Regularizing with Bregman–Moreau envelopes

Heinz H. Bauschke
COCANA
University of British Columbia
heinz.bauschke@ubc.ca

Minh N. Dao
CARMA
University of Newcastle
daonminh@gmail.com

Scott B Lindstrom
CARMA
University of Newcastle
scott.lindstrom@uon.edu.au

AMSI Optimise 2017

Revised July 4, 2017
Definition 1 (Moreau envelope)

Moreau envelope with parameter $\gamma \in \mathbb{R}^{++}$

$$\text{env}_{\theta}^{\gamma} : x \mapsto \inf_{y \in X} \theta(y) + \frac{1}{2\gamma} \|x - y\|^2. \quad (1)$$

- A special case of infimal convolution:

$$\theta \Box f : \mathbb{R}^n \to [-\infty, \infty] : x \mapsto \inf_{y \in \mathbb{R}^n} (\theta(y) + f(x - y))$$

(“exact” if $\theta \Box f(x) = \min_{y \in \mathbb{R}^n} (\theta(y) + f(x - y)) \forall x \in \text{dom } \theta \Box f$

- Moreau only considered $\gamma = 1$

- Systematic study involving $\gamma$ originated with Attouch [2][3]
Think: smoothing through epigraph addition

$\text{epi}(\theta \Box f) = \text{epi } \theta + \text{epi } f^1$ is always true when $\theta \Box f$ is exact.
As $\gamma \to 0$ we recover $\theta$
As $\gamma \to \infty$ we recover $\min \theta$
Varying the Parameter

Varying the parameter
Throughout: \( \theta \) is a lower semicontinuous convex function of Legendre type

**Definition 2 (Prox Operator)**

\( \text{Prox}_{\gamma \theta}(x) \) is the unique point satisfying

\[
\text{env}_{\gamma \theta}(x) = \min_{y \in \mathbb{R}^n} \left( \theta(y) + \frac{1}{2\gamma} \|x - y\|^2 \right)
\]

\[
= \theta(\text{Prox}_{\gamma \theta}(x)) + \frac{1}{2\gamma} \|x - \text{Prox}_{\gamma \theta}(x)\|^2
\]
Prox Operators: Geometric Intuition

Figure: Where $\theta = |y + x - 1|$ and $\gamma = 1/2$

Figure: The net Prox $\gamma \theta(1, 2)$ where $\gamma \in ]0, \infty[$
Prox Operators: Limiting Cases

**Figure:** \( \lim_{\gamma \to \infty} \text{Prox} \, \gamma \theta(x) = \text{P}_{\text{argmin}} \theta(x) \)
Figure: $\lim_{\gamma \to 0} \text{Prox}_{\gamma \theta}(x) = x$
Remark 1 (Special case: projection)

When $\theta = \nu_C$ is the indicator of $C$

$$\text{Prox}_{\gamma \theta}(x) = P_C(x)$$

is the projection operator.
Definition 3 (Bregman Distance)

The Bregman Distance of a function $f$ between two points $x, y$ is

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

Figure: Bregman distance where the function $f$ is the Boltzmann-Shannon Entropy $x \mapsto x \log(x) - x$
Our assumptions on $f$

We assume:

1. $f$ is a lower semicontinuous convex function of Legendre type and $U := \text{int dom } f$.
2. $\nabla^2 f$ exists and is continuous on $U$;
3. $D_f$ is jointly convex, i.e., convex on $X \times X$;
4. $(\forall x \in U) \ D_f(x, \cdot)$ is strictly convex on $U$;
5. $(\forall x \in U) \ D_f(x, \cdot)$ is coercive, i.e., $D_f(x, y) \to +\infty$ as $\|y\| \to +\infty$. 
Our demo functions

Where $x, y \in \mathbb{R}^J$:

1. **Energy:** If $f : x \mapsto \frac{1}{2} \| x \|^2$, then
   \[ D_f(x, y) = \frac{1}{2} \| x - y \|^2 \]

