数理解析研究所講究録
1004巻 1997年 71-89

Information Geometric Analysis of a Interior-Point Method for Semidefinite Programming

大阪大学基礎工学部　小原 敦美　(Atsumi Ohara)

1 Introduction

Since Karmarkar proposed an interior point algorithm with polynomial-time complexity to solve linear programming (LP) problems [11], many works have contributed to the progress and applications of interior-point methods of mathematical programming (See for some text books [16, 5, 21, 22]).

Among them, some researchers have brought differential geometric points of view and elucidated mathematical structures behind the mechanism of interior-point methodology [2, 12, 20, 8, 6].

In particular, (continuous version of) affine scaling trajectories in LP has been known integrable via Legendre transformation and regarded as geodesics for a certain connection [20, 6]. These results are crucially relying on several dualistic properties of the problem and one of the key tool unifying them has been so-called *information geometry* [1]. As for the integrability of affine scaling trajectories, we should also refer to [10, 4, 15, 7].

On the other hand, recent study is developing the applicable area of polynomial-time interior-point framework to broader class of convex programming problems. One of the most significant in the engineering application is semidefinite programming (SDP), which is actually useful in system and control theory [3] and combinatorial optimization [21]. Further, the set of positive definite matrices is one of the examples where dualistic nature appears in the simplest way and is easy to analyze [18], e.g., Legendre transformation turns out to be essentially matrix inversion. This fact implies that analyzing SDP enables us to exploit abundant dualistic structure of interior-point methodology via direct calculations.

This paper first introduce preliminary results of information geometry in Section 2.

Section 3 discusses the interior-point machinery for the general convex programming in terms of the frame work of information geometry. In this case, we again find that various dualistic structure on a considering convex region, such as Legendre transformation, dual connection and so on, naturally appear and play important roles. This part can be regarded as a simple extension of the work given by [20, 6] in LP case.

Next, by examining the above results, we consider in Section 4 the possibility of a new algorithm using Legendre transformation. Consequently, we show some class of nontrivial problems in SDP can be solved without any iterations. This class is characterized by geometric term: ∇^*-autoparallelism. Further this class turns out to be related to Jordan subalgebra of symmetric matrices under a certain circumstance.

Finally in Section 5, we exploit the above result quantitatively, i.e., we show a certain geometric quantities called the second fundamental form (, or Euler-Schouten embedding curvature) is directly related to predictor step size we can take without increasing computational complexity in the succeeding corrector phase. This result shows one approach to analyze how over all computational complexity is dependent on structure of each problem to be solved.

Notation: $Sym(n)$: the set of n by n real symmetric matrices, $PD(n)$: the set of n by n real positive definite matrices, $N : (= n(n+1)/2)$ dimension of the vector space $Sym(n)$, δ_i^j: Kronekar's delta.

Further, we obey the following convention for the simplicity of notation:
Convention: When Super- or subscripts are used without any range, Italic letters p, q, r are supposed to index integers from 1 to N, e.g., $p = 1, \ldots, N$. Similarly, the other Italic and Greek super- or subscripts are respectively supposed to index integers from 1 to m and from $m+1$ to N, e.g., $i = 1, \ldots, m$ and $\kappa = m+1, \ldots, N$.

2 Preliminaries for Information Geometry

In this section, we give a brief introduction of some results in dualistic geometry to be used in the following sections. Those who are interested in the details and further results can refer to [1, 6, 18].

Let $\mathcal{M}$ be an arbitrary open convex set in $\mathbf{R}^m$. One of the simplest way to define information geometry on $\mathcal{M}$ is using a convex potential function. As is shown in the followings, we can obtain key geometric quantities which define dualistic structures as derivatives of the function.

Let $\psi(x)$ be any smooth convex function on $\mathcal{M}$ that has positive definite Hessian matrix. , where (x^i) is a coordinate system for $\mathbf{R}^m$. Now consider a Riemannian manifold $(\mathcal{M}, g)$, where its Riemannian metric is given by the Hessian matrix of $\psi(x)$, i.e.,

$$g_{ij}(x) := \partial_i \partial_j \psi(x), \text{ where } \partial_i := \frac{\partial}{\partial x^i}. \tag{2.1}$$

Here $g_{ij}(x)$ represent the components of Riemannian metric g.

Next we will introduce a pair of connections ∇ and ∇^* on our Riemannian manifold $(\mathcal{M}, g)$. While in mathematical physics the Levi-Civita connection on Riemannian manifolds plays an important role, non-Levi-Civita connections are crucial in our frame work. Let denote the components of ∇ and ∇^* by

$$\Gamma_{ijk}(x) = g(\nabla_{\partial_i} \partial_j, \partial_k), \quad \Gamma^*_{ijk}(x) = g(\nabla^*_{\partial_i} \partial_j, \partial_k). \tag{2.2}$$

and define these connections respectively by their components:

$$\Gamma_{ijk}(x) := [ij:k](x) - \frac{1}{2} T_{ijk}(x) = 0, \quad \Gamma^*_{ijk}(x) := [ij;k](x) + \frac{1}{2} T_{ijk}(x) = T_{ijk}(x) = \partial_i g_{jk}(x). \tag{2.3}$$

Here, $[ij;k]$ represent the components of the Riemannian (Levi-Civita) connection:

$$[ij:k] := \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}).$$

and

$$T_{ijk}(x) := \partial_i \partial_j \partial_k \psi(x), \tag{2.4}$$

Since $\Gamma_{ijk}(x) = 0$, the coordinate system x is called ∇-*affine*. Although ∇ and ∇^* are not metric-preserving, (2.3) implies

$$\partial_i g_{jk} = \Gamma_{ijk} + \Gamma^*_{ikj}, \quad \text{i.e.,} \quad Ag(B, C) = g(\nabla_A B, C) + g(B, \nabla^*_A C) \tag{2.5}$$

for any vector fields A, B and C on $\mathcal{M}$. Due to this property, we say ∇ and ∇^* are *mutually dual* with respect to g.

Thus, we can easily derive the structure of information geometry $(\mathcal{M}, g, \nabla, \nabla^*)$ from the potential function $\psi(x)$ on $\mathcal{M}$.

In case that dual connections ∇ and ∇^* are derived from a potential function in the manner of (2.3), it is known that torsion and curvature tensors on $\mathcal{M}$ with respect to ∇ and ∇^* vanish. If this is the case, we call $\mathcal{M}$ *dually flat*.

When $\mathcal{M}$ dually flat, we can introduce a new coordinate system (y_i) and convex (with respect to y) function $\phi(y)$ via Legendre transform:

$$y_i := \partial_i \psi(x), \quad \phi(y) := \psi^*(y) = \sup_{x \in \mathcal{M}} \{x^i y_i - \psi(x)\}. \tag{2.6}$$

Note that we use subscripts for the components of y. Let us define as $\partial^i := \partial/\partial y_i$, then the components of Riemannian metric g and the dual connections ∇ and ∇^* with respect to y are represented in the dual manner as

$$x^i = \partial^i \phi(y), \tag{2.7}$$

$$g^{ij}(y) = \partial^i \partial^j \phi(y), \tag{2.8}$$

$$\Gamma^{ijk}(y) = \partial^i \partial^j \partial^k \phi(y), \qquad \Gamma^{*ijk}(y) = 0. \tag{2.9}$$

