Information Geometry and Primal-Dual Interior-point Algorithms

Satoshi Kakihara * Atsumi Ohara † Takashi Tsuchiya‡

November 2009

Abstract

In this paper, we study polynomial-time interior-point algorithms in view of information geometry. We introduce an information geometric structure for a conic linear program based on a self-concordant barrier function. Riemannian metric is defined with the Hessian of the barrier function. We introduce two connections ∇ and ∇^* which roughly corresponds to the primal and the dual problem. The dual feasible region is embedded in the primal cone and thus we consider the primal and dual problems in the same space. Characterization of the central trajectory and its property in view of the curvature is studied. A predictor-corrector primal path-following algorithm is represented based on this geometry and (its asymptotic) iteration-complexity is related to an integral involving the embedding curvature. Then we focus on the classical linear program and primal-dual algorithm. We will study an integral over the central trajectory which represents the number of iterations of the Mizuno-Todd-Ye predictor-corrector (MTY-PC) algorithm. We will show that this integral admits a rigorous differential geometric expression involving the embedding curvature. Connecting this expression to an integral bound previously obtained by Monteiro and Tsuchiya in relation to the layered interior-point algorithm by Vavasis and Ye, we prove a global geometric theorem on the central trajectory. Finally, we demonstrate that the integral value by itself provides an accurate estimate of the number of iterations of the MTY-PC algorithm through numerical experiments with fairly large practical instances from Netlib problems such as DFL001 and PILOT87. This leads to an interesting conclusion: *the number of iterations of the standard primal-dual algorithm is the value of a differential geometric curvature integral over the central trajectory.* This paper is a revised version of the paper "A. Ohara and T. Tsuchiya: an imformation geometric approach to interior-point algorithms: complexity estimate via curvature integral (December, 2007)."

Key words: interior-point methods, information geometry, polynomial-time algorithm, linear programming, semidefinite programming, embedding curvature, computational complexity, differential geometry, convex programming

1 Introduction

1.1 Setting and background

Let $\mathbf{E}$ be an n-dimensional vector space, let $\Omega \subset \mathbf{E}$ be a proper open convex cone, and let $\mathbf{E}^*$ be the space of linear functional on $\mathbf{E}$. We denote by $\langle \cdot, \cdot \rangle$ the duality product between elements in $\mathbf{E}$ and

*Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. (Email: Satoshi_Kakihara@mist.i.u-tokyo.ac.jp).

†Graduate School of Engineering Science, Osaka University, 1-3 Machikane-yama, Toyonaka, Osaka 560-8531, Japan. (Email: ohara@sys.es.osaka-u.ac.jp).

‡The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8562, Japan. (Email: tsuchiya@ism.ac.jp). This author is supported in part with Grant-in-Aid for Scientific Research (B), 2007, 19360192 and with Grant-in-Aid for Scientific Research (B), 2008, 20340024 from the Japan Society for the Promotion of Sciences.

$\mathbf{E}^*$. Let $\Omega^* \subset \mathbf{E}^*$ be the open dual cone $\{s \in \mathbf{E}^* \mid \langle s, x\rangle > 0,\ \forall x \in \mathrm{cl}(\Omega)\backslash\{0\}\}$ of Ω. In regard to Ω and Ω^*, we consider the following dual pair of convex linear programs:

$$\min\ \langle c, x\rangle, \quad \text{s.t. } x \in (d + \mathbf{T}) \cap \mathrm{cl}(\Omega) \tag{1}$$

and

$$\min\ \langle s, d\rangle, \quad \text{s.t. } s \in (c + \mathbf{T}^*) \cap \mathrm{cl}(\Omega^*), \tag{2}$$

where $c \in \mathbf{E}^*$, $d \in \mathbf{E}$, and $\mathbf{T} \subset \mathbf{E}$ is $(n-m)$-dimensional linear subspace, and $\mathbf{T}^* \subset \mathbf{E}^*$ is m-dimensional linear subspace. Note that $\mathbf{T}$ and $\mathbf{T}^*$ are mutually "orthogonal", i.e., for any $x \in \mathbf{T}$ and $s \in \mathbf{T}^*$, we have $\langle s, x\rangle = 0$. This is a generic framework of conic linear programming including classical linear programming and semidefinite programming.

The theory of polynomial-time interior-point algorithms for convex optimization established by Nesterov and Nemirovski [21] and information geometry by Amari and Nagaoka [1, 2, 3, 20], are among several innovative developments in computational mathematics and mathematical informatics in the last two decades. While interior-point methods brought a relevant and fruitful progress to the world of optimization in both theoretical and practical aspects, information geometry provided a generic differential geometric framework in studying the various disciplines dealing with information and uncertainty, including statistics, information theory, learning theory, control theory, signal processing, etc. Recently, Ohara and Tsuchiya [28] developed a theory aiming at establishing a direct connection between these two theories. This paper is a revised version of the original paper [28], with more emphasis on the primal-dual interior-point algorithms, plus additional new numerical experiments on the Netlib LP instances.

This paper consists of three parts. First we introduce information geometry and present our differential geometric framework. We consider an information geometric structure for a conic linear program based on a self-concordant barrier function. Riemannian metric is defined with the Hessian of the barrier. We also introduce mutually dual two connections ∇ and ∇^* based on the gradient map. These connections roughly correspond to the primal and the dual problems. The dual feasible region is embedded in the primal cone and thus we consider the primal and dual problem in the same space. Based on this framework, characterization of the central trajectory and its curvature is studied. A predictor-corrector primal path-following algorithm based on this geometry is represented and its iteration-complexity is related to an integral involving the embedding curvature over the central trajectory in an asymptotic sense.

In the second part, we focus on classical linear programs and an integral over the central trajectory which represents the number of iterations of the Mizuno-Todd-Ye primal-dual predictor-corrector (MTY-PC) algorithm. This integral was firstly introduced by Sonnevend, Stoer and Zhao [29] and later studied in detail by Stoer and Zhao [43] (see also [42]) and Monteiro and Tsuchiya [19]. We

will show that this integral is expressed exactly as a differential geometric quantity involving the curvature under our framework. This characterization is a merit of adapting information geometry rather than the traditional Riemannian geometry. We link this result to Vavasis and Ye's layered-step interior-point algorithm [41] whose complexity just depends on A (but neither b nor c) through aforementioned Monteiro and Tsuchiya's work, and present a *global theorem* on the central trajectory. To our best knowledge, this is one of the few global theorems in information geometry, and probably the first one to connect an algorithmic theory and differential geometry.

In the third part, we demonstrate through numerical experiments that the number of iterations of the MTY-PC algorithm with the neighborhood of opening up to one agrees quite well with the aforementioned integral by solving fairly large practical instances such as DFL001 Netlib problem (dimension of A: 6072×12230). In other words, we establish the following statement connecting algorithmic complexity and differential geometry: *The number of iterations of the standard interior-point algorithm is the value of a differential geometric curvature integral on the central trajectory.*

1.2 Differential geometric approach to interior-point algorithms

One of the first literatures of differential geometric approach to interior-point algorithms is due to Bayer and Lagarias [4, 5]. Karmarkar [10] suggested a fundamental idea of analyzing the interior-point algorithm based on Riemannian geometry and the central trajectory curvature integral. After Nesterov and Nemirovski's seminal work [21], geometrical studies of interior-point method were conducted by several authors including [7, 23, 22] in the context of general convex programs, where structure of geodesic and Riemannian curvature, etc, are studied based on the Riemannian geometrical structure defined by the Hessian of the self-concordant barrier function as the Riemannian metric.

In particular, the paper [23] provides an estimate of the iteration-complexity of the short-step primal interior-point algorithm based on an integral over the central trajectory, namely, the number of iterations of the short-step algorithm is essentially estimated by "the total length" of the portion of the central trajectories with respect to the Riemannian metric. They also demonstrate a certain near-optimality of the primal-dual central trajectory in the sense that the trajectory is $\sqrt{2}$-geodesic in the Riemannian space defined by the direct product of the primal and dual problem and their associated self-concordant barriers. We note that their results are for the short-step method which, roughly, updates the homotopy parameter with a fixed fraction. We will work with the predictor-corrector algorithm which updates the homotopy parameter adaptively and express the iteration-complexity with curvature based on information geometry.

The connection between information geometry and interior-point algorithms was firstly pointed out in Tanabe and Tsuchiya [36], in the context of linear programming. One of the key observations here is that the trajectory appearing in the primal interior-point method is a "straight line" under

the gradient map of the logarithmic barrier function. This fact is observed by Bayer and Lagarias [4, 5], and Tanabe [32, 35]. Tanabe further studied fiber structure of primal and dual problems in linear programming from information geometric viewpoint [34] (see also [35]). Subsequently, Ohara introduced an information geometric framework for symmetric positive semidefinite matrices defined by the logarithmic determinant function. See, for example, [27] for details. In [24, 25], Ohara studied a class of the directly solvable semidefinite programs in view of Euclidean Jordan algebra and doubly autoparallel submanifold in information geometry [26, 40]. He further developed a predictor-corrector type path-following algorithm for semidefinite programming based on information geometry. A generalization of this algorithm is developed in [28] and reviewed in this paper.

1.3 Main Results

Now we outline the structure and the main results of this paper. Sections 2 is a preliminary section to introduce information geometry and the Nesterov-Nemirovski self-concordant barrier. In particular, the information geometric structure of the cone Ω equipped with a θ-normal barrier function $\psi(x)$ is described.

In Section 3, we introduce information geometric structure of the dual pair of conic linear programs (1) and (2) based on θ-normal barrier ψ on the domain Ω. This structure is summarized as follows. The Riemannian metric is defined as the Hessian of ψ. The gradient map $\bar{s}(x) = -\partial\psi(x)/\partial x$ defines a one-to-one mapping from Ω to Ω^*. Thus, $\bar{s}(\cdot)$ defines another coordinate on Ω and under this coordinate Ω can be regarded as Ω^*. Therefore, we identify Ω with Ω^* through the gradient map and see them as a unique manifold with two distinct coordinate systems. The interior $\mathcal{P}$ of the primal feasible region and the interior $\mathcal{D}$ of the dual feasible region are naturally regarded as submanifolds of Ω.

We also introduce mutually dual two connections ∇ and ∇^*. Roughly, a connection is a structure which determines straightness of a curve and flatness of a submanifold. The analogues of straight line and affine space on a manifold are called geodesic and autoparallel submanifold, respectively. In our case, ∇-geodesic is a straight line in the original coordinate, and ∇^*-geodesic is a "straight line" in the gradient coordinate, or equivalently, in Ω^*. Furthermore, $\mathcal{P}$ is a ∇-autoparallel submanifold, and $\mathcal{D}$ is a ∇^*-autoparallel submanifold.

