Signal Processing and Networks Optimization
Part VI: Duality

Pierre Borgnat[1], **Jean-Christophe Pesquet**[2], **Nelly Pustelnik**[1]

[1] ENS Lyon – Laboratoire de Physique – CNRS UMR 5672
pierre.borgnat@ens-lyon.fr, nelly.pustelnik@ens-lyon.fr

[2] LIGM – Univ. Paris-Est – CNRS UMR 8049
jean-christophe.pesquet@univ-paris-est.fr

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

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

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

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

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 $\mathrm{int}\,(\mathrm{dom}\, g) \cap L(\mathrm{dom}\, f) \neq \varnothing$ or $\mathrm{dom}\, g \cap \mathrm{int}\,\big(L(\mathrm{dom}\, f)\big) \neq \varnothing$, 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

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

is associated with the the dual problem

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

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

Proof: 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

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

The sharing problem is given by

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

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

Example 2: Consensus and sharing

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

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

The sharing problem is given by

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

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

Proof: Set $L = \mathrm{Id}$ and $\left(\forall x = (x_1, \dots, x_m) \in \mathcal{H}^N\right)$ $\begin{cases} f(x) = \iota_{\Lambda_m}(x), \\ g(x) = \sum_{i=1}^m g_i(x_i) \end{cases}$

where $\Lambda_m = \left\{(x_1, \dots, x_m) \in \mathcal{H}^m \;\middle|\; x_1 = \dots = x_m\right\}$.

Fenchel-Rockafellar duality

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 $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$.
If $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$, then

$$(\forall x \in \mathcal{H}) \qquad \partial f(x) + L^*\partial g(Lx) \subset \partial(f + g \circ L)(x)$$

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 $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ and $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, then

$$\partial f + L^*\partial g L = \partial(f + g \circ L)$$

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 $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

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

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 $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\mathrm{zer}\,(\partial f + L\partial g L^*) \neq \varnothing \quad \Leftrightarrow \quad \mathrm{zer}\,\big((-L)\partial f^*(-L^*) + \partial g^*\big) \neq \varnothing$$

Proof:

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

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 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 $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, 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

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

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

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

<u>Proof</u>:

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

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.
Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, $g\colon \mathcal{G} \to \,]-\infty,+\infty]$, 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

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

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 $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, $g\colon \mathcal{G} \to \left]-\infty,+\infty\right]$, 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

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

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

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

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 $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$.

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

$$\Updownarrow$$

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

Link with Lagrange duality

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

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

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

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

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

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

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 $(\widehat{x}, \widehat{y}, \widehat{z})$:

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

But the convergence is not guaranteed in general !

Alternating-direction method of multipliers

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

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

But the convergence is not guaranteed in general !

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

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

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 $\mathrm{int}\,(\mathrm{dom}\, g) \cap L(\mathrm{dom}\, f) \neq \varnothing$ or $\mathrm{dom}\, g \cap \mathrm{int}\,\big(L(\mathrm{dom}\, f)\big) \neq \varnothing$.

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

$$\Updownarrow$$

$(\widehat{x}, \gamma\widehat{z})$ 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 $\widetilde{L}$, then

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

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

Alternating-direction method of multipliers

Proof ($\Leftarrow$): 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,

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

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

Augmented Lagrangian method

ADMM algorithm (*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 $\gamma \in \,]0, +\infty[$.

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

Augmented Lagrangian method

ADMM algorithm (*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 $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, and that $\operatorname{Argmin}(f + g \circ L) \neq \varnothing$. Let $(y_0, z_0) \in \mathcal{G}^2$ and

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

We have:

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

Augmented Lagrangian method

ADMM algorithm (*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 $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}\big(L(\operatorname{dom} f)\big) \neq \varnothing$, and that $\operatorname{Argmin}(f + g \circ L) \neq \varnothing$. Let $(y_0, z_0) \in \mathcal{G}^2$ and

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

We have:

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

Sketch of the proof: ADMM can be shown to be equivalent to Douglas-Rachford applied to the dual problem.

Augmented Lagrangian method

- Extension: Splitting more than 2 functions

$$\underset{x\in\mathcal{H}}{\text{minimize}} \sum_{i=1}^{M} f_i(x)$$

where $\mathcal{H}$ a real Hilbert space and, for every $i \in \{1,\ldots,m\}$, $f_i \in \Gamma_0(\mathcal{H})$.

