Convex analysis on Hadamard spaces and scaling problems

Hiroshi HIRAI
Graduate School of Mathematics,
Nagoya University, Nagoya, 464-8602, Japan.
hirai.hiroshi@math.nagoya-u.ac.jp

June 1, 2023

Abstract

In this paper, we address the bounded/unbounded determination of geodesically convex optimization on Hadamard spaces. In Euclidean convex optimization, the recession function is a basic tool to study the unboundedness, and provides the domain of the Legendre-Fenchel conjugate of the objective function. In a Hadamard space, the asymptotic slope function (Kapovich, Leeb, and Millson 2009), which is a function on the boundary at infinity, plays a role of the recession function. We extend this notion by means of convex analysis and optimization, and develop a convex analysis foundation for the unbounded determination of geodesically convex optimization on Hadamard spaces, particularly on symmetric spaces of nonpositive curvature. We explain how our developed theory is applied to operator scaling and related optimization on group orbits, which are our motivation.

Keywords: Convex analysis, Hadamard space, CAT(0) space, recession function, Legendre-Fenchel conjugate, symmetric space, Euclidean building, matrix and operator scaling, null-cone membership, moment polytope, submodular function, Busemann function

1 Introduction

Hadamard spaces are complete geodesic metric spaces having nonpositive curvature. In such a space, a geodesic connecting any two points is uniquely determined. A function on a Hadamard space is called \textit{(geodesically) convex} if it is convex along any geodesics. The theory of convex optimization on Hadamard spaces is a promising direction of research, though it has just started and is still undeveloped; see e.g., [5]. Since the influential paper [22] by Garg, Gurvits, Oliveira, and Wigderson on \textit{operator scaling} [25], apparently unrelated problems in diverse fields of mathematical sciences have been formulated and partially/completely solved via geodesically convex optimization on Riemannian manifolds; see [2, 11, 15, 19, 23, 28] and references therein. These manifolds are, in fact, Hadamard manifolds (Riemannian manifolds that are Hadamard spaces), more specifically, \textit{symmetric spaces of nonpositive curvature}. 

In these problems, as well as finding near-optimal solutions, deciding boundedness of the optimization problem,

\[ \inf_{x \in X} f(x) > -\infty \text{ or } = -\infty \quad (1.1) \]

becomes an important issue. Examples are the approximate scalability in operator scaling and its invariant theoretic generalizations (null-cone membership, moment polytope membership); see the above references.

The present paper addresses this bounded/unbounded determination by means of convex analysis. For explaining our approach, let us recall the Euclidean situation. In Euclidean convex optimization, recession functions (also called asymptotic functions) are a basic tool to study the boundedness property; see [29] Section 3.2 and [40] Section 8. For a convex function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), the recession function \( f^\infty : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\} \) is defined by

\[ f^\infty(u) := \lim_{t \to \infty} f(x + tu)/t \quad (u \in \mathbb{R}^n), \quad (1.2) \]

where \( f^\infty(u) \) is independent of \( x \in \mathbb{R}^n \). The recession function \( f^\infty \) is a positively homogeneous convex function, and links with Legendre-Fenchel duality as follows. Recall the Legendre-Fenchel conjugate \( f^* : \mathbb{R}^n \rightarrow \mathbb{R}^n \cup \{\infty\} \) of \( f \), which is defined by \( f^*(p) := \sup_{x \in \mathbb{R}^n} \langle p, x \rangle - f(x) \). Its domain \( \text{dom } f^* := \{p \in \mathbb{R}^n \mid f^*(p) < \infty\} \) is precisely the set of vectors \( p \in \mathbb{R}^n \) for which \( \inf_{x \in \mathbb{R}^n} f(x) - \langle p, x \rangle \) is bounded below. Then, the recession function \( f^\infty \) equals the support function of the domain \( \text{dom } f^* \). Namely, \( f^\infty(u) = \sup_{p \in \text{dom } f^*} \langle u, p \rangle \), and the closure \( \overline{\text{dom } f^*} \) of \( \text{dom } f^* \) equals

\[ B(f^\infty) := \{p \in \mathbb{R}^n \mid \langle u, p \rangle \leq f^\infty(u) \quad (u \in \mathbb{R}^n)\}. \quad (1.3) \]

See [29] Example 2.4.6 and [40] Theorem 13.3. In particular, \( f^\infty \) provides an inequality description of \( \overline{\text{dom } f^*} \).

Suppose further that \( f \) is smooth. The image \( \nabla f(\mathbb{R}^n) \) of gradient map \( x \mapsto \nabla f(x) \) is precisely the set of vectors \( p \) for which the infimum of \( \inf_{x \in \mathbb{R}^n} f(x) - \langle p, x \rangle \) is attained. Then, the gradient space \( \nabla f(\mathbb{R}^n) \) contains the relative interior \( \text{ri } B(f^\infty) \) of \( B(f^\infty) = \overline{\text{dom } f^*} \) [40] Corollary 26.4.1]. Thus, the following relation holds:

\[ \text{ri } B(f^\infty) \subseteq \nabla f(\mathbb{R}^n) \subseteq \text{dom } f^* \subseteq B(f^\infty). \quad (1.4) \]

Particularly, \( \nabla f(\mathbb{R}^n) = \overline{\text{dom } f^*} = B(f^\infty) \) holds. If \( \text{dom } f^* \) is closed, then the following conditions are equivalent for \( p \in \mathbb{R}^n \): \( c \) \( (a) \inf_{x \in \mathbb{R}^n} \|\nabla f(x) - p\| = 0. \]

\( (b) \ -f^*(p) = \inf_{x \in \mathbb{R}^n} f(x) - \langle p, x \rangle > -\infty. \]

\( (c) \ p \in B(f^\infty). \]

This equivalence plays fundamental roles in matrix scaling [44]—the origin of operator scaling and related group orbits optimization. Given an \( n \times n \) nonnegative matrix \( A = (a_{ij}) \) and positive vectors \( r, c \in \mathbb{R}^n \) with the same sum \( \sum_i r_i = \sum_i c_i = l \), the matrix scaling problem is to ask positive diagonal matrices \( R, C \) such that \( ((RAC)^{\top} 1, RAC 1)^T = (r, c) \), where \( 1 \) denotes the all-ones vector. The matrix \( A \) is said to be scalable if there exist such \( R, C \), and approximately scalable if there exist \( R, C \)
such that \((RAC)^{\top}1, RAC1\) is arbitrary close to \((r, c)\). In fact, the approximate scalability is written as the condition (a) for convex function \(f_B(s, t) := \log \sum_{i,j} e^{s_i a_{ij}} e^{t_j}\) and vector \(p = (r, c)\). The recession function \(f_A^\infty\) is given by \(f_A^\infty(u, v) = \ell \max \{ u_i + v_j \mid i, j : a_{ij} \neq 0 \}\). The condition (c) is written as a network flow LP and efficiently verified. The combinatorial characterization of the approximate scalability by Rothblum and Schneider [41] was proved via \(f^\infty\) in this way. From (b) and the fact that \(\text{dom} f_A^\infty\) is closed, approximate scaling \(RAC\) is obtained by solving convex optimization \(\inf_{s,t} f_A(s, t) - \langle r, s \rangle - \langle c, t \rangle\). Sinkhorn algorithm [41] is viewed as an alternating minimization algorithm for this problem.

The goal of this paper is to establish analogues of the equivalence of (a), (b), and (c) for geodesically convex optimization on Hadamard spaces, and to provide a convex analysis foundation to the above mentioned problems. We particularly focus on an analogous notion of the recession function on a Hadamard space. In fact, such a notion was already introduced by Kapovich, Leeb, and Millson [31] (see also [46, Section 5.4]), who defined the asymptotic slope of a convex function via a geodesic analogue of \((1.2)\). By utilizing the asymptotic slope, they studied the boundedness (semistability) for a class of convex functions related to a generalization of Horn’s problem. Our presented theory reformulates and extends some of their arguments from viewpoints of convex analysis and optimization.

