Doubly autoparallel structure and curvature integrals

- An application to iteration-complexity analysis of convex optimization -

Atsumi Ohara　University of Fukui

Collaborators:
H. Ishi (Osaka Metropolitan Univ.),
T. Tsuchiya (GRIPS),
K. Uohashi (Tohoku Gakuin Univ.)

統計多様体の幾何学とその周辺(14)
November 2-4, 2022 @北大

Outline

- Introduction: Doubly autoparallel submanifolds
- Preliminaries
  - Dually flat structure on a symmetric cones
- Characterization of DA submfd in sym. cones
  - Several applications
- Conic linear program on convex cones Ω
  - Central trajectory
  - Geometric predictor-corrector method in Ω*
- Curvature integral and iteration-complexity
- Application to primal-dual path following methods
- Concluding remark

Introduction

Doubly autoparallel submanifolds

Def. Statistical manifold: $(\mathcal{S}, g, \nabla, \nabla^*)$

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

X, Y and Z : arbitrary vector fields on $\mathcal{S}$

★ g : Riemannian metric

★ (∇, ∇^*) : torsion-free affine connections

$R^{\nabla} = 0, \; R^{\nabla^*} = 0$ ➡ dually flat

★ $\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^*$: α-connections

- <u>Def.</u> Let (S, g, ∇, ∇^*) be a statistical manifold and M be its submanifold. We call M a doubly autoparallel submanifold in S when the followings hold:

  - $\forall X, Y \in \mathcal{X}(M),\ \nabla_X Y \in \mathcal{X}(M)$

    i.e. $H_M(X, Y) = 0$

  - $\forall X, Y \in \mathcal{X}(M),\ \nabla^*_X Y \in \mathcal{X}(M)$

    i.e. $H^*_M(X, Y) = 0$

Important Properties

<u>Proposition</u> The following statements are equivalent:

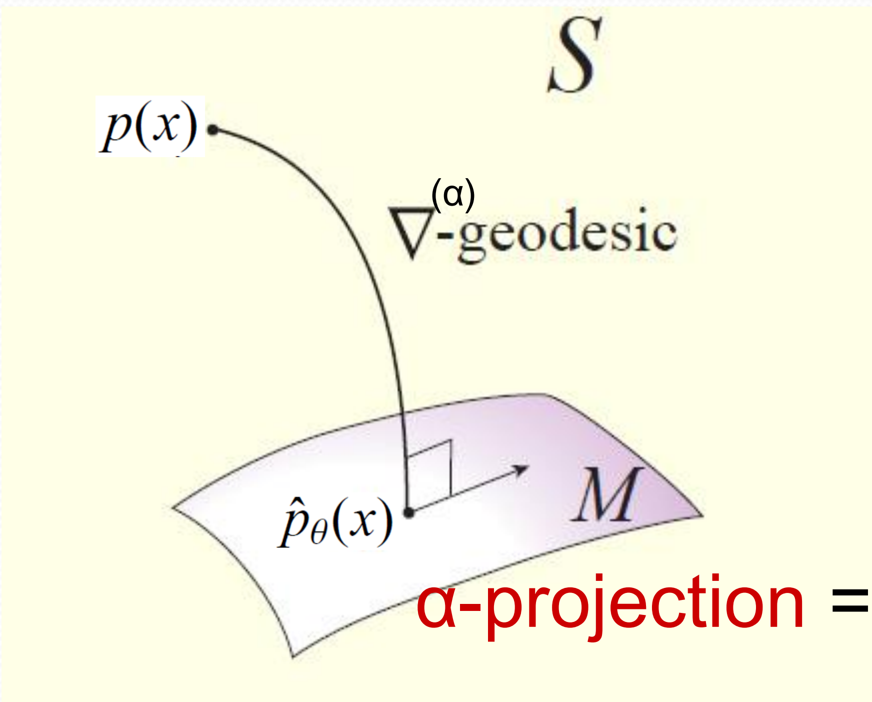
- 1) A submanifold M is doubly autoparallel (DA)
- 2) M is autoparallel w.r.t. the α-connections

  $$\nabla^{(\alpha)} = \{(1+\alpha)\nabla + (1-\alpha)\nabla^*\}/2$$

  for two different α's.
- 3) M is autoparallel w.r.t. all the α-connections.
- 4) all the α-geodesics connecting two points on M lay in M (if it is simply connected).
- 5) M is affinely constrained in both ∇- and ∇^*-affine coordinates if S is dually flat.

Furthermore, for a parametric statistical model S

- If M is DA in S, then α-projections (q-MaxEnt) from p to M are unique for all α if they exist.



M is simultaneously an exp. and mix. family

α-projection =arg min $D^{(\alpha)}(p, M)$

Related topics and applications

Symmetric cones

- MLE for structured covariance matrices is tractable (cast to convex program: inversely linear structure)
  [Anderson 70, Malley 94]
- Explicitly solvable Semi-Definite Programs [O 99]
- Structure of α-power means on symmetric cones [O 04]

Related topics and applications

Probability simplex

- Statistical models Markov-isomorphic to the probability simplex [Nagaoka 17]
- Characterization and classification of DA submfds in prob. simplex via Hadamard algebra [O&Ishi 18]
- Learning theory [Mutus&Ay 03]

Miscellaneous

- The self-similar (*Barenblatt–Pattle*) solution for the porous medium equation [O&Wada 10]

General statistical manifolds

- Purely geometric study [Satoh *et al*. 21]

Preliminaries

[Faraut&Korani 94]

- Symmetric cone Ω in an Euclidian space E
  - Homogeneous

    $G(\Omega) = \{\tau \in GL(E) \mid \tau(\Omega) = \Omega\}$ acts transitively
  - self-dual w.r.t. an inner product of E

    $\Omega = \Omega^*, \qquad \Omega^* = \{y \in E \mid (x|y) > 0, \forall x \in \overline{\Omega} \setminus \{0\}\}$
- Euclidean Jordan algebra $(V, *)$
  - Commutative
  - $x^2 * (x * y) = x * (x^2 * y), \quad$ where $x^2 = x * x$
  - Associative inner-product $(x * y|z) = (y|x * z)$

Prop. $\Omega = \mathrm{int}\{x^2 \mid x \in V\}$ is a symmetric cone in V.

Ex. the set of real symmetric p. d. matrices $PD(n, \mathbf{R})$

$$V = \mathrm{Sym}(n; \mathbf{R}), \qquad X * Y = (XY + YX)/2$$

$$\tau_G(X) = GXG^T, \qquad G \in GL(n, \mathbf{R})$$

$$(X|Y) = \mathrm{tr}(XY),$$ the unit: I, the inverse: X^{-1}

- $L(x) : V \to V, \quad L(x)y = x * y$
- $P(x, y) := L(x)L(y) + L(y)L(x) - L(x * y)$
- Mutation: $x \perp_a y := P(x, y)a$

  isomorphic to $*$, the unit element: a^{-1}

Ex. $X \perp_A Y = (XAY + YAX)/2$

Preliminaries (Dually flat structure on Ω)

- Logarithmic characteristic function on Ω

$$\psi(x) := \log \int_{\Omega^*} e^{-\langle s,x \rangle} ds,$$

  - positive definite Hessian on Ω
  - $x^{-1} = -\mathrm{grad}\,\psi(x)$, $\qquad (\mathrm{grad}\, f(x)|u) = D_u f(x)$
- Ex. $\psi(x) = -\sum_{i=1}^{n} \log x^i$ on $\mathbf{R}^n_{++}$, $\quad \psi(x) = -\log\det X$ on $\mathrm{PD}(n)$

