Perspective Functions and Applications

Patrick L. Combettes

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

McGill University – Applied Mathematics Seminar
February 15, 2021

Optimization in data analysis

- Fitting geodesic data by the method of least deviations:
  - R. J. Boscovich, De literaria expeditione per pontificiam ditionem et synopsis... *Bononiensi Scientiarum et Artum Inst. atque Acad. Comment.*, 1757.
- Fitting astronomical data by the method of least squares:
  - A. M. Legendre, *Nouvelles Méthodes pour la Détermination de l'Orbite des Comètes.* Courcier, Paris, 1805.
- The gradient method was invented to deal with for astronomical data processing:
  - A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simmultanées, *C. R. Acad. Sci. Paris*, 1847.

Some formulations arising in data analysis

- 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 formulations arising in data analysis

- 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 formulations arising in data analysis

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

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

- A strengthened form of the central limit theorem (Lions and Toscani, 1995) involving functionals of $x > 0$ of the form

$$\int_{\mathbb{R}} |x(t)|^{(1-p)} |x^{(k)}(t)|^p dt, \quad \text{with} \quad p > 1$$

Some formulations arising in data analysis

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

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

- A strengthened form of the central limit theorem (Lions and Toscani, 1995) involving functionals of $x > 0$ of the form

$$\int_{\mathbb{R}} \big|x(t)\big|^{(1-p)} \big|x^{(k)}(t)\big|^p dt, \quad \text{with} \quad p > 1$$

- Rey's “fair” function (1983) in robust statistics:

$$f(y,\eta) = \begin{cases} |y| + \eta \ln \eta - \eta \ln\big(\eta + |y|\big), & \text{if } \eta > 0; \\ |y|, & \text{if } \eta = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Some formulations arising in data analysis

- Minimization roblems 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 formulations arising in data analysis

- Minimization roblems 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: $\int_{\mathbb{R}^N} x(t)^{\alpha} y(t)^{1-\alpha} dt$
  - Pearson: $\int_{\mathbb{R}^N} \frac{|x(t) - y(t)|^2}{y(t)} dt$
- The heteroscedastic M-estimation model (PLC & Müller, 2020)

$$\underset{s\in\mathbb{R}^N,\, t\in\mathbb{R}^P,\, b\in\mathbb{R}^p}{\text{minimize}} \quad \varsigma(s)+\varpi(t)+\theta(b)+\sum_{i=1}^{N}\sigma_i\varphi_i\left(\frac{X_i b - y_i}{\sigma_i}\right)+\sum_{i=1}^{P}\tau_i\psi_i\left(\frac{L_i b}{\tau_i}\right)$$

Some formulations arising in data analysis

- Minimization roblems 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: $\displaystyle\int_{\mathbb{R}^N} \frac{|x(t) - y(t)|^2}{y(t)} dt$
- The heteroscedastic M-estimation model (PLC & Müller, 2020)

$$\underset{s\in\mathbb{R}^N,\, t\in\mathbb{R}^P,\, b\in\mathbb{R}^p}{\text{minimize}} \;\; \varsigma(s)+\varpi(t)+\theta(b)+\sum_{i=1}^{N}\sigma_i\varphi_i\left(\frac{X_i b - y_i}{\sigma_i}\right)+\sum_{i=1}^{P}\tau_i\psi_i\left(\frac{L_i b}{\tau_i}\right)$$

- ***What is the common structure underlying these formulations?***

Some formulations arising in data analysis

- Minimization roblems involving various notions of divergence between $x > 0$ and $y > 0$:
  - pth order Hellinger: $\displaystyle\int_{\mathbb{R}^N} \big|x(t)^{1/p} - y(t)^{1/p}\big|^p dt$
  - 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: $\displaystyle\int_{\mathbb{R}^N} \frac{|x(t) - y(t)|^2}{y(t)} dt$
- The heteroscedastic M-estimation model (PLC & Müller, 2020)

$$\underset{s\in\mathbb{R}^N,\, t\in\mathbb{R}^P,\, b\in\mathbb{R}^p}{\text{minimize}} \quad \varsigma(s)+\varpi(t)+\theta(b)+\sum_{i=1}^{N}\sigma_i\varphi_i\left(\frac{X_i b - y_i}{\sigma_i}\right)+\sum_{i=1}^{P}\tau_i\psi_i\left(\frac{L_i b}{\tau_i}\right)$$

- ***What is the common structure underlying these formulations?***
- ***Is there some underlying convexity?***

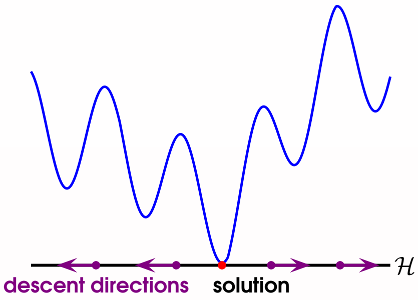
A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