Let $x_{\mathcal{P}}(t)$ and $s_{\mathcal{D}}(t)$ be the points on the central trajectories of (1) and (2) with parameter t as defined by Nesterov and Nemirovski [21] (the trajectory approaches optimal sets as $t \to \infty$). As the central trajectory is one dimensional submanifold in Ω, we denote it for the primal problem as $\gamma_{\mathcal{P}} \equiv \{\gamma_{\mathcal{P}}(t) | t \in (0, \infty]\}$, where $x_{\mathcal{P}}(t) \equiv x(\gamma_{\mathcal{P}}(t))$, and the central trajectory for the dual problem as $\gamma_{\mathcal{D}} \equiv \{\gamma_{\mathcal{D}}(t) | t \in (0, \infty]\}$, where $s_{\mathcal{D}}(t) \equiv s(\gamma_{\mathcal{D}}(t))$. The central trajectory is characterized as $\gamma_{\mathcal{P}} = \mathcal{P} \cap \mathrm{Hom}(\mathcal{D})$, i.e., the intersection of ∇-autoparallel submanifold $\mathcal{P}$ and ∇^*-autoparallel submanifold $\mathrm{Hom}(\mathcal{D})$, where $\mathrm{Hom}(\mathcal{D})$ is a "homogenization" of $\mathcal{D}$. Similarly, we establish $\gamma_{\mathcal{D}} = \mathcal{D} \cap \mathrm{Hom}(\mathcal{P})$.

Then a predictor-corrector algorithm to follow the primal central trajectory $\gamma_{\mathcal{P}}$ based on this characterization is briefly explained. The number of iterations of this algorithm when the neighborhood is sufficiently small is estimated with a curvature integral involving the directional embedding curvature $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})$ of $\mathcal{P}$ along the direction $\dot\gamma_{\mathcal{P}}$ of the central trajectory with respect to the connection ∇^*. If we follow the dual central trajectory $\gamma_{\mathcal{D}}$, then iteration complexity is written with the dual counterpart $H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}})$, the directional embedding curvature of $\mathcal{D}$ along the dual central trajectory. Since the quantities $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})$ and $H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}})$ are important in connecting computational complexities and geometric structures of the problem, we derive a few properties of $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})$ and $H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}})$.

In Section 4, we focus on the concrete case of classical linear programming, and analyze the primal-dual algorithm under this geometric framework. In [29], Sonnevend, Stoer and Zhao introduced an integral $I_{PD}(t_1, t_2)$ over the primal-dual central trajectory which essentially expresses the iteration-complexity of the MTY-PC algorithm. Let $\#_{PD}(t_1, t_2, \beta)$ be the number of iterations to follow the central trajectory from $(\gamma_{\mathcal{P}}(t_2), \gamma_{\mathcal{D}}(t_2))$ to the neighorhood of $(\gamma_{\mathcal{P}}(t_1), \gamma_{\mathcal{D}}(t_1))$ when the opening of the neighborhood is β. In [43] and [42] they showed that

$$\#_{PD}(t_1, t_2, \beta) \le c_1(\beta)\frac{1}{\sqrt{\beta}} I_{PD}(t_1, t_2) + c_2(\beta)\log(t_2/t_1),$$

where c_1 and c_2 are universal constants just depending on β. (For simplicity, we assume that the algorithm starts from an exact center point, but essentially the same result holds if we allows to start from an off-center point.) In [19], Monteiro and Tsuchiya studied this integral in detail in connection with the seminal work by Vavasis and Ye, and proved that

$$\lim_{\beta\to 0} \frac{\sqrt{\beta}\times \#_{PD}(t_1, t_2, \beta)}{I_{PD}(t_1, t_2)} = 1 \quad \text{and} \quad I_{PD}(0, \infty) = O(n^{3.5}\log(\bar\chi^*_A + n)), \tag{3}$$

where $\bar\chi^*_A$ is a scaling-invariant condition number of the coefficient matrix $A \in R^{m\times n}$ in the standard form linear program [18, 19]. Note that the bound does not depend on b nor c. The condition number $\bar\chi^*_A$ is known to be $O(2^{L_A})$ where L_A is the input size of A, and for the class of combinatorial linear programs including network problems where A is a 0-1 matrix, we have the bound

$$I_{PD}(0, \infty) = O(n^{4.5}m),$$

just depending on the dimension of the problem.

In this paper, we show that the integral I_{PD} is expressed precisely as an information geometric quantity as follows:

$$I_{PD}(t_1, t_2) = \int_{t_1}^{t_2} \left(\frac{1}{4}\,\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^2_{\gamma_{\mathcal{P}}(t)} + \frac{1}{4}\,\|H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}})\|^2_{\gamma_{\mathcal{D}}(t)}\right)^{1/4} dt.$$

From this result and (3), we immediately obtain that

$$\max\left\{\int_0^\infty \frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^{1/2}_{\gamma_{\mathcal{P}}(t)} dt,\ \int_0^\infty \frac{1}{\sqrt{2}}\|H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}})\|^{1/2}_{\gamma_{\mathcal{D}}(t)} dt\right\} = O(n^{3.5}\log(\chi^*_A + n)),$$

and if A is a 0-1 matrix, we have

$$\max\left\{\int_0^\infty \frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}},\dot\gamma_{\mathcal{P}})\|^{1/2}dt,\ \int_0^\infty \frac{1}{\sqrt{2}}\|H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}},\dot\gamma_{\mathcal{D}})\|^{1/2}dt\right\} = O(n^{4.5}m)$$

just depending on the dimension of the problem. This result has its own value as a global theorem in information geometry. The curvature integrals

$$\int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}},\dot\gamma_{\mathcal{P}})\|^{1/2}_{\gamma_{\mathcal{P}}(t)}dt \quad\text{and}\quad \int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}},\dot\gamma_{\mathcal{D}})\|^{1/2}_{\gamma_{\mathcal{D}}(t)}dt$$

are counterparts of $I_{PD}(\cdot)$ in the primal and dual algorithms [28] in the sense that the number of iterations of the predictor-corrector type primal/dual algorithms is described with this quantity.

Here we should note that the embedding curvature is an essential ingredient of the iteration-complexity estimation we mentioned above. This makes a good contrast with [23] where the iteration-complexity is estimated with the Riemannian length along the central trajectory for the short-step algorithm. The short-step algorithm iterates with an approximately fixed step length in terms of Riemannian distance regardless to whether the trajectory is curved or straight. The predictor-corrector algorithm changes the step length adaptively taking account of the curvature of the central trajectory; it can take a very long step if the trajectory is straight. Information geometry provides a suitable differential geometric framework to capture this point.

In Section 5, we conduct numerical experiments with Netlib LP instances. We implement precisely the feasible Mizuno-Todd-Ye predictor-corrector algorithm with Monteiro and Adler's technique to convert the original problem into the equivalent one with a trivial centered feasible interior solution, and observed that the number of iterations is predicted by the integral quite well even with $\beta = 1$ or $\beta = 0.5$ for the instances of several thousands of variables such as DFL001 (dimension of A: 6072×12230) and PILOT87 (dimension of A: 2031×4883), where β is the opening of the neighborhood. Since the number of iterations with $\beta = 1$ is the integral itself, these numerical results confirm that the iteration count of the MTY-PC algorithm is a differential geometric quantity. Section 6 is a concluding discussion.

2 Information Geometry and Dually Flat Space

In this section, we briefly describe a framework of information geometry. For details see [2, 3]. We mention [11, 9] for general differential geometry text books.

2.1 Dually flat spaces

Let $\mathbf{E}$ and $\mathbf{E}^*$ be an n-dimensional vector space and its dual space, respectively. 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 basis vectors of $\mathbf{E}$ and $\{e_1^*, \ldots, e_n^*\}$ be its dual basis vectors of $\mathbf{E}^*$ satisfying $\langle e_j^*, e_i\rangle = \delta_{ij}$, i.e., we take a pair of biorthogonal basis. We

consider the affine coordinate system $(x^1, \ldots, x^n)$ based on $\{e_1, \ldots, e_n\}$, i.e, we represent $x \in \mathbf{E}$ as $x = \sum x^i e_i$. Similarly, for $s \in \mathbf{E}^*$ we consider the affine coordinate system $(s^1, \ldots, s^n)$ based on $\{e_1^*, \ldots, e_n^*\}$, i.e., we represent $s \in \mathbf{E}^*$ as $s = \sum s^i e_i^*$. With this setting, we identify vectors $x \in \mathbf{E}$ and $s \in \mathbf{E}^*$ and their coordinate expressions in $\mathbf{R}^n$. Then we have $\langle x, s \rangle = \sum_{i=1}^n x^i s^i = x^\mathsf{T} s$.

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., $\mathcal{C}^x = \{\xi \in \mathbf{R}^n \mid \sum_{i=1}^n \xi^i e_i \in \Omega\}$. 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 $\mathbf{R}^n$ under the coordinate system $(x^1, \ldots, x^n)$. Then the gradient mapping

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

is smoothly invertible on its image $\mathcal{C}^s \equiv \bar{s}(\mathcal{C}^x) \subseteq \mathbf{R}^n$ because the Hessian matrix of Ψ is positive definite. (The gradient map was referred to as Legendre transformation in [28].) 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

$$\mathcal{G}(\mathcal{C}) \equiv \left\{ \sum_{i=1}^n s^i e_i^* \;\middle|\; s \in \mathcal{C}^s \right\} = \left\{ \sum_{i=1}^n \bar{s}^i(x) e_i^* \;\middle|\; x \in \mathcal{C}^x \right\} \tag{4}$$

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. We illustrate the situation in Fig. 1. In the following, the original coordinate is referred to as x-coordinate and the dual coordinate is referred to as s-coordinate.

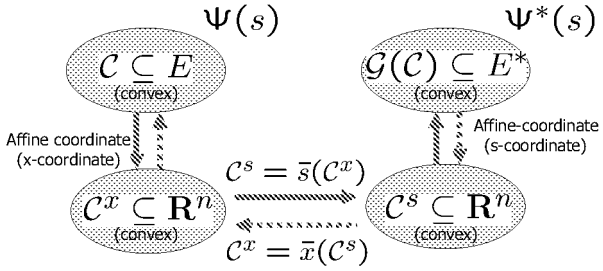

Figure 1: Dually flat space.

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 $\bar{x}(\cdot) : \mathcal{G}(\mathcal{C}) \to \mathcal{C}^x$ is given by

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

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

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

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 $\mathcal{C}^s$ are 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 to $\mathcal{G}(\mathcal{C})$ through s-coordinate and the relation (4), 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:

$$\bar{x}(s(p)) = x(p), \quad \bar{s}(x(p)) = s(p). \tag{5}$$

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 $\mathcal{M}^x \equiv \{x(p) \in \mathbf{R}^n \mid p \in \mathcal{M}\}$ and $\mathcal{M}^s \equiv \{s(p) \in \mathbf{R}^n \mid p \in \mathcal{M}\}$.

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:

$$G_p(V_p, W_p) \equiv \sum_{i,j} (G_{x(p)})_{ij} V^i_{x(p)} W^j_{x(p)} = V^{\mathsf{T}}_{x(p)} G_{x(p)} W_{x(p)}, \quad (G_{x(p)})_{ij} = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x(p)).$$

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

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

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 $G_p(V_p, V_p) = V_x^{\mathsf{T}} G_x V_x = V_s^{\mathsf{T}} G_s V_s$ and $G_s = G_x^{-1}$.

