Fonctions Perspectives et Statistique en Grande Dimension

Patrick L. Combettes

Department of Mathematics
North Carolina State University
Raleigh, NC 27695, USA

Basé sur un travail conjoint avec C. L. Müller du Flatiron Institute, New York

Journées MAS-MODE de la SMAI, Institut Henri Poincaré, Paris, 9 janvier 2017

Some optimization problems in statistics

- Standard finite-dimensional linear model: Observation $z = Xb + \sigma e = (\zeta_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^n$, unknown $b = (\beta_j)_{1 \leqslant j \leqslant p} \in \mathbb{R}^p$
  - Belloni et al.'s square-root lasso (2011):

$$\underset{b \in \mathbb{R}^p}{\text{minimize}} \; \|Xb - z\|_2 + \alpha \|b\|_1$$

Some optimization problems in statistics

- Standard finite-dimensional linear model: Observation $z = Xb + \sigma e = (\zeta_i)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$, unknown $b = (\beta_j)_{1\leqslant j\leqslant p} \in \mathbb{R}^p$
  - Belloni et al.'s square-root lasso (2011):

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \;\; \|Xb - z\|_2 + \alpha\|b\|_1$$

  - Sun and Zhang's scaled lasso (2012):

$$\underset{b\in\mathbb{R}^p,\, \sigma>0}{\text{minimize}} \;\; \frac{1}{2n}\frac{\|Xb - z\|_2^2}{\sigma} + \frac{\sigma}{2} + \alpha\|b\|_1$$

Some optimization problems in statistics

- Standard finite-dimensional linear model: Observation $z = Xb + \sigma e = (\zeta_i)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$, unknown $b = (\beta_j)_{1\leqslant j\leqslant p} \in \mathbb{R}^p$
  - Belloni et al.'s square-root lasso (2011):

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \;\; \|Xb - z\|_2 + \alpha\|b\|_1$$

  - Sun and Zhang's scaled lasso (2012):

$$\underset{b\in\mathbb{R}^p,\, \sigma>0}{\text{minimize}} \;\; \frac{1}{2n}\frac{\|Xb - z\|_2^2}{\sigma} + \frac{\sigma}{2} + \alpha\|b\|_1$$

  - Lederer&Müller TREX estimator (2015):

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \;\; \frac{\|Xb - z\|_2^2}{\|X^\top(Xb - z)\|_\infty} + \alpha\|b\|_1$$

Some optimization problems in statistics

- Standard finite-dimensional linear model: Observation $z = Xb + \sigma e = (\zeta_i)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$, unknown $b = (\beta_j)_{1\leqslant j\leqslant p} \in \mathbb{R}^p$
  - Belloni et al.'s square-root lasso (2011):

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \;\; \|Xb - z\|_2 + \alpha\|b\|_1$$

  - Sun and Zhang's scaled lasso (2012):

$$\underset{b\in\mathbb{R}^p,\, \sigma>0}{\text{minimize}} \;\; \frac{1}{2n}\frac{\|Xb - z\|_2^2}{\sigma} + \frac{\sigma}{2} + \alpha\|b\|_1$$

  - Lederer&Müller TREX estimator (2015):

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \;\; \frac{\|Xb - z\|_2^2}{\|X^\top(Xb - z)\|_\infty} + \alpha\|b\|_1$$

  - Owen's penalized concomitant M-estimators (2007):

$$\underset{b,\sigma,\tau}{\text{minimize}} \;\; n\sigma + \sigma\sum_{i=1}^{n}\text{Huber}\left(\frac{\zeta_i - \langle b \mid x_i\rangle}{\sigma}\right) + p\tau + \tau\sum_{j=1}^{p}\text{Berhu}\left(\frac{\beta_j}{\tau}\right)$$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\displaystyle\int_{\mathbb{R}^N} \left|x(t)^{1/p} - y(t)^{1/p}\right|^p dt$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\int_{\mathbb{R}^N} \left| x(t)^{1/p} - y(t)^{1/p} \right|^p dt$
  - Kullback-Leibler: $\int_{\mathbb{R}^N} x(t) \ln \left(\frac{x(t)}{y(t)} \right) dt$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\int_{\mathbb{R}^N} \left|x(t)^{1/p} - y(t)^{1/p}\right|^p dt$
  - Kullback-Leibler: $\int_{\mathbb{R}^N} x(t) \ln\left(\frac{x(t)}{y(t)}\right) dt$
  - Rényi: $\int_{\mathbb{R}^N} x(t)^{\alpha} y(t)^{1-\alpha} dt$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)} dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\displaystyle\int_{\mathbb{R}^N} \left|x(t)^{1/p} - y(t)^{1/p}\right|^p dt$
  - Kullback-Leibler: $\displaystyle\int_{\mathbb{R}^N} x(t) \ln\left(\frac{x(t)}{y(t)}\right) dt$
  - Rényi: $\displaystyle\int_{\mathbb{R}^N} x(t)^\alpha y(t)^{1-\alpha} dt$
  - Pearson: $\displaystyle\int_{\mathbb{R}^N} \frac{|x(t) - y(t)|^2}{y(t)} dt$