- a coordinate system (x^i) $\quad x = \sum_{i=1}^{n} x^i e_i$, $\{e_i\}_{i=1}^n$: a basis of E
- a dual coordinate system (s_i)

$x^{-1} = \sum_{i=1}^n s_i e^i$, $\quad \{e^i\}_{i=1}^n$: a basis of E with $(e^i|e_j) = -\delta^i_j$

- D : the canonical flat affine connection on E
- $\{x^1, \cdots, x^n\}$: affine coordinate system, i.e., $D_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = 0$
- g : Riemannian metric on Ω

$$g = Dd\psi = \sum_{i,j} \frac{\partial^2 \psi}{\partial x^i \partial x^j} dx^i dx^j .$$

- D' : the dual affine connection on Ω

$$Xg(Y, Z) = g(D_X Y, Z) + g(Y, D'_X Z)$$

(g, D, D') : dually flat structure on Ω

Pleriminaries and ex. on PD(n)

Dually flat structure on Ω [Uohashi&O 04]

- Potential: –log det x,

  Ex. -log det X, ($X = \sum_{i=1}^{N} x^i E_i$)

- Riemanian metric: $g_x\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = (P(x)^{-1} e_i | e_j)$, $P(x) := P(x, x)$

$$g_X\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \mathrm{tr}(X^{-1} E_i X^{-1} E_j)$$

- α-connections: $\left(\nabla^{(\alpha)}_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}\right)_x = (\alpha - 1)(e_i \perp_{x^{-1}} e_j)$

$$\left(\nabla^{(\alpha)}_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}\right)_X = \frac{\alpha - 1}{2}(E_i X^{-1} E_j + E_j X^{-1} E_i)$$

Characterization of DA submfds in Ω

Let W be a linear subspace in Jordan algebra $(V, *)$ and $p = q * q$ in Ω.

<u>Thm.</u> [OIT] The following 1)-3) are equivalent:

1) A Submanifold $M = (W + p) \cap \Omega$ is DA, where

$$W + p = \{w + p \mid w \in W\}$$

2) For all x in M, $u \perp_{x^{-1}} v \in W, \ (u, v \in W)$

3) The subspace $P(q)^{-1}W$ is a Jordan subalgebra.

<u>Rem.</u> (a) 3) is able to be checked at the single point p

(b) $M = \{(W' + p^{-1}) \cap \Omega\}^{-1}$ with $W' = P(p)^{-1}W$

The proof is based on 5) in the Proposition

(c) Implication: Classification of DA submflds in Ω reduces to that of Jordan subalgs of $(V, *)$.
(For V=Sym(n,**R**) →[Jacobson 87], [Malley 87])

- Ex. - Jordan subalgebras in Sym(n,**R**)
  1) fixed eigen vectors, 2) doubly symmetric, etc.
  - Two bases $\{E_i\}_{i=1}^m$ and $\{F^i\}_{i=1}^m$ of Sym(n,**R**)

$$\mathcal{M} = \{P \mid P = E_0 + \sum_{i=1}^{m} x^i E_i,\ \exists x = (x^i) \in \mathbf{R}^m\} \cap \mathrm{PD}(n)$$

$$\mathcal{M} = \{P \mid P^{-1} = F^0 + \sum_{i=1}^{m} s_i F^i,\ \exists s = (s_i) \in \mathbf{R}^m\} \cap \mathrm{PD}(n).$$

Application(1) Means on Positive Operators

[Kubo & Ando 80]

- <u>Def.</u> (Axioms of means)

  σ is a mean on self-conjugate positive operators

  - i) $A \leq C,\ B \leq D \Rightarrow A\sigma B \leq C\sigma D$
  - ii) $C(A\sigma B)C = (CAC)\sigma(CBC)$
  - iii) $A_n \downarrow A,\ B_n \downarrow B \Rightarrow A_n\sigma B_n \downarrow A\sigma B$

    where $A_n \downarrow A \overset{\text{def}}{\Leftrightarrow} (A_i \geq A_{i+1}, \forall i)\ \&\ (A_n \to A)$
  - iv) $I\sigma I = I$

α-geodesics on PD(*n*)

- α-geodesic $P(s)$ boundary conds. : $P(0)=A$, $P(1)=B$

$$P^{(\alpha)}(s) = A^{1/2}\left\{[(A^{-1/2}BA^{-1/2})^{\alpha} - I]s + I\right\}^{1/\alpha} A^{1/2}$$

$$\alpha = 1 \qquad P(s) = A + s(B - A)$$

$$\alpha = 0 \qquad \hat{P}(s) = A^{1/2}\exp(s \log A^{-1/2}BA^{-1/2})A^{1/2}$$

$$\alpha = -1 \qquad P^{*}(s) = \{A^{-1} + s(B^{-1} - A^{-1})\}^{-1}$$

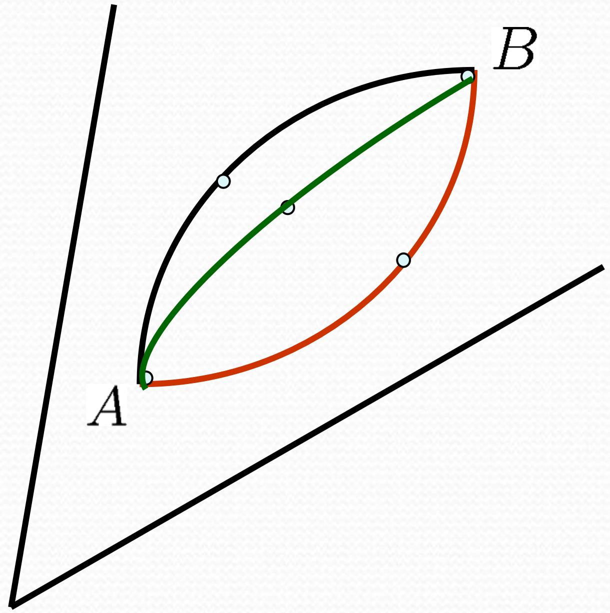
$$\begin{aligned} A\mathrm{a}B &:= P(1/2) \\ A\mathrm{g}B &:= \hat{P}(1/2) \\ A\mathrm{h}B &:= P^{*}(1/2) \end{aligned}$$

$P^{(\alpha)}(1/2)$: a power mean

Means and α-geodesics on PD(n) [O 04]

Thm. Points on α-geodesics for s in [0,1] and α in [-1,1] are 2-param. family of means, i.e.,

$$A\sigma_s^{(\alpha)}B = P^{(\alpha)}(s)$$



In particular, for fixed s in [0, 1]

$$P^{(\alpha)}(s) > P^{(\beta)}(s),$$
$$1 \geq \alpha > \beta \geq -1$$ AGH ineq. (s=1/2)

Cor. A and B are in a DA submanifold M

$$\Rightarrow \quad A\sigma_s^{(\alpha)}B \in M, s \in [0,1], \alpha \in [-1,1]$$

App.(2) MLE for structured covariance matrices

- Sample covariance S in $\mathrm{PD}(n,\mathbf{R})$
- a zero-mean Gaussian p.d.f. with covariance mtx. Σ

$$p(x) = (2\pi)^{-n/2}(\det\Sigma)^{-1/2}\exp\{-\tfrac{1}{2}x^T\Sigma^{-1}x\}$$

- structured covariance mtx. (with linear constraints)

$$\Sigma \in \mathcal{M} = (E_0 + \mathcal{W}) \cap \mathrm{PD}(n, \mathbf{R})$$

- <u>Ex.</u>
  - Toeplitz matrices: $\{T = (t_{ij}) | t_{ij} = t_{ji} = y_{|i-j|}\}$
  - zero-patterns : $\{\Sigma = (\sigma_{ij}) | \sigma_{ij} = \sigma_{ji} = 0,\ (i,j) \in \mathcal{E}\}$
  - etc…

MLE for structured covariance matrices

- Negative logarithmic likelihood func (up to const.):

$$\ell(\Sigma) := -\log\det\Sigma^{-1} + \mathrm{tr}(\Sigma^{-1}S) \rightarrow \min$$

  - Rem Note that $-\log\det$ is a convex function.

