Generalized Bregman Distances

Regina S. Burachik

University of South Australia

Workshop on Optimization and Operator Theory

15 November 2021

Dedicated to Professor Lev Bregman

University of South Australia

Outline

University of South Australia

This talk contains joint work with:

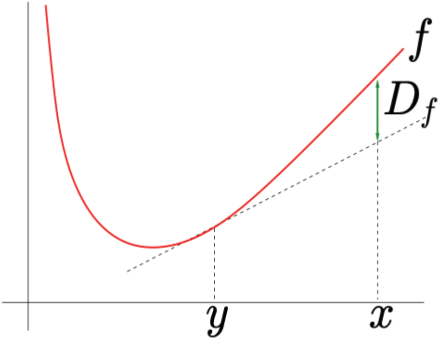
Juan Enrique Martínez Legaz
Universitat Autónoma de Barcelona, Spain

Minh N. Dao
Federation University, Ballarat, Australia

Scott B. Lindstrom
Hong Kong Polytechnic and Curtin University, Perth, Australia

$f : X \to \mathbb{R}_{+\infty}$ strictly convex, smooth, (Bregman, 1967)

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



Using specific choices of f, we obtain:

Kullback-Liebler divergence A distance between positive vectors, used in information theory, statistics, portfolio selection,...

Itakura-Saito divergence A non-symmetric measure of difference between probability distributions. Used to measure sound quality and speech processing

Squared Euclidean (ℓ_2) distance for $f(\cdot) := (1/2)\|\cdot\|^2$ we have $D_f(x, y) = (1/2)\|x - y\|^2$

University of South Australia

Using specific choices of f, we obtain:

Kullback-Liebler divergence A distance between positive vectors, used in information theory, statistics, portfolio selection,...

Itakura-Saito divergence A non-symmetric measure of difference between probability distributions. Used to measure sound quality and speech processing

Squared Euclidean (ℓ_2) distance for $f(\cdot) := (1/2)\|\cdot\|^2$ we have $D_f(x, y) = (1/2)\|x - y\|^2$

University of South Australia

What is the Bregman distance really measuring?

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

$$(Fenchel - Young) = f(x) + f^*(\nabla f(y)) - \langle \nabla f(y), x \rangle$$

Hence, $\boxed{D_f(x,y) = 0 \iff \nabla f(x) = \nabla f(y)}$

so D_f measures distance between images of the gradient map!

If f is smooth, then

$$D_f(x,y) \quad \geq \quad 0 \;\; \forall x,y \iff f \text{ is convex}$$

Due to a "Fenchel-Young" property for the graph of ∇f:

$$f(x) + f^*(v) \geq \langle x, v\rangle \qquad \forall x \in X,\; v \in X^*$$

$$f(x) + f^*(v) = \langle x, v\rangle \iff v = \nabla f(x)$$

An analogous property holds for max-mon T!

The family $\mathcal{H}(T)$ Introduced in [RSB-Svaiter, 2002]

For $T : X \rightrightarrows X^*$ max-mon, the Fitzpatrick family $\mathcal{H}(T)$ consists of functions $h : X \times X^* \to \mathbb{R}_{+\infty}$ convex and norm-weak* lsc, s.t.

$$h(x, v) \geq \langle x, v\rangle \quad \forall x \in X,\ v \in X^*$$
$$h(x, v) = \langle x, v\rangle \iff v \in Tx,$$

Fenchel-Young gives $h(x, v) := f(x) + f^*(v) \in \mathcal{H}(\partial f)$

University of South Australia

$\mathcal{H}(T)$ and the Fitzpatrick function

$\mathcal{H}(T)$ has a smallest and a biggest element. The smallest one is the *Fitzpatrick function*[1]:

$$\mathcal{F}_T(x,v) := \sup_{(z,w)\in G(T)} \langle z-x, v-w\rangle + \langle x, v\rangle.$$

The biggest is $\sigma_T = cl\,conv(\pi + \delta_{G(T)})$ and in fact

$$\sigma_T(x,v) := \mathcal{F}_T{}^*(v,x),$$

where $\pi := \langle\cdot,\cdot\rangle$.

