Algorithmes proximaux pour les problèmes d'optimisation structurés

Patrick L. Combettes

Laboratoire Jacques-Louis Lions
Faculté de Mathématiques
Université Pierre et Marie Curie – Paris 6
75005 Paris, France

Horizon Maths, La Défense, 19 décembre 2012

Notation 1

- $\mathcal{H}, \mathcal{G}_1, \dots, \mathcal{G}_m$: real Hilbert spaces
- $\mathcal{B}(\mathcal{H}, \mathcal{G})$: bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$
- $\rightharpoonup$: weak convergence; $\to$: strong convergence
- $\Gamma_0(\mathcal{H})$: lower semicontinuous convex fuctions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ with $\operatorname{dom} f = \big\{x \in \mathcal{H} \mid f(x) < +\infty\big\} \neq \varnothing$.
- The conjugate of $f\colon \mathcal{H} \to [-\infty, +\infty]$ is

$$f^*\colon \mathcal{H}^* \to [-\infty, +\infty]\colon x^* \mapsto \sup_{x \in \mathcal{H}} \big(\langle x \mid x^* \rangle - f(x)\big)$$

- The subdifferential of $f\colon \Gamma_0(\mathcal{H})$ is

$$\partial f\colon x \mapsto \big\{u \in \mathcal{H} \mid f(x) + f^*(u) = \langle x \mid u \rangle\big\}$$

  f differentiable at $x \Rightarrow \partial f(x) = \{\nabla f(x)\}$;
- Infimal convolution of $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$:

$$f \,\square\, g\colon \mathcal{X} \to [-\infty, +\infty]\colon x \mapsto \inf_{y \in \mathcal{X}} \big(f(y) + g(x - y)\big)$$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\ f(x)$ i.e., solve $\mathbf{0} \in \partial \boldsymbol{f}(\boldsymbol{x})$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x)$ i.e., solve $\mathbf{0} \in \partial \boldsymbol{f}(\boldsymbol{x})$
- $f, k \in \Gamma_0(\mathcal{H})$: $\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + k(x)$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x)$ i.e., solve $\mathbf{0} \in \partial \boldsymbol{f}(\boldsymbol{x})$
- $f, k \in \Gamma_0(\mathcal{H})$: $\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x) + k(x)$
- $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$: $\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x) + g(Lx)$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x)$ i.e., solve $\mathbf{0} \in \partial \boldsymbol{f}(\boldsymbol{x})$
- $f, k \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + k(x)$
- $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + g(Lx)$
- $f \in \Gamma_0(\mathcal{H})$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + \sum_{i=1}^{m} g_i(L_i x)$$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)$ i.e., solve $\mathbf{0} \in \partial \boldsymbol{f}(\boldsymbol{x})$
- $f, k \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + k(x)$
- $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + g(Lx)$
- $f \in \Gamma_0(\mathcal{H})$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \sum_{i=1}^{m} g_i(L_i x)$$

- $f \in \Gamma_0(\mathcal{H})$, $\varphi_i, \ell_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$, $r_i \in \mathcal{G}_i$, $h\colon \mathcal{H} \to \mathbb{R}$ convex and smooth:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \sum_{i=1}^{m} (\varphi_i \,\square\, \ell_i)(L_i x - r_i) + h(x) \qquad (1)$$

Structured convex optimization problems

- $f \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x)$ i.e., solve $\mathbf{0 \in \partial f(x)}$
- $f, k \in \Gamma_0(\mathcal{H})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + k(x)$
- $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$: $\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + g(Lx)$
- $f \in \Gamma_0(\mathcal{H})$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + \sum_{i=1}^{m} g_i(L_i x)$$

- $f \in \Gamma_0(\mathcal{H})$, $\varphi_i, \ell_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$, $r_i \in \mathcal{G}_i$, $h\colon \mathcal{H} \to \mathbb{R}$ convex and smooth:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\; f(x) + \sum_{i=1}^{m} (\varphi_i \,\square\, \ell_i)(L_i x - r_i) + h(x) \qquad (1)$$

- ***How can we solve* (1) *reliably, with an implementable algorithm?***

Supervised learning

- The goal is to infer the functional form of a system acting from an input space $\mathcal{S}$ to an output space $\mathcal{T}$.
- Let (S, T) be a random element on $\mathcal{S} \times \mathcal{T}$ with unknown probability measure P. The training set consists of n realizations $(s_j, t_j)_{1 \leqslant j \leqslant n}$ in $\mathcal{S} \times \mathcal{T}$ of independent P-distributed observations.
- Given some loss function $\ell \colon \mathcal{T} \times \mathcal{T} \to \left]-\infty, +\infty\right]$, a plausible ideal goal is to find a predictor $\Phi \colon \mathcal{S} \to \mathcal{T}$ that minimizes the risk $\mathsf{E}^{\mathsf{P}} \ell(\Phi(S), T)$. Since P is unknown, one aims at minimizing a regularized empirical risk, say

$$\underset{\Phi}{\text{minimize}} \;\; \frac{1}{n} \sum_{j=1}^{n} \ell(\Phi(s_j), t_j) + \varrho(\Phi), \tag{2}$$

where the regularization term ϱ is intended to prevent overfitting and to promote prior information on Φ.

Supervised learning

- Given a countable dictionary $(e_k)_{k\in\mathbb{K}}$ of "features" mapping $\mathcal{S}$ to $\mathcal{T}$, one looks for a predictor of the form $\Phi = \sum_{k\in\mathbb{K}} \xi_k e_k$.
- Under suitable assumptions, (2) can be recast as a convex minimization problem in a Hilbert sequence space $\mathcal{H}$

$$\underset{x=(\xi_k)_{k\in\mathbb{K}}\in\mathcal{H}}{\text{minimize}} \quad \frac{1}{n}\sum_{j=1}^{n} \ell\Bigg(\sum_{k\in\mathbb{K}} \xi_k e_k(s_j), t_j\Bigg) + h(x).$$

- Conceptually, such problems assume the form

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \underbrace{f(x)}_{\text{data fidelity}} + \underbrace{h(x)}_{\text{prior information}} \quad ,$$

where $f \in \Gamma_0(\mathcal{H})$, $h \in \Gamma_0(\mathcal{H})$.

Inverse problems

- The ideal object $\overline{x}$ lies in a Hilbert space $\mathcal{H}$ and is to be recovered from noisy measurements.
- A priori information is available on the problem: properties of $\overline{x}$, data formation model, noise properties, etc.
- Basic example: Constrained linear inverse problem. Data formation model $z = L\overline{x} + w$, convex constraint $\overline{x} \in C$.

$$\underset{x \in C}{\text{minimize}} \ \|Lx - z\|.$$

- Conceptually, such problems are also reducible to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \underbrace{f(x)}_{\text{data fidelity}} + \underbrace{h(x)}_{\text{prior information}},$$

where f and h are in $\Gamma_0(\mathcal{H})$.

Summary

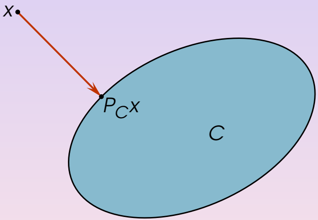
- Common features in learning and inverse problems:
  - The available data (training set in statistical learning or physical measurements in inverse problems) is not sufficient to determine meaningful solutions.
  - Prior information enforcement plays a crucial role in obtaining reliable solutions.
  - They can often be reduced to Hilbertian convex minimization problems.
  - Increasing need for modeling and enforcing sparsity.
- The general variational model under consideration

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{i=1}^{m} (g_i \,\square\, \ell_i)(L_i x - r_i) + h(x),$$

where $f \in \Gamma_0(\mathcal{H})$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $\ell_i \in \Gamma_0(\mathcal{G}_i)$, $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$, $r_i \in \mathcal{G}_i$, $\mathcal{H} \colon \mathcal{H} \to \mathbb{R}$ convex smooth.