Let K be the symmetric tensor field whose value at p is defined as, for $(X_p, Y_p, Z_p) \in T_p\mathcal{C} \times T_p\mathcal{C} \times T_p\mathcal{C}$:

$$K_p(X_p, Y_p, Z_p) = \sum_{i,j,k} (K_x)_{ijk} X^i_x Y^j_x Z^k_x, \quad (K_x)_{ijk} \equiv \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k},$$

Similarly, let K^* be the tensor field whose value at p is defined as, for $(X_p, Y_p, Z_p) \in T_p\mathcal{C} \times T_p\mathcal{C} \times T_p\mathcal{C}$:

$$K^*_p(X, Y, Z) = \sum_{i,j,k} (K^*_s)_{ijk} X^i_s Y^j_s Z^k_s, \quad (K^*_s)_{ijk} = \frac{\partial^3 \Psi^*}{\partial s^i \partial s^j \partial s^k}$$

defined by Ψ^* analogous to K. The following remarkable relation holds between K and K^*.

Proposition 2.1. *We have $K^* = -K$.*

Proof. Let $p \in \mathcal{C}$, and let $X_p, Y_p, Z_p \in T_p\mathcal{C}$. We consider the curve $\gamma(t)$ such that $x(\gamma(t)) = x(p) + tZ_x$ which goes through p. Then,

$$\begin{aligned} K_p(X_p, Y_p, Z_p) &= \sum_{i,j,k} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k} X^i_x Y^j_x Z^k_x = X_x^{\mathsf{T}} \left(\frac{d}{dt} G_x \right) Y_x = -X_x^{\mathsf{T}} \left(\frac{d}{dt} G_s^{-1} \right) Y_x \\ &= -X_x^{\mathsf{T}} G_s^{-1} \left(\frac{d}{dt} G_s \right) G_s^{-1} Y_x = -X_s^{\mathsf{T}} \left(\frac{d}{dt} G_s \right) Y_s = -K^*(X_p, Y_p, Z_p). \end{aligned}$$

■

Now we introduce affine connections and covariant derivatives on $\mathcal{C}$, which determine the structure of the manifold such as torsions and curvatures. See the appendix for a quick introduction to connections and covariant derivatives. 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.,

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

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 covariate derivative with respect to ∇ and the ordinary derivative in s-coordinate coincides with the covariate 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 $\gamma \equiv \{\gamma(t) \in \mathcal{C} \mid t \in [a,b] \subset \mathbf{R}\}$ be a curve in $\mathcal{C}$. Then, for any $t \in [a,b]$, we have

$$(\nabla_{\dot\gamma}\dot\gamma)_{x(\gamma(t))} = \frac{d^2 x(\gamma(t))}{dt^2} \quad \text{and} \quad (\nabla_{\dot\gamma}^*\dot\gamma)_{s(\gamma(t))} = \frac{d^2 s(\gamma(t))}{dt^2}. \tag{6}$$

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

We also have

$$(\nabla_{\dot\gamma} G)_x = \frac{dG_{x(\gamma(t))}}{dt} \quad \text{and} \quad (\nabla_{\dot\gamma}^* G)_s = \frac{dG_{s(\gamma(t))}}{dt}$$

as Γ and Γ^* vanishes in x-coordinate and s-coordinate, respectively. See also the appendix.

Let $U_{\gamma(t)}, V_{\gamma(t)} \in T_{\gamma(t)}\mathcal{C}$, then

$$\begin{aligned} U_{x(\gamma(t))}^{\mathsf{T}} \frac{dG_{x(\gamma(t))}}{dt} V_{x(\gamma(t))} &= \sum_{i,j} \frac{d(G_{x(\gamma(t))})_{ij}}{dt} U_{x(\gamma(t))}^i V_{x(\gamma(t))}^j = \sum_{i,j,k} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k} U_{x(\gamma(t))}^i V_{x(\gamma(t))}^j \frac{dx^k(\gamma(t))}{dt} \\ &= K_{\gamma(t)}(\dot\gamma(t), U_{\gamma(t)}, V_{\gamma(t)}), \end{aligned}$$

and

$$U^{\mathsf{T}}_{s(\gamma(t))}\frac{dG_{s(\gamma(t))}}{dt}V_s = \sum_{i,j}\frac{d(G_{s(\gamma(t))})_{ij}}{dt}U^i_{s(\gamma(t))}V^j_{s(\gamma(t))} = \sum_{i,j,k}\frac{\partial^3\Psi^*}{\partial s^i\partial s^j\partial s^k}U^i_{s(\gamma(t))}V^j_{s(\gamma(t))}\frac{ds^k(\gamma(t))}{dt}$$
$$= K^*_p(\dot\gamma(t), U_{\gamma(t)}, V_{\gamma(t)}), \tag{7}$$

These relations will be used later.

Before proceeding, we motivate the two connections in view of the inner product. It is well-known that in Riemannian geometry the inner product of two vectors is preserved in parallel translation along a curve under the Levi-Civita connection ∇^{LC}. That is, let X and Y be parallel vector fields along the curve γ, i.e., $\nabla^{\mathrm{LC}}_{\dot\gamma}X_{\gamma(t)} = 0$ and $\nabla^{\mathrm{LC}}_{\dot\gamma}Y_{\gamma(t)} = 0$, then its inner product $G(X,Y)$ does not change along the curve. In the dually flat space of information geometry, the following analogous result holds; let X be parellel along γ with respect to the connection ∇, and Y be parellel along γ with respect to the connection ∇^*, i.e., $\nabla_{\dot\gamma}X = 0$ and $\nabla^*_{\dot\gamma}Y = 0$, then its inner product $G(X,Y)$ does not change along the curve. In this respect, the mutually dual connections are a natural extension of the well-known Levi-Civita connection.

2.2 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 $\mathbf{R}^n$:

$$\mathcal{M} \subseteq \{p \in \mathcal{C} | x(p) = c_0 + \sum_{i=1}^k y^i c_i,\ c_i \in \mathbf{E}, y^i \in \mathbf{R}\},\ \text{and } \mathcal{M} \text{ is an open set}$$
or, equivalently
$$\mathcal{M}^x \subseteq \{x \in \mathcal{C}^x | x = c_0 + \sum_{i=1}^k y^i c_i,\ c_i \in \mathbf{R}^n, y^i \in \mathbf{R}\} \text{ and } \mathcal{M}^x \text{ is an open set.}$$

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:

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

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

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

$\mathrm{Hom}(\mathcal{M})$ 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. We adapt x-coordinate, and suppose that $T_p\mathcal{M} \subset T_p\mathcal{C}$ is represented by the kernel of a certain linear operator $A : \mathbf{R}^n \to \mathbf{R}^m$, i.e.,

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

Then we have

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

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

$$\Pi^{\perp}_{s(p)} = G_{x(p)} \Pi^{\perp}_{x(p)} G^{-1}_{x(p)} = A^{\mathsf{T}} (A G_{s(p)} A^{\mathsf{T}})^{-1} A G_{s(p)}. \tag{8}$$

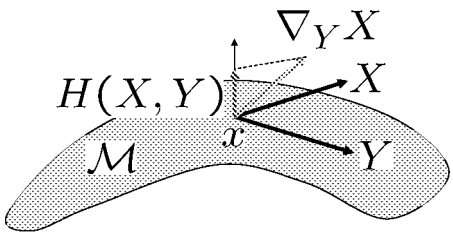

Figure 2: Embedding curvature.

2.3 Self-concordant functions and information geometry

If

$$\left|\sum_{i,j,k}^{n}\frac{\partial^3\Psi(x)}{\partial x^i\partial x^j\partial x^k}X^iX^jX^k\right| \leq 2\left(\sum_{i,j}^{n}\frac{\partial^2\Psi(x)}{\partial x^i\partial x^j}X^iX^j\right)^{3/2} \tag{9}$$

holds for all $x \in \mathcal{C}^x$ and $X \in \mathbf{R}^n$, 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

$$\sum_{i}\left|\frac{\partial\Psi(x)}{\partial x^i}X^i\right| \leq \sqrt{\theta}\left(\sum_{i,j}^{n}\frac{\partial^2\Psi(x)}{\partial x^i\partial x^j}X^iX^j\right)^{1/2} \tag{10}$$

for all $x \in \mathcal{C}^x$ and $X \in \mathbf{R}^n$, 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.

Due to (9) and (10), we have

$$|K_p(X,X,X)| \leq 2(G_p(X,X))^{3/2} = 2\|X\|_p^3 \text{ and } \langle\bar{s}(x), X\rangle \leq \sqrt{\theta}(G_p(X,X))^{1/2} = \sqrt{\theta}\|X\|_p$$

for all $X \in T_p\mathcal{C}$ at all $p \in \mathcal{C}$, and the following property holds (see Appendix 1 of [21]):

$$|K_p(X,Y,Z)| \leq 2\|X\|_p\|Y\|_p\|Z\|_p. \quad \forall X,\ Y,\ Z \in T_p\mathcal{C} \text{ and } p \in \mathcal{C}. \tag{11}$$

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

$$\Omega^*_{\mathbf{R}} = \{s \in \mathbf{R}^n |\ \sum_i s^i e_i^* \in \Omega^*\} = \{s \in \mathbf{R}^n |\ s^{\mathsf{T}}x > 0\ \forall\ \Omega^x/\{0\}\}.$$

The next result plays a fundamental role in our 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.This fact is already known e.g., [21], and we tailored it suitable for our presentation.

Theorem 2.2. *Under the notations as above,* $\Omega^s = \Omega^*_{\mathbf{R}}$, *and hence* $\mathcal{G}(\Omega) = \Omega^*$. Ω^x *and* Ω^s *are coordinate representations of* Ω *and* Ω^* *such that*

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

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 $\Omega^s = \Omega^*_{\mathbf{R}}$. Other statements are obvious from previous discussions and definitions. First we observe that $\Omega^s \subseteq \Omega^*_{\mathbf{R}}$. Let $s \in \Omega^s$. This implies that the optimization problem: $\min_{x \in \Omega^x} -s^{\mathsf{T}}x + \psi(x)$ have an optimal solution. Since $\psi(x)$ is strongly convex, the set $\{x \in \mathbf{R}^n |\ s^{\mathsf{T}}x = 1\} \cap \Omega_{\mathbf{R}}$ is bounded, and hence $\{x \in \mathbf{R}^n |\ s^{\mathsf{T}}x = 0\} \cap \Omega_{\mathbf{R}} = \{0\}$. Due to the convex conic version of Gordan's theorem, we have $s \in \Omega^*_{\mathbf{R}}$.