2. **Boltzmann–Shannon** entropy: If $f : x \mapsto \sum_{j=1}^{J} x_j \ln(x_j) - x_j$, then one obtains the Kullback–Leibler divergence
   \[ D_f(x, y) = \begin{cases} \sum_{j=1}^{J} x_j \ln(x_j/y_j) - x_j + y_j, & \text{if } x \geq 0 \text{ and } y > 0; \\ +\infty, & \text{otherwise.} \end{cases} \]

3. **Fermi–Dirac entropy:** If $f : x \mapsto \sum_{j=1}^{J} x_j \ln(x_j) + (1 - x_j) \ln(1 - x_j)$, then
   \[ D_f(x, y) = \begin{cases} \sum_{j=1}^{J} x_j \ln(x_j/y_j) + (1 - x_j) \log \left( \frac{(1 - x_j)}{(1 - y_j)} \right), & \text{if } 0 \leq x \leq 1 \text{ and } 0 < y < 1; \\ +\infty, & \text{otherwise.} \end{cases} \]

\[ \text{With Boltzmann–Shannon entropy and Fermi–Dirac entropy, we use convention } 0 \cdot \ln(0) := 0. \]
What Changes?

We lose triangle Inequality.

\[
D_f(x, y) > D_f(z, y) + D_f(x, z).
\]

**Figure:** Where \( f \) is the Boltzmann-Shannon Entropy,
What Changes?

We also lose symmetry... and translation invariance.

\[ D_f\left(x, \frac{1}{2}\right), \quad D_f\left(\frac{1}{2}, x\right) \]

\[ D_f\left(x, \frac{1}{4}\right), \quad D_f\left(x, \frac{3}{4}\right) \]
Except when using the energy, of course.
Definition 4

For a given $\theta, f$ where int dom $f \cap$ dom $\theta \neq \emptyset$:

- The left Bregman envelope is

$$\hat{\text{env}}_{\theta}^\gamma : X \rightarrow [-\infty, +\infty] : y \mapsto \inf_{x \in X} \theta(x) + \frac{1}{\gamma} D_f(x, y) \quad (2)$$

- The right Bregman envelope is

$$\text{env}_{\theta}^\gamma : X \rightarrow [-\infty, +\infty] : x \mapsto \inf_{y \in X} \theta(y) + \frac{1}{\gamma} D_f(x, y), \quad (3)$$

If $f = \frac{1}{2} \| \cdot \|^2$, then $D_f : (x, y) \mapsto \frac{1}{2} \| x - y \|^2$, and

$$\hat{\text{env}}_{\theta}^\gamma = \text{env}_{\theta}^\gamma = \theta \square \left( \frac{1}{2\gamma} \| \cdot \|^2 \right)$$

is the classical Moreau envelope of $\theta$ of parameter $\gamma$. 


Consider the following properties:

(a) \( U \cap \text{dom}\ \theta \) is bounded.
(b) \( \inf \theta(U) > -\infty \).
(c) \( f \) is supercoercive, i.e., \( f(x)/\|x\| \to +\infty \) as \( \|x\| \to +\infty \).
(d) \( (\forall x \in U) \ D_f(x, \cdot) \) is supercoercive.

Then the following hold (and we suppose them moving forward):

- If any of (a), (b), or (c) holds, then
  \[
  (\forall y \in U) \quad \theta(\cdot) + \frac{1}{\gamma} D_f(\cdot, y) \text{ is coercive}
  \]

- If any of (a), (b), or (d) holds, then
  \[
  (\forall x \in U) \quad \theta(\cdot) + \frac{1}{\gamma} D_f(x, \cdot) \text{ is coercive.}
  \]
Definition 5 (Bregman Proximity Operators)

For a given $\theta, f$ where $\text{int dom } f \cap \text{dom } \theta \neq \emptyset$:

1. The *left* prox operator is

$$\overleftarrow{P}_\theta : \text{int dom } f \to \text{int dom } f : y \mapsto \arg\min_{x \in X} \theta(x) + \frac{1}{\gamma} D_f(x, y).$$

