Information Geometry and Interior-Point Algorithms in SDP and Symmetric Cone Programs

Satoshi Kakihara * Atsumi Ohara † Takashi Tsuchiya‡

September 30, 2011

Abstract

This paper is a continuation of the paper [3] by the authors where they demonstrated that the number of iterations of Mizuno-Todd-Ye predictor-corrector primal-dual interior-point methods for SDP and more generally symmetric cone programs is (asymptotically) expressed with an integral over the central trajectory called "curvature integral." It was shown that the number of the iterations of the algorithm is approximated surprisingly well with the integral for fairly large problems of LP and SDP with thousands of variables. In this paper, we demonstrate that the curvature integral admits a rigorous differential geometric expression in view of information geometry. Our result is a direct extension of [10] in linear programming to SDP and symmetric cone programs, using the results for curvature integrals of primal-dual algorithms in SDP and symmetric cone programs in [3]. Together with the numerical evidence in [3], we claim that *the number of iterations of the interior-point algorithm is expressed as a differential geometric quantity.*

Keywords: interior-point methods, primal/dual algorithms, primal-dual algorithms, iteration complexities, curvature integral, semidefinite programming, symmetric cone programs, information geometry

1 Introduction

Interplay between interior-point methods and differential geometry is an interesting topic studied by several authors. It was shown in [3] that the substantial portion of the iteration complexity of Mizuno-Todd-Ye predictor-corrector primal-dual interior-point method for SDP and symmetric cone programs is represented with an integral along the central trajectory called "curvature integral." The curvature integral is introduced by Sonnevend et al. for LP and further studied by Monteiro and Tsuchiya, and recently extended to symmetric cone programming including SDP. The objective of this paper is to establish a rigorous differential geometric expression for this curvature integral in view of information geometry. This is done by exploiting a geometric structure of the problems, called a dually flat space, and relating the primal and dual problems each other by Legendre transformation. In application to convex programming, by focusing on the duality of the constraint cones, the primal space and the dual space are naturally associated with an identical manifold through Legendre transformation. As a result, we can link their curvature integrals together through a certain projection of curvature integrals of primal-dual algorithms to ones of primal/dual algorithms.

Curvature integrals have been playing an important role in the complexity analysis of interior-point methods. The concepts of curvature integrals was firstly introduced by Karmarkar [4] in the context of Riemannian geometry in special LP problems, and then developed by Sonnevend, Stoer and Zhao [12]

*National Graduate Institute for Policy Studies 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan. (Email: s-kakihara@grips.ac.jp).

†Electorical and Electronics Engineering, Graduate School of Engineering, University of Fukui, 3-9-1 Bunkyo, Fukui City, Fukui 910-8507, Japan. (Email: ohara@fuee.u-fukui.ac.jp).

‡National Graduate Institute for Policy Studies 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan. (Email: tsuchiya@grips.ac.jp). This author is supported in part with Grant-in-Aid for Scientific Research (B), 2007, 19360192 from the Japan Society for the Promotion of Sciences.

and Zhao and Stoer [13], who gave a concrete formula for LP and gave a lower and upper bound of its iteration complexity. Curvature integrals are defined along the central path, reflecting the geometric structure of the central path; they become large in curved parts of the central path and become nearly zero in approximately straight parts. Path-following methods, which are of our concern in this paper, trace the central path approximately to seek an optimal solution. Thus intuitively it is considered that curvature integrals represent the iteration complexities of path-following algorithms. In fact, besides [12, 13], several researches also have been done in regards to these subjects: for primal-dual algorithms [5] deals with this in LP case and [3] deals with this in SDP and symmetric cone programs cases, and for primal/dual algorithms [10] deal with this in a general conic linear program.

Information geometry is a differential geometric framework specifically tailored to deal with convexity naturally arising in information science, initially developed by Amari and Nagaoka [8] to study the probability theory and statistics. In [10], Ohara and Tsuchiya developed an information geometric framework of interior-point methods for conic linear programs based on the grand work Nesterov and Nemirovski [9]. See also [3] where they focus on information geometric framework and primal-dual interior-point algorithms for LP providing extensive numerical experiments. According to this framework, we can consider the central path and subsequent curvature integrals of primal and dual problems simultaneously in the same manifold and thus associate iteration complexities of interior-point algorithms to an information geometric integral over the central trajectory involving both problems.

To sum up, the objective of this paper, to put it more precisely, is to geometrically relate the curvature integrals of primal-dual algorithms in SDP and symmetric cone programs obtained in [3], with curvature integrals of primal/dual algorithms obtained in [10]. And this is a direction extension of [10, 3] in LP case. Together with the numerical evidences in [3, 3], we claim that *the number of iterations of the interior-point algorithm is expressed as a differential geometric quantity.*

The organization of this paper is as follows. In section 2, we discuss the primal-dual algorithms in SDP and symmetric cone programs and then summarize the relationship between curvature integrals and iteration complexities in LP, SDP, and symmetric cone programs. In section 3, we summarize the information geometry and then in section 4, discuss its framework for conic linear programs. In section 5, we establish Pythagorean relationship in SDP case and in section 6 we do in symmetric cone programs' case. We end with conclusions in section 7.

2 SDP and Symmetric Cone Programs and their Curvature Integrals

In this section, we lay out minimal foundations of primal-dual algorithms in SDP and symmetric cone programs in section 2.1 and 2.2 respectively, and then state the relationship between curvature integrals and iteration complexities in LP [5] and in SDP and symmetric cone programs [3].

2.1 SDP

Let $\mathbb{S}^n$ be a real vector space generated by $n \times n$ real matrices, and its associated inner product is defined by $A \bullet B := \mathrm{Tr}(AB)$ for any $A, B \in \mathbb{S}^n$. $X \succeq 0$ and $X \succ 0$ denote the positive semidefiniteness and positive definiteness of X, respectively. Let $\mathbb{R}^m$ be a m-dimensional real vector space.

The standard primal form of SDP is formulated as

$$\begin{array}{llrcl} (P) & \min_X & C \bullet X & & \\ & & A_i \bullet X & = & b_i, \quad i = 1, \ldots, m, \\ & & X & \succeq & 0, \end{array} \tag{1}$$

and its dual is formulated as

$$\begin{array}{llrcrcl} (D) & \max_{y,S} & b^T y & & & & \\ & & \sum_{i=1}^m y_i A_i & + & S & = & C, \\ & & & & S & \succeq & 0, \end{array} \tag{2}$$

where $C \in \mathbb{S}^n, A_i \in \mathbb{S}^n,\ i = 1, \ldots, m$ and $b = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ are the given data, and $X \in \mathbb{S}^n$ and $(y, S) \in \mathbb{R}^m \times \mathbb{S}^n$ are the primal and dual variables, respectively. We assume that $A_i, i = 1, \ldots, m$ are linearly independent.

The *feasible regions* to (1) and (2) are defined by

$$\begin{aligned} \mathcal{P} &:= \{X \in \mathbb{S}^n \mid A_i \bullet X = b_i,\ i = 1, \ldots, m, \quad X \succeq 0\}, && (3)\\ \mathcal{D} &:= \{(y, S) \in \mathbb{R}^m \times \mathbb{S}^n \mid \textstyle\sum_{i=1}^m y_i A_i + S = C, \quad S \succeq 0\}, && (4) \end{aligned}$$

and its corresponding *strictly feasible* regions are defined by

$$\begin{aligned} \mathcal{P}^+ &:= \{X \in \mathcal{P} \mid X \succ 0\}, && (5)\\ \mathcal{D}^+ &:= \{(y, S) \in \mathcal{D} \mid S \succ 0\}. && (6) \end{aligned}$$

We assume that $\mathcal{P}^+ \neq \emptyset$ and $\mathcal{D}^+ \neq \emptyset$.

Using the *(normalized) duality gap* defined by $\mu(X, S) := X \bullet S/n$, we define the distance measure by

$$d_F(X, S) \quad := \quad \left\|X^{1/2} S X^{1/2} - \mu(X, S) I\right\|_F \quad = \quad \left[\textstyle\sum_{i=1}^n (\lambda_i(XS) - \mu)^2\right]^{1/2}, \tag{7}$$

where $X^{1/2}$ for $X \succ 0$ is the positive semidefinite square root, For a given constant $0 < \beta < 1$, we then define the neighborhood

$$\mathcal{N}(\beta) \quad := \quad \{(X, y, S) \in \mathcal{P} \times \mathcal{D} \mid d_F(X, S) \leq \beta \mu(X, S)\}.$$

Let us simplify the notation slightly; we define the operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ by

$$(\mathcal{A}X)_i := A_i \bullet X, \quad i = 1, \ldots, m.$$

Then the adjoint of this operator is $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{S}^n$ satisfying

$$\mathcal{A}^* y = \textstyle\sum_{i=1}^m y_i A_i.$$

It is known that the system

$$\begin{array}{rcccl} & \mathcal{A}^* y & + & S & = & C, \\ \mathcal{A}X & & & & = & b, \\ & XS & & & = & \nu I. \end{array} \tag{8}$$

has a unique solution in $(X, y, S) \in \mathcal{P}^+ \times \mathcal{D}^+$ for any $\nu > 0$ and the set of such solutions are called a *central path.*

Now we introduce Monteiro-Zhang (MZ) family of directions [6], which are obtained from the system:

$$\begin{array}{rcccl} & \mathcal{A}^* \Delta y & +\Delta S & = & C - S - \mathcal{A}^* y, \\ \mathcal{A}\Delta X & & & = & b - \mathcal{A}X, \\ \mathcal{H}_P(\Delta X S & + & X \Delta S) & = & \sigma\mu I - \mathcal{H}_P(XS), \end{array} \tag{9}$$

using an operator $\mathcal{H}_P : \mathbb{R}^{n \times n} \to \mathbb{S}^n$,

$$\mathcal{H}_P(M) = \frac{1}{2}\left[PMP^{-1} + (PMP^{-1})^T\right], \tag{10}$$

with a nonsingular scaling matrix P. σ is called the centering parameter; when $\sigma = 0$, the derived direction is the Newton direction or an affine scaling direction and when $\sigma = 1$, it is a centering direction.

We will now state the scheme we will consider in this paper.

MTY Predictor-Corrector Algorithm in SDP:

Fix some constants $\beta \in (0,1)$.

Let $w^0 = (X^0, y^0, S^0) \in \mathcal{N}(\beta^2)$ and $\mu_F < \mu_0 := (X^0 \bullet S^0)/n$ be given.

Repeat until $\mu_k \leq \mu_f$, **do**

(1) Choose a nonsingular scaling matrix P^k;

(2) Compute the solution $\Delta w^k = (\Delta X^k, \Delta y^k, \Delta S^k)$ of the system (9) in MZ direction with $P = P^k$, $\mu = \mu_k$, $\sigma = 0$ and $(X, y, S) = (X^k, y^k, S^k)$;

(3) Set $z^{k+1} := w^k + \alpha_k \Delta w^k$, where $\alpha_k > 0$ is the largest $\alpha > 0$ such that $w^k + \tilde{\alpha}\Delta w^k \in \mathcal{N}(\beta)$ for all $\tilde{\alpha} \in [0, \alpha]$;

(4) From z^{k+1} compute a point $w^{k+1} = (X^{k+1}, y^{k+1}, S^{k+1}) \in \mathcal{N}(\beta^2)$ with the same duality gap as w^{k+1};

(4) Set $\mu_{k+1} := (X^{k+1} \bullet S^{k+1})/n$ and increment k by 1.

End do

End

2.2 Symmetric Cone Programs

Since symmetric cone programs are described by Euclidean Jordan algebra, we firstly summarize Euclidean Jordan algebra and then state the problem and its algorithm. More detailed discussions are found in [2, 7, 11].

Let V be an n-dimensional real vector space with a bilinear mapping $(x, y) \mapsto x \circ y$ from $V \times V$ into V. *Jordan algebra* is a vector space V with a multiplication $\circ$ satisfying

1. $x \circ y = y \circ x$
2. $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$

where $x^2 = x \circ x$. Note that in general it is not associative, i.e., $x \circ (y \circ z) \neq (x \circ y) \circ z$, but it is power associative, i.e., $x^{k+l} = x^k \circ x^l$ for nonnegative integers k and l. In this paper, we assume that there exists a unique identity $e \in V$ such that $x \circ e = e \circ x = x$ for all $x \in V$.

If there exists a positive definite bilinear form $\langle \cdot, \cdot \rangle$ on V which is associative, namely,

$$\langle x \circ y, z \rangle = \langle x, y \circ z \rangle \ \text{ for } \forall x, y, z \in V, \tag{11}$$

then, Jordan algebra is *Euclidean*. Since Euclidean Jordan algebra is power associative, the notion of *rank* and spectral decomposition are naturally defined. Using spectral decomposition, *symmetric cones* are defined as the set of elements in V whose spectral values are all nonnegative. Or equivalently, they are defined as the cones of squares of Euclidean Jordan algebra V. We denote that the interior of the symmetric cone by Ω. It is known that symmetric cones are self-dual, i.e.,

$$\Omega = \Omega^* := \{ y \in V \mid \langle y, x \rangle > 0, \quad \forall x \in V \backslash \{0\} \}.$$

We introduce a linear form L_x and two types of quadratic forms Q_x and $Q_{x,y}$ as follows. L_x is a linear mapping of V defined by $L_x y := x \circ y$ for $\forall x, y \in V$ and $Q_x := 2L_x^2 - L_{x^2}$ and $Q_{x,y} := L_x L_y + L_y L_x - L_{x \circ y}$.