Some optimization problems in statistics

- Problems involving the Fisher information of a multi-dimensional density $x > 0$ (Fisher, 1925):

$$\int_{\mathbb{R}^N} \frac{\|\nabla x(t)\|_2^2}{x(t)}\,dt$$

- Problems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\displaystyle\int_{\mathbb{R}^N} \left|x(t)^{1/p} - y(t)^{1/p}\right|^p dt$
  - Kullback-Leibler: $\displaystyle\int_{\mathbb{R}^N} x(t) \ln\left(\frac{x(t)}{y(t)}\right) dt$
  - Rényi: $\displaystyle\int_{\mathbb{R}^N} x(t)^{\alpha} y(t)^{1-\alpha}\, dt$
  - Pearson: $\displaystyle\int_{\mathbb{R}^N} \frac{|x(t) - y(t)|^2}{y(t)}\, dt$
- *What is the common structure underlying these formulations?*

Perspective functions: Definition

- $\mathcal{H}$, $\mathcal{G}$ real Hilbert spaces
- $\Gamma_0(\mathcal{G})$: set of lower semicontinuous convex functions from $\mathcal{G}$ to $]-\infty,+\infty]$ with $\operatorname{dom}\varphi = \{x \in \mathcal{G} \mid \varphi(x) < +\infty\}$.
- $\varphi \in \Gamma_0(\mathcal{G})$
- $\operatorname{rec}\varphi$ is the recession function of φ: given $z \in \operatorname{dom}\varphi$,

$$(\forall y \in \mathcal{G}) \quad (\operatorname{rec}\varphi)(y) = \sup_{x \in \operatorname{dom}\varphi} \big(\varphi(x+y) - \varphi(y)\big)$$

- (Lower semicontinuous envelope of the) Perspective function of φ:

$$\widetilde{\varphi}\colon \mathbb{R} \times \mathcal{G} \to \,]-\infty,+\infty]\colon (\eta, y) \mapsto \begin{cases} \eta\varphi(y/\eta), & \text{if } \eta > 0;\\ (\operatorname{rec}\varphi)(y), & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0. \end{cases}$$

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$
- $\widetilde{\varphi}$ is positively homogeneous

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$
- $\widetilde{\varphi}$ is positively homogeneous
- Let $C = \{(\mu, u) \in \mathbb{R} \times \mathcal{G} \mid \mu + \varphi^*(u) \leqslant 0\}$. Then $\widetilde{\varphi} = \sigma_C$ and $(\widetilde{\varphi})^* = \iota_C$

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- $\widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$
- $\widetilde{\varphi}$ is positively homogeneous
- Let $C = \big\{(\mu, u) \in \mathbb{R} \times \mathcal{G} \mid \mu + \varphi^*(u) \leqslant 0\big\}$. Then $\widetilde{\varphi} = \sigma_C$ and $(\widetilde{\varphi})^* = \iota_C$
- Let $\eta \in \mathbb{R}$ and $y \in \mathcal{G}$. Then $\partial\widetilde{\varphi}(\eta, y) =$

$$\begin{cases} \big\{\big(\varphi(y/\eta) - \langle y \mid u\rangle/\eta, u\big) \mid u \in \partial\varphi(y/\eta)\big\}, & \text{if } \eta > 0; \\ \big\{(\mu, u) \in C \mid \sigma_{\operatorname{dom}\varphi^*}(y) = \langle u \mid y\rangle\big\}, & \text{if } \eta = 0 \text{ and } y \neq 0; \\ C, & \text{if } \eta = 0 \text{ and } y = 0; \\ \varnothing, & \text{if } \eta < 0 \end{cases}$$

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\lambda \in \,]0, +\infty[$. Then $[\lambda\varphi + \psi]^{\sim} = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\lambda \in\,]0,+\infty[$. Then $[\lambda\varphi + \psi]^{\sim} = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.
- Let $\Lambda\colon \mathcal{H} \to \mathcal{G}$ be linear, bounded, and such that $\operatorname{ran}\Lambda \cap \operatorname{dom}\varphi \neq \varnothing$. Set $\widetilde{\Lambda}\colon \mathbb{R} \oplus \mathcal{H} \to \mathbb{R} \oplus \mathcal{G}\colon (\xi, x) \mapsto (\xi, \Lambda x)$. Then $[\varphi \circ \Lambda]^{\sim} = \widetilde{\varphi} \circ \widetilde{\Lambda} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Perspective functions: Properties

Let $\varphi \in \Gamma_0(\mathcal{G})$. Then

- Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\lambda \in \,]0,+\infty[$. Then $[\lambda\varphi + \psi]^{\sim} = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.
- Let $\Lambda \colon \mathcal{H} \to \mathcal{G}$ be linear, bounded, and such that $\operatorname{ran}\Lambda \cap \operatorname{dom}\varphi \neq \varnothing$. Set $\widetilde{\Lambda} \colon \mathbb{R} \oplus \mathcal{H} \to \mathbb{R} \oplus \mathcal{G} \colon (\xi, x) \mapsto (\xi, \Lambda x)$. Then $[\varphi \circ \Lambda]^{\sim} = \widetilde{\varphi} \circ \widetilde{\Lambda} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.
- Let $\psi \in \Gamma_0(\mathcal{G})$ and let C be a closed convex subset of $\mathcal{G}$ such that $C \cap \operatorname{dom}\psi \neq \varnothing$. Set

$$g \colon (\eta, y) \mapsto \begin{cases} \eta\psi(y/\eta), & \text{if } \eta > 0 \text{ and } y \in \eta(C \cap \operatorname{dom}\psi); \\ (\operatorname{rec}\psi)(y), & \text{if } \eta = 0 \text{ and } y \in \operatorname{rec} C; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $g = [\iota_C + \psi]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Perspective functions: Examples