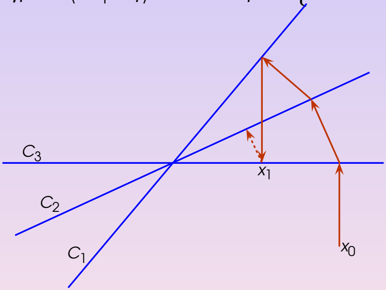
Information modeling and enforcement via convex projection operators

- Projection onto a nonempty closed convex subset C of $\mathcal{H}$: the best approximation to an object $x \in \mathcal{H}$ from the class C of objects is $P_C x = \operatorname{argmin}_{y \in C} \|y - x\|$.



Example 1: Algebraic reconstruction techniques (ART) in computer-aided tomography (1970)

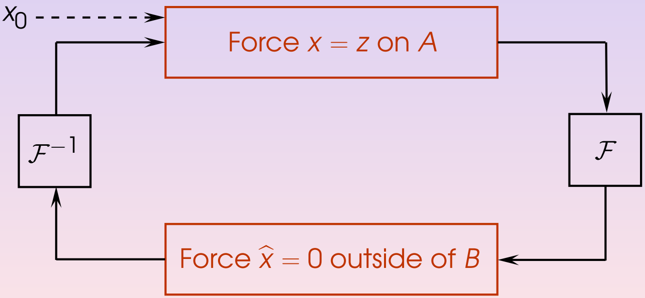
Goal: Reconstruct an image $\overline{x}$ from m scalar measurements $\eta_i = \langle \overline{x} \mid u_i \rangle$. Set $C_i = \big\{ x \in \mathcal{H} \mid \langle x \mid u_i \rangle = \eta_i \big\}$.



This algorithm goes back to Kaczmarz (1937).

Example 2: Band-limited extrapolation (1974-1975)

- The original signal $\overline{x}$ is band-limited (its Fourier transform has compact support B around 0) and it is observed over some region A.
- Gerchberg-Papoulis algorithm:



Example 2: Band-limited extrapolation (1974-1975)

- The set of signals with Fourier support B is the closed vector subspace

$$C_1 = \{x \in L^2 \mid \widehat{x}|_{\complement B} = 0\}$$

- Projecting x onto C_1 amounts to forcing $\widehat{x}$ to 0 outside of B: $\widehat{P_1 x} = \widehat{x}1_B$.
- The set of signals which coincide with z on A is the closed affine subspace

$$C_2 = \{x \in L^2 \mid x|_A = z\}.$$

- Projecting x onto C_2 amounts to forcing $x = z$ on A : $P_2 x = z1_A + x1_{\complement A}$.
- Gerchberg-Papoulis is an alternating projection algorithm: $x_{n+1} = P_1 P_2 x_n$ (Youla,1978).

Projection methods in affine feasibility problems

Given *affine subspaces* $(C_i)_{1\leqslant i\leqslant m}$ of $\mathcal{H}$,

$$\text{Find} \quad x \in C = \bigcap_{i=1}^{m} C_i,$$

i.e.,

$$\underset{x\in\mathcal{H}}{\text{minimize}} \ \sum_{i=1}^{m} \iota_{C_i}(x), \quad \text{where} \quad \iota_{C_i} \colon x \mapsto \begin{cases} 0 & \text{if } \ x \in C_i \\ +\infty & \text{if } \ x \notin C_i \end{cases}$$

Theorem (von Neumann (1933, $m = 2$) - Halperin (1962))

Suppose that $C \neq \varnothing$ and let $x_0 \in \mathcal{H}$. Then

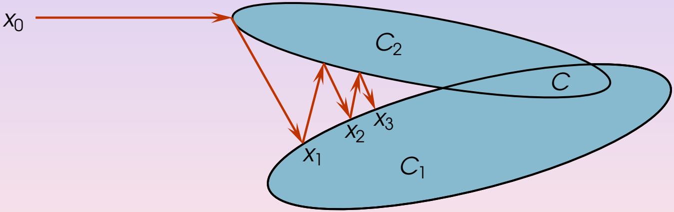
$$x_n = (P_1 \cdots P_m)^n x_0 \to P_C x_0.$$

Projection methods in convex feasibility problems

Given closed *convex subsets* $(C_i)_{1\leqslant i\leqslant m}$ of $\mathcal{H}$, find $x \in C = \bigcap_{i=1}^m C_i$.

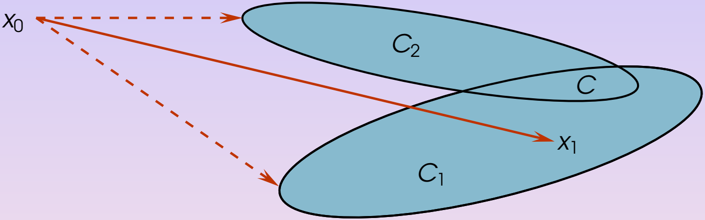
Theorem (Bregman, 1965)

Suppose that $C \neq \varnothing$ and let $x_0 \in \mathcal{H}$. Then (POCS algorithm) $x_n = (P_1 \cdots P_m)^n x_0 \rightharpoonup x \in C$.



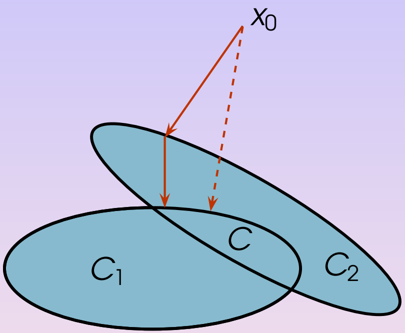
Projection methods in convex feasibility problems

- Variant exist in the form of extrapolated block-iterative parallel algorithms.



- In these convex projection methods the limit is an undetermined feasible point.

Projection methods in best approximation problems



The alternating projection algorithm fails to provide the closest point to x_0 in $C = C_1 \cap C_2$.

Projection methods in best feasible approximation problems

- Problem: compute $P_C x_0$, i.e.,

$$\min_{x \in C = \bigcap_{i=1}^m C_i} \|x - x_0\|, \quad \text{i.e.,} \quad \underset{x \in \mathcal{H}}{\text{minimize}} \;\; \|x - x_0\| + \sum_{i=1}^{m} \iota_{C_i}(x)$$

- Examples:
  - Minimum energy feasible solution ($x_0 = 0$)
  - Least feasible deviation from a nominal function x_0
  - Constrained signal/image denoising: $x_0 = \overline{x} + w$
- Projection algorithms:
  - Anchor point method
  - Haugazeau's method
  - Boyle-Dykstra's method

Best feasible approximation problems

Theorem (Boyle/Dykstra, 1986)

Let $z \in \mathcal{H}$, let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$, and set

$$\left\lfloor \begin{array}{l} v_0 = z \\ \\ \end{array} \right. \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} x_n & = P_D(v_n \qquad) \\ v_{n+1} & = P_C(x_n \qquad) \end{array} \right. .$$

Then $x_n \rightharpoonup x \in C \cap D$ (Bregman (1965)).

Best feasible approximation problems

Theorem (Boyle/Dykstra, 1986)

Let $z \in \mathcal{H}$, let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$, and set

$$\left\lfloor \begin{array}{l} v_0 = z \\ p_0 = 0 \\ q_0 = 0 \end{array} \right. \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} x_n & = P_D(v_n + q_n) \\ q_{n+1} & = v_n + q_n - x_n \\ v_{n+1} & = P_C(x_n + p_n) \\ p_{n+1} & = x_n + p_n - v_{n+1}. \end{array} \right.$$

Then $x_n \to P_{C \cap D} z$.

Best feasible approximation problems

Theorem (Boyle/Dykstra, 1986)

Let $z \in \mathcal{H}$, let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$, and set

$$\left\lfloor \begin{array}{l} v_0 = z \\ p_0 = 0 \\ q_0 = 0 \end{array} \right. \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} x_n & = P_D(v_n + q_n) \\ q_{n+1} & = v_n + q_n - x_n \\ v_{n+1} & = P_C(x_n + p_n) \\ p_{n+1} & = x_n + p_n - v_{n+1}. \end{array} \right.$$

Then $x_n \to P_{C\cap D} z$. (Corollary: von-Neumann-Halperin.)