Let V be an n-dimensional Euclidean Jordan algebra with rank r and Ω its associated symmetric cone. We consider a primal-dual pair of the symmetric cone program:

$$(P) \qquad \min_x \langle c, x \rangle \qquad \text{s.t.} \quad \begin{array}{l} \langle a_i, x \rangle = b_i, \quad i = 1, \ldots, m, \\ x \in \mathrm{cl}(\Omega), \end{array} \tag{12}$$

and

$$(D) \qquad \min_{y,s} b^T y \qquad \text{s.t.} \quad \begin{array}{l} \sum_{i=1}^m y_i a_i + s = c, \\ s \in \mathrm{cl}(\Omega), \end{array} \tag{13}$$

where $c \in V$, $a_i \in V$, $i = 1, \dots, m$ and $b = (b_1, \dots, b_m)^T \in \mathbb{R}^m$ are the given data, and $x \in V$ and $(y, s) \in \mathbb{R}^m \times V$ are the primal and dual variables, respectively.

The feasible regions of (12) and (13) are defined as

$$\mathcal{P} \;:=\; \{x \in V \mid \langle a_i, x\rangle = b_i,\ i = 1, \dots, m, \quad x \in \mathrm{cl}(\Omega)\}, \tag{14}$$

$$\mathcal{D} \;:=\; \left\{(y, s) \in \mathbb{R}^m \times V \mid \sum\nolimits_{i=1}^{m} y_i a_i + s = c, \quad s \in \mathrm{cl}(\Omega)\right\}, \tag{15}$$

and its corresponding strictly feasible regions are defined as

$$\mathcal{P}^+ \;:=\; \{x \in \mathcal{P} \mid x \in \Omega\}, \tag{16}$$

$$\mathcal{D}^+ \;:=\; \{(y, s) \in \mathcal{D} \mid s \in \Omega\}. \tag{17}$$

We assume that $\mathcal{P}^+ \neq \emptyset$ and $\mathcal{D}^+ \neq \emptyset$.

Using the normalized duality gap $\mu(x, s) := \langle s, x\rangle / r$, the distance measure is defined as

$$d(x, s) \;:=\; \|Q_{x^{1/2}} s - \mu(x, s)e\| \;=\; \left[\sum\nolimits_{i=1}^{r} (\lambda_i(x \circ s) - \mu)^2\right]^{1/2}, \tag{18}$$

and, for a given constant $0 < \beta < 1$, we then define the neighborhood:

$$\mathcal{N}(\beta) \;:=\; \{(x, y, s) \in \mathcal{P}^+ \times \mathcal{D}^+ \mid d(x, s) \leq \beta\mu(x, s)\}.$$

Using the notations:

$$(\mathcal{A}x)_i := \langle a_i, x\rangle, \quad i = 1, \dots, m,$$

and letting $\mathcal{A}^*$ be an adjoint operator of $\mathcal{A}$, the system:

$$\begin{array}{rcrcl} & & \mathcal{A}^* y + s & = & c, \\ \mathcal{A}x & & & = & b, \\ & & x \circ s & = & \nu e, \end{array} \tag{19}$$

has a unique solution in $(x, y, s) \in \mathcal{P}^+ \times \mathcal{D}^+$ for any $\nu > 0$ and the set of such solutions are called the *central path*.

Based on this, we introduce Monteiro-Zhang (MZ) family of directions [6], which are given by

$$\begin{array}{rcrcl} & \mathcal{A}^*\Delta y & +\Delta s & = & c - s - \mathcal{A}^* y, \\ \mathcal{A}\Delta x & & & = & b - \mathcal{A}x, \\ (g^{-*}s) \circ (g\Delta x) & + & (gx) \circ (g^{-*}\Delta s) & = & \sigma\mu e - (gx) \circ (g^{-*}s). \end{array} \tag{20}$$

Note that σ is a centering parameter where the solution is the Newton direction when $\sigma = 0$ and one is the centering direction when $\sigma = 1$. The MTY-PC algorithm in symmetric cone programs is identical to one in SDP except that it solves (20) instead of (9) for computing the direction.

2.3 Curvature Integrals and iteration Complexities in MTY-PC algorithms

We briefly lay out the results in regard to the relationship between curvature integrals and iteration complexities in SDP and symmetric cone programs discussed in [5] and [3].

Theorem 2.1. *Let $\beta \in (0, 1/2]$. For given $w^0 \in \mathcal{N}(\beta)$ and $0 < \nu_f < \mu(w^0)$ denote by $\#_{PD}(\mu(w^0), \nu_f, \beta)$ the number of iterations of the MTY-PC algorithm with $\beta \in (0, 1/2]$ needed to reduce the duality gap from $\nu_i := \mu(w^0)$ to ν_f. Then, using curvature integral:*

$$I_{PD}(\nu_f, \nu_i) = \int_{\nu_f}^{\nu_i} h_{PD}^{1/2}(\nu)\mathrm{d}\nu, \tag{21}$$

with the square of the integrands:

$$h_{PD}(\nu) \;:=\; \frac{\|\dot{x} \circ \dot{s}\|}{\nu} \quad \textit{(LP case)},$$

$$h_{PD}(\nu) \;:=\; \left\|X^{-1/2}\dot{X}X^{-1/2} \circ S^{-1/2}\dot{S}S^{-1/2}\right\|_F \quad \textit{(SDP case)},$$

$$h_{PD}(\nu) \;:=\; \|Q_{x^{-1/2}}\dot{x} \circ Q_{s^{-1/2}}\dot{s}\| \quad \textit{(symmetric cone programs case)},$$

we have

$$\lim_{\beta \downarrow 0} \frac{I_{PD}(\nu_f, \nu_i)/\sqrt{\beta}}{\#_{PD}(\nu_i, \nu_f, \beta)} = 1.$$

3 Brief Overview of Information Geometry

In this section, we lay out a minimal foundation of information geometry and its relation to interior-point methods. We firstly introduce the structure of informaion geometry and then discuss how information geometry fits into the framework of convex programming. For more concise information, please refer to [1] for information geometry and [9] for a general theory of interior-point methods. The contents of this section has considerable overlap with [3]. To make the paper self-contained, we repeat them here.

3.1 Setting and Background

3.1.1 Dually Flat Space

Let $\mathbf{E}$ and $\mathbf{E}^*$ be an n-dimensional real vector space and its dual space. We denote by $\langle s, x\rangle$ the duality product of $x \in \mathbf{E}$ and $s \in \mathbf{E}^*$. Let $\{e_1, \ldots, e_n\}$ be the basis of $\mathbf{E}$ and $\{e_1^*, \ldots, e_n^*\}$ be its dual basis of $\mathbf{E}^*$ with a biorthonormal property, i.e., $\langle e_i, e_j^*\rangle = \delta_{ij}$. We consider the affine coordinate system $(x^1, \ldots, x^n)$ based on $\{e_1, \ldots, e_n\}$ so that any $x \in \mathbf{E}$ is represented by $x = \sum_{i=1}^n x^i e_i$. Similarly, we consider $(s^1, \ldots, s^n)$ based on $\{e_1^*, \ldots, e_n^*\}$ with any $s = \sum_{i=1}^n s^i e_i^*$. In this setting, the inner product is now computed as $\langle s, x\rangle = \sum_{i=1}^n x^i s^i$.

Let $\mathcal{C}$ be an open convex set in $\mathbf{E}$ with nonempty interior and let $\mathcal{C}^x$ be its x-coordinate expression, i.e., $\mathcal{C}^x = \{\xi \in \mathbf{R}^n \mid \sum_{i=1}^n \xi^i e_i \in \mathcal{C}\}$. We introduce dually flat structure on $\mathcal{C}$. Let $\Psi(x)$ be a strongly convex smooth function on $\mathcal{C}$. In the following, Ψ is also regarded as a function on $\mathbb{R}^n$ under the coordinate system $(x^1, \ldots, x^n)$. Then the gradient mapping

$$\bar{s}(\cdot) : x \in \mathcal{C}^x \mapsto \bar{s} \in \mathbf{R}^n, \quad \bar{s}^i = -\partial\Psi/\partial x^i \tag{22}$$

is smoothly invertible on its image $\mathcal{C}^s := \bar{s}(\mathcal{C}^x) \subseteq \mathbf{R}^n$ because the Hessian matrix of Ψ is positive definite. We call $(\bar{s}^1, \ldots, \bar{s}^n)$ the *dual coordinate* of $\mathcal{C}$. $\mathcal{C}^s$ becomes convex in $\mathbf{R}^n$ under appropriate regularity conditions, e.g., $\Psi(x) \to \infty$ as $x \to \partial\mathcal{C}^x$. Obviously, the set

$$\mathcal{G}(\mathcal{C}) := \left\{\sum\nolimits_{i=1}^n s^i e_i^* \;\middle|\; s \in \mathcal{C}^s\right\} = \left\{\sum\nolimits_{i=1}^n \bar{s}^i(x) e_i^* \;\middle|\; x \in \mathcal{C}^x\right\} \tag{23}$$

is a convex set in $\mathbf{E}^*$. The set $\mathcal{G}(\mathcal{C})$ does not depend on the choice of the basis $\{e_1, \ldots, e_n\}$ (nor $\{e_1^*, \ldots, e_n^*\}$). Later we will take $\mathcal{C}$ to be Ω and Ψ to be a normal barrier function on Ω. Then $\mathcal{G}(\mathcal{C})$ will be the dual cone Ω^*. This is a fundamental fact which connects information geometry and interior-point algorithms. In the following, the original coordinate is referred to as x-coordinate and the dual coordinate is referred to as s-coordinate.

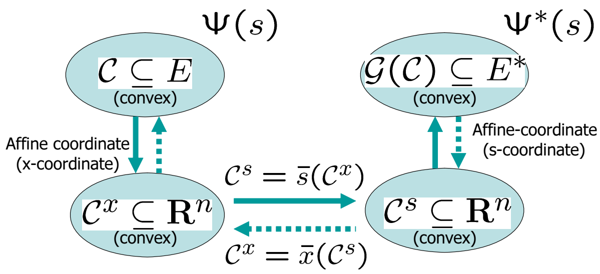

Figure 1: Dually flat space.

The conjugate function Ψ^* is defined by

$$\Psi^*(s) = \sup_{x \in \mathcal{C}^x} \{-\langle s, x\rangle - \Psi(x)\}.$$

The domain of $\Psi^*(s)$ is $\mathcal{C}^s$. As to Ψ^*, we use a similar convention as Ψ, namely, Ψ^* is regarded as a function on $\mathbf{E}^*$ by associating $(s^1, \ldots, s^n) \in \mathbf{R}^n$ with $\sum_{i=1}^n s^i e_i^* \in \mathbf{E}^*$. Obviously the domain of Ψ^* as a function on $\mathbf{E}^*$ is $\mathcal{G}(\mathcal{C})$. If we start our discussion from $\mathcal{G}(\mathcal{C}) \subset \mathbf{E}^*$ and Ψ^*, we will recover $\mathcal{C} \subset \mathbf{E}$ and Ψ exactly in the same manner, i.e., the dual coordinate $\bar{x}(\cdot) : \mathcal{G}(\mathcal{C})^s \to \mathcal{C}^x$ is given by

$$\bar{x}(\cdot) : s \in \mathcal{G}(\mathcal{C})^s \mapsto \bar{x} \in \mathbf{R}^n, \quad \bar{x}^i = -\partial\Psi^*/\partial s^i \tag{24}$$

and we have $\mathcal{C}^x = \bar{x}(\mathcal{C}^s)$. $\bar{s}(\cdot)$ and $\bar{x}(\cdot)$ are mutually inverse maps. This fact readily follows by notifying that

$$\Psi^*(\bar{s}(x)) = -x^\top \frac{\partial \Psi}{\partial x} + \Psi(x)$$

and differentiating the both sides with respect to x.

The sets $\mathcal{C}^x$ and $\mathcal{C}^s$ are coordinate expressions of $\mathcal{C}$ in x- and s-coordinate, respectively. While x-coordinate is an affine coordinate on $\mathcal{C}$, s-coordinate is a nonlinear coordinate on it. On the other hand, we may regard $\mathcal{C}^x$ and $\mathcal{C}^s$ as coordinate expressions of $\mathcal{G}(\mathcal{C})$ as well, with s-coordinate being an affine coordinate and x-coordinate being a nonlinear coordinate on it. Thus, $\mathcal{C}$ and $\mathcal{G}(\mathcal{C})$ share $\mathcal{C}^x$ and $\mathcal{C}^s$ in common as coordinate representations. This is a remarkable primal-dual symmetry which should be kept in mind throughout this paper. The information geometric structure induced on $\mathcal{C}$ by Ψ is translated into $\mathcal{G}(\mathcal{C})$ through s-coordinate and the relation (23), and this structure is exactly the same as the information geometric structure induced on $\mathcal{G}(\mathcal{C})$ with Ψ^*.