Since $\Gamma^{*ijk}(y) = 0$, y is similarly called ∇^*-*affine*. A crucial point is that Jacobian matrix of the transformation between x and y is just g, i.e.,

$$\partial_i = g_{ij} \partial^j, \quad \partial^i = g^{ij} \partial_j, \tag{2.10}$$

and $(g_{ij}(x))$ is an inverse matrix of $(g^{ij}(y))$ at any same point specified by x and y: $p = p(x) = p(y)$. Note that it follows from (2.10)

$$g(\partial_i, \partial^j) = \delta_i^j. \tag{2.11}$$

Finally, we should note that using the components of the connections, we can represent differential equations of their geodesics as follows: For a geodesic with respect to the connection ∇, we have

$$\ddot{x}^i(t) = 0, \quad \sum_{j=1}^{m} g^{ij} \ddot{y}_j(t) + \sum_{j,k=1}^{m} \Gamma^{ijk} \dot{y}_j(t) \dot{y}_k(t) = 0,$$

which are represented in x and y coordinate system, respectively. Similarly for a geodesic with respect to the connection ∇^*,

$$\sum_{j=1}^{m} g_{ij} \ddot{x}^j(t) + \sum_{j,k=1}^{m} \Gamma^*_{ijk} \dot{x}^j(t) \dot{x}^k(t) = 0, \quad \ddot{y}_i(t) = 0.$$

Here, $\dot{}$ denotes derivative by the parameter t.

3 Some Interior Point Methods and Geodesics

Here, we show the property of continuous trajectories associated with some (primal) interior point methods in terms of information geometry. Next, we discuss the possibility of new algorithm based on this result.

Given a vector $c \in \mathbf{R}^m$ and convex functions $f_i(x), i = 1, \ldots, h$, consider the following convex programming problem with a linear objective function:

$$\min c^T x, \quad \text{s.t. } x \in \mathcal{M} \subset \mathbf{R}^m, \mathcal{M} = \{x | f_i(x) \leq 0, i = 1, \ldots, h\}. \tag{3.1}$$

Note that we can transform general convex programming problems with arbitrary convex objective function $f_0(x)$:

$$\min f_0(x), \quad \text{s.t. } x \in \mathcal{M} \subset \mathbf{R}^m, \mathcal{M} = \{x | f_i(x) \leq 0, i = 1, \ldots, h\} \tag{3.2}$$

to the the form (3.1). By introducing new variable x^{m+1} and convex function

$$f_{h+1}(x, x^{m+1}) := f_0(x) - x^{m+1}, \tag{3.3}$$

we obtain the equivalent problem

$$\begin{gathered} \min x^{m+1} = c^T \tilde{x}, \quad \tilde{x} \in \tilde{\mathcal{M}} \subset \mathbf{R}^{m+1}, \\ c^T = [\, 0 \ \ldots \ 1 \,], \quad \tilde{x}^T = [\, x^T \ x^{m+1} \,], \\ \tilde{\mathcal{M}} = \{\tilde{x} | f_i(x) \leq 0, i = 1, \ldots, h+1\}, \quad f_i(x) : \text{convex}, \quad i = 0, \ldots, h+1. \end{gathered}$$

Thus, we can regard (3.1) as one of standard forms for convex programming problems. In this section, we will assume $\mathcal{M}$ is bounded.

Let us consider a barrier function $\phi(x)$ for $\mathcal{M}$ that satisfies the following conditions:

B1) $\psi(x)$ is three times continuously differentiable on $\mathcal{M}$,

B2) Hessian matrix of $\psi(x)$ is positive definite on int$\mathcal{M}$,

B3) $\psi(x) \to \infty, (x \to \partial\mathcal{M})$.

For example, when $f_i(x)$ is linear or quadratic (, more generally, *relatively Lipschitz* [17])

$$\psi(x) := -\sum_{i=1}^{h} \log(-f_i(x)) \tag{3.4}$$

is known to satisfy the above conditions. Such a function $\psi(x)$ is said *logarithmic barrier* for $\mathcal{M}$ and its unique minimizer on $\mathcal{M}$, called *analytic center*, plays an important role in interior-point methodology.

Now among some classes of interior-point algorithms, we consider *path following* and *affine scaling* methods for the problem (3.1) and analyze their associated continuous trajectories.

One of the simplest path following methods is *barrier method*, which involves the minimization problem for the following weighted sum of the barrier and objective functions:

$$\Psi_t(x) := tc^T x + \psi(x) \to \min, \tag{3.5}$$

where $t > 0$ is a weighting parameter. Let $x^\sharp(t)$ and $x^\sharp$ be a minimizer of $\Psi_t(x)$ and the original convex problem (3.1), then we obtain

$$x^\sharp(t) \to x^\sharp \quad (t \to +\infty).$$

For the implementation, by discretizing properly t as an increasing sequence $\{t_i\}$ and finding, for each t_i, approximant of $x^\sharp(t_i)$ by Newton method, we obtain a sequence that follows near the path $x^\sharp(t)$ converging to $x^\sharp$. The path $x^\sharp(t)$ is called *central path.*

In addition to this method, there are also some variants of the path following methods, which minimize, for example,

$$\Psi_s(x) := -\zeta \log(s - c^T x) + \psi(x) \to \min, \quad \text{for given } \zeta > 0 \tag{3.6}$$

or

$$\psi(x) \to \min, \quad \text{s.t.} \quad c^T x = \tau. \tag{3.7}$$

Let $x^\sharp(s)$ and $x^\sharp(\tau)$ be the minimizers of (3.6) and (3.7), respectively, then we find they converge to $x^\sharp$ when the parameters t or τ approach the optimal value of $c^T x$ from the above [16]. Note that, however, $x^\sharp(s)$ and $x^\sharp(\tau)$ coincide with the central path $x^\sharp(t)$ with different parameterizations.

Thus, the above three types of path-following methods finally generate sequences that converge to the optimal solution $x^\sharp$ following the central path. Further, when the barrier function $\psi(x)$ has a property called *self-concordance* [16], i.e.,

$$|\sum_{i,j,k=1}^{m} \Gamma^*_{ijk} h^i h^j h^k| \le a(\sum_{i,j=1}^{m} g_{ij} h^i h^j)^{3/2}, \quad \forall h = (h^i) \in \mathbf{R}^m, \exists a \in \mathbf{R}, \tag{3.8}$$

these methods are known to work efficiently in terms of worst-case computational complexity.

On the other hand, affine scaling method is essentially a gradient method along the vector field:

$$\dot{x} = g^{-1}(x)c, \tag{3.9}$$

where $g(x)$ is the Hessian matrix of $\psi(x)$, i.e.,

$$g(x) := (g_{ij}(x)), \quad g_{ij}(x) := \partial^2 \psi(x)/\partial x^i \partial x^j.$$

The right-hand side of (3.9) can be interpreted as gradient vector for the objective function $c^T x$ with respect to the Riemannian metric g defined on $\mathcal{M}$.