- Idea: Reformulate the problem as a consensus one in the product space $\boldsymbol{\mathcal{H}} = \mathcal{H}^m$

$$\underset{x=(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}}{\text{minimize}} \sum_{i=1}^{M} f_i(x_i) + \iota_{\Lambda_m}(x)$$

where

$$\Lambda_m = \left\{(x_1,\ldots,x_m) \in \boldsymbol{\mathcal{H}} \;\middle|\; x_1 = \cdots = x_m\right\}.$$

Augmented Lagrangian method

- Use of ADMM: set

$$\begin{cases} (\forall x = (x_1, \ldots, x_m) \in \boldsymbol{\mathcal{H}}) \qquad f(x) = \sum_{i=1}^{M} f_i(x_i) \\ g = \iota_{\Lambda_m}, \\ L = \mathrm{Id}. \end{cases}$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

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

Augmented Lagrangian method

- Use of ADMM: set

$$\begin{cases} (\forall x = (x_1, \dots, x_m) \in \boldsymbol{\mathcal{H}}) \qquad f(x) = \sum_{i=1}^{M} f_i(x_i) \\ g = \iota_{\Lambda_m}, \\ L = \mathrm{Id}. \end{cases}$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \operatorname{prox}_{\frac{f}{\gamma}}(y_n - z_n) \\ y_{n+1} = P_{\Lambda_m}(z_n + x_n) \\ z_{n+1} = z_n + x_n - y_{n+1}. \end{cases}$$

$$\Rightarrow P_{\Lambda_m} z_{n+1} = P_{\Lambda_m}(z_n + x_n) - P_{\Lambda_m} y_{n+1} = y_{n+1} - y_{n+1} = 0$$

Augmented Lagrangian method

- Use of ADMM: set

$$\begin{cases} (\forall x = (x_1, \ldots, x_m) \in \boldsymbol{\mathcal{H}}) \qquad f(x) = \sum_{i=1}^{M} f_i(x_i) \\ g = \iota_{\Lambda_m}, \\ L = \mathrm{Id}. \end{cases}$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \mathrm{prox}_{\frac{f}{\gamma}}(y_n - z_n) \\ y_{n+1} = P_{\Lambda_m} x_n \\ z_{n+1} = z_n + x_n - y_{n+1} \end{cases}$$

provided that $P_{\Lambda_m} z_0 = 0$.

Augmented Lagrangian method

- Projection onto Λ_m:

$$(\forall u = (u_1, \dots, u_m) \in \boldsymbol{\mathcal{H}}) \qquad P_{\Lambda_m}(u) = (\overline{u}, \dots, \overline{u})$$

where

$$\overline{u} = \frac{1}{m} \sum_{i=1}^{m} u_i.$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_n = \operatorname{prox}_{\frac{f}{\gamma}}(y_n - z_n) \\ y_{n+1} = P_{\Lambda_m} x_n \\ z_{n+1} = z_n + x_n - y_{n+1} \end{cases}$$

provided that $P_{\Lambda_m} z_0 = 0$.

Augmented Lagrangian method

- Projection onto Λ_m:

$$(\forall u = (u_1, \ldots, u_m) \in \boldsymbol{\mathcal{H}}) \qquad P_{\Lambda_m}(u) = (\overline{u}, \ldots, \overline{u})$$

where

$$\overline{u} = \frac{1}{m} \sum_{i=1}^{m} u_i.$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n,i} = \mathrm{prox}_{\frac{f_i}{\gamma}} \left(\overline{x}_n - z_{n,i} \right), \quad i \in \{1, \ldots, m\} \\ \overline{x}_n = \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} x_{n,i} \\ z_{n+1,i} = z_{n,i} + x_{n,i} - \overline{x}_n, \quad i \in \{1, \ldots, m\} \end{cases}$$

provided that $\sum_{i=1}^{m} z_{0,i} = 0$.

Augmented Lagrangian method

- Projection onto Λ_m:

$$(\forall u = (u_1, \ldots, u_m) \in \mathcal{H}) \qquad P_{\Lambda_m}(u) = (\overline{u}, \ldots, \overline{u})$$

where

$$\overline{u} = \frac{1}{m} \sum_{i=1}^{m} u_i.$$

- Resulting SDMM (*Simultaneous Direction Method of Multipliers*):

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n,i} = \operatorname{prox}_{\frac{f_i}{\gamma}} \left(\overline{x}_n - z_{n,i}\right), \quad i \in \{1, \ldots, m\} \\ \overline{x}_n = \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} x_{n,i} \\ z_{n+1,i} = z_{n,i} + x_{n,i} - \overline{x}_n, \quad i \in \{1, \ldots, m\} \end{cases}$$

$\rightsquigarrow$ Parallel computing of proximity operator, but centralized computation of average.