The structure and results of this paper are outlined as follows. In Section 2, we provide necessary backgrounds on a Hadamard space \(X\), particularly, the boundary \(X^\infty\) at infinity and its Euclidean cone \(CX^\infty\). The cone \(CX^\infty\) of the boundary is also a Hadamard space, and plays a role of the “dual space” of \(X\). With a point \(p \in CX^\infty\), we associate the Busemann function \(b_p\) as a correspondent of a linear function, and regard the boundary \(CX^\infty\) as the space of Busemann functions. This viewpoint leads to the notion of the asymptotic Legendre-Fenchel conjugate \(f^*\) of \(f\), which is a function on \(CX^\infty\) defined by \(p \mapsto f^*(p) := \sup_{x \in X} - b_p(x) - f(x)\). The asymptotic slope is a function on \(X^\infty\). We extend it, in a homogeneous way, to introduce the recession function \(f^\infty\) on \(CX^\infty\). We also define the associated subset \(B(f^\infty)\) via an analogue of \((1.3)\). We show that \(f^\infty\) is positively homogeneous convex on \(CX^\infty\) and that \(p \in B(f^\infty)\) is a necessary condition for \(f^*(p) < \infty\), i.e., \(\text{dom} f^* \subseteq B(f^\infty)\).

We then move to smooth convex optimization on a Hadamard manifold \(M\). The boundary cone \(CM^\infty\) is identified with the tangent space \(T_x\) at any point \(x\), though \(CM^\infty\) has a different topology from the usual one \(T_x \simeq \mathbb{R}^n\), and is not a manifold in general. We define the asymptotic gradient \(\nabla^\infty f(x)\) as the image of the gradient \(\nabla f(x)\) in \(CM^\infty\). This notion fits into our purpose: We verify a weaker analogue of \((1.4)\) in which \(\nabla\) is replaced by \(\nabla^\infty\) and \(ri\) is placed by the interior. We also verify an analogue of the equivalence of (a) and (c).

We further restrict our study to symmetric spaces of nonpositive curvature—a representative class of Hadamard manifolds having rich group symmetry. The boundary \(CM^\infty\) has a polyhedral cone complex structure, known as a Euclidean building. We utilize the symmetry property and building structure to show that the conjugate \(f^*\) is convex on each cone (Weyl chamber). We also provide some fundamental results on \(\text{dom} f^*\) and \(B(f^\infty)\). Then we present detailed calculations and specializations of several concepts for the symmetric space \(P_n = GL(n, \mathbb{C})/U(n)\) of positive definite matrices. The boundary \(CP^\infty_n\) is viewed as the order complex of flags of vector subspaces, which naturally links submodular functions on the lattice of vector subspaces and their
Lovász extension. Via the Lovász extension, submodular functions correspond to convex functions on $CP_n^\infty$ that are linear on each cone. For a convex function $f$ on $M$ inducing a submodular function at infinity, called an asymptotically submodular function, $B(f^\infty)$ is an analogue of the base polyhedron and the membership of $B(f^\infty)$ becomes discrete convex optimization (i.e., submodular function minimization) over the modular lattice of all vector subspaces of $\mathbb{C}^n$.

In Section 3, we explain how the developed theory is applied to operator scaling and related optimization on group orbits. These problems are viewed as convex optimization on symmetric spaces of nonpositive curvature. We take up the operator scaling with marginals by Franks [19]. We compute the recession function, from which Franks’ characterization of the approximate triangular scalability is naturally deduced.

We finally consider the null-cone and moment polytope membership for a linear action of a reductive group $G$, formulated by Bürgisser, Franks, Garg, Oliveira, Walter, and Wigderson [15]. These are convex optimization of the Kempf-Ness function $f_v$ over symmetric space $M = G/K$. In this setting, asymptotic gradients give rise to the moment map and moment polytope, for which $f_v^\infty$ gives an inequality description. We establish the link with Hilbert-Mumford criterion and Kempf-Ness theorem, and verify from our approach shifting trick and convexity theorem for the moment polytope.

The main implication of this paper is that finding a one-parameter subgroup in the Hilbert-Mumford criterion and membership of the moment polytope (after randomization) reduce to minimization of recession functions $f_v^\infty$ over Euclidean building $CM^\infty$. This is a far-reaching generalization of the approach by Hamada and Hirai [26] for the noncommutative-rank computation (= the null-cone problem of the left-right action), in which their algorithm is now viewed as minimizing $f_v^\infty$. The current approach (e.g., [15, 28]) for these problems is mainly based on smooth convex optimization (b) on $G/K$. We hope that our results will motivate to develop nonsmooth convex optimization techniques on non-manifold Hadamard spaces, for verifying (c).

2 Convex analysis on Hadamard spaces

2.1 Hadamard spaces

Here we introduce Hadamard spaces; see [5, 6, 11] for details. Let $X$ be a metric space with distance function $d$. A path in $X$ is a continuous map from interval $[0, l] \subseteq \mathbb{R}$ to $X$, where $l \geq 0$. We say that a path $c : [0, l] \to X$ connects $c(0)$ and $c(l)$. A path $c : [0, l] \to X$ is said to be geodesic if $d(c(s), c(t)) = |s - t|$ for every $s, t \in [0, l]$. A geodesic metric space is a metric space $X$ in which any pair of two points is connected by a geodesic path.

Suppose that $X$ is a geodesic metric space. Let $x, y, z \in X$. A geodesic triangle of $x, y, z$ is the union of geodesic paths connecting $x, y$, $x, z$, and $y, z$. The comparison triangle is the triangle in Euclidean plane $\mathbb{R}^2$ with vertices $\bar{x}, \bar{y}, \bar{z}$ such that $d(x, y) = \|\bar{x} - \bar{y}\|_2$, $d(y, z) = \|\bar{y} - \bar{z}\|_2$ and $d(z, x) = \|\bar{z} - \bar{x}\|_2$. In the geodesic triangle, suppose that points $x, y$ and $x, z$ are connected by geodesic paths $c$ and $c'$, respectively, with $c(0) = c'(0) = x$. For $t, t' \in [0, 1]$, let $p := c(td(x, y))$ and $q := c'(td(x, z))$. Let $p := (1 - t)\bar{x} + t\bar{y}$ and $q := (1 - t')\bar{x} + t'\bar{z}$ be the corresponding points in $\mathbb{R}^2$. Then the $CAT(\theta)$-inequality is given by

$$d(p, q) \leq \|p - q\|_2.$$  (2.1)
A geodesic metric space $X$ is said to be $\text{CAT}(0)$ if the $\text{CAT}(0)$ inequality \((2.1)\) holds for every choice of a geodesic triangle and $t \in [0, 1]$. There are several ways of defining $\text{CAT}(0)$ spaces. A useful one is the following: $X$ is $\text{CAT}(0)$ if for every triple of points $x, y, z \in X$, geodesic path $c$ with $c(0) = x$ and $c(d(x, y)) = y$, and $t \in [0, 1]$, it holds
\[ d(pt, z)^2 \leq (1 - t)d(x, z)^2 + td(y, z)^2 - t(1 - t)d(x, y)^2, \]  \((2.2)\)
where $pt = c(td(x, y))$. Notice that the RHS equals the squared comparison distance between $\overline{xy} = (1 - t)\overline{xz} + t\overline{yz}$ and $\overline{xy}$.

It is known [11, II.1.4] that $\text{CAT}(0)$ spaces are uniquely geodesic, i.e., a geodesic path connecting any two points is unique. In this case, for points $x, y \in X$, let $[x, y]$ denote the image of the unique geodesic path connecting $x, y$. For $t \in [0, 1]$, let $(1 - t)x + ty$ denote the point $z \in [x, y]$ with $d(x, z)/d(x, y) = t$.

A Hadamard space is a $\text{CAT}(0)$ space that is complete as a metric space. Let $X$ be a Hadamard space. A subset $S \subset X$ is called convex if $x, y \in X$ implies $[x, y] \subset X$. The smallest convex set including a given $S \subset X$ is called the convex hull of $S$. A function $f : X \to \mathbb{R} \cup \{\infty\}$ is said to be convex if for all $x, y \in X, t \in [0, 1]$ it satisfies
\[ (1 - t)f(x) + tf(y) \geq f((1 - t)x + ty). \] \((2.3)\)

$f$ is called strictly convex if $< \text{ holds in } (2.3)$ for any $t \in (0, 1)$. If $-f$ is convex, then $f$ is said to be concave. An affine function is a convex and concave function. A function $f : X \to \mathbb{R}$ is said to be $L$-Lipschitz with parameter $L \geq 0$ if it satisfies $|f(x) - f(y)| \leq Ld(x, y)$ for all $x, y \in X$.

