi/2$-averaged nonexpansive.
- Let $\gamma > 0$. Then $\operatorname{prox}_{\gamma\psi}$ is $1/2$-averaged nonexpansive.
- If S_1 and S_2 are averaged nonexpansive, respectively, then $T = S_1 \circ S_2$ is averaged nonexpansive for some $\alpha \in \,]0,1[$ (Appendix).

$$T_{\gamma,\psi,\varphi} = \operatorname{prox}_{\gamma\psi} \circ G_{\gamma\varphi} \text{ is averaged nonexpansive.}$$

Forward-backward splitting (FBS)

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma\psi}(x_n - \gamma\nabla\varphi(x_n))$$

Then $x_n \to x^* \in \arg\min(\psi + \varphi)$.

Dual FBS

Dual TV-ℓ_1 regularization

$$\min_{u\in\mathbb{R}^N} \underbrace{\frac{1}{2}\|z - D^\top u\|_2^2}_{\varphi(u)} + \underbrace{\iota_{[-\lambda,\lambda]^N}(u)}_{\psi(u)}$$

Dual FBS

Let $\gamma \in \left[0, \frac{2}{\|D\|_2^2}\right[$.

$$(\forall n \in \mathbb{N}) \quad u_{n+1} = P_{[-\lambda,\lambda]^N}(u_n - \gamma D(D^\top u_n - z))$$

Then $u_n \to u^*$ solution to (D) and $x^* = z - D^\top u^* \in \arg\min F$.

Python. Image-TV-l2-grad-theo-vs-num

TV ℓ_1 deblurring

- Suppose that $\overline{x} \in \mathbb{R}^N$ is the signal to be recovered.
- $z = \Phi\overline{x} + \varepsilon \in \mathbb{R}^N$, where ε is a Gaussian noise and Φ is a blur operator.
- $\lambda > 0$

Deblurring with TV-ℓ_1 regularization

$$\min_{x \in \mathbb{R}^N} F(x) = \frac{1}{2}\|\Phi x - z\|_2^2 + \lambda\|Dx\|_1$$

- Non-smooth convex function. Gradient method not available.
- $F \in \Gamma_0(\mathbb{R}^N)$ not strongly convex unless Φ injective.
- Fermat and Moreau-Rockafellar theorems yields

$$0 \in \Phi^\top(\Phi x^* - z) + \lambda D^\top \partial(\|\cdot\|_1)(Dx^*)$$

- Primal-dual algorithms !

Primal-dual splitting (PDS)

Problem

- $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$
- D is a $M \times N$ real matrix
- h is convex, differentiable, and ∇h is L-Lipschitz cont.

$$(P) \qquad \min_{x \in \mathbb{R}^N} f(x) + g(Dx) + h(x)$$

Condat-Vũ (2013)

- $x_0 \in \mathbb{R}^N$, $u_0 \in \mathbb{R}^M$, and $\sigma \|D\|^2 < \frac{1}{\tau} - \frac{L}{2}$

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + D^\top u_n)) \\ u_{n+1} = \mathrm{prox}_{\sigma g^*}(u_n + \sigma L(2x_{n+1} - x_n)). \end{cases}$$

Then, there exists x^* solution to (P) and u^* solution to (D) such that $x_n \to x^*$ and $u_n \to u^*$.

Application: TV-ℓ_1 deblurring

Deblurring with TV-ℓ_1 regularization

$$\min_{x\in\mathbb{R}^N} F(x) = \underbrace{\frac{1}{2}\|\Phi x - z\|_2^2}_{h(x)} + \underbrace{\lambda\|Dx\|_1}_{g(Dx)}$$

PDS to TV-ℓ_1 deblurring

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = x_n - \tau(\Phi^\top(\Phi x_n - z) + D^\top u_n)) \\ u_{n+1} = \operatorname{prox}_{[-\sigma\lambda,\sigma\lambda]^N}(u_n + \sigma D(2x_{n+1} - x_n)). \end{cases}$$

Python. TV-l2-image-deblurring-l1.

Concluding remarks

- In this lecture, we have presented different algorithms (and their convergence) for image denoising/deblurring.
- As the model becomes more complex, the algorithms for solving it increase their complexity.
- The algorithms presented for the more complex models also can be used for simpler ones.
- Recent comparisons show that the use of proximity operators instead of gradient steps leads to more efficient algorithms.
- The regularization $\|Dx\|_{1,2}$ can be replaced by other regularization terms, including wavelets, nuclear norm, etc. which need more complicated algorithmic structures.

References: Part I

- Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
- Nesterov, Y. Introductory lectures on convex optimization: A basic course (Vol. 87). Springer Science & Business Media. (2003).
- https://npusteln.github.io/ProxImage/

References: Part II&III

- Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
- Nesterov, Y. Introductory lectures on convex optimization: A basic course (Vol. 87). Springer Science & Business Media.
- Briceño-Arias, L. M. and Pustelnik, N., Convergence rate comparison of proximal algorithms for non-smooth convex optimization. Application to texture segmentation. IEEE Signal Processing Letters, vol. 22, pp. 1337–1341, 2022.
- Aujol, J.-F., Gilboa, G., Chan, T., Osher, S.J.: Structure-texture image decomposition—Modeling, algorithms, and parameter selection. Int. J. Comput. Vis. 67, 111–136 (2006)
- Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. 57, 1413–1457 (2004)
- Briceño-Arias, L. M., Combettes, P. L., Pesquet, J.-C., and Pustelnik, N., Proximal algorithms for multicomponent image processing, J. Math. Imaging Vision, vol. 41, pp. 3–22 (2011).
- Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forwardbackward splitting. Multiscale Model. Simul. 4, 1168–1200 (2005)

References: Part II&III

- Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones, Israel J. Math. 26 137–150, (1977)
- Combettes, P. L., Solving monotone inclusions via compositions of nonexpansive averaged operators, Optimization, vol. 53, no. 5–6, pp. 475–504, 2004.
- Briceño-Arias, L. M., Pustelnik, N., Theoretical and numerical comparison of first-order algorithms for cocoercive equations and smooth convex optimization, preprint, 2022 https://arxiv.org/abs/2101.06152
- Combettes, P. L., Dũng, D., Vũ, B. C., Dualization of signal recovery problems, Set-Valued and Variational Analysis, vol. 18, pp. 373–404, 2010.
- Lions, P.-L., Mercier, B., Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal., 16, pp. 964–979, (1979).
- Condat, L. A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. J Optim Theory Appl 158, 460–479 (2013).
- Vũ., B. C., A splitting algorithm for dual monotone inclusions involving cocoercive operators, Adv. Comput. Math. 38, pp. 667–681, (2013).

http://lbriceno.mat.utfsm.cl/

luis.briceno@usm.cl