Now we show both central path and affine scaling trajectories are characterized as ∇-geodesics of dualistic geometry derived from the barrier function $\psi(x)$. This fact was generalization of [20] for linear programming case. Let us consider

$$\Psi_t(x) := tc^T x + \psi(x), \quad c^T := [\, c_1 \quad c_2 \quad \dots \quad c_m \,]. \tag{3.10}$$

Since the Hessian matrix of $\Psi_t(x)$ is also positive definite, $x^\sharp(t)$ is a unique solution of the optimality condition:

$$\frac{\partial \Psi_t(x)}{\partial x^i} = tc_i + \frac{\partial \psi(x)}{\partial x^i} = 0, \quad i = 1, \dots, m. \tag{3.11}$$

This condition can be represented in dual coordinate system as follows:

$$y_i(t) = -c_i t, \quad t > 0, \quad i = 1, \ldots, m. \tag{3.12}$$

Thus, the central path $y^\sharp(t)$ turns out to be *a half straight line* in the dual coordinate system, where its initial point is the origin. Since $y_i(t)$ in (3.12) satisfies

$$\frac{dy_i}{dt} = -c_i, \quad i = 1, \ldots, m, \tag{3.13}$$

we have differential equations for the central path $x^\sharp(t)$ using (2.10)

$$\frac{dx^j}{dt} = -\sum_{j=1}^{m} g^{ji} c_i = -\sum_{j=1}^{m} g^{ji} \partial_i (c^T x), \quad j = 1, \ldots, m, \tag{3.14}$$

with its initial point $x(0) = x_{\mathrm{AC}}$, i.e., analytic center for $\mathcal{M}$. This is the same differential equation with (3.9) for the affine scaling trajectory, which implies the central path is also a solution of the gradient system of the object function $c^T x$ under the Riemannian metric g.

Finally, multiply both sides of (3.14) by g_{ij} and differentiate them by t, then we obtain the differential equation of ∇^*-geodesics:

$$\sum_{j=1}^{m} g_{ij} \frac{d^2 x^j}{dt^2} + \Gamma^*_{ijk} \frac{dx^j}{dt} \frac{dx^k}{dt} = 0, \quad i = 1, \ldots, n, \quad \text{where} \sum_{j,k=1}^{m} \Gamma^*_{ijk} = \partial_i \partial_k \partial_j \psi = \partial_k g_{ij}. \tag{3.15}$$

Hence both central path and affine scaling trajectories are proved to be ∇^*-geodesics. Further, in contrast with the primal coordinate case, following these trajectories in the dual coordinate system is a very easy task with no iteration, i.e., just extending a straight line as shown in (3.13).

Therefore, one intuitive idea that arises from this fact is that following the central path or affine scaling trajectories in the dual coordinate system, via Legendre transformation, may have better performance than doing so in the primal coordinate system, in terms of computational complexity. Additionally when $\psi(x)$ is a self-concordant, there also exists the relation between t and an error [16]

$$c^T x(t) - c^T x^\sharp \leq \vartheta / t, \tag{3.16}$$

where the constant ϑ is a parameter depending only on $\psi(x)$. Consequently, the rough sketch of the algorithm is as follows:

Algorithm

Step 1. Find the initial feasible point $x^\sharp(0) \in \mathcal{M}$.

Step 2. Calculate $y^\sharp(0)$, the Legendre transform of $x^\sharp(0)$, using (2.6).

Step 3. For sufficiently large t (decided by the accuracy estimate (3.16)), calculate $y^\sharp(t)$ by (3.12).

Step 4. Find $x^\sharp(t)$, *inverse* Legendre transform of $y^\sharp(t)$, by (2.7).

However, this optimistic idea, of course, turns out false. Generally, obtaining the Legendre conjugate $\psi^*(y)$ explicitly as a function of y is difficult. Hence, to find $x^\sharp(t)$ for given $y(t)$ in Step 4, we must solve, instead of using (2.7) directly, the following nonlinear equation:

given $y_i(t), i = 1, \ldots, m$, find $x_i(t)$, s.t.,

$$\frac{\partial \psi}{\partial x^i}(x) = y_i(t), \quad i = 1, \ldots, m, \tag{3.17}$$

or equivalently the following convex problems:

$$\arg \min_x \{\psi(x) - \sum_{i=1}^m y_i(t) x^i\}, \quad i = 1, \ldots, m.$$

Even if $\psi(x)$ is self-concordant, computing $x^\sharp(t)$ in such a way might cost more, without good initial estimates for $x^\sharp$, than the path-following methods mentioned before.

4 Semidefinite Programming

4.1 Information Geometry for Positive Definite Matrices

In this section, we specialize the previous discussions to semidefinite programming (SDP) case.

Let $c \in \mathbf{R}^m$ and $E_i \in Sym(n), i = 0, \ldots, m$, where $\{E_i\}_{i=1}^m$ are linearly independent basis of $Sym(n)$. Consider the following form of SDP problems:

$$\min_x c^T x, \text{ s.t. } P(x) = E_0 + \sum_{i=1}^m x^i E_i \geq 0. \tag{4.1}$$

We introduce notation for some sets associated with the above SDP problem. Denote by $\mathcal{V}$ and $E_0 + \mathcal{V}$, vector and affine subspaces in $Sym(n)$ respectively defined as

$$\mathcal{V} := \{X | X \in \text{span}\{E_i\}_{i=1}^m\}, \quad E_0 + \mathcal{V} := \{X | X - E_0 \in \mathcal{V}\}. \tag{4.2}$$

Then the constrained set of $P(x)$ in (4.1) is the closure of $\mathcal{L}$ which is defined as so-called *conic form* [16], i.e.,

$$\mathcal{L} := PD(n) \cap (E_0 + \mathcal{V}). \tag{4.3}$$

Note that $\mathcal{V}$ coincides with tangent space of $\mathcal{L}$.

In SDP case, path-following algorithms commonly use a self-concordant function

$$\psi(x) = -\log \det P(x) \tag{4.4}$$

as a barrier. Then, according to the section 3, we can introduce information geometric structure $(\mathcal{L}, g, \nabla, \nabla^*)$ to define dual coordinate, Riemannian metric and ∇^*-connection as follows [18]:

$$y_i = -\text{tr}(P(x)^{-1} E_i), \tag{4.5}$$

$$g_{ij}(x) = \text{tr}(P(x)^{-1} E_i P(x)^{-1} E_j), \tag{4.6}$$

$$\Gamma^*_{ijk}(x) = -2\text{tr}(P^{-1}(x) E_i P(x)^{-1} E_j P(x)^{-1} E_k). \tag{4.7}$$

The affine-scaling trajectories and central path are still a solution of the ∇^*-geodesic equation (3.15) on the interior of $\mathcal{L}$. As described in the previous section, Legendre conjugate $\psi^*(y)$ and transform $x = x(y)$ can not be generally expressed as functions of y.

Now we turn our points of view to treating the problem in $PD(n)$. Let $\{E_i\}_{i=m+1}^N$ be any basis matrices of a subspace complementary to $\mathcal{V}$ and define the set of bi-orthogonal basis matrices $\{E^j\}_{j=1}^N$ as follows:

$$-\operatorname{tr}\{E^i E_j\} = \delta_j^i, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N. \tag{4.8}$$

Then we can express any $P \in Sym(n)$ as

$$P = P(\theta) := \sum_{i=1}^{N} \theta^i E_i. \tag{4.9}$$

Hence, $\theta = (\theta^1 \ldots \theta^N)^T$ is regarded as a coordinate system for $PD(n)$. Note that θ coordinate for $P(x) \in \mathcal{L}$ is

$$\theta^i = x^i + \theta_0^i, i = 1, \ldots, m, \quad \theta^i = \theta_0^i, i = m+1, \ldots, N, \tag{4.10}$$

$$\theta = \begin{pmatrix} x \\ 0 \end{pmatrix} + \theta_0, \tag{4.11}$$

where $\theta_0 = (\theta_0^1 \ldots \theta_0^N)^T$ is a θ coordinate for E_0, i.e., $E_0 = \sum_{i=1}^N \theta_0^i E_i$.

