Hessian Riemannian Gradient Flows in Convex Programming

Felipe Alvarez, Jérôme Bolte, Olivier Brahic

INTERNATIONAL CONFERENCE ON MODELING AND OPTIMIZATION
MODOPT 2004
Universidad de La Frontera, Temuco, Chile
January 19-22, 2004.

Outline

1. Motivation: scaling the Euclidean gradient.
2. Riemannian gradient flows on convex sets.
3. Hessian metrics, existence, convergence and examples.

1. Motivation: gradient method

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function, $x_0 \in \mathbb{R}^n$.

$$x_{k+1} = x_k + \ \lambda_k \ d_k,$$

1. Motivation: gradient method

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function, $x_0 \in \mathbb{R}^n$.

$$x_{k+1} = x_k + \lambda_k d_k,$$

where

- $\lambda_k > 0$ is a stepsize

1. Motivation: gradient method

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a smooth function, $x_0 \in \mathbb{R}^n$.

$$x_{k+1} = x_k + \lambda_k d_k,$$

where

- $\lambda_k > 0$ is a stepsize

and

- $d_k = -\nabla f(x_k)$ is the steepest descent direction.

1. Motivation: gradient method

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function, $x_0 \in \mathbb{R}^n$.

$$x_{k+1} = x_k + \lambda_k \; d_k,$$

where

- $\lambda_k > 0$ is a stepsize

and

- $d_k = -\nabla f(x_k)$ is the steepest descent direction.

Continuous gradient method: $\quad \dfrac{dx}{dt} = -\nabla f(x), \; t > 0.$

Continuous flow $\leftrightsquigarrow$ discrete method.

1. Scaling and Newton's method

Newton's correction: $x_{k+1} = x_k - \lambda_k \nabla^2 f(x_k)^{-1} \nabla f(x_k).$

Continuous Newton's method:

$$\frac{dx}{dt} = -\nabla^2 f(x)^{-1} \nabla f(x).$$

1. Scaling and Newton's method

Newton's correction: $x_{k+1} = x_k - \lambda_k \nabla^2 f(x_k)^{-1} \nabla f(x_k).$

Continuous Newton's method:

$$\frac{dx}{dt} = -\nabla^2 f(x)^{-1} \nabla f(x).$$

$$\Downarrow$$

$$\frac{d}{dt} \nabla f(x) = -\nabla f(x).$$

1. Scaling and Newton's method

Newton's correction: $x_{k+1} = x_k - \lambda_k \nabla^2 f(x_k)^{-1} \nabla f(x_k)$.

Continuous Newton's method:

$$\frac{dx}{dt} = -\nabla^2 f(x)^{-1} \nabla f(x).$$

$$\Downarrow$$

$$\frac{d}{dt} \nabla f(x) = -\nabla f(x).$$

$$\Downarrow$$

$$\nabla f(x(t)) = e^{-t} \nabla f(x_0)$$

Scale invariant rate of convergence on a straight line !

back.

1. Scaling and constraints

- Problem: $\min\{f(x) \mid x \geq 0\}$.

1. Scaling and constraints

- Problem: $\min\{f(x) \mid x \geq 0\}$.
- ODE approach: $\dfrac{dx_i}{dt} = -x_i \dfrac{\partial f}{\partial x_i}(x), \; i = 1, \ldots, n.$

1. Scaling and constraints

- Problem: $\min\{f(x) \mid x \geq 0\}$.
- ODE approach: $\dfrac{dx_i}{dt} = -x_i \dfrac{\partial f}{\partial x_i}(x), \; i = 1, \ldots, n.$
- Properties:
  - $\frac{d}{dt} f(x) = -\sum_{i=1}^{n} x_i \left| \frac{\partial f}{\partial x_i}(x) \right|^2 \leq 0 \rightsquigarrow$ descent method on $\mathbb{R}^n_+$.
  - $x_i(0) > 0 \Rightarrow \forall t > 0$, $x_i(t) > 0 \rightsquigarrow$ interior point trajectory.