In the following, we will use the letters p, p_1, p_2, etc to represent points in $\mathcal{C}$. We denote by $x(p)$ and $s(p)$ the coordinate functions to give the coordinate values of $p \in \mathcal{C}$ in x- and s-coordinates, respectively. The following relation is obvious but worth mentioning:

$$\bar{x}(s(p)) = x(p), \quad \bar{s}(x(p)) = s(p). \tag{25}$$

For a submanifold $\mathcal{M}$ of $\mathcal{C}$, we denote by $T_p\mathcal{M}$ the tangent space of $\mathcal{M}$ at $p \in \mathcal{M}$, and by $T\mathcal{M}$ the set $\cup_{p\in\mathcal{M}} T_p\mathcal{M}$. The representations of $\mathcal{M}$ in x- and s-coordinates are written as $\mathcal{M}^x$ and $\mathcal{M}^s$, respectively, namely, we let $\mathcal{M}^x := \{x(p) \in \mathbb{R}^n \mid p \in \mathcal{M}\}$ and $\mathcal{M}^s := \{s(p) \in \mathbb{R}^n \mid p \in \mathcal{M}\}$.

A vector (or a tensor) at $p \in \mathcal{C}$ is written V_p, say, with the lower subscript. If we consider a vector (or a tensor) field V over $\mathcal{C}$ or its submanifold, then its value at p is also written as V_p. We also use notations $V_{x(p)}$ and $V_{s(p)}$ to represent a vector (or a tensor) V_p in x- and s- coordinate, respectively.

Now we introduce a Riemannian metric on $\mathcal{C}$. Let $p \in \mathcal{C}$ and $V_p, W_p \in T_p\mathcal{C}$. Riemannian metric G_p is given by the Hessian matrix $G_{x(p)}$ of Ψ in x-coordinate and using this $G_{x(p)}$, the inner product of two vectors $V_p, W_p \in T_p\mathcal{M}$ is defined as follows via x-coordinate:

$$G_p(V_p, W_p) := \sum\nolimits_{i,j} (G_{x(p)})_{ij} V^i_{x(p)} W^j_{x(p)} = V^\top_{x(p)} G_{x(p)} W_{x(p)}, \quad (G_{x(p)})_{ij} = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x(p)).$$

The Jacobian matrix of the gradient map is

$$\frac{\partial \bar{s}}{\partial x}(x(p)) = -G_{x(p)}.$$

Thus, for a vector V_p, we have $V_{s(p)} = -G_{x(p)} V_{x(p)}$. Therefore, the metric G_p in s-coordinate becomes the inverse of $G_{x(p)}$, i.e., $G_{s(p)} = G^{-1}_{x(p)}$.

To confirm that the information geometric structures introduced by $(\mathcal{C}, \Psi)$ and $(\mathcal{G}(\mathcal{C}), \Psi^*)$ are consistent, we observe that the Riemannian metric G^*_p defined by the Hessian matrix of Ψ^* coincides with G_p. Let

$$(G^*_{s(p)})_{ij} := \frac{\partial^2 \Psi^*}{\partial s^i \partial s^j}(s(p)).$$

By differentiating the both sides of $\bar{x}(\bar{s}(x)) = x$ with x, we obtain $G^*_{s(p)} = G^{-1}_{x(p)} = G_{s(p)}$.

For $V_p \in T_p\mathcal{C}$, the length $\sqrt{G_p(V_p, V_p)}$ of V_p is denoted by $\|V_p\|_p$. Let $V_p \in T_p\mathcal{C}$, and let V_x and V_s be its expressions in x- and s- coordinate, respectively. Then we have $G_p(V_p, V_p) = V_x^\top G_x V_x = V_s^\top G_s V_s$ and $G_s = G_x^{-1}$.

Now we introduce affine connections and covariant derivatives on $\mathcal{C}$, which determine the structure of the manifold such as torsions and curvatures. One of the distinct features of information geometry is that it invokes two affine connections ∇ and ∇^*, which accord with dualities in convex analysis, rather than the Levi-Civita connection in Riemannian geometry. The connections ∇ and ∇^* are defined so that the straight lines in x- and s-coordinates become geodesics.

Formally, the connections are determined by the associated Christoffel symbols. The Christoffel symbols Γ and Γ^* associated with the connections ∇ and ∇^* become zero in x- and s-coordinates, i.e.,

$$\Gamma^k_{ij} = 0, \text{ (in } x\text{-coordinate)}, \quad \Gamma^{*k}_{ij} = 0, \text{ (in } s\text{-coordinate)}.$$

As was mentioned before, this means that a ∇-geodesic is nothing but a straight line in x-coordinate and so is a ∇^*-geodesic in s-coordinate.

Next we derive formulas for covariant derivatives. Since the Christoffel symbols associated with the connections ∇ and ∇^* vanish in x- and s-coordinates, respectively, the ordinary derivative in x-coordinate coincides with the covariant derivative with respect to ∇ and the ordinary derivative in s-coordinate coincides with the covariant derivative with respect to ∇^*. Consequently, we have the following simple expressions of the covariant derivatives of a vector field X at p with respect to $Y \in T_p\mathcal{C}$, namely,

$$(\nabla_Y X)_{x(p)} = \frac{\partial X_{x(p)}}{\partial x} Y_{x(p)},$$

in x-coordinate where $X_{x(p)}$ and $Y_{x(p)}$ are their x-coordinate expressions, and

$$\left(\frac{\partial X_{x(p)}}{\partial x}\right)_{ij} = \frac{\partial X_x^i}{\partial x^j}.$$

Similarly, we have

$$(\nabla^*_Y X)_{s(p)} = \frac{\partial X_{s(p)}}{\partial s} Y_{s(p)},$$

in s-coordinate, where $X_{s(p)}$ and $Y_{s(p)}$ are their s-coordinate expressions.

Furthermore, let $\gamma := \{\gamma(t) \in \mathcal{C} \mid t \in [a,b] \subset \mathbb{R}\}$ be a curve in $\mathcal{C}$. Then, for any $t \in [a,b]$, we have

$$(\nabla_{\dot{\gamma}}\dot{\gamma})_{x(\gamma(t))} = \frac{\mathrm{d}^2 x(\gamma(t))}{\mathrm{d}t^2} \quad \text{and} \quad (\nabla^*_{\dot{\gamma}}\dot{\gamma})_{s(\gamma(t))} = \frac{\mathrm{d}^2 s(\gamma(t))}{\mathrm{d}t^2}. \tag{26}$$

These relations follows since $\Gamma = 0$ and $\Gamma^* = 0$ in x-coordinate and s-coordinate, respectively.

We also have

$$(\nabla_{\dot{\gamma}} G)_x = \frac{\mathrm{d}G_{x(\gamma(t))}}{\mathrm{d}t} \quad \text{and} \quad (\nabla^*_{\dot{\gamma}} G)_s = \frac{\mathrm{d}G_{s(\gamma(t))}}{\mathrm{d}t} \tag{27}$$

as Γ and Γ^* vanishes in x-coordinate and s-coordinate, respectively.

3.2 Autoparallel Submanifolds and Embedding Curvature

Let $\mathcal{M}$ be a submanifold of $\mathcal{C}$. $\mathcal{M}$ is a ∇-autoparallel submanifold of $\mathcal{C}$ if $\mathcal{M}^x$ ($\mathcal{M}$ in x-coordinate) is written as an open subset of the intersection of $\mathcal{C}^x$ and an affine subspace in $\mathbb{R}^n$:

$$\begin{aligned} \mathcal{M} &\subseteq \{p \in \mathcal{C} | x(p) = c_0 + \textstyle\sum_{i=1}^k y^i c_i,\ c_i \in \mathbf{E}, y^i \in \mathbb{R}\}, \text{ or equivalently,} \\ \mathcal{M}^x &\subseteq \{x \in \mathcal{C}^x | x = c_0 + \textstyle\sum_{i=1}^k y^i c_i,\ c_i \in \mathbb{R}^n, y^i \in \mathbb{R}\}. \end{aligned}$$

Similarly, a ∇^*-*autoparallel* submanifold $\mathcal{M}$ is defined as a manifold which is represented in s-coordinate as an open subset of the intersection of $\mathcal{C}^s$ and an affine subspace.

Let $\mathcal{M}$ be a ∇-autoparallel submanifold of $\mathcal{C}$ and consider its homogenization in x-coordinate:

$$\begin{aligned} \mathrm{Hom}(\mathcal{M}) &:= \textstyle\bigcup_{t>0} t\mathcal{M} = \{p \in \mathcal{C} |\ x(p) = tc_0 + \sum_{l=1}^k y^i c_i, t > 0,\ x(p)/t \in \mathcal{M}^x\}, \\ t\mathcal{M} &:= \{p \in \mathcal{C} | x(p) = tx(p'), p' \in \mathcal{M}\}. \end{aligned}$$

Since $\mathrm{Hom}(\mathcal{M})^x$ is an open subset of $\mathcal{C}^x$ contained in the affine space $\{x \in \mathbb{R}^n | x = tc_0 + \sum_i y^i c_i\}$, $\mathrm{Hom}(\mathcal{M})$ is a ∇-autoparallel submanifold of $\mathcal{C}$.

An analogous notation is applied to a ∇^*-autoparallel submanifold in $\mathcal{C}$ using s-coordinate, namely, if $\mathcal{M}$ is ∇^*-autoparallel submanifold, we let

$$\mathrm{Hom}(\mathcal{M}) := \textstyle\bigcup_{t>0} t\mathcal{M}, \quad t\mathcal{M} = \{p \in \mathcal{C} | s(p) = ts(p'), p' \in \mathcal{M}\}.$$

$\mathrm{Hom}(\mathcal{M})$ is an ∇^*-autoparallel submanifold.

Let $\mathcal{M}$ be a k-dimensional submanifold in the dually flat manifold $\mathcal{C}$. We define the embedding curvature $H_{\mathcal{M}}(\cdot,\cdot)$ of $\mathcal{M}$ as follows. Since the tangent space $T_p\mathcal{C}$ at $p \in \mathcal{M}$ has the orthogonal decomposition with respect to the Riemannian metric G, i.e.,

$$T_p\mathcal{C} = T_p\mathcal{M} \oplus (T_p\mathcal{M})^{\perp},$$

we can define the orthogonal projection $\Pi_p^{\perp} : T_p\mathcal{C} \to (T_p\mathcal{M})^{\perp}$ at each p. For tangent vector fields X and Y on $\mathcal{M}$, let $H_{\mathcal{M}}(X,Y)$ be a normal vector field on $\mathcal{M}$ defined by

$$(H_{\mathcal{M}}(X,Y))_p = \Pi_p^{\perp}(\nabla_X Y)_p \in (T_p\mathcal{M})^{\perp},$$

at each p. Such a tensor field $H_{\mathcal{M}}$ is called the *(Euler-Schouten) embedding curvature* or the *second fundamental form* of $\mathcal{M}$ with respect to ∇. Similarly, we can introduce the dual embedding curvature $H^*_{\mathcal{M}}$ by replacing ∇ with ∇^*, i.e.,

$$(H^*_{\mathcal{M}}(X,Y))_p = \Pi_p^{\perp}(\nabla^*_X Y)_p \in (T_p\mathcal{M})^{\perp}.$$

It is shown that $\mathcal{M}$ is ∇-autoparallel (∇^*-autoparallel) *iff* $H_{\mathcal{M}} = 0$ ($H^*_{\mathcal{M}} = 0$).

For later use, we provide a concrete formula of $\Pi_p^{\perp}$ in x-coordinate and s-coordinate. We will denote them as $\Pi^{\perp}_{x(p)}$ and $\Pi^{\perp}_{s(p)}$, respectively. In x-coordinate, suppose that $T_p\mathcal{M} \subset T_p\mathcal{C}$ is represented by the kernel of a certain linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, i.e.,

$$V_p \in T_p\mathcal{M} \iff AV_{x(p)} = 0.$$

Then we have

$$\Pi^{\perp}_{x(p)} = G^{-1}_{x(p)} A^{\top} (A G^{-1}_{x(p)} A^{\top})^{-1} A,$$

and since $G_{x(p)} = G^{-1}_{s(p)}$ and $-G_{x(p)}$ is the operator of the coordinate transformation from x-coordinate to s-coordinate at the tangent space $T_p\mathcal{C}$, we have

$$\Pi^{\perp}_{s(p)} = G_{x(p)} \Pi^{\perp}_{x(p)} G^{-1}_{x(p)} = A^{\top}(A G_{s(p)} A^{\top})^{-1} A G_{s(p)}. \tag{28}$$

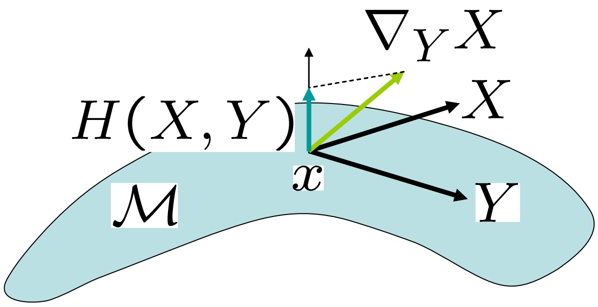

Figure 2: Embedding curvature.