To introduce information geometric structure on $PD(n)$ [18], we use a barrier (potential) function on the whole $PD(n)$

$$\psi(\theta) = -\log\det P(\theta). \tag{4.12}$$

Then the Legendre transform of θ, denoted by η, is

$$\eta_i(\theta) = -\operatorname{tr}(P(\theta)^{-1} E_i). \tag{4.13}$$

Note that for all P on $\mathcal{L}$

$$y_i = \eta_i, \quad i = 1, \ldots, m \tag{4.14}$$

holds from (4.5) and (4.13). The dual coordinate $\eta = (\eta_1 \ldots \eta_N)^T$ represents any $P \in PD(n)$ as follows:

$$P = P(\eta) := \left(\sum_{i=1}^{N} \eta_i E^i \right)^{-1}. \tag{4.15}$$

Riemannian metric g and coefficients of ∇^*-connection are also in the same form as (4.6) and (4.7) except x replaced by θ. We will use the same notation g, ∇, ∇^* for both $\mathcal{L}$ and $PD(n)$. Information geometric structure $(\mathcal{L}, g, \nabla, \nabla^*)$ coincides with induced one from $(PD(n), g, \nabla, \nabla^*)$ defined like this.

A significant point for the case of $PD(n)$ is that we can give explicit expressions of Legendre conjugate $\psi^*(\eta)$ and inverse Legendre transform as functions of η:

$$\psi^*(\eta) = -\log\det P(\eta)^{-1}, \tag{4.16}$$

$$\theta^i(\eta) = -\operatorname{tr}(P(\eta) E^i). \tag{4.17}$$

Thus, from (4.13) and (4.17) the (inverse) Legendre transformation on $PD(n)$ from one coordinate to the other is essentially inverting symmetric positive definite matrices and is very cheaply executed without iterative optimization procedures as in general cases.

4.2 ∇^*-autoparallel submanifold in $PD(n)$

Now computational difficulty of the inverse Legendre transform is avoided in case of $PD(n)$. Further, the central path and affine-scaling trajectories are ∇^*-geodesics in $\mathcal{L}$ and there holds a relation (4.14). Then the natural question that arises is as follows:
Question: Does the algorithm considered in the previous section work without iterative optimization procedures in $PD(n)$?

Unfortunately, the answer is generally **NO**. The reason is that ∇^*-geodesic of the submanifold $\mathcal{L}$ is not always that of the ambient manifold $PD(n)$. However, if both ∇-geodesics coincide, the answer for the question is affirmative. Such a property is characterized by a geometric term: *totally geodesic* or *autoparallel*.

We use the following definition and fact:
Definition [13] : Let $\mathcal{N}$ be a manifold equipped with a certain connection $\bar{\nabla}$ and $\mathcal{M}$ be a submanifold with induced connection from $\bar{\nabla}$ of ambient manifold N. The submanifold $\mathcal{M}$ is said to be $\bar{\nabla}$-*autoparallel* if for any vector fields X and Y on $\mathcal{M}$, covariant derivative $\bar{\nabla}_X Y$ is again tangent to $\mathcal{M}$ at every point $x \in \mathcal{M}$.
Proposition 4.1 [13] : Let $(\mathcal{N}, \bar{\nabla}$ is torsion-free, then a submanifold $\mathcal{M}$ is autoparallel in $\mathcal{N}$ iff $\mathcal{M}$ is totally geodesic, i.e., an every geodesic of $\mathcal{N}$ starting from a point on $\mathcal{M}$ with an initial vector tangent to $\mathcal{M}$ stays in $\mathcal{M}$.

Recalling that the ∇^* connection is torsion-free, we immediately obtain a main result of this section:
Theorem 4.2: If $\mathcal{L}$ is ∇^*-autoparallel, we can solve SDP problem (4.1) without any optimization procedure by executing the algorithm in the section 3.
Remark: i) The converse is not true, i.e., there exist SDP problems that can be solved by the algorithm while $\mathcal{L}$ associated with them are not ∇^*-autoparallel. This is due to combination of the initial point and c (See example 2 below). ii) The submanifold $\mathcal{L}$ is always ∇-autoparallel by its definition [18]. Hence ∇^*-autoparallel submanifold $\mathcal{L}$ is dually autoparallel with respect to both ∇ and ∇^* connections.

To check whether the $\mathcal{L}$ is ∇^*-autoparallel, we have the following condition:
Lemma 4.3: A submanifold $\mathcal{L}$ of (4.3) in $PD(n)$ is ∇^*-autoparallel if and only if

$$-E_i P^{-1} E_j - E_j P^{-1} E_i \in \mathcal{V}, \quad \text{for } \forall P \in \mathcal{L}, \quad 1 \leq i, j \leq m. \tag{4.18}$$

Proof: Since the tangent space of $PD(n)$ is isomorphic to $Sym(n)$, we can identify the tangent vector ∂_i at P and the symmetric matrix E_i. Then ∇^*-covariant derivative $\nabla^*_{\partial_i} \partial_j$ can be shown to have the following matrix representation [18]:

$$\nabla^*_{\partial_i} \partial_j \equiv -E_i P^{-1} E_j - E_j P^{-1} E_i,$$

where the symbol $\equiv$ denotes the above identification. Hence, the statements immediately follows. Q.E.D.

Especially, when $E_0 = 0$, i.e., $\mathcal{L}$ is convex subcone of $PD(n)$, the tangent space $\mathcal{V}$ of ∇^*-autoparallel submanifold $\mathcal{L}$ has a relation with *Jordan subalgebra* of $Sym(n)$.

Definition [14, 9](Jordan algebra of $Sym(n)$): Let us define the product $*$ on the vector space $Sym(n)$ by means of usual matrix product as follows:

$$X * Y := \frac{1}{2}(XY + YX). \tag{4.19}$$

We call the vector space $Sym(n)$ equipped with the usual matrix sum $+$ and the product $*$ *Jordan algebra* of $Sym(n)$.

Theorem 4.4: Assume $E_0 = 0$ and the identity matrix I is an element of submanifold $\mathcal{L}$ of the form (4.3). Then $\mathcal{L}$ is ∇^*-autoparallel (and hence is dually autoparallel in the sense of both ∇ and ∇^* connections) if and only if $T_P\mathcal{L}(= \mathcal{V})$ is Jordan subalgebra of $Sym(n)$.

Proof: First note that $\mathcal{L} \subset \mathcal{V}$ due to $E_0 = 0$. Let $A, B \in T_P\mathcal{L} = \mathcal{V}$ be represented as

$$A = \sum_{i=1}^{m} a^i E_i, \quad B = \sum_{i=1}^{m} b^i E_i.$$

Suppose $\mathcal{L}$ is ∇^*-autoparallel, then, using Lemma 4.3, we have

$$AP^{-1}B + BP^{-1}A \in \mathcal{V}$$

for all $P \in \mathcal{L}$. By setting $P = I$, $\mathcal{V}$ proves to be Jordan subalgebra of $Sym(n)$.

Conversely, let $\mathcal{V}$ be Jordan subalgebra and assume $P \in \mathcal{L}$ and $A, B \in \mathcal{V}$. Since

$$A * (A * B) = \frac{1}{4}(A^2B + BA^2 + 2ABA) \in \mathcal{V}$$

and

$$(A * A) * B = \frac{1}{2}(A^2B + BA^2) \in \mathcal{V},$$

it follows $ABA \in \mathcal{V}$. Using this and $P^{-1} \in \mathcal{L}$ (See [14]), we conclude

$$(A + B)P^{-1}(A + B) - AP^{-1}A - BP^{-1}B = AP^{-1}B + BP^{-1}A \in \mathcal{V}$$

This implies $\mathcal{L}$ is ∇^*-autoparallel. Q.E.D.