### 2.1.1 Boundary at infinity and its Euclidean cone

To study the asymptotic behavior of convex functions, we consider the boundary of a Hadamard space $X$; see [3] Chapter II and [11] Chapter II.8 for the boundary.

A geodesic ray is a continuous map $c : [0, \infty) \to X$ such that $d(c(s), c(t)) = |s - t|$ for $s, t \in [0, \infty)$. With a geodesic ray $c$ and $r \geq 0$, the map written as $[0, \infty) \ni t \mapsto c(rt)$ is called a constant-speed ray, or simply, a ray. If $c(0) = x$, we say that ray $c$ issues from $x$. Two geodesic rays $c, c'$ are said to be asymptotic if there is a positive constant $K > 0$ such that $d(c(t), c'(t)) \leq K$ for all $t \geq 0$. The asymptotic relation is an equivalence relation on the set of all geodesic rays. Let $X^\infty$ denote the set of all equivalence classes, which is called the boundary of $X$ at infinity. The equivalence class of $c$ is called the asymptotic class of $c$, and is denoted by $c(\infty)$. For any point $x \in X$, and each $\xi \in X^\infty$ there is a unique geodesic ray $c$ issuing from $x$ such that $c(\infty) = \xi$. Therefore, $X^\infty$ can be identified with the set of all geodesic rays issuing from any fixed $x$.

We next introduce a metric on $X^\infty$. For points $x, y, z \in X$ (with $y \neq x \neq z$), the comparison angle $\overline{xyz}$ between $y$ and $z$ at $x$ is the inner angle of the comparison triangle of $x, y, z$ at $x$. For two geodesic rays $c, c'$ issuing from $x$, let the angle $\angle_x(c, c')$ of $c, c'$ at $x$ be defined by
\[ \angle_x(c, c') := \lim_{t \to 0} \overline{Z}_x(c(t), c'(t)), \] \((2.4)\)
where the limit indeed exists [11, II.3.1]. For $x, x' \in X^\infty$, the angle $\angle(x, x') \in [0, \pi]$ is defined by
\[ \angle(x, x') := \sup_{x \in X} \angle_x(c, c') = \lim_{t, t' \to \infty} \overline{Z}_x(c(t), c'(t')) = \sup_{t, t' > 0} \overline{Z}_x(c(t), c'(t')), \] \((2.5)\)
where \(c, c'\) are geodesic rays issuing from \(x\) with \(c(\infty) = \xi, c'(\infty) = \xi'\). The equalities in (2.5) follow from [11 II. 9.8 (1)] (where \(x\) is arbitrary). Here \((\xi, \xi') \mapsto \angle(\xi, \xi')\) defines a metric on \(X^\infty\), which is called the angular metric. Then \(X^\infty\) becomes a metric space, and a topological space accordingly.

In addition to \(X^\infty\), we consider its Euclidean cone \(CX^\infty\); see [11 Chapter I.5] for generalities of the Euclidean cone construction. Let \(\mathbb{R}_+\) denote the set of nonnegative numbers. The set \(CX^\infty\) is the quotient of \(\mathbb{R}_+ \times X^\infty\) by the equivalence relation \((r, \xi) \simeq (r', \xi')\) if \(r = r' = 0\) or \((r, \xi) = (r', \xi')\). The equivalence class of \((r, \xi)\) is denoted by \(r \xi\). By identifying \(\xi\) with 1, we regard \(X^\infty\) as a subset of \(CX^\infty\). Let 0 denote the class of \((0, \xi)\). The angle \(\angle(p, p')\) of (nonzero) points \(p = r \xi, p' = r' \xi' \in CX^\infty\) is defined as \(\angle(\xi, \xi')\).

The space \(CX^\infty\) is viewed as the space of all constant-speed rays issuing from any fixed point. Indeed, \(r \xi \in CX^\infty\) is associated with a constant-speed ray \(t \mapsto c^r(t) = c(rt)\), where \(c\) is a geodesic ray with \(c(\infty) = \xi\). In this case, we let \(c^r(\infty) := r \xi\). The space \(CX^\infty\) is metrized by the following distance \(d^\infty:\)

\[
d^\infty(r \xi, r' \xi')^2 := r^2 + r'^2 - 2rr' \cos \angle(\xi, \xi') \quad (r \xi, r' \xi' \in CX^\infty).
\]

The topology of \(CX^\infty\) is given accordingly, which is called the \(d^\infty\)-topology.

**Lemma 2.1** (see [6 II.4.8]). \(CX^\infty\) is a Hadamard space.

A flat triangle in a Hadamard space is a geodesic triangle whose convex hull is isometric to the convex hull of their comparison triangle in \(\mathbb{R}^2\). It is known (see [11 II.2.9]) that if one of the vertex angles of the triangle is equal to the corresponding angle of the comparison triangle, then it is flat. By construction of \(CX^\infty\), the angle at 0 is always equal to the corresponding angle of the comparison triangle. Hence we have:

**Lemma 2.2.** In \(CX^\infty\), any three points containing 0 form a flat triangle.

Therefore, the convex hull of \(\{0, tp, t'p'\}_{t, t' \in \mathbb{R}_+}\) is isometric to the convex cone \(C\) in \(\mathbb{R}^2\). Then, via the isometry to \(C \subseteq \mathbb{R}^2\), we can consider nonnegative combinations in two nonzero points in \(CX^\infty\). For a point \(p = t \xi \in CX^\infty\) and \(a \in \mathbb{R}_+\), define \(a p := (at) \xi \in CX^\infty\). It is clear that \(p \mapsto a p\) is continuous and \(a(bp) = (ab)p\). For two points \(p, q \in CX^\infty\), the sum \(p + q\) of \(p, q\) is defined by

\[
p + q := 2((1/2)p + (1/2)q).
\]

Recall that \((1/2)p + (1/2)q\) is the midpoint of the geodesic path between \(p\) and \(q\). Map \((p, q) \mapsto p + q\) is continuous; this follows from the fact [11 II.1.4] that the geodesic segment in a Hadamard space varies continuously with its endpoints. Notice that + is not associative in general; so there may be many “inverses” \(q\) of \(p + q = 0\). From (2.6), it holds \(d^\infty(\alpha p, \alpha q) = \alpha d^\infty(p, q)\). This means that by \(p \mapsto \alpha p\), geodesic segment \([p, q]\) is mapped to geodesic segment \([\alpha p, \alpha q]\). Then we have a linearity relation

\[
\alpha(p + q) = \alpha p + \alpha q \quad (\alpha \in \mathbb{R}_+).
\]

Define the inner product \(\langle p, q \rangle\) of \(p, q \in CX^\infty\) by

\[
2\langle p, q \rangle := d^\infty(0, p)^2 + d^\infty(0, q)^2 - d^\infty(p, q)^2.
\]
For $p = t\xi \in CX^\infty$, define the norm $\|p\|$ by $\|p\| := t$. Observe from definitions (2.6) (2.8) that $\|p\| = d^\infty(0, p) = \sqrt{\langle p, p \rangle}$. Then we have

$$\langle p, q \rangle = \|p\|\|q\| \cos \angle(p, q).$$

(2.9)

A function $f : CX^\infty \to \mathbb{R} \cup \{\infty\}$ is called positively homogeneous if it holds

$$f(\alpha p) = \alpha f(p) \quad (a \in \mathbb{R}_+).$$

Lemma 2.3. For any $q \in CX^\infty$, the function $p \mapsto \langle p, q \rangle$ is continuous and positively homogeneous concave.

Proof. The continuity follows from (2.8) that $\langle , \rangle$ is written by continuous function $d^\infty$. The positive homogeneity follows from (2.9). For concavity, it suffices to show

$$\langle p + p', q \rangle \geq \langle p, q \rangle + \langle p', q \rangle.$$  (2.10)

Twice the RHS is equal to

$$-d^\infty(p, q)^2 - d^\infty(p', q)^2 + \|p\|^2 + \|p'\|^2 + 2\|q\|^2.$$  (2.11)

On the other hand, twice the LHS is equal to

$$4((p + p')/2, q) = -2d^\infty((p + p')/2, q)^2 + 2\|(p + p')/2\|^2 + 2\|q\|^2.$$  (2.12)

Since $0, p, p'$ form a flat triangle (Lemma 2.2), the CAT(0) inequality (2.2) with $(x, y, z) = (p, p', 0)$ and $t = 1/2$ holds in equality:

$$2\|(p + p')/2\|^2 = \|p\|^2 + \|p'\|^2 - d^\infty(p, p')^2/2.$$  (2.13)

From (2.11), (2.12), and (2.13), we see that (2.10) is equivalent to

$$2d^\infty((p + p')/2, q)^2 \leq d^\infty(p, q)^2 + d^\infty(p', q)^2 - d^\infty(p, p')^2/2.$$  (2.14)

This is the CAT(0)-inequality (2.2) with $(x, y, z) = (p, p', q)$ and $t = 1/2$, which holds by Lemma 2.1.