3.3 Self-Concordant Functions and Information Geometry

If

$$\left| \sum\nolimits_{i,j,k}^{n} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} X^i X^j X^k \right| \le 2 \left(\sum\nolimits_{i,j}^{n} \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} X^i X^j \right)^{3/2} \tag{29}$$

holds for all $x \in \mathcal{C}^x$ and $X \in \mathbb{R}^n$, then Ψ is called a *self-concordant* function on $\mathcal{C}$.

A self-concordant function Ψ is said to be a *self-concordant barrier* if it satisfies

$$\Psi(x) \to \infty \text{ as } x \to \partial\mathcal{C}.$$

If, in addition, the self-concordant barrier satisfies the condition

$$\sum_i \left|\frac{\partial \Psi(x)}{\partial x^i} X^i\right| \le \sqrt{\theta}\left(\sum_{i,j}^n \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} X^i X^j\right)^{1/2}, \tag{30}$$

for all $x \in \mathcal{C}^x$ and $X \in \mathbb{R}^n$, then we call $\Psi(x)$ *θ-self-concordant barrier* on $\mathcal{C}$.

Let Ω be a proper open convex cone, and let $\Omega^* = \{s \in \mathbf{E}^* | \ \langle x, s\rangle > 0, \ \forall x \in \Omega\backslash\{0\}\}$ be the open dual cone. A barrier function $\Psi(x)$ on Ω is called θ-logarithmically homogeneous if $\Psi(tx) = \Psi(x) - \theta \log t$ holds for $t > 0$. If $\Psi(x)$ is a θ-logarithmically homogeneous barrier on Ω, so is Ψ^* on Ω^*. A self-concordant barrier is called a *θ-normal barrier* if it is θ-logarithmically homogeneous. A θ-normal barrier is known to be a θ-self-concordant barrier. If $\Psi(x)$ is a θ-normal barrier function on Ω, so is Ψ^* on Ω^*.

Now, let us consider the information geometry introduced on Ω with a θ-self-concordant barrier ψ as the potential function. This means that we take Ω as $\mathcal{C}$ and consider ψ as the potential function Ψ. In the following, we denote the common coordinate representations $\mathcal{C}^x$ and $\mathcal{C}^s$ of $\Omega(= \mathcal{C})$ and $\mathcal{G}(\Omega)(= \mathcal{G}(\mathcal{C}))$ by Ω^x and Ω^s, respectively.

Let $\Omega^*_{\mathbb{R}}$ be the coordinate representation of Ω^*, i.e.,

$$\Omega^*_{\mathbb{R}} = \{s \in \mathbb{R}^n | \ \sum_i s^i e_i^* \in \Omega^*\} = \{s \in \mathbb{R}^n | \ s^\top x > 0 \ \forall x \in \Omega^x \setminus \{0\}\}.$$

The next result plays a fundamental role in this paper. The theorem claims that the primal cone Ω can be identified with the dual cone Ω^* through the gradient map. We illustrate the situation in Fig. 3.

Theorem 3.1. *Under the notations as above, $\Omega^s = \Omega^*_{\mathbb{R}}$, and hence $\mathcal{G}(\Omega) = \Omega^*$. Ω^x and Ω^s are coordinate representations of Ω and Ω^* such that*

$$\text{(a) } \Omega = \{x \in \mathbf{E} | \ x = \sum_{i=1}^n \xi^i e_i, \ \xi \in \Omega^x\} \qquad \text{(b) } \Omega^* = \{s \in \mathbf{E}^* | \ s = \sum_{i=1}^n \sigma^i e_i^*, \ \sigma \in \Omega^s\}, \tag{31}$$

that is, x-coordinate and s-coordinate are affine coordinate systems on Ω and Ω^*, respectively. On the other hand, s-coordinate is a nonlinear coordinate system of Ω and x-coordinate is a nonlinear coordinate system of Ω^*. Thus, Ω^x and Ω^s are common coordinate systems of Ω and Ω^*.

Proof. What we need to show here is $\Omega^s = \Omega^*_{\mathbb{R}}$. Other statements are obvious from previous discussions and definitions. First we observe that $\Omega^s \subseteq \Omega^*_{\mathbb{R}}$. Let $s \in \Omega^s$. This implies that the optimization problem: $\min_{x\in\Omega^x} -s^\top x + \psi(x)$ have an optimal solution. Since $\psi(x)$ is strongly convex, the set $\{x \in \mathbf{R}^n | \ s^\top x = 1\} \cap \Omega_{\mathbf{R}}$ is bounded, and hence $\{x \in \mathbf{R}^n | \ s^\top x = 0\} \cap \Omega_{\mathbf{R}} = \{0\}$. Due to the convex conic version of Gordan's theorem, we have $s \in \Omega^*_{\mathbb{R}}$.

Next we show $\Omega^*_{\mathbb{R}} \subseteq \Omega^s$. If $s \in \Omega^*_{\mathbb{R}}$, then the set $X = \{x \in \Omega^x | \ s^\top x = 1\}$ is bounded because s is an interior point of $\mathrm{cl}(\Omega^*_{\mathbb{R}})$ which is the dual cone of $\mathrm{cl}(\Omega^x)$. Therefore, $\min_{x\in X} \psi(x)$ have an optimal solution $\hat{x}$ which satisfies $\lambda s = -\partial\psi(\hat{x})/\partial x = \bar{s}(\hat{x})$. Due to logarithmically homogeneous property of ψ, we have $s = \bar{s}(\hat{x}/\lambda)$, and therefore $s \in \Omega^s$. $\square$

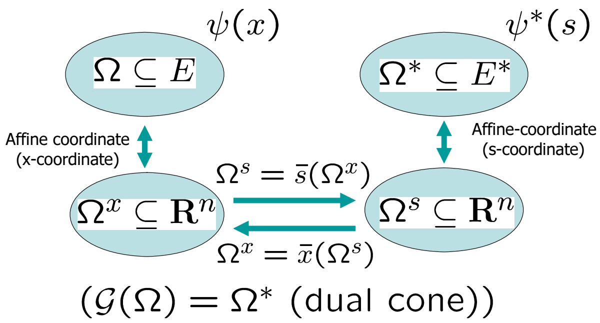

Figure 3: Dually flat space (Conic case).

Proposition 3.2. *Let $\psi(x)$ be a θ-logarithmically homogeneous barrier on Ω, and consider information geometry of Ω based on ψ. Let $p, p' \in \mathcal{C}$, and let $t > 0$. Then the following statement holds:*

(a) $ts(p') = s(p)$ *iff* $x(p') = tx(p)$.

(b) $G_{x(p')} = t^{-2} G_{x(p)}$ *if* $x(p') = tx(p)$, *and* $G_{s(p')} = t^{-2} G_{s(p)}$ *if* $s(p') = ts(p)$.

(c) $s(p)^{\mathsf{T}} x(p) = \theta$.

(d) $s(p) = G_{x(p)} x(p)$ *and* $x(p) = G_{s(p)} s(p)$.

Proof. Basically the proof is the same as Proposition 2.3.4 of [9] (just the notation is different). $\square$

In the end of this section, we provide a list of symbols and notations which will frequently appear in the following sections:

$x(p), s(p)$:	x- and s-coordinate representations of a point p in a dually flat manifold,
$\bar{x}(s)$:	Gradient map (Legendre transformation) from $\mathcal{C}^s$ to $\mathcal{C}^x$,
$\bar{s}(x)$:	Gradient map (Legendre transformation) from $\mathcal{C}^x$ to $\mathcal{C}^s$,
G_x :	Riemannian metric in x-coordinate,
G_s :	Riemannian metric in s-coordinate,
$\mathcal{M}^x$:	x-coordinate representation of a submanifold $\mathcal{M}$: $\{\xi \in \mathbf{R}^n \mid \xi = x(p),\ p \in \mathcal{M}\}$,
$\mathcal{M}^s$:	s-coordinate representation of a submanifold $\mathcal{M}$: $\{\sigma \in \mathbf{R}^n \mid \sigma = s(p),\ p \in \mathcal{M}\}$,
V_p :	Vector or tensor at p in a manifold,
$V_{x(p)},\ V_{s(p)}$:	Vector V at p represented in x-coordinate and s-coordinate, respectively.

4 Information Geometric Framework of Conic Linear Programs

In this section, we first introduce the canonical form of conic linear programs and then state the information geometric framework for conic linear programs, their algorithms, and the relation between the embedding curvature and iteration complexity.

4.1 Conic Linear Programs

Let $\Omega \in \mathbf{E}$ be a proper open convex cone and $\Omega^* \in \mathbf{E}^*$ be its dual cone. The standard pair of conic linear programs are formulated as

$$(P) \qquad \min_x \langle c, x \rangle \qquad \text{s.t.} \quad \begin{array}{l} \langle a_i, x \rangle = b_i, \quad i = 1, \ldots, m, \\ x \in \mathrm{cl}(\Omega), \end{array} \tag{32}$$

and

$$(D) \qquad \min_{y,s} b^T y \qquad \text{s.t.} \quad \begin{array}{l} \sum_{i=1}^m y_i a_i + s = c, \\ s \in \mathrm{cl}(\Omega^*), \end{array} \tag{33}$$

where $c \in \mathbf{E}^*$, $a_i \in \mathbf{E}^*$, $i = 1, \ldots, m$ and $b = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ are the given data, and $x \in \mathbf{E}$ and $(y, s) \in \mathbb{R}^m \times \mathbf{E}^*$ are the primal and dual variables, respectively. The operator cl returns the closure of a given set. We assume that $a_i, i = 1, \ldots, m$ are linearly independent. Using the notations:

$$(\mathcal{A}x)_i := \langle a_i, x \rangle, \quad i = 1, \ldots, m,$$

and letting $\mathcal{A}^*$ be an adjoint operator of $\mathcal{A}$, $\ker \mathcal{A} = \mathcal{L}$, $\mathcal{L}^*$ be its orthogonal complement, and d be the vector satisfying $\mathcal{A}d = b$, we can rewrite (32) and (33) in the geometrical forms:

$$\min \ \langle c, x \rangle, \quad \text{s.t. } x \in (d + \mathcal{L}) \cap \mathrm{cl}(\Omega), \tag{34}$$

and

$$\min \ \langle x, d\rangle\ , \quad \text{s.t. } s \in (c + \mathcal{L}^*) \cap \mathrm{cl}(\Omega^*), \tag{35}$$

where $\mathcal{L}$ is $(n-m)$-dimensional linear subspace and $\mathcal{L}^*$ is m-dimensional linear subspace.

We consider the standard (biorthogonal) bases $\{e_1, \ldots, e_n\}$ and $\{e_1^*, \ldots, e_n^*\}$ in $\mathbf{E}$ and $\mathbf{E}^*$, respectively, and assume that (32) and (33) are represented with respect to these bases. We use the same letters for vectors and subspaces and their coordinate expressions. In particular, $c, d, \mathcal{L}$ and $\mathcal{L}^*$ below mean vectors and linear subspaces in $\mathbb{R}^n$ representing $c, d, \mathcal{L}$, and $\mathcal{L}^*$ in (34) and in (35) with respect to these bases. Note that if $x \in \mathbf{E}$ and $s \in \mathbf{E}^*$, we have $\langle x, s\rangle = x^\mathsf{T} s$, where x and s on the righthand side are their coordinate expressions.

Let $\Omega_\mathbb{R} \subseteq \mathbb{R}^n$ and $\Omega_\mathbb{R}^* \subseteq \mathbb{R}^n$ be coordinate representations of Ω and Ω^* with respect to their bases, respectively, and we define

$$\mathcal{P}^x := (d + \mathbf{T}) \cap \Omega_\mathbb{R} \text{ and } \mathcal{D}^s := (c + \mathbf{T}^*) \cap \Omega_\mathbb{R}^*.$$

Then (34) and (35) are written, as optimization problems on $\mathbf{R}^n$, as

$$\min \ c^\mathsf{T} x, \quad \text{s.t. } x \in \mathrm{cl}(\mathcal{P}^x), \tag{36}$$

and

$$\min \ s^\mathsf{T} d, \quad \text{s.t. } s \in \mathrm{cl}(\mathcal{D}^s), \tag{37}$$

respectively.

We assume that both (34) and (35) have interior feasible solutions. Then we have $\mathcal{P}^x \neq \emptyset$ and $\mathcal{D}^s \neq \emptyset$. Under this assumption, (36) and (37) have an optimal solution satisfying the following conditions:

$$s^\mathsf{T} x = 0, \ x \in \mathrm{cl}(\mathcal{P}^x), \ s \in \mathrm{cl}(\mathcal{D}^s).$$

Let $\psi(x)$ be a θ-normal barrier whose domain is Ω. The conjugate function $\psi^*(s)$ of $\psi(x)$ is a θ-normal barrier whose domain is Ω^* as noted before. Based on $\psi(x)$ and $\psi^*(s)$, we introduce the central trajectories of (36) and (37).

As for (36), we consider the following optimization problem with parameter t

$$\min \ tc^\mathsf{T} x + \psi(x) \ \text{ s.t. } x \in \mathcal{P}^x. \tag{38}$$

The optimality condition of this problem is written as:

$$\left(tc + \frac{\partial \psi}{\partial x} =\right) tc - \bar{s}(x) \in \mathbf{T}^*, \quad x \in d + \mathbf{T}, \quad x \in \Omega.$$

or equivalently,

$$\bar{s}(x) \in t\mathcal{D}^s, \quad x \in \mathcal{P}^x. \tag{39}$$

Let $x_\mathcal{P}(t)$ be the unique optimal solution to (38). The set of $x_\mathcal{P}(t)$ with parameter t varying from 0 to infinity is called the central trajectory of (36).