- If $\mathcal{M}$ is DA (inversely linear structure), then the minimization problem of $\ell(\Sigma)$ (MLE) s.t. $\Sigma \in \mathcal{M}$ is a strictly convex program.

  ⇒ Unique solution,
  Numerically tractable (optimality eq. is linear)

App.(3) Convex program

Affine-scaling method and IG

- General convex program: Convex set $\mathcal{M} \subset \mathbf{R}^n$, $c \in \mathbf{R}^n$

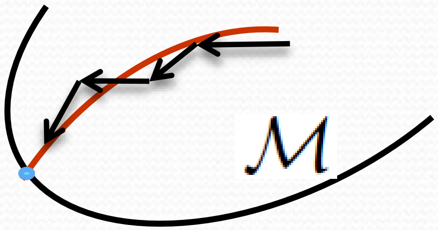
$$\text{minimize } c^T x, \quad \text{s.t. } x \in \overline{\mathcal{M}}$$

- Ψ: a *good* convex barrier func. for $\mathcal{M}$,

  1) $\Psi(x) \to +\infty$ ($x \to \text{bd}\,\mathcal{M}$), 2) h: p.d. Hessian, 3) + α

- Gradient flow for Riemannian mfd $(\mathcal{M}, h)$

$$\dot{x} = \frac{dx}{dt} = -h(x)^{-1} c, \quad x(0) \in \mathcal{M}$$



$x(t)$: affine-scaling trajectory

(numerically traced)

- Legendre transform ⇒ linearized

$$\dot{s} = -c, \quad s_i = \frac{\partial \Psi}{\partial x^i}, \; i = 1, \ldots, n, \quad \widehat{s} := -\lim_{t \to +\infty} ct + s(0)$$

- Opt. sol.: $\widehat{x} = \mathrm{grad}\Psi^*(\widehat{s})$ (inverse Legendre trans.)
- Red underlined: needs the explicit form of Ψ^*

  (or solving the nonlinear eq.: $\hat{s} = \mathrm{grad}\Psi(\hat{x})$)

Idea

Ψ^* is known for a good barrier $\Psi \Rightarrow$ an explicit opt. sol. $\widehat{x}$

- 1) Ω: sym. cones $\Rightarrow$ $\psi(x)$ = -log det x, $\psi^*(s)$=-log det s,

  Legendre transform: $x \mapsto s = x^{-1}$
- 2) $\mathcal{M}$ realized by $\mathcal{M} = (a+W)\cap\Omega$ is DA in Ω

  $\Rightarrow$a) convexity of $\mathcal{M}$, b) linearized traj. belongs to $\mathcal{M}$

- Ex. SemiDefinite Program (SDP)

$$\underset{P}{\text{minimize}}\,(C|P), \text{ s.t. } P = E_0 + \sum_{i=1}^{m} x^i E_i \in \overline{\mathcal{M}} = \overline{(E_0 + \mathcal{W}) \cap \mathrm{PD}(n)}$$

- If $\mathcal{M}$ is DA in PD(n) and $P \in \mathcal{M}$
  - 1. Set $F^0 = P^{-1}, F^i = -P^{-1} E_i F^{-1}$, then

$$\mathcal{M} = \{P \,|\, P^{-1} = F^0 + \sum_{i=1}^{m} s_i F^i,\ \exists s = (s_i) \in \mathbf{R}^m\} \cap \mathrm{PD}(n)$$

  - 2. Solve $\widetilde{C} \in \mathrm{span}\{F^i\}_{i=1}^{m}$ meeting

$$\forall P \in \mathcal{M}, \quad (C|P) = (\widetilde{C}|P) + \mathrm{const.}$$

- 3. Spectral decomposition

$$\widetilde{C} = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & O \\ O & O \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = V_1 \Sigma_1 V_1^T$$

- 4. For $\forall P_0 \in \mathcal{M}$ with $S_0 = P_0^{-1}$, the opt. sol. Is

$$\widehat{P} = \lim_{t\to\infty} S(t)^{-1} = \lim_{t\to\infty} (-\widetilde{C}t + S_0)^{-1} = P_0 - P_0 V_1 (V_1^T P_0 V_1)^{-1} V_1^T P_0$$

Rem. Independent of the objective function $(C|P)$ and an initial value P_0

Interior point method (IP) for Conic linear program

Conic linear program -Notation-

- Vector space E of dimension *n* $E \ni x$
- The dual vector space E^* $E^* \ni s$
- $\langle s, x \rangle$: Paring
- Ω : proper open convex cone in E
- Ω^* : the dual cone of Ω

$$\Omega^* := \{s \in E^* | \langle s, x \rangle > 0, \forall x \in \overline{\Omega} \setminus \{0\}\}$$

- T^*: (Orthogonal) dual subspace of $T \subset E$

$$T^* = \{s \in E^* | \langle s, x \rangle = 0, \ \forall x \in T\}$$

Conic Linear Program

Given

$$c \in E^*, \; f \in E \text{ and } T \subset E$$

- Primal problem

(P)
$$\text{minimize } \langle c, x \rangle, \text{ s.t. } x \in \overline{\mathcal{P}}, \quad \text{where } \mathcal{P} := (f + T) \cap \Omega,$$

- Dual problem

(D)
$$\text{maximize } \langle s, f \rangle, \text{ s.t. } s \in \overline{\mathcal{D}}, \quad \text{where } \mathcal{D} := (c + T^*) \cap \Omega^*.$$

Typical Examples

- Linear program (LP):

$$E = E^* = \mathbf{R}^n, \ \Omega = \Omega^* = \mathbf{R}^n_{++}$$

- Semidefinite program (SDP):

  $E = E^*$: the set of real symmetric matrices

  $\Omega = \Omega^*$: the set of positive definite matrices

- Second order cone (Lorentz cone) program (SOCP)
- Mixture of the aboves

ϑ-normal barrier on an open convex cone Ω

- <u>Def.</u> θ-normal barrier ψ on Ω (⇐+α)
  - A (smooth) convex function ψ satisfying, at each x in Ω,

$$\psi(tx) = \psi(x) - \vartheta \log t,$$
$$|(D^2 d\psi)_x(X, X, X)| \leq 2((Dd\psi)_x(X, X))^{3/2}$$

  for $\vartheta \geq 1$, $\forall t > 0$ and $\forall X \in T_x\Omega \cong E$
  - $\psi(x) \to +\infty$ ($x \to$ bd Ω),

<u>Rem.</u> [Nesterov & Nemirovski 94] (1) Existence for all Ω (but not explicit forms), (2) the Hessian is p.d., (3) self-concordance (⇒the Newton method is efficient).