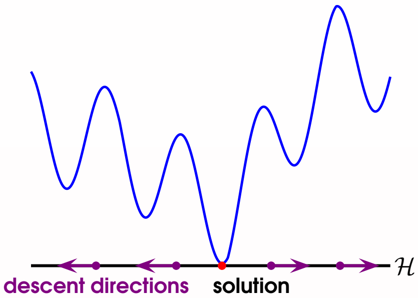
2. Moving permanently *away* from solutions in descent methods:



A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

2. Moving permanently *away* from solutions in descent methods:

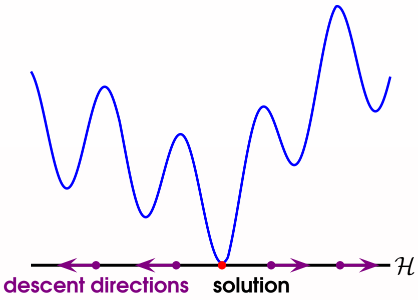


3. Loose connections with other branches of nonlinear analysis

A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

2. Moving permanently *away* from solutions in descent methods:



3. Loose connections with other branches of nonlinear analysis

4. Algorithms may yield trivial solutions:

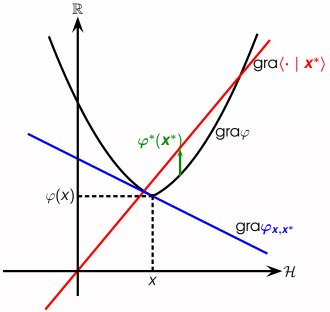
Let $f\colon \mathcal{H} \to \{0,\ldots,p\}$ be l.s.c. (e.g., rank etc.), let $C \neq \varnothing$. Then ***any*** point in C is a local minimizer of:

$$\underset{x\in C}{\text{minimize}}\ \ f(x)$$

J.-B. Hiriart-Urruty, When only global optimization matters, *J. Global Optim.*, vol. 56, pp. 761–763, 2013

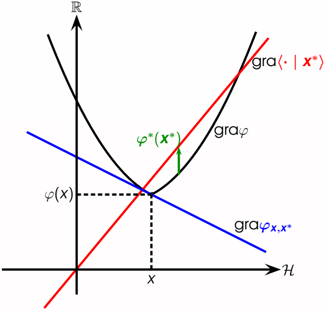
A few facts from convex analysis

- $\mathcal{H}$: a real Hilbert space
- $\varphi \in \Gamma_0(\mathcal{H})$: $\varphi\colon \mathcal{H} \to \,]-\infty,+\infty]$ is lower semicontinuous, convex, and $\operatorname{dom}\varphi = \{x \in \mathcal{H} \mid \varphi(x) < +\infty\} \neq \varnothing$



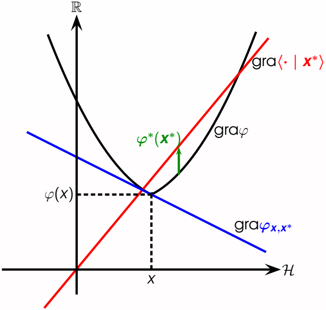
A few facts from convex analysis

- $\mathcal{H}$: a real Hilbert space
- $\varphi \in \Gamma_0(\mathcal{H})$: $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is lower semicontinuous, convex, and $\operatorname{dom}\varphi = \{x \in \mathcal{H} \mid \varphi(x) < +\infty\} \neq \varnothing$
- If $\varphi(x) \to +\infty$ as $\|x\| \to +\infty$, then $\operatorname{Argmin}\varphi \neq \varnothing$



A few facts from convex analysis

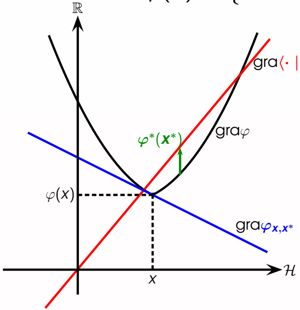
- $\mathcal{H}$: a real Hilbert space
- $\varphi \in \Gamma_0(\mathcal{H})$: $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is lower semicontinuous, convex, and $\operatorname{dom}\varphi = \big\{x \in \mathcal{H} \mid \varphi(x) < +\infty\big\} \neq \varnothing$
- If $\varphi(x) \to +\infty$ as $\|x\| \to +\infty$, then $\operatorname{Argmin}\varphi \neq \varnothing$
- $\varphi^*\colon x^* \mapsto \sup_{x\in\mathcal{H}} \langle x \mid x^*\rangle - \varphi(x)$ is the Legendre conjugate of φ; if $\varphi \in \Gamma_0(\mathcal{H})$, then $\varphi^* \in \Gamma_0(\mathcal{H})$ and $\varphi^{**} = \varphi$