Distributed optimization

- $\mathbb{V} = \{1, \ldots, m\}$ set of indices of the vertices of an undirected nonreflexive graph
  $x \in \boldsymbol{\mathcal{H}}$ vector of node weights
  $E = \{e_{i,j} \mid (i,j) \in \mathbb{E}\}$ (with $i < j$) set of edges.

Distributed optimization

- $\mathbb{V} = \{1, \ldots, m\}$ set of indices of the vertices of an undirected nonreflexive graph
  $x \in \boldsymbol{\mathcal{H}}$ vector of node weights
  $E = \big\{ e_{i,j} \;\big|\; (i,j) \in \mathbb{E} \big\}$ (with $i < j$) set of edges.

- The graph is assumed to be connected:

$$x \in \Lambda_m \qquad \Leftrightarrow \qquad \big(\forall (i,j) \in \mathbb{E}\big) \quad (x_i, x_j) \in \Lambda_2$$

Distributed optimization

- $\mathbb{V} = \{1, \dots, m\}$ set of indices of the vertices of an undirected nonreflexive graph
  $x \in \boldsymbol{\mathcal{H}}$ vector of node weights
  $E = \big\{e_{i,j} \;\big|\; (i,j) \in \mathbb{E}\big\}$ (with $i < j$) set of edges.

- The graph is assumed to be connected:

$$x \in \Lambda_m \qquad \Leftrightarrow \qquad \big(\forall (i,j) \in \mathbb{E}\big) \quad (x_i, x_j) \in \Lambda_2$$

- For every $(i,j) \in \mathbb{E}$, we define
  - decimation operators: $L_{i,j} \colon \boldsymbol{\mathcal{H}} \to \mathcal{H}^2 \colon x \mapsto (x_i, x_j)$
  - interpolation operators: $L_{i,j}^* \colon \mathcal{H}^2 \to \boldsymbol{\mathcal{H}} \colon (y_1, y_2) \to x = (x_{i'})_{1 \le i' \le m}$
    where

$$(\forall i' \in \{1, \dots, m\}) \qquad x_{i'} = \begin{cases} y_1 & \text{if } i' = i \\ y_2 & \text{if } i' = j \\ 0 & \text{otherwise.} \end{cases}$$

Distributed optimization

- $\mathbb{V} = \{1, \ldots, m\}$ set of indices of the vertices of an undirected nonreflexive graph
  $x \in \boldsymbol{\mathcal{H}}$ vector of node weights
  $E = \{e_{i,j} \mid (i,j) \in \mathbb{E}\}$ (with $i < j$) set of edges.
- The graph is assumed to be connected:

$$x \in \Lambda_m \qquad \Leftrightarrow \qquad \big(\forall (i,j) \in \mathbb{E}\big) \quad (x_i, x_j) \in \Lambda_2$$

- For every $(i,j) \in \mathbb{E}$, we define
  - decimation operators: $L_{i,j} \colon \boldsymbol{\mathcal{H}} \to \mathcal{H}^2 \colon x \mapsto (x_i, x_j)$
  - interpolation operators: $L_{i,j}^* \colon \mathcal{H}^2 \to \boldsymbol{\mathcal{H}} \colon (y_1, y_2) \to x = (x_{i'})_{1 \leq i' \leq m}$
    where

$$(\forall i' \in \{1, \ldots, m\}) \qquad x_{i'} = \begin{cases} y_1 & \text{if } i' = i \\ y_2 & \text{if } i' = j \\ 0 & \text{otherwise.} \end{cases}$$

  The concatenated decimation operator $L = (L_{i,j})_{(i,j) \in \mathbb{E}}$ is such that

$$L^* L = \sum_{(i,j) \in \mathbb{E}} L_{i,j}^* L_{i,j} = \text{Diag}(d_1 \mathrm{Id}, \ldots, d_m \mathrm{Id})$$

  where, for every $i \in \{1, \ldots, m\}$, d_i is the degree of vertex of index i.