Example 2.4. Consider the case of $X = \mathbb{R}^n$. Two geodesic rays $t \to \xi t + b$ and $t \to \xi' t + b'$ are asymptotic if and only if $\xi = \xi'$. The angle of the corresponding asymptotic classes is given by $\cos^{-1}\langle \xi, \xi' \rangle$. Therefore, $X^\infty$ is isometric to the sphere $S^{n-1}$. The Euclidean cone $CX^\infty$ is isometric to the Euclidean space $\mathbb{R}^n = \mathbb{R}_+ \times S^{n-1} \simeq$. Also the distance $d^\infty$ and product $\langle , \rangle$ coincides with the Euclidean ones.

2.1.2 Busemann functions and asymptotic Legendre-Fenchel conjugate

For a geodesic ray $c$, the Busemann function $b_c : X \to \mathbb{R}$ is defined by

$$b_c(x) := \lim_{t \to \infty} d(x, c(t)) - t \quad (x \in X).$$  (2.15)

Lemma 2.5 (see [11, II.8.22]). Busemann functions for geodesic rays are 1-Lipschitz convex functions.
Two asymptotic geodesic rays \(c, c'\) yields the same Busemann functions up to additive constant:

**Lemma 2.6** (see [11 II.8.20]). \(c\) and \(c'\) are asymptotic if and only if \(b_c - b_{c'}\) is a constant function.

The Busemann function \(b_c\) is extended for a constant-speed ray \(c : t \mapsto c'(rt)\) by \(b_c := rb_{c'}\), where \(c'\) is a geodesic ray. Fix \(x_0 \in X\). For \(p \in CX^\infty\), there is a unique ray \(c\) issuing \(x_0\) with \(c(\infty) = p\), and hence \(b_c\) is also written as \(b_{x_0, p} = b_p\). In this way, we can identify \(CX^\infty\) with the space of Busemann functions.

**Example 2.7.** In the case of \(X = \mathbb{R}^n\), the Busemann function \(b_c\) for a geodesic ray \(c(t) = \xi t + x_0\) is given by
\[
b_c(x) = -\langle \xi, x - x_0 \rangle.
\]
This follows from \(\|x - \xi t - x_0\|_2 - t = t - \langle \xi, x - x_0 \rangle + O(1/t) - t\). Thus, Busemann functions (for constant-speed rays) are precisely affine functions. If the origin 0 is chosen as \(x_0\), the space \(CX^\infty(\simeq \mathbb{R}^n)\) is identified with the space of linear functions \(x \mapsto -\langle p, x \rangle\), i.e., the dual space of \(\mathbb{R}^n\).

For a (convex) function \(f : X \to \mathbb{R}\), define the *asymptotic Legendre-Fenchel conjugate* \(f^* : CX^\infty \to \mathbb{R} \cup \{\infty\}\) of \(f\) by
\[
f^*(p) := \sup_{x \in X} -b_p(x) - f(x) \quad (p \in CX^\infty).
\]
(2.16)

Note that this notion of Legendre-Fenchel conjugate is rather different from those introduced by [11 36]. If \(X = \mathbb{R}^n\), then this matches the usual Legendre-Fenchel conjugate (see Examples 2.4 and 2.7). Notice that \(f^*\) is not necessarily convex in \(CX^\infty\). We give a simple lemma providing explicit examples of asymptotic Legendre-Fenchel conjugates.

**Lemma 2.8.** For a function \(h : \mathbb{R}_+ \to \mathbb{R}\), define \(f : X \to \mathbb{R}\) by
\[
f(x) := h(d(x, x_0)) \quad (x \in X).
\]
Then the asymptotic Legendre-Fenchel conjugate \(f^*\) is given by
\[
f^*(p) := h^*(\|p\|) \quad (p \in CX^\infty),
\]
where \(h^*\) is the Legendre-Fenchel conjugate of \(h\), i.e., \(h^*(\xi) := \sup_{r \in \mathbb{R}} \xi r - h(r)\) with \(h(r) := \infty\) for \(r < 0\).

**Proof.** By definitions of \(f^*\) and \(b_p\), we have
\[
f^*(p) = \sup_{x \in X} -\|p\| \lim_{t \to \infty} (d(x, c(t)) - t) - h(d(x, x_0)),
\]
where \(c\) is a geodesic ray with \(c(0) = x_0\) and \(c(\infty) = p/\|p\|\). For \(x \in X\), letting \(r := d(x, x_0)\), we have \(h(d(x, x_0)) = h(r)\), and \(d(x, c(t)) \geq d(c(t), x_0) - d(x, x_0) = t - r = d(c(r), c(t))\) for every large \(t\). This implies that \(\sup\) can be restricted to \(c\). Thus \(f^*(p) = \sup_{r \in \mathbb{R}_+} \|p\| r - h(r) = h^*(\|p\|)\). \(\square\)

**Example 2.9.** Suppose that \(f(x) = \frac{1}{2}d(x, x_0)^2\). By the CAT(0)-inequality (2.2), \(f\) is (strongly) convex. By the above lemma, we have \(f^*(p) = \frac{1}{2}\|p\|^2\). In this case, \(f^*\) is also convex. The convexity is seen by considering the flat triangle of vertices 0, \(p, q\).

Further investigation of \(f^*\) is left for future research. As mentioned in the introduction, our central interest is how to describe the domain \(\text{dom} f^* = \{p \in CX^\infty \mid f^*(p) < \infty\}\) of the conjugate \(f^*\).
2.1.3 Recession functions (asymptotic slope functions)

Let $f : X \rightarrow \mathbb{R}$ be a continuous convex function. The recession function $f^\infty : CX^\infty \rightarrow \mathbb{R} \cup \{\infty\}$ of $f$ is defined by

$$f^\infty(p) := \lim_{t \rightarrow \infty} \{f(c(t)) - f(c(0))\}/t = \lim_{t \rightarrow \infty} f(c(t))/t \quad (p \in CX^\infty) \quad (2.17)$$

where $c$ is a constant-speed ray with $c(\infty) = p$. The recession function is just a homogeneous extension of the asymptotic slope function by Kapovich, Leeb, and Millson [31], which is defined on $X^\infty$. By convexity, $(f(c(t)) - f(c(0)))/t$ is monotone nondecreasing, and it converges to a finite value or $\infty$. The recession function is indeed independent of the choice of a ray $c$.

**Lemma 2.10** ([31, Lemma 2.10]). $\lim_{t \rightarrow \infty} f(c(t))/t = \lim_{t \rightarrow \infty} f(c'(t))/t$ holds for two asymptotic geodesic rays $c, c'$.

It is easy to see the positive homogeneity of $f^\infty$ (by the change of variable in [2.17]). In particular, $f^\infty(r\xi) = rf^\infty(\xi)$ for $\xi \in X^\infty$. A partial convexity property of asymptotic slope functions is obtained in [31, Lemma 3.2 (ii)]. In the setting of recession functions the following general convexity holds.

**Theorem 2.11.** The recession function $f^\infty$ is positively homogeneous convex.

**Proof.** We have already seen the positive homogeneity. Hence it suffices to show

$$f^\infty(p) + f^\infty(p') \geq f^\infty(p + p') \quad (p, p' \in CX^\infty).$$

Let $p, p' \in CX^\infty$. We can assume that both $p$ and $p'$ are nonzero and both $f^\infty(p)$ and $f^\infty(p')$ are finite. Suppose that $p = a\xi$ and $p' = a'\xi'$ for $a \geq a' > 0$ and $\xi, \xi' \in X^\infty$. Let $x \in X$ and let $\sigma, \sigma'$ be geodesic rays issuing from $x$ with $\sigma(\infty) = \xi$ and $\sigma'(\infty) = \xi'$.