Best feasible approximation problems

Theorem (Boyle/Dykstra, 1986)

Let $z \in \mathcal{H}$, let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$, and set

$$\left\lfloor \begin{aligned} v_0 &= z \\ p_0 &= 0 \\ q_0 &= 0 \end{aligned} \right. \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{aligned} x_n &= P_D(v_n + q_n) \\ q_{n+1} &= v_n + q_n - x_n \\ v_{n+1} &= P_C(x_n + p_n) \\ p_{n+1} &= x_n + p_n - v_{n+1}. \end{aligned} \right.$$

Then $x_n \to P_{C \cap D} z$. *(Corollary: von-Neumann-Halperin.)*

Unfortunately, not much more we can do with projections operators beyond such best approximation problems.

Notation 2

- Conjugate of $f \in \Gamma_0(\mathcal{H})$:

$$f^*\colon \mathcal{H} \to \left]-\infty, +\infty\right]\colon u \mapsto \sup_{x \in \mathcal{H}} \langle x \mid u \rangle - f(x).$$

- Subdifferential of $f \in \Gamma_0(\mathcal{H})$:

$$\begin{aligned} \partial f\colon \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\mapsto \big\{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid u \rangle + f(x) \leqslant f(y)\big\} \\ &= \big\{u \in \mathcal{H} \mid f(x) + f^*(u) = \langle x \mid u \rangle\big\}. \end{aligned}$$

- Distance, indicator, and support function of $C \subset \mathcal{H}$: $d_C, \iota_C, \sigma_C = \iota_C^*$.

J.-J. Moreau's proximity operator (1962)

- $C \subset \mathcal{H}$ nonempty closed and convex. Then $\iota_C \in \Gamma_0(\mathcal{H})$,

$$P_C \colon x \mapsto \underset{y \in C}{\operatorname{argmin}} \frac{1}{2}\|x-y\|^2 = \underset{y \in \mathcal{H}}{\operatorname{argmin}}\, \iota_C(y) + \frac{1}{2}\|x-y\|^2.$$

J.-J. Moreau's proximity operator (1962)

- $C \subset \mathcal{H}$ nonempty closed and convex. Then $\iota_C \in \Gamma_0(\mathcal{H})$,

$$P_C\colon x \mapsto \underset{y\in C}{\operatorname{argmin}}\ \frac{1}{2}\|x-y\|^2 = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\ \iota_C(y) + \frac{1}{2}\|x-y\|^2.$$

- The proximity operator of $f \in \Gamma_0(\mathcal{H})$ is

$$\operatorname{prox}_f\colon x \mapsto \underset{y\in\mathcal{H}}{\operatorname{argmin}}\ f(y) + \frac{1}{2}\|x-y\|^2;$$

the Moreau envelope of f is $\widetilde{f} = f \,\square\, (\|\cdot\|^2/2)$.

J.-J. Moreau's proximity operator (1962)

- $C \subset \mathcal{H}$ nonempty closed and convex. Then $\iota_C \in \Gamma_0(\mathcal{H})$,

$$P_C \colon x \mapsto \underset{y \in C}{\operatorname{argmin}} \frac{1}{2}\|x-y\|^2 = \underset{y \in \mathcal{H}}{\operatorname{argmin}} \, \iota_C(y) + \frac{1}{2}\|x-y\|^2.$$

- The proximity operator of $f \in \Gamma_0(\mathcal{H})$ is

$$\operatorname{prox}_f \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \, f(y) + \frac{1}{2}\|x-y\|^2;$$

  the Moreau envelope of f is $\widetilde{f} = f \,\square\, (\|\cdot\|^2/2)$.
- Basic properties:
  - $\operatorname{prox}_f = (\operatorname{Id} + \partial f)^{-1} = \nabla \widetilde{f^*}$.
  - $\operatorname{Fix} \operatorname{prox}_f = \operatorname{Argmin} f$.
  - $\operatorname{prox}_f + \operatorname{prox}_{f^*} = \operatorname{Id}$ (Moreau's decomposition).
  - $\|\operatorname{prox}_f x - \operatorname{prox}_f y\| \leqslant \|x-y\|$.

J.-J. Moreau's proximity operator (1962)

- $C \subset \mathcal{H}$ nonempty closed and convex. Then $\iota_C \in \Gamma_0(\mathcal{H})$,

$$P_C \colon x \mapsto \underset{y \in C}{\operatorname{argmin}} \frac{1}{2}\|x-y\|^2 = \underset{y \in \mathcal{H}}{\operatorname{argmin}} \iota_C(y) + \frac{1}{2}\|x-y\|^2.$$

- The proximity operator of $f \in \Gamma_0(\mathcal{H})$ is

$$\operatorname{prox}_f \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2}\|x-y\|^2;$$

the Moreau envelope of f is $\widetilde{f} = f \,\square\, (\|\cdot\|^2/2)$.

- Basic properties:
  - $\operatorname{prox}_f = (\operatorname{Id} + \partial f)^{-1} = \nabla \widetilde{f^*}$.
  - $\operatorname{Fix} \operatorname{prox}_f = \operatorname{Argmin} f$.
  - $\operatorname{prox}_f + \operatorname{prox}_{f^*} = \operatorname{Id}$ (Moreau's decomposition).
  - $\|\operatorname{prox}_f x - \operatorname{prox}_f y\|^2 \leqslant \|x-y\|^2 - \|\operatorname{prox}_{f^*} x - \operatorname{prox}_{f^*} y\|^2$.

What do proximity operators do?

- Translation: $\operatorname{prox}_{\langle\cdot\mid z\rangle}x = x - z$.
- Scaling: $\operatorname{prox}_{\alpha\|\cdot\|^2/2}x = (\alpha+1)^{-1}x$.
- Best approximation: $\operatorname{prox}_{\iota_C} = P_C$.
- Underrelaxed projection:

$$\operatorname{prox}_{\frac{1}{2\alpha}d_C^2} = \operatorname{Id} + \frac{1}{1+\alpha}(P_C - \operatorname{Id}).$$

- Landweber-Tykhonov transformation:

$$\operatorname{prox}_{\alpha\|L\cdot-r\|^2/2}x = (\operatorname{Id} + \alpha L^*L)^{-1}(x + \alpha L^*r).$$

- Maximum a posteriori denoising with log-concave densities: $z = \overline{x} + w$, $\overline{x} \sim \alpha\exp(-f(\overline{x}))$. Denoised signal $x = \operatorname{prox}_f z$.

What do proximity operators do?

- Soft-thresholding: $\operatorname{prox}_{\alpha\|\cdot\|} x = \begin{cases} \left(1 - \dfrac{\alpha}{\|x\|}\right) x & \text{if } \|x\| > \alpha \\ 0 & \text{if } \|x\| \leqslant \alpha. \end{cases}$

What do proximity operators do?

- Soft-thresholding: $\operatorname{prox}_{\alpha\|\cdot\|} x = \begin{cases} \left(1 - \dfrac{\alpha}{\|x\|}\right) x & \text{if } \|x\| > \alpha \\ 0 & \text{if } \|x\| \leqslant \alpha. \end{cases}$
- Generalized soft-thresholding: $C \neq \varnothing$ closed and convex, $\phi \in \Gamma_0(\mathbb{R})$ even, $\Omega = \big\{x \in \mathcal{H} \mid d_C(x) \in \partial\phi(0)\big\}$. Then

$$\operatorname{prox}_{\sigma_C + \phi\circ\|\cdot\|} x = \begin{cases} \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x - P_C x), & \text{if } x \notin \Omega; \\ 0, & \text{if } x \in \Omega. \end{cases}$$

The standard case is recovered with $C = \{0\}$ and $\phi = \alpha|\cdot|$.

What do proximity operators do?