Example 1: The set of symmetric matrices whose eigenvectors are fixed is one of examples of Jordan subalgebra of $Sym(n)$. Another example is the set of *doubly symmetric matrices*, which is symmetric with respect with both main- and anti-diagonals [14]. Hence, when $\mathcal{V}$ coincides to these set, SDP can be solved without any iteration for any vector c and initial point $x(0)$ using the algorithm discussed above.

Example 2: Consider the following SDP problem,

$$\min x^1, \text{ s.t., } P(x) := \begin{pmatrix} x^1 & x^2 \\ x^2 & x^2 + 1 \end{pmatrix} \geq 0.$$

If we take any initial point satisfying $x^2(0) = 0$ or $x^1(0) = x^2(0)$, the problem is readily solved with the above algorithm.

5 General Case

5.1 Predictor-Corrector type algorithm in Dual Coordinate

In this section, we consider a simple modification of the algorithm given in the section 4 to solve general SDP problems even in the case when $\mathcal{L}$ is not ∇^*-autoparallel. By examining behaviors of the modified algorithm, we show its computational complexity is directly related to a geometric quantity called *the second fundamental form* (or sometimes called *Euler-Schouten embedding curvature*).

We here denote, by $P^\sharp(t)$, an affine scaling trajectory or central path on $\mathcal{L}$, with a given initial interior point $P_0 \in \mathcal{L}$, and represent it by θ and η coordinates rather than x to discuss its behavior in $PD(n)$. Note that when $P^\sharp(t)$ is the central path, P_0 is the analytic center for $\mathcal{L}$. The relation between the coordinates θ and x has been given in (4.10).

Let us partition coordinate vector θ, η and Riemannian metric matrix $G = (g_{pq})$ compatibly:

$$\theta = \begin{pmatrix} \overline{\theta} \\ \underline{\theta} \end{pmatrix}, \quad \eta = \begin{pmatrix} \overline{\eta} \\ \underline{\eta} \end{pmatrix}, \quad G = \begin{pmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{pmatrix}, \tag{5.1}$$

where $\overline{\theta} = (\theta^i)$, $\underline{\theta} = (\theta^\kappa)$, $\overline{\eta} = (\eta_i)$, $\underline{\eta} = (\eta_\kappa)$, $G_1 = (g_{ij})$, $G_2 = (g_{i\kappa})$, $G_3 = (g_{\kappa\nu})$(5.2)

From (3.14) and the fact that $P^\sharp(t)$ stays in $\mathcal{L}$ for all $t \geq 0$, the θ coordinate of $P^\sharp(t)$ denoted by $\theta^\sharp(t)$ is the solution for

$$\dot{\theta} = \begin{pmatrix} \dot{\overline{\theta}} \\ \dot{\underline{\theta}} \end{pmatrix} = -\begin{pmatrix} G_1(\theta)^{-1}c \\ 0 \end{pmatrix}, \quad \theta(0) = \theta_0. \tag{5.3}$$

where $\theta_0 = (\theta_0^i, \theta_0^\kappa)$ is θ coordinate of an initial point P_0 for the affine trajectory on $\mathcal{L}$.

Since $(\partial\eta_p/\partial\theta^q)$, the Jacobian matrix of the Legendre transformation from η to θ, is equal to G as in (2.10), η coordinate of $P^\sharp(t)$ denoted by $\eta^\sharp(t)$ is the solution of

$$\dot{\eta} = G\dot{\theta} = -\begin{pmatrix} c \\ G_2^T(\eta)G_1^{-1}(\eta)c \end{pmatrix}, \qquad \eta(0) = \eta_0. \tag{5.4}$$

where $\eta_0 = (\eta_{i0}, \eta_{\kappa 0})$ is η coordinate of an initial point for P_0.

Now we show a rough sketch of a method to follow the path $P^\sharp(t)$. Assume we are given a certain point $P_a^\sharp := (\theta^\sharp(t_a)) = (\eta^\sharp(t_a))$ on $P^\sharp(t)$ corresponding to the parameter value $t = t_a$. Let $P^*(t)$ be the ∇^*-geodesic of $PD(n)$ which passes through $P_a^\sharp$ with the direction there equal to $\dot{P}^\sharp(t_a)$. If we define constant vector $\underline{c} = (c_\kappa) \in \mathbf{R}^{N-m}$ as

$$\underline{c} := -G_2(\eta^\sharp(t_a))G_1^{-1}(\eta^\sharp(t_a))c, \text{ i.e., } c_\kappa := \sum_{i,j=1}^{m} g_{\kappa j}(\eta^\sharp(t_a))g_1^{ij}(\eta^\sharp(t_a))c_i, \text{ where } (g_1^{ij}) = G_1^{-1}, \tag{5.5}$$

then η coordinate of $P^*(t)$ denoted by $\eta^*(t)$ is the solution of

$$\dot{\eta} = -\begin{pmatrix} c \\ \underline{c} \end{pmatrix}, \qquad \eta(t_a) = \eta^\sharp(t_a). \tag{5.6}$$

Thus, by viewing $P^*(t)$ and $P^\sharp(t)$ via η coordinate, we can regard $P^*(t)$ as a linearization of $P^\sharp(t)$ at the point $P_a^\sharp$. Hence, the point $P_b^\sharp := P^\sharp(t_b)$ on the affine scaling trajectory corresponding to the new parameter $t = t_b$ has first order approximation $P_b^* := P^*(t_b)$. This is the predicting step to follow the path $P^\sharp(t)$. However, when $\mathcal{L}$ is *not* ∇^*-autoparallel, we need a correcting step returning from P_b^* to $P_b^\sharp$ due to the approximation error. Iterating each step alternatively leads us to so-called *predictor-corrector* type algorithm. Note that the predictor P_b^* is outside of the feasible region $\mathcal{L}$.

The predicting step is easily executed because the $\eta^*(t_b)$ is obtained as

$$\eta^*(t_b) = \eta^\sharp(t_a) - \begin{pmatrix} c \\ \underline{c} \end{pmatrix} (t_b - t_a) \tag{5.7}$$

or in a matrix form

$$P^{*-1}(t_b) = P^{\sharp-1}(t_a) - C\delta t,$$

where C is a constant matrix, the η coordinate of which coincides with $(c^T \ \underline{c}^T)^T$.

Next, before we state about the correcting step, it is convenient to introduce the following submanifold:

Definition: Given $P \in PD(n)$, define $\mathcal{L}^\perp(P) \in PD(n)$ as follows:

$$\mathcal{L}^\perp(P) := \{P(\underline{\eta}) | P^{-1}(\underline{\eta}) = P^{-1} + \sum_{\kappa=m+1}^{N} \eta_\kappa E^\kappa, P(\underline{\eta}) \in PD(n)\}.$$

Note that $\underline{\eta}$ is a coordinate system of this submanifold. See [18] for various properties of $\mathcal{L}^\perp(P)$. For the correcting step, we have the following result:

Theorem 5.1: [18] The corrector $P_b^\sharp$ is the unique solution for the problem of minimizing the quantity so-called *divergence* $D(\bullet, \bullet) : PD(n) \times PD(n) \to \mathbf{R}$, i.e.,

$$P_b^\sharp = \arg \min_{P \in \mathcal{L}^\perp(P_b^*)} D(P_0^\sharp, P), \tag{5.8}$$

$$\text{where } D(P_1, P_2) := \log\det P_2 - \log\det P_1 + \mathrm{tr}(P_2^{-1}P_1) - n \tag{5.9}$$

By representing $P \in \mathcal{L}^\perp(P_b^*)$ by $\underline{\eta}$ and $\log\det P_0^\sharp$ is a constant, the above minimization problem is reduced to convex programming problem of

$$\eta^\sharp(t_b) = \arg\min_{\underline{\eta}} \varphi(\underline{\eta}), \quad \text{s.t. } P(\underline{\eta}) \in \mathcal{L}^\perp(P_b^*), \tag{5.10}$$

$$\text{where} \quad \varphi(\underline{\eta}) := -\log\det\left(P_b^{*-1} + \sum_{\kappa=m+1}^{N} \eta_\kappa E^\kappa\right) - \sum_{\kappa=m+1}^{N} \eta_\kappa \theta_0^{\sharp\kappa}. \tag{5.11}$$

Since the first term of the objective function φ is self-concordant and the second one is linear with respect to $\underline{\eta}$, the function φ is also self-concordant due to (3.8). Hence, we can solve efficiently this convex optimization problem by dumped Newton method [16].