1. Scaling and constraints

- Problem: $\min\{f(x) \mid x \geq 0\}$.
- ODE approach: $$\frac{dx_i}{dt} = -x_i \frac{\partial f}{\partial x_i}(x),\ i = 1, \ldots, n.$$
- Properties:
  - $\frac{d}{dt} f(x) = -\sum_{i=1}^{n} x_i \left| \frac{\partial f}{\partial x_i}(x) \right|^2 \leq 0 \rightsquigarrow$ descent method on $\mathbb{R}^n_+$.
  - $x_i(0) > 0 \Rightarrow \forall t > 0$, $x_i(t) > 0 \rightsquigarrow$ interior point trajectory.
  - The equation may be written as

$$\frac{d}{dt} \log(x_i) = -\frac{\partial f}{\partial x_i}(x)$$

Scaling $\rightsquigarrow$ logarithmic barrier to force $x(t) > 0$!

1. Hessian type scaling

$$\frac{d}{dt}\log(x_i) = -\frac{\partial f}{\partial x_i}(x)$$

1. Hessian type scaling

$$\frac{d}{dt}\log(x_i) = -\frac{\partial f}{\partial x_i}(x)$$

$$\Updownarrow$$

$$\frac{d}{dt}\nabla h(x) = -\nabla f(x),$$

where

$$h(x) = \sum_{i=1}^{n} x_i \log(x_i) - x_i$$

1. Hessian type scaling

$$\frac{d}{dt}\log(x_i) = -\frac{\partial f}{\partial x_i}(x)$$

$$\Updownarrow$$

$$\frac{d}{dt}\nabla h(x) = -\nabla f(x),$$

where

$$h(x) = \sum_{i=1}^{n} x_i \log(x_i) - x_i$$

Thus

$$\frac{dx_i}{dt} = -x_i \frac{\partial f}{\partial x_i}(x) \Leftrightarrow \frac{dx}{dt} = -\nabla^2 h(x)^{-1} \nabla f(x)$$

Remark the analogy with Newton's method

2. Riemannian gradient flows

- Problem: $\min_{x \in C} f(x)$, with $C = \{x \in \mathbb{R}^n \mid x \in \overline{Q},\ Ax = b\}$.
  - $Q \subset \mathbb{R}^n$ nonempty, open and convex.
  - $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ with $m \leq n$.

2. Riemannian gradient flows

- Problem: $\min_{x \in C} f(x)$, with $C = \{x \in \mathbb{R}^n \mid x \in \overline{Q},\ Ax = b\}$.
  - $Q \subset \mathbb{R}^n$ nonempty, open and convex.
  - $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ with $m \leq n$.
- Strategy:
  - Introduce a *Riemannian metric* $H(x) \in \mathbb{S}^n_{++}$ on Q,

$$(u, v)_x = \langle H(x)u, v \rangle = \sum_{i=1}^{n} H_{ij}(x) u_i v_j,\ x \in Q.$$

  - Consider the gradient flow

$$\frac{dx}{dt}(t) = -\nabla_H f(x(t)), t > 0,$$

2. Riemannian gradient

- Let $M, (\cdot,\cdot)$ be a Riemannian manifold.
- T_xM: *tangent space* to M at $x \in M$.

2. Riemannian gradient

- Let M, $(\cdot, \cdot)$ be a Riemannian manifold.
- $T_x M$: *tangent space* to M at $x \in M$.

The gradient $\mathrm{grad} f$ of $f \in \mathcal{C}^1(M; \mathbb{R})$ is uniquely determined by

- Tangency: for all $x \in M$,

$$\mathrm{grad} f(x) \in T_x M.$$

- Duality: for all $x \in M$, $v \in T_x M$,

$$df(x)v = (\mathrm{grad} f(x), v)_x,$$

where $df(x) : T_x M \to \mathbb{R}$ is the differential of f.

2. Riemannian gradient in our case

- $M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$ with Q open set, then

$$\Rightarrow \quad T_x M \simeq \mathrm{Ker}\, A.$$

- $(\cdot,\cdot)_x = \langle H(x)\cdot,\cdot\rangle$ with a barrier/penalty effect near ∂Q.