Similarly, we consider the following optimization problem with parameter t associated with (37).

$$\min \ ts^\mathsf{T} d + \psi^*(s) \ \text{ s.t. } s \in \mathcal{D}^s. \tag{40}$$

The optimality condition for this problem is

$$\left(td + \frac{\partial \psi^*}{\partial s} =\right) td - \bar{x}(s) \in \mathbf{T}, \quad s \in c + \mathbf{T}^*, \quad s \in \Omega^*. \tag{41}$$

Let $s_\mathcal{D}(t)$ be the unique optimal solution to (40). The set of $s_\mathcal{D}(t)$ with parameter t varying from 0 to infinity is referred to as the central trajectory of (37).

4.2 Framework

Now we consider information geometric structure on Ω induced by the θ-normal barrier ψ on the domain Ω. We choose Ω as $\mathcal{C}$ and ψ as the potential function Ψ. The primal problem (34) is an optimization problem in Ω, and is expressed as (36) in x-coordinate. We formulate the dual problem (35) as an optimization problem in Ω as follows. The dual problem is equivalent to (37) as an optimization problem on $\Omega^*_{\mathbb{R}}$. Due to Theorem 3.1, we have $\Omega_{\mathbb{R}} = \Omega^x$ and $\Omega^*_{\mathbb{R}} = \Omega^s$. Furthermore, Ω^s is a global coordinate of Ω defined by the gradient map. Therefore, (37) can be considered as an optimization over Ω by associating each feasible solution $s \in \mathcal{D}^s$ with $\bar{s}^{-1}(s) \in \Omega^x$.

Under the setting above, we define the feasible region $\mathcal{P}$ and the dual feasible region $\mathcal{D}$ as submanifolds in Ω as follows:

$$\mathcal{P} = \{p \in \Omega |\ x(p) \in \mathcal{P}^x\}, \quad \mathcal{D} = \{p \in \Omega |\ s(p) \in \mathcal{D}^s\}.$$

By definition, $\mathcal{P}$ is in x-coordinate the intersection of Ω^x and the affine space $d + \mathbf{T}$, and $\mathcal{D}$ is in s-coordinate the intersection of Ω^s and the affine space $c + \mathbf{T}^*$. Therefore, in reference to the definition in Section 2.2, $\mathcal{P}$ is an $(n-m)$-dimensional ∇-autoparallel submanifold and $\mathcal{D}$ is an m dimensional ∇^*-autoparallel submanifold. Note that a ∇-autoparallel (∇^*-autoparallel) submanifold is not necessarily ∇^*-autoparallel (∇-autoparallel). Thus, in view of x-coordinate and the connection ∇, the primal feasible region $\mathcal{P}$ is an autoparallel manifold and the dual feasible region $\mathcal{D}$ is a curved submanifold, while, in view of s-coordinate and ∇^*-connection, $\mathcal{D}$ is an autoparallel manifold and $\mathcal{P}$ is a curved submanifold. $\mathcal{P}$ and $\mathcal{D}$ intersect at a unique point, which is a point on the central trajectory as discussed below. See Fig. 4.

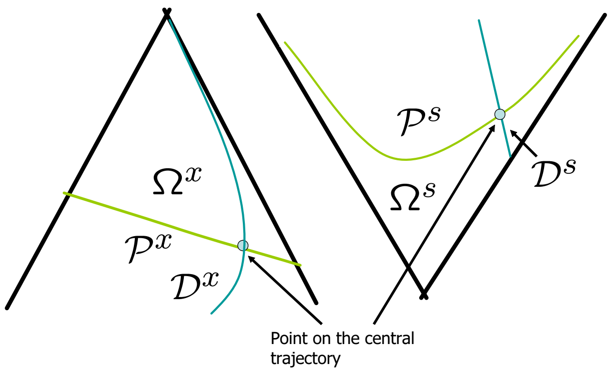

Figure 4: Primal and dual feasible regions in x-coordinate (left) and s-coordinate (right) .

Let $\gamma_{\mathcal{P}}(t)$ be the point in Ω expressed as $x_{\mathcal{P}}(t) = x(\gamma_{\mathcal{P}}(t))$ in x-coordinate. We define the central trajectory $\gamma_{\mathcal{P}}$ of (34)/(36) as one dimensional submanifold $\gamma_{\mathcal{P}} = \{\gamma_{\mathcal{P}}(t) |\ t \in [0, \infty)\}$ in Ω. $x(\gamma_{\mathcal{P}}(t))$ converges to the optimal solution of (36) as $t \to \infty$.

Let $s_{\mathcal{D}}(t)$ be the unique optimal solution to (38), and $\gamma_{\mathcal{D}}(t)$ be the point in Ω expressed as $s_{\mathcal{D}}(t) = s(\gamma_{\mathcal{D}}(t))$ in s-coordinate. We define the central trajectory $\gamma_{\mathcal{D}}$ of (35)/(37) as one dimensional submanifold $\gamma_{\mathcal{D}} = \{\gamma_{\mathcal{D}}(t) |\ t \in [0, \infty)\}$ in Ω. $s(\gamma_{\mathcal{D}}(t))$ converges to the optimal solution of (37) as $t \to \infty$.

Proposition 4.1. *$\gamma_{\mathcal{P}}$ and $\gamma_{\mathcal{D}}$ are represented as the intersection of two submanifolds:*

$$\gamma_{\mathcal{P}}(t) = \mathcal{P} \cap t\mathcal{D} \text{ and } \gamma_{\mathcal{P}} = \mathcal{P} \cap \mathrm{Hom}(\mathcal{D}), \tag{42}$$

and

$$\gamma_{\mathcal{D}}(t) = \mathcal{D} \cap t\mathcal{P} \text{ and } \gamma_{\mathcal{D}} = \mathcal{D} \cap \mathrm{Hom}(\mathcal{P}). \tag{43}$$

Proof. Let $p = \gamma_{\mathcal{P}}(t)$. Then $x(p)$ satisfies (39). Since $\bar{s}(x(p)) \in t\mathcal{D}^s$ and $x(p) \in \mathcal{P}$, the relation (42) follows from the definition of $\mathrm{Hom}(\mathcal{D})$. The relation (43) follows similarly. $\square$

The following proposition shows that $\gamma_{\mathcal{P}}(t)$ ($\gamma_{\mathcal{D}}(t)$) is also characterized as a convex optimization problem in s-coordinate (x-coordinate).

Proposition 4.2. *The following holds.*

1. $s(\gamma_{\mathcal{P}}(t))$ *is the optimal solution of the following problem:*

$$\min \; s^{\mathsf{T}} d + \psi^*(s) \quad s.t. \; s \in t\mathcal{D}^s. \tag{44}$$

2. $x(\gamma_{\mathcal{D}}(t))$ *is the optimal solution of the following problem:*

$$\min \; c^{\mathsf{T}} x + \psi(x) \quad s.t. \; x \in t\mathcal{P}^x.$$

Proof. We just prove the first relation. The optimality condition of (44) is:

$$-\bar{x}(s) + d \in T, \;\; s \in t\mathcal{D}^s,$$

which is equivalent to $\bar{x}(s) \in \mathcal{P}^x, \;\; s \in t\mathcal{D}^s$. Comparing this condition with (39), we obtain the result. □

In our framework, $\gamma_{\mathcal{P}}$ and $\gamma_{\mathcal{D}}$ are two different curves in Ω. But they are related to each other in the sense that, for any $t > 0$, $\gamma_{\mathcal{P}}(t)$ and $\gamma_{\mathcal{D}}(t)$ exist on the same ray in both x- and s-coordinate. We have the following proposition.

Proposition 4.3.

$$x(\gamma_{\mathcal{D}}(t)) = tx(\gamma_{\mathcal{P}}(t)) \;\; \textit{and} \;\; s(\gamma_{\mathcal{P}}(t)) = ts(\gamma_{\mathcal{D}}(t)). \tag{45}$$

Proof. We show the second relation of (45). Once this is done, the first one immediately follows from proposition 3.2(a). The point $s(\gamma_{\mathcal{P}}(t))$ satisfies that $\bar{x}(s) \in \mathcal{P}^x, \;\; s \in t\mathcal{D}^s$. Now, we let p' be a point such that $s(p') = s(\gamma_{\mathcal{P}}(t))/t$, and show that $p' = \gamma_{\mathcal{D}}(t)$. It is enough to check that $p' \in t\mathcal{P} \cap \mathcal{D}$. We have $p' \in \mathcal{D}$ since $s(p') = s(\gamma_{\mathcal{P}}(t))/t \in \mathcal{D}^s$. Since $ts(p') = s(\gamma_{\mathcal{P}}(t))$, we have $x(p') = tx(\gamma_{\mathcal{P}}(t))$. Since $x(\gamma_{\mathcal{P}}(t)) \in \mathcal{P}^x$, we have $x(p') \in t\mathcal{P}^x$ as we desire. □

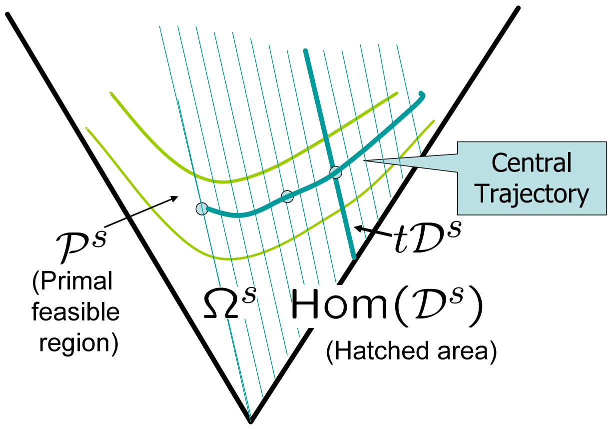

Figure 5: Central trajectory.

In the following, we assume that $\mathcal{D}$ is written as, through s-coordinate,

$$\mathcal{D} = \{p \in \Omega | \; s(p) = c - A^{\mathsf{T}} y, \;\; y \in \mathbb{R}^m\},$$

where $A \in \mathbb{R}^{m \times n}$ and the rows of A are linearly independent. On the other hand, for $b \in \mathbb{R}^m$ satisfying $Ad = -b$ we can express $\mathcal{P}$ as

$$\mathcal{P} = \{p \in \Omega | \; Ax(p) = b\}.$$

Now we derive the differential equation of the central trajectory $\gamma_{\mathcal{P}}$ written through s-coordinate. In the following, we let $x(t) = x(\gamma_{\mathcal{P}}(t))$ and $s(t) = s(\gamma_{\mathcal{P}}(t))$. The point $\gamma_{\mathcal{P}}(t)$ on the central trajectory $\gamma_{\mathcal{P}}$ in x-coordinate is the optimal solution of

$$\min \;\; tc^{\mathsf{T}} x + \psi(x) \text{ s.t. } Ax = b, \;\; x \in \Omega.$$

The optimality condition implies that

$$tc - s = A^\top y,\ Ax = b,\ s = -\frac{\partial \psi(x)}{\partial x}.$$

One more differentiation with respect to t yields that

$$c + G_x \dot{x} = A^\top \dot{y}, \quad A\dot{x} = 0.$$

Multiplying the first equation by G_x^{-1} and then A from left, we have $\dot{y} = (AG_x^{-1}A^\top)^{-1}AG_x^{-1}c$ and

$$\dot{s} = -G_x\dot{x} = (G_x - A^\top(AG_x^{-1}A^\top)^{-1}A)G_x^{-1}c = G_x(I - \Pi_x^\perp)G_x^{-1}c = (I - \Pi_s^\perp)c. \tag{46}$$

Observe that the rightmost hand side of (46) is well-defined for any $p \in \Omega$. Therefore, we consider the vector field $V^{\mathrm{ct}(\mathcal{P})}$ defined on Ω which is written, in s-coordinate, as the rightmost hand side of (46), i.e.,

$$V_s^{\mathrm{ct}(\mathcal{P})} := (I - \Pi_s^\perp)c.$$

The differential equation of the central trajectory $\gamma_{\mathcal{D}}$ in x-coordinate is written as:

$$\dot{x} = \Pi_x^\perp d.$$

The righthand side is well-defined over Ω and defines a vector field. We will denote this vector field as $V^{\mathrm{ct}(\mathcal{D})}$.

4.3 A Geometric Predictor-Corrector Algorithm and Curvature Integral

In [10], we developed a geometric predictor-corrector algorithm which follows the central trajectory $\gamma_{\mathcal{P}}$ in s-coordinate. We briefly outline the algorithm and explain how the complexity of the algorithm is related to the curvature integral. In the following, we denote $s(\gamma_{\mathcal{P}}(t))$ by $s_{\mathcal{P}}(t)$.