- Ex. $\psi(x) = -\sum_{i=1}^{n} \log x^i$ (LP), $\psi(x) = -\log\det X$ (SDP)

Dually flat structure on Ω (revisited)

- D : the canonical flat affine connection on E
- $\{x^1, \cdots, x^n\}$: affine coordinate system, i.e., $D_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = 0$
- g : Riemannian metric on Ω

$$g = Dd\psi = \sum_{i,j} \frac{\partial^2 \psi}{\partial x^i \partial x^j} dx^i dx^j .$$

- D' : the dual affine connection on Ω

$$Xg(Y,Z) = g(D_X Y, Z) + g(Y, D'_X Z)$$

(g, D, D') : dually flat structure on Ω

Remark

- $\{s_1, \cdots, s_n\}$: dual coordinate system on E^* , s.t.

$$\langle s, x\rangle = \sum_i s_i(s)x^i(x)$$

- Gradient map $\iota : \Omega \to \Omega^*$ defined by

$$s_i \circ \iota = -\frac{\partial\psi}{\partial x^i}$$

induces dually flat structure on Ω^* from (g, D, D')

(1) D^* : the canonical flat affine connection on E^*

$$D^*_{\iota_*(X)}\iota_*(Y) = \iota_*(D'_X Y) \qquad (\iota^* D^* = D')$$

D^* -autoparallel in Ω^* ⟺ D' -autoparallel in Ω

Remark

(2) Riemannian metric $g^* := D^* d\psi^*$ on Ω^*

$$g = \iota^* g^*$$

(3) $\langle \iota_*(X), Y \rangle = -g_x(X, Y)$

Hessian norm : We denote the length of *X* in $T_x\Omega \cong E$ by

$$\|X\|_x := \|Z\|_s := \sqrt{g_x(X, X)} = \sqrt{g^*_s(Z, Z)},$$

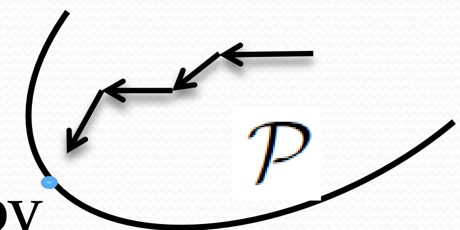
where $s = \iota(x)$ and $Z = \iota_*(X)$.

Curvature integral and iteration-complexity of IP

One of important computational performance indices for optimization algorithms is the iteration-complexity.

- Ω: sym. cone and $\mathcal{P} := (f + T) \cap \Omega$ is DA

  $\Rightarrow$ iteration-complexity=o for (P)



- General case? Iter.-comp. is characterized by
  - Curvature integrals along the central trajectory $\gamma_{\mathcal{P}}$

$$\int_{t_1}^{t_2} \|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}(t), \dot{\gamma}_{\mathcal{P}}(t))\|^{1/2}_{\gamma_{\mathcal{P}}(t)} dt$$

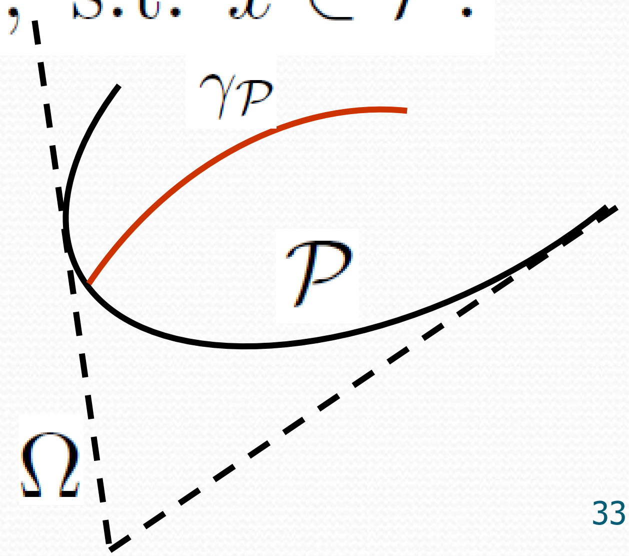
  - Similarly, for (D) curvature integrals along the dual c. t. $\gamma_{\mathcal{D}}$

$$\int_{t_1}^{t_2} \|H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}}(t), \dot{\gamma}_{\mathcal{D}}(t))\|^{1/2}_{\gamma_{\mathcal{D}}(t)} dt$$

Central trajectory $\gamma_{\mathcal{P}}$

- Primal problem: $\text{minimize } \langle c, x \rangle, \text{ s.t. } x \in \overline{\mathcal{P}},$
  $\text{where } \mathcal{P} := (f + T) \cap \Omega,$

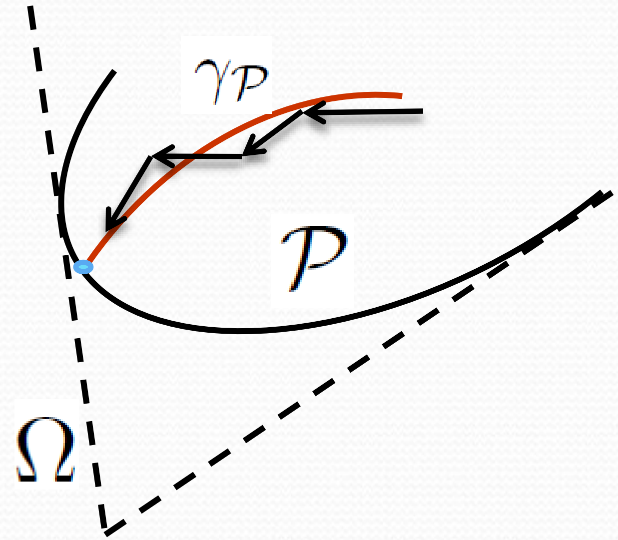
- $x(t) := \gamma_{\mathcal{P}}(t)$: the unique minimizer of
  $$\text{minimize } \underline{\underline{t}}\langle c, x \rangle + \psi(x), \text{ s.t. } x \in \overline{\mathcal{P}}.$$
  for each $t > 0$
- $\gamma_{\mathcal{P}} := \{\gamma_{\mathcal{P}}(t) | t > 0\}$:
  (Primal) central trajectory



Central trajectory

- Homotopy path to the opt. sol. of the primal problem, i.e., $x(t)$ converges when $t \to \infty$.
- Numerically tracing $\gamma_{\mathcal{P}}$ is the standard and efficient way to solve the primal problem.