Next we show $\Omega^*_{\mathbf{R}} \subseteq \Omega^s$. If $s \in \Omega^*_{\mathbf{R}}$, then the set $X = \{x \in \Omega^x |\ s^{\mathsf{T}}x = 1\}$ is bounded because s is an interior point of $\mathrm{cl}(\Omega^*_{\mathbf{R}})$ 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 homogenious property of ψ, we have $s = \bar{s}(\lambda\hat{x})$, and therefore $s \in \Omega^s$. ∎

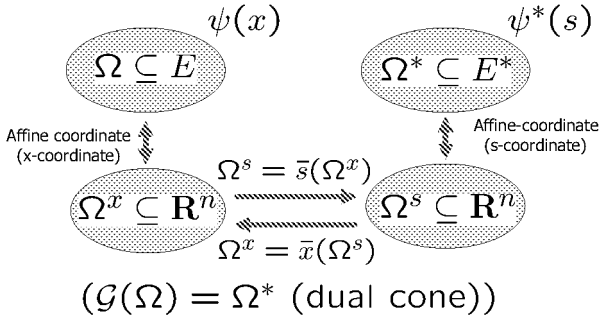

Figure 3: Dually flat space (Conic case).

Proposition 2.3. *Let* $\psi(x)$ *be a* θ*-logarithmically homogeneous barrier on* Ω, *and consider information geometry of* Ω *based on* ψ. *Let* $p, p' \in \Omega$, *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 [21] (just the notation is different). For the sake of completeness, we provide a proof below.

To show (a), we assume $x(p') = tx(p)$ and prove that $ts(p') = s(p)$. Due to (5), it is enough to show that $t\bar{s}(x(p')) = \bar{s}(x(p))$. Under the assumption, this is equivalent to $t\bar{s}(tx(p)) = \bar{s}(x(p))$. Differentiation of $\psi(tx) = \psi(x) - \theta \log t$ with respect to x applying the chain rule yields the result. The reverse part can be proved exactly in the same manner using ψ^*. The first statement of (b) can be proved by differentiating the relation $\psi(tx) = \psi(x) - \theta \log t$ twice with respect to x. The second statement is proved in the similar manner by using ψ^*. (c) is proved by differentiating $\psi(tx) = \psi(x) - \theta \log t$ with respect to t and by letting $t = 1$. The first statement of (d) is proved by differentiating $\bar{x}(s)^{\mathsf{T}} s = \theta$. Since $G_s = G_x^{-1}$, we have the second statements. ∎

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-coordinate 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

3 Information Geometric Framework of Conic Linear Programs

Now we introduce information geometric framework on (1) and (2). Following Section 2, 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 (1) and (2) are represented with respect to these bases. We use the same letters for vectors and subspaces and their coordinate expressions. In particular, $c, d, \mathbf{T}$ and $\mathbf{T}^*$ below mean vectors and linear subspaces in $\mathbf{R}^n$ representing $c, d, \mathbf{T}$, and $\mathbf{T}^*$ in (1) and in (2) 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.

3.1 Conic Linear Programs

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

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

Then (1) and (2) are written, as optimization problems on $\mathbf{R}^n$, as

$$\min \ c^{\mathsf{T}} x, \quad \text{s.t. } x \in \mathrm{cl}(\mathcal{P}^x), \tag{12}$$

and

$$\min \ s^{\mathsf{T}} d, \quad \text{s.t. } s \in \mathrm{cl}(\mathcal{D}^s), \tag{13}$$

respectively.

We assume that both (1) and (2) have interior feasible solutions. Then we have $\mathcal{P}^x \neq \emptyset$ and $\mathcal{D}^s \neq \emptyset$. Under this assumption, (12) and (13) 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 (12) and (13).

As for (12), we consider the following optimization problem with parameter t

$$\min \ tc^{\mathsf{T}} x + \psi(x) \ \text{ s.t. } x \in \mathcal{P}^x. \tag{14}$$

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,

$$\bar{s}(x) \in t\mathcal{D}^s, \quad x \in \mathcal{P}^x. \tag{15}$$

Let $x_{\mathcal{P}}(t)$ be the unique optimal solution to (14). The set of $x_{\mathcal{P}}(t)$ with parameter t varying from 0 to infinity is called the central trajectory of (12).

Similarly, we consider the following optimization problem with parameter t associated with (13).

$$\min \ ts^{\mathsf{T}} d + \psi^*(s) \ \text{ s.t. } s \in \mathcal{D}^s. \tag{16}$$

The optimality condition for this problem is

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

Let $s_{\mathcal{D}}(t)$ be the unique optimal solution to (16). The set of $s_{\mathcal{D}}(t)$ with parameter t varying from 0 to infinity is referred to as the central trajectory of (13).

3.2 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 (1) is an optimization problem in Ω, and is expressed as (12) in x-coordinate. We formulate the dual problem (2) as an optimization problem in Ω as follows. The dual problem is equivalent to (13) as an optimization problem on $\Omega^*_{\mathbf{R}}$. Due to Theorem 2.2, we have $\Omega_{\mathbf{R}} = \Omega^x$ and $\Omega^*_{\mathbf{R}} = \Omega^s$. Furthermore, Ω^s is a global coordinate of Ω defined by the gradient map. Therefore, (13) 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 intersecation 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.

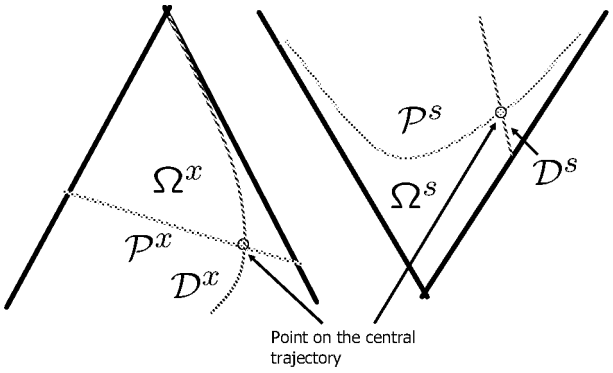

Figure 4: Primal and dual feasible regions in x-coordinate (left) and s-coordinate (right) .

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 (1)/(12) 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 (12) as $t \to \infty$.

Let $s_{\mathcal{D}}(t)$ be the unique optimal solution to (14), 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 (2)/(13) 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 (13) as $t \to \infty$.

Proposition 3.1. *$\gamma_{\mathcal{P}}$ and $\gamma_{\mathcal{D}}$ are represented as the intersection of two submanifolds:*

$$\gamma_{\mathcal{P}}(t) = \mathcal{P} \cap t\mathcal{D} \text{ and } \gamma_{\mathcal{P}} = \mathcal{P} \cap \mathrm{Hom}(\mathcal{D}), \tag{17}$$

and

$$\gamma_{\mathcal{D}}(t) = \mathcal{D} \cap t\mathcal{P} \text{ and } \gamma_{\mathcal{D}} = \mathcal{D} \cap \mathrm{Hom}(\mathcal{P}). \tag{18}$$

(See Fig. 5)

Proof. Let $p = \gamma_{\mathcal{P}}(t)$. Then $x(p)$ satisfies (15). Since $\bar{s}(x(p)) \in t\mathcal{D}^s$ and $x(p) \in \mathcal{P}$, the relation (17) follows from the definition of $\mathrm{Hom}(\mathcal{D})$. The relation (18) follows similarly. ∎

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 3.2. *The following holds.*

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

$$\min\ \ s^{\mathsf{T}} d + \psi^*(s) \quad s.t.\ s \in t\mathcal{D}^s. \tag{19}$$

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 (19) 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 (15), 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 3.3.

$$x(\gamma_{\mathcal{D}}(t)) = tx(\gamma_{\mathcal{P}}(t)) \text{ and } s(\gamma_{\mathcal{P}}(t)) = ts(\gamma_{\mathcal{D}}(t)). \tag{20}$$

Proof. We show the second relation of (20). Once this is done, the first one immediately follows from Proposition 2.3(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. ∎

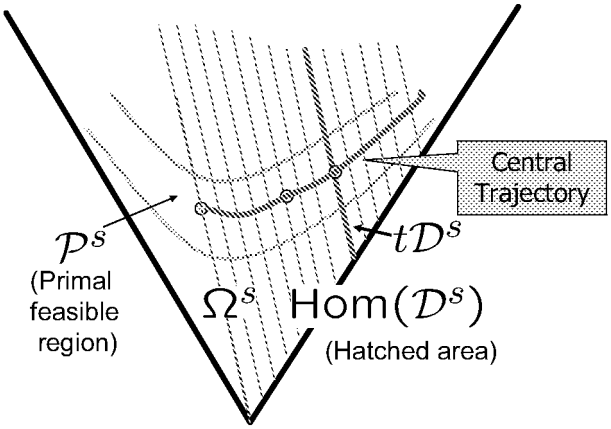

Figure 5: Central trajectory.

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

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

where $A \in \mathbf{R}^{m \times n}$ and the rows of A are linearly independent. On the other hand, for $b \in \mathbf{R}^m$ 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

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

One more differentiation with respect to t yields that

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

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

$$\dot{s} = -G_x\dot{x} = (G_x - A^{\mathsf{T}}(AG_x^{-1}A^{\mathsf{T}})^{-1}A)G_x^{-1}c = G_x(I - \Pi_x^{\perp})G_x^{-1}c = (I - \Pi_s^{\perp})c. \tag{21}$$

Observe that the rightmost hand side of (21) 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 (21), 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 $V^{\mathrm{ct}(\mathcal{D})}$.

3.3 A Geometric Predictor-Corrector Algorithm and Curvature Integral

In [28], 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 $s_{\mathcal{P}}(t)$.

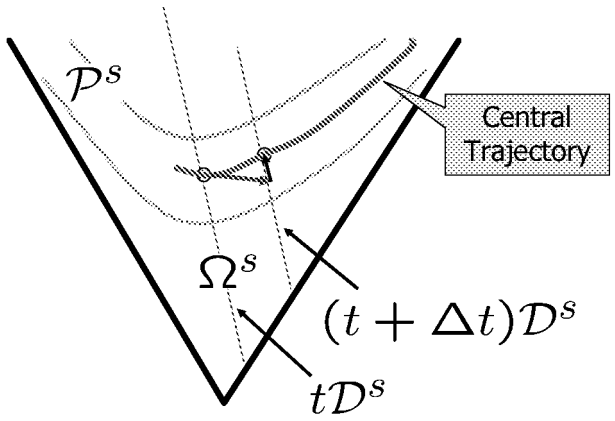

Figure 6: Path-following algorithm.

[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

$$\hat{s}_L(t') = \hat{s} + (t' - t)(I - \Pi_{\hat{s}}^{\perp})c. \tag{22}$$

Since $(I - \Pi_{\hat{s}}^{\perp})c = c - A^{\mathsf{T}}y'$ for some $y' \in \mathbf{R}^m$, we have

$$s_L(t') \in t'\mathcal{D} \text{ (as long as } s_L(t') \in \Omega^s). \tag{23}$$

As we observed in (23), 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 (19) (with $t := t'$) and $\hat{s}'$ is the feasible solution to (19). The corrector step at $\hat{s}'$ is the Newton step for the point $\hat{s}(t')$ to solve the problem (19). Let $N_{\hat{s}'}$ be the Newton displacement vector for (19) 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 $t'\mathcal{D}^s$:

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

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

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

After the predictor-step is performed, we have

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

As we discussed above, the corrector-step is the Newton step for the convex optimization problem (19) (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 (22). 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

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

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

$$\lim_{\beta \to 0} \frac{\sqrt{\beta} \times \#_P(t_1, t_2, \beta)}{I_{\mathcal{P}}(t_1, t_2)} = 1, \tag{24}$$

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 [28] 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 (24) 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

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

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 next subsection, we have $\nabla^*_{\dot\gamma_{\mathcal{P}}} \dot\gamma_{\mathcal{P}} = H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})$. Therefore,

$$\frac{1}{\sqrt{2}} \|H_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^{1/2}_{\gamma_{\mathcal{P}}(t)} \Delta t = \frac{\|\nabla^*_{\dot\gamma} \dot\gamma\|^{1/2}}{\sqrt{2}} \Delta t \sim \sqrt{\beta} \tag{25}$$

holds at one iteration of the predictor-corrector algorithm.