- Soft-thresholding: $\operatorname{prox}_{\alpha\|\cdot\|} x = \begin{cases} \left(1 - \dfrac{\alpha}{\|x\|}\right) x & \text{if } \|x\| > \alpha \\ 0 & \text{if } \|x\| \leqslant \alpha. \end{cases}$
- Generalized soft-thresholding: $C \neq \varnothing$ closed and convex, $\phi \in \Gamma_0(\mathbb{R})$ even, $\Omega = \big\{x \in \mathcal{H} \mid d_C(x) \in \partial\phi(0)\big\}$. Then

$$\operatorname{prox}_{\sigma_C + \phi\circ\|\cdot\|} x = \begin{cases} \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x - P_C x), & \text{if } x \notin \Omega; \\ 0, & \text{if } x \in \Omega. \end{cases}$$

  The standard case is recovered with $C = \{0\}$ and $\phi = \alpha|\cdot|$.
- Componentwise thresholding: if $(e_k)_{k\in\mathbb{K}}$ is an orthonormal basis of $\mathcal{H}$, $(\forall k \in \mathbb{K})\ \Gamma_0(\mathbb{R}) \ni \phi_k \geqslant \phi_k(0) = 0$, and $f\colon x \mapsto \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle)$, then $\operatorname{prox}_f x = \sum_{k\in\mathbb{K}} (\operatorname{prox}_{\phi_k}\langle x \mid e_k\rangle) e_k$.

Proximal soft thresholders on the real line

Characterization and computation (PLC-Pesquet, 2007)

Let $\phi \in \Gamma_0(\mathbb{R})$ and let $\varnothing \neq \Omega \subset \mathbb{R}$ be a closed interval. Set

$$\mathrm{soft}_\Omega = \mathrm{prox}_{\sigma_\Omega} : \xi \mapsto \begin{cases} \xi - \underline{\omega}, & \text{if } \xi < \underline{\omega}; \\ 0, & \text{if } \xi \in \Omega; \\ \xi - \overline{\omega}, & \text{if } \xi > \overline{\omega}, \end{cases} \quad \text{with} \quad \begin{cases} \underline{\omega} = \inf \Omega, \\ \overline{\omega} = \sup \Omega. \end{cases}$$

Then the following are equivalent.

1. prox_ϕ is a soft thresholder on Ω.
2. $\phi = \psi + \sigma_\Omega$, where $\psi \in \Gamma_0(\mathbb{R})$ is differentiable at 0 and $\psi'(0) = 0$.

Moreover, $\mathrm{prox}_\phi = \mathrm{prox}_\psi \circ \mathrm{soft}_\Omega$.

Example (triangular density)

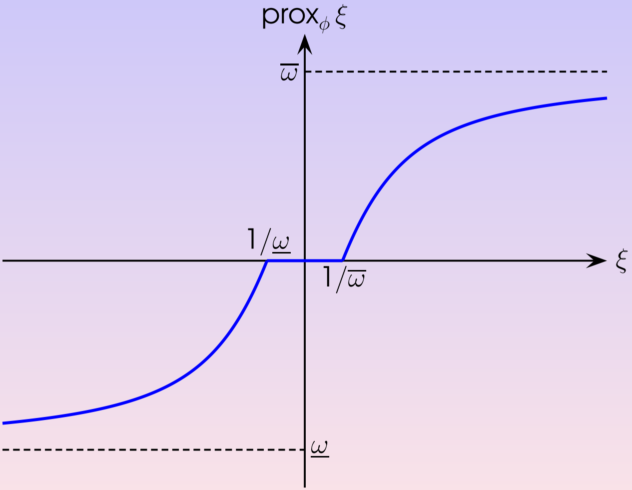
- $\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right]\colon \xi \mapsto \begin{cases} \ln(\omega) - \ln(\omega - |\xi|), & \text{if } |\xi| < \omega; \\ +\infty, & \text{otherwise.} \end{cases}$
- Set $\Omega = [-1/\omega, 1/\omega]$ and

$$\psi(\xi) = \begin{cases} \ln(\omega) - \ln(\omega - |\xi|) - |\xi|/\omega, & \text{if } |\xi| < \omega; \\ +\infty, & \text{otherwise.} \end{cases}$$

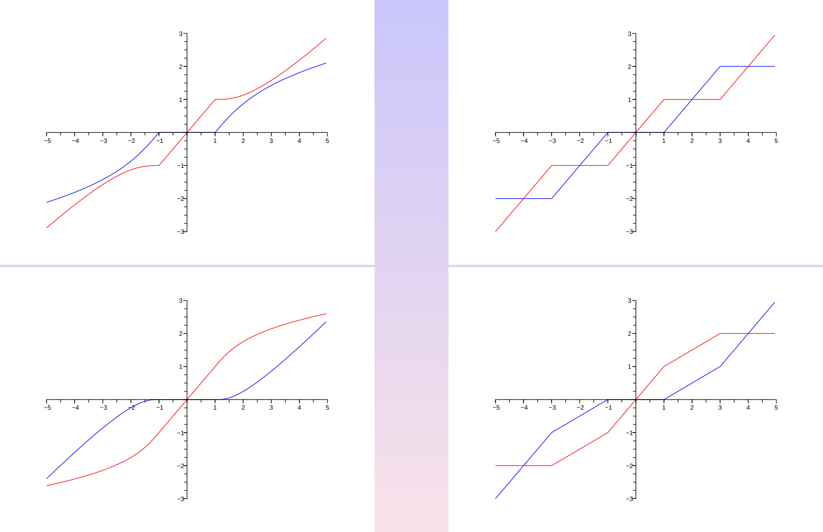
- Then $\psi'(0) = 0$ and $\phi = \psi + \sigma_\Omega$.
- $\operatorname{prox}_\phi \xi =$

$$\begin{cases} \operatorname{sign}(\xi)\, \dfrac{|\xi| + \omega - \sqrt{\big||\xi| - \omega\big|^2 + 4}}{2}, & \text{if } |\xi| > 1/\omega; \\ 0, & \text{otherwise.} \end{cases}$$

Example (triangular density)



More dual proximal thresholder pairs on $[-1, 1]$...



More properties of proximity operators...

property	$\psi(x)$	$\operatorname{prox}_{\psi} x$
shift	$\varphi(x - z), z \in \mathcal{H}$	$z + \operatorname{prox}_{\varphi}(x - z)$
scaling	$\varphi(x/\rho), \rho \in \mathbb{R} \smallsetminus \{0\}$	$\rho \operatorname{prox}_{\varphi/\rho^2}(x/\rho)$
reflection	$\varphi(-x)$	$-\operatorname{prox}_{\varphi}(-x)$
quadratic perturbation	$\varphi(x) + \alpha\|x\|^2/2 + \beta\langle x \mid u\rangle + \gamma$ $u \in \mathcal{H}, \alpha > 0, (\beta, \gamma) \in \mathbb{R}^2$	$\operatorname{prox}_{\varphi/(\alpha+1)}\big((x - \beta u)/(\alpha + 1)\big)$
conjugation	$\varphi^*(x)$	$x - \operatorname{prox}_{\varphi} x$
squared distance	$d_C^2(x)/2$	$(x + P_C x)/2$
Moreau envelope	$\widetilde{\varphi}(x) = \inf_{y \in \mathcal{H}} \varphi(y) + \|x - y\|^2/2$	$\big(x + \operatorname{prox}_{2\varphi} x\big)/2$
decomposition in an orthonormal basis	$\sum_{k=1}^N \phi_k(\langle b_k \mid x\rangle)$ $\phi_k \in \Gamma_0(\mathbb{R}), (b_k)_{1 \leq k \leq N}$ orthonormal basis of $\mathcal{H}$	$\sum_{k=1}^N \operatorname{prox}_{\phi_k}(\langle b_k \mid x\rangle) b_k$
semi-orthogonal linear transform	$\varphi(Lx)$ $L \in \mathbb{R}^{M \times N}, LL^* = \nu I, \nu > 0$	$x + \nu^{-1} L^*\big(\operatorname{prox}_{\nu\varphi}(Lx) - Lx\big)$
quadratic function	$\gamma\|Lx - z\|^2/2$ $L \in \mathbb{R}^{M \times N}, \gamma > 0, z \in \mathcal{G}$	$(I + \gamma L^* L)^{-1}(x + \gamma L^* z)$
indicator function	$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}$	$P_C x$
distance function	$\gamma d_C(x), \gamma > 0$	$\begin{cases} x + \frac{\gamma}{d_C(x)}(P_C x - x) & \text{if } d_C(x) > \gamma \\ P_C x & \text{otherwise} \end{cases}$
function of distance	$\phi(d_C(x))$ $\phi \in \Gamma_0(\mathbb{R})$ even, differentiable at 0 with $\phi'(0) = 0$	$\begin{cases} x + \left(1 - \frac{\operatorname{prox}_{\phi} d_C(x)}{d_C(x)}\right)(P_C x - x) & \text{if } x \notin C \\ x & \text{otherwise} \end{cases}$
support function	$\sigma_C(x)$	$x - P_C x$
extended thresholding	$\sigma_C(x) + \phi(\|x\|)$ $\phi \in \Gamma_0(\mathbb{R})$ even and not constant	$\begin{cases} \frac{\operatorname{prox}_{\phi} d_C(x)}{d_C(x)}(x - P_C x) & \text{if } d_C(x) > \max \operatorname{Argmin} \phi \\ x - P_C x & \text{otherwise} \end{cases}$