2. The *right* prox operator is

$$\overrightarrow{P}_\theta : \text{int dom } f \to \text{int dom } f : x \mapsto \arg\min_{y \in X} \theta(y) + \frac{1}{\gamma} D_f(x, y).$$
Shown: left envelope with Boltzman-Shannon entropy
Still regularizes
Addition changes
Shown: *right* envelope with Boltzmann-Shannon entropy

Think: what about limiting cases?
With $x, y \in \text{int dom } f$ the following hold:

- **As $\gamma \downarrow 0$:**
  1. **Left case:** $\tilde{\text{env}}_\gamma(y) \uparrow \theta(y)$, $\theta(\tilde{P}_\gamma(y)) \uparrow \theta(y)$,
     \[ \frac{1}{\gamma} D_f(\tilde{P}_\gamma(y), y) \rightarrow 0, \text{ and } \tilde{P}_\gamma(y) \rightarrow y. \]
  2. **Right case:** $\tilde{\text{env}}_\gamma(x) \uparrow \theta(x)$, $\theta(\tilde{P}_\gamma(x)) \uparrow \theta(x)$,
     \[ \frac{1}{\gamma} D_f(x, \tilde{P}_\gamma(x)) \rightarrow 0, \text{ and } \tilde{P}_\gamma(x) \rightarrow x. \]

- **As $\gamma \uparrow \infty$:**
  1. **Left case:** $\tilde{\text{env}}_\gamma(y) \downarrow \inf \theta(X)$ and if $\text{argmin } \theta \subseteq \text{int dom } f$,
     then $\tilde{P}_\gamma(y) \rightarrow \tilde{P}_{\text{argmin } \theta} y$ as $\gamma \uparrow +\infty$.
  2. **Right case:** $\tilde{\text{env}}_\gamma(x) \downarrow \inf \theta(X)$ and if $\text{argmin } \theta \subseteq \text{int dom } f$,
     then $\tilde{P}_\gamma(x) \rightarrow \tilde{P}_{\text{argmin } \theta} x$ as $\gamma \uparrow +\infty$. 

\[24/39\]
Example: Boltzmann-Shannon Entropy
Example: Fermi-Dirac Entropy
**Prox Operators: Geometric Intuition**

**Figure:** \( \theta = |y + x - 1| \) and 
\( D_f(\cdot, (1, 2)) \) where \( \gamma = 1 \) and \( f \) is Boltzmann-Shannon entropy

**Figure:** The net \( \hat{P}_{\gamma \theta}(1, 2) \) where 
\( \gamma \in ]0, \infty[ \)
**Definition 6 (Bregman projectors)**

Let $C$ be a closed convex subset of $X$, $\text{int} \text{ dom } f \cap C \neq \emptyset$.

- $\leftarrow P_C := \leftarrow P_{\iota_C}$ is the *left* Bregman projector onto $C$.
- $\rightarrow P_C := \rightarrow P_{\iota_C}$ is the *right* Bregman projector onto $C$.

Bregman and Euclidean projections may differ in $\mathbb{R}^n$ for $n > 1$. 
Figure: Bregman projection with Boltzmann-Shannon entropy and with the energy (Euclidean case)
Prox Operators: Limiting Cases

\[ \lim_{\gamma \to \infty} \overrightarrow{P}_{\gamma \theta}(x) = \overrightarrow{P}_{\text{argmin } \theta}(x) \]

(Analogously: \( \lim_{\gamma \to \infty} \overrightarrow{P}_{\gamma \theta}(x) = \overrightarrow{P}_{\text{argmin } \theta}(x) \))
Prox Operators: Limiting Cases

Figure: $\lim_{\gamma \to 0} \overleftarrow{P}_{\gamma \theta}(x) = x$

(Analogously: $\lim_{\gamma \to 0} \overrightarrow{P}_{\gamma \theta}(x) = x$)
Contributions of this work include:

1. Limit as the parameter goes to infinity
2. Exploration of right prox
3. Computed examples prototypical of polyhedral adaptation
Thanks for listening!


References II


