Numerical Optimization Methods in Imaging

Part IV: Primal-dual methods

jean-christophe@pesquet.eu

PhD Summer School MMLIA – Bologna

Fenchel-Rockafellar duality

Primal problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to\ ]-\infty,+\infty]$, $g\colon \mathcal{G} \to\ ]-\infty,+\infty]$. Let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
We want to

$$\underset{x\in\mathcal{H}}{\operatorname{minimize}}\ \ f(x) + g(Lx).$$

Dual problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to\ ]-\infty,+\infty]$, $g\colon \mathcal{G} \to\ ]-\infty,+\infty]$. Let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
We want to

$$\underset{v\in\mathcal{G}}{\operatorname{minimize}}\ \ f^*(-L^*v) + g^*(v).$$

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let f be a proper fonction from $\mathcal{H}$ to $]-\infty,+\infty]$, g be a proper function from $\mathcal{G}$ to $]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Let

$$\mu = \inf_{x\in\mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v\in\mathcal{G}} f^*(-L^*v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the duality gap.

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let f be a proper fonction from $\mathcal{H}$ to $]-\infty,+\infty]$, g be a proper function from $\mathcal{G}$ to $]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Let

$$\mu = \inf_{x\in\mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v\in\mathcal{G}} f^*(-L^*v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the **duality gap**.

Proof: According to Fenchel-Young inequality, for every $x \in \mathcal{H}$ and $v \in \mathcal{G}$,

$$f(x) + g(Lx) + f^*(-L^*v) + g^*(v) \geq \langle x \mid -L^*v\rangle + \langle Lx \mid v\rangle = 0.$$

Fenchel-Rockafellar duality

Strong duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
If $\mathrm{int}\,(\mathrm{dom}\, g - L(\mathrm{dom}\, f)) \neq \varnothing$, then

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) = -\min_{v \in \mathcal{G}} f^*(-L^* v) + g^*(v) = -\mu^* .$$

Example 1: Linear programming

Let $L \in \mathbb{R}^{K\times N}$, $b \in \mathbb{R}^{K}$, and $c \in \mathbb{R}^{N}$.
The primal problem

$$\text{Primal-LP}: \qquad \underset{x\in[0,+\infty[^N}{\text{minimize}} \ \langle c \mid x\rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP}: \qquad \underset{y\in[0,+\infty[^K}{\text{maximize}} \ \langle b \mid y\rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

Example 1: Linear programming

Let $L \in \mathbb{R}^{K \times N}$, $b \in \mathbb{R}^K$, and $c \in \mathbb{R}^N$.
The primal problem

$$\text{Primal-LP}: \qquad \underset{x \in [0,+\infty[^N}{\operatorname{minimize}} \ \langle c \mid x \rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP}: \qquad \underset{y \in [0,+\infty[^K}{\operatorname{maximize}} \ \langle b \mid y \rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

<u>Proof</u>: Set
$$\begin{cases} (\forall x \in \mathcal{H} = \mathbb{R}^N) \quad f(x) = \langle c \mid x \rangle + \iota_{[0,+\infty[^N}(x), \\ (\forall z \in \mathcal{G} = \mathbb{R}^K) \quad g(z) = \iota_{[0,+\infty[^K}(z - b), \\ y = -v. \end{cases}$$

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.
For every $i \in \{1, \dots, m\}$, let $g_i \colon \mathcal{H} \to]-\infty, +\infty]$ and $h_i \colon \mathcal{H} \to]-\infty, +\infty]$.
The consensus problem is given by

$$\underset{\substack{(x_1,\dots,x_m)\in\mathcal{H}^m \\ x_1=\cdots=x_m}}{\text{minimize}} \quad \sum_{i=1}^{m} g_i(x_i).$$

The sharing problem is given by

$$\underset{\substack{(u_1,\dots,u_m)\in\mathcal{H}^m \\ u_1+\cdots+u_m=\overline{u}}}{\text{maximize}} \quad \sum_{i=1}^{m} h_i(u_i), \qquad \overline{u} \in \mathcal{H}.$$

If, for every $i \in \{1, \dots, m\}$, $h_i = -g_i^*(\cdot - \overline{u}/m)$, then sharing is the dual problem of consensus.

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.
For every $i \in \{1, \ldots, m\}$, let $g_i \colon \mathcal{H} \to]-\infty, +\infty]$ and $h_i \colon \mathcal{H} \to]-\infty, +\infty]$.
The consensus problem is given by

$$\underset{\substack{(x_1,\ldots,x_m)\in\mathcal{H}^m\\ x_1=\cdots=x_m}}{\text{minimize}} \quad \sum_{i=1}^{m} g_i(x_i).$$

The sharing problem is given by

$$\underset{\substack{(u_1,\ldots,u_m)\in\mathcal{H}^m\\ u_1+\cdots+u_m=\overline{u}}}{\text{maximize}} \quad \sum_{i=1}^{m} h_i(u_i), \qquad \overline{u} \in \mathcal{H}.$$

If, for every $i \in \{1, \ldots, m\}$, $h_i = -g_i^*(\cdot - \overline{u}/m)$, then sharing is the dual problem of consensus.