Suppose first that $p + q \neq 0$. Then $\angle(\xi, \xi') < \pi$, or $\angle(\xi, \xi') = \pi$ and $a > a'$. Let $m(t)$ be the midpoint of $\sigma(at)$ and $\sigma'(at)$. We will consider geodesic segment $[x, m(t)]$ with $t \rightarrow \infty$. We show $e := \lim_{t \rightarrow \infty} d(x, m(t))/t > 0$. Indeed, by triangle inequality $d(x, m(t)) \geq d(x, \sigma(at)) - d(\sigma(at), m(t))$ with $d(x, \sigma(at)) = at$ and $d(\sigma(at), m(t)) = d(\sigma(at), \sigma'(at))/2$ we have

$$d(x, m(t))/t \geq a - d(\sigma(at), \sigma'(at))/2t = a - \frac{1}{2} \sqrt{a^2 + a'^2 - 2aa' \cos \angle(x, \sigma(at), \sigma'(at))},$$

where the equality follows from the law of cosine in the comparison triangle of $x, \sigma(at), \sigma(a't')$. Letting $t \rightarrow \infty$, we have $\overline{Z}_x(\sigma(at), \sigma'(at')) \rightarrow \angle(\xi, \xi')$ and

$$e \geq a - \frac{1}{2} \sqrt{a^2 + a'^2 - 2aa' \cos \angle(\xi, \xi')} \geq (a - a')/2.$$

By $a > a'$ or $\angle(\xi, \xi') < \pi$, we have $e > 0$. By [3, II.4.4], it holds $p + p' = 2e\eta$ for some $\eta \in X^\infty$. Let $\rho$ denote the geodesic ray with $\rho(0) = x$ and $\rho(\infty) = \eta$.

We can assume $f(x) = 0$ by replacing $f$ with $f - f(x)$. Consider the constant-speed path $\rho_t : [0, et] \rightarrow X$ with $\rho(0) = x$ and $\rho(et) = m(t)$. Then, by [3, II.4.4], the path $\rho_t$ converges to geodesic ray $\rho$ for $t \rightarrow \infty$. For every $\epsilon > 0$ and $s \geq 0$, there is $t_0$ such that for every $t \geq t_0$ it holds

$$f(\rho(s)) \leq \epsilon + f(\rho_t(s)) \leq \epsilon + f(m(t))d(x, \rho_t(s))/d(x, m(t)).$$
where the second inequality follows from \( f(x) = 0 \) and the convexity of \( f \) along \([x, m(t)]\).

For all large \( s, t \) we have

\[
\frac{f(\sigma(at))}{t} + \frac{f'(\sigma(a't))}{t} \geq 2 \frac{f(m(t))}{t} \geq 2 \left( \frac{f(\rho(s))}{s} - \frac{\epsilon}{s} \right) \frac{s}{d(x, \rho_t(s))} \frac{d(x, m(t))}{t}
\]

By \( t \to \infty \), we have \( d(x, \rho_t(s)) \to s \) and \( d(x, m(t))/t \to \epsilon \). By \( s \to \infty \), we have

\[
f^\infty(p) + f^\infty(p') \geq \lim_{s \to \infty} f(\rho(s))2\epsilon/s = 2\epsilon f^\infty(\eta) = f^\infty(p + p').
\]

Finally, consider the case \( p + p' = 0 \), i.e., \( \angle(\xi, \xi') = \pi \) and \( a = a' \). For \( \epsilon > 0 \), it holds \( p + (1 + \epsilon)p' \neq 0 \). Then, from the above case, we have

\[
f^\infty(p) + (1 + \epsilon)f^\infty(p') = f^\infty(p) + f^\infty((1 + \epsilon)p') \geq f^\infty(p + (1 + \epsilon)p') = \epsilon f^\infty(p').
\]

For the last equality, observe from (2.6) that the midpoint between \( p, (1 + \epsilon)p' \) is \((\epsilon/2)p'\). Hence \( p + (1 + \epsilon)p' = \epsilon p' \) and \( f^\infty(\epsilon p') = \epsilon f^\infty(p') \). By \( \epsilon \to 0 \), we obtain the desired inequality. \( \square \)

It is known \[31\] p. 318] that the asymptotic slope of a Busemann function is given by \( b_\xi(\eta)^\infty = -\cos \angle(\xi, \eta) \) for \( \xi, \eta \in X^\infty \). Hence we have:

**Lemma 2.12.** It holds \( (b_\eta)^\infty(q) = -\langle p, q \rangle \) for \( p, q \in CX^\infty \).

Motivated by \[1.3\], for any positively homogeneous function \( h : CX^\infty \to \mathbb{R} \cup \{\infty\} \) we define a subset \( B(h) \subseteq CX^\infty \) by

\[
B(h) := \{ p \in CX^\infty \mid \langle u, p \rangle \leq h(u) \ (u \in X^\infty \subseteq CX^\infty) \}.
\]  

Since \( p \mapsto \langle u, p \rangle \) is continuous (Lemma 2.3), we have:

**Lemma 2.13.** \( B(h) \) is a closed subset in \( CX^\infty \).

We mainly consider \( B(f^\infty) \) for recession function \( f^\infty \). Belonging to \( B(f^\infty) \) is a necessary condition for \( p \in CX^\infty \) for which \( f + b_p \) is bounded below.

**Lemma 2.14.** \( \text{dom } f^* \subseteq B(f^\infty) \).

**Proof.** Suppose that \( \langle u, p \rangle > f^\infty(u) \) for some \( u \in X^\infty \). Then \( f^\infty(u) - \langle u, p \rangle = (f + b_p)^\infty(u) = \lim_{t \to \infty}(f(c(t)) + b_p(c(t)))/t \) is negative, where \( c \) is a ray with \( c(\infty) = u \). Therefore, for some \( \alpha > 0 \), it holds \( f(c(t)) + b_p(c(t)) < -\alpha t \) for every large \( t > 0 \). This means that \( \inf_{x \in X} f(x) + b_p(x) = -\infty \), and \( p \notin \text{dom } f^* \). \( \square \)

For a subset \( R \subseteq CX^\infty \), let \( \overline{R} \) denote the closure of \( R \) with respect to the \( d^\infty \)-topology. By the above lemma, it holds \( \text{dom } \overline{f^*} \subseteq B(f^\infty) \). We do not know whether the equality holds in general.

---

10
2.2 Hadamard manifolds

From here, we restrict our study to smooth convex optimization on a Hadamard manifold, i.e., a simply-connected complete Riemannian manifold having nonpositive sectional curvature. We utilize elementary concepts in Riemannian geometry; see e.g., [42]. A recent book [9] for optimization perspectives is also useful. For Hadamard manifolds, we consult [7, 17].

Let $M$ be an $n$-dimensional Hadamard manifold. For $x \in M$, let $\langle \cdot, \cdot \rangle_x$ denote the inner product of the tangent space $T_x$. Let $d$ denote the distance function on $M$ obtained from the Riemannian connection. The metric space $M$ is known to be a Hadamard space. A geodesic ray issuing from $x$ is given by the exponential map $\exp_x : T_x \to M$. Namely, for $v \in T_x$, the map $t \mapsto \exp_x(vt)$ is a constant-speed ray with speed $\|v\|_x := \sqrt{\langle v, v \rangle_x}$. The map $\exp_x$ is a diffeomorphism from $T_x$ to $M$.

Consider the boundary $M^\infty$ and its Euclidean cone $CM^\infty$ of $M$. Via the exponential map, the boundary $M^\infty$ is identified with the unit sphere at $T_x$, and the cone $CM^\infty$ is identified with $T_x \simeq \mathbb{R}^n$. This identification gives another topology to $M^\infty$ and to $CM^\infty$, which is called the standard topology, and is independent of the choice of $x$. In the standard topology, $M^\infty$ and $CM^\infty$ are homeomorphic to sphere $S^{n-1}$ and Euclidean space $\mathbb{R}^n$, respectively. It is known [11, II.9.7(1)] that the identity map on $CM^\infty$ from the $d^\infty$-topology to the standard topology is continuous (and is not homeomorphic in general).