2. Riemannian gradient in our case

- $M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$ with Q open set, then

$$\Rightarrow \quad T_x M \simeq \text{Ker } A.$$

- $(\cdot, \cdot)_x = \langle H(x)\cdot, \cdot \rangle$ with a barrier/penalty effect near ∂Q.

$$\Downarrow$$

$$\boxed{\nabla_H f = H^{-1}[I - A^T(AH^{-1}A^T)^{-1}AH^{-1}]\nabla f,}$$

where ∇_H stands for grad to stress the dependence on H.

2. Riemannian gradient in our case

- $M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$ with Q open set, then

$$\Rightarrow \quad T_x M \simeq \text{Ker}\, A.$$

- $(\cdot,\cdot)_x = \langle H(x)\cdot,\cdot\rangle$ with a barrier/penalty effect near ∂Q.

$$\Downarrow$$

$$\boxed{\nabla_H f = H^{-1}[I - A^T(AH^{-1}A^T)^{-1}AH^{-1}]\nabla f,}$$

where ∇_H stands for grad to stress the dependence on H.

- Projection $\rightsquigarrow$ vector field in the tangent space.
- Scaling $\rightsquigarrow$ interior point method $x(t) \in Q$.

2. Example

$$C = \Delta_{n-1} := \{x \in \mathbb{R}^n \mid x \geq 0,\ \sum_{i=1}^n x_i = 1\}$$

2. Example

$C = \Delta_{n-1} := \{x \in \mathbb{R}^n \mid x \geq 0, \ \sum_{i=1}^n x_i = 1\}$

- $Q = \mathbb{R}^n_{++}$, $A = (1, \ldots, 1) \in \mathbb{R}^{1 \times n}$ and $b = 1$.
- $M = \{x \in \mathbb{R}^n \mid x > 0, \ \sum_{i=1}^n x_i = 1\}$.
- $T_x M = \{v \in \mathbb{R}^n \mid \sum_{i=1}^n v_i = 0\}$.

2. Example

$C = \Delta_{n-1} := \{x \in \mathbb{R}^n \mid x \geq 0,\ \sum_{i=1}^n x_i = 1\}$

- $Q = \mathbb{R}^n_{++}$, $A = (1, \dots, 1) \in \mathbb{R}^{1\times n}$ and $b = 1$.
- $M = \{x \in \mathbb{R}^n \mid x > 0,\ \sum_{i=1}^n x_i = 1\}$.
- $T_x M = \{v \in \mathbb{R}^n \mid \sum_{i=1}^n v_i = 0\}$.

Take $\boxed{H(x) = \mathrm{diag}(1/x_1, \dots, 1/x_n)}$, then

- $(u, v)_x = \sum_{i=1}^{n} \frac{u_i v_j}{x_i} \rightsquigarrow$ *Shahshahani metric*.
- $\boxed{\frac{dx_i}{dt} = -x_i \frac{\partial f}{\partial x_i} + \sum_{j=1}^{n} x_i x_j \frac{\partial f}{\partial x_j}} \rightsquigarrow$ *Lotka-Volterra type eq.*

Karmarkar '90, Faybusovich '91,...

2. Barrier effect: Legendre functions

We focus on the case $H(x) = \nabla^2 h(x)$, $x \in Q$ with

(H_0)

- $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, convex and proper.
- int dom $h = Q$.
- h is of Legendre type.
- $h_{|_Q} \in \mathcal{C}^2(Q; \mathbb{R})$ and $\nabla^2 h(x) > 0$.
- $\nabla^2 h$ is locally Lipschitz.

2. Barrier effect: Legendre functions

We focus on the case $H(x) = \nabla^2 h(x)$, $x \in Q$ with

(H_0)

- $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, convex and proper.
- $\text{int dom } h = Q$.
- h is of Legendre type.
- $h_{|_Q} \in \mathcal{C}^2(Q; \mathbb{R})$ and $\nabla^2 h(x) > 0$.
- $\nabla^2 h$ is locally Lipschitz.