<u>Exercice</u>: Prove this result.

Example 2: Consensus and sharing

Proof: Set $\big(\forall x = (x_1, \ldots, x_m) \in \mathcal{H}^m\big)$ $g(x) = \sum_{i=1}^m g_i(x_i)$ and
$\Lambda_m = \big\{(x_1, \ldots, x_m) \in \mathcal{H}^m \;\big|\; x_1 = \ldots = x_m\big\}$.
The consensus problem reads

$$\underset{x=(x_1,\ldots,x_m)\in\mathcal{H}^m}{\text{minimize}} \quad \underbrace{\iota_{\Lambda_m}(x)}_{f(x)} + g(x).$$

The dual problem is thus $(L = \mathrm{Id})$

$$\underset{v\in\mathcal{H}^m}{\text{minimize}} \;\; f^*(-v) + g^*(v).$$

where, for every $v = (v_i)_{1\le i\le m} \in \mathcal{H}^m$,

$$f^*(v) = \sup_{x\in\Lambda_m} \langle v \mid x\rangle = \sup_{x_1\in\mathcal{H}} \sum_{i=1}^m \langle v_i \mid x_1\rangle = \sup_{x_1\in\mathcal{H}} \left\langle \sum_{i=1}^m v_i \;\middle|\; x_1 \right\rangle = \iota_{\{0\}}\Big(\sum_{i=1}^m v_i\Big).$$

Example 2: Consensus and sharing

Proof: The dual problem is thus ($L = \mathrm{Id}$)

$$\underset{v \in \mathcal{H}^m}{\text{minimize}} \; f^*(-v) + g^*(v).$$

where, for every $v = (v_i)_{1 \le i \le m} \in \mathcal{H}^m$,

$$f^*(v) = \sup_{x \in \Lambda_m} \langle v \mid x \rangle = \sup_{x_1 \in \mathcal{H}} \sum_{i=1}^m \langle v_i \mid x_1 \rangle = \sup_{x_1 \in \mathcal{H}} \left\langle \sum_{i=1}^m v_i \mid x_1 \right\rangle = \iota_{\{0\}}\Big(\sum_{i=1}^m v_i \Big).$$

The dual problem reads

$$\underset{\substack{(v_1,\dots,v_m) \in \mathcal{H}^m \\ v_1 + \cdots + v_m = 0}}{\text{minimize}} \; \sum_{i=1}^m g_i^*(v_i).$$

Setting $(\forall i \in \{1,\dots,m\}) u_i = v_i + \overline{u}/m$ and $h_i(u_i) = -g_i^*(u_i - \overline{u}/m)$ yields the sharing problem.

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\operatorname{zer}\left(\partial f + L^* \circ \partial g \circ L\right) \neq \varnothing \quad \Leftrightarrow \quad \operatorname{zer}\left((-L) \circ \partial f^* \circ (-L^*) + \partial g^*\right) \neq \varnothing.$$

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\operatorname{zer}\left(\partial f + L^* \circ \partial g \circ L\right) \neq \varnothing \quad \Leftrightarrow \quad \operatorname{zer}\left((-L) \circ \partial f^* \circ (-L^*) + \partial g^*\right) \neq \varnothing.$$

Proof:

$$\begin{aligned}
(\exists x \in \mathcal{H})\ \ 0 \in \partial f(x) + L^* \partial g(Lx) \quad &\Leftrightarrow \quad (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} -L^* v \in \partial f(x) \\ v \in \partial g(Lx) \end{cases} \\
&\Leftrightarrow \quad (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} x \in \partial f^*(-L^* v) \\ Lx \in \partial g^*(v) \end{cases} \\
&\Leftrightarrow \quad (\exists v \in \mathcal{G}) \quad\quad 0 \in -L \partial f^*(-L^* v) + \partial g^*(v).
\end{aligned}$$

Fenchel-Rockafellar duality

Duality theorem (2)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

- If there exists $\widehat{x} \in \mathcal{H}$ such that $0 \in \partial f(\widehat{x}) + L^* \partial g(L\widehat{x})$, then $\widehat{x}$ is a solution to the primal problem. Moreover, there exists a solution $\widehat{v}$ to the dual problem such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$.
- If there exists $(\widehat{x}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}$ such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$ then $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem.

If $(\widehat{x}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}$ is such that $-L^*\widehat{v} \in \partial f(\widehat{x})$ and $L\widehat{x} \in \partial g^*(\widehat{v})$, then $(\widehat{x}, \widehat{v})$ is called a **Kuhn-Tucker point**.