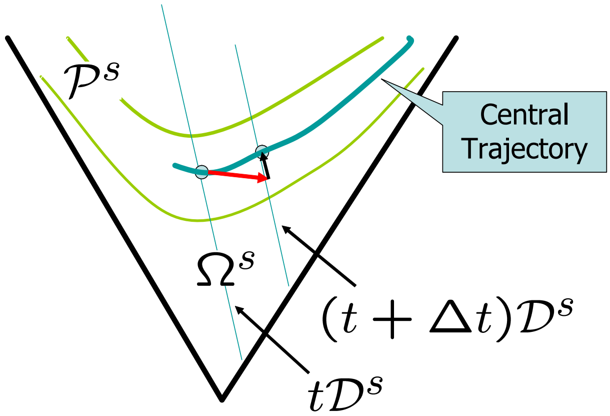

Figure 6: Path-following algorithm.

[Predictor-step]
The central trajectory $\gamma_{\mathcal{P}}$ which we follow is an integral curve of the vector field $V^{\mathrm{ct}(\mathcal{P})}$ we introduced in the last section. Therefore, we take $V^{\mathrm{ct}(\mathcal{P})}$ as the direction in the predictor step. Let $\hat{s} \in t\mathcal{D}^s$, and let

$$\hat{s}_L(t') := \hat{s} + (t' - t)(I - \Pi_{\hat{s}}^\perp)c. \tag{47}$$

Since $(I - \Pi_{\hat{s}}^\perp)c = c - A^\top y'$ for some $y' \in \mathbf{R}^m$, we have

$$\hat{s}_L(t') \in t'\mathcal{D} \text{ (as long as } s_L(t') \in \Omega^s). \tag{48}$$

As we observed in (48), we have $\hat{s}_L(t') \in t'\mathcal{D}^s$ We choose a step Δt and adopt $\hat{s}_L(t + \Delta t)$ as the result of the predictor step.

[Corrector-step and the Neighborhood of the Central Trajectory]
Let $\hat{s}' \in t'\mathcal{D}^s$. Recall that $s(\gamma_{\mathcal{P}}(t'))$ is characterized as the optimal solution to the problem (44) (with $t := t'$) and $\hat{s}'$ is the feasible solution to (44). The corrector step at $\hat{s}'$ is the Newton step for the point $\hat{s}'$ to solve the problem (44). Let $N_{\hat{s}'}$ be the Newton displacement vector for (44) at $\hat{s}' \in t'\mathcal{D}^s$. The Newton decrement is written as $\sqrt{N_{\hat{s}'}^{\mathsf{T}} G_{\hat{s}'} N_{\hat{s}'}}$. We introduce the neighborhood $\mathcal{N}_{t'}(\beta)$ of the center point $\gamma_{\mathcal{P}}(t')$ as the following subset on the slice $t'\mathcal{D}^s$:

$$\mathcal{N}_{t'}(\beta) := \{\hat{s} \in t'\mathcal{D}^s \mid \sqrt{N_{\hat{s}'}^{\mathsf{T}} G_{\hat{s}'} N_{\hat{s}'}} \leq \beta\}.$$

The neighborhood $\mathcal{N}(\beta)$ of the central trajectory is determined as

$$\mathcal{N}(\beta) := \cup_{t\in[0,\infty)}\mathcal{N}_t(\beta) = \{\hat{s} \in \mathrm{Hom}(\mathcal{D}) \mid \sqrt{N_{\hat{s}}^T G_{\hat{s}} N_{\hat{s}}} \leq \beta\}.$$

After the predictor-step is performed, we have

$$\bar{s}^{\mathrm{P}} := \bar{s}_L(t+\Delta t) \in (t+\Delta t)\mathcal{D} \cap \mathcal{N}(\beta).$$

As we discussed above, the corrector-step is the Newton step for the convex optimization problem (44) (with $t := t+\Delta t$) whose optimal solution is $s(\gamma_{\mathcal{P}}(t+\Delta t))$. The point $\bar{s}^{\mathrm{P}}$ is a feasible solution to this problem, and we apply a single step of the Newton method. This is the corrector step.

Now we are ready to describe the algorithm.

[A Predictor-Corrector Algorithm]

1. Let $\beta \leq 1/4$.
2. Let $s \in t\mathcal{D}$ such that $s \in \mathcal{N}(\frac{16}{9}\beta^2)$.
3. (Predictor step) Let $\Delta t > 0$ be such that
$$\sqrt{N_{s_L(t+\Delta t)}^{\mathsf{T}} G_{s_L(t+\Delta t)} N_{s_L(t+\Delta t)}} = \beta,$$
where $s_L(\cdot)$ is as defined in (47). Let $s^{\mathrm{P}} := s_L(t+\Delta t)$.
4. (Corrector step) At s^{P}, compute the Newton direction for the corrector-step $N_{s_L(t+\Delta t)}$ as above, and let $s^+ := s^{\mathrm{P}} + N_{s_L(t+\Delta t)}$.
5. $t := t+\Delta t,\ s := s^+$ and return to step 1.

[Asymptotic Iteration-complexity and the Curvature Integral]
Let

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

It can be shown that the algorithm follows the central trajectory from $\gamma_{\mathcal{P}}(t_1)$ to $\gamma_{\mathcal{P}}(t_2)$ $(t_1 < t_2)$ in $O(\sqrt{\theta}\log(t_2/t_1))$ iterations and the number of iterations is approximated with the integral $\frac{1}{\sqrt{\beta}}I_{\mathcal{P}}(t_1,t_2)$ involving the embedding curvature $H_{\mathcal{P}}^*$ as β goes to zero. In other words, we have

$$\lim_{\beta\to 0}\frac{\sqrt{\beta}\times \#_P(t_1,t_2,\beta)}{I_{\mathcal{P}}(t_1,t_2)} = 1, \tag{49}$$

where $\#_P(t_1,t_2,\beta)$ is the number of iterations of this algorithm to follow the central trajectory from $\gamma_{\mathcal{P}}(t_1)$ to $\gamma_{\mathcal{P}}(t_2)$ with the neighborhood opening β. See [10] for the proof. This is an asymptotic result which would hold by fixing t_1 and t_2 and then letting $\beta \to 0$.

The proof of (49) is outlined as follows. We consider an ideal algorithm with perfect centering for simplicity. Then, due to Step 3, if β is sufficiently small, then the Newton decrement is approximated well as

$$\beta^2 = N_{s_L(t+\Delta t)}^{\mathsf{T}} G_{s_L(t+\Delta t)} N_{s_L(t+\Delta t)} \sim (s_L(t+\Delta t) - s_{\mathcal{P}}(t+\Delta t))^{\mathsf{T}} G_{s_L(t+\Delta t)}(s_L(t+\Delta t) - s_{\mathcal{P}}(t+\Delta t)),$$

as the corrector-step at $s_{L(t+\Delta t)}$ is almost pointing $s_{\mathcal{P}}(t+\Delta t)$ (if β is sufficiently small.) We apply Taylor expansion to $s_{\mathcal{P}}(t+\Delta t)$ and ignore the second order of Δt. By using the fact that the change of the metric G is the order of Δt, we see that

$$\begin{aligned}(s_L(t+\Delta t)-s_{\mathcal{P}}(t+\Delta t))^{\mathsf{T}}G_{s_L(t+\Delta t)}(s_L(t+\Delta t)-s_{\mathcal{P}}(t+\Delta t)) &\sim \left(\frac{\Delta t^2}{2}\ddot{s}_{\mathcal{P}}\right)^{\mathsf{T}} G_{s(t)}\frac{\Delta t^2}{2}\ddot{s}_{\mathcal{P}} \\ &= \frac{\Delta t^4}{4}\|\nabla^*_{\dot{\gamma}_{\mathcal{P}}}\dot{\gamma}_{\mathcal{P}}\|^2_{\gamma_{\mathcal{P}}(t)} \sim \beta^2\end{aligned}$$

holds. Furthermore, as we will see in the lemme below, we have

$$\nabla^*_{\dot{\gamma}_{\mathcal{P}}}\dot{\gamma}_{\mathcal{P}} = H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}}). \tag{50}$$

Therefore,

$$\frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\|^{1/2}_{\gamma_{\mathcal{P}}(t)}\Delta t = \frac{\|\nabla^*_{\dot{\gamma}}\dot{\gamma}\|^{1/2}}{\sqrt{2}}\Delta t \sim \sqrt{\beta} \tag{51}$$

holds at one iteration of the predictor-corrector algorithm.

Taking summation of the both sides of (51) throughout the iterations, we see that the lefthand side is approximated with the integral and the righthand side is $\sqrt{\beta}$ multiplied by the number of iterations and therefore (49) follows.

In spite of its asymptotic nature, we consider that (49) is a relevant relation which connects directly the number of iterations of interior-point algorithms to the curvature structure of the central trajectory in a differential geometric framework.

Let

$$I_{\mathcal{D}}(t_1,t_2) := \frac{1}{\sqrt{2}}\int_{t_1}^{t_2}\|H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}},\dot{\gamma}_{\mathcal{D}})\|^{1/2}_{\gamma_{\mathcal{D}}(t)}\mathrm{d}t.$$

Then an analogous result holds for the predictor-corrector algorithm which follows γ_D. We call $I_{\mathcal{P}}$ and $I_{\mathcal{D}}$ *the curvature integrals* of the central trajectories.

In the end of this section, we prove (50). This relation plays an important role in connecting geometrical structure to computational complexity.

Lemma 4.4. *We have the following relations:*

(a) $\nabla^*_{\dot{\gamma}_{\mathcal{P}}}\dot{\gamma}_{\mathcal{P}} = H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})$, $\quad (H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}}))_s = -\Pi^{\perp}_s G_s^{-1}\dot{G}_s\dot{s}$.

(b) $\nabla_{\dot{\gamma}_{\mathcal{D}}}\dot{\gamma}_{\mathcal{D}} = H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}},\dot{\gamma}_{\mathcal{D}})$, $\quad (H^*_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}},\dot{\gamma}_{\mathcal{D}}))_x = -(I-\Pi^{\perp}_x)G_x^{-1}\dot{G}_x\dot{x}$.

Proof. We prove (a). The proof of (b) is similar. Let $s(t) = s(\gamma_{\mathcal{P}}(t))$. Recall that $(\nabla^*_{\dot{\gamma}_{\mathcal{P}}}\dot{\gamma}_{\mathcal{P}})_s = \ddot{s}$ (see (26)). Due to (46), $\nabla^*_{\dot{\gamma}_{\mathcal{P}}}\dot{\gamma}_{\mathcal{P}} = \ddot{s} = -\dot{\Pi}^{\perp}_s c$. We derive an expression for $\dot{\Pi^{\perp}}_s$ below. Recall that $\Pi^{\perp}_{s(p)} = A^{\mathsf{T}}(AG_{s(p)}A^{\mathsf{T}})^{-1}AG_{s(p)}$. (see (28)). Since

$$\begin{aligned}A^{\mathsf{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}(AG_sA^{\mathsf{T}})^{-1}\right)AG_s &= -A^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}(AG_sA^{\mathsf{T}})\right)(AG_sA^{\mathsf{T}})^{-1}AG_s \\ &= -\Pi^{\perp}_s G_s^{-1}\dot{G}\Pi^{\perp}_s,\end{aligned}$$

we have $\dot{\Pi^{\perp}}_s = \Pi^{\perp}_s G_s^{-1}\dot{G}_s(I-\Pi^{\perp}_s)$. Therefore,

$$(\nabla^*_{\dot{\gamma}_P}\dot{\gamma}_P)_s = \ddot{s} = -\Pi^{\perp}_s G_s^{-1}\dot{G}_s\dot{s}.$$

From this relation, it immediately follows that

$$(H^*_{\mathcal{P}}(\dot{\gamma}_P,\dot{\gamma}_P))_s = (\Pi^{\perp}\nabla^*_{\dot{\gamma}_P}\dot{\gamma}_P)_s = -\Pi^{\perp}_s\Pi^{\perp}_s G_s^{-1}\dot{G}_s\dot{s} = -\Pi^{\perp}_s G_s^{-1}\dot{G}_s\dot{s} = (\nabla^*_{\dot{\gamma}_P}\dot{\gamma}_P)_s.$$

This completes the proof of (a). The proof of (b) is similar by using that $\ddot{x} = \dot{\Pi}^{\perp}_x d$ and $\Pi^{\perp}_x = (\Pi^{\perp}_s)^{\mathsf{T}}$ and $\dot{G}_s G_s^{-1} = -G_x^{-1}\dot{G}_x$. $\square$

5 Pythagorean Relationship in SDP case

In this section, we derive the information geometric expression of the curvature integral (21) of primal-dual algorithms in SDP, proving the SDP part of the main theorem of this paper.

The vectorization operator vec is defined by $\mathrm{vec}\, M := (M_{*,1}, \ldots, M_{*,n})^T$ for $M \in \mathbb{R}^{n\times n}$, where $M_{*,j}$ denotes the jth column of the matrix M. Let $\mathcal{A}^T := (\mathrm{vec}\, A_1 \cdots \mathrm{vec}\, A_m)$, $x := \mathrm{vec}\, X$, $s := \mathrm{vec}\, S$. Also note that we use Jordan product $\circ$ in $\mathbb{S}^n$ for both matrix forms and vector forms in the same sense, i.e., $x \circ y = 1/2 \cdot \mathrm{vec}(XY + YX)$ for $\forall X, Y \in \mathbb{S}^n$.