[1] Fitzpatrick, 1988.

University of South Australia

Extending the Bregman distance to a generic T

For T max-mon, $x, y \in D(T)$, and $h \in \mathcal{H}(T)$, recall that

$$\begin{aligned} h(x,v) &\geq \langle x, v\rangle \quad \forall x \in X,\ v \in X^* \\ h(x,v) &= \langle x, v\rangle \iff v \in Tx, \end{aligned}$$

The "sharp" version $D_T^{\sharp,h}(x,y) := \sup_{v\in Ty} h(x,v) - \langle x, v\rangle$

The "flat" version $D_T^{\flat,h}(x,y) := \inf_{v\in Ty} h(x,v) - \langle x, v\rangle$

Note that the distance depends on the choice of h!

Extending the Bregman distance to a generic T

For T max-mon, $x, y \in D(T)$, and $h \in \mathcal{H}(T)$, recall that

$$h(x,v) \geq \langle x, v\rangle \quad \forall x \in X,\ v \in X^*$$
$$h(x,v) = \langle x, v\rangle \iff v \in Tx,$$

The "sharp" version $D_T^{\sharp,h}(x,y) := \sup_{v \in Ty} h(x,v) - \langle x, v\rangle$

The "flat" version $D_T^{\flat,h}(x,y) := \inf_{v \in Ty} h(x,v) - \langle x, v\rangle$

Note that the distance depends on the choice of h!

Extending the Bregman distance to a generic T

For T max-mon, $x, y \in D(T)$, and $h \in \mathcal{H}(T)$, recall that

$$\begin{aligned} h(x,v) &\geq \langle x, v\rangle \quad \forall x \in X,\ v \in X^* \\ h(x,v) &= \langle x, v\rangle \iff v \in Tx, \end{aligned}$$

The "sharp" version $D_T^{\sharp,h}(x,y) := \sup_{v\in Ty} h(x,v) - \langle x, v\rangle$

The "flat" version $D_T^{\flat,h}(x,y) := \inf_{v\in Ty} h(x,v) - \langle x, v\rangle$

Note that the distance depends on the choice of h!

University of South Australia

Extending the Bregman distance to a generic T

For T max-mon, $x, y \in D(T)$, and $h \in \mathcal{H}(T)$, recall that

$$h(x,v) \geq \langle x, v\rangle \quad \forall x \in X,\ v \in X^*$$
$$h(x,v) = \langle x, v\rangle \iff v \in Tx,$$

The "sharp" version $D_T^{\sharp,h}(x,y) := \sup_{v\in Ty} h(x,v) - \langle x, v\rangle$

The "flat" version $D_T^{\flat,h}(x,y) := \inf_{v\in Ty} h(x,v) - \langle x, v\rangle$

Note that the distance depends on the choice of h!

University of South Australia

Particular case 1: $T = \nabla f$ and $h = h^{FY} = f + f^*$

If f convex and smooth, choose then

$$D^{\sharp,h^{FY}}_{\nabla f}(x,y) = D^{\flat,h^{FY}}_{\nabla f}(x,y) = D_f(x,y),$$

for $(x,y) \notin (\operatorname{dom} f \setminus \operatorname{dom} \nabla f) \times \operatorname{dom} \nabla f$ (region where coincides with the classical Bregman distance)

University of South Australia

Particular case 2: T max-mon and $h = \mathcal{F}_T$

The "sharp" version $D_T^{\sharp,h}(x,y) := \sup_{v \in Ty} \mathcal{F}_T(x,v) - \langle x, v \rangle$

The "flat" version $D_T^{\flat,h}(x,y) := \inf_{v \in Ty} \mathcal{F}_T(x,v) - \langle x, v \rangle$

where $\mathcal{F}_T(x,v) := \sup_{(z,w) \in G(T)} \langle z - x, v - w \rangle + \langle x, v \rangle.$