Since $\mathcal{L}^\perp(P_b^*) = \mathcal{L}^\perp(P_b^\sharp)$ by its definition, we can say, to sum up, this modified algorithm is alternating predictors and correctors in the submanifold $\cup_{P \in P^\sharp(t), 0 \le t} \mathcal{L}^\perp(P)$. Note that the above predicting and correcting procedures can be completely executed with only η coordinate without calculating θ coordinate or matrix inversion.

5.2 Step Length

Finally, let us discuss a estimate of the possible long step we can take in each predicting direction, while keeping the number of iterations less than some prescribed upper bound in the succeeding corrector step. In our settings this is equivalent to discuss how large increment $\delta t := t_b - t_a$ we can take at each predicting step. Our analysis may be useful in design a long-step type algorithm.

We first show the η coordinate representation of the predicting error can be expressed by so-called the second fundamental form. Let us define new coordinates $\tilde{\eta}_p$ on $PD(n)$ such that their vector fields are defined as

$$\frac{\partial}{\partial \tilde{\eta}_i} := \sum_{j=1}^{m} g_1^{ij} \frac{\partial}{\partial \theta^j}, \tag{5.12}$$

$$\frac{\partial}{\partial \tilde{\eta}_\kappa} := \frac{\partial}{\partial \eta_\kappa}. \tag{5.13}$$

Note that at any point on $\mathcal{L}$, $\partial/\partial\tilde{\eta}_i$ and $\partial/\partial\tilde{\eta}_\kappa$ are tangential and orthogonal to $\mathcal{L}$, respectively, because each $\partial/\partial\tilde{\eta}_i$ is a linear combination of $\partial/\partial\theta^i$ and $g(\partial/\partial\theta^i, \partial/\partial\eta_\kappa) = \delta_i^\kappa$ due to (2.11). Since $(\partial\eta_p/\partial\theta^q) = G$ as is shown in (2.10), we have the following Jacobian matrices:

$$\left(\frac{\partial \eta_p}{\partial \tilde{\eta}_q}\right) = \begin{pmatrix} I & G_1^{-1}G_2 \\ 0 & I \end{pmatrix}, \quad \left(\frac{\partial \tilde{\eta}_q}{\partial \eta_p}\right) = \begin{pmatrix} I & -G_1^{-1}G_2 \\ 0 & I \end{pmatrix} \tag{5.14}$$

and

$$\left(\frac{\partial \theta^p}{\partial \tilde{\eta}_q}\right) = \begin{pmatrix} G_1^{-1} & 0 \\ -S^{-1}G_2^T G_1^{-1} & S^{-1} \end{pmatrix}, \quad \text{where } S := G_3 - G_2^T G_1^{-1} G_2.$$

Let us denote $\tilde{\partial}^p := \partial/\partial\tilde{\eta}_p$. Since $\tilde{\partial}^i$ and $\tilde{\partial}^\kappa$ are respectively tangential and orthogonal to $\mathcal{L}$ at every $P \in \mathcal{L}$, we can represent the components of *the second fundamental form* [13] of $\mathcal{L}$ for ∇^*-connection, denoted by $\widetilde{H}_\kappa^{*ij}$, as a Christoffel's symbol of ∇^*-connection with respect to $\tilde{\eta}$ coordinate system:

$$\widetilde{H}_\kappa^{*ij} = \tilde{\Gamma}_\kappa^{*ij} = \sum_{p=1}^{N} \tilde{g}_{\kappa p} \tilde{\Gamma}^{*ijp}, \quad i,j = 1,\ldots,m, \quad \kappa = m+1,\ldots,N. \tag{5.15}$$

Here $\tilde{g}_{pq}$ is the components of the inverse matrix of $(\tilde{g}^{pq})$ such that

$$\tilde{g}^{pq} = g(\tilde{\partial}^p, \tilde{\partial}^q) \tag{5.16}$$

and

$$\tilde{\Gamma}^{*pqr} = g(\nabla^*_{\tilde{\partial}^p} \tilde{\partial}^q, \tilde{\partial}^r). \tag{5.17}$$

The quantities $\widetilde{H}_\kappa^{*ij}$ provide information about how locally curved the shape of $\mathcal{L}$ is in $PD(n)$ in the sense of ∇^*-connection.

The following result shows η coordinate of the error $P_b^* - P_b^\sharp$ can be represented by $\widetilde{H}_\kappa^{*ij}$.

Lemma 5.2: For all $t_b = t_a + \delta t$ such that $\delta t > 0$, the η coordinate representation of the predicting error $P_b^* - P_b^\sharp$ has the following expression:

$$\begin{aligned} \eta_i^*(t_b) - \eta_i^\sharp(t_b) &= 0, \\ \eta_\kappa^*(t_b) - \eta_\kappa^\sharp(t_b) &= \frac{\delta t^2}{2} \sum_{i,j=1}^{m} \widetilde{H}_\kappa^{*ij}(t_0 + \xi\delta t) c_i c_j. \end{aligned} \tag{5.18}$$

Proof: Using (5.4) and (5.5), we have the Taylor series of $\eta^\sharp(t)$ for some $\xi \in (0,\ 1)$

$$\begin{cases} \eta_i^\sharp(t_a + \delta t) = \eta_i^\sharp(t_a) - c_i \delta t, & i = 1, \ldots, m \\ \eta_\kappa^\sharp(t_a + \delta t) = \eta_\kappa^\sharp(t_a) - c_\kappa \delta t + \dfrac{1}{2}\ddot{\eta}_\kappa^\sharp(t_0 + \xi\delta t)\delta t^2, & \kappa = m+1, \ldots, N, \end{cases} \tag{5.19}$$

Hence, from (5.6),

$$\begin{aligned} \eta_i^*(t_a + \delta t) - \eta_i^\sharp(t_a + \delta t) &= 0, \\ \eta_\kappa^*(t_a + \delta t) - \eta_\kappa^\sharp(t_a + \delta t) &= -\frac{1}{2}\ddot{\eta}_\kappa^\sharp(t_0 + \xi\delta t)\delta t^2. \end{aligned} \tag{5.20}$$

Since $P^\sharp(t)$ stays in $\mathcal{L}$ for all t, its derivative for the direction orthogonal to $\mathcal{L}$ is zero, i.e., $\dot{\tilde{\eta}}_\kappa = 0,\ \kappa = m+1, \ldots, N$ on $P^\sharp(t)$. Accordingly, $\ddot{\eta}^\sharp$ has the following expression:

$$\ddot{\eta}_\kappa^\sharp = \sum_{i=1}^{m} \frac{\partial \dot{\eta}_\kappa^\sharp}{\partial \tilde{\eta}_i} \dot{\tilde{\eta}}_i = -\sum_{i,j,l=1}^{m} \frac{\partial g_{\kappa l} g_1^{lj}}{\partial \tilde{\eta}_i} c_j \dot{\tilde{\eta}}_i. \tag{5.21}$$