Path-following method



Idea: consider the problem in Ω^* and relate the complexity with the curvature

(1)Representation of feasible region

- A linear surj. operator $A : E \to \mathbf{R}^m$ s.t. Ker A =T

$$\mathcal{P} = \{x \in \Omega | Ax = b\},$$
$$\mathcal{D} = \{s \in \Omega^* | s = c - A^* y,\ y \in \mathbf{R}^m\}$$

where $A^* : \mathbf{R}^m \to E^*$ satisfying $y^T(Ax) = \langle A^* y, x \rangle$,

$b := Af \in \mathbf{R}^m$

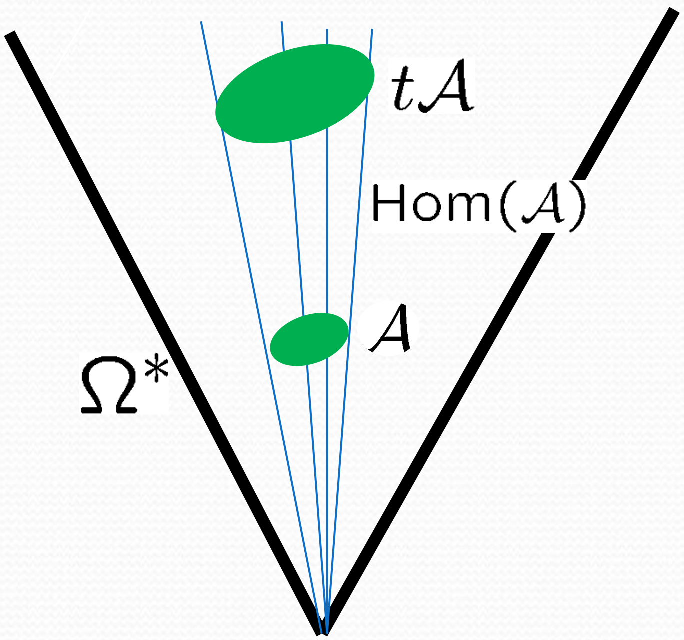
- dim $\mathcal{P}$ =n-m, dim $\mathcal{D}$ =m
- $\mathcal{P}$ is D-autoparallel and $\mathcal{D}$ is D^*-autoparallel

(2)Homogenization (conic hull)

- homogenization of $\mathcal{D}$ in Ω^*

$$\mathrm{Hom}(\mathcal{D}) := \bigcup_{t>0} t\mathcal{D},\ \ t\mathcal{D} := \{s \in \Omega^* | s = t\tilde{s},\ \ \tilde{s} \in \mathcal{D}\}$$

  - D^*-autoparallel because $\mathcal{D}$ is.
  - dim $\mathrm{Hom}(\mathcal{D})$=$m$+1



Homogenization

Lemma

The following relations hold in Ω^*:

$$\iota(\gamma_{\mathcal{P}}) = \iota(\mathcal{P}) \cap \mathrm{Hom}(\mathcal{D})$$

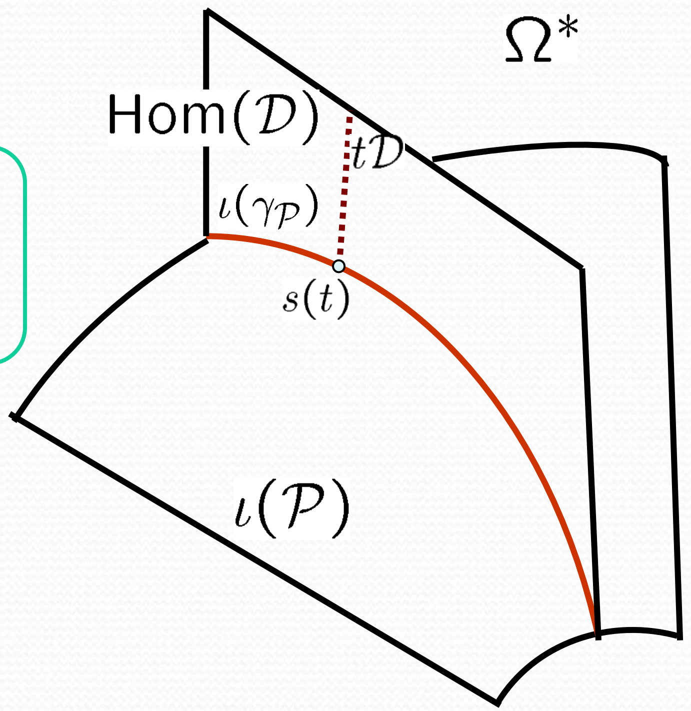
$$s(t) := \iota(x(t)) = \iota(\mathcal{P}) \cap t\mathcal{D}$$

$$L(x,y) := t\langle c, x\rangle + \psi(x) + y^T(b - Ax)$$

$$\partial L/\partial x = 0 \ \rightarrow s \in t\mathcal{D}$$

Remark

$\iota(\mathcal{P})$ and $t\mathcal{D}$ are orthogonal w.r.t. g^* at $s(t)$ by definition.



3. Geometric predictor-corrector algorithm (tracing $\gamma_{\mathcal{P}}$ in $\mathrm{Hom}(\mathcal{D})$)

Ideal case

- Predictor

  From $s(t) \in \iota(\gamma_{\mathcal{P}})$

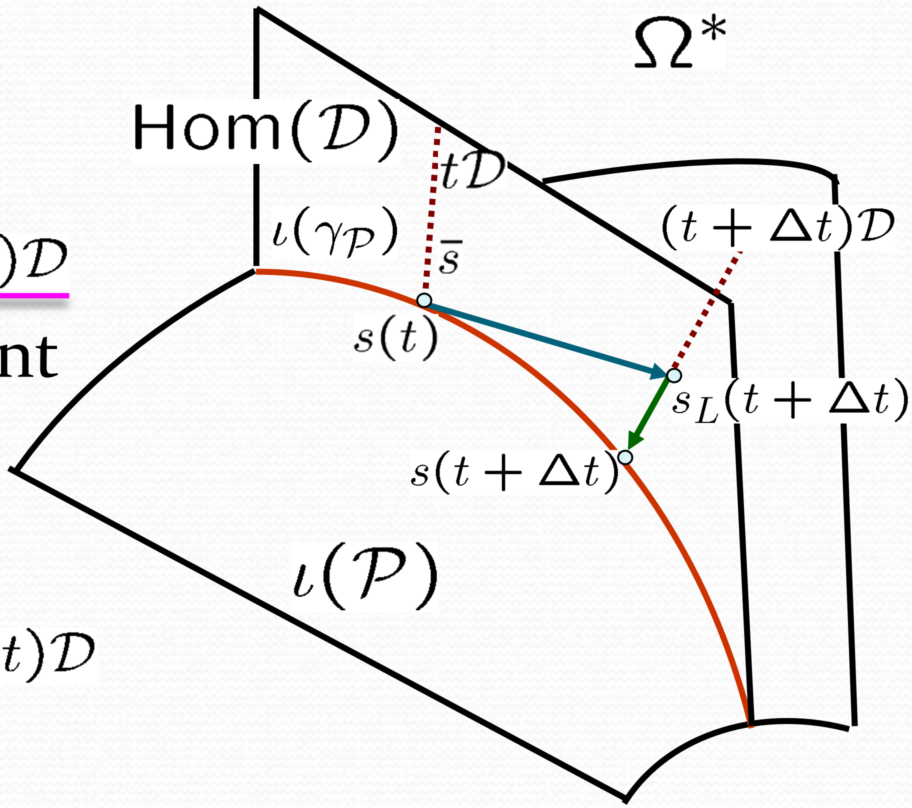
  to $s_L(t+\Delta t) \in (t+\Delta t)\mathcal{D}$

  with the direction tangent to $\iota(\gamma_{\mathcal{P}})$

- Corrector

  From $s_L(t+\Delta t) \in (t+\Delta t)\mathcal{D}$

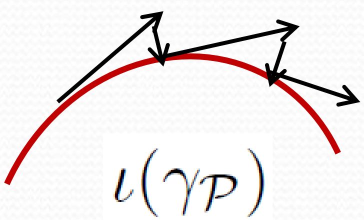
  to $s(t+\Delta t) \in \iota(\gamma_{\mathcal{P}})$

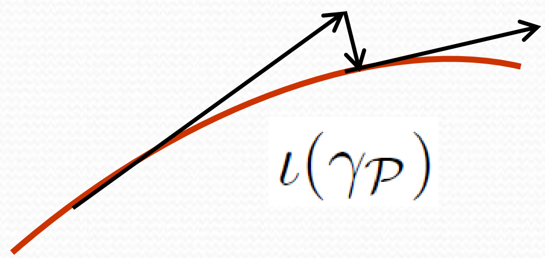


Intuitive observation

- $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t), \dot\gamma_{\mathcal{P}}(t))$: the Euler-Schouten embedding curvature (second fundamental form) of $\iota(\gamma_{\mathcal{P}})$ with respect to D^*

- If $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t), \dot\gamma_{\mathcal{P}}(t))$ is small at t, so is expected the iteration number !?