Example 2.15. Suppose that $M$ is a hyperbolic space. For distinct $\xi, \eta \in M^\infty$ there is a geodesic line $c : \mathbb{R} \to M$ such that $c(\infty) = \xi$ and $c(-\infty) = \eta$. This means that $\angle(\xi, \eta) = \pi$ for every distinct $\xi, \eta \in M^\infty$. Therefore, the boundary $M^\infty$ is a discrete topological space. See [11] II.9.6 (ii)]. The cone $CM^\infty$ is a star obtained from infinitely many half-lines $\mathbb{R}_+$, each associated with $\xi \in M^\infty$, by identifying the origin of all $\mathbb{R}_+$. See Figure 1. This can be seen from the distance formula (2.6) as $d^\infty(r\xi, s\eta) = r + s$ if $\xi \neq \eta$ and $|r - s|$ if $\xi = \eta$.

Let $f : M \to \mathbb{R}$ be a smooth convex function.

Lemma 2.16. (1) $f^\infty$ is lower semicontinuous in the standard topology [33, Lemma 2.11].

(2) $(p, q) \mapsto \langle p, q \rangle$ is upper semicontinuous in the standard topology.
Proof. (1). Take an arbitrary \( x \in M \), and identify \( CM^\infty \) with \( T_x \) (by \( v \mapsto \exp_x v \infty \)). Let \( f_t(p) := (f(\exp_x(p)) - f(x))/t \) for \( p \in T_x \), which is monotone nondecreasing in \( t \) and continuous on \( T_x \) (in the standard topology). Therefore \( f^\infty(p) = \sup_{t \geq 0} f_t(p) \). It is well-known that the supremum of continuous functions \( f_t \) is lower semicontinuous.

(2). It is known \([11, \text{II.9.5]}\) that \( (\xi, \eta) \mapsto \angle(\xi, \eta) \) is lower semicontinuous. Then \( (p, q) \mapsto \langle p, q \rangle \) is upper semicontinuous, since \( p \mapsto d^\infty(0, p) = \| p \|_x \) is continuous in the standard topology and cosine is a decreasing function.

As in the Euclidean case, a minimizer of \( f \) is characterized by the gradient. The gradient \( \nabla f(x) \in T_x \) of \( f \) at \( x \) is defined via

\[
\langle \nabla f(x), v \rangle_x = df(x)(v) = \lim_{t \to 0} \{f(\exp_x tv) - f(x)\}/t \quad (v \in T_x),
\]

where \( df(x) : T_x \to \mathbb{R} \) is the differential of \( f \) at \( x \). It is easy to see:

**Lemma 2.17** (see \([9, \text{Corollary 11.22}]\)). \( x \in M \) is a minimizer of \( f \) if and only if \( \nabla f(x) = 0 \).

In Euclidean case \( M = \mathbb{R}^n \), \( x \) is a minimizer of \( x \mapsto f(x) - \langle p, x \rangle \) if and only if \( \nabla f(x) = p \). To extend it, we consider the gradient of Busemann functions. Note that Busemann functions are continuously differentiable; see \([17, \text{1.10.2 (1)}]\). As in the previous subsection, we fix \( x_0 \in M \) and identify \( CM^\infty \) with the space of Busemann functions for rays issuing from \( x_0 \).

**Lemma 2.18** (see \([17, \text{1.10.2 (2)}]\)). For \( x \in M \) and \( p \in CM^\infty \), it holds \( \nabla b_p(x) = -u \) for \( u \in T_x \) with \( p = \exp_x u \infty \).

The asymptotic gradient \( \nabla^\infty f(x) \) for \( x \in M \) is defined as the asymptotic class of the ray \( t \mapsto \exp_x t \nabla f(x) \), that is,

\[
\nabla^\infty f(x) := \exp_x \infty \nabla f(x).
\]

Notice that \( \| \nabla^\infty f(x) \| = d^\infty(0, \nabla^\infty f(x)) \) is the speed of \( t \mapsto \exp_x t \nabla f(x) \). Therefore we have

\[
\| \nabla^\infty f(x) \| = \| \nabla f(x) \|_x.
\]

Then we have the following analogue of the one in Euclidean convex analysis.

**Lemma 2.19.** For \( p \in CM^\infty \), the following conditions are equivalent:

(i) \( x \) is a minimizer of \( f + b_p \) over \( M \).

(ii) \( \nabla^\infty f(x) = p \).

(iii) \( f(x) + f^*(p) = -b_p(x) \)

**Proof.** (i) \( \iff \nabla(f + b_p)(x) = 0 \iff \nabla f(x) = u \) for \( u = \exp_x u \infty \iff (ii). (i) \iff (iii) \) is obvious from the definition \([2.16]\) of conjugate \( f^* \).

As in the Euclidean case, the conjugate of a (smooth) convex function recovers the original function via the inverse transformation.

**Lemma 2.20.** \( f(x) = \sup_{p \in CM^\infty} -b_p(x) - f^*(p) \quad (x \in M) \).
Proof. By definition, $f^*(p) \geq -b_p(x) - f(x)$ for every $x \in M$ and $p \in CM^\infty$. Therefore $f(x) \geq \sup_{p \in CM^\infty} -b_p(x) - f^*(p)$. For $p = \nabla^\infty f(x)$, the equality is attained by Lemma 2.19.

This gives rise to an interesting question of characterizing the class of functions $g$ on $CM^\infty$ for which $g^*(x) := \sup_{p \in CM^\infty} -b_p(x) - g(p)$ is convex on $M$. However this is beyond the theme of this paper, and we leave it for future research.

For a set $R \subseteq CM^\infty$, let $\text{int } R$ denote the interior of $R$ in the $d^\infty$-topology.

Lemma 2.21. Let $h : CM^\infty \rightarrow \mathbb{R} \cup \{\infty\}$ be a positively homogeneous function. Suppose that $h$ is lower semicontinuous in the standard topology. Then it holds

$$\text{int } B(h) = \{p \in CM^\infty \mid \langle \xi, p \rangle < h(\xi) \ (\xi \in M^\infty \subseteq CM^\infty)\}. \quad (2.20)$$

Proof. Let $p \in B(h)$. Suppose that $\langle \xi, p \rangle = h(\xi)$ for some $\xi \in M^\infty$. Then, for arbitrary $\epsilon > 0$, $\langle \xi, p + \epsilon \xi \rangle \geq \langle \xi, p \rangle + \epsilon \|\xi\|^2 > h(\xi)$ (Lemma 2.3). This means $p + \epsilon \xi \not\in B(h)$. Since $\langle . \rangle$ is continuous (Lemma 2.3), $p$ is never an interior point of $B(h)$.

Suppose that $\langle \xi, p \rangle < h(\xi)$ for all $\xi \in M^\infty$. For $q \in CM^\infty$, let $\Delta(q) := \inf_{\xi \in M^\infty} h(\xi) - \langle \xi, p \rangle$. Since $M^\infty$ is compact and $\xi \mapsto h(\xi) - \langle \xi, p \rangle$ is lower semicontinuous in the standard topology (Lemma 2.16 (2)), the infimum is always attained. Let $\epsilon := \Delta(p) > 0$. Since $\langle q, \xi \rangle \mapsto h(\xi) - \langle \xi, q \rangle$ is lower semicontinuous, for each $\xi \in M^\infty$ there is an open neighborhood $U_\xi \times V_\xi$ of $(p, \xi)$ (in the standard topology) such that for each $(q, \eta) \in U_\xi \times V_\xi$ we have $h(\eta) - \langle \eta, q \rangle \geq h(\xi) - \langle \xi, p \rangle - \epsilon \geq \Delta(p) - \epsilon = 0$. Since $M^\infty$ is compact, there are $\xi_1, \xi_2, \ldots, \xi_m$ with $M^\infty = \bigcup_{i=1}^m V_{\xi_i}$. Let $U := \bigcap_{i=1}^m U_{\xi_i}$, which is an open neighborhood of $p$ (in the standard topology). For any $q \in U$, $\Delta(q) = h(\xi') - \langle \xi', q \rangle$ for some $\xi'$. Since $\xi'$ belongs to $V_{\xi_i}$ for some $i$, it holds $\Delta(q) = h(\xi') - \langle \xi', q \rangle \geq 0$. Hence, we have $p \in U \subseteq B(h)$. Since the identify map on $CM^\infty$ from the $d^\infty$-topology to the standard topology is continuous, $U$ is an open neighborhood of $p$ in $d^\infty$-topology.

As the proof shows, the RHS of (2.20) is the interior of $B(h)$ also in the standard topology, although $B(h)$ may not be closed in this topology.

We apply this lemma to the recession function $f^*$ and the associated subset $B(f^*)$.