Fenchel-Rockafellar duality

Proof:

$$0 \in \partial f(\widehat{x}) + L^* \partial g(L\widehat{x}) \subset \partial(f + g \circ L)(\widehat{x}).$$

Then, according to Fermat's rule, $\widehat{x}$ is a solution to the primal problem.
In addition, there exists $\widehat{v} \in \mathcal{G}$ such that

$$\begin{cases} 0 \in \partial f(\widehat{x}) + L^* \widehat{v} \\ \widehat{v} \in \partial g(L\widehat{x}) \end{cases} \quad \Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ L\widehat{x} \in \partial g^*(\widehat{v}). \end{cases}$$

We have also $\widehat{x} \in \partial f^*(-L^* \widehat{v})$, which implies that

$$0 \in -L \partial f^*(-L^* \widehat{v}) + \partial g^*(\widehat{v}).$$

On the other hand,

$$0 \in -L \partial f^*(-L^* \widehat{v}) + \partial g^*(\widehat{v}) \subset \partial\big(f^* \circ (-L^*) + g^*\big)(\widehat{v})$$

$\Rightarrow$ $\widehat{v}$ solution to the dual problem.
The second assertion is shown in a similar manner.

Fenchel-Rockafellar duality

Particular case:
If $f = \varphi + \frac{1}{2}\| \cdot - z\|^2$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $z \in \mathcal{H}$, then

$$\begin{aligned} -L^*\widehat{v} \in \partial f(\widehat{x}) &\Leftrightarrow -L^*\widehat{v} \in \partial\varphi(\widehat{x}) + \widehat{x} - z \\ &\Leftrightarrow 0 \in \widehat{x} + L^*\widehat{v} - z + \partial\varphi(\widehat{x}). \end{aligned}$$

Hence,

$$\widehat{x} = \mathrm{prox}_{\varphi}(-L^*\widehat{v} + z).$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, $g\colon \mathcal{G} \to \left]-\infty,+\infty\right]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.

The primal problem is equivalent to finding

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$\big(\forall (x,y,v) \in \mathcal{H}\times\mathcal{G}^2\big) \quad \mathcal{L}(x,y,v) = f(x) + g(y) + \langle v \mid Lx - y\rangle \,.$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The primal problem is equivalent to finding

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$\left(\forall (x,y,v) \in \mathcal{H}\times\mathcal{G}^2\right) \quad \mathcal{L}(x,y,v) = f(x) + g(y) + \langle v \mid Lx - y\rangle\,.$$

Proof:

$$\begin{aligned}
\mu = \inf_{x\in\mathcal{H}} f(x) + g(Lx) &= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x) + g(y) + \iota_{\{0\}}(Lx - y)\\
&= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} f(x) + g(y) + \sup_{v\in\mathcal{G}} \langle v \mid Lx - y\rangle\\
&= \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y\rangle\,.
\end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$, $g : \mathcal{G} \to \left]-\infty, +\infty\right]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
The dual problem is equivalent to finding

$$-\mu^* = \sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(x, y, v) = f(x) + g(y) + \langle v \mid Lx - y \rangle .$$

Proof:

$$\begin{aligned}
\mu^* = \inf_{v \in \mathcal{G}} f^*(-L^* v) + g^*(v) &= \inf_{v \in \mathcal{G}} \Big(\sup_{x \in \mathcal{H}} \langle x \mid -L^* v \rangle - f(x) \Big) + \Big(\sup_{y \in \mathcal{G}} \langle y \mid v \rangle - g(y) \Big) \\
&= \inf_{v \in \mathcal{G}} -\Big(\inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y \rangle \Big) \\
&= -\sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y \rangle .
\end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
The dual problem is equivalent to finding

$$-\mu^* = \sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$\big(\forall(x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \quad \mathcal{L}(x,y,v) = f(x) + g(y) + \langle v \mid Lx - y\rangle\,.$$

Remark: v is called the Lagrange multiplier associated with the constraint $Lx = y$.

Link with Lagrange duality

Let $(\widehat{x}, \widehat{y}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}^2$.
$(\widehat{x}, \widehat{y}, \widehat{v})$ is a saddle point of the Lagrange function $\mathcal{L}$ if

$$\big(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2\big) \qquad \mathcal{L}(\widehat{x}, \widehat{y}, v) \leq \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) \leq \mathcal{L}(x, y, \widehat{v}).$$

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $(\widehat{x}, \widehat{y}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}^2$.
Assume that $\operatorname{int}(\operatorname{dom} g - L(\operatorname{dom} f)) \neq \varnothing$.

$$(\widehat{x}, \widehat{y}, \widehat{v}) \quad \text{is a saddle point of the Lagrange function}$$

$$\Updownarrow$$

$$(\widehat{x}, \widehat{v}) \quad \text{is a Kuhn-Tucker point and } \widehat{y} = L\widehat{x}.$$

Link with Lagrange duality

Proof ($\Rightarrow$): If $(\widehat{x}, \widehat{y}, \widehat{v})$ is a saddle point of $\mathcal{L}$, then

$$\begin{cases} 0 \in \partial_x \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = \partial f(\widehat{x}) + L^* \widehat{v} \\ 0 \in \partial_y \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = \partial g(\widehat{y}) - \widehat{v} \\ 0 = \nabla_v \mathcal{L}(\widehat{x}, \widehat{y}, \widehat{v}) = L\widehat{x} - \widehat{y} \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ \widehat{v} \in \partial g(\widehat{y}) \\ \widehat{y} = L\widehat{x} \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} -L^* \widehat{v} \in \partial f(\widehat{x}) \\ L\widehat{x} \in \partial g^*(\widehat{v}) \\ \widehat{y} = L\widehat{x}. \end{cases}$$