- Let $\psi \in \Gamma_0(\mathcal{G})$ and let $\operatorname{env}\psi\colon y \mapsto \inf_{x\in\mathcal{G}}(\psi(x) + \|y-x\|^2/2)$ be the Moreau envelope of ψ. Set

$$g\colon (\eta, y) \mapsto \begin{cases} \dfrac{\|y\|^2}{2\eta} - \eta(\operatorname{env}\psi)(y/\eta), & \text{if } \ \eta > 0;\\ \sigma_{\operatorname{dom}\psi}(y), & \text{if } \ \eta = 0;\\ +\infty, & \text{if } \ \eta < 0. \end{cases}$$

Then $g = [\operatorname{env}(\psi^*)]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Perspective functions: Examples

- Take $\psi = \iota_{B(0;1)}$ in previous example and set

$$g\colon (\eta, y) \mapsto \begin{cases} \rho\|y\| - \dfrac{\eta}{2}, & \text{if } \|y\| > \eta \text{ and } \eta > 0;\\ \dfrac{\|y\|^2}{2\eta}, & \text{if } \|y\| \leqslant \eta \text{ and } \eta > 0;\\ \|y\|, & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0. \end{cases}$$

  Then $g = [\varphi]^{\sim}$, where $\varphi = \text{env}\,\|\cdot\| = \|\cdot\|^2/2 - d^2_{B(0;1)}/2$ is the generalized Huber function.

- In computer vision, g is called the bivariate Huber function. It also shows up in Owen's concomitant M-estimator formulation.

Perspective functions: Examples

- Let C and D be nonempty closed convex subsets of $\mathcal{G}$, and let $\rho \in \,]0, +\infty[$.
- Set

$$g\colon (\eta, y) \mapsto \begin{cases} \dfrac{\eta d_C^2(y/\eta)}{2\rho} + \sigma_D(y), & \text{if } \eta > 0 \text{ and } y \notin \eta C; \\ \sigma_D(y), & \text{if } \eta > 0 \text{ and } y \in \eta C; \\ \sigma_D(y), & \text{if } \eta = 0 \text{ and } y \in \operatorname{rec} C; \\ +\infty, & \text{otherwise} \end{cases}$$

- Then $g = \widetilde{\varphi} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$, where $\varphi = d_C^2/(2\rho) + \sigma_D$.
- A special case of g appears in computer vision.
- If $\mathcal{G} = \mathbb{R}$ and $D = [-1, 1]$, φ is the Berhu (reversed Huber) function used in mechanics and in Owen's concomitant M-estimator formulation

Perspective functions: Examples

- Let $\psi\colon \mathcal{G} \to [0,+\infty]$ be a proper lower semicontinuous positively homogeneous convex function, let $\delta \in \mathbb{R}$, let $\rho \in [0,+\infty[$, let $p \in [1,+\infty[$, and set

$$g\colon (\eta, y) \mapsto \begin{cases} \delta\eta + \left|\rho\eta^p + \psi^p(y)\right|^{1/p}, & \text{if } \eta \geqslant 0;\\ +\infty, & \text{if } \eta < 0. \end{cases}$$

  Then $g = [\delta + |\rho + \psi^p|^{1/p}]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

- Let $\phi \in \Gamma_0(\mathbb{R})$ be an even function, let $v \in \mathcal{G}$, let $\delta \in \mathbb{R}$, and set

$$g\colon (\eta, y) \mapsto \begin{cases} \eta\phi(\|y\|/\eta) + \langle y \mid v\rangle + \delta\eta, & \text{if } \eta > 0;\\ (\operatorname{rec}\phi)(\|y\|) + \langle y \mid v\rangle, & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0. \end{cases}$$

  Then $g = [\phi \circ \|\cdot\| + \langle \cdot \mid v\rangle + \delta]^{\sim} \in \Gamma_0(\mathbb{R} \oplus \mathcal{G})$.