A few facts from convex analysis

- $\mathcal{H}$: a real Hilbert space
- $\varphi \in \Gamma_0(\mathcal{H})$: $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is lower semicontinuous, convex, and $\operatorname{dom}\varphi = \{x \in \mathcal{H} \mid \varphi(x) < +\infty\} \neq \varnothing$
- If $\varphi(x) \to +\infty$ as $\|x\| \to +\infty$, then $\operatorname{Argmin}\varphi \neq \varnothing$
- $\varphi^*\colon x^* \mapsto \sup_{x\in\mathcal{H}} \langle x \mid x^*\rangle - \varphi(x)$ is the Legendre conjugate of φ; if $\varphi \in \Gamma_0(\mathcal{H})$, then $\varphi^* \in \Gamma_0(\mathcal{H})$ and $\varphi^{**} = \varphi$
- The subdifferential of φ at $x \in \mathcal{H}$ is

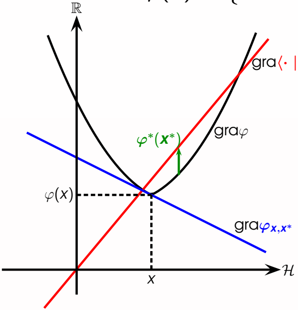
$$\partial\varphi(x) = \big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\;\; \underbrace{\langle y - x \mid x^*\rangle + \varphi(x)}_{\varphi_{x,x^*}(y)} \leqslant \varphi(y)\big\}$$



A few facts from convex analysis

- $\mathcal{H}$: a real Hilbert space
- $\varphi \in \Gamma_0(\mathcal{H})$: $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is lower semicontinuous, convex, and $\operatorname{dom}\varphi = \{x \in \mathcal{H} \mid \varphi(x) < +\infty\} \neq \varnothing$
- If $\varphi(x) \to +\infty$ as $\|x\| \to +\infty$, then $\operatorname{Argmin}\varphi \neq \varnothing$
- $\varphi^*\colon x^* \mapsto \sup_{x\in\mathcal{H}} \langle x \mid x^*\rangle - \varphi(x)$ is the Legendre conjugate of φ; if $\varphi \in \Gamma_0(\mathcal{H})$, then $\varphi^* \in \Gamma_0(\mathcal{H})$ and $\varphi^{**} = \varphi$
- The subdifferential of φ at $x \in \mathcal{H}$ is

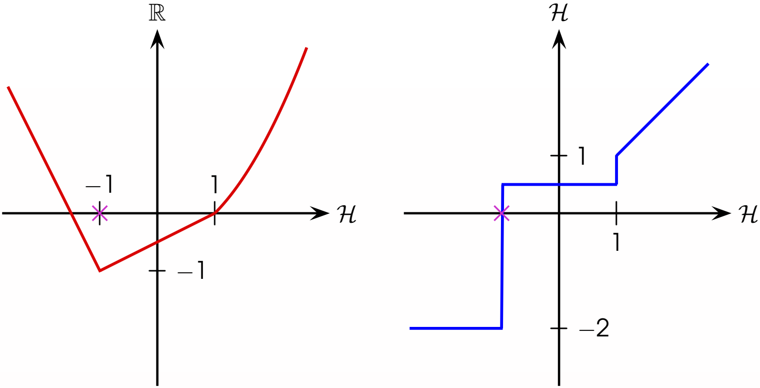
$$\partial\varphi(x) = \Big\{x^* \in \mathcal{H} \;\Big|\; (\forall y \in \mathcal{H})\;\; \underbrace{\langle y - x \mid x^*\rangle + \varphi(x)}_{\varphi_{x,x^*}(y)} \leqslant \varphi(y)\Big\}$$



Fermat's rule:

x minimizes $\varphi \Leftrightarrow 0 \in \partial\varphi(x)$

Convex functions: Subdifferentiability



Left: Graph of a function φ defined on $\mathcal{H} = \mathbb{R}$.
Right: Graph of its subdifferential $\partial\varphi$.

Perspective functions: Definition

- $\mathcal{H}, \mathcal{G}$ real Hilbert spaces
- $\varphi \in \Gamma_0(\mathcal{G})$
- $\operatorname{rec}\varphi$ is the recession function of φ:

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

- *Perspective function* of φ:

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

Perspective functions: Example

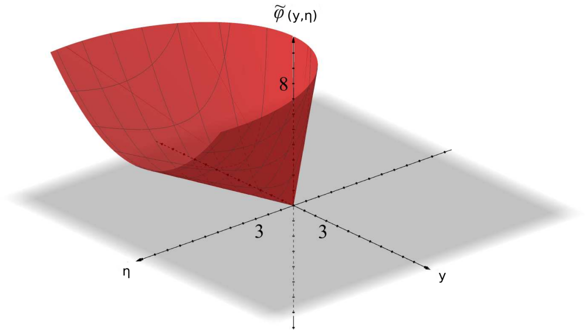


Figure: Perspective of $\varphi = |\cdot|^2 + 1/2$.