$\star$ h is strictly convex and $\mathcal{C}^1$ on $\text{int dom } h$.

$\star$ $\text{int dom } h \ni x^j \to x \in \partial \text{int dom } h$, $\|\nabla h(x^j)\| \to +\infty$.

back

2. Example of Legendre function

$$h(x) = \sum_{i \in I} \theta(g_i(x)).$$

where

- $I = \{1, \ldots, p\}$, $g_i \in \mathcal{C}^3(\mathbb{R}^n)$ concave.
- $Q = \{x \in \mathbb{R}^n \mid g_i(x) > 0, i \in I\} \neq \emptyset$.
- $\forall x \in Q, \text{ span } \{\nabla g_i(x) \,|\, i \in I\} = \mathbb{R}^n$,

and

$$\begin{cases} \text{(i) } (0,\infty) \subseteq \text{dom}\theta \subseteq [0,\infty), \; \theta \in \mathcal{C}^3(0,\infty). \\ \text{(ii) } \lim_{s \to 0^+} \theta'(s) = -\infty \text{ and } \forall s > 0, \; \theta''(s) > 0. \\ \text{(iii) Either } \theta \text{ is nonincreasing or } \forall i \in I, g_i \text{ is affine.} \end{cases}$$

3. Questions

Given $x_0 \in M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$:

$$\frac{dx}{dt}(t) = -\nabla_H f(x(t)), t > 0,$$

3. Questions

Given $x_0 \in M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$:

$$\frac{dx}{dt}(t) = -\nabla_H f(x(t)), t > 0,$$

- Well posedness: global existence for all $t > 0$.
- Asymptotic behavior: convergence to an equilibrium as $t \to \infty$, rate of convergence,...

3. Questions

Given $x_0 \in M = Q \cap \{x \in \mathbb{R}^n \mid Ax = b\}$:

$$\frac{dx}{dt}(t) = -\nabla_H f(x(t)), t > 0,$$

- Well posedness: global existence for all $t > 0$.
- Asymptotic behavior: convergence to an equilibrium as $t \to \infty$, rate of convergence,...

Main difficulty: singular behavior near ∂Q.

$\Rightarrow$ *Classical results do not apply.*

3. Well posedness: global existence

Thm. 2 The trajectory $x(t)$ is defined for all $t \geq 0$ under any of the following conditions:

(C_1) $\{x \in C \mid f(x) \leq f(x_0)\}$ is bounded.

$$(C_2) \quad \begin{cases} (i)\ \mathrm{dom}\, h = \overline{Q} \\ (ii)\ \forall y \in \overline{Q}, \forall \gamma \in \mathbb{R}, \{x \in C \mid D_h(y,x) \leq \gamma\} \text{ is bdd.} \\ (iii)\ \mathrm{Argmin}_C\, f \neq \emptyset \text{ and } f \text{ quasiconvex.} \end{cases}$$

(C_3) $\exists K \geq 0$, $L \in \mathbb{R}$ such that

$$\forall x \in Q,\ ||H(x)^{-1}|| \leq K|x| + L.$$

3. Why Hessian metrics ?

Suppose f is convex and $A = 0$ and $b = 0$.

Euclidean case: $y \in \operatorname{Argmin}_C f \Leftrightarrow \forall x \in C,\ \langle \nabla f(x), x - y \rangle \geq 0.$

3. Why Hessian metrics ?

Suppose f is convex and $A = 0$ and $b = 0$.

Euclidean case: $y \in \text{Argmin}_C f \Leftrightarrow \forall x \in C,\ \langle \nabla f(x), x - y \rangle \geq 0.$

$$\varphi_y(x) = \frac{1}{2}\|x - y\|^2$$

$$\Downarrow$$

$$\dot{x} = -\nabla f(x) \Rightarrow \frac{d}{dt}\varphi_y(x) = \langle \nabla \varphi_y(x), \dot{x} \rangle = \langle x - y, -\nabla f(x) \rangle \leq 0.$$

3. Why Hessian metrics ?

Suppose f is convex and $A = 0$ and $b = 0$.