Perspective functions: Examples

- The divergences between $x > 0$ and $y > 0$ discussed earlier are of the form

$$\int_{\mathbb{R}^N} \widetilde{\varphi}(y(t), x(t))\, dt,$$

where

  - pth order Hellinger: $\varphi(\xi) = \begin{cases} \left|t^{1/p} - 1\right|^p, & \text{if } t > 0; \\ +\infty, & \text{otherwise} \end{cases}$
  - Kullback-Leibler: $\varphi(\xi) = \begin{cases} \xi \ln \xi, & \text{if } \xi > 0; \\ +\infty, & \text{otherwise} \end{cases}$
  - Rényi: $\varphi(\xi) = \begin{cases} \xi^\alpha, & \text{if } \xi > 0; \\ +\infty, & \text{otherwise} \end{cases}$
  - Pearson: $\varphi(\xi) = |\xi - 1|^2$

Composite perspective functions

- Let $L\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $\varphi \in \Gamma_0(\mathcal{G})$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and set

$$f\colon x \mapsto \begin{cases} (\langle x \mid u\rangle - \rho)\varphi\left(\dfrac{Lx - r}{\langle x \mid u\rangle - \rho}\right), & \text{if } \langle x \mid u\rangle > \rho;\\ (\operatorname{rec}\varphi)(Lx - r), & \text{if } \langle x \mid u\rangle = \rho;\\ +\infty, & \text{if } \langle x \mid u\rangle < \rho. \end{cases}$$

Suppose that there exists $z \in \mathcal{H}$ such that

$$Lz \in r + (\langle z \mid u\rangle - \rho)\operatorname{dom}\varphi \quad \text{and} \quad \langle z \mid u\rangle \geqslant 0,$$

and set $A\colon \mathcal{H} \to \mathbb{R} \oplus \mathcal{G}\colon x \mapsto (\langle x \mid u\rangle - \rho, Lx - r)$. Then $f = \widetilde{\varphi} \circ A \in \Gamma_0(\mathcal{H})$.

Composite perspective functions: Examples

- Let $L\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $|||\cdot|||$ be a norm on $\mathcal{G}$ such that, for some $\chi \in \,]0,+\infty[$, $|||\cdot||| \geqslant \chi\|\cdot\|$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and let q and s be in $]1,+\infty[$. Set

$$h\colon x \mapsto \begin{cases} \dfrac{|||Lx - r|||^{qs}}{|\langle x \mid u\rangle - \rho|^{(q-1)s}}, & \text{if } \langle x \mid u\rangle > \rho;\\ 0, & \text{if } Lx = r \text{ and } \langle x \mid u\rangle = \rho;\\ +\infty, & \text{otherwise.} \end{cases}$$

Then $h \in \Gamma_0(\mathcal{H})$.

Composite perspective functions: Examples

- Let $L\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded, let $|||\cdot|||$ be a norm on $\mathcal{G}$ such that, for some $\chi \in \,]0,+\infty[$, $|||\cdot||| \geqslant \chi\|\cdot\|$, let $r \in \mathcal{G}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, and let q and s be in $]1,+\infty[$. Set

$$h\colon x \mapsto \begin{cases} \dfrac{|||Lx-r|||^{qs}}{|\langle x \mid u\rangle - \rho|^{(q-1)s}}, & \text{if } \langle x \mid u\rangle > \rho;\\ 0, & \text{if } Lx = r \text{ and } \langle x \mid u\rangle = \rho;\\ +\infty, & \text{otherwise.} \end{cases}$$

Then $h \in \Gamma_0(\mathcal{H})$.

- Let $(\Omega,\mathcal{F},\mathsf{P})$ be a probability space, let $\mathcal{H} = L^2(\Omega,\mathcal{F},\mathsf{P})$, let $p \in \,]1,2]$, and let q and s be in $]1,+\infty[$. Set

$$h\colon X \mapsto \begin{cases} \dfrac{\mathsf{E}^{qs/p}|X|^p}{\mathsf{E}^{(q-1)s}X}, & \text{if } \mathsf{E}X > 0;\\ 0, & \text{if } X = 0 \text{ a.s.};\\ +\infty, & \text{otherwise.} \end{cases}$$

Then $h \in \Gamma_0(\mathcal{H})$.

Composite perspective functions: Examples

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let G be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{G})$. Set $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathbb{R})$ and $\mathcal{G} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{G})$, and suppose that $\mu(\Omega) < +\infty$ or $\varphi \geqslant \varphi(0) = 0$. For every $x \in \mathcal{H}$, set $\Omega_0(x) = \{\omega \in \Omega \mid x(\omega) = 0\}$ and $\Omega_+(x) = \{\omega \in \Omega \mid x(\omega) > 0\}$. Define

$$\Phi \colon \mathcal{H} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] \colon (x, y) \mapsto$$
$$\begin{cases} \displaystyle\int_{\Omega_0(x)} (\operatorname{rec} \varphi)\big(y(\omega)\big)\mu(d\omega) + \int_{\Omega_+(x)} x(\omega)\varphi\left(\frac{y(\omega)}{x(\omega)}\right)\mu(d\omega), \\ \qquad\qquad \text{if } \begin{cases} x \geqslant 0 \text{ a.e.} \\ (\operatorname{rec} \varphi)(y)1_{\Omega_0(x)} + x\varphi(y/x)1_{\Omega_+(x)} \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \end{cases} \\ +\infty, \qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Then $\Phi \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$.