Distributed optimization

- Distributed consensus

$$\underset{x=(x_1,\ldots,x_m)\in\mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^{M} f_i(x_i) + \underbrace{\iota_{\Lambda_m}(x)}_{\displaystyle\sum_{(i,j)\in\mathbb{E}} \iota_{\Lambda_2}\big((x_i, x_j)\big)}$$

Distributed optimization

- Distributed consensus

$$\underset{x=(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}}{\text{minimize}} \quad \sum_{i=1}^{M} f_i(x_i) + \sum_{(i,j)\in\mathbb{E}} \iota_{\Lambda_2}\big(L_{i,j}x\big)$$

Distributed optimization

- Distributed consensus

$$\underset{x=(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}}{\text{minimize}} \quad \sum_{i=1}^{M} f_i(x_i) + \sum_{(i,j)\in\mathbb{E}} \iota_{\Lambda_2}(L_{i,j}x)$$

- Use of ADMM: Set

$$(\forall x=(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}) \qquad f(x)=\sum_{i=1}^{M} f_i(x_i)$$

$$(\forall y=(y_{i,j})_{(i,j)\in\mathbb{E}}\in\mathcal{H}^{|\mathbb{E}|}) \qquad g(y)=\sum_{(i,j)\in\mathbb{E}} \iota_{\Lambda_2}(y_{i,j}).$$

Distributed optimization

- Use of ADMM: Set

$$(\forall x = (x_1, \ldots, x_m) \in \boldsymbol{\mathcal{H}}) \qquad f(x) = \sum_{i=1}^{M} f_i(x_i)$$

$$\left(\forall y = (y_{i,j})_{(i,j)\in\mathbb{E}} \in \mathcal{H}^{|\mathbb{E}|}\right) \qquad g(y) = \sum_{(i,j)\in\mathbb{E}} \iota_{\Lambda_2}\big(y_{i,j}\big).$$

- Distributed ADMM (1):

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

Distributed optimization

- Distributed ADMM (1):

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

- Distributed ADMM (2): provided that $\big(\forall (i,j) \in \mathbb{E}\big)\ P_{\Lambda_2} z_{0,i,j} = 0$,

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} 0 \in L^*(Lx_n - y_n + z_n) + \frac{1}{\gamma} \partial f(x_n) \\ s_{n,i,j} = (x_{n,i}, x_{n,j}), \quad (i,j) \in \mathbb{E} \\ y_{n+1,i,j} = P_{\Lambda_2} (z_{n,i,j} + s_{n,i,j}) = P_{\Lambda_2} s_{n,i,j}, \quad (i,j) \in \mathbb{E} \\ z_{n+1,i,j} = z_{n,i,j} + s_{n,i,j} - y_{n+1,i,j}, \quad (i,j) \in \mathbb{E}. \end{cases}$$

Distributed optimization

- Distributed ADMM (2): provided that $\big(\forall (i,j)\in\mathbb{E}\big)\; P_{\Lambda_2} z_{0,i,j}=0$,

$$(\forall n\in\mathbb{N})\qquad \begin{cases} 0\in L^*(Lx_n-y_n+z_n)+\frac{1}{\gamma}\partial f(x_n)\\ s_{n,i,j}=(x_{n,i},x_{n,j}),\;\;(i,j)\in\mathbb{E}\\ y_{n+1,i,j}=P_{\Lambda_2}(z_{n,i,j}+s_{n,i,j})=P_{\Lambda_2}s_{n,i,j},\;\;(i,j)\in\mathbb{E}\\ z_{n+1,i,j}=z_{n,i,j}+s_{n,i,j}-y_{n+1,i,j},\;\;(i,j)\in\mathbb{E}. \end{cases}$$

- Distributed ADMM (3):

$$(\forall n\in\mathbb{N})\quad \begin{cases} 0\in d_i x_{n,i}-\big(L^*(y_n-z_n)\big)_i+\frac{1}{\gamma}\partial f_i(x_{n,i}),\;\;i\in\mathbb{V}\\ s_{n,i,j}=(x_{n,i},x_{n,j}),\;\;(i,j)\in\mathbb{E}\\ y_{n+1,i,j}=P_{\Lambda_2}s_{n,i,j},\;\;(i,j)\in\mathbb{E}\\ z_{n+1,i,j}=z_{n,i,j}+s_{n,i,j}-y_{n+1,i,j},\;\;(i,j)\in\mathbb{E}. \end{cases}$$