Perspective functions: Example

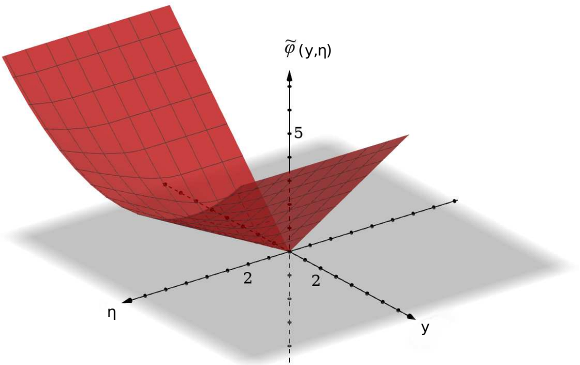


Figure: Perspective of $\varphi = h_1 + 1/2$, where h_1 is the Huber function.

Perspective functions: Properties

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

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

Perspective functions: Properties

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

- $\widetilde{\varphi} \in \Gamma_0(\mathcal{G} \times \mathbb{R})$
- Let $C = \big\{(u, \mu) \in \mathcal{G} \times \mathbb{R} \mid \mu + \varphi^*(u) \leqslant 0\big\}$. Then

$$\widetilde{\varphi} = \sigma_C \quad \text{and} \quad (\widetilde{\varphi})^* = \iota_C$$

Perspective functions: Properties

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

- $\widetilde{\varphi} \in \Gamma_0(\mathcal{G} \times \mathbb{R})$
- Let $C = \big\{(u,\mu) \in \mathcal{G} \times \mathbb{R} \;\big|\; \mu + \varphi^*(u) \leqslant 0\big\}$. Then

$$\widetilde{\varphi} = \sigma_C \quad \text{and} \quad (\widetilde{\varphi})^* = \iota_C$$

- Let $y \in \mathcal{G}$ and $\eta \in \mathbb{R}$. Then $\partial\widetilde{\varphi}(y,\eta) =$

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

Perspective functions: Properties ($\varphi \in \Gamma_0(\mathcal{G})$)

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

Perspective functions: Properties ($\varphi \in \Gamma_0(\mathcal{G})$)

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

Perspective functions: Properties ($\varphi \in \Gamma_0(\mathcal{G})$)

- Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\lambda > 0$. Then $[\lambda\varphi + \psi]^{\sim} = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathcal{G} \times \mathbb{R})$.
- Let $\Lambda\colon \mathcal{H} \to \mathcal{G}$ be linear, bounded, and such that $\operatorname{ran}\Lambda \cap \operatorname{dom}\varphi \neq \varnothing$. Set $\widetilde{\Lambda}\colon \mathcal{H} \times \mathbb{R} \to \mathcal{G} \times \mathbb{R}\colon (x, \xi) \mapsto (\Lambda x, \xi)$. Then $[\varphi \circ \Lambda]^{\sim} = \widetilde{\varphi} \circ \widetilde{\Lambda} \in \Gamma_0(\mathcal{H} \times \mathbb{R})$.
- Suppose that φ is positively homogeneous with $\operatorname{dom}\varphi = \mathcal{G}$, let $\phi \in \Gamma_0(\mathbb{R})$ be increasing on $\operatorname{ran}\varphi$ and such that $0 \in \operatorname{dom}\phi$, let $\eta \in \mathbb{R}$, and let $y \in \mathcal{G}$. Then $\Gamma_0(\mathcal{G} \times \mathbb{R}) \ni [\phi \circ \varphi]^{\sim}\colon (y, \eta) \mapsto \widetilde{\phi}(\varphi(y), \eta)$.

Perspective functions: Properties ($\varphi \in \Gamma_0(\mathcal{G})$)

- Let $\psi \in \Gamma_0(\mathcal{G})$ be such that $\operatorname{dom}\varphi \cap \operatorname{dom}\psi \neq \varnothing$, and let $\lambda > 0$. Then $[\lambda\varphi + \psi]^{\sim} = \lambda\widetilde{\varphi} + \widetilde{\psi} \in \Gamma_0(\mathcal{G} \times \mathbb{R})$.
- Let $\Lambda \colon \mathcal{H} \to \mathcal{G}$ be linear, bounded, and such that $\operatorname{ran}\Lambda \cap \operatorname{dom}\varphi \neq \varnothing$. Set $\widetilde{\Lambda} \colon \mathcal{H} \times \mathbb{R} \to \mathcal{G} \times \mathbb{R} \colon (x, \xi) \mapsto (\Lambda x, \xi)$. Then $[\varphi \circ \Lambda]^{\sim} = \widetilde{\varphi} \circ \widetilde{\Lambda} \in \Gamma_0(\mathcal{H} \times \mathbb{R})$.
- Suppose that φ is positively homogeneous with $\operatorname{dom}\varphi = \mathcal{G}$, let $\phi \in \Gamma_0(\mathbb{R})$ be increasing on $\operatorname{ran}\varphi$ and such that $0 \in \operatorname{dom}\phi$, let $\eta \in \mathbb{R}$, and let $y \in \mathcal{G}$. Then $\Gamma_0(\mathcal{G} \times \mathbb{R}) \ni [\phi \circ \varphi]^{\sim} \colon (y, \eta) \mapsto \widetilde{\phi}(\varphi(y), \eta)$.
- 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