Taking summation of the both sides of (25) 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 (24) follows.

In spite of its asymptotic nature, we consider that (24) 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

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

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.

3.4 Embedding Curvature

As was suggested in the previous subsection, the embedding curvature along the central trajectories $H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})$ and $H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}}, \dot{\gamma}_{\mathcal{D}})$ plays an important role in connecting geometrical structure to computational complexity. In this subsection, we deal with the embedding curvature.

Lemma 3.4. *We have the following relations:*

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

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

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 (6)). Due to (21), $\nabla^*_{\dot{\gamma}_{\mathcal{P}}} \dot{\gamma}_{\mathcal{P}} = \ddot{s} = -\dot{\Pi}_s^{\perp} c$. We derive an expression for $\dot{\Pi}_s^{\perp}$ below. Recall that $\Pi^{\perp}_{s(p)} = A^{\mathsf{T}}(AG_{s(p)}A^{\mathsf{T}})^{-1}AG_{s(p)}$. (see (8)). Since

$$A^{\mathsf{T}} \left(\frac{d}{dt}(AG_sA^{\mathsf{T}})^{-1} \right) AG_s \;\; = \;\; -A^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1} \left(\frac{d}{dt}(AG_sA^{\mathsf{T}}) \right) (AG_sA^{\mathsf{T}})^{-1}AG_s = -\Pi_s^{\perp} G_s^{-1} \dot{G} \Pi_s^{\perp},$$

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

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

From this relation, it immediately follows that

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

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

In the following, we prove that $\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2}$ is bounded by $\sqrt{\theta}/t$. The result implies that

$$I_{\mathcal{P}}(t_1, t_2) \leq \sqrt{\theta} \log \frac{t_2}{t_1},$$

which is naturally expected from the standard complexity analysis of interior-point algorithms.

Proposition 3.5. *We have*

$$\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|_{\gamma_{\mathcal{P}}(t)} \leq \frac{2\theta}{t^2}, \quad \|H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}}, \dot{\gamma}_{\mathcal{D}})\|_{\gamma_{\mathcal{D}}(t)} \leq \frac{2\theta}{t^2},$$

Proof. We just prove the first relation. The proof of the second relation is similar. Since $G_p(V_p, \Pi_p^{\perp} V_p) = G_p(\Pi_p^{\perp} V_p, \Pi_p^{\perp} V_p)$ holds for $V_p \in T_p\Omega$, we have

$$\begin{aligned} \|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^2_{\gamma_{\mathcal{P}}(t)} &= (\Pi_s^{\perp} G_s^{-1} \dot G_s \dot s)^{\mathsf{T}} G_s (\Pi_s^{\perp} G_s^{-1} \dot G_s \dot s) = (\Pi_s^{\perp} G_s^{-1} \dot G_s \dot s)^{\mathsf{T}} G_s (G_s^{-1} \dot G_s \dot s) \\ &= -H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})_s^{\mathsf{T}} \dot G_s \dot s = -K^*(H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}}), \dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}}) = K(H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}}), \dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}}), \end{aligned}$$

where the second last equality and the last equality follow from (7) and Proposition 2.1, respectively. Then it follows from (11) that

$$\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^2_{\gamma_{\mathcal{P}}} \le 2\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|_{\gamma_{\mathcal{P}}} \|\dot\gamma_{\mathcal{P}}\|_{\gamma_{\mathcal{P}}} \|\dot\gamma_{\mathcal{P}}\|_{\gamma_{\mathcal{P}}}.$$

Therefore we obtain $\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|_{\gamma_{\mathcal{P}}} \le 2\|\dot\gamma_{\mathcal{P}}\|^2_{\gamma_{\mathcal{P}}}$. Since $s = tc - A^{\mathsf{T}} y$ for some y and $(I - \Pi_s^{\perp}) A^{\mathsf{T}} y = 0$, we have $t(\dot\gamma_{\mathcal{P}})_s = t\dot s = (I - \Pi_s^{\perp}) s$. The norm of $\dot\gamma_{\mathcal{P}}$ is bounded as follows:

$$\|\dot\gamma_{\mathcal{P}}\|^2_{\gamma_{\mathcal{P}}(t)} = t^{-2}((I - \Pi_s^{\perp})s)^{\mathsf{T}} G_s (I - \Pi_s^{\perp}) s \le s^{\mathsf{T}} G_s s = s^{\mathsf{T}} x = t^{-2}\theta.$$

This completes the proof. ∎

4 Linear Programming

In this section, we focus on classical linear programming. Let us consider the dual pair of linear programs:

$$\min\ c^{\mathsf{T}} x \text{ s.t. } Ax = b,\ x \ge 0, \quad \text{and} \quad \max\ b^{\mathsf{T}} y \text{ s.t. } c - A^{\mathsf{T}} y = s,\ s \ge 0,$$

where $A \in R^{m \times n}$, $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$. We assume that the rows of A are linearly independent. It is easy to see that the problem fits into the general setting in (1) and (2), if we take $\Omega = \Omega^* = \mathbf{R}^n_{++}$ and take d satisfying $Ad = -b$. We will consider the situation where we choose $\psi(x) = -\sum_{i=1}^n \log x_i$ which is an n-normal barrier.

Let

$$\bar\chi_A = \max_B \|A_B^{-1} A\|,$$

where B is the set of indices such that A_B is nonsingular. Furthermore, let

$$\bar\chi_A^* = \inf_D\ \bar\chi_{AD},$$

where D is the positive definite diagonal matrix. The quantity $\bar\chi_A$ is the condition number of the coefficient matrix A studied in, for example, [6, 8, 30, 37, 38]. This quantity plays an important role in the polynomial-time layered-step interior-point algorithm by Vavasis and Ye [41] whose complexity just depends on A, and the subsequent analysis by Monteiro and Tsuchiya [19]. The quantity $\bar\chi_A^*$ is

a scaling-invariant version of $\bar{\chi}_A$ introduced in [19]. If A is integral, then, $\bar{\chi}_A$ is bounded by $2^{O(L_A)}$, where L_A is the input size of A.

The main goal of this section is to represent the iteration-complexity of the MTY-PC algorithm in terms of information geometry and establish a bound on the total curvature of the central trajectory. Before going to prove the results, we introduce a few notations here. Given two vectors u and v, we denote the elementwise product as $u \circ v$. The unit element of this product is the vector of all ones and denoted by e. The vectors whose components are inverse and square root of the elements of a vector u are denoted by u^{-1} and $u^{1/2}$, respectively. The elementwise product $u \circ v$ is the Euclidean Jordan algebra associated with the cone $\mathbf{R}^n_{++}$. As to the order of operations, we promise that the product $\circ$ is weaker than the ordinary product of matrix and vectors, i.e., $Ax \circ y$, say, is interpreted as $(Ax) \circ y$ and not as $A(x \circ y)$.

We have $\bar{x}(s) = s^{-1}$, $\bar{s}(x) = x^{-1}$, $G_x = \mathrm{diag}(x)^{-2}$, $G_s = \mathrm{diag}(s)^{-2}$. At $p \in \Omega$, we define the projection matrix Q as follows:

$$Q(p) = G_{s(p)}^{1/2} A^{\mathsf{T}} (A G_{s(p)} A^{\mathsf{T}})^{-1} A G_{s(p)}^{1/2}.$$

By using Proposition 2.3(b), it is easy to see that

$$Q(p') = Q(p) \text{ if } s(p') = t s(p) \text{ holds for some } t > 0.$$

We also use the notation $\|\cdot\|_2$ for the ordinary Euclidean norm defined by $\|u\|_2 = \sqrt{\sum_i u_i^2}$ for a vector u, say.

Let $(x_{PD}(\nu), s_{PD}(\nu), y_{PD}(\nu))$ be the point of the central trajectory with parameter ν which is defined as the unique solution to the following system of equations.

$$x \circ s = \nu e, \quad Ax = b, \quad c - A^{\mathsf{T}} y = s, \quad x \geq 0, \quad s \geq 0.$$

It is well-known that

$$x_{PD}(\nu) = x\left(\gamma_{\mathcal{P}}\left(\frac{1}{\nu}\right)\right), \quad s_{PD}(\nu) = s\left(\gamma_{\mathcal{D}}\left(\frac{1}{\nu}\right)\right).$$

We define, for $0 \leq t_1 \leq t_2$

$$I_{PD}(t_1, t_2) = \int_{1/t_2}^{1/t_1} \frac{1}{\sqrt{\nu}} \left\| \left(\frac{dx_{PD}}{d\nu}\right) \circ \left(\frac{ds_{PD}}{d\nu}\right) \right\|_2^{1/2} d\nu.$$

This integral was firstly introduced by Sonnevend, Stoer and Zhao [29] for the purpose of approximating the number of iterations of interior-point algorithms with an integral over the central trajectory. Let $\#(t_1, t_2, \beta)$ be the number of iterations of the MTY-PC algorithm with the neighborhood width β to reduce the normalized duality gap $x^{\mathsf{T}} s/n$ from t_1^{-1} to t_2^{-1}, starting from the center point $(x_{PD}(t_1^{-1}), s_{PD}(t_1^{-1}), y_{PD}(t_1^{-1}))$. Zhao and Stoer [43] proved $\#(t_1, t_2, \beta)$ is bounded as follows:

$$\#(t_1, t_2, \beta) \leq c_1(\beta) I_{PD}(t_1, t_2) + c_2(\beta) \log(t_2/t_1),$$

where c_1 and c_2 are constants just depend on β (see also [42]). Furthermore, the following result was obtained by Monteiro and Tsuchiya [19].

Theorem 4.1. *We have the following:*

(a) $\lim_{\beta\to 0} \dfrac{\sqrt{\beta}\times \#(t_1,t_2,\beta)}{I_{PD}(t_1,t_2)} = 1$

(except for the special case where the central trajectory happens to be a straight line in the primal-dual space).

(b) $I_{PD}(0,\infty) = O(n^{3.5}\log(\bar{\chi}_A^* + n))$.

We note that if the central trajectory is a straight line, then the problem is essentially homogenious and the MTY-PC algorithm can solve the problem in one predictor step. We also note that even in such a special case, the statement (b) holds since in the case the integrant of I_{PD} becomes zero on the central trajectory, see Lemma 2.1 of [19].