Distributed optimization

- Distributed ADMM (3):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} 0 \in d_i x_{n,i} - \big(L^*(y_n - z_n)\big)_i + \frac{1}{\gamma}\partial f_i(x_{n,i}), & i \in \mathbb{V} \\ s_{n,i,j} = (x_{n,i}, x_{n,j}), & (i,j) \in \mathbb{E} \\ y_{n+1,i,j} = P_{\Lambda_2} s_{n,i,j}, & (i,j) \in \mathbb{E} \\ z_{n+1,i,j} = z_{n,i,j} + s_{n,i,j} - y_{n+1,i,j}, & (i,j) \in \mathbb{E}. \end{cases}$$

- Distributed ADMM (4):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n,i} = \mathrm{prox}_{\frac{f_i}{\gamma d_i}}\Big(d_i^{-1}\big(L^*(y_n - z_n)\big)_i\Big) & i \in \mathbb{V} \\ \overline{y}_{n+1,i,j} = \frac{1}{2}(x_{n,i} + x_{n,j}), & (i,j) \in \mathbb{E} \\ y_{n+1,i,j} = (\overline{y}_{n+1,i,j}, \overline{y}_{n+1,i,j}) \\ z_{n+1,i,j,1} = z_{n,i,j,1} + x_{n,i} - \overline{y}_{n+1,i,j}, & (i,j) \in \mathbb{E} \\ z_{n+1,i,j,2} = z_{n,i,j,2} + x_{n,j} - \overline{y}_{n+1,i,j}, & (i,j) \in \mathbb{E}. \end{cases}$$

Distributed optimization

- Distributed ADMM (4):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}} \left(d_i^{-1} \big(L^*(y_n - z_n) \big)_i \right) & i \in \mathbb{V} \\ \overline{y}_{n+1,i,j} = \frac{1}{2}(x_{n,i} + x_{n,j}), \quad (i,j) \in \mathbb{E} \\ y_{n+1,i,j} = (\overline{y}_{n+1,i,j}, \overline{y}_{n+1,i,j}) \\ z_{n+1,i,j,1} = z_{n,i,j,1} + x_{n,i} - \overline{y}_{n+1,i,j}, \quad (i,j) \in \mathbb{E} \\ z_{n+1,i,j,2} = z_{n,i,j,2} + x_{n,j} - \overline{y}_{n+1,i,j}, \quad (i,j) \in \mathbb{E}. \end{cases}$$

- Distributed ADMM (5):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}} \left(d_i^{-1} \Big(\sum_{j:(i,j)\in\mathbb{E}} (\overline{y}_{n,i,j} - z_{n,i,j,1}) + \sum_{j:(j,i)\in\mathbb{E}} (\overline{y}_{n,j,i} - z_{n,j,i,2}) \Big) \right), \quad i \in \mathbb{V} \\ \overline{y}_{n+1,i,j} = \frac{1}{2}(x_{n,i} + x_{n,j}), \quad (i,j) \in \mathbb{E} \\ z_{n+1,i,j,1} = z_{n,i,j,1} + x_{n,i} - \overline{y}_{n+1,i,j}, \quad (i,j) \in \mathbb{E} \\ z_{n+1,j,i,2} = z_{n,j,i,2} + x_{n,i} - \overline{y}_{n+1,j,i}, \quad (j,i) \in \mathbb{E}. \end{cases}$$

$$\Rightarrow \begin{cases} \overline{y}_{n+1,i,j} - z_{n+1,i,j,1} = 2\overline{y}_{n+1,i,j} - x_{n,i} - z_{n,i,j,1} = x_{n,j} - z_{n,i,j,1} \\ \overline{y}_{n+1,j,i} - z_{n+1,j,i,2} = x_{n,j} - z_{n,j,i,2} \end{cases}$$

Distributed optimization

- Distributed ADMM (5):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}} \Big(d_i^{-1} \big(\sum_{j:(i,j)\in\mathbb{E}} (\overline{y}_{n,i,j} - z_{n,i,j,1}) + \sum_{j:(j,i)\in\mathbb{E}} (\overline{y}_{n,j,i} - z_{n,j,i,2}) \big) \Big), \quad i \in \mathbb{V} \\ \overline{y}_{n+1,i,j} = \frac{1}{2}(x_{n,i} + x_{n,j}), \quad (i,j) \in \mathbb{E} \\ z_{n+1,i,j,1} = z_{n,i,j,1} + x_{n,i} - \overline{y}_{n+1,i,j}, \quad (i,j) \in \mathbb{E} \\ z_{n+1,j,i,2} = z_{n,j,i,2} + x_{n,i} - \overline{y}_{n+1,j,i}, \quad (j,i) \in \mathbb{E}. \end{cases}$$

$$\Rightarrow \begin{cases} \overline{y}_{n+1,i,j} - z_{n+1,i,j,1} = 2\overline{y}_{n+1,i,j} - x_{n,i} - z_{n,i,j,1} = x_{n,j} - z_{n,i,j,1} \\ \overline{y}_{n+1,j,i} - z_{n+1,j,i,2} = x_{n,j} - z_{n,j,i,2} \end{cases}$$