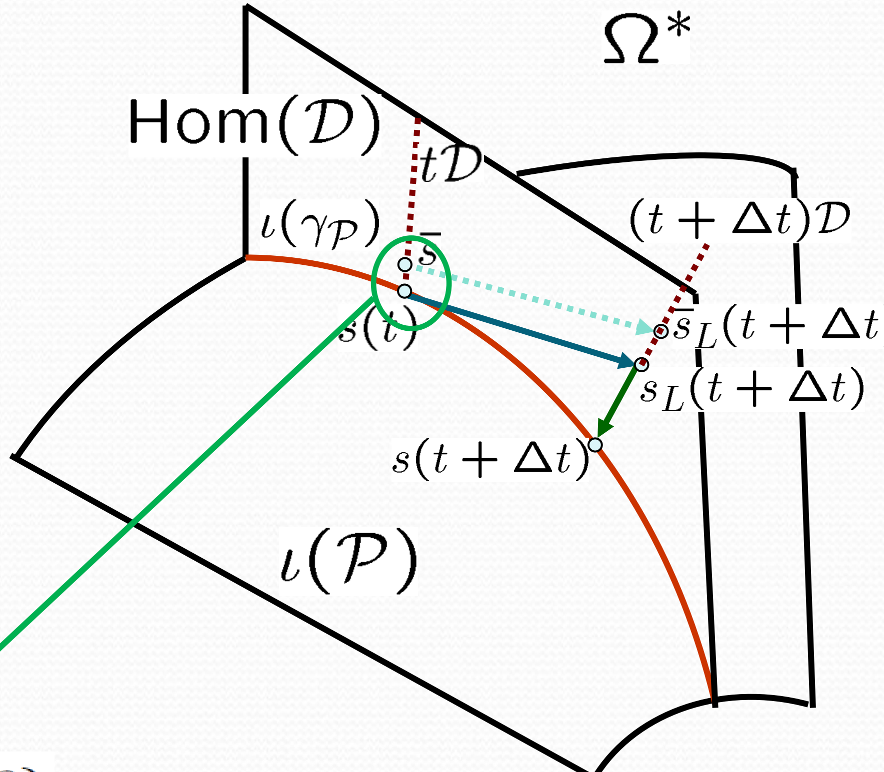


Actually,

$$\ddot s = D^*_{\dot s}\dot s = \iota_*(H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}}))$$

Remark: practical case

- Cannot expect that the corrector returns precisely on $\iota(\gamma_{\mathcal{P}})$
- Consider the point $\bar{s}$ in the neighborhood of $s(t) \in \iota(\gamma_{\mathcal{P}})$ in the sense of Riemannian metric



$$\mathcal{N}_t(\beta) := \{s \in t\mathcal{D} | \delta(s) \leq \beta\}$$

Predictor

- The differential equation expressing $\iota(\gamma_{\mathcal{P}})$:

$$\dot{s} = (\text{id} - \Pi_s^{\perp})c = \frac{1}{t}(\text{id} - \Pi_s^{\perp})s$$

where $\Pi_s^{\perp}$ is the orthogonal projection w.r.t. g^* from E^* to $T^* = \text{Range}A^*$ at *s*.

Note: $\Pi_s^{\perp}$ =0 ⇒ ODE for the A-S traj. (up to sign)

- Hence, the predictor is defined by

$$\bar{s}_L(t + \Delta t) := \bar{s} + \Delta t(I - \Pi_{\bar{s}}^{\perp})c \in (t + \Delta t)\mathcal{D}$$

Corrector

- Reduces to the following convex optimization on $t\overline{\mathcal{D}}$:

$$\text{minimize } F(s) := \langle s, f \rangle + \psi^*(s), \ \text{s.t. } s \in t\overline{\mathcal{D}}$$

- Newton direction *N* for this opt. problem:

$$D^* dF(X, N) = -dF(X), \ \forall X \in \mathcal{X}(t\mathcal{D})$$

  - Newton decrement: measure of approximation of *s*

$$\delta(s) := \|N\|_s$$

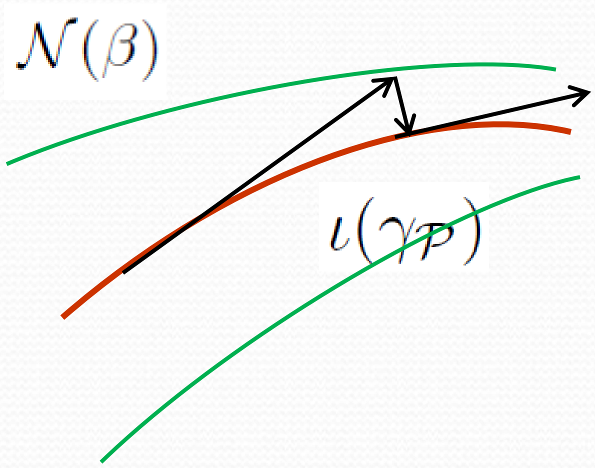
- We define the corrector with a single Newton step by:

$$\bar{s}_L^+(t + \Delta t) := \bar{s}_L(t + \Delta t) + N_{\bar{s}_L(t+\Delta t)}$$

Tubular neighborhood

- The standard analysis technique in IP ensures the polynomiality of the complexity for this path-following strategy if all the generated points are near to $\iota(\gamma_{\mathcal{P}})$.

- Introduce the tubular neighborhood $\mathcal{N}(\beta)$ of $\iota(\gamma_{\mathcal{P}})$



$$\mathcal{N}(\beta) := \bigcup_{t\in(0,\infty)} \mathcal{N}_t(\beta),$$

where $\mathcal{N}_t(\beta) := \{s \in t\mathcal{D} | \delta(s) \leq \beta\}$.

4. Curvature integral and asymptotic iteration-complexity (Main result)

- Assumption: $\iota(\gamma_{\mathcal{P}})$ is not D^*-autoparallel, i.e.,

  $\beta \rightarrow 0$ implies that $\Delta t \rightarrow 0$ (If it is ?)