Link with Lagrange duality

<u>Proof</u> ($\Leftarrow$): Conversely, assume that $(\widehat{x}, \widehat{v})$ is a Kuhn-Tucker point and $\widehat{y} = L\widehat{x}$. Since $\widehat{x}$ (resp. $\widehat{v}$) is a solution to the primal (resp. dual) problem, then

$$\mu = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \sup_{v\in\mathcal{G}} \mathcal{L}(x,y,v) = \sup_{v\in\mathcal{G}} \mathcal{L}(\widehat{x},\widehat{y},v)$$

$$-\mu^* = \sup_{v\in\mathcal{G}} \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,v) = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,\widehat{v}).$$

By strong duality, $\sup_{v\in\mathcal{G}} \mathcal{L}(\widehat{x},\widehat{y},v) = \inf_{(x,y)\in\mathcal{H}\times\mathcal{G}} \mathcal{L}(x,y,\widehat{v})$, which can be rewritten as

$$\big(\forall (x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \qquad \mathcal{L}(\widehat{x},\widehat{y},v) \le \mathcal{L}(x,y,\widehat{v})$$

or equivalently

$$\big(\forall (x,y,v)\in\mathcal{H}\times\mathcal{G}^2\big) \qquad \mathcal{L}(\widehat{x},\widehat{y},v) \le \mathcal{L}(\widehat{x},\widehat{y},\widehat{v}) \le \mathcal{L}(x,y,\widehat{v}).$$

Alternating-direction method of multipliers

Idea: iterations for finding a saddle point $(\widehat{x}, \widehat{y}, \widehat{v})$:

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n \in \operatorname{Argmin} \mathcal{L}(\cdot, y_n, v_n) \\ y_{n+1} \in \operatorname{Argmin} \mathcal{L}(x_n, \cdot, v_n) \\ v_{n+1} \text{ such that } \mathcal{L}(x_n, y_{n+1}, v_{n+1}) \geq \mathcal{L}(x_n, y_{n+1}, v_n). \end{cases}$$

But the convergence is not guaranteed in general !

Alternating-direction method of multipliers

Idea: iterations for finding a saddle point $(\widehat{x}, \widehat{y}, \widehat{v})$:

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n \in \operatorname{Argmin} \mathcal{L}(\cdot, y_n, v_n) \\ y_{n+1} \in \operatorname{Argmin} \mathcal{L}(x_n, \cdot, v_n) \\ v_{n+1} \text{ such that } \mathcal{L}(x_n, y_{n+1}, v_{n+1}) \geq \mathcal{L}(x_n, y_{n+1}, v_n). \end{cases}$$

But the convergence is not guaranteed in general !

Solution: introduce an Augmented Lagrange function .
Let $\gamma \in \,]0, +\infty[$, we define

$$(\forall (x, y, z) \in \mathcal{H} \times \mathcal{G}^2) \;\; \widetilde{\mathcal{L}}(x, y, z) = f(x) + g(y) + \gamma \langle z \mid Lx - y \rangle + \frac{\gamma}{2} \|Lx - y\|^2$$

The Lagrange multiplier is $v = \gamma z$.

Alternating-direction method of multipliers

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $(\widehat{x}, \widehat{y}, \widehat{v}) \in \mathcal{H} \times \mathcal{G}^2$.
Assume that $\operatorname{int}(\operatorname{dom} g - L(\operatorname{dom} f)) \neq \varnothing$.

$$(\widehat{x}, \widehat{y}, \widehat{z}) \text{ is a saddle point of the augmented Lagrange function}$$

$$\Updownarrow$$

$$(\widehat{x}, \gamma\widehat{z}) \quad \text{is a Kuhn-Tucker point and } \widehat{y} = L\widehat{x}.$$

Alternating-direction method of multipliers

Proof ($\Rightarrow$): If $(\widehat{x}, \widehat{y}, \widehat{z})$ is a saddle point of $\widetilde{\mathcal{L}}$, then

$$\begin{cases} 0 \in \partial_x \widetilde{\mathcal{L}}(\widehat{x}, \widehat{y}, \widehat{z}) = \partial f(\widehat{x}) + \gamma L^* \widehat{z} + \gamma L^*(L\widehat{x} - \widehat{y}) \\ 0 \in \partial_y \widetilde{\mathcal{L}}(\widehat{x}, \widehat{y}, \widehat{z}) = \partial g(\widehat{y}) - \gamma \widehat{z} + \gamma(\widehat{y} - L\widehat{x}) \\ 0 = \nabla_z \widetilde{\mathcal{L}}(\widehat{x}, \widehat{y}, \widehat{z}) = \gamma(L\widehat{x} - \widehat{y}) \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} 0 \in \partial_x \mathcal{L}(\widehat{x}, \widehat{y}, \gamma\widehat{z}) \\ 0 \in \partial_y \mathcal{L}(\widehat{x}, \widehat{y}, \gamma\widehat{z}) \\ 0 = \nabla_v \mathcal{L}(\widehat{x}, \widehat{y}, \gamma\widehat{z}). \end{cases}$$