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

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

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

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

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

Perspective functions: Examples

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

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

  Then $g = [\varphi]^{\sim}$, where $\varphi = \operatorname{env}\|\cdot\| = \|\cdot\|^2/2 - d^2_{B(0;1)}/2$ 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 $\rho \in \left]0,+\infty\right[$.
- Set

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

- Then $g = \widetilde{\varphi} \in \Gamma_0(\mathcal{G} \times \mathbb{R})$, where $\varphi = d_C^2/(2\rho) + \sigma_D$
- A special case of g appears in computer vision
- If $\mathcal{G} = \mathbb{R}$, $C = [-\rho, \rho]$, and $D = [-1, 1]$, φ is the Berhu (reverse 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 $\rho > 0$, let $p \in [1,+\infty[$, and set

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

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

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 $\rho > 0$, let $p \in [1,+\infty[$, and set

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

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

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

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

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

Perspective functions: Examples

- Let $\rho \in\ ]0,+\infty[$, let $p \in [1,+\infty[$, and set $g\colon (y,\eta) \mapsto$

$$\begin{cases} \dfrac{\rho\|y\|^p}{\eta^{p-1}} + p\eta\ln\eta - \eta\ln\left(\eta^p + \rho\|y\|^p\right), & \text{if } \ \eta > 0;\\ \rho\|y\|, & \text{if } \ \eta = 0 \text{ and } p = 1;\\ 0, & \text{if } \ \eta = 0,\ y = 0, \text{ and } p > 1;\\ +\infty, & \text{otherwise.} \end{cases}$$

- Then $g = [\phi \circ \|\cdot\|]^{\sim}$, where

$$\phi\colon \mathbb{R} \to\ ]-\infty,+\infty]\colon t \mapsto \rho|t|^p - \ln\left(1 + \rho|t|^p\right)$$

- For $\mathcal{G} = \mathbb{R}$ and $\rho = p = 1$, g is called the “fair” function in robust statistics; it also arises in least-squares regularization

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

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

and set $A\colon \mathcal{H} \to \mathcal{G} \times \mathbb{R}\colon x \mapsto (Lx - r, \langle x \mid u\rangle - \rho)$. Then

$$f = \widetilde{\varphi} \circ A \in \Gamma_0(\mathcal{H}).$$

Composite perspective functions: Examples

Example

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, let $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathsf{P})$, let $\varphi \in \Gamma_0(\mathcal{H})$, and set

$$f\colon \mathcal{H} \to \,]-\infty, +\infty] \colon X \mapsto \begin{cases} \mathsf{E}X\, \varphi\left(\dfrac{X}{\mathsf{E}X}\right), & \text{if } \mathsf{E}X > 0; \\ (\operatorname{rec} \varphi)(X), & \text{if } \mathsf{E}X = 0; \\ +\infty, & \text{if } \mathsf{E}X < 0. \end{cases}$$

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

Composite perspective functions: Examples

Example

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

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

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

Composite perspective functions: Examples

- The Fisher information

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

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

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 $I_0(x) = \{i \in I \mid \xi_i = 0\}$, $I_+(x) = \{i \in I \mid \xi_i > 0\}$, $J(x,y) = \{i \in I \,|\, \xi_i \geqslant 0 \text{ and } \eta_i < 0\}$, and $D_\phi(x,y) =$

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

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

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

- An essential tool in the design of splitting algorithms to solve convex minimization problems, especially in data science
  - PLC and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, 2005
  - PLC and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, 2011

Perspective functions: Proximity operator

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

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

- An essential tool in the design of splitting algorithms to solve convex minimization problems, especially in data science
  - PLC and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, 2005
  - PLC and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, 2011
- Basic properties:
  - $\mathrm{prox}_g + \mathrm{prox}_{g^*} = \mathrm{Id}$ (Moreau's decomposition)
  - $(\mathrm{prox}_g x, x - \mathrm{prox}_g x) = (\mathrm{prox}_g x, \mathrm{prox}_{g^*} x) \in \mathrm{gra}\, \partial g$
  - $\mathrm{Fix}\, \mathrm{prox}_g = \mathrm{Argmin} g$
  - $\|\mathrm{prox}_g x - \mathrm{prox}_g y\| \leqslant \|x - y\|$