- Theorem

  For $0 < t_1 < t_2$ and $s_1 \in \mathcal{N}(\beta) \cap t_1\mathcal{D}$, let $\sharp(s_1, t_2, \beta)$ be the iteration number to find $s_2 \in \mathcal{N}(\beta) \cap t_2\mathcal{D}$. Then,

$$\lim_{\beta \to 0} \frac{\sqrt{\beta} \times \sharp(s_1, t_2, \beta)}{I_{\mathcal{P}}(t_1, t_2)} = 1,$$

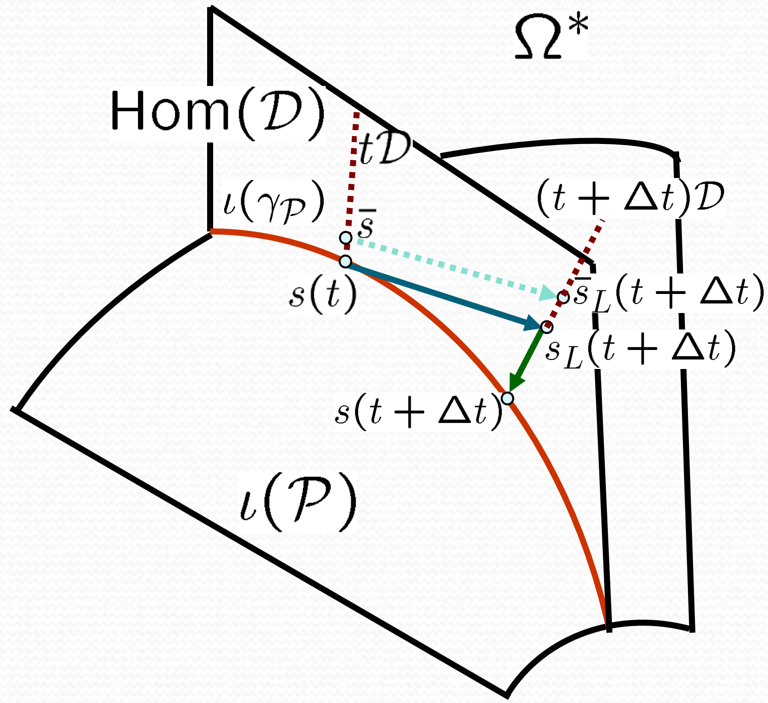
where

$$I_{\mathcal{P}}(t_1, t_2) := \frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \| H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}(t), \dot{\gamma}_{\mathcal{P}}(t)) \|^{1/2}_{\gamma_{\mathcal{P}}(t)} dt.$$

Outline of the proof

- Evaluate the Newton dec. of the predictor $\bar{s}_L(t+\Delta t)$ by $\|\ddot{s}(t)\|_{s(t)}$ (For each iteration)

$$\begin{aligned}
&\delta(\bar{s}_L(t+\Delta t)) \\
&= \|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_{\bar{s}_L(t+\Delta t)} + r_4 \\
&= \frac{(\Delta t)^2}{2} \underline{\|\ddot{s}(t)\|}_{s(t)} + \delta(\bar{s}) + r_1 + r_2 + r_3,
\end{aligned}$$



Outline of the proof

- Intermediate two relations for sufficiently small Δt and β.

(For each iteration)

- $$\sqrt{(1-\eta)\beta}(1-O(\sqrt{\beta})) \le \sqrt{w} - \sqrt{M_3}\delta(\bar{s}) \le \frac{\Delta t}{\sqrt{2}} \|\ddot{s}(t)\|_{s(t)}^{1/2} + \sqrt{|r_1|} + \sqrt{M_3}(\Delta t)^2,$$
- $$\frac{\Delta t}{\sqrt{2}} \|\ddot{s}(t)\|_{s(t)}^{1/2} - \sqrt{|r_1|} - \sqrt{M_3}(\Delta t)^2 \le \sqrt{w} + \sqrt{M_3}\delta(\bar{s}) \le \sqrt{\beta}(1+O(\sqrt{\beta}))$$

Outline of the proof

- Take summations of iterations
  - $$\sqrt{(1-\eta)\beta} \sum_{k=1}^{\sharp(s_1,t_2,\beta)} (1 - O(\sqrt{\beta})) \le \frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \|\ddot{s}(t)\|_{s(t)}^{1/2}\, dt + M'\sqrt{\Delta t_{\max}},$$
  - $$\frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \|\ddot{s}(t)\|_{s(t)}^{1/2}\, dt - M'\sqrt{\Delta t_{\max}} \le \sqrt{\beta} \sum_{k=1}^{\sharp(s_1,t_2,\beta)} (1 + O(\sqrt{\beta}))$$

$$\ddot{s} = D_{\dot{s}}^{*}\dot{s} = \iota_{*}(H_{\mathcal{P}}^{*}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}}))$$

Remark

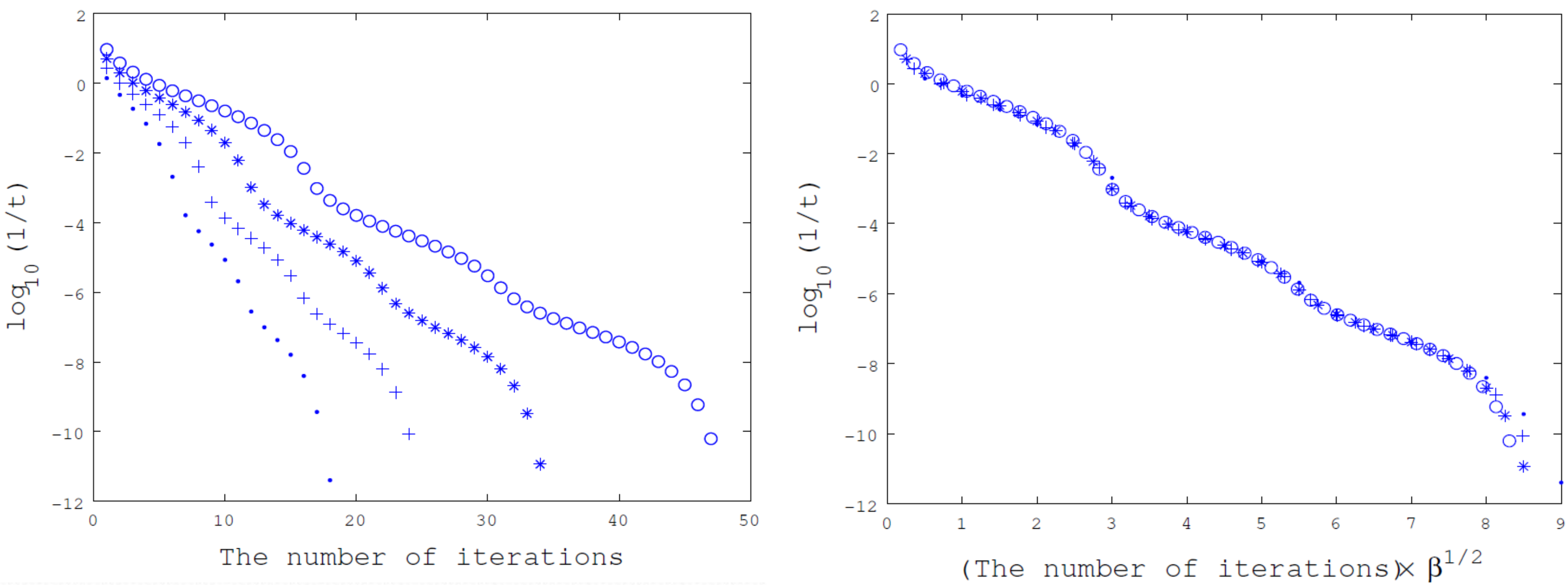
- An asymptotic result for $\beta \to 0$ (and hence, $\Delta t \to 0$)
  - $\mathcal{P}$ is DA $\Rightarrow$ $\iota(\gamma_{\mathcal{P}})$ is DA (D^*-autoparallel) $\Rightarrow$ $\Delta t \to \infty$ $\Rightarrow$ explicit sol.