Examples of proximity operators

$\phi(\xi)$	$\operatorname{prox}_\phi \xi$
$\sigma_{[\underline{\omega},\overline{\varpi}]}(\xi) = \begin{cases} \underline{\omega}\xi & \text{if } \xi < 0 \\ 0 & \text{if } \xi = 0 \\ \overline{\varpi}\xi & \text{otherwise} \end{cases}$	$\operatorname{soft}_{[\underline{\omega},\overline{\varpi}]}(\xi) = \begin{cases} \xi - \underline{\omega} & \text{if } \xi < \underline{\omega} \\ 0 & \text{if } \xi \in [\underline{\omega}, \overline{\varpi}] \\ \xi - \overline{\varpi} & \text{if } \xi > \overline{\varpi} \end{cases}$
$\psi(\xi) + \sigma_{[\underline{\omega},\overline{\varpi}]}(\xi)$ $\psi \in \Gamma_0(\mathbb{R})$ differentiable at 0 with $\psi'(0) = 0$	$\operatorname{prox}_\psi\left(\operatorname{soft}_{[\underline{\omega},\overline{\varpi}]}(\xi)\right)$
$\tau\lvert\xi\rvert^2$	$\xi/(2\tau+1)$
$\kappa\lvert\xi\rvert^p$	$\operatorname{sign}(\xi)\varrho$, where $\varrho \geq 0$ and $\varrho + p\kappa\varrho^{p-1} = \lvert\xi\rvert$
$\begin{cases} \tau\xi^2 & \text{if } \lvert\xi\rvert \leqslant \omega/\sqrt{2\tau} \\ \omega\sqrt{2\tau}\lvert\xi\rvert - \omega^2/2 & \text{otherwise} \end{cases}$	$\begin{cases} \xi/(2\tau+1) & \text{if } \lvert\xi\rvert \leqslant \omega(2\tau+1)/\sqrt{2\tau} \\ \xi - \omega\sqrt{2\tau}\operatorname{sign}(\xi) & \text{otherwise} \end{cases}$
$\omega\lvert\xi\rvert + \tau\lvert\xi\rvert^2 + \kappa\lvert\xi\rvert^p$	$\operatorname{sign}(\xi)\operatorname{prox}_{\kappa\lvert\cdot\rvert^p/(2\tau+1)}\left(\max\{\lvert\xi\rvert - \omega, 0\}/(2\tau+1)\right)$
$\omega\lvert\xi\rvert - \ln(1+\omega\lvert\xi\rvert)$, $\omega > 0$	$(2\omega)^{-1}\operatorname{sign}(\xi)\left(\omega\lvert\xi\rvert - \omega^2 - 1 + \sqrt{\lvert\omega\lvert\xi\rvert - \omega^2 - 1\rvert^2 + 4\omega\lvert\xi\rvert}\right)$
$\begin{cases} \omega\xi & \text{if } \xi \geqslant 0 \\ +\infty & \text{otherwise} \end{cases}$	$\begin{cases} \xi - \omega & \text{if } \xi \geqslant \omega \\ 0 & \text{otherwise} \end{cases}$
$\begin{cases} -\kappa\ln(\xi) + \omega\xi & \text{if } \xi > 0 \\ +\infty & \text{otherwise} \end{cases}$	$\left(\xi - \omega + \sqrt{\lvert\xi - \omega\rvert^2 + 4\kappa}\right)/2$
$\begin{cases} -\kappa\ln(\xi) + \xi^2/2 & \text{if } \xi > 0 \\ +\infty, & \text{otherwise} \end{cases}$	$\left(\xi + \sqrt{\xi^2 + 8\kappa}\right)/4$
$\iota_{[\underline{\omega},\overline{\varpi}]}(\xi)$	$P_{[\underline{\omega},\overline{\varpi}]}\xi$
$\begin{cases} -\ln(\xi - \underline{\omega}) + \ln(-\underline{\omega}) & \text{if } \xi \in]\underline{\omega}, 0] \\ -\ln(\overline{\varpi} - \xi) + \ln(\overline{\varpi}) & \text{if } \xi \in]0, \overline{\varpi}[\\ +\infty & \text{otherwise} \end{cases}$ $\underline{\omega} < 0 < \overline{\varpi}$	$\begin{cases} \left(\xi + \underline{\omega} + \sqrt{\lvert\xi - \underline{\omega}\rvert^2 + 4}\right)/2 & \text{if } \xi < 1/\underline{\omega} \\ \left(\xi + \overline{\varpi} - \sqrt{\lvert\xi - \overline{\varpi}\rvert^2 + 4}\right)/2 & \text{if } \xi > 1/\overline{\varpi} \\ 0 & \text{otherwise} \end{cases}$
$\begin{cases} -\kappa\ln(\xi) + \omega\xi^p & \text{if } \xi > 0 \\ +\infty & \text{otherwise} \end{cases}$	$\pi > 0$ such that $p\omega\pi^p + \pi^2 - \xi\pi = \kappa$
$\begin{cases} -\kappa\ln(\xi) + \omega\xi + \rho/\xi & \text{if } \xi > 0 \\ +\infty, & \text{otherwise} \end{cases}$	$\pi > 0$ such that $\pi^3 + (\omega - \xi)\pi^2 - \kappa\pi = \rho$
$\begin{cases} -\underline{\kappa}\ln(\xi - \underline{\omega}) - \overline{\kappa}\ln(\overline{\varpi} - \xi) & \text{if } \xi \in]\underline{\omega}, \overline{\varpi}[\\ +\infty & \text{otherwise} \end{cases}$	$\pi \in]\underline{\omega}, \overline{\varpi}[$ such that $\pi^3 - (\underline{\omega} + \overline{\varpi} + \xi)\pi^2 + \left(\underline{\omega}\overline{\varpi} - \underline{\kappa} - \overline{\kappa} + (\underline{\omega} + \overline{\varpi})\xi\right)\pi = \underline{\omega}\overline{\varpi}\xi - \underline{\omega}\overline{\kappa} - \overline{\varpi}\underline{\kappa}$

Proximal splitting methods for minimization

- Problem:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+\sum_{i=1}^{m}(g_i\,\Box\,\ell_i)(L_ix-r_i)+h(x),$$

- Traditional splitting methods restricted to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+h(x),\;\text{ where } f \text{ and } h \text{ are in } \Gamma_0(\mathcal{H}).$$

  - Douglas-Rachford splitting requires the proximity operator of $h=\sum_{i=1}^m g_i\circ L_i$: hopeless.
  - Forward-backward splitting is applicable if each g_i (hence $h=\sum_{i=1}^m g_i\circ L_i$) is smooth with a Lipschitzian gradient: too restrictive for our purposes.

- Key: duality in monotone operator theory.

Forward-backward splitting

- h finite, differentiable, ∇h $1/\beta$-Lipschitz-continuous.
- x minimizes $f + h \Leftrightarrow x = \text{prox}_{\gamma f}\big(x - \gamma \nabla h(x)\big)$, $\gamma > 0$.
- The sequence constructed by the algorithm

$$x_{n+1} = \quad \text{prox}_{\gamma \ f}\big(x_n - \gamma \quad (\nabla h(x_n) \quad)\big) \quad ,$$

where

- $0 < \gamma < 2\beta$ (Mercier, 1979)

converges weakly to a minimizer of $f + h$.