Euclidean case: $y \in \operatorname{Argmin}_C f \Leftrightarrow \forall x \in C,\ \langle \nabla f(x), x - y \rangle \geq 0.$

$$\boxed{\varphi_y(x) = \tfrac{1}{2}\|x - y\|^2}$$

$$\Downarrow$$

$$\dot{x} = -\nabla f(x) \Rightarrow \frac{d}{dt}\varphi_y(x) = \langle \nabla \varphi_y(x), \dot{x} \rangle = \langle x - y, -\nabla f(x) \rangle \leq 0.$$

$$\Downarrow$$

$\varphi_y(x)$ is a *Lyapunov function* for the gradient flow

3. Characterization of Hessian metrics

Riemannian case:

$$y \in \mathrm{Argmin}_{\overline{Q}}\, f \Leftrightarrow \forall x \in Q,\ (\nabla_H f(x), x - y)_x \geq 0.$$

3. Characterization of Hessian metrics

Riemannian case:

$$y \in \mathrm{Argmin}_{\overline{Q}}\, f \Leftrightarrow \forall x \in Q,\ (\nabla_H f(x), x - y)_x \geq 0.$$

Thm. 1 $H \in \mathcal{C}^1(Q; \mathbb{S}^n_{++})$ *satisfies*

$$\forall y \in Q,\ \exists \varphi_y \in \mathcal{C}^1(Q; \mathbb{R}),\ \nabla_H \varphi_y(x) = x - y$$

$$\Updownarrow$$

$$?$$

3. Characterization of Hessian metrics

Riemannian case:

$$y \in \mathrm{Argmin}_{\overline{Q}}\, f \Leftrightarrow \forall x \in Q,\ (\nabla_H f(x), x - y)_x \geq 0.$$

Thm. 1 $H \in \mathcal{C}^1(Q; \mathbb{S}^n_{++})$ *satisfies*

$$\forall y \in Q,\ \exists \varphi_y \in \mathcal{C}^1(Q; \mathbb{R}),\ \nabla_H \varphi_y(x) = x - y$$

$$\Updownarrow$$

$\exists h \in \mathcal{C}^3(Q)$ *such that* $H = \nabla^2 h$ *on* Q *and*

$$\begin{aligned} \varphi_y(x) &= D_h(y, x) = h(y) - h(x) - \langle \nabla h(x), y - x \rangle \\ &= \text{Bregman pseudo-distance induced by } h. \end{aligned}$$

3. Implicit proximal iteration

$$x_{k+1} \in \text{Argmin}\left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x_k) \mid Ax = b \right\},$$

$$\Updownarrow$$

$$\frac{1}{\lambda_k}[\nabla h(x_{k+1}) - \nabla h(x_k)] \in -\nabla f(x_{k+1}) + \text{Im } A^T, \ Ax_{k+1} = b$$

Bregman 67, Censor-Zenios 92, Teboulle 92, Eckstein 93, Kiwiel 97,...

3. Implicit proximal iteration

$$x_{k+1} \in \text{Argmin}\left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x_k) \mid Ax = b \right\},$$

$$\Updownarrow$$

$$\frac{1}{\lambda_k}[\nabla h(x_{k+1}) - \nabla h(x_k)] \in -\nabla f(x_{k+1}) + \text{Im } A^T, \;\; Ax_{k+1} = b$$

Bregman 67, Censor-Zenios 92, Teboulle 92, Eckstein 93, Kiwiel 97,...

But $$\frac{dx}{dt} = -\nabla_H f(x) \Leftrightarrow \left\{ \begin{array}{rcl} \frac{d}{dt}\nabla h(x) & \in & -\nabla f(x) + \text{Im } A^T, \\ Ax(t) & = & b, \; t \geq 0. \end{array} \right.$$

$\star$ This link was already noticed by Iusem-Svaiter-Da Cruz Neto ’99, together with convergence results for a *linear* objective function.

3. Convergence: Bregman functions

A Legendre function h with $\mathrm{dom} h = \overline{Q}$ is of *Bregman type* if

(i) $\{x \in Q \mid D_h(y,x) \leq \gamma\}$ is bdd. $\forall y \in \overline{Q},\ \forall \gamma \in \mathbb{R}$.