Proposition 2.22. For $p \in \text{int } B(f^*)$, there is a minimizer of $f + b_p$. In particular, any point in $B(f^*) \setminus \text{dom } f^*$ belongs to the boundary of $B(f^*)$.

See [31] Lemma 3.2 (iv)] for a related argument.

Proof. Since $\xi \mapsto f^*(\xi) - \langle \xi, p \rangle$ is lower semicontinuous on compact set $M^\infty$ in the standard topology, the minimum value $\alpha > 0$ exists. Let $\alpha' \in (0, \alpha)$ and let $g := f + b_p$. Fix an arbitrary $x \in M$. Then, for every $\xi \in M^\infty$ there is $t_\xi \in \mathbb{R}_+$ such that $g(\exp_x t_\xi) - g(x) > \alpha' t$ for all $t \geq t_\xi$. Define $h : M^\infty \rightarrow \mathbb{R}_+$ by

$$h(\xi) := \inf\{t \geq 0 \mid g(\exp_x t_\xi) - g(x) > \alpha' t\} \quad (\xi \in M^\infty).$$

Then $h$ is upper semicontinuous, since the epigraph $\{(\xi, t) \in M^\infty \times \mathbb{R} \mid t \leq h(\xi)\}$ is the closed set $\{(\xi, t) \in M^\infty \times \mathbb{R}_+ \mid g(\exp_x t_\xi) - g(x) \leq \alpha' t\} \cup \{(\xi, t) \in M^\infty \times \mathbb{R} \mid t \leq 0\}$. Since $M^\infty$ is compact, the maximum $t^*$ of $h$ over $M^\infty$ exists. Then, for every $\xi \in M^\infty$, we have $g(\exp_x t_\xi) - g(x) \geq \alpha' t$ for all $t \geq t^*$. This means that the level set $\{y \in M \mid g(y) \leq g(x) + \alpha' t^*\}$ belongs to the metric ball at center $x$ with radius $t^*$, which is compact by Hopf-Rinow theorem (see [42] III.1)). A minimizer of $g$ exists in this set.

\[\square\]
Thus we have
\[ \text{int } B(f^\infty) \subseteq \nabla^\infty f(M) \subseteq \text{dom } f^* \subseteq B(f^\infty). \] (2.21)

We next provide a characterization of \( B(f^\infty) \), which sharpens the following important result by Kapovich, Leeb, and Millson [31].

**Theorem 2.23** ([31 Lemma 3.4]). If \( f^\infty(u) \geq 0 \) for all \( u \in M^\infty \), i.e., \( 0 \in B(f^\infty) \), then \( \inf_{x \in M} \| \nabla f(x) \|_x = 0 \).

An outline of the proof is as follows: If \( \inf_{x \in M} \| \nabla f(x) \|_x > 0 \), then a trajectory of the normalized gradient flow of \( f \) goes to \( u \in M^\infty \) with \( f^\infty(u) < 0 \); See also [16 section 5.4].

We now obtain an analogue of the equivalence between (a) and (c) in the introduction.

**Theorem 2.24.** For \( p \in CM^\infty \), the following are equivalent:

(a) \( \inf_{x \in M} \| \nabla (f + b_p)(x) \|_x = 0 \).

(c) \( p \in B(f^\infty) \).

**Proof.** (c) \( \Rightarrow \) (a) follows from applying the above theorem to \( f + b_p \). We verify (a) \( \Rightarrow \) (c). Let \( g := f + b_p \). For arbitrary \( \epsilon > 0 \), there is \( x \in M \) such that \( \| \nabla g(x) \|_x < \epsilon \). Consider \( u \in M^\infty \) and the geodesic ray \( t \mapsto \exp_x vt \) with \( \exp_x v\infty = u \). Then \( \lim_{t \to 0} (g(\exp_x vt) - g(x))/t = (\nabla g(x), v)_x = \| \nabla g(x) \|_x \cos \theta > -\epsilon \), where \( \theta \) is the angle between \( \nabla g(x) \) and \( v \) in \( T_x \). Since \( (g(\exp_x vt) - g(x))/t \) is monotone nondecreasing, by \( t \to \infty \) we have \( g^\infty(u) = f^\infty(u) - \langle u, p \rangle > -\epsilon \). Thus, for every \( \epsilon > 0 \) and \( u \in M^\infty \), it holds \( \langle u, p \rangle < f^\infty(u) + \epsilon \). This implies \( p \in B(f^\infty) \). \( \square \)

The condition (a) \( \inf_{x \in M} \| \nabla (f + b_p)(x) \|_x = 0 \) may be viewed as a correspondent of \( \inf_{x \in \mathbb{R}^n} \| \nabla f(x) - p \| = 0 \) of the Euclidean case. If \( \nabla^\infty f(x_i) \) \((i = 1, 2, \ldots)\) converges to \( p \in CM^\infty \) in the \( d^\infty \)-topology, then (a) holds. Indeed, by Lemma 2.18 and (2.19) it holds

\[
\| \nabla (f + b_p)(x) \|_x^2 = \| \nabla f(x) \|_x^2 + \| u \|_x^2 - 2 \nabla f(x) \|_x \| u \|_x \cos \theta_x \\
\leq \| \nabla^\infty f(x) \|_x^2 + \| p \|^2 - 2 \| \nabla^\infty f(x) \| \| p \| \cos \theta \\
= d^\infty(\nabla^\infty f(x), p)^2,
\]

where \( u := -\nabla b_p(x) \), \( \theta_x \) is the angle between \( \nabla f(x) \) and \( u \) in \( T_x \), and \( \theta := \angle(\nabla^\infty f(x), p) \geq \theta_x \) by definition (2.5). However, the converse is not true. Also (a) does not mean the convergence in the standard topology.

### 2.3 Symmetric spaces of nonpositive curvature

Here we consider symmetric spaces of nonpositive curvature, which constitute a fundamental class of Hadamard manifolds. Our argument basically consults [17 Chapters 2 and 3]. In a Hadamard manifold \( M \) and a point \( x \), the geodesic symmetry \( \sigma_x : M \to M \) at \( x \) is defined by \( \sigma_x(\exp_x(v)) = \exp_x(-v) \) for \( v \in T_x \). A symmetric space of nonpositive curvature is a Hadamard manifold \( M \) such that for every \( x \in M \) the geodesic symmetry \( \sigma_x \) is an isometry on \( M \). Via the de Rham decomposition theorem (see [12 III.6]), \( M \) is (uniquely) decomposed as Riemannian product \( M = M_0 \times N \), where \( M_0 \) is isometric.
to Euclidean space $\mathbb{R}^k (k \geq 0)$ (called the Euclidean de Rham factor of $M$) and a sym-
metric space $N$ of noncompact type, i.e., it is given by $N = G/K$ for a (real) semisimple
Lie group $G$ and its maximal compact subgroup $K$. Then $N$ has a trivial Euclidean de
Rham factor, and its Riemannian structure is given by a $G$-invariant metric. We will
see more concrete constructions in Sections 2.4 and 3.2.

Let $M$ be a symmetric space of nonpositive curvature, and let $x_0 \in M$. A $k$-
dimensional flat is a submanifold of $M$ isometric to $\mathbb{R}^k$. A maximal flat is a flat that
is not contained in another flat of a larger dimension. It is a basic fact that all maximal
flats have the same dimension $d$, which is called the rank of $M$. By a geodesic line
we mean a map $l : \mathbb{R} \rightarrow M$ with $d(l(s), l(t)) = |s - t|$ for $s, t \in \mathbb{R}$, which is just a 1-
dimensional flat. A geodesic line is called regular if it is contained by a unique maximal
flat. Let $x \in M$ and let $F$ be a maximal flat containing $x$. A Weyl chamber at tip $x$
is the closure of a connected component of the set of points $y \in F \setminus \{x\}$ such that the
unique geodesic line containing $x, y$ is regular. When $F$ is viewed as $\mathbb{R}^d$ with origin $x$,
Weyl chambers are polyhedral cones. They have the same shape, since the group $G$
acts transitively on them.