- Distributed ADMM (6):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} z_{n+1,i,j,1} = z_{n,i,j,1} + \frac{1}{2}(x_{n,i} - x_{n,j}), \quad (i,j) \in \mathbb{E} \\ z_{n+1,j,i,2} = z_{n,j,i,2} + \frac{1}{2}(x_{n,i} - x_{n,j}), \quad (j,i) \in \mathbb{E} \\ \overline{x}_{n,i} = d_i^{-1} \big(\sum_{j:(i,j)\in\mathbb{E}} x_{n,j} + \sum_{j:(j,i)\in\mathbb{E}} x_{n,j} \big) \quad i \in \mathbb{V} \\ \overline{z}_{n,i} = d_i^{-1} \big(\sum_{j:(i,j)\in\mathbb{E}} z_{n,i,j,1} + \sum_{j:(j,i)\in\mathbb{E}} z_{n,j,i,2} \big) \quad i \in \mathbb{V} \\ x_{n+1,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}} \Big(\overline{x}_{n,i} - \overline{z}_{n,i} \Big), \quad i \in \mathbb{V} \end{cases}$$

Distributed optimization

- Distributed ADMM (6):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} z_{n+1,i,j,1} = z_{n,i,j,1} + \frac{1}{2}(x_{n,i} - x_{n,j}), & (i,j) \in \mathbb{E} \\ z_{n+1,j,i,2} = z_{n,j,i,2} + \frac{1}{2}(x_{n,i} - x_{n,j}), & (j,i) \in \mathbb{E} \\ \overline{x}_{n,i} = d_i^{-1}\big(\sum_{j:(i,j)\in\mathbb{E}} x_{n,j} + \sum_{j:(j,i)\in\mathbb{E}} x_{n,j}\big) & i \in \mathbb{V} \\ \overline{z}_{n,i} = d_i^{-1}\big(\sum_{j:(i,j)\in\mathbb{E}} z_{n,i,j,1} + \sum_{j:(j,i)\in\mathbb{E}} z_{n,j,i,2}\big) & i \in \mathbb{V} \\ x_{n+1,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}}\Big(\overline{x}_{n,i} - \overline{z}_{n,i}\Big), & i \in \mathbb{V} \end{cases}$$

- Distributed ADMM (final form):

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \overline{x}_{n,i} = d_i^{-1}\big(\sum_{j:(i,j)\in\mathbb{E}} x_{n,j} + \sum_{j:(j,i)\in\mathbb{E}} x_{n,j}\big) & i \in \mathbb{V} \\ \overline{z}_{n+1,i} = \overline{z}_{n,i} + d_i x_{n,i} - \overline{x}_{n,i} \\ x_{n+1,i} = \operatorname{prox}_{\frac{f_i}{\gamma d_i}}\Big(\overline{x}_{n,i} - \overline{z}_{n,i}\Big), & i \in \mathbb{V}. \end{cases}$$

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 have 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

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

- Arrow-Hurwitz 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

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

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

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

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

Primal-dual methods

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

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

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.

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

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 $\tau^{-1} - \sigma \|L\|^2 \geq \beta/2$ and $\operatorname{zer}\left(\partial f + \nabla h + L^* \partial g L\right) \neq \varnothing$.
Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

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

We have:

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

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 $\tau^{-1} - \sigma\|L\|^2 \geq \beta/2$ and $\operatorname{zer}\big(\partial f + \nabla h + L^*\partial g L\big) \neq \varnothing$.
Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

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

We have:

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

Sketch of the proof: rewrite the algorithm as a Forward-Backward iteration for solving a saddle point problem.

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 $i \in \{1,\dots,m\}$, let $g_i \in \Gamma_0(\mathcal{G})$ and $L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$.
It is assumed that h is differentiable and have a β-Lipschitzian gradient with $\beta \in \,]0,+\infty[$. Propose a primal-dual algorithm to solve

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