Forward-backward splitting

- h finite, differentiable, ∇h $1/\beta$-Lipschitz-continuous.
- x minimizes $f + h \Leftrightarrow x = \operatorname{prox}_{\gamma f}\big(x - \gamma \nabla h(x)\big)$, $\gamma > 0$.
- The sequence constructed by the algorithm

$$x_{n+1} = \operatorname{prox}_{\gamma_n f}\big(x_n - \gamma_n (\nabla h(x_n))\big),$$

where

- $0 < \gamma < 2\beta$ (Mercier, 1979)
- $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < 2\beta$ (Tseng, 1990)

converges weakly to a minimizer of $f + h$.

Forward-backward splitting

- h finite, differentiable, ∇h $1/\beta$-Lipschitz-continuous.
- x minimizes $f + h \Leftrightarrow x = \text{prox}_{\gamma f}\big(x - \gamma \nabla h(x)\big)$, $\gamma > 0$.
- The sequence constructed by the algorithm

$$x_{n+1} = \text{prox}_{\gamma_n f}\big(x_n - \gamma_n (\nabla h(x_n) + b_n)\big) + a_n ,$$

where

- $0 < \gamma < 2\beta$ (Mercier, 1979)
- $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < 2\beta$ (Tseng, 1990)
- $\sum_{n\in\mathbb{N}} \|a_n\| < +\infty$, $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$ (PLC, 2004)

converges weakly to a minimizer of $f + h$.

Forward-backward splitting

- h finite, differentiable, ∇h $1/\beta$-Lipschitz-continuous.
- x minimizes $f + h \Leftrightarrow x = \text{prox}_{\gamma f}\big(x - \gamma \nabla h(x)\big)$, $\gamma > 0$.
- The sequence constructed by the algorithm

$$x_{n+1} = x_n + \lambda_n\Big(\text{prox}_{\gamma_n f}\big(x_n - \gamma_n \quad (\nabla h(x_n) + b_n)\big) + a_n - x_n\Big),$$

where

- $0 < \gamma < 2\beta$ (Mercier, 1979)
- $0 < \inf_{n \in \mathbb{N}} \gamma_n \leqslant \sup_{n \in \mathbb{N}} \gamma_n < 2\beta$ (Tseng, 1990)
- $\sum_{n \in \mathbb{N}} \|a_n\| < +\infty$, $\sum_{n \in \mathbb{N}} \|b_n\| < +\infty$ (PLC, 2004)
- $(\lambda_n)_{n \in \mathbb{N}}$ in $]0, 1]$, $\inf_{n \in \mathbb{N}} \lambda_n > 0$ (PLC, 2004)

converges weakly to a minimizer of $f + h$.

Forward-backward splitting

- h finite, differentiable, ∇h $1/\beta$-Lipschitz-continuous.
- x minimizes $f + h \Leftrightarrow x = \text{prox}_{\gamma f}\big(x - \gamma \nabla h(x)\big)$, $\gamma > 0$.
- The sequence constructed by the algorithm

$$x_{n+1} = x_n + \lambda_n\Big(\text{prox}^{U_n}_{\gamma_n f}\big(x_n - \gamma_n U_n^{-1}(\nabla h(x_n) + b_n)\big) + a_n - x_n\Big),$$

where

- $0 < \gamma < 2\beta$ (Mercier, 1979)
- $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < 2\beta$ (Tseng, 1990)
- $\sum_{n\in\mathbb{N}} \|a_n\| < +\infty$, $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$ (PLC, 2004)
- $(\lambda_n)_{n\in\mathbb{N}}$ in $]0,1]$, $\inf_{n\in\mathbb{N}} \lambda_n > 0$ (PLC, 2004)
- $(1+\eta_n)U_{n+1} \succcurlyeq U_n = U_n^* \succcurlyeq \alpha \text{Id}$, $\alpha > 0$, $\eta_n \geqslant 0$, $\sum_{n\in\mathbb{N}} \eta_n < +\infty$ (PLC&Vũ, 2012)

converges weakly to a minimizer of $f + h$.

Strongly convex problems (PLC-Dinh Dũng-Vũ, 2011)

- $z \in \mathcal{H}$, $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $\omega_i \in \left]0,1\right]$, $\sum_{i=1}^m \omega_i = 1$, $(r_i)_{1 \leqslant i \leqslant m} \in \operatorname{sri} \big\{ (L_i x - y_i)_{1 \leqslant i \leqslant m} \mid x \in \mathcal{H}, y_i \in \operatorname{dom} g_i \big\}$.
- Primal problem:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^m \omega_i g_i(L_i x - r_i) + \frac{1}{2}\|x - z\|^2.$$

- Dual problem:

$$\underset{v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m}{\text{minimize}} \quad \frac{1}{2}\left\| z - \sum_{i=1}^m \omega_i L_i^* v_i \right\|^2 + \sum_{i=1}^m \omega_i \big(g_i^*(v_i) + \langle v_i \mid r_i \rangle \big),$$

- The dual problem has a Lipschitz-differentiable component + a separable component: apply the forward backward algorithm to it.

Strongly convex problems

- Algorithm ($\rho = 2(\max_{1\leqslant i\leqslant m} \|L_i\|)^{-2}$):

$$
\left\lfloor
\begin{array}{l}
x_n = z - \sum_{i=1}^m \omega_i L_i^* v_{i,n} \\
\gamma_n \in [\varepsilon, 2\rho - \varepsilon] \\
\text{For } i = 1, \ldots, m \\
\quad \left\lfloor v_{i,n+1} \approx \text{prox}_{\gamma_n g_i^*}\big(v_{i,n} + \gamma_n(L_i x_n - r_i)\big). \right.
\end{array}
\right.
$$

- $x_n \to \bar{x} = \text{prox}_f z$, where $f\colon x \mapsto \sum_{i=1}^m \omega_i g_i(L_i x - r_i)$.

Example: Multiview image recovery

- The image space is $\mathcal{H} = \mathrm{H}_0^1(\Omega)$, where Ω is a nonempty bounded open domain in $\mathbb{R}^2$
- The original image $\overline{x}$ admits a sparse decomposition in an orthonormal basis $(e_k)_{k\in\mathbb{N}}$ of $\mathcal{H}$. This property can be promoted by the *elastic net* potential

$$x \mapsto \sum_{k\in\mathbb{N}} \phi_k(\langle x \mid e_k\rangle)$$

  where $\phi_k\colon \xi \mapsto \alpha|\xi| + \beta|\xi|^2$, with $\alpha > 0$ and $\beta > 0$
- The original image $\overline{x}$ is piecewise smooth. This property is promoted by the total variation potential

$$\mathrm{tv}(x) = \int_\Omega |\nabla x(\omega)|_2 d\omega.$$

Example: Multiview image recovery

- p noisy linear observations of the original image are available : $r_i = T_i\widetilde{x} + u_i \in \mathcal{G}_i$, $1 \leqslant i \leqslant p$
- Problem:

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^{p} \omega_i \|T_i x - r_i\|_{\mathcal{G}_i} + \sum_{k\in\mathbb{N}} \left(\omega_{p+1}|\langle x \mid e_k\rangle| + \frac{1}{2}|\langle x \mid e_k\rangle|^2\right) + \omega_{p+2}\,\mathrm{tv}(x)$$

- A special case of the previous problem with:
  - $z = 0$.
  - $L_i = T_i$ and $g_i = \|\cdot\|_{\mathcal{G}_i}$, hence $\mathrm{prox}_{\gamma g_i^*} = P_{B_i(0;1)}$ $(1 \leqslant i \leqslant p)$
  - $\mathcal{G}_{p+1} = \ell^2(\mathbb{N})$, $g_{p+1} = \|\cdot\|_{\ell^1}$, $L_{p+1}^* \colon (\nu_k)_{k\in\mathbb{N}} \mapsto \sum_{k\in\mathbb{N}} \nu_k e_k$
  - $\mathcal{G}_{p+2} = \mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega)$, $L_{p+2} = \nabla$, $L_{p+2}^* = -\,\mathrm{div}$, and $g_{p+2} = \sigma_K$, where $K = \{y \in \mathcal{G}_{p+2} \mid |y|_2 \leqslant 1 \text{ a.e.}\}$