By means of the formula (A.1), it follows that the relation between $\tilde{\Gamma}_\kappa^{*ij}$ and Γ_r^{*pq}, Christoffel's symbols with respect to $\tilde{\eta}$ and η coordinates, is

$$\widetilde{H}_\kappa^{*ij} = \tilde{\Gamma}_\kappa^{*ij} = \sum_{p,q,r=1}^{N} \frac{\partial \eta_p}{\partial \tilde{\eta}_i} \frac{\partial \eta_q}{\partial \tilde{\eta}_j} \frac{\partial \tilde{\eta}_\kappa}{\partial \eta_r} \Gamma_r^{*pq} + \sum_{q=1}^{N} \frac{\partial^2 \eta_q}{\partial \tilde{\eta}_i \partial \tilde{\eta}_j} \frac{\partial \tilde{\eta}_\kappa}{\partial \eta_q}.$$

Recall η is ∇^*-affine, then it follows $\Gamma_r^{*pq} = 0$. Using this fact and substituting (5.14) into the above equation, we obtain

$$\sum_{l=1}^{m} \frac{\partial g_{\kappa l} g_1^{lj}}{\partial \tilde{\eta}_i} = \widetilde{H}_\kappa^{*ij}, \quad \text{for} \quad i, j = 1, \ldots, m,\ \kappa = m+1, \ldots, N. \tag{5.22}$$

Since $\dot{\tilde{\eta}}_i = \dot{\eta}_i = -c_i,\ i = 1, \ldots, m$ on $P^\sharp(t)$, the statement holds. Q.E.D.

Remark: Due to (5.2) and the formula (A.1), $\widetilde{H}_\kappa^{*ij}$ can be calculated using the quantities with respect to θ coordinate.

$$\widetilde{H}_\kappa^{*ij} = \tilde{\Gamma}_\kappa^{*ij} = \sum_{l,k=1}^{m} \sum_{\mu=m+1}^{N} g_1^{il} g_1^{jk} s_{\kappa\mu} \Gamma_{lk}^{*\mu},$$

where

$$(s_{\kappa\mu}) = S, \quad \Gamma_{lk}^{*\mu} = \sum_{p=1}^{N} g^{\mu p} \Gamma_{lkp}^{*}.$$

From Lemma 5.2, when the second fundamental form $\widetilde{H}_\kappa^{*ij}$ is small in some sense, the predicting error represented in η coordinate proves to be small and, consequently, we can expect longer predictor step.

Next, we will discuss the possible long increment δt under the following assumptions:
A1) On $P^\sharp(t) \in \mathcal{L}$, the second fundamental form is small in the sense that

$$\max_{\kappa=m+1,\ldots,N}\{|\sum_{i,j=1}^{m} \widetilde{H}_\kappa^{*ij} c_i c_j|\} \leq M,$$

A2) The corrector returns to $P^\sharp(t) \in \mathcal{L}$ exactly.

The assumption A2) is not practical but just for the sake of simplicity. This is unnecessary if we replace $P^\sharp(t) \in \mathcal{L}$ by some suitable neighborhood along $P^\sharp(t) \in \mathcal{L}$ in A1).

The crucial point is keeping P_b^* in the neighborhood of $P_b^\sharp$ (preferably the quadratic convergence region) not to increase computational cost in the subsequent correcting step.

The suitable measures for this purpose are the error of the objective function

$$\epsilon(\underline{\eta}) := \varphi(\underline{\eta}) - \min_{\mathcal{L}^\perp(P_b^*)} \varphi(\underline{\eta}) = \varphi(\underline{\eta}) - \varphi(\underline{\eta^\sharp})$$

Since the dumped Newton method can decrease the value of self-concordant functions greater than some constant, ϵ can be a measure for number of iterations in corrector step.

Another measure is *Newton decrement* [16] for $\varphi(\underline{\eta})$:

$$\lambda(\underline{\eta}) := \left(\varphi'(\underline{\eta})^T[\varphi''(\underline{\eta})]^{-1}\varphi'(\underline{\eta})\right)^{1/2},$$

where $\varphi'(\underline{\eta})$ and $[\varphi''(\underline{\eta})]$ are the gradient vector and the Hessian matrix of $\varphi(\underline{\eta})$, respectively. Smaller value of $\lambda(\eta^*(t_b))$, which is evaluated at P_b^*, implies fewer (dumped) Newton steps to reach $P_b^\sharp$, minimizer of φ. Both measures are of course equivalent in the following sense [16]:

$$\lambda < 1/3 \Rightarrow \epsilon \leq \frac{\omega^2(\lambda)(1+\omega(\lambda))}{2(1-\omega(\lambda))}, \quad \text{where} \quad \omega(\lambda) := 1-(1-3\lambda)^{1/3},$$

$$\lambda \leq \rho^{-1}(\epsilon), \quad \text{where } \rho(\lambda) := \lambda - \ln(1+\lambda).$$

To evaluate these measures, define $e = (e_\kappa)$, where $e_\kappa(\delta t) := \eta_\kappa^*(t_a+\delta t) - \eta_\kappa^\sharp(t_a+\delta t)$ and introduce the following quantity:

$$r(\underline{\eta}) := (e^T S^{-1}(\underline{\eta}) e)^{1/2}.$$

Lemma 5.3: If $r(\underline{\eta^*}) < 1$, then

$$\epsilon \leq \frac{r^2(\underline{\eta^*})}{2(1-r(\underline{\eta^*}))^2}. \tag{5.23}$$

Further, if $r(\underline{\eta^*}) < 1/2$ then

$$\lambda \leq \frac{r(\underline{\eta^*})(1-r(\underline{\eta^*}))}{(1-2r(\underline{\eta^*}))^2}. \tag{5.24}$$

Proof: Since $\underline{\eta}^\sharp$ is the minimizer of $\varphi(\underline{\eta})$, we have the following Taylor series at $\underline{\eta}^\sharp$,

$$\varphi(\underline{\eta}^*) = \varphi(\underline{\eta}^\sharp) + \frac{1}{2}e^T\varphi''(\underline{\eta}^\sharp + \xi e)e, \quad \exists \xi \in (0,\ 1). \tag{5.25}$$

Combining Theorem 2.1.1 of [16] and $\varphi''(\underline{\eta}) = S^{-1}(\underline{\eta})$, we obtain

$$\epsilon = \varphi(\underline{\eta}^*) - \varphi(\underline{\eta}^\sharp) = \frac{1}{2}e^T\varphi''(\underline{\eta}^* + (\xi - 1)e)e = \frac{1}{2}r(\underline{\eta}^* + (\xi - 1)e) \leq \frac{r^2}{2(1-r)^2}, \tag{5.26}$$

For (5.24), we first recall the following result by [16, p. 24]:

$$\lambda \leq \varrho(r(\underline{\eta}^\sharp)) := \frac{r(\underline{\eta}^\sharp)}{(1 - r(\underline{\eta}^\sharp))^2}. \tag{5.27}$$

Since the function $\varrho(\bullet)$ is monotone increasing on the interval [0,1), the right hand side is not larger than $\varrho(r(\underline{\eta}^*)/(1 - r(\underline{\eta}^*)))$ due to Theorem 2.1.1 of [16]. Hence, we have (5.24). Q.E.D.