Perspective functions: Proximity operator

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

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

- An essential tool in the design of splitting algorithms to solve convex minimization problems, especially in data science
  - PLC and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, 2005
  - PLC and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* Springer, New York, 2011
- Basic properties:
  - $\operatorname{prox}_g + \operatorname{prox}_{g^*} = \operatorname{Id}$ (Moreau's decomposition)
  - $(\operatorname{prox}_g x, x - \operatorname{prox}_g x) = (\operatorname{prox}_g x, \operatorname{prox}_{g^*} x) \in \operatorname{gra} \partial g$
  - $\operatorname{Fix} \operatorname{prox}_g = \operatorname{Argmin} g$
  - $\|\operatorname{prox}_g x - \operatorname{prox}_g y\|^2 \leqslant \|x - y\|^2 - \|\operatorname{prox}_{g^*} x - \operatorname{prox}_{g^*} y\|^2$

Perspective functions: Proximity operator

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

- Suppose that $\eta + \gamma\varphi^*(y/\gamma) \leqslant 0$. Then $\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = (0,0)$.
- Suppose that $\operatorname{dom}\varphi^*$ is open and that $\eta + \gamma\varphi^*(y/\gamma) > 0$. Then

  $$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = \big(y - \gamma p, \eta + \gamma\varphi^*(p)\big),$$

  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

Example

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

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

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

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

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

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

Perspective functions: Proximity operator

We can also handle cases when $\operatorname{dom}\varphi^*$ is not open.

Proposition

Let $\phi \in \Gamma_0(\mathbb{R})$ *be even, set* $\varphi = \phi \circ \|\cdot\|$, *let* $\gamma \in \left]0,+\infty\right[$, *let* $\eta \in \mathbb{R}$, *and let* $y \in \mathcal{G}$. *Set* $\mathcal{R} = \big\{(\nu,\chi) \in \mathbb{R}^2 \mid \chi + \phi^*(\nu) \leqslant 0\big\}$.

- *Suppose that* $\eta + \gamma\phi^*(\|x\|/\gamma) \leqslant 0$. *Then* $\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = (0,0)$.
- *Suppose that* $\eta > \gamma\phi(0)$ *and* $y = 0$. *Then*

$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = (y, \eta - \gamma\phi(0)).$$

- *Suppose that* $\eta + \gamma\phi^*(\|y\|/\gamma) > 0$ *and* $y \neq 0$, *and set* $(\nu,\chi) = \operatorname{proj}_{\mathcal{R}}(\|y\|/\gamma, \eta/\gamma)$. *Then*

$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = \left(\left(1 - \frac{\gamma\nu}{\|y\|}\right) y, \eta - \gamma\chi\right).$$

Perspective functions: Proximity operator

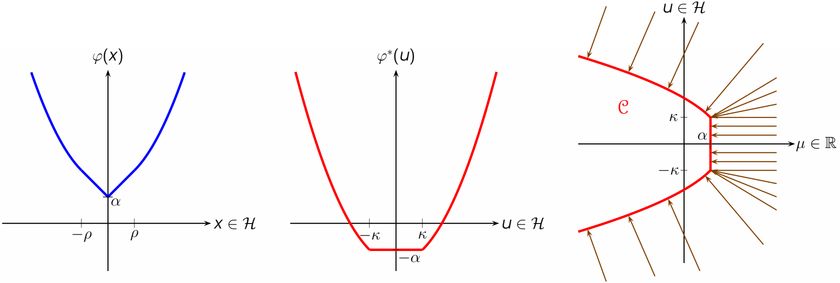


Figure: Geometry of the computation of $\operatorname{prox}_{\widetilde{\varphi}} = \operatorname{Id} - \operatorname{proj}_{\mathcal{C}}$.

$$\mathcal{C} = \big\{(u, \mu) \in \mathcal{G} \times \mathbb{R} \mid \mu + \varphi^*(u) \leqslant 0\big\}$$

Example: Generalized Huber function

- Let α, γ, and ρ be in $]0,+\infty[$, let $q \in \,]1,+\infty[$ and $q^* = q/(q-1)$.
- Define

$$\varphi\colon \mathcal{G} \to \mathbb{R}\colon y \mapsto \begin{cases} \alpha - \dfrac{\rho^{q^*}}{q^*} + \rho\|y\|, & \text{if } \|y\| > \rho^{q^*/q};\\[2ex] \alpha + \dfrac{\|y\|^q}{q}, & \text{if } \|y\| \leqslant \rho^{q^*/q}. \end{cases}$$

- Let $y \in \mathcal{G}$ and $\eta \in \mathbb{R}$. Then

$$\widetilde{\varphi}(y,\eta) = \begin{cases} \left(\alpha - \dfrac{\rho^{q^*}}{q^*}\right)\eta + \rho\|y\|, & \text{if } \eta > 0 \text{ and } \|y\| > \eta\rho^{q^*/q};\\[2ex] \alpha\eta + \dfrac{\|y\|^q}{q\eta^{q-1}}, & \text{if } \eta > 0 \text{ and } \|y\| \leqslant \eta\rho^{q^*/q};\\[1ex] \rho\|y\|, & \text{if } \eta = 0;\\ +\infty, & \text{if } \eta < 0. \end{cases}$$