Example: Multiview image recovery

- Algorithm:

$$
\left\lfloor
\begin{array}{l}
x_n = z - \sum_{i=1}^{p} \omega_i T_i^* v_{i,n} - \omega_{p+1} \sum_{k\in\mathbb{N}} \nu_{k,n} e_k + \omega_{p+2} \operatorname{div} v_{p+2,n} \\
\gamma_n \in [\varepsilon, 2\rho - \varepsilon] \\
\text{For } i = 1, \ldots, p \\
\left\lfloor \; v_{i,n+1} = \dfrac{v_{i,n} + \gamma_n (T_i x_n - r_i)}{\max\{1, \|v_{i,n} + \gamma_n (T_i x_n - r_i)\|_{\mathcal{G}_i}\}} \right. \\
\text{For every } k \in \mathbb{N},\ \nu_{k,n+1} = \dfrac{\nu_{k,n} + \gamma_n \langle x_n \mid e_k \rangle}{\max\{1, |\nu_{k,n} + \gamma_n \langle x_n \mid e_k \rangle|\}} \\
\text{For almost every } \omega \in \Omega, \\
\qquad v_{p+2,n+1}(\omega) = \dfrac{v_{p+2,n}(\omega) + \gamma_n \nabla x_n(\omega)}{\max\{1, |v_{p+2,n}(\omega) + \gamma_n \nabla x_n(\omega)|_2\}}
\end{array}
\right.
$$

- $(x_n)_{n\in\mathbb{N}}$ converges strongly to the solution

Back to the general problem

- $z \in \mathcal{H}$, $f \in \Gamma_0(\mathcal{H})$, and $h\colon \mathcal{H} \to \mathbb{R}$ convex and differentiable with a μ-Lipschitzian gradient
- For every $i \in \{1, \ldots, m\}$, $r_i \in \mathcal{G}_i$, $g_i \in \Gamma_0(\mathcal{G}_i)$, $\ell_i \in \Gamma_0(\mathcal{G}_i)$ $1/\nu_i$-strongly convex, $L_i\colon \mathcal{H} \to \mathcal{G}_i$ linear and bounded
- **Primal problem:**

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{i=1}^{m} (g_i \,\Box\, \ell_i)(L_i x - r_i) + h(x) - \langle x \mid z\rangle$$

- **Dual problem:**

$$\underset{v_1\in\mathcal{G}_1,\ldots,v_m\in\mathcal{G}_m}{\text{minimize}} \; (f^* \,\Box\, h^*)\Big(z - \sum_{i=1}^{m} L_i^* v_i\Big) + \sum_{i=1}^{m} \big(g_i^*(v_i) + \ell_i^*(v_i) + \langle v_i \mid r_i\rangle\big)$$

Structure of the problem

- $\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \sum_{i=1}^m (g_i \,\square\, \ell_i)(L_i x - r_i) + h(x) - \langle x \mid z\rangle$
- $\underset{v_1\in\mathcal{G}_1,\ldots,v_m\in\mathcal{G}_m}{\text{minimize}}\;\; (f^* \,\square\, h^*)\Big(z - \sum_{i=1}^m L_i^* v_i\Big) + \sum_{i=1}^m \big(g_i^*(v_i) + \ell_i^*(v_i) + \langle v_i \mid r_i\rangle\big)$
- $A = \partial f$, $C = \nabla h$, $B_i = \partial g_i$, and $D_i = \partial \ell_i$
- $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m$
- $\boldsymbol{M}\colon \mathcal{K} \to 2^{\mathcal{K}}\colon (x, v_1, \ldots, v_m) \mapsto (-z+Ax) \times (r_1+B_1^{-1}v_1) \times \cdots \times (r_m+B_m^{-1}v_m)$
- $\boldsymbol{Q}\colon \mathcal{K} \to \mathcal{K}\colon (x, v_1, \ldots, v_m) \mapsto \Big(Cx + \sum_{i=1}^m L_i^* v_i, -L_1x+D_1^{-1}v_1, \ldots, -L_mx+D_m^{-1}v_m\Big)$
- $\boldsymbol{M}$ and $\boldsymbol{Q}$ are maximally monotone, $\boldsymbol{Q}$ is Lipschitzian, the zeros of $\boldsymbol{M}+\boldsymbol{Q}$ are primal-dual solutions

Structure of the problem

- Apply the forward-backward-forward algorithm in $\mathcal{K}$

$$
\begin{array}{l}
\text{For } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{s}_n \approx \boldsymbol{w}_n - \gamma_n \boldsymbol{Q} \boldsymbol{w}_n \\
\boldsymbol{p}_n \approx J_{\gamma_n \boldsymbol{M}}\, \boldsymbol{s}_n \\
\boldsymbol{q}_n \approx \boldsymbol{p}_n - \gamma_n \boldsymbol{Q} \boldsymbol{p}_n \\
\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - \boldsymbol{s}_n + \boldsymbol{q}_n,
\end{array}
\right.
\end{array}
$$

to get a solution to $\mathbf{0} \in \boldsymbol{M}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{x}$, which yields ...

General problem

Splitting algorithm

$$
\left\lfloor
\begin{array}{l}
y_{1,n} \approx x_n - \gamma_n\big(\nabla h(x_n) + \sum_{i=1}^m L_i^* v_{i,n}\big) \\
p_{1,n} \approx \operatorname{prox}_{\gamma_n f}(y_{1,n} + \gamma_n z) + b_{1,n} \\
\text{For } i = 1,\ldots,m \\
\left\lfloor
\begin{array}{l}
y_{2,i,n} \approx v_{i,n} + \gamma_n\big(L_i x_n - \nabla \ell_i^*(v_{i,n})\big) \\
p_{2,i,n} \approx \operatorname{prox}_{\gamma_n g_i^*}(y_{2,i,n} - \gamma_n r_i) \\
q_{2,i,n} \approx p_{2,i,n} + \gamma_n\big(L_i p_{1,n} - \nabla \ell_i^*(p_{2,i,n})\big) \\
v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n}
\end{array}
\right. \\
q_{1,n} \approx p_{1,n} - \gamma_n\big(\nabla h(p_{1,n}) + \sum_{i=1}^m L_i^* p_{2,i,n}\big) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array}
\right.
$$

General mixed problems

Theorem (PLC-Pesquet, 2012)

Set $\beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^m \|L_i\|^2}$ *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be in* $[\varepsilon, (1-\varepsilon)/\beta]$. *Suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^* \circ (\partial g_i \,\square\, \partial \ell_i) \circ (L_i \cdot - r_i) + \nabla h\right).$$

Then there exist a primal solution $\overline{x}$ *and a dual solution* $(\overline{v}_i)_{1\leqslant i\leqslant m}$ *such that*

- $z - \sum_{j=1}^m L_j^* \overline{v_j} \in \partial f(\overline{x}) + \nabla h(\overline{x})$ *and* $(\forall i \in \{1,\ldots,m\})$ $L_i\overline{x} - r_i \in \partial g_i^*(\overline{v_i}) + \nabla \ell_i^*(\overline{v_i})$ *(Kuhn-Tucker on steroids)*
- $x_n \rightharpoonup \overline{x}$ *and* $p_{1,n} \rightharpoonup \overline{x}$
- $(\forall i \in \{1,\ldots,m\})$ $v_{i,n} \rightharpoonup \overline{v_i}$ *and* $p_{2,i,n} \rightharpoonup \overline{v_i}$

General mixed problems

- Full splitting is achieved for f, h, and each L_i, g_i, ℓ_i
- All Lipschitzian operators are activated via explicit steps.