Composite perspective functions: Examples

Corollary: Let Ω be a nonempty open subset of $\mathbb{R}^N$ and let $\mathcal{H}$ be the Sobolev space $H^1(\Omega)$, i.e., $\mathcal{H} = \{x \in L^2(\Omega) \mid \nabla x \in (L^2(\Omega))^N\}$. For every $x \in \mathcal{H}$, set $\Omega_-(x) = \{t \in \Omega \mid x(t) < 0\}$, $\Omega_0(x) = \{t \in \Omega \mid x(t) = 0\}$, and $\Omega_+(x) = \{t \in \Omega \mid x(t) > 0\}$. Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ be such that $\varphi \geqslant \varphi(0) = 0$, and define

$$
\begin{aligned}
f \colon \mathcal{H} &\to \left]-\infty, +\infty\right] \\
x &\mapsto \begin{cases} \displaystyle\int_{\Omega_0(x)} (\operatorname{rec}\varphi)\big(\nabla x(t)\big)\,dt + \int_{\Omega_+(x)} x(t)\varphi\left(\frac{\nabla x(t)}{x(t)}\right) dt, & \text{if } x \geqslant 0 \\ +\infty, & \text{else} \end{cases}
\end{aligned}
$$

Then $f \in \Gamma_0(\mathcal{H})$.

Composite perspective functions: Examples

- The Fisher information

$$f\colon H^1(\Omega) \to \left]-\infty,+\infty\right]$$
$$x \mapsto \begin{cases} \displaystyle\int_{\Omega_+(x)} \frac{\|\nabla x(t)\|_2^2}{x(t)}dt, & \text{if } \begin{cases} x \geqslant 0 \text{ a.e.} \\ [x=0 \Rightarrow \nabla x = 0] \text{ a.e.;} \end{cases} \\ +\infty, & \text{otherwise} \end{cases}$$

is in $\Gamma_0(H^1(\Omega))$.

- For $(x,y) \in \mathbb{R}^{2N}$, set $I_0(x,y) = \{i \in I \mid \xi_i = 0 \text{ and } \eta_i < 0\}$ and

$$d_\phi(x,y) = \begin{cases} \displaystyle\sum_{i \in I_0(x)} \eta_i + \sum_{i \in I_+(x)} \left|\eta_i^{1/p} - \xi_i^{1/p}\right|^p, & \text{if } I_-(x) \cup I_0(x,y) = \varnothing; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $d_\phi \in \Gamma_0(\mathbb{R}^{2N})$. We recover the Kolmogorov variational divergence for $p = 1$ and the Hellinger divergence for $p = 2$.

Perspective functions: Proximity operator

- The Moreau proximity operator of $g \in \Gamma_0(\mathcal{G})$ is

$$\operatorname{prox}_g \colon \mathcal{G} \to \mathcal{G} \colon x \mapsto \underset{y \in \mathcal{G}}{\operatorname{argmin}} \left(g(y) + \frac{1}{2}\|x - y\|^2 \right).$$

- It is an essential tool in the design of splitting algorithms to solve a variety of convex minimization problems, especially in data science over the past dodecade
  - PLC and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, 2005

Proximity operators

- Many common convex functions in data processing (statistics, machine learning, image recovery, data denoising, support vector machine, signal processing) have explicit proximity operators:
  - ℓ_1 norm
  - Shatten norm
  - nuclear norm
  - Huber's function
  - Berhu function
  - elastic net regularizer
  - hinge loss
  - Fisher information
  - distance function
  - Vapnik's ε-insensitive loss
  - Burg's entropy
  - etc.

Proximity operators

- Basic properties:
  - $p = \operatorname{prox}_f x \Leftrightarrow x - p \in \partial f(p)$
  - $\operatorname{prox}_f + \operatorname{prox}_{f^*} = \operatorname{Id}$ (Moreau's decomposition)
    - For $f = \iota_V$, V a closed vector subspace: $P_V + P_{V^\perp} = \operatorname{Id}$
    - $\operatorname{prox}_{\rho|\cdot|} = \operatorname{Id} - \operatorname{prox}_{(\rho|\cdot|)^*} = \operatorname{Id} - P_{[-\rho,\rho]} = \operatorname{soft}_\rho$
  - $(\operatorname{prox}_f x, x - \operatorname{prox}_f x) = (\operatorname{prox}_f x, \operatorname{prox}_{f^*} x) \in \operatorname{gra} \partial f$
  - $\operatorname{Fix} \operatorname{prox}_f = \operatorname{Argmin} f$
  - $\|\operatorname{prox}_f x - \operatorname{prox}_f y\| \leqslant \|x - y\|$

Proximity operators

- Basic properties:
  - $p = \mathrm{prox}_f x \Leftrightarrow x - p \in \partial f(p)$
  - $\mathrm{prox}_f + \mathrm{prox}_{f^*} = \mathrm{Id}$ (Moreau's decomposition)
    - For $f = \iota_V$, V a closed vector subspace: $P_V + P_{V^\perp} = \mathrm{Id}$
    - $\mathrm{prox}_{\rho|\cdot|} = \mathrm{Id} - \mathrm{prox}_{(\rho|\cdot|)^*} = \mathrm{Id} - P_{[-\rho,\rho]} = \mathrm{soft}_\rho$
  - $(\mathrm{prox}_f x, x - \mathrm{prox}_f x) = (\mathrm{prox}_f x, \mathrm{prox}_{f^*} x) \in \mathrm{gra}\,\partial f$
  - $\mathrm{Fix}\ \mathrm{prox}_f = \mathrm{Argmin}\, f$
  - $\|\mathrm{prox}_f x - \mathrm{prox}_f y\|^2 \leqslant \|x - y\|^2 - \|\mathrm{prox}_{f^*} x - \mathrm{prox}_{f^*} y\|^2$
- The last two properties suggest the conceptual algorithm

$$x_{n+1} = \mathrm{prox}_f x_n$$

to minimize f, which is at the root of proximal splitting algorithms.