University of South Australia

Particular case 2: T max-mon and $h = \mathcal{F}_T$

The “sharp” version $D_T^{\sharp,h}(x,y) := \sup_{v \in Ty} \mathcal{F}_T(x,v) - \langle x, v\rangle$

The “flat” version $D_T^{\flat,h}(x,y) := \inf_{v \in Ty} \mathcal{F}_T(x,v) - \langle x, v\rangle$

where $\mathcal{F}_T(x,v) := \sup_{(z,w) \in G(T)} \langle z - x, v - w\rangle + \langle x, v\rangle.$

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 1. Negative Burg entropy

$f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(t) = \begin{cases} -\log t, & t > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x, y) = \begin{cases} \left(\sqrt{\frac{x}{y}} - 1\right)^2, & x \geq 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

University of South Australia

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 1. Negative Burg entropy

$f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(t) = \begin{cases} -\log t, & t > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x, y) = \begin{cases} \left(\sqrt{\frac{x}{y}} - 1\right)^2, & x \geq 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

University of South Australia

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 1. Negative Burg entropy

$f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(t) = \begin{cases} -\log t, & t > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x, y) = \begin{cases} \left(\sqrt{\frac{x}{y}} - 1\right)^2, & x \geq 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

University of South Australia

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 2: negative Boltzmann-Gibbs-Shannon/Kullback-Leibler/ entropy

$$f : \mathbb{R} \to \mathbb{R} \text{ defined as } f(t) = \begin{cases} t\log t - t, & t > 0 \\ 0 & t = 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x,y) = \begin{cases} \frac{y}{e}, & x = 0,\ y \geq 0 \\ x\left[W(e\,\frac{x}{y}) + \frac{1}{W(e\,\frac{x}{y})} - 2\right], & x > 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$W : \mathbb{R}_{++} \to \mathbb{R}_{++}$ inverse of $t \to te^t$ (Lambert-W function). Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

University of South Australia

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 2: negative Boltzmann-Gibbs-Shannon/Kullback-Leibler/ entropy

$$f : \mathbb{R} \to \mathbb{R} \text{ defined as } f(t) = \begin{cases} t \log t - t, & t > 0 \\ 0 & t = 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x,y) = \begin{cases} \dfrac{y}{e}, & x = 0,\ y \geq 0 \\ x\left[W(e\,\tfrac{x}{y}) + \dfrac{1}{W(e\,\tfrac{x}{y})} - 2\right], & x > 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$W : \mathbb{R}_{++} \to \mathbb{R}_{++}$ inverse of $t \to te^t$ (Lambert-W function). Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

$T = \nabla f$ and $h = \mathcal{F}_{\nabla f}$

Example 2: negative Boltzmann-Gibbs-Shannon/Kullback-Leibler/ entropy

$$f : \mathbb{R} \to \mathbb{R} \text{ defined as } f(t) = \begin{cases} t \log t - t, & t > 0 \\ 0 & t = 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$$D_{\nabla f}^{\mathcal{F}_{\nabla f}}(x,y) = \begin{cases} \dfrac{y}{e}, & x = 0,\ y \geq 0 \\ x\left[W(e\,\frac{x}{y}) + \dfrac{1}{W(e\,\frac{x}{y})} - 2\right], & x > 0,\ y > 0 \\ +\infty, & \text{c.c.} \end{cases}$$

$W : \mathbb{R}_{++} \to \mathbb{R}_{++}$ inverse of $t \to te^t$ (Lambert-W function). Used $\mathcal{F}_{\nabla f}$ from [Bauschke-McLaren-Sendov, 2005]

Prox-envelopes: Moreau, Bregman, Fitzpatrick

The Moreau prox-envelope of a convex $\theta : \mathbb{R} \to \mathbb{R}$:

$$e^M_\gamma(\theta)(x) := \inf_w \left\{ \theta(w) + \frac{1}{\gamma}\|w - x\|^2 \right\}$$

The Bregman prox-envelope:

$$e^B_\gamma(\theta)(x) := \inf_w \left\{ \theta(w) + \frac{1}{\gamma} D_f(w, x) \right\},$$

where $D_f(x, y) = D^{h^{FY}_f}_{\nabla f}(x, y) = D^{(f+f^*)}_{\nabla f}(x, y)$ classical Bregman distance