(ii) $\forall y \in \overline{Q}$, $\forall y^j \to y$ with $y^j \in Q$, $D_h(y, y^j) \to 0$.

3. Convergence: Bregman functions

A Legendre function h with $\mathrm{dom} h = \overline{Q}$ is of *Bregman type* if
(i) $\{x \in Q \mid D_h(y,x) \le \gamma\}$ is bdd. $\forall y \in \overline{Q}$, $\forall \gamma \in \mathbb{R}$.
(ii) $\forall y \in \overline{Q}$, $\forall y^j \to y$ with $y^j \in Q$, $D_h(y, y^j) \to 0$.

Thm. 3 *Suppose*

- (H_0) *with* h *of Bregman type.*
- f *is quasiconvex and* $\mathrm{Argmin}_C\, f \neq \emptyset$.

Then $\exists x_\infty \in C$ *such that* $x(t) \to x_\infty$ *as* $t \to +\infty$ *with*

$$-\nabla f(x_\infty) \in N_{\overline{Q}}(x_\infty) + \mathrm{Ker}\, A^\perp,$$

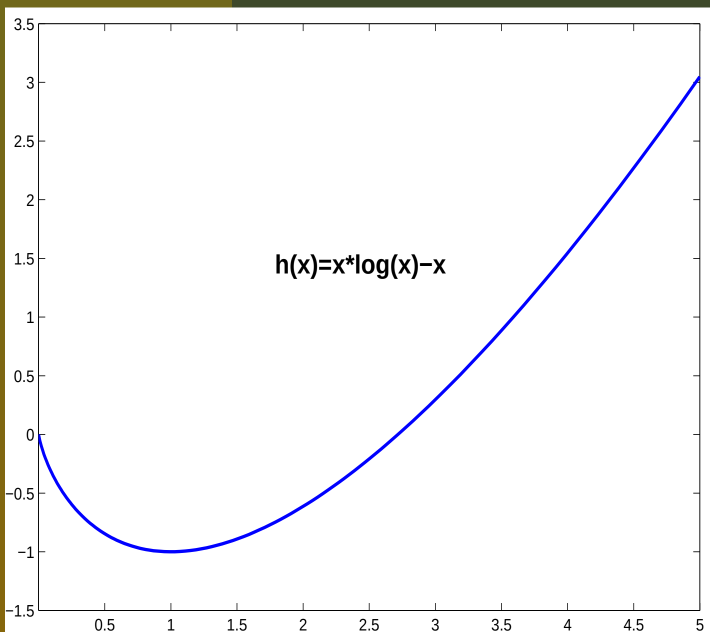
where $N_{\overline{Q}}(x_\infty)$ *is the normal cone to* $\overline{Q}$ *at* x_∞.