Example: Generalized Huber function

In addition, the following hold:

- Suppose that $\|y\| \leqslant \gamma\rho$ and $\|y\|^{q^*} \leqslant \gamma^{q^*} q^*(\alpha - \eta/\gamma)$. Then $\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = (0,0)$.
- Suppose that $\eta \leqslant \gamma(\alpha - \rho^{q^*}/q^*)$ and $\|y\| > \gamma\rho$. Then
$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = \left(\left(1 - \frac{\gamma\rho}{\|y\|}\right)y, 0\right).$$
- Suppose that $\eta > \gamma(\alpha - \rho^{q^*}/q^*)$ and $\|y\| \geqslant \gamma\rho^{q^*-1}(\eta/\gamma + \rho^{2-q^*} + \rho^{q^*}/q^* - \alpha)$. Then
$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = \left(\left(1 - \frac{\gamma\rho}{\|y\|}\right)y, \eta + \gamma\left(\frac{\rho^{q^*}}{q^*} - \alpha\right)\right).$$
- Suppose that $\|y\|^{q^*} > q^*\gamma^{q^*}(\alpha - \eta/\gamma)$ and $\|y\| < \gamma\rho^{q^*-1}(\eta/\gamma + \rho^{2-q^*} + \rho^{q^*}/q^* - \alpha)$. If $y \neq 0$, let t be the unique solution in $]0,+\infty[$ to the equation
$$\gamma t^{2q^*-1} + q^*(\eta - \gamma\alpha)t^{q^*-1} + \gamma q^* t - q^*\|y\| = 0.$$
Set $p = ty/\|y\|$ if $y \neq 0$, and $p = 0$ if $y = 0$. Then
$$\operatorname{prox}_{\gamma\widetilde{\varphi}}(y,\eta) = \begin{cases} (y - \gamma p, \eta + \gamma(t^{q^*}/q^* - \alpha)), & \text{if } q^*\gamma^{q^*-1}\eta + \|y\|^{q^*} > q^*\gamma^{q^*}\alpha; \\ (0,0), & \text{if } q^*\gamma^{q^*-1}\eta + \|y\|^{q^*} \leqslant q^*\gamma^{q^*}\alpha. \end{cases}$$

Maximum-likelihood-type estimation

- **Data model:** The vector $y = (\eta_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^n$ of observations is

$$y = X\overline{b} + \overline{o} + Ce,$$

where $X \in \mathbb{R}^{n \times p}$ is a known design matrix with rows $(x_i)_{1 \leqslant i \leqslant n}$, $\overline{b} \in \mathbb{R}^p$ is the unknown regression vector (location), $\overline{o} \in \mathbb{R}^n$ is the unknown mean shift vector containing outliers, $e \in \mathbb{R}^n$ is a vector of realizations of i.i.d. zero mean random variables, and $C \in [0, +\infty[^{n \times n}$ is a diagonal matrix the diagonal of which are the (unknown) standard deviations.

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

$$\underset{b \in \mathbb{R}^p, \sigma > 0, \tau > 0}{\text{minimize}} \quad n\sigma + \sigma \sum_{i=1}^{n} \text{Huber}\left(\frac{\zeta_i - \langle b \mid x_i \rangle}{\sigma}\right) + p\tau + \tau \sum_{j=1}^{p} \text{Berhu}\left(\frac{\beta_j}{\tau}\right)$$

Maximum-likelihood-type estimation

- Let $\varsigma \in \Gamma_0(\mathbb{R}^N)$, let $\varpi \in \Gamma_0(\mathbb{R}^P)$, let $\theta \in \Gamma_0(\mathbb{R}^p)$, let $(n_i)_{1\leqslant i\leqslant N}$ be strictly positive integers such that $\sum_{i=1}^N n_i = n$, and let $(p_i)_{1\leqslant i\leqslant P}$ be strictly positive integers. For every $i \in \{1,\ldots,N\}$, let $\varphi_i \in \Gamma_0(\mathbb{R}^{n_i})$, let $X_i \in \mathbb{R}^{n_i\times p}$, and let $y_i \in \mathbb{R}^{n_i}$ be such that

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}.$$

  For every $i \in \{1,\ldots,P\}$, let $\psi_i \in \Gamma_0(\mathbb{R}^{p_i})$, and let $L_i \in \mathbb{R}^{p_i\times p}$.