Alternating-direction method of multipliers

Proof (⇐): Conversely, if $(\widehat{x}, \gamma\widehat{z})$ is a Kuhn-Tucker point and $\widehat{y} = L\widehat{x}$, then it is a saddle point of $\mathcal{L}$. In addition,

$$\big(\forall (x, y, z) \in \mathcal{H} \times \mathcal{G}^2\big) \quad \widetilde{\mathcal{L}}(x, y, z) = \mathcal{L}(x, y, \gamma z) + \frac{\gamma}{2}\|Lx - y\|^2.$$

It can be deduced that

$$\begin{aligned}
\widetilde{\mathcal{L}}(\widehat{x}, \widehat{y}, z) = \mathcal{L}(\widehat{x}, \widehat{y}, \gamma z) &\leq \mathcal{L}(\widehat{x}, \widehat{y}, \gamma\widehat{z}) = \widetilde{\mathcal{L}}(\widehat{x}, \widehat{y}, \widehat{z}) \\
&\leq \mathcal{L}(x, y, \gamma\widehat{z}) \leq \widetilde{\mathcal{L}}(x, y, \widehat{z}).
\end{aligned}$$

Alternating-direction method of multipliers

Algorithm for finding a saddle point:

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \ \widetilde{\mathcal{L}}(x, y_n, z_n) \\ y_{n+1} = \underset{y \in \mathcal{G}}{\operatorname{argmin}} \ \widetilde{\mathcal{L}}(x_n, y, z_n) \\ z_{n+1} \text{ such that } \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_{n+1}) \geq \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_n). \end{cases}$$

By performing a gradient ascent on the Lagrange multiplier,

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \ f(x) + \gamma \left\langle z_n \mid Lx - y_n \right\rangle + \frac{\gamma}{2} \|Lx - y_n\|^2 \\ y_{n+1} = \underset{y \in \mathcal{G}}{\operatorname{argmin}} \ g(y) + \gamma \left\langle z_n \mid Lx_n - y \right\rangle + \frac{\gamma}{2} \|Lx_n - y\|^2 \\ z_{n+1} = z_n + \frac{1}{\gamma} \nabla_z \widetilde{\mathcal{L}}(x_n, y_{n+1}, z_n) \end{cases}$$

$$\Leftrightarrow \quad (\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \ \frac{1}{2} \left\| Lx - y_n + z_n \right\|^2 + \frac{1}{\gamma} f(x) \\ y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}} (z_n + Lx_n) \\ z_{n+1} = z_n + Lx_n - y_{n+1}. \end{cases}$$

Unrelaxed Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.
Let $\gamma \in \,]0, +\infty[$.
We assume that $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. Let $u_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = \operatorname{prox}_{\gamma g} u_n \\ w_n = \operatorname{prox}_{\gamma f}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n. \end{cases}$$

The following properties are satisfied:

1. $u_n \rightharpoonup \widehat{u}$
2. $v_n \rightharpoonup \widehat{v} = \operatorname{prox}_{\gamma g} \widehat{u} \in \operatorname{Argmin}(f + g)$.

Dual form of unrelaxed Douglas-Rachford algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be Hilbert spaces.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.
Let Ł $\in \mathcal{B}(\mathcal{G}, \mathcal{H})$. Let $\gamma \in \,]0, +\infty[$.
We assume that $\operatorname{zer}(\partial(f^* \circ (-L^*)) + \partial g^*) \neq \varnothing$. Let $u_0 \in \mathcal{G}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = \operatorname{prox}_{\gamma g^*} u_n \\ w_n = \operatorname{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n. \end{cases}$$

The following properties are satisfied:

1. $u_n \rightharpoonup \widehat{u}$
2. $v_n \rightharpoonup \widehat{v} = \operatorname{prox}_{\gamma g^*} \widehat{u} \in \operatorname{Argmin}(f^* \circ (-L^*) + g^*)$.

Equivalence between DR and ADMM

Let $\mathcal{H}$ et $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ et $g \in \Gamma_0(\mathcal{G})$.
Let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be such that L^*L is an isomorphism and let $\gamma \in \,]0,+\infty[$.
The dual form of the unrelaxed Douglas-Rachford algorithm

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} v_n = \operatorname{prox}_{\gamma g^*} u_n \\ w_n = \operatorname{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

can be reexpressed as

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \; \frac{1}{2}\|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ s_n = Lx_n \\ y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}}(z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1} \end{cases}$$

by setting $y_n = \gamma^{-1}(u_n - v_n)$ and $z_n = \gamma^{-1} v_n$.