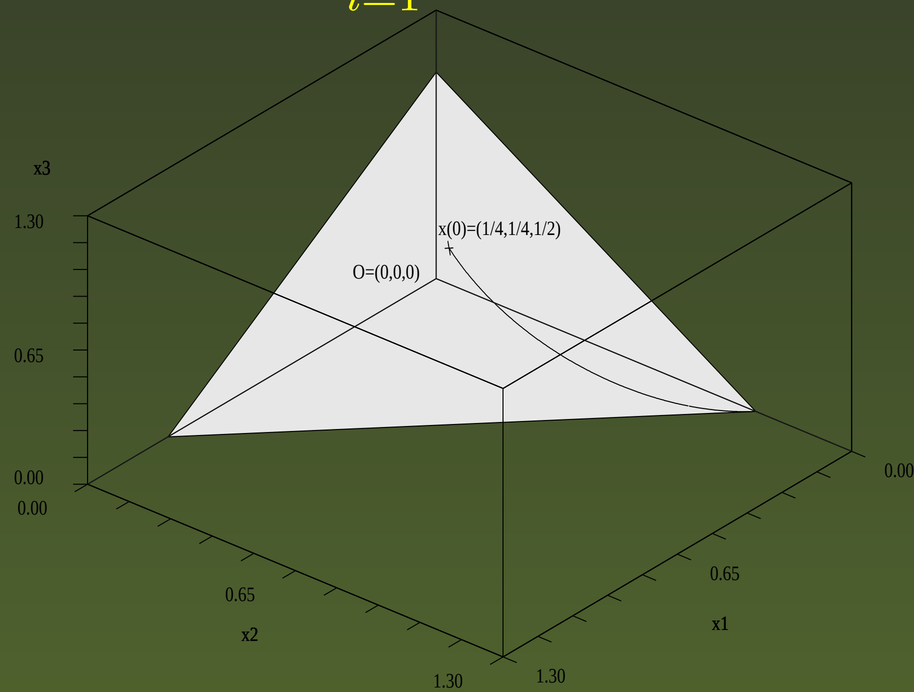
3. Examples on Δ_{n-1}

Boltzmann-Shanon entropy: $h(x) = \sum_{i=1}^{n} x_i \log(x_i) - x_i.$

Shahshahani metric: $H(x) = \nabla^2 h(x) = \text{diag}(1/x_1, \dots, 1/x_n).$

Kullback-Liebler divergence: $D_h(y, x) = \sum_{i=1}^{n} y_i \log(y_i/x_i) + x_i - y_i.$





Lotka-Volterra type flow

3. Other examples

$$h(x) = -2\sum_{i=1}^{n}\sqrt{x_i}$$

- $H(x) = \nabla^2 h(x) = \frac{1}{2}\mathrm{diag}(1/x_1^{3/2}, \ldots, 1/x_n^{3/2})$.
- $D_h(y,x) = \sum_{i=1}^{n}(\sqrt{y_i} - \sqrt{x_i})/\sqrt{x_i}$.
- *Flow*: $\frac{dx_i}{dt} = -2x_i^{3/2}\left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{n}\frac{x_j^{3/2}}{\sum_{k=1}^{n}x_k^{3/2}}\frac{\partial f}{\partial x_j}\right)$.

3. Other examples

$$h(x) = -2\sum_{i=1}^n \sqrt{x_i}$$

- $H(x) = \nabla^2 h(x) = \frac{1}{2}\mathrm{diag}(1/x_1^{3/2}, \dots, 1/x_n^{3/2})$.
- $D_h(y,x) = \sum_{i=1}^n (\sqrt{y_i} - \sqrt{x_i})/\sqrt{x_i}$.
- *Flow:* $\frac{dx_i}{dt} = -2x_i^{3/2}\left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^n \frac{x_j^{3/2}}{\sum_{k=1}^n x_k^{3/2}} \frac{\partial f}{\partial x_j}\right)$.

$h(x) = -\sum_{i=1}^n \log(x_i)$ ($h(0) = +\infty$ so that h is not Bregman).

- $H(x) = \nabla^2 h(x) = \mathrm{diag}(1/x_1^2, \dots, 1/x_n^2)$.
- $D_h(y,x) = \sum_{i=1}^n \log(x_i/y_i) + (y_i - x_i)/x_i$.
- *Flow:* $\frac{dx_i}{dt} = -x_i^2\left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^n \frac{x_j^2}{\sum_{k=1}^n x_k^2} \frac{\partial f}{\partial x_j}\right)$.

3. Further developments

- *Rate of convergence.*
- *Dual trajectory and its convergence.*
- *Geodesic type characterization of trajectories.*
- *Connections with completely integrable Hamiltonian systems.*

Reference: A.-Bolte-Brahic, to appear in SIAM J. on Control Optim.

3. Further developments

- *Rate of convergence.*
- *Dual trajectory and its convergence.*
- *Geodesic type characterization of trajectories.*
- *Connections with completely integrable Hamiltonian systems.*

Reference: A.-Bolte-Brahic, to appear in SIAM J. on Control Optim.

Continuous version of similar results for proximal iterations: Iusem-Monteiro '00.