Thus, both measures can be bounded by $r(\underline{\eta}^*)$. The right-hand sides of (5.23) and (5.24) are monotone increasing with respect to r when $r < 1$ and $r < 1/2$, respectively. Hence, by setting a suitable constant $r^* \leq 1/2$ and solving δt such that $r(\underline{\eta}^*(t_a+\delta t)) \leq r^*$, we can obtain the long predictor step $\underline{\eta}^*(t_a+\delta t)$, where the number of iterations necessary in the succeeding corrector phase is bounded by $i(r^*)$, a certain number decided by r^*.

Since $S^{-1} = (s^{\kappa\mu})$ is a submatrix of G^{-1}, we have the following expression [18]:

$$s^{\kappa\mu}(\underline{\eta}^*(t_a + \delta t)) = \mathrm{tr}\{P^*(t_a + \delta t)E^\kappa P^*(t_a + \delta t)E^\mu\}.$$

Recall that $P^*(t_a + \delta t) = (P_a^{\sharp -1} + \delta t C)^{-1}$ and let T be nonsingular matrix such that

$$P_a^{\sharp -1} = T\Sigma_a T^T, \quad C = T\Sigma_c T^T,$$

where $\Sigma_a = \mathrm{diag}\{\alpha_1, \ldots, \alpha_n\} > 0$ and $\Sigma_c = \mathrm{diag}\{\gamma_1, \ldots, \gamma_n\}$. Then we have

$$s^{\kappa\mu}(\underline{\eta}^*(t_a + \delta t)) = \mathrm{tr}\{(\Sigma_a + \delta t\Sigma_c)^{-1}\tilde{E}^\kappa(\Sigma_a + \delta t\Sigma_c)^{-1}\tilde{E}^\mu\}, \tag{5.28}$$

where $\tilde{E}^\kappa := T^{-1}E^\kappa T^{-T}$. Since (5.28) is a quadratic form of $1/(\alpha_i+\delta t\gamma_i)$ and $P^*(t_a+\delta t) > 0$ implies $\alpha_i + \delta t\gamma_i > 0, i = 1, \ldots, n$, there exist n by n symmetric matrices $A_{\kappa\mu}, \kappa, \mu = m+1, \ldots, N$ such that

$$\begin{aligned} |s^{\kappa\mu}(\underline{\eta}^*(t_a + \delta t))| &= \left| \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix}^T A_{\kappa\mu} \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix} \right| \\ &\leq \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix}^T |A_{\kappa\mu}| \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix}, \end{aligned} \tag{5.29}$$

where $|A_{\kappa\mu}|$ is a matrix, each component of which is the absolute value of corresponding component of $A_{\kappa\mu}$. Note that $A_{\kappa\mu}$ is dependent only on $\tilde{E}^{\mu}$ and not on δt.

Let us define

$$\chi(\delta t) := \left\{ \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix}^T \left(\sum_{\kappa,\mu=m+1}^{N} |A_{\kappa\mu}| \right) \begin{pmatrix} \frac{1}{\alpha_1+\delta t\gamma_1} \\ \vdots \\ \frac{1}{\alpha_n+\delta t\gamma_n} \end{pmatrix} \right\}^{1/2}, \tag{5.30}$$

Theorem 5.4: Assume A1 and A2. When the increment δt in the predicting step satisfies

$$\delta t \leq \delta t^*,$$

the number of iterations necessary in the succeeding corrector phase is less than $i(r^*)$. Here δt^* is the largest solution for the inequality

$$\frac{\delta t^2}{2}\chi(\delta t)M \leq r^*.$$

Proof) Straightforward from the following inequalities:

$$r(\underline{\eta}^*) \leq \left(\sum_{\kappa,\mu=m+1}^{N} |e_\kappa||e_\mu||s^{\kappa\mu}| \right)^{1/2} \leq \frac{\delta t^2}{2}\chi(\delta t)M.$$

Q.E.D.

References

[1] S. Amari, *Differential-Geometrical Methods in Statistics* Springer-Verlag, (1985).

[2] D. A. Bayer and J. C. Lagarias, The Nonlinear Geometry of Linear Programming I, *Trans. Amer. Math. Soc.*, 314, 499-526 (1989), II, *Trans. Amer. Math. Soc.*, 314, 527-581 (1989).

[3] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM (1994).

[4] R. W. Brockett, Dynamical Systems That Sort Lists and Solve Linear Programming Problems, *Proc. 27th IEEE CDC*, 779-803 (1988)

[5] D. den Hertog, *Interior Point Approach to Linear, Quadratic and Convex Programming*, Kluwer Academic Publishers (1994)

[6] A. Fujiwara and S. Amari, Gradient Systems in view of Information Geometry, *Physica*, D80, 317-327 (1995)

[7] L. Faybusovich, A Hamilton Structure for Generalized Affine-scaling Vector Fields, *J. Nonlinear Science*, 5, 11-28 (1995)

[8] L. Faybusovich, Jordan Algebras, Symmetric Cones and Interior-point Methods, preprint, (1995)

[9] J. Faraut & A. Koranyi, *Analysis on Symmetric Cones*, Oxford (1994)

[10] M. Iri, Integrability of Vector and Multivector Fields Associated with Interior Point Methods for Linear Programming, *Mathematical Programming*, 511-525 (1991)

[11] N. Karmarkar, A New Polynomial-Time Algorithm for Linear Programming, *Combinatorica*, 4, 373-395 (1984)

[12] N. Karmarkar, Riemannian Geometry Underlying Interior-Point Methods for Linear Programming, *Contemporary Mathematics*, 114, 51-75 (1990)

[13] S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry I*, John Wiley and Sons (1963).

[14] S. D. Morgera, The Role of Abstract Algebra in Structured Estimation Theory, *IEEE Trans. on Inf. Theory*, 38-3, 1053-1065 (1992)

[15] Y. Nakamura, Lax Pair and Fixed Point Analysis of Karmarkar's Scaling Trajectory for Linear Programming, *Japan J. Indust. Appl. Math.*, 11, 1-9 (1994)

[16] Yu. Nesterov and A. Nemirovsky, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM (1994).

[17] F. Jarre, Interior-Point Methods for Convex Programming, *Applied Mathematics and Optimization*, Vol.26, 287-311 (1992)

[18] A. Ohara, N. Suda and S. Amari, Dualistic Differential Geometry of Positive Definite Matrices and Its Applications to Related Problems, *Linear Algebra and Its Applications*, 247, 31-53 (1996)

[19] K. Tanabe, Centered Newton Method for Mathematical Programming, *System Modeling and Optimization (eds. M. Iri, et al.)* Springer-Verlag (1988)

[20] K. Tanabe and T. Tsuchiya, New Geometry for Linear Programming, Mathematical Science, 303, 32-37 (1988) *in Japanese*

[21] T. Terlaky *eds.*, *Interior Point Methods of Mathematical Programming*, Kluwer Academic Publishers (1996)

[22] S. J. Wright, *Primal-Dual Interior Point Methods*, SIAM, (1997)

A Appendix

Let x and $\tilde{x}$ be local coordinate systems on a manifold equipped with a linear connection. When the coordinate system is changed from x to $\tilde{x}$, associated Christoffel's symbol is subject to the following transformation rule:

$$\tilde{\Gamma}^c_{ab} = \sum_{i,j,k=1}^{n} \frac{\partial x^i}{\partial \tilde{x}^a} \frac{\partial x^j}{\partial \tilde{x}^b} \frac{\partial \tilde{x}^c}{\partial x^k} \Gamma^k_{ij} + \sum_{j=1}^{n} \frac{\partial^2 x^j}{\partial \tilde{x}^a \partial \tilde{x}^b} \frac{\partial \tilde{x}^c}{\partial x^j}. \tag{A.1}$$