Equivalence between DR and ADMM

Proof: We have

$$w_n = \operatorname{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \Leftrightarrow 2v_n - u_n - w_n \in \gamma \partial\big(f^* \circ (-L^*)\big)(w_n)$$

L^*L isomorphism $\Rightarrow \operatorname{ran} L^* = \mathcal{H}$ and
$\operatorname{dom} f^* \cap \operatorname{int}(\operatorname{ran} L^*) = \operatorname{dom} f^* \neq \varnothing$.
D'où $\partial\big(f^* \circ (-L^*)\big) = -L \circ \partial f^* \circ (-L^*)$ and we can define x_n such that

$$\begin{cases} 2v_n - u_n - w_n = -\gamma L x_n \Leftrightarrow w_n = 2v_n - u_n + \gamma L x_n \\ x_n \in \partial f^*(-L^* w_n) \Leftrightarrow -L^* w_n \in \partial f(x_n) \end{cases}$$
$$\Rightarrow \quad L^*(u_n - 2v_n - \gamma L x_n) \in \partial f(x_n)$$

Since $y_n = (u_n - v_n)/\gamma$ et $z_n = v_n/\gamma$,

$$L^*(y_n - z_n - Lx_n) \in \frac{1}{\gamma}\partial f(x_n) \Leftrightarrow \quad 0 \in L^*(Lx_n - y_n + z_n) + \frac{1}{\gamma}\partial f(x_n)$$
$$\Leftrightarrow \quad x_n = \operatorname*{argmin}_{x \in \mathcal{H}} \frac{1}{2}\|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x).$$

Equivalence between DR and ADMM

On the other hand,

$$v_n = \operatorname{prox}_{\gamma g^*} u_n \Leftrightarrow v_n = u_n - \gamma \operatorname{prox}_{\frac{g}{\gamma}} \left(\frac{u_n}{\gamma} \right)$$
$$\Leftrightarrow y_n = \frac{u_n - v_n}{\gamma} = \operatorname{prox}_{\frac{g}{\gamma}} \left(\frac{u_n}{\gamma} \right).$$

Futhermore,

$$\begin{cases} u_{n+1} = u_n + w_n - v_n \\ 2v_n - u_n - w_n = -\gamma L x_n \Leftrightarrow u_n + w_n = 2v_n + \gamma L x_n \end{cases}$$
$$\Rightarrow \quad \frac{u_{n+1}}{\gamma} = \frac{v_n}{\gamma} + L x_n = z_n + L x_n.$$

Hence

$$y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}} \left(\frac{u_{n+1}}{\gamma} \right) = \operatorname{prox}_{\frac{g}{\gamma}} (z_n + L x_n).$$

Finally

$$z_{n+1} = \frac{v_{n+1}}{\gamma} = \frac{u_{n+1}}{\gamma} - y_{n+1} = z_n + L x_n - y_{n+1}.$$

Augmented Lagrangian method

ADMM (*Alternating-direction method of multipliers*)

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.
Let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ such that L^*L is an isomorphism and let $\gamma \in \,]0,+\infty[$.
We assume that $\mathrm{int}\,(\mathrm{dom}\, g) \cap L(\mathrm{dom}\, f) \neq \varnothing$ or $\mathrm{dom}\, g \cap \mathrm{int}\,\big(L(\mathrm{dom}\, f)\big) \neq \varnothing$, and that $\mathrm{Argmin}(f + g \circ L) \neq \varnothing$. Let $(y_0, z_0) \in \mathcal{G}^2$ and

$$
(\forall n \in \mathbb{N}) \qquad
\begin{cases}
x_n = \underset{x \in \mathcal{H}}{\mathrm{argmin}} \; \frac{1}{2}\|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\
s_n = L x_n \\
y_{n+1} = \mathrm{prox}_{\frac{g}{\gamma}}(z_n + s_n) \\
z_{n+1} = z_n + s_n - y_{n+1}.
\end{cases}
$$

We have:

- $x_n \rightharpoonup \widehat{x}$ where $\widehat{x} \in \mathrm{Argmin}(f + g \circ L)$
- $\gamma z_n \rightharpoonup \widehat{v}$ where $\widehat{v} \in \mathrm{Argmin}(f^* \circ (-L^*) + g^*)$.

Primal-dual methods

Problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces, $f \in \Gamma_0(\mathcal{H})$, $h \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
It is assumed that h is differentiable and has a β-Lipschitzian gradient with $\beta \in \,]0, +\infty[$.
We want to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx) + h(x).$$

Primal-dual methods

- ADMM algorithm: Let $\gamma \in \,]0,+\infty[$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \ \frac{1}{2}\|Lx - y_n + z_n\|^2 + \frac{1}{\gamma}\big(f(x) + h(x)\big) \\ y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}}(z_n + Lx_n) \\ z_{n+1} = z_n + Lx_n - y_{n+1}. \end{cases}$$

- Limitations:
  - Computation of x_n at iteration $n \in \mathbb{N}$ may be complicated.
  - Convergence requires L^*L to be invertible.
  - The smoothness of h is not exploited.