In the following, we consider $Q(p)$ when $p = \gamma_{\mathcal{P}}(t)$ or $p = \gamma_{\mathcal{D}}(t)$. Since $s(\gamma_{\mathcal{P}}(t)) = ts(\gamma_{\mathcal{D}}(t))$, we have $Q(\gamma_{\mathcal{P}}(t)) = Q(\gamma_{\mathcal{D}}(t))$. Therefore, slightly abusing the notation, we let

$$Q(t) = G_s^{1/2}A^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1}AG_s^{1/2}$$

with $s = s(\gamma_{\mathcal{P}}(t))$ (and/or $s = s(\gamma_{\mathcal{D}}(t))$), and define

$$h_{PD}(t) = \frac{\|(I - Q(t))e \circ Q(t)e\|_2}{t^2}.$$

h_{PD} is referred to as the primal-dual curvature. We have the following alternative expression of I_{PD}.

Proposition 4.2. *We have*

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

Proof. Since $x_{PD}(\nu) = x(\gamma_{\mathcal{P}}(1/\nu))$ and $s_{PD}(\nu) = s(\gamma_{\mathcal{D}}(1/\nu))$, it is not difficult to see that

$$\nu x_{PD}^{-1}\circ\left(\frac{dx_{PD}}{d\nu}\right) = \left(I - Q\left(\frac{1}{\nu}\right)\right)e, \quad \nu s_{PD}^{-1}\circ\left(\frac{ds_{PD}}{d\nu}\right) = Q\left(\frac{1}{\nu}\right)e.$$

This implies that

$$\nu\left(\frac{dx_{PD}}{d\nu}\right)\circ\left(\frac{ds_{PD}}{d\nu}\right) = \left(I - Q\left(\frac{1}{\nu}\right)\right)e \circ Q\left(\frac{1}{\nu}\right)e$$

and hence

$$\int_{1/t_2}^{1/t_1}\frac{1}{\sqrt{\nu}}\left\|\left(\frac{dx_{PD}}{d\nu}\right)\circ\left(\frac{ds_{PD}}{d\nu}\right)\right\|_2^{1/2}d\nu = \int_{1/t_2}^{1/t_1}\frac{\left\|\left(I - Q\left(\frac{1}{\nu}\right)\right)e\circ Q\left(\frac{1}{\nu}\right)e\right\|_2^{1/2}}{\nu}d\nu.$$

We make change of variables $t = \nu^{-1}$ in the integral. Then the proposition immediately follows. ∎

Now we are ready to prove the main results in this section.

Theorem 4.3. *Let $\gamma_{\mathcal{P}}(t)$ and $\gamma_{\mathcal{D}}(t)$ be the points on the primal and the dual central trajectory with parameter t. We have the following Pythagoras relation among the primal embedding curvature $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t), \dot\gamma_{\mathcal{P}}(t))$ at $\gamma_{\mathcal{P}}(t)$ and the dual embedding curvatures $H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}(t), \dot\gamma_{\mathcal{D}}(t))$ at $\gamma_D(t)$, and the primal-dual curvature $h_{PD}(t)$.*

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

and therefore, I_{PD} is expressed exactly in terms of information geometry.

$$I_{PD} = \int_{t_1}^{t_2} \left[\left\| \frac{1}{2} H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t), \dot\gamma_{\mathcal{P}}(t)) \right\|^2_{\gamma_{\mathcal{P}}(t)} + \left\| \frac{1}{2} H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}(t), \dot\gamma_{\mathcal{D}}(t)) \right\|^2_{\gamma_{\mathcal{D}}(t)} \right]^{1/4} dt.$$

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

$$G_s^{1/2}\dot s = G_s^{1/2}(I - \Pi_s^{\perp})c = t^{-1}G_s^{1/2}(I-\Pi_s^{\perp})tc = t^{-1}G_s^{1/2}(I-\Pi_s^{\perp})s = t^{-1}(I-Q)e,$$

where the last equatlity follows from $G_s^{1/2}s = \mathrm{diag}(s)^{-1}s = e$. Therefore, we have

$$\begin{aligned}
H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})_{s(\gamma_{\mathcal{P}}(t))} &= -\Pi_s^{\perp} G_s^{-1}\dot G_s \dot s = \Pi_s^{\perp}\left(\frac{d}{dt}G_s^{-1}\right) G_s \dot s \\
&= 2\Pi_s^{\perp}(\mathrm{diag}(\dot s))(\mathrm{diag}(s))(\mathrm{diag}(s))^{-2}\dot s \\
&= 2G_s^{-1/2} Q G_s^{1/2}(\mathrm{diag}(\dot s))(\mathrm{diag}(s))^{-1}\dot s \\
&= 2G_s^{-1/2} Q((G_s^{1/2}\dot s) \circ (G_s^{1/2}\dot s)) \\
&= \frac{2}{t^2} G^{-1/2}_{s(\gamma_{\mathcal{P}}(t))} Q(\gamma_{\mathcal{P}}(t))((I - Q(\gamma_{\mathcal{P}}(t)))e) \circ ((I - Q(\gamma_{\mathcal{P}}(t)))e)).
\end{aligned}$$

Similarly, we have

$$H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}}))_{x(\gamma_{\mathcal{D}}(t))} = \frac{2}{t^2} G^{-1/2}_{x(\gamma_{\mathcal{D}(t)})}(I - Q(\gamma_{\mathcal{D}(t)}))(Q(\gamma_{\mathcal{D}}(t))e \circ Q(\gamma_{\mathcal{D}}(t))e).$$

Therefore,

$$\begin{aligned}
&\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^2_{\gamma_{\mathcal{P}}(t)} + \|H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}, \dot\gamma_{\mathcal{D}}))\|^2_{\gamma_{\mathcal{D}}(t)} \\
&= \frac{4}{t^4}\|Q(\gamma_{\mathcal{P}(t)})((I - Q(\gamma_{\mathcal{P}(t)}))e \circ (I - Q(\gamma_{\mathcal{P}(t)}))e)\|_2^2 + \frac{4}{t^4}\|(I - Q(\gamma_{\mathcal{D}(t)})(Q(\gamma_{\mathcal{D}(t)})e \circ Q(\gamma_{\mathcal{D}(t)})e)\|_2^2. \\
&= \frac{4}{t^4}\|Q(t)((I - Q(t))e \circ (I - Q(t))e)\|_2^2 + \frac{4}{t^4}\|(I - Q(t))(Q(t)e \circ Q(t)e)\|_2^2.
\end{aligned}$$

To complete the proof, we show

$$\|Q(t)((I - Q(t))e \circ (I - Q(t))e)\|_2^2 + \|(I - Q(t))(Q(t)e \circ Q(t)e)\|_2^2 = \|Q(t) \circ (I - Q(t))\|_2^2. \quad (26)$$

Let $Q := Q(t)$. Since $Q(I-Q)e = (I-Q)Qe = 0$, we have

$$Q((I-Q)e \circ (I-Q)e) = -Q((I-Q)e \circ Qe) \text{ and } (I-Q)(Qe \circ Qe) = -(I-Q)((I-Q)e \circ Qe).$$

This implies that

$$Q((I-Q)e \circ (I-Q)e) + (I-Q)(Qe \circ Qe) = -Q((I-Q)e \circ Qe) - (I-Q)((I-Q)e \circ Qe) = -Qe \circ (I-Q)e.$$

Taking $\|\cdot\|_2$ norm of the both sides of the equation and using $Q(I-Q)=0$, we obtain (26) as we desire. ∎

Finally, we derive the following global theorem on the central trajectories for linear programming.

Theorem 4.4. (Total curvature of the central trajectory in the case of "classical linear programming")

If $\Omega = \mathbf{R}^n_+$ and $\psi(x) = -\sum_{i=1}^n \log x_i$, then the total curvature of the central trajectory is finite (exists in the improper sense) and is bounded as follows:

$$\max[I_{\mathcal{P}}(0,\infty), I_{\mathcal{D}}(0,\infty)] \leq \int_0^\infty h_{PD}(t)^{1/2} dt = O(n^{3.5}\log(\bar{\chi}^*_A + n)).$$

In particular, if A is integral, then

$$\max[I_{\mathcal{P}}(0,\infty), I_{\mathcal{D}}(0,\infty)] = O(n^{3.5} L_A),$$

where L_A is the input bit size of A, and in particular, if A is a 0-1 matrix, then

$$\max[I_{\mathcal{P}}(0,\infty), I_{\mathcal{D}}(0,\infty)] = O(n^{4.5} m).$$

Proof. The theorem readily follows from Theorem 4.1(b), Theorem 4.3 and the fact that $\bar{\chi}^*_A = 2^{O(L_A)}$. ∎

Remark: These total curvature bounds just depend on A but neither b nor c, while, as a geometric objects, these central trajectories depends on A, b and c. This is a remarkable property from a geometric viewpoint.

5 The number of iterations as an integral over the central trajectory

We proved in the last section that the integral $I_{PD}(\cdot)$ is an information geometric integral along the central trajectory. In this section, we demonstrate that the number of iterations is indeed predicted quite well with this integral even in the practical situation, through a *fully classical implementation of the MTY-PC algorithm.*

We implemented the MTY-PC algorithm precisely so that the polynomial convergence is guaranteed. We adopted the formulation by Monteiro and Adler to convert any LP problem into a primal-dual interior-feasible instance with a trivial interior feasible solution on central trajectory, and solved them with this algorithm. We run the algorithm changing β including $\beta = 1$, and plot (a) the number of iterations vs the duality gap, and (b) $\sqrt{\beta}\times$(the number of iterations) vs the duality gap. We show the results for six instances:

- AGG3 (Dimension of A: 517×302; Optimal value: 1.0312115935×10^7 (Figure 6))
- BNL1 (Dimension of A: 644×1175; Optimal value: 1.9776292856×10^3 (Figure 7))
- CYCLE (Dimension of A: 1904×2857; Optimal value: $-5.2263930249 \times 10^0$ (Figure 8))
- D2Q06C (Dimension of A: 2172×5167; Optimal value: 1.2278423615×10^5 (Figure 9))
- PILOT87 (Dimension of A: 2031×4883; Optimal value: 3.0171072827×10^2 (Figure 10))
- DFL001 (Dimension of A: 6072×12230; Optimal value: 1.12664×10^7 (Figure 11))

It is seen that all the curves in (a) overlaps in (b). From these figures, we conclude that the number of iterations of the MTY-PC algorithm is a very good approximation to a differential geometric integral reflecting the geometric nature of the problem.

6 Concluding discussion

In this paper, we made an attempt to bridge the computational complexity and information geometry. We developed a suitable geometrical machinery for studying complexity of the polynomial-time interior-point algorithms. Based on this framework, it was shown that the number of iterations of the primal-dual interior-point algorithms are written as the differential geometric integral over the central trajectory, and through numerical experiments that the integral explains the number of iterations of the interior-point algorithms quite well in a practical situation. As we mentioned in the introduction,

the number of iterations of interior-point algorithms is a differential geometric curvature integral.