Using the above notations, the central path equation (8) is rewritten as

$$\begin{cases} \mathcal{A}x = b, \\ \mathcal{A}^T y + s = c, \\ x \circ s = \nu e, \end{cases} \tag{52}$$

where $e := \mathrm{vec}\, I$. By differentiating (52) w.r.t. ν, we have

$$\begin{cases} \mathcal{A}\dot{x} = 0, \\ \mathcal{A}^T \dot{y} + \dot{s} = 0, \\ L_s \dot{x} + L_x \dot{s} = e, \end{cases} . \tag{53}$$

where $L_x := \frac{1}{2}(X \otimes I + I \otimes X)$ and $L_s := \frac{1}{2}(S \otimes I + I \otimes S)$. Then we have

$$\begin{cases} \dot{y} = -\left(\mathcal{A}L_S^{-1}L_X\mathcal{A}^T\right)^{-1}\mathcal{A}L_S^{-1}e, \\ \dot{s} = \mathcal{A}^T\left(\mathcal{A}L_S^{-1}L_X\mathcal{A}^T\right)^{-1}\mathcal{A}L_S^{-1}e, \\ \dot{x} = L_S^{-1}\left(e - L_X\mathcal{A}^T\left(\mathcal{A}L_S^{-1}L_X\mathcal{A}^T\right)^{-1}\mathcal{A}L_S^{-1}e\right). \end{cases} . \tag{54}$$

Noting that the relation $L_s^{-1}L_x = \nu(S^{-1} \otimes S^{-1})$ holds on the central path, it follows that

$$\begin{aligned} (X^{-1/2} \otimes X^{-1/2})\dot{x} &= (X^{-1/2} \otimes X^{-1/2})S^{-1}e \\ &\quad -(X^{-1/2} \otimes X^{-1/2})S^{-1}X\mathcal{A}^T\left(\mathcal{A}S^{-1}X\mathcal{A}^T\right)^{-1}\mathcal{A}S^{-1}e \\ &= \frac{1}{\nu}(S^{1/2} \otimes S^{1/2})S^{-1}e \\ &\quad -\frac{1}{\nu}(S^{1/2} \otimes S^{1/2})\nu(S^{-1} \otimes S^{-1})\mathcal{A}^T\left(\mathcal{A}\nu(S^{-1} \otimes S^{-1})\mathcal{A}^T\right)^{-1}\mathcal{A}S^{-1}e \\ &= \frac{1}{\nu}e - \frac{1}{\nu}(S^{-1/2} \otimes S^{-1/2})\mathcal{A}^T\left(\mathcal{A}(S^{-1/2} \otimes S^{-1/2})(S^{-1/2} \otimes S^{-1/2})\mathcal{A}^T\right)^{-1} \\ &\quad \cdot\mathcal{A}(S^{-1/2} \otimes S^{-1/2})e \\ &= \frac{1}{\nu}(\mathcal{I} - \mathcal{P})e, \\ (S^{-1/2} \otimes S^{-1/2})\dot{s} &= \frac{1}{\nu}(S^{-1/2} \otimes S^{-1/2})\mathcal{A}^T\left(\mathcal{A}(S^{-1/2} \otimes S^{-1/2})(S^{-1/2} \otimes S^{-1/2})\mathcal{A}^T\right)^{-1} \\ &\quad \cdot\mathcal{A}(S^{-1/2} \otimes S^{-1/2})e \\ &= \frac{1}{\nu}\mathcal{P}e, \end{aligned}$$

where $\mathcal{P}$ is a projection onto $\mathrm{Im}((S^{-1/2} \otimes S^{-1/2})\mathcal{A}^T)$.

Therefore, by changing the variable $t = 1/\nu$ and letting $t_1 := 1/\nu_f$ and $t_2 := 1/\nu_i$, we obtain

$$\begin{aligned} I_{PD}(1/\nu_f, 1/\nu_i) &= \int_{1/\nu_f}^{1/\nu_i} \left\|(X^{-1/2} \otimes X^{-1/2})\dot{x} \circ (S^{-1/2} \otimes S^{-1/2})\dot{s}\right\|_F^{1/2} \mathrm{d}\nu \\ &= \int_{1/\nu_f}^{1/\nu_i} \frac{1}{\nu}\|(\mathcal{I} - \mathcal{P})e \circ \mathcal{P}e\|_2^{1/2}\mathrm{d}\nu \\ &= \int_{t_2}^{t_1} \frac{1}{t}\|\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e\|_2^{1/2}\mathrm{d}t. \end{aligned}$$

The first equality is the main result of [3]. Thus, we have the square of integrand of I_{PD} by h_{PD}:

$$h_{PD}(t) = \frac{1}{t^2}\|\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e\|_2. \tag{55}$$

Then, the following geometrical relationship of iteration complexities between the primal-dual algorithms and primal/dual algorithms holds.

Theorem 5.1 (Pythagorean Relationship).

$$[h_{PD}(t)]^2 = \left\|\frac{1}{2}H_P^*(\dot\gamma_P(t), \dot\gamma_P(t))\right\|^2_{\gamma_P(t)} + \left\|\frac{1}{2}H_D(\dot\gamma_D(t), \dot\gamma_D(t))\right\|^2_{\gamma_D(t)}. \tag{56}$$

In integral forms, I_{PD} is expressed in terms of information geometry as

$$\int_{t_2}^{t_1} h_{PD}(t)^{1/2}\mathrm{d}t = \int_{t_2}^{t_1}\left[\left\|\frac{1}{2}H_P^*(\dot\gamma_P(t), \dot\gamma_P(t))\right\|_{\gamma_P(t)} + \left\|\frac{1}{2}H_D(\dot\gamma_D(t), \dot\gamma_D(t))\right\|_{\gamma_D(t)}\right]^{1/4}\mathrm{d}t.$$

Proof. Since $G_s = S^{-1}\otimes S^{-1}$, $\Pi_s^\perp = \mathcal{A}^T(\mathcal{A}G_s\mathcal{A}^T)^{-1}\mathcal{A}G_s = G_s^{-1/2}\mathcal{P}G_s^{1/2}$, and $(\mathcal{I}-\mathcal{P})G_s^{1/2}(tc - s(\gamma_P(t))) = 0$, we have $G_s^{1/2}\dot s = G_s^{1/2}(\mathcal{I}-\Pi_s^\perp)c = (\mathcal{I}-\mathcal{P})G_s^{1/2}c = t^{-1}(\mathcal{I}-\mathcal{P})e$. Note that $\dot s$ and all the differential operations in this proof are done in terms of t. Also we have

$$\dot G_s = (S^{-1}\otimes S^{-1})' = -(S^{-1}\otimes S^{-1})(\dot S\otimes S + S\otimes \dot S)(S^{-1}\otimes S^{-1}).$$

Therefore, we obtain

$$\begin{aligned}
(H_P^*(\dot\gamma_P, \dot\gamma_P))_s &= -\Pi_s^\perp G_s^{-1}\dot G_s \dot s\\
&= G_s^{-1/2}\mathcal{P}(S^{-1/2}\otimes S^{-1/2})(\dot S\otimes S + S\otimes\dot S)(S^{-1/2}\otimes S^{-1/2})t^{-1}(\mathcal{I}-\mathcal{P})e\\
&= t^{-1}G_s^{-1/2}\mathcal{P}(S^{-1/2}\dot S S^{-1/2}\otimes I + I\otimes S^{-1/2}\dot S S^{-1/2})(\mathcal{I}-\mathcal{P})e\\
&= \frac{2}{t^2}G_s^{-1/2}\mathcal{P}((\mathcal{I}-\mathcal{P})e\circ(\mathcal{I}-\mathcal{P})e)\\
&= -\frac{2}{t^2}G_s^{-1/2}\mathcal{P}(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e).
\end{aligned}$$

The last equation is due to the fact that

$$\mathcal{P}(e\circ(\mathcal{I}-\mathcal{P})e) = \mathcal{P}(\mathcal{I}-\mathcal{P})e = 0.$$

Thus, we have the embedding curvature of the primal problem as

$$\frac{1}{2}\|H_P^*(\dot\gamma_P, \dot\gamma_P)\|_{\gamma_P} = \frac{1}{2}\sqrt{(H_P^*(\dot\gamma_P, \dot\gamma_P))_s^T G_s (H_P^*(\dot\gamma_P, \dot\gamma_P))_s} = \left\|\frac{1}{t^2}\mathcal{P}(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e)\right\|_2. \tag{57}$$

Similarly, on the central path, the following identities hold: $\Pi_x^\perp = G_x^{-1}\mathcal{A}^T(\mathcal{A}G_x^{-1}\mathcal{A}^T)^{-1}\mathcal{A} = G_x^{-1/2}\mathcal{P}G_x^{1/2}$, $G_s^{1/2}\dot x = t^{-1}G_s^{1/2}\Pi_x^\perp td = t^{-1}\mathcal{P}G_x^{-1/2}(x+z) = t^{-1}\mathcal{P}e$ where $z\in\ker\mathcal{A}$. Therefore, we have

$$\begin{aligned}
(H_D(\dot\gamma_D, \dot\gamma_D))_x &= -(I-\Pi_x^\perp)G_x^{-1}\dot G_x\dot x\\
&= -(\mathcal{I}-G_x^{-1/2}\mathcal{P}G_x^{1/2})G_x^{-1}(G_x^{-1})'\dot x\\
&= -G_x^{-1/2}(\mathcal{I}-\mathcal{P})G_x^{1/2}G_x^{-1}G_x(\dot X\otimes X + X\otimes\dot X)G_x^{1/2}G_x^{1/2}\dot x\\
&= G_x^{-1/2}(\mathcal{I}-\mathcal{P})(X^{-1/2}\dot X X^{-1/2}\otimes I + I\otimes X^{-1/2}\dot X X^{-1/2})t^{-1}\mathcal{P}e\\
&= \frac{2}{t^2}G_x^{-1/2}(\mathcal{I}-\mathcal{P})(\mathcal{P}e\circ\mathcal{P}e)\\
&= -\frac{2}{t^2}G_x^{-1/2}(\mathcal{I}-\mathcal{P})(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e).
\end{aligned}$$

The last equation is due to the fact that

$$(\mathcal{I}-\mathcal{P})(e\circ\mathcal{P}e). = (\mathcal{I}-\mathcal{P})\mathcal{P}e = 0.$$

Thus, we have the embedding curvature of the dual problem as

$$\frac{1}{2}\left\|H_D(\dot{\gamma}_D,\dot{\gamma}_D)\right\|_{\gamma_D} = \frac{1}{2}\sqrt{\left(H_D(\dot{\gamma}_D,\dot{\gamma}_D)\right)_x^T G_x \left(H_D(\dot{\gamma}_D,\dot{\gamma}_D)\right)_x} = \left\|\frac{1}{t^2}(\mathcal{I}-\mathcal{P})(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e)\right\|_2. \quad (58)$$

Since $\mathcal{P}$ is a projection, it is easy to verify the following equation:

$$\|\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e\|_2^2 = \|(\mathcal{I}-\mathcal{P})\left(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e\right)\|_2^2 + \|\mathcal{P}\left(\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e\right)\|_2^2. \quad (59)$$

Lastly, combining (55), (57), (58) and (59), we obtain (56). □

6 Pythagorean Relationship in Symmetric Cone Programs

In this section, we discuss the geometric relationship of iteration complexities for predictor-corrector type methods between primal/dual algorithms and primal-dual algorithms in symmetric cone programs. This result includes the ones in LP and SDP as special cases.