A new family of Prox-envelopes

The Fitzpatrick prox-envelope:

$$e_\gamma^F(\theta)(x) := \inf_w \left\{ \theta(w) + \frac{1}{\gamma} D_{\nabla f}^{\mathcal{F}_{\nabla f}}(w,x) \right\}$$

or, in general, taking $D_T^{\star,h} \in \{D_T^{\sharp,h}, D_T^{\flat,h}\}$:

$$e_\gamma^{\star,h,T}(\theta)(x) := \inf_w \left\{ \theta(w) + \frac{1}{\gamma} D_T^{\star,h}(w,x) \right\}$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \to y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \to y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \to y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \to y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \to y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \rightharpoonup y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \rightharpoonup y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \rightharpoonup y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \downarrow 0$:

- θ convex, proper, lsc, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$
- $\varphi_\mu(\cdot) := \theta(\cdot) + \frac{1}{\mu}\mathcal{D}_T^{\sharp,h}(\cdot, y)$ is coercive for some $\mu \in \mathbb{R}_{++}$.

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \downarrow 0$,

- $\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0$.
- If T is str. mon. over $\operatorname{dom} T \cap \operatorname{dom} \theta$,

$$s_\gamma \rightharpoonup y, \quad e_\gamma^{\sharp,h,T}(\theta)(y) \uparrow \theta(y), \quad \theta(s_\gamma) \to \theta(y), \quad \frac{1}{\gamma}\mathcal{D}_T^{\sharp,h}(s_\gamma, y) \to 0.$$

University of South Australia

Asymptotic behaviour when $\gamma \uparrow +\infty$:

- θ convex, proper, lsc, and coercive, $y \in \mathrm{dom}\, T \cap \mathrm{dom}\, \theta$
- $(\mathrm{dom}\, T \cap \mathrm{dom}\, \theta) \times \{y\} \subseteq \mathrm{dom}\, \mathcal{D}_T^{\sharp,h}$

Let $s_\gamma \in \mathrm{Argmin}\, \varphi_\gamma$. Then, as $\gamma \uparrow +\infty$,

$$e_\gamma^{\sharp,h,T}(\theta)(y) \downarrow \inf \theta(\mathrm{dom}\, T)$$

University of South Australia

Asymptotic behaviour when $\gamma \uparrow +\infty$:

- θ convex, proper, lsc, and coercive, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \uparrow +\infty$,

$$e_\gamma^{\sharp,h,T}(\theta)(y) \downarrow \inf \theta(\operatorname{dom} T)$$

Asymptotic behaviour when $\gamma \uparrow +\infty$:

- θ convex, proper, lsc, and coercive, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \uparrow +\infty$,

$$e_\gamma^{\sharp,h,T}(\theta)(y) \downarrow \inf \theta(\operatorname{dom} T)$$

Asymptotic behaviour when $\gamma \uparrow +\infty$:

- θ convex, proper, lsc, and coercive, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \uparrow +\infty$,

$$e_\gamma^{\sharp,h,T}(\theta)(y) \downarrow \inf \theta(\operatorname{dom} T)$$

University of South Australia

Asymptotic behaviour when $\gamma \uparrow +\infty$:

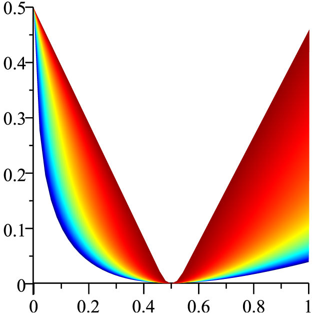
- θ convex, proper, lsc, and coercive, $y \in \operatorname{dom} T \cap \operatorname{dom} \theta$
- $(\operatorname{dom} T \cap \operatorname{dom} \theta) \times \{y\} \subseteq \operatorname{dom} \mathcal{D}_T^{\sharp,h}$

Let $s_\gamma \in \operatorname{Argmin} \varphi_\gamma$. Then, as $\gamma \uparrow +\infty$,