We also derived a theorem on the central trajectory. In [28], Ohara and Tsuchiya analyzed the integral $I_{\mathcal{P}}$ ($I_{\mathcal{D}}$) represents the asymptotic iteration complexity of a polynomial-time primal interior-point algorithm. An extended analysis of this result is now under development.

Appendix: Connection and Covariant Derivatives

This appendix provides a quick introduction to connection and covariant derivatives. Connection is a quantity which determines parallel translation of vectors (and tensors) along a vector or a curve on a manifold. A covariant derivative is defined based on the connection and is an extension of the ordinary derivative. The covariant derivative of a vector is a vector while the ordinary derivative is not. See text books of general differential geometry for more detailed and formal description (e.g. [9, 11]).

Let $\mathcal{M}$ be a manifold and let $p \in \mathcal{M}$ be a point. Let $(x^1, \ldots, x^n)$ be its local coordinate in the neighborhood of p. Let $X_p, Y_p \in T_p\mathcal{M}$. We introduce a parallel tranlation operator $\Pi_{p,p+\delta t Y_p}$ to

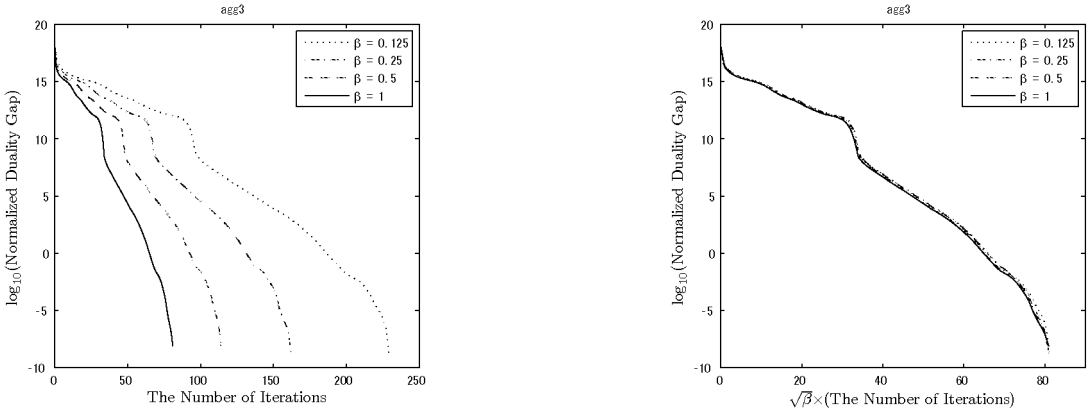

Figure 7: AGG3 (Dimension of A: 517×302; Optimal value: 1.0312115935×10^7)

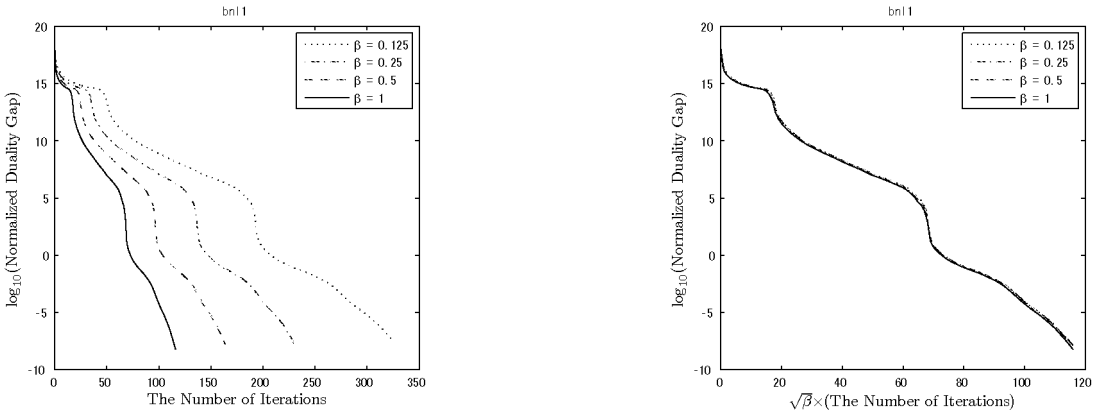

Figure 8: BNL1 (Dimension of A: 644×1175; Optimal value: 1.9776292856×10^3)

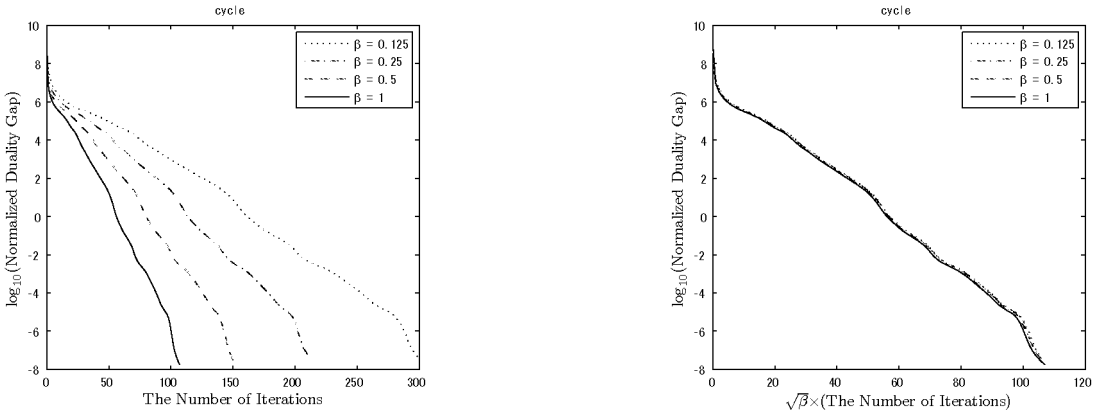

Figure 9: CYCLE (Dimension of A: 1904×2857; Optimal value: $-5.2263930249 \times 10^0$)

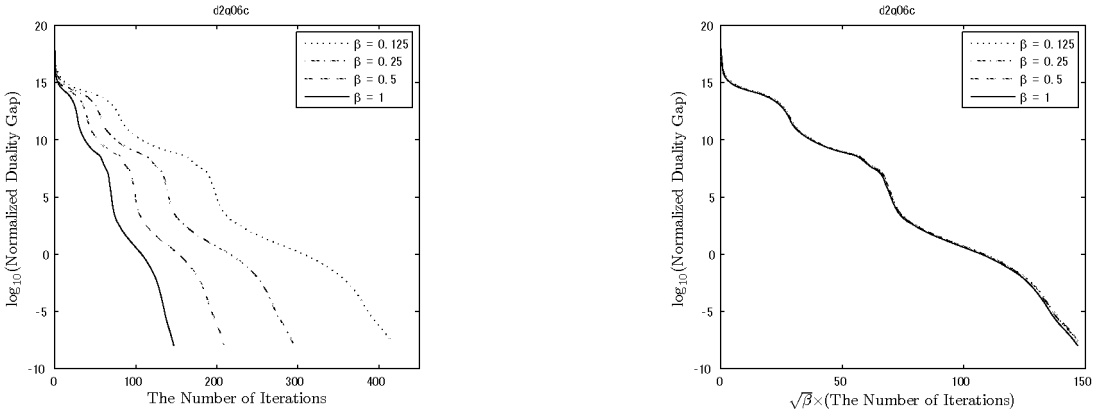

Figure 10: D2Q06C (Dimension of A: 2172×5167; Optimal value: 1.2278423615×10^5)

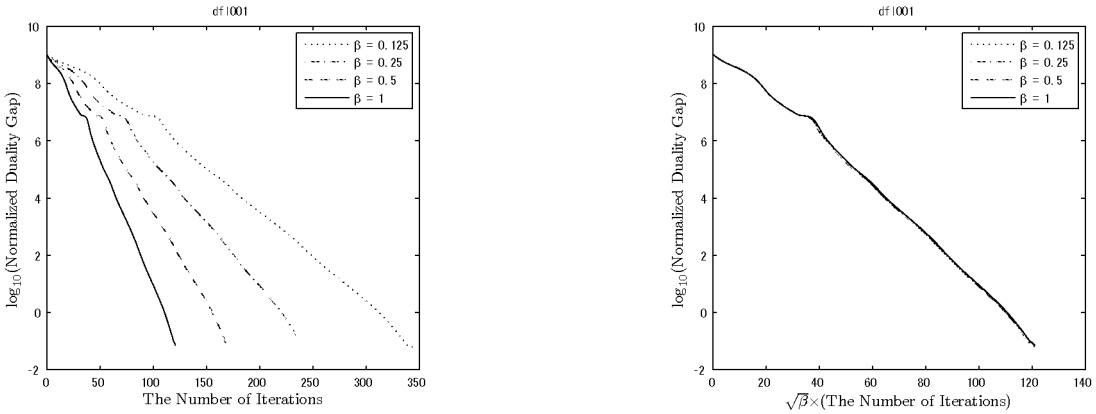

Figure 11: DFL001 (Dimension of A: 6072×12230; Optimal value: 1.12664×10^7)

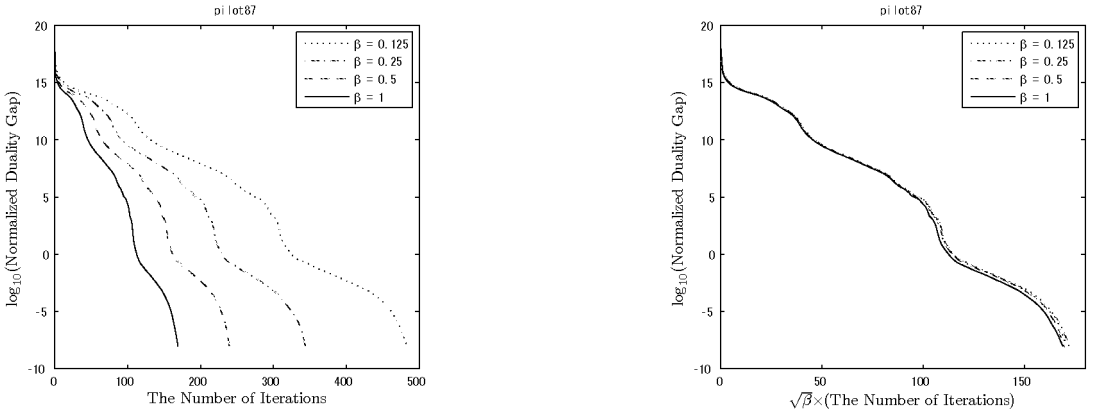

Figure 12: PILOT87 (Dimension of A: 2030×4883; Optimal value: 3.0171072827×10^2)

determine parallel translation of a tangent vector $X_p \in T_p\mathcal{M}$ to a tangent vector at $p + \delta t Y_p$, which is infinitesimally close to p (δt is "infinitesimally small"):

$$(\Pi_{p,p+\delta t Y_p} X_p)^i = X_x^i - \delta t \Gamma_{jk}^i X_x^j Y_x^k.$$

The function Γ is called the Christoffel symbol and determines a connection. Let $(y^1, \ldots, y^n)$ be another local coordinate in the neighborhood of p, and let $\widetilde{\Gamma}_{jk}^i$ be the Christoffel symbol in the coordinate system y. We require that Γ_{jk}^i obeys to the following rule in the change of the coordinate systems from x to y.