The Gurin-Polyak-Raik problem (1967)

- C_1, C_2: nonempty closed convex subsets of $\mathcal{H}$, C_1 bounded.
- Alternating projections algorithm:

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x_{2n+1} = P_2 x_{2n} \\ x_{2n+2} = P_1 x_{2n+1}. \end{array} \right.$$

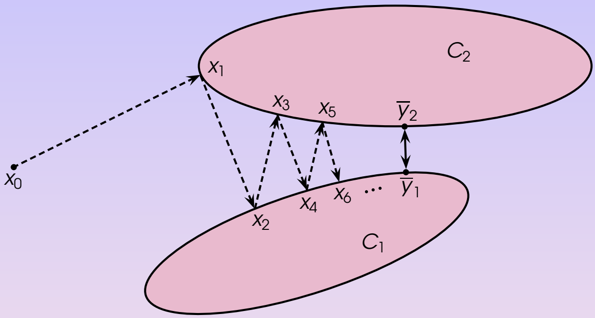
- Convergence (Cheney/Goldstein, 1959 (C_1 compact); Levitin/Polyak, 1966): $x_{2n} \rightharpoonup \overline{y}_1$, $x_{2n+1} \rightharpoonup \overline{y}_2$, and $(\overline{y}_1, \overline{y}_2)$ is a *cycle*:

$$\overline{y}_1 = P_1 \overline{y}_2 \quad \text{and} \quad \overline{y}_2 = P_2 \overline{y}_1.$$

- Variational interpretation: $(\overline{y}_1, \overline{y}_2)$ solves

$$\underset{y_1 \in C_1,\, y_2 \in C_2}{\text{minimize}} \quad \|y_1 - y_2\|.$$

Cyclic projection methods



- **Open problem (1967):** In the case of $m = 2$ sets, the method of alternating projections produces a cycle $(\overline{y}_1, \overline{y}_2)$ that achieves the minimal distance between the two sets. **What about $m \geqslant 3$ sets?**

Cyclic projection methods

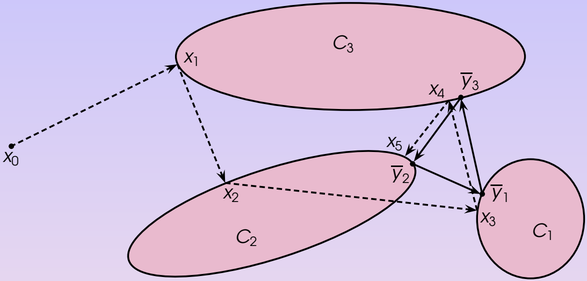
- Algorithm:

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} x_{mn+1} & = P_m x_{mn} \\ x_{mn+2} & = P_{m-1} x_{mn+1} \\ & \vdots \\ x_{mn+m} & = P_1 x_{mn+m-1}. \end{array} \right.$$

- Convergence (Gurin-Polyak-Raik, 1967): $x_{mn} \rightharpoonup \overline{y}_1$, $x_{mn+1} \rightharpoonup \overline{y}_m, \ldots, x_{mn+m-1} \rightharpoonup \overline{y}_2$, and $(\overline{y}_1, \ldots, \overline{y}_m)$ is a *cycle*, i.e.,

$$\overline{y}_1 = P_1 \overline{y}_2, \; \ldots, \; \overline{y}_{m-1} = P_{m-1} \overline{y}_m, \; \overline{y}_m = P_m \overline{y}_1.$$

Cyclic projection methods



Periodic projections produce the cycle $(\overline{y}_1, \overline{y}_2, \overline{y}_3)$, i.e., a Nash equilibrium of the game in which, for every $i \in \{1, \ldots, m\}$, the strategies of player i, belong to C_i and its penalty function is $(x_i)_{1 \leqslant i \leqslant m} \mapsto \|x_i - x_{i+1}\|^2$, that is, player i wants to have strategies as close as possible to the strategies of player $i+1$.

Cyclic projection methods

- Denote by $\text{cyc}(C_1,\ldots,C_m)$ is the set of cycles of $(C_1,\ldots,C_m)$, i.e.,

$$\text{cyc}(C_1,\ldots,C_m)=\big\{(\overline{y}_1,\ldots,\overline{y}_m)\in\mathcal{H}^m \,\big|\, \overline{y}_1=P_1\overline{y}_2,\ldots,\\ \overline{y}_{m-1}=P_{m-1}\overline{y}_m,\ \overline{y}_m=P_m\overline{y}_1\big\}.$$

- **Problem 5 (Gurin-Polyak-Raik, 1967):** Let $m\geqslant 3$ be an integer. Does there exist a function $\Phi\colon\mathcal{H}^m\to\mathbb{R}$ such that, for every ordered family of nonempty closed convex subsets $(C_1,\ldots,C_m)$ of $\mathcal{H}$, $\text{cyc}(C_1,\ldots,C_m)$ is the set of solutions to the variational problem

$$\underset{y_1\in C_1,\ldots,\,y_m\in C_m}{\text{minimize}}\quad \Phi(y_1,\ldots,y_m)\,?$$

Cyclic projection methods

Since the seminal work Gurin-Polyak-Raik in 1967 that first established the existence of cycles, little progress has been made beyond the observation that simple candidates such as

$$\Phi\colon (y_1,\ldots,y_m) \mapsto \|y_1-y_2\|+\cdots+\|y_{m-1}-y_m\|+\|y_m-y_1\|$$

and the like fail. See for instance

- Kosmol (1987)
- PLC (1994)
- Bauschke and Borwein (1994,1997)
- etc.

Cyclic projection methods

Theorem (Baillon-PLC-Cominetti, 2012)

Suppose that $\dim \mathcal{H} \geqslant 2$ and let $\mathbb{N} \ni m \geqslant 3$. There exists ***no*** function $\Phi \colon \mathcal{H}^m \to \mathbb{R}$ such that, for every ordered family of nonempty closed convex subsets $(C_1, \ldots, C_m)$ of $\mathcal{H}$, $\operatorname{cyc}(C_1, \ldots, C_m)$ is the set of solutions to the variational problem

$$\underset{y_1 \in C_1, \ldots, y_m \in C_m}{\text{minimize}} \quad \Phi(y_1, \ldots, y_m).$$

The "right" extension of alternating projections

Theorem (Baillon-PLC-Cominetti, 2012)

Let m be an integer at least equal to 3. For every $i \in I = \{1, \ldots, m\}$, let C_i be a nonempty closed convex subset of $\mathcal{H}$ with projection operator P_i, and let $x_{i,0} \in \mathcal{H}$. Suppose that C_1 is bounded and set

$$(\forall n \in \mathbb{N})(\forall i \in I) \quad x_{i,n+1} = P_i\left(\frac{1}{m-1}\sum_{j \in I \smallsetminus \{i\}} x_{j,n}\right).$$

Then for every $i \in I$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $\overline{y}_i \in C_i$, and $(\overline{y}_i)_{i\in I}$ is a solution to the variational problem

$$\underset{y_1 \in C_1, \ldots, y_m \in C_m}{\text{minimize}} \sum_{(i,j) \in I^2, i<j} \|y_i - y_j\|^2.$$

Moreover, $\overline{y} = (1/m)\sum_{i\in I} \overline{y}_i$ minimizes $\varphi \colon y \mapsto \sum_{i\in I} \|y - P_i y\|^2$.

References

- H. H. Bauschke and PLC, ***Convex Analysis and Monotone Operator Theory in Hilbert Spaces.*** Springer, New York, 2011.
- PLC, Convexité et signal, *Proc. Congrès SMAI'01.* Pompadour, France, May 28-June 1, 2001 (6 pages).
- PLC and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, 2005.
- PLC and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM J. Optim.*, vol. 18, 2007.
- PLC, Dinh Dũng, and B. C. Vũ, Proximity for sums of composite functions, *J. Math. Anal. Appl.*, vol. 380, 2011.
- PLC and J.-C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators, *Set-Valued Var. Anal.*, vol. 20, 2012.
- PLC and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, 2013.
- J.-B. Baillon, P. L. Combettes, and R. Cominetti, There is no variational characterization of the cycles in the method of periodic projections, *J. Funct. Anal.*, vol. 262, 2012.