Proximal splitting methods in convex optimization

- $f \in \Gamma_0(\mathcal{H})$, $\varphi_k \in \Gamma_0(\mathcal{G}_k)$, $\ell_k \in \Gamma_0(\mathcal{G}_k)$ strongly convex, $L_k \colon \mathcal{H} \to \mathcal{G}_k$ linear bounded, $\|L_k\| = 1$, $h \colon \mathcal{H} \to \mathbb{R}$ convex and smooth:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} (\varphi_k \,\square\, \ell_k)(L_k x - r_k) + h(x)$$

  where: $\varphi_k \,\square\, \ell_k \colon x \mapsto \inf_{y \in \mathcal{H}} \big(\varphi_k(y) + \ell_k(x - y)\big)$

- Example: multiview total variation image recovery from observations $r_k = L_k \overline{x} + w_k$:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \sum_{k \in \mathbb{N}} \phi_k(\langle x \mid e_k \rangle) + \sum_{k=1}^{p-1} \alpha_k \underbrace{d_{C_k}}_{\iota_C \,\square\, \|\cdot\|} (L_k x - r_k) + \beta \|\nabla x\|_{1,2}$$

- A splitting algorithm activates each function and each linear operator individually

Proximal splitting methods in convex optimization

- Algorithm:

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \left(\nabla h(x_n) + \sum_{k=1}^{m} L_k^* v_{k,n}\right) \\
p_{1,n} = \operatorname{prox}_f y_{1,n} \\
\text{For } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{2,k,n} = v_{k,n} + (L_k x_n - \nabla \ell_k^*(v_{k,n})) \\
p_{2,k,n} = \operatorname{prox}_{g_k^*}(y_{2,k,n} - r_k) \\
q_{2,k,n} = p_{2,k,n} + (L_k p_{1,n} - \nabla \ell_k^*(p_{2,k,n})) \\
v_{k,n+1} = v_{k,n} - y_{2,k,n} + q_{2,k,n}
\end{array}
\right. \\
q_{1,n} = p_{1,n} - \left(\nabla h(p_{1,n}) + \sum_{k=1}^{m} L_k^* p_{2,k,n}\right) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}
\end{array}
\right.
\end{array}
$$

- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution
  - PLC, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, 2013

Perspective functions: Proximity operator

Let $\varphi \in \Gamma_0(\mathcal{G})$, let $\gamma \in \,]0,+\infty[$, let $\eta \in \mathbb{R}$, and let $y \in \mathcal{G}$.

- Suppose that $\eta + \gamma\varphi^*(y/\gamma) \leqslant 0$. Then $\operatorname{prox}_{\gamma\widetilde{\varphi}}(\eta, y) = (0, 0)$.
- Suppose that $\operatorname{dom}\varphi^*$ is open and that $\eta + \gamma\varphi^*(y/\gamma) > 0$. Then
$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(\eta, y) = \big(\eta + \gamma\varphi^*(p), y - \gamma p\big),$$
where p is the unique solution to the inclusion
$$y \in \gamma p + \big(\eta + \gamma\varphi^*(p)\big)\partial\varphi^*(p).$$
If φ^* is differentiable at p, then p is characterized by
$$y = \gamma p + (\eta + \gamma\varphi^*(p))\nabla\varphi^*(p).$$

Perspective functions: Proximity operator

Let $v \in \mathcal{G}$, let $\delta \in \mathbb{R}$, and let $\phi \in \Gamma_0(\mathbb{R})$ be an even function such that ϕ^* is differentiable on $\mathbb{R}$. Define

$$g\colon (\eta, y) \mapsto \begin{cases} \eta\phi(\|y\|/\eta) + \delta\eta + \langle y \mid v\rangle, & \text{if } \ \eta > 0;\\ 0, & \text{if } \ y = 0 \ \text{ and } \ \eta = 0;\\ +\infty, & \text{otherwise.} \end{cases}$$

Let $\gamma \in \left]0, +\infty\right[$, let $\eta \in \mathbb{R}$, let $y \in \mathcal{G}$, and set

$$\psi\colon s \mapsto \left(\phi^*(s) + \frac{\eta}{\gamma} - \delta\right)\phi^{*\prime}(s) + s.$$

Then ψ is invertible. Moreover, if $\eta + \gamma\phi^*(\|y/\gamma - v\|) > \gamma\delta$, set

$$t = \psi^{-1}(\|y/\gamma - v\|) \quad \text{and} \quad p = v + \frac{t}{\|y - \gamma v\|}(y - \gamma v).$$