- The same argument holds for the dual problem.

- The results are valid for general convex cones

Numerical experiment

- Curvature structure of CT for a certain LP



$$(\cdot : \beta = 1/4,\ + : \beta = 1/8,\ * : \beta = 1/16,\ \circ : \beta = 1/32)$$

Curved part is Straight and Straight part is Curved?(1)

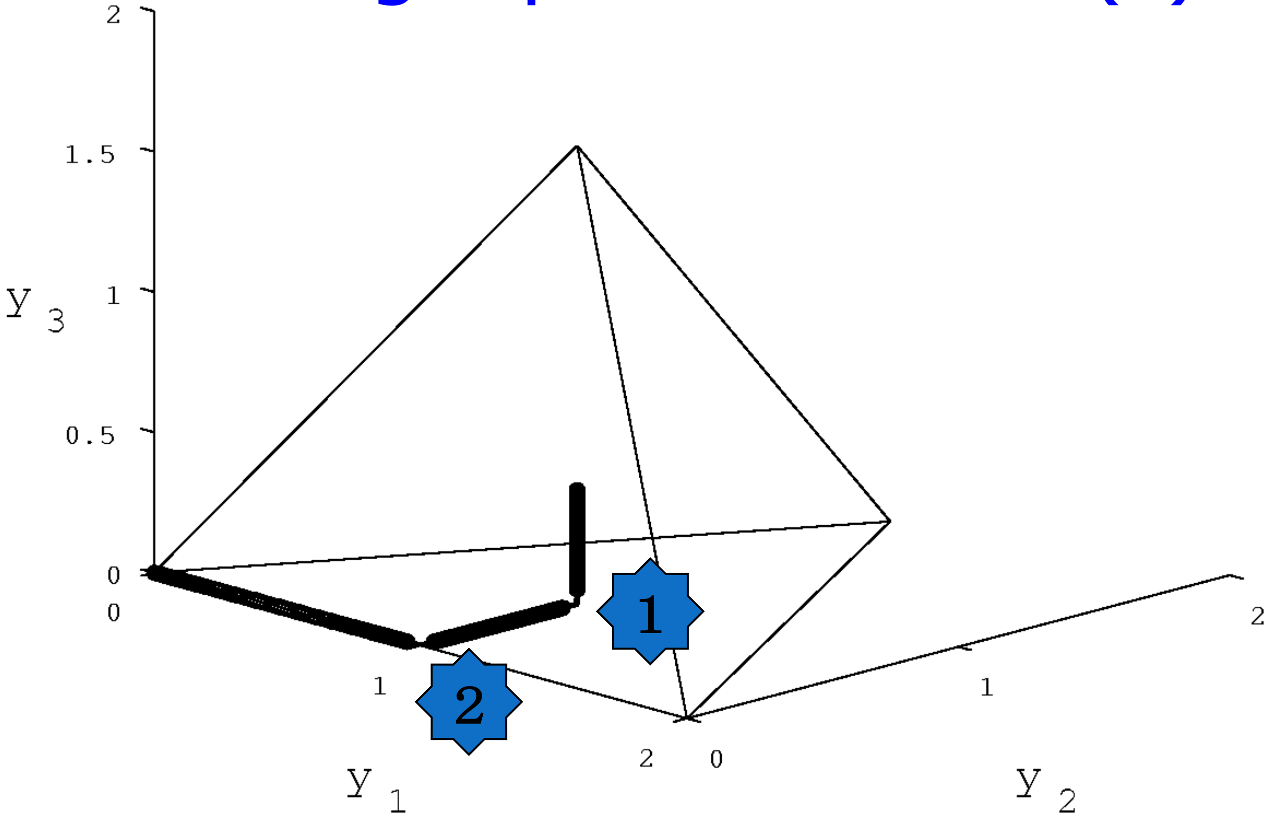


Figure 7: Exaple Figure

Proposition

It holds that $\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2}_{\gamma_{\mathcal{P}}(t)} \leq \frac{\sqrt{2\vartheta}}{t}$

ϑ : a constant determined by $\psi(x)$

- Remark

  The above proposition gives the upper bound:

$$I_{\mathcal{P}}(t_1, t_2) \leq \sqrt{\vartheta} \log(t_2/t_1)$$

Further study for LP case

- Primal and Dual Linear Program:

$$\min \; c^T x$$

$$\text{s.t.} \quad Ax = b, \quad x \geq 0, \; A \in R^{m \times n}, b \in R^n$$

$$\max \; b^T y$$

$$\text{s.t.} \quad s = c - A^T y, \quad s \geq 0,$$

Application to Primal-dual path-following (PDPF) method

- current main-stream IP (cheap in each iteration)
- The following quantity has been known to play an important and similar role in complexity analysis of PDPF method:

$$I_{PD}(t_1, t_2) = \int_{t_1}^{t_2} h_{PD}(t)^{1/2} dt$$

where $h_{PD}(t)$ is given by

$$h_{PD}(t) := \frac{1}{t^2}((I_n - Q(t))e) * (Q(t)e).$$

e: the unit element of Jordan product *

$Q(t)$: a certain projection matrix

Proposition

It holds that

$$h_{PD}(t)^2 = \left(\frac{1}{2}\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}(t), \dot{\gamma}_{\mathcal{P}}(t))\|_{\gamma_{\mathcal{P}}(t)}\right)^2 + \left(\frac{1}{2}\|H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}}(t), \dot{\gamma}_{\mathcal{D}}(t))\|_{\gamma_{\mathcal{D}}(t)}\right)^2$$

Remark :

- geometric implication of the quantity of $I_{PD}(t_1, t_2)$

- inequalities $\max\{I_{\mathcal{P}}(t_1, t_2), I_{\mathcal{D}}(t_1, t_2)\} \leq I_{PD}(t_1, t_2) \leq I_{\mathcal{P}}(t_1, t_2) + I_{\mathcal{D}}(t_1, t_2).$

Concluding Remark

- Tractable characterization of DA submfds in symmetric cones Ω
- Application to conic linear programs
  - Explicit sol. when the feasible region M is DA in Ω.
    - M is DA $\Rightarrow$ AS (CT) traj. is DA (D^*-autoparallel) $\Rightarrow$ $\Delta t \to \infty$ $\Rightarrow$ explicit sol.
- Extension: # of iterations and curvature integral of CT
  - Asymptotic analysis ($\beta \to 0$)
    - Complemented by numerical experiment for finite β
  - Geometric structure of CT has a influence on complexity of the IP algorithm

- Relation among iteration-complexities of P. D. and PD algorithm.
- DA submanifolds in a certain submfd in Jordan algebras [OIT]

- Future work: Geometrical study for general stat. mfd.
  - Various geometrical concepts for mutually dual connections and their characterizations (Furuhata *et al.*)
  - Classifications
  - Families of continuous probability densities
  - Applications (Ex. Study of ODE's on manifolds?)

Thank you for your attention

References:

[OIT] A. Ohara, H. Ishi and T. Tsuchiya,

Doubly autoparallel structure and curvature integrals

- An application to iteration-complexity analysis of convex optimization -, Information Geometry, to appear.