- The objective of *perspective M-estimation* is to

$$\underset{s\in\mathbb{R}^N,\, t\in\mathbb{R}^P,\, b\in\mathbb{R}^p}{\text{minimize}} \quad \varsigma(s) + \varpi(t) + \theta(b) + \sum_{i=1}^N \widetilde{\varphi}_i\big(X_i b - y_i, \sigma_i\big) + \sum_{i=1}^P \widetilde{\psi}_i\big(L_i b, \tau_i\big)$$

Maximum-likelihood-type estimation

We recover a wide array of statistical problem formulations, including:

- Huber *M*-estimation (Huber, 1981)
- Fused lasso model (Tibshirani, 2005)
- Scaled lasso model (Antoniadis, 2010)
- Owen's concomitant estimation (Owen, 2007)
- Group Lasso (Bach et al, 2011)
- Adaptive BerHu robust regression (Lambert-Lacroix et al, 2016)
- Trend filtering (Tibshirani, 2014)
- Scaled square-root elastic net estimation (Raninen and E. Ollila, 2017)
- etc.

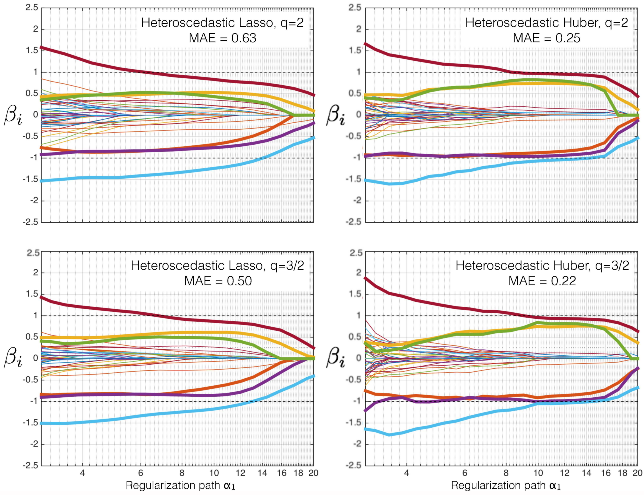
Maximum-likelihood-type estimation: Algorithm

for $k = 0, 1, \ldots$

$$
\begin{array}{l}
q_{s,k} = x_{s,k} - h_{s,k}\\
q_{t,k} = x_{t,k} - h_{t,k}\\
q_{b,k} = Ax_{b,k} - h_{b,k}\\
\text{for } i = 1, \ldots, N\\
\quad q_{i,k} = X_i x_{b,k} - h_{i,k}\\
\text{for } i = 1, \ldots, P\\
\quad q_{N+i,k} = L_i x_{b,k} - h_{N+i,k}\\
s_k = x_{s,k} - q_{s,k}/2\\
t_k = x_{t,k} - q_{t,k}/2\\
b_k = x_{b,k} - Qq_{b,k}\\
z_{s,k} = \text{prox}_{\gamma\varsigma}(2s_k - x_{s,k})\\
z_{t,k} = \text{prox}_{\gamma\varpi}(2t_k - x_{t,k})\\
z_{b,k} = \text{prox}_{\gamma\theta}(2b_k - x_{b,k})\\
x_{s,k+1} = x_{s,k} + \mu_k(z_{s,k} - s_k)\\
x_{t,k+1} = x_{t,k} + \mu_k(z_{t,k} - t_k)\\
x_{b,k+1} = x_{b,k} + \mu_k(z_{b,k} - b_k)\\
\text{for } i = 1, \ldots, N\\
\quad c_{i,k} = X_i b_k\\
\quad (\delta_{i,k}, d_{i,k}) = (0, y_i) + \text{prox}_{\gamma\widetilde{\varphi}_i}(2\sigma_{i,k} - \eta_{i,k}, 2c_{i,k} - h_{i,k} - y_i)\\
\text{for } i = 1, \ldots, P\\
\quad c_{N+i,k} = L_i b_k\\
\quad (\delta_{N+i,k}, d_{N+i,k}) = \text{prox}_{\gamma\widetilde{\psi}_i}(2\tau_{i,k} - \eta_{N+i,k}, 2c_{N+i,k} - h_{N+i,k})\\
h_{s,k+1} = h_{s,k} + \mu_k(d_{s,k} - s_k)\\
h_{t,k+1} = h_{t,k} + \mu_k(d_{t,k} - t_k)\\
h_{b,k+1} = h_{b,k} + \mu_k(d_{b,k} - c_{b,k})
\end{array}
$$

Maximum-likelihood-type estimation: Algorithm

See paper for details of these experiments.



References

- PLC, Perspective functions: Properties, constructions, and examples, *Set-Valued Var. Anal.*, vol. 26, pp. 247–264, 2018.
- PLC and C. L. Müller, Perspective functions: Proximal calculus and applications in high-dimensional statistics, *J. Math. Anal. Appl.*, vol. 457, pp. 1283–1306, 2018.
- PLC and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Elec. J. Stat.*, vol. 14, pp. 207-238, 2020.
- PLC and C. L. Müller, Regression models for compositional data: General log-contrast formulations, proximal optimization, and microbiome data applications, *Stat. Biosci.*, 2020.
- H. H. Bauschke and PLC, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed., corrected printing. Springer, New York, 2019.
- Chierchia, Chouzenoux, PLC, Pesquet, *Proximity Operator Repository*, `http://proximity-operator.net/`