Then

$$\operatorname{prox}_{\gamma g}(\eta, y) = \begin{cases} (\eta + \gamma(\phi^*(t) - \delta), y - \gamma p), & \text{if } \ \eta + \gamma\phi^*(\|y/\gamma - v\|) > \gamma\delta;\\ (0, 0), & \text{if } \ \eta + \gamma\phi^*(\|y/\gamma - v\|) \leqslant \gamma\delta. \end{cases}$$

Perspective functions: Proximity operator

Let $v \in \mathcal{G}$, let $\delta \in \mathbb{R}$, let $\alpha \in \,]0,+\infty[$, let $q \in \,]1,+\infty[$, and consider the function

$$g\colon (\eta, y) \mapsto \begin{cases} \dfrac{\|y\|^q}{\alpha\eta^{q-1}} + \delta\eta + \langle y \mid v\rangle, & \text{if } \eta > 0;\\ 0, & \text{if } y = 0 \text{ and } \eta = 0;\\ +\infty, & \text{otherwise.} \end{cases}$$

Let $\gamma \in \,]0,+\infty[$, set $q^* = q/(q-1)$, set $\varrho = (\alpha(1-1/q^*))^{q^*-1}$, and take $\eta \in \mathbb{R}$ and $y \in \mathcal{G}$. If $q^*\gamma^{q^*-1}\eta + \varrho\|y\|^{q^*} > \gamma\delta$, let $t \in [0,+\infty[$ be the unique solution to the equation

$$t^{2q^*-1} + \frac{q^*(\eta-\gamma\delta)}{\gamma\varrho}t^{q^*-1} + \frac{q^*}{\varrho^2}t - \frac{q^*\|y-\gamma v\|}{\gamma\varrho^2} = 0$$

and set $p = v + t(y-\gamma v)/\|y-\gamma v\|$. Then $\operatorname{prox}_{\gamma g}(\eta, y) =$

$$\begin{cases} \left(\eta + \gamma(\varrho t^{q^*} - \delta)/q^*, y - \gamma p\right), & \text{if } q^*\gamma^{q^*-1}\eta + \varrho\|y\|^{q^*} > \gamma\delta;\\ (0,0), & \text{if } q^*\gamma^{q^*-1}\eta + \varrho\|y\|^{q^*} \leqslant \gamma\delta. \end{cases}$$

Perspective functions: Proximity operator

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let G be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{G})$. Set $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathbb{R})$ and $\mathcal{G} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{G})$, and suppose that $\mu(\Omega) < +\infty$ or $\varphi \geqslant \varphi(0) = 0$. For every $x \in \mathcal{H}$, set $\Omega_0(x) = \{\omega \in \Omega \mid x(\omega) = 0\}$ and $\Omega_+(x) = \{\omega \in \Omega \mid x(\omega) > 0\}$. Define

$$\Phi \colon \mathcal{H} \oplus \mathcal{G} \to \,]-\infty, +\infty] \colon (x, y) \mapsto$$
$$\begin{cases} \displaystyle\int_{\Omega_0(x)} (\operatorname{rec} \varphi)\big(y(\omega)\big)\mu(d\omega) + \int_{\Omega_+(x)} x(\omega)\varphi\left(\frac{y(\omega)}{x(\omega)}\right)\mu(d\omega), \\ \qquad\qquad \text{if } \begin{cases} x \geqslant 0 \text{ a.e.} \\ (\operatorname{rec} \varphi)(y)1_{\Omega_0(x)} + x\varphi(y/x)1_{\Omega_+(x)} \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \end{cases} \\ +\infty, \quad \text{otherwise.} \end{cases}$$

Now let $x \in \mathcal{H}$ and $y \in \mathcal{G}$, and set, for μ-almost every $\omega \in \Omega$, $(p(\omega), q(\omega)) = \operatorname{prox}_{\widetilde{\varphi}}(x(\omega), y(\omega))$. Then $\operatorname{prox}_\Phi(x, y) = (p, q)$.

Perspective functions: Proximity operator

- We can also handle cases when $\operatorname{dom}\varphi^*$ is not open.
- Consider the perspective function

$$\widetilde{\varphi}\colon \mathbb{R}^2 \to \,]-\infty,+\infty]\colon (\eta,y)\mapsto \begin{cases} d_{[-\varepsilon\eta,\varepsilon\eta]}(y), & \text{if } \eta \geqslant 0;\\ +\infty, & \text{if } \eta < 0\end{cases}$$

  of the Vapnik loss function $\varphi = \max\{|\cdot| - \varepsilon, 0\}$.
- $\varphi^* = \varepsilon|\cdot| + \iota_{[-1,1]}$.
- Let $\eta \in \mathbb{R}$, let $y \in \mathbb{R}$, and set $(\chi, q) = \operatorname{prox}_{\gamma\widetilde{\varphi}}(\eta, y)$. Then
  - If $\eta + \varepsilon|y| \leqslant 0$ and $|y| \leqslant \gamma$, $(\chi, q) = (0, 0)$.
  - If $\eta \leqslant -\gamma\varepsilon$ and $|y| > \gamma$, $(\chi, q) = (0, y - \gamma\,\operatorname{sign}(y))$.
  - If $\eta > -\gamma\varepsilon$ and $|y| > \varepsilon\eta + \gamma(1+\varepsilon^2)$, $(\chi, q) = (\eta + \gamma\varepsilon, y - \gamma\,\operatorname{sign}(y))$.
  - If $|y| > -\eta/\varepsilon$ and $\varepsilon\eta \leqslant |y| \leqslant \varepsilon\eta + \gamma(1+\varepsilon^2)$, $(\chi, q) = ((\eta + \varepsilon|y|)/(1+\varepsilon^2), \varepsilon(\eta + \varepsilon|y|)\operatorname{sign}(y)/(1+\varepsilon^2))$.
  - If $\eta \geqslant 0$ and $|y| \leqslant \varepsilon\eta$, $(\chi, q) = (\eta, y)$.