3. Further developments

- *Rate of convergence.*
- *Dual trajectory and its convergence.*
- *Geodesic type characterization of trajectories.*
- *Connections with completely integrable Hamiltonian systems.*

Reference: A.-Bolte-Brahic, to appear in SIAM J. on Control Optim.

→ Continuous version of similar results for proximal iterations: Iusem-Monteiro ’00.

→ Generalizations of results on the log-metric in linear programming: Bayer-Lagarias ’89.

4. Duality when $Q = \mathbb{R}^n_{++}$

$$(P) \qquad \min\{f(x) \mid x \geq 0,\ Ax = b\}$$

Assume:

- f is convex and $S(P) \neq \emptyset$.
- $\exists x_0 \in \mathbb{R}^n,\ x_0 > 0,\ Ax_0 = b.$

Dual problem:

$$(D) \qquad \min\{p(\lambda) \mid \lambda \geq 0\}$$

where $p(\lambda) = \sup\{\langle \lambda, x\rangle - f(x) \mid Ax = b\}$. Then

$$S(D) = \{\lambda \in \mathbb{R}^n \mid \lambda \geq 0,\ \lambda \in \nabla f(x^*) + \operatorname{Im} A^T, \langle \lambda, x^* \rangle = 0\},$$

where x^* is any solution of (P).

4. Dual trajectory

Integrating the differential inclusion

$$\frac{d}{dt}\nabla h(x(t)) \in -\nabla f(x(t)) + \operatorname{Im} A^T,$$

we obtain

$$\lambda(t) \in c(t) + \operatorname{Im} A^T,$$

where $c(t) = \frac{1}{t}\int_0^t \nabla f(x(\tau))d\tau$ and

$$\lambda(t) = \frac{1}{t}[\nabla h(x^0) - \nabla h(x(t))].$$

If $h(x) = \sum_{i=1}^{n} \theta(x_i)$, then $\lambda_i(t) = \frac{1}{t}[\theta'(x_i^0) - \theta'(x_i(t))]$.

4. Dual penalty scheme

We have: $\lambda(t) = \frac{1}{t}[\nabla h(x^0) - \nabla h(x(t))]$.

But

$$h \text{ is Legendre } \Rightarrow \nabla h^{-1} = \nabla h^*,$$

with $h^*(\lambda) = \sum_{i=1}^n \theta^*(\lambda_i)$ being the *Fenchel conjugate* of h.

Hence $x(t) = \nabla h^*(\nabla h(x^0) - t\lambda(t))$, where

Take $A\widetilde{x} = b$. Since $Ax(t) = b$, we have

$$\widetilde{x} - \nabla h^*(\nabla h(x^0) - t\lambda(t)) \in \mathrm{Ker}\, A.$$

Then, $\lambda(t)$ solves

$$\min_{\lambda} \left\{ \langle \widetilde{x}, \lambda \rangle + \frac{1}{t} \sum_{i=1}^n \theta^*(\theta'(x_i^0) - t\lambda_i) \,\middle|\, \lambda \in c(t) + \mathrm{Im} A^T \right\}$$

4. Dual trajectory: convergence

Example: $\theta(s) = s\log(s) - s \Rightarrow \theta^*(s^*) = \exp(s^*)$, $s^* \in \mathbb{R}$.

Then

$$\min_{\lambda} \left\{ \langle \widetilde{x}, \lambda \rangle + \frac{1}{t} \sum_{i=1}^{n} x_i^0 \exp(-t\lambda_i) \middle| \lambda \in c(t) + \mathrm{Im} A^T \right\}$$

Convergence:

- $f(x) = \langle c, x \rangle \Rightarrow c(t) = \frac{1}{t}\int_0^t \nabla f(x(\tau))d\tau \equiv c$

$\Rightarrow$ by Cominetti-San Martin ´96, Auslender et al. ´97, Cominetti '00,... convergence to the θ^*-center of $S(D)$.

- Otherwise, $x(t)$ bounded $\Rightarrow \nabla f(x(t)) \to \nabla f(x^*)$ for $x^* \in S(P)$

$\Rightarrow c(t) \to \nabla f(x^*) \Rightarrow$ convergence by Iusem-Monteiro '00.