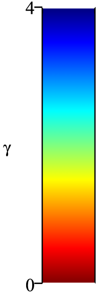
$$e_\gamma^{\sharp,h,T}(\theta)(y) \downarrow \inf \theta(\operatorname{dom} T)$$

University of South Australia

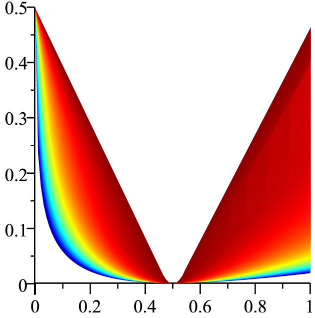
Bregman vs. Fitzpatrick envelope for $\theta(t) := |t - (1/2)|$ for Kullback-Leibler f, $T = \log(\cdot)$



Bregman envelope
$h = f + f^*$



color key



Fitzpatrick envelope
$h = \mathcal{F}_{\nabla f}$

*Credit for the figures: Scott Lindstrom

University of South Australia

Bregman distance between two set-valued maps

$T, S : X \rightrightarrows X^*$, T max-mon, S gral., $h_T \in \mathcal{H}(T)$:

The "sharp" version

$$D_S^{\sharp, h_T}(x, y) := \sup_{v \in Sy} h_T(x, v) - \langle x, v \rangle$$

The "flat" version

$$D_S^{\flat, h_T}(x, y) := \inf_{v \in Sy} h_T(x, v) - \langle x, v \rangle$$

University of South Australia

A new measure of degree of overlap between Tx and Sy

T max-mon, $h \in \mathcal{H}(T)$, $(x, y) \in D(T) \times D(S)$:

(a) If S loc. bded in $\mathrm{int} D(S)$, weakly-closed valued,

$(x, y) \notin \mathrm{bdry} D(S) \times \mathrm{bdry} D(T)$.

$$D_S^{\flat,h}(x, y) = 0 \iff Sy \cap Tx \neq \emptyset$$

(b) $D_S^{\sharp,h}(x, y) = 0 \iff Sy \subset Tx$

University of South Australia

A new measure of degree of overlap between Tx and Sy

T max-mon, $h \in \mathcal{H}(T)$, $(x,y) \in D(T) \times D(S)$:

(a) If S loc. bded in $\mathrm{int} D(S)$, weakly-closed valued,

$(x,y) \notin \mathrm{bdry} D(S) \times \mathrm{bdry} D(T)$.

$$D_S^{\flat,h}(x,y) = 0 \quad \Longleftrightarrow \quad Sy \cap Tx \neq \emptyset$$

(b) $$D_S^{\sharp,h}(x,y) = 0 \quad \Longleftrightarrow \quad Sy \subset Tx$$

University of South Australia

T max-mon, $h \in \mathcal{H}(T)$, $u \in X^*$: Find x s.t. $u \in Tx$ (P_T)

If (P_T) difficult, find y sol. of the "better conditioned" problem

$$(P_S) \qquad u \in Sy$$

For this y, find x s.t. $D_S^{\sharp,h}(x,y) \leq \varepsilon$, then x solves

$$(P_{T,\varepsilon}) \qquad u \in T^e(\varepsilon, x),$$

where

$$T^e(\varepsilon, x) := \{v \in X^* \ : \ \langle x - y, v - w\rangle \geq -\varepsilon \ \forall (y,w) \in G(T)\}$$

Note: $D_S^{\sharp,h}(\cdot, y)$ is convex and x solves an ε-approximation of (P_T)!

University of South Australia

An enlargement of T induced by $h \in \mathcal{H}(T)$

T max-mon, the *enlargement* L^h of T is

$$L^h(\varepsilon, x) := \{v \in X^* : h(x, v) - (x, v) \leq \varepsilon\}.$$

Consider the problem:

(PS) find $x \in X$ s.t. $0 \in Sx + Tx$.

GBDs define approximate solutions of (PS) using L^h

University of South Australia

T max-mon, S point-to-set, $h \in \mathcal{H}(T)$

Consider the following statements:

(a) $0 \in L^h(\varepsilon, x) + Sx$.