Applications in high-dimensional statistics

- Linear data model: $z = Xb + \sigma e$
- Penalized concomitant M-estimators:

$$\underset{\sigma\in\mathbb{R},\,\tau\in\mathbb{R},\,b\in\mathbb{R}^p}{\text{minimize}} \quad \sum_{i=1}^{n} \widetilde{\varphi}_i(\sigma, X_{i:}b - \zeta_i) + \sum_{j=1}^{p} \widetilde{\psi}_j(\tau, a_j^\top b).$$

- This model unifies various robust regression procedures
- Can be solved efficiently by the block-iterative proximal splitting method of
  - PLC and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Mathematical Programming,* published online 2016-07-05
- Other model of interest: generalized TREX

$$\underset{b\in\mathbb{R}^p}{\text{minimize}} \quad \frac{\|Xb - z\|_2^q}{\alpha\|X^\top(Xb - z)\|_\infty^{q-1}} + \|b\|_1$$

Applications in high-dimensional statistics

- The nonconvex generalized TREX problem can be rewritten as a system of $2p$ convex problems

$$\underset{\substack{b\in\mathbb{R}^p\\ x_j^\top(Xb-z)>0}}{\text{minimize}}\ \frac{\|Xb-z\|_2^q}{\alpha|x_j^\top(Xb-z)|^{q-1}}+\|b\|_1,\ \text{where } x_j=sX_{:j},\ s\in\{-1,1\}.$$

- Each subproblem involves the (shifted) perspective function

$$g_j\colon(\eta,y)\mapsto\begin{cases}\dfrac{\|y-z\|_2^2}{\alpha\big(\eta-x_j^\top z\big)}, & \text{if }\ \eta>x_j^\top z;\\ 0, & \text{if }\ y=z \text{ and } \eta=x_j^\top z;\\ +\infty, & \text{otherwise}\end{cases}$$

of $\|\cdot\|_2^q$ composed with the linear operator $b\mapsto\big(x_j^\top Xb, Xb\big)$, and $h=[\|\cdot\|_1]^\sim=\|\cdot\|_1$.

- It can be solved (for instance), by a Douglas-Rachford-like algorithm.

Applications in high-dimensional statistics

- prox_h is the standard soft thresholding operator
- We have $\mathrm{prox}_{\gamma g_j}(\eta, y)$

$$= \begin{cases} (\eta + \gamma\varrho t^{q^*}/q^*, y - \gamma p), & \text{if } q^*\gamma^{q^*-1}(\eta - x_j^\top z) + \varrho\|y - z\|_2^{q^*} > 0; \\ (x_j^\top z, z), & \text{if } q^*\gamma^{q^*-1}(\eta - x_j^\top z) + \varrho\|y - z\|_2^{q^*} \leqslant 0, \end{cases}$$

where $\varrho = (\alpha(1 - 1/q^*))^{q^*-1}$,

$$p = \begin{cases} \dfrac{t}{\|y - z\|}(y - z), & \text{if } y \neq z; \\ 0, & \text{if } y = z, \end{cases}$$

and t is the unique solution in $]0, +\infty[$ to the reduced equation

$$s^{2q^*-1} + \frac{q^*(\eta - x_j^\top z)}{\gamma\varrho} s^{q^*-1} + \frac{q^*}{\varrho^2} s - \frac{q^*\|y - z\|}{\gamma\varrho^2} = 0.$$

Applications in high-dimensional statistics

- Algorithm for the jth generalized TREX subproblem

$$\left\lfloor \begin{array}{l} q_k = M_j x_k - y_k \\ b_k = x_k - R_j q_k \\ c_k = M_j b_k \\ z_k = \operatorname{prox}_{\gamma h}(2b_k - x_k) \\ t_k = \operatorname{prox}_{\gamma g_{j,q}}(2c_k - y_k) \\ x_{k+1} = x_k + \mu_k(z_k - b_k) \\ y_{k+1} = y_k + \mu_k(t_k - c_k). \end{array} \right.$$

- $(b_k)_{k\in\mathbb{N}}$ converges to a solution b to the subproblem.
- See paper for detailed numerical application to sparse regression.

References

- PLC and C. L. Müller, Perspective functions: Proximal calculus and applications in high-dimensional statistics, *J. Math. Anal. Appl.*, 2017 (published online)
- PLC, Perspective functions: Properties, constructions, and examples, 2016. https://arxiv.org/abs/1610.01552
- H. H. Bauschke and PLC, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011. (second edition: February 2017)