Primal-dual methods

- Idea 1: the optimization problem is reformulated as finding

$$\inf_{x\in\mathcal{H}} f(x) + h(x) + \underbrace{\sup_{v\in\mathcal{G}} \langle v \mid Lx\rangle - g^*(v)}_{g(Lx)} .$$

Primal-dual methods

- Idea 1: the optimization problem is reformulated as finding

$$\inf_{x\in\mathcal{H}} \sup_{v\in\mathcal{G}} f(x) + h(x) + \langle v \mid Lx\rangle - g^*(v).$$

- Arrow-Hurwicz method: Let $(\tau_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ be sequences in $]0,+\infty[$.

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} t_n \in \partial f(x_n) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_n) \\ v_{n+1} = v_n - \sigma_n(s_n - Lx_{n+1}) \end{cases}$$

$\rightsquigarrow$ requires stringent conditions on the choice of the step-size (e.g. decaying to zero)

Primal-dual methods

- Idea 2: Use implicit updates

$$
(\forall n \in \mathbb{N}) \qquad \begin{cases} t_n \in \partial f(x_{n+1}) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_{n+1}) \\ v_{n+1} = v_n - \sigma_n(s_n - L x_{n+1}) \end{cases}
$$

$$
\Leftrightarrow \quad \begin{cases} 0 \in x_{n+1} - x_n + \tau_n(\nabla h(x_n) + L^* v_n) + \tau_n \partial f(x_{n+1}) \\ 0 \in v_{n+1} - v_n - \sigma_n L x_{n+1} + \sigma_n \partial g^*(v_{n+1}) \end{cases}
$$

$$
\Leftrightarrow \quad \begin{cases} x_{n+1} = \operatorname{prox}_{\tau_n f}\big(x_n - \tau_n(\nabla h(x_n) + L^* v_n)\big) \\ v_{n+1} = \operatorname{prox}_{\sigma_n g^*}\big(v_n + \sigma_n L x_{n+1}\big) \end{cases}
$$

$\rightsquigarrow$ still does not converge for constant values of the step-size.

Primal-dual methods

- Idea 3: Use the approximation $x_{n+1} \simeq 2x_{n+1} - x_n$

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau_n f}\big(x_n - \tau_n(\nabla h(x_n) + L^* v_n)\big) \\ v_{n+1} = \mathrm{prox}_{\sigma_n g^*}\big(v_n + \sigma_n L(2x_{n+1} - x_n)\big). \end{cases}$$

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β-Lipschitzian gradient.

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \operatorname{prox}_{\tau f}\big(x_n - \tau\big(\nabla h(x_n) + L^* v_n\big)\big) \\ v_{n+1} = \operatorname{prox}_{\sigma g^*}\big(v_n + \sigma L(2x_{n+1} - x_n)\big). \end{cases}$$

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β-Lipschitzian gradient.
Let $\tau \in \,]0, +\infty[$ and $\sigma \in \,]0, +\infty[$.
We assume that $\tau^{-1} - \sigma\|L\|^2 > \beta/2$ and $\mathrm{zer}\left(\partial f + \nabla h + L^* \partial g L\right) \neq \varnothing$.
Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau f}\big(x_n - \tau\big(\nabla h(x_n) + L^* v_n\big)\big) \\ v_{n+1} = \mathrm{prox}_{\sigma g^*}\big(v_n + \sigma L(2x_{n+1} - x_n)\big). \end{cases}$$

We have:

- $x_n \rightharpoonup \widehat{x} \in \mathrm{Argmin}\big(f + h + g \circ L\big)$
- $v_n \rightharpoonup \widehat{v} \in \mathrm{Argmin}\big((f + h)^* \circ (-L^*) + g^*\big)$.

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β-Lipschitzian gradient.
Let $\tau \in \,]0, +\infty[$ and $\sigma \in \,]0, +\infty[$.
We assume that $\boxed{\tau^{-1} - \sigma\|L\|^2 > \beta/2}$ and $\mathrm{zer}\left(\partial f + \nabla h + L^* \partial g L\right) \neq \varnothing$.
Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau f}\big(x_n - \tau\big(\nabla h(x_n) + L^* v_n\big)\big) \\ v_{n+1} = \mathrm{prox}_{\sigma g^*}\big(v_n + \sigma L(2x_{n+1} - x_n)\big). \end{cases}$$

We have:

- $\boxed{x_n \rightharpoonup \widehat{x} \in \mathrm{Argmin}\big(f + h + g \circ L\big)}$
- $v_n \rightharpoonup \widehat{v} \in \mathrm{Argmin}\big((f + h)^* \circ (-L^*) + g^*\big)$.