$$\widetilde{\Gamma}_{jk}^i = \sum_{a,b,c} \frac{\partial y^i}{\partial x^c} \frac{\partial x^a}{\partial y^j} \frac{\partial x^b}{\partial y^k} \Gamma_{ab}^c + \sum_a \frac{\partial y^i}{\partial x^a} \frac{\partial^2 x^a}{\partial y^j \partial y^k}.$$

Covariant derivative $\nabla_Y X$ is defined as an extension of partial derivative taking account of parallel translation of vectors.

$$(\nabla_Y X)_x^i = \lim_{\delta t \to 0} \frac{X_x^i(p + \delta t Y) - (\Pi_{p,p+\delta t Y} X(p))_x^i}{\delta t} = \frac{dX_x^i}{dt} + \Gamma_{jk}^i X_x^j Y_x^k.$$

Let $\gamma(t)$ be a curve, and $\dot{\gamma}(t)$ be the its tangent vector at $\gamma(t)$. We let $Y := \dot{\gamma}(t)$, and the covariant derivative of X along the curve γ in x-coordinate becomes:

$$(\nabla_{\dot{\gamma}} X(\gamma(t)))_x^i = \lim_{t \to 0} \frac{X_x^i(\gamma(t) + \delta t \dot{\gamma}(t)) - (\Pi_{\gamma(t),\gamma(t)+\delta t \dot{\gamma}(t)} X_{\gamma(t)})_x^i}{\delta t} = \frac{dX_x^i(\gamma(t))}{dt} + \Gamma_{jk}^i X_x^j \dot{x}^k(\gamma(t)).$$

By using the transformation rule of the Christoffel symbols, we verify that

$$(\nabla_{\dot{\gamma}} X(\gamma(t)))_y^a = \sum_i \frac{\partial y^a}{\partial x^i} (\nabla_{\dot{\gamma}} X(\gamma(t)))_x^i,$$

thus, $\nabla_{\dot{\gamma}} X$ is a vector.

The covariant derivative of a tensor is defined analogous to the covariant derivative of a vector. Here we just derive the form of the covariant derivative of the metric tensor G. For vectors X and Y, $G(X, Y)$ determines the inner product. Let $G_x = (G_{ij})$, $X_x = (X_x^i), Y_x = (Y_x^j)$ be vectors G, X, Y represented in the local coordinate x, and we consider the covariant derivative of G along the curve $\gamma(t)$. Since $G(X, Y)$ is a scalar function along $\gamma(t)$, we have

$$\frac{d}{dt} G_{\gamma(t)}(X_{\gamma(t)}, Y_{\gamma(t)}) = \frac{d}{dt}(X_x^\mathsf{T} G_x Y_x) = (\nabla_{\dot{\gamma}} X)_x^\mathsf{T} G_x Y_x + X_x^\mathsf{T} (\nabla_{\dot{\gamma}} G)_x Y_x + X_x^\mathsf{T} G_x (\nabla_{\dot{\gamma}} Y).$$

Based on this observation, we see that $(\nabla_{\dot{\gamma}} G)_x$ is given by

$$(\nabla_{\dot{\gamma}} G)_{ij} = \frac{dG_{ij}}{dt} - \sum_{l,k} G_{ki} \Gamma_{jl}^k \dot{x}^l - \sum_{u,l} G_{il} \Gamma_{ju}^l \dot{x}^u$$

Generally, we cannot elliminate the Cristoffel symbols globally, however, since a connection is not a tensor field, at each point $p \in \mathcal{M}$ we can take a local coordinate in which the Christoffel

symbols vanishes, i.e., $\Gamma^k_{ij} = 0$. If $\mathcal{M}$ admits a coordinate in which the Christoffel symbols vanishes globally, $\mathcal{M}$ is called a flat manifold (with respect to the connection) and the coordinate in which Γ^i_{jk} vanishes globally is called "an affine coordinate". If we fix a connection on $\mathcal{M}$ and take the associated affine coordinate, then the covariant derivative coincides with the ordinary derivative. x-coordinate and s-coordinate in this paper are affine coordinates associated with the connections ∇ and ∇^*, respectively.

References

[1] S. Amari: Differential geometry of curved exponential families — curvatures and information loss. *Annals of Statistics*, Vol.10 (1982), No. 2, pp. 357-385.

[2] S.-I. Amari: *Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics*, Vol. 28, Springer-Verlag, 1985.

[3] S.-I. Amari and H. Nagaoka: *Methods of Information Geometry.* AMS, Province, RI, USA. (An extended English translation of the book in Japanese published under the same title in 1993 from Iwanami-shoten, Tokyo, Japan.)

[4] D. Bayer, J. Lagarias: The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories. *Transactions of American Mathematical Society*, Vol. 314 (1989), No. 2, 499-526.

[5] D. Bayer, J. Lagarias: The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories. *Transactions of American Mathematical Society*, Vol. 314 (1989), No. 2, 527-581.

[6] I. Dikin: On the convergence of an iterative process (in Russian). *Upravlyaemye Sistemi*, Vol. 12 (1974), pp. 54-60.

[7] J. J. Duistermaat: On the boundary behaviour of the Riemannian structure of a self-concordant barrier function. *Asymptotic Analysis*, Vol. 27 (2001), No. 1, 9–46.

[8] C. C. Gonzaga and H. J. Lara: A note on properties of condition numbers. *Linear Algebra and its Applications.* Vol. 261 (1997), pp. 269–273.

[9] S. Helgason: *Differential Geometry and Symmetric Spaces* (2nd Revised edition). AMS, 2001.

[10] N. Karmarkar: Riemannian geometry underlying interior-point methods for linear programming. In *Mathematical Developments Arising From Linear Programming* (eds. J. C. Lagarias and M. J. Todd), pp. 51–75, *Contemporary Mathematics*, Vol. 114, AMS, Providence, 1990.

[11] S. Kobayashi and K. Nomizu: *Foundations of Differential Geometry Volume I,II.* John-Wiley and Sons, 1963, 1969.

[12] M. Kojima, S. Mizuno and A. Yoshise: A primal-dual interior point algorithm for linear programming. In *Progress in Mathematical Programming, Interior-point and Related Methods* (ed. N. Megiddo), Springer-Verlag, New York, 1989, pp. 29–47.

[13] M. Kojima, S. Mizuno and A. Yoshise: A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, Vol. 44 (1989), pp. 1–26.

[14] N. Megiddo, S. Mizuno and T. Tsuchiya: A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, Vol. 82 (1998), pp. 339-355.

[15] S. Mizuno, M. J. Todd, and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, Vol. 18 (1993), pp. 964–981.

[16] R. D. C. Monteiro and I. Adler: Interior path following primal-dual algorithms. Part I: Linear programming. *Mathematical Programming*, Vol. 44 (1989), pp. 27–41.

[17] R. D. C. Monteiro and T. Tsuchiya: A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, Vol. 13 (2003), pp. 1054–1079.

[18] R. D. C. Monteiro and T. Tsuchiya: A new iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM Journal on Optimization*, Vol. 15 (2004), pp. 319–347.

[19] 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. *Mathematical Programming* (2007) (published on-line).

[20] 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, Tokyo, Japan, October 1982.

[21] Yu. Nesterov and A. Nemirovskii: *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, 1994.

[22] Yu. Nesterov and A. Nemirovski: Primal central paths and Riemannian distances for convex sets. *Foundations of Computational Mathematics*, Vol. 8, pp.533-560.

[23] Yu. Nesterov and M. J. Todd: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, Vol. 2 (2002), pp.333–361.

[24] A. Ohara: Information geometric analysis of semidefinite programming problems (in Japanese). *Proceedings of the Institute of Statistical Mathematics*, Vol. 46 (1998), no. 2, 317–334.

[25] A. Ohara: Information geometric analysis of an interior point method for semidefinite programming. In *Geometry in Present Day Science* (eds. O. E. Barndorff-Nielsen and E. B. V. Jensen), pp. 49–74, World Scientific, 1999.

[26] A. Ohara: Geodesics for dual connections and means on symmetric cones *Integral Equations and Operator Theory*, Vol.50 (2004), pp. 537–548.

[27] A. Ohara, N. Suda and S.-I. Amari: Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra and its Applications*, Vol. 247 (1996), pp. 31–53.

[28] A. Ohara and T. Tsuchiya: An Information Geometric Approach to Polynomial-time Interior-point Algorithms: Complexity Bound via Curvature Integral. *Research Memorandum* No. 1055, The Institute of Statistical Mathematics, December 2007.

[29] G. Sonnevend, J. Stoer and G. Zhao: On the complexity of following the central path of linear programs by linear extrapolation: II. *Mathematical Programming*, Vol. 52 (1991), pp. 527–553.

[30] G. W. Stewart: On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, Vol. 112 (1989) pp. 189–193.

[31] K. Tanabe: Geometric method in nonlinear programming. *Journal of Optimization Theory and Applications*, Vol. 30 (1980), pp. 181–210.

[32] K. Tanabe: Center flattening transformation and a centered Newton method for linear programming. (Manuscript presented at MP seminar, The Operations Research Society of Japan, July, 1987.)

[33] K. Tanabe: Centered Newton method for mathematical programming. In *System Modeling and Optimization* (eds. M. Iri and K. Yajima), pp. 197-206, Springer-Verlag, Berlin, 1988.

[34] K. Tanabe: Differential geometry of optimization (in Japanese). *Preliminary issue of the Bulletin of the Japan Society for Industrial and Applied Mathematics*, No. 3 (1990), pp.39–50.

[35] K. Tanabe: *Centered Newton Methods and Differential Geometry of Optimization. Cooperative Research Report*, Vol. 89, The Institute of Statistical Mathematics, November, 1996.

[36] K. Tanabe and T. Tsuchiya: New geometry of linear programming (in Japanese). *Suurikagaku*, No. 303 (1988), pp. 32–37.

[37] M. J. Todd: A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, Vol. 38 (1990) pp. 1006–1018.

[38] M. J. Todd, L. Tunçel and Y. Ye: Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Mathematical Programming*, Vol. 90 (2001), No. 1, pp. 59–70.

[39] L. Tunçel: Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. *Mathematical Programming*, Vol. 86 (1999), No. 1, pp. 219–223.

[40] K. Uohashi and A. Ohara: Jordan algebras and dual affine connections on symmetric cones. *Positivity*, Vol. 8 (2004), pp. 369–378.

[41] S. Vavasis and Y. Ye: A primal-dual accelerated interior point method whose running time depends only on A. *Mathematical Programming*, Vol. 74 (1996), pp. 79–120.

[42] G. Zhao: On the relationship between the curvature integral and the complexity of pathfollowing methods in linear programming. *SIAM Journal on Optimization*, Vol. 6 (1996), pp. 57–73.

[43] G. Zhao and J. Stoer: Estimating the complexity of a class of path-following methods for solving linear programs by curvature integrals. *Applied Mathematics and Optimization*, Vol. 27 (1993), pp. 85–103.