(b) $\mathcal{D}^{\flat,h}_{-S}(x,x) \leq \varepsilon$.

Then (a) $\Longrightarrow$ (b) (necessary condition for optimality). Moreover, if $\operatorname{dom} S$ is open and S is locally bounded with weakly closed images, then the two statements are equivalent.

T max-mon, S point-to-set, $h \in \mathcal{H}(T)$

Consider the following statements:

(a) $0 \in L^h(\varepsilon, x) + Sx$.

(b) $\mathcal{D}^{\flat,h}_{-S}(x, x) \leq \varepsilon$.

Then (a) $\implies$ (b) (necessary condition for optimality). Moreover, if $\operatorname{dom} S$ is open and S is locally bounded with weakly closed images, then the two statements are equivalent.

University of South Australia

Some Questions

(a) How can we use these new distances for obtaining more efficient solution techniques for variational inequalities/inclusion problems

(b) When $T = S = \partial f$, will the Fitzpatrick distances play a role similar to that of the classical Bregman distances (Bregman projections, convergence analysis, etc)?

(c) Can these distances be used to regularize/penalize in prox-like iterations for variational inequalities?

Some Questions

(a) How can we use these new distances for obtaining more efficient solution techniques for variational inequalities/inclusion problems

(b) When $T = S = \partial f$, will the Fitzpatrick distances play a role similar to that of the classical Bregman distances (Bregman projections, convergence analysis, etc)?

(c) Can these distances be used to regularize/penalize in prox-like iterations for variational inequalities?

Some Questions

(a) How can we use these new distances for obtaining more efficient solution techniques for variational inequalities/inclusion problems

(b) When $T = S = \partial f$, will the Fitzpatrick distances play a role similar to that of the classical Bregman distances (Bregman projections, convergence analysis, etc)?

(c) Can these distances be used to regularize/penalize in prox-like iterations for variational inequalities?

Conclusions

Convex functions appear naturally when studying maximally monotone operators.

Some notions involving convex functions (such as classical Bregman distances), can be extended to maximally monotone operators, thus producing new tools both for convex analysis and for maximal monotone theory.

University of South Australia

Conclusions

Convex functions appear naturally when studying maximally monotone operators.

Some notions involving convex functions (such as classical Bregman distances), can be extended to maximally monotone operators, thus producing new tools both for convex analysis and for maximal monotone theory.

University of South Australia

References used in this talk

- RSB, Juan Enrique Martínez Legaz, On Bregman-type distances for convex functions and maximally monotone operators, Set-Valued Var. Anal., **26**(2), 2018.
- RSB, Minh N. Dao, Scott B. Lindstrom, Generalized Bregman envelopes and proximity operators, J. Optim. Theory Appl., **190**(3), 2021.
- RSB, Minh N. Dao, Scott B. Lindstrom, The generalized Bregman distance, SIAM J. Optim., **31**(1), 2021.

University of South Australia

References used in this talk

- RSB, Juan Enrique Martínez Legaz, On Bregman-type distances for convex functions and maximally monotone operators, Set-Valued Var. Anal., **26**(2), 2018.
- RSB, Minh N. Dao, Scott B. Lindstrom, Generalized Bregman envelopes and proximity operators, J. Optim. Theory Appl., **190**(3), 2021.
- RSB, Minh N. Dao, Scott B. Lindstrom, The generalized Bregman distance, SIAM J. Optim., **31**(1), 2021.

References used in this talk

- RSB, Juan Enrique Martínez Legaz, On Bregman-type distances for convex functions and maximally monotone operators, Set-Valued Var. Anal., **26**(2), 2018.
- RSB, Minh N. Dao, Scott B. Lindstrom, Generalized Bregman envelopes and proximity operators, J. Optim. Theory Appl., **190**(3), 2021.
- RSB, Minh N. Dao, Scott B. Lindstrom, The generalized Bregman distance, SIAM J. Optim., **31**(1), 2021.

University of South Australia