By differentiating the central path equation (19) w.r.t. ν, we have

$$\left\{\begin{array}{l} \mathcal{A}\dot{x}=0,\\ \mathcal{A}^T\dot{y}+\dot{s}=0,\\ L_s\dot{x}+L_x\dot{s}=e. \end{array}\right. \quad (60)$$

By solving this system of equations w.r.t. $(\dot{x},\dot{y},\dot{s})$, we have

$$\left\{\begin{array}{l} \dot{y} = -\left(\mathcal{A}L_s^{-1}L_x\mathcal{A}^T\right)^{-1}\mathcal{A}L_s^{-1}e,\\ \dot{s} = \mathcal{A}^T\left(\mathcal{A}L_s^{-1}L_x\mathcal{A}^T\right)^{-1}\mathcal{A}L_s^{-1}e,\\ \dot{x} = L_s^{-1}\left(e - L_x\mathcal{A}^T\left(\mathcal{A}L_s^{-1}L_x\mathcal{A}^T\right)^{-1}\mathcal{A}L_s^{-1}e\right). \end{array}\right. \quad (61)$$

Since the following identities hold on the central path: $L_s^{-1}L_x = \nu Q_s^{-1}$, $L_s^{-1}e = Q_s^{-1/2}e$, $G_x = Q_{x^{-1}} = \frac{1}{\nu^2}Q_s$, $G_s = Q_{s^{-1}} = \frac{1}{\nu^2}Q_x$, we have

$$\begin{aligned} G_s^{1/2}\dot{s} &= G_s^{1/2}\mathcal{A}^T\left(\mathcal{A}L_s^{-1}L_x\mathcal{A}^T\right)^{-1}\mathcal{A}L_s^{-1}e\\ &= Q_s^{-1/2}\mathcal{A}^T\left(\mathcal{A}\nu Q_s^{-1}\mathcal{A}^T\right)^{-1}\mathcal{A}Q_s^{-1/2}e\\ &= \frac{1}{\nu}\mathcal{P}e, \end{aligned}$$

and

$$\begin{aligned} G_x^{1/2}\dot{x} &= G_x^{1/2}L_s^{-1}\left(e - L_x\mathcal{A}^T\left(\mathcal{A}L_s^{-1}L_x\mathcal{A}^T\right)^{-1}\mathcal{A}L_s^{-1}e\right)\\ &= \frac{1}{\nu}Q_s^{1/2}L_s^{-1}\left(e - L_x\mathcal{A}^T\left(\mathcal{A}\nu Q_s^{-1}\mathcal{A}^T\right)^{-1}\mathcal{A}Q_s^{-1/2}e\right)\\ &= \frac{1}{\nu}\left(Q_s^{1/2}Q_s^{-1/2}e - Q_s^{1/2}\nu Q_s^{-1}\mathcal{A}^T\left(\mathcal{A}\nu Q_s^{-1}\mathcal{A}^T\right)^{-1}\mathcal{A}Q_s^{-1/2}e\right)\\ &= \frac{1}{\nu}(\mathcal{I}-\mathcal{P})e, \end{aligned}$$

where $\mathcal{P}$ is the projection onto $\operatorname{Im} Q_s^{-1/2}\mathcal{A}^T$. Then using the identity $G_x^{1/2}\dot{x}\circ G_s^{1/2}\dot{s} = Q_{x^{-1/2}}\dot{x}\circ Q_{s^{-1/2}}\dot{s}$ Therefore, by changing the variable $t=1/\nu$ and letting $t_1 := 1/\nu_f$ and $t_2 := 1/\nu_i$, we have

$$\begin{aligned} I_{PD}(1/\nu_f,1/\nu_i) &= \int_{1/\nu_f}^{1/\nu_i}\left\|G_x^{1/2}\dot{x}\circ G_s^{1/2}\dot{s}\right\|^{1/2}\,\mathrm{d}\nu\\ &= \int_{1/\nu_f}^{1/\nu_i}\frac{1}{\nu}\left\|(\mathcal{I}-\mathcal{P})e\circ\mathcal{P}e\right\|^{1/2}\,\mathrm{d}\nu\\ &= \int_{t_2}^{t_1}\frac{1}{t}\left\|\mathcal{P}e\circ(\mathcal{I}-\mathcal{P})e\right\|^{1/2}\,\mathrm{d}t. \end{aligned}$$

Thus we have the square of integrand of I_{PD}:

$$h_{PD}(t) = \frac{1}{t^2} \|\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e\| . \tag{62}$$

Then we have the final theorem in this paper.

Theorem 6.1 (Pythagorean Theorem).

$$[h_{PD}(t)]^2 = \left\| \frac{1}{2} H_P^*(\dot{\gamma}_P(t), \dot{\gamma}_P(t)) \right\|_{\gamma_P}^2 + \left\| \frac{1}{2} H_D(\dot{\gamma}_D(t), \dot{\gamma}_D(t)) \right\|_{\gamma_D}^2 . \tag{63}$$

Integral forms, we have

$$\int_{t_2}^{t_1} I_{PD}(t)\mathrm{d}t = \int_{t_2}^{t_1} \left[\left\| \frac{1}{2} H_P^*(\dot{\gamma}_P(t), \dot{\gamma}_P(t)) \right\|_{\gamma_P}^2 + \left\| \frac{1}{2} H_D(\dot{\gamma}_D(t), \dot{\gamma}_D(t)) \right\|_{\gamma_D}^2 \right]^{1/4} \mathrm{d}t.$$

Proof. Since $tc - s(\gamma_P(t)) \in \mathrm{Im}(\mathcal{A}^T)$ and $(\mathcal{I} - \mathcal{P})G_s^{1/2}\mathcal{A}^T = 0$, we have

$$(\mathcal{I} - \mathcal{P})G_s^{1/2}(tc - s(\gamma_P(t))) = 0.$$

Using the above fact and identity: $\Pi_s^{\perp} = \mathcal{A}^T(\mathcal{A}L_s^{-1}\mathcal{A}^T)^{-1}\mathcal{A}Q_s^{-1} = G_s^{-1/2}\mathcal{P}G_s^{1/2}$, we have

$$\begin{aligned} Q_s^{-1/2}\dot{s} &= t^{-1}G_s^{1/2}(\mathcal{I} - \Pi_s^{\perp})tc \\ &= t^{-1}(\mathcal{I} - \mathcal{P})G_s^{1/2}tc = t^{-1}(\mathcal{I} - \mathcal{P})e. \end{aligned}$$

Noting the fact that

$$\begin{aligned} Q_s^{-1/2}Q_s'Q_s^{-1/2} &= Q_s^{-1/2}\, 2Q_{s,\dot{s}}\, Q_s^{-1/2} = 2Q_s^{-1/2}Q_s^{1/2}L_{\left(Q_{s^{1/2}}^{-1}\right)\dot{s}}Q_s^{1/2}Q_s^{-1/2} \\ &= 2L_{Q_{s^{1/2}}^{-1}\dot{s}} = 2L_{\frac{(\mathcal{I}-\mathcal{P})e}{t}} = \frac{2}{t}L_{(\mathcal{I}-\mathcal{P})e}, \end{aligned}$$

it follows that

$$\begin{aligned} (H_P^*(\dot{\gamma}_P, \dot{\gamma}_P))_{\gamma_P(t)} &= -\Pi_s^{\perp}Q_s\left(Q_s^{-1}\right)'\dot{s} \\ &= -(Q_{s^{1/2}}\mathcal{P}Q_{s^{-1/2}})Q_s(-Q_s^{-1}Q_s'Q_s^{-1})\dot{s} \\ &= Q_s^{1/2}\mathcal{P}(Q_s^{-1/2}Q_s'Q_s^{-1/2})(Q_s^{-1/2}\dot{s}) \\ &= \frac{1}{t}Q_s^{1/2}\mathcal{P}\left(\frac{2}{t}L_{\mathcal{I}-\mathcal{P}e}\right)(t^{-1}(\mathcal{I} - \mathcal{P})e) \\ &= \frac{2}{t^2}G_s^{-1/2}\mathcal{P}((\mathcal{I} - \mathcal{P}e) \circ (\mathcal{I} - \mathcal{P})e) \\ &= -\frac{2}{t^2}G_s^{-1/2}\mathcal{P}(\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e). \end{aligned}$$

Therefore, we have the embedding curvature of the primal problem as

$$\frac{1}{2} \|H_P^*(\dot{\gamma}_P, \dot{\gamma}_P)\|_{\gamma_P} = \left\| \frac{1}{t^2}\mathcal{P}(\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e) \right\|_2 . \tag{64}$$

In the similar way, noting the identities on the central path: $\Pi_x^{\perp} = G_x^{-1}\mathcal{A}^T(\mathcal{A}G_x^{-1}\mathcal{A}^T)^{-1}\mathcal{A} = G_x^{-1/2}\mathcal{P}G_x^{1/2}$,

$Q_x^{-1/2}\dot{x} = t^{-1}Q_x^{-1/2}\Pi_x^{\perp}td = t^{-1}\mathcal{P}Q_x^{-1/2}(x+z) = t^{-1}\mathcal{P}e$ where $z \in \ker\mathcal{A}$, we have

$$\begin{aligned}
(H_D(\dot{\gamma}_D, \dot{\gamma}_D))_{\gamma_D} &= -(I - \Pi_x^{\perp})G_x^{-1}\dot{G}_x\dot{x} \\
&= -(\mathcal{I} - Q_x^{1/2}\mathcal{P}Q_x^{-1/2})Q_x(Q_x^{-1})'\dot{x} \\
&= -Q_x^{1/2}(\mathcal{I} - \mathcal{P})Q_x^{-1/2}Q_x(-Q_x^{-1}Q_x'Q_x^{-1})\dot{x} \\
&= Q_x^{1/2}(\mathcal{I} - \mathcal{P})(Q_x^{-1/2}Q_x'Q_x^{-1/2})(Q_x^{-1/2}\dot{x}) \\
&= Q_x^{1/2}(\mathcal{I} - \mathcal{P})\left(\frac{2}{t}L_{\mathcal{P}e}\right)(t^{-1}\mathcal{P}e) \\
&= \frac{2}{t^2}G_x^{-1/2}(\mathcal{I} - \mathcal{P})\,(\mathcal{P}e \circ \mathcal{P}e) \\
&= -\frac{2}{t^2}G_x^{-1/2}(\mathcal{I} - \mathcal{P})\,(\mathcal{P}e \circ (I - \mathcal{P})e)\,.
\end{aligned}$$

Therefore, we have the embedding curvature of the dual problem as

$$\frac{1}{2}\left\|H_D(\dot{\gamma}_D, \dot{\gamma}_D)\right\|_{\gamma_D} = \left\|\frac{1}{t^2}(\mathcal{I} - \mathcal{P})(\mathcal{P}e \circ (\mathcal{I} - \mathcal{P})e)\right\|_2. \tag{65}$$

Since $\mathcal{P}$ is a projection, the relation (59) again holds here. Thus, combining (62), (64), (65) and (59), we obtain (63). ☐

7 Conclusions

In this paper, we extends the following geometrical relationship in linear programming, originally developed in [10], to those in SDP and symmetric cone programs:

$$[h_{PD}(t)]^2 = \left\|\frac{1}{2}H_P^*(\dot{\gamma}_P(t), \dot{\gamma}_P(t))\right\|_{\gamma_P}^2 + \left\|\frac{1}{2}H_D(\dot{\gamma}_D(t), \dot{\gamma}_D(t))\right\|_{\gamma_D}^2, \tag{66}$$

where $h_{PD}(t)$ is the square of the integrand of curvature integrals in primal-dual algorithms with changing variable $t = 1/\nu$ discussed in section 2.3 and $1/\sqrt{2}\cdot\|H_P^*(\dot{\gamma}_P(t), \dot{\gamma}_P(t))\|_{\gamma_P}^{1/2}$ represents the integrand of the curvature integral in primal algorithms and $1/\sqrt{2}\cdot\|H_D(\dot{\gamma}_D(t), \dot{\gamma}_D(t))\|_{\gamma_D}^{1/2}$ represents the integrand of the curvature integrals in dual algorithms. In integral forms, above relation is translated into

$$\int_{t_2}^{t_1} I_{PD}(t)\mathrm{d}t = \int_{t_2}^{t_1}\left[\left\|\frac{1}{2}H_P^*(\dot{\gamma}_P(t), \dot{\gamma}_P(t))\right\|_{\gamma_P}^2 + \left\|\frac{1}{2}H_D(\dot{\gamma}_D(t), \dot{\gamma}_D(t))\right\|_{\gamma_D}^2\right]^{1/4}\mathrm{d}t.$$

References

[1] S.-i. Amari and H. Nagaoka. *Methods of information geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2000. Translated from the 1993 Japanese original by Daishi Harada.

[2] L. Faybusovich. Linear systems in Jordan algebras and primal-dual interior-point algorithms. *J. Comput. Appl. Math.*, 86(1):149–175, 1997. Special issue dedicated to William B. Gragg (Monterey, CA, 1996).

[3] S. Kakihara, A. Ohara, and T. Tsuchiya. Curvature integrals and iteration complexities in SDP and symmetric cone programs. *Optimization Online*, 2011.

[4] N. Karmarkar. Riemannian geometry underlying interior-point methods for linear programming. In *Mathematical developments arising from linear programming (Brunswick, ME, 1988)*, volume 114 of *Contemp. Math.*, pages 51–75. Amer. Math. Soc., Providence, RI, 1990.

[5] R. D. C. Monteiro and T. Tsuchiya. A strong bound on the integral of the central path curvature and its relationship with the iteration-complexity of primal-dual path-following LP algorithms. *Math. Program.*, 115(1, Ser. A):105–149, 2008.

[6] R. D. C. Monteiro and Y. Zhang. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Program.*, 81(3, Ser. A):281–299, 1998.

[7] M. Muramatsu. On a commutative class of search directions for linear programming over symmetric cones. *J. Optim. Theory Appl.*, 112(3):595–625, 2002.

[8] H. Nagaoka and S.-i. Amari. Differential geometry of smooth families of probability distributions. *Mathematical Engineering Technical Reports, Department of Mathematical Engineering and Instrumentation Physics, University of Tokyo*, 1982.

[9] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[10] A. Ohara and T. Tsuchiya. An information geometric approach to polynomial-time interior-point algorithms: Complexity bound via curvature integral. *Research Memorandum The Institute of Statistical Mathematics*, 1055, 2007.

[11] S. H. Schmieta and F. Alizadeh. Extension of primal-dual interior point algorithms to symmetric cones. *Math. Program.*, 96(3, Ser. A):409–438, 2003.

[12] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation. II. *Math. Program.*, 52(3, Ser. B):527–553 (1992), 1991. Interior point methods for linear programming: theory and practice (Scheveningen, 1990).

[13] G. Zhao and J. Stoer. Estimating the complexity of a class of path-following methods for solving linear programs by curvature integrals. *Appl. Math. Optim.*, 27(1):85–103, 1993.