Remark: when $g = 0$: Forward-Backward algorithm
when $h = 0$: Chambolle-Pock algorithm

Reformulation

Let

- $\mathbf{K} = \mathcal{H} \oplus \mathcal{G}$
- $\mathbf{A} \colon \mathbf{K} \to 2^{\mathbf{K}} \colon (x, v) \mapsto \big(\partial f(x) + L^* v\big) \times \big(- Lx + \partial g^*(v)\big)$
- $\mathbf{B} \colon \mathbf{K} \to \mathbf{K} \colon (x, v) \mapsto \big(\nabla h(x), 0\big)$
- $\mathbf{V} \colon \mathbf{K} \to \mathbf{K} \colon (x, v) \mapsto \big(\tau^{-1} x - L^* v, -Lx + \sigma^{-1} v\big)$ with $(\rho, \sigma) \in]0, +\infty[$ and $\tau\sigma\|L\|^2 < 1$.

In the renormed space $(\mathbf{K}, \|\cdot\|_{\mathbf{V}})$, $\mathbf{V}^{-1}\mathbf{A}$ is maximally monotone and $\mathbf{V}^{-1}\mathbf{B}$ is cocoercive with constant $\beta^{-1}(\tau^{-1} - \sigma\|L\|^2)$
In addition, finding a zero of the sum of these operators is equivalent to finding a pair of primal-dual solutions.
FB algorithm:

$$(x_{n+1}, v_{n+1}) = J_{\mathbf{V}^{-1}\mathbf{A}}\big((x_n, v_n) - \mathbf{V}^{-1}\mathbf{B}(x_n, v_n)\big).$$

Rescaled form of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.
Let $h \in \Gamma_0(\mathcal{H})$ have a β-Lipschitzian gradient with $\beta \in \,]0, +\infty[$
Let $\tau \in \,]0, +\infty[$ and $\sigma \in \,]0, +\infty[$.
We assume that $\tau^{-1} - \sigma\|L\|^2 \geq \beta/2.$ and $\mathrm{zer}\left(\partial f + \nabla h + L^*\partial g L\right) \neq \varnothing$.

Let $x_0 \in \mathcal{H}$, $v'_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau f}\left(x_n - \tau\left(\nabla h(x_n) + \sigma L^* v'_n\right)\right) \\ v'_{n+1} = (\mathrm{Id} - \mathrm{prox}_{\sigma^{-1} g})\left(v'_n + L(2x_{n+1} - x_n)\right) \end{cases}$$

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of $f + h + g \circ L$.

Normalize: $v'_n = v_n/\sigma$.

Relaxed form of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.
Let $h \in \Gamma_0(\mathcal{H})$ have a β-Lipschitzian gradient with $\beta \in \,]0, +\infty[$
Let $\tau \in \,]0, +\infty[$ and $\sigma \in \,]0, +\infty[$.
We assume that $\tau^{-1} - \sigma\|L\|^2 \geq \beta/2.$ and $\operatorname{zer}\left(\partial f + \nabla h + L^* \partial g L\right) \neq \varnothing$.
Let $\delta = 2 - \beta\left(\tau^{-1} - \sigma\|L\|^2\right)^{-1}/2 \in [1, 2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} p_n = \operatorname{prox}_{\tau f}\left(x_n - \tau\left(\nabla h(x_n) + L^* v_n\right)\right) \\ q_n = \operatorname{prox}_{\sigma g^*}\left(v_n + \sigma L(2p_n - x_n)\right) \\ (x_{n+1}, v_{n+1}) = (x_n, v_n) + \lambda_n(p_n - x_n, q_n - y_n). \end{cases}$$

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of $f + h + g \circ L$.

Exercice

Let $\mathcal{H}$ and $(\mathcal{G}_i)_{1\leq i\leq m}$ be real Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$, let $h \in \Gamma_0(\mathcal{H})$, and, for every $i \in \{1, \ldots, m\}$, let $g_i \in \Gamma_0(\mathcal{G}_i)$ and $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$.
It is assumed that h is differentiable and has a β-Lipschitzian gradient with $\beta \in \,]0, +\infty[$. Propose a primal-dual algorithm to solve

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

Solution

By setting
$\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$
$(\forall y = (y_1, \dots, y_m) \in \mathcal{G}) \begin{cases} \|y\| = \sqrt{\sum_{i=1}^m \omega_i \|y_i\|^2}, \\ g(y) = \sum_{i=1}^m g_i(y_i) \end{cases}$
with $(\omega_i)_{1 \le i \le m} \in [0, +\infty[^m$
$L\colon x \mapsto (L_1 x, \dots, L_m x)$.
The problem reads:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx) + h(x).$$

Algorithm: Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\tau f}\big(x_n - \tau\big(\nabla h(x_n) + \sum_{i=1}^m \sigma_i L_i^* v'_{n,i}\big)\big) \\ v'_{n+1,i} = \big(\mathrm{Id} - \mathrm{prox}_{(\sigma_i)^{-1} g_i}\big)\big(v'_{n,i} + L_i(2x_{n+1} - x_n)\big) \end{cases}$$

where $(\forall i \in \{1, \dots, m\})\ \sigma_i = \omega_i \sigma$.
Convergence condition: $\tau^{-1} - \|\sum_{i=1}^m \sigma_i L_i^* L_i\| \ge \beta/2$