We next explain the building structure of the boundary $M^\infty$ and its cone $CM^\infty$; [7
Appendix 5] is a useful reference. It is clear (from Example 2.4) that the boundary $F^\infty$
of a maximal flat $F$ is isometric to sphere $S^{d-1}$. The boundary $C^\infty$ of a Weyl chamber
$C$ (at some tip) is called an (asymptotic) Weyl chamber. For two Weyl chambers
$C, D$ (at possibly different tips), the asymptotic ones $C^\infty, D^\infty$ are the same or have
disjoint interiors. Then, the set of all Weyl chambers and their faces give rise to a
cell-complex structure on $M^\infty$, where each cell is isometric to a polyhedral cell in a
sphere (the intersection of a sphere and a polyhedral cone). This structure is an $M_1$-
polyhedral complex in the sense of [11 Chapter I.7], and forms a spherical building,
where an apartment is precisely the subcomplex formed by the boundary of a maximal
flat. Although Weyl chambers here may not be simplices, one can subdivide them to
obtain a simplicial complex $\mathcal{C}$ so that each apartment is a spherical Coxeter complex.
Then the apartments are glued nicely. That is, they satisfy, as an abstract simplicial
complex, the axiom of building:

- Any two simplices in $\mathcal{C}$ are contained in a common apartment.
- For two apartments $A, A'$ including simplices $C, C'$, there is an isomorphism $A \rightarrow
A'$ fixing $C, C'$ pointwise.

See e.g., [1] and [11 II.10. Appendix] for (formal) theory of building. By considering
the Euclidean cone, $CM^\infty$ has the structure of a Euclidean building. A maximal cone
is also called a Weyl chamber. Apartments are the subcomplexes induced by $CF^\infty$
for all maximal flats $F$. We simply call $CF^\infty$ an apartment. Each apartment is a convex
subspace of $CM^\infty$ isometric to $\mathbb{R}^d$. We can identify an apartment $E \subseteq CM^\infty$ with $\mathbb{R}^d$
so that the origins coincide. In this identification, $\langle , \rangle$ in $E$ is precisely the Euclidean
inner product of $\mathbb{R}^d$.

Suppose that $M = \mathbb{R}^k \times N$, where $N$ is a symmetric space $G/K$ of noncompact type.
Then $CM^\infty = \mathbb{R}^k \times CN^\infty$. The structure of building $CM^\infty$ is determined by $CN^\infty$.
We can suppose (as in [17]) that $G$ is the identity component of the group of isometries
of $N$, where $G$ acts isometrically on $M = \mathbb{R}^k \times N$ by $\mathbb{R}^k \times N \ni (x', x) \mapsto (x', gx)$.
Since any isometry on $M$ induces an isometry on $M^\infty$, the group $G$ acts isometrically
on $M^\infty$ and on $CM^\infty$ by $gp = ||p||g(p/||p||)$. Also $G$ acts on the set of Weyl chambers
and their faces. The facial incidence structure of the building $M^\infty$ is described by the inclusion relation of all parabolic subgroups of $G$. Here a **parabolic subgroup** is the subgroup consisting of $g \in G$ with $g\xi = \xi$ for some $\xi \in N^\infty$, which is denoted by $G_\xi$. If $\xi, \xi'$ belong to the relative interior of the same face of a Weyl chamber, then $G_\xi = G_{\xi'}$. Therefore, minimal parabolic subgroups correspond to Weyl chambers. Fix a Weyl chamber $C_0$, and regard it as a polyhedral cone in $\mathbb{R}^d$. Let $P$ be the minimal parabolic subgroup for $C_0$. The set of Weyl chambers is identified with the flag variety $G/P$. If $p \in CM^\infty$ belongs to a Weyl chamber corresponding to $F \in G/P$ and the orbit of $p \in CM^\infty$ by $G$-action meets a (unique) point $\lambda$ in $C_0 \subseteq \mathbb{R}^d$, then we denote $p$ by

$$p = \lambda \cdot F.$$  

(2.22)

This notation designates the coordinate $\lambda$ of $p$ in the Weyl chamber indexed by $F$.

Two geodesic lines $c, c' : \mathbb{R} \to M$ are said to be parallel if $d(c(t), c'(t))$ ($t \in \mathbb{R}$) is bounded above. For a geodesic line $c$, let $F(c)$ denote the union of all geodesic lines parallel to $c$. It is known that $F(c)$ is a totally geodesic submanifold of $M$. If $c$ is regular, $F(c)$ is a maximal flat.

For $\xi \in N^\infty$, the **horospherical subgroup** $N_\xi$ of $G_\xi$ consists of $n \in G$ such that

$$\lim_{t \to -\infty} d(n(c(t)), c(t)) = 0,$$

where $c$ is the geodesic ray with $c(\infty) = \xi$ and $c(0) = x_0$ (independent of $x_0$). Then $N_\xi$ keeps the Busemann function $b_\xi$ as $b_\xi(nx) = b_\xi(x)$, since

$$|b_\xi(nx) - b_\xi(x)| = \lim_{t \to -\infty} |d(x, n^{-1}c(t)) - d(x, c(t))| \leq \lim_{t \to -\infty} d(c(t), n^{-1}c(t)) = 0.$$

The **generalized Iwasawa decomposition** [L7, 2.17.5 (5)] implies that $M$ is diffeomorphic to $N_\xi \times F(c)$ by $(n, y) \mapsto n(y)$.

In this setting, let us start our convex analysis on $M$. Let $f : M \to \mathbb{R}$ be a smooth convex function. For a maximal flat $F$, let $f_F : F \to \mathbb{R}$ denote the restriction of $f$ to $F$. The asymptotic Legendre-Fenchel conjugate $f_F^* : CF^\infty \to \mathbb{R} \cup \{\infty\}$ is defined by $f_F^*(p) := \sup_{x \in F} -b_p(x) - f(x)$, where $x_0$ may be outside of $F$. By $p \in CF^\infty$ and Example 2.7, $b_p$ is an affine function on $F$. Therefore, $f_F^*$ is viewed as the ordinary Legendre-Fenchel conjugate (up to an additive constant) under identification $CF^\infty = \mathbb{R}^d$. Hence we have:

**Lemma 2.25.** For any maximal flat $F$, $f_F^*$ is a convex function on $CF^\infty$.

Still, $f^*$ may be nonconvex but “partially” convex in the following sense:

**Proposition 2.26.** For a Weyl chamber $C \subseteq CM^\infty$, it holds

$$f^*(p) = \sup_{F} f_F^*(p) \quad (p \in C),$$

(2.23)

where sup is taken over all maximal flats $F$ with $CF^\infty \supseteq C$. In particular, $f^*$ is a convex function on $C$, and $C \cap \text{dom } f^*$ is a convex set contained by a convex set $\cap_F C \cap \text{dom } f_F^*$:

$$C \cap \text{dom } f^* \subseteq \cap_F C \cap \text{dom } f_F^*.$$  

(2.24)

Note that $\subseteq$ may be strict since $f^*(p) = \sup_{F} f_F^*(p)$ may be $\infty$ even if $f_F^*(p) < \infty$ for every $F$.

**Proof.** Consider the Iwasawa decomposition $M = N \times F_0$, where $F_0$ is a maximum flat containing $C$ at infinity and $N$ is the horospherical subgroup for $C$. Then $NF_0$ ranges
over all maximal flats containing $C$ at infinity. Thus $M$ is the (disjoint) union of all maximal flats $F$ such that $F^\infty \supseteq C \ni p$. Then we have $f^*(p) = \sup_{x \in F} \sup_{b \in \mathbb{R}} -b_p(x) - f(x) = \sup_F f_F^*(p)$.

Next, we consider the recession function $f^\infty$ and the associated subset $B(f^\infty)$. For each apartment $E$, we also consider

$$B_E(f^\infty) := \{ p \in E \mid \langle u, p \rangle \leq f^\infty(u) \ (u \in E) \}.$$  \ (2.25)

**Proposition 2.27.** For a Weyl chamber $C \subseteq E$, it holds

$$C \cap B(f^\infty) = \bigcap_F C \cap B_{C^F}(f^\infty) = \bigcap_F C \cap \text{dom} f_{F^*}^\infty.$$  \ (2.26)

where $\bigcap$ is taken over all maximal flats $F$ with $CF^\infty \supseteq C$. In particular, $C \cap B(f^\infty)$ is a convex set in $C$.

**Proof.** Since $CM^\infty$ is the union of all apartments $E$ containing any fixed Weyl chamber $C$, we have the first equality. When $E$ is viewed as Euclidean space $\mathbb{R}^n$, then $C$ is a convex cone, $\langle u, p \rangle \leq f^\infty(u)$ is a linear inequality, and therefore $C \cap B_E(f^\infty)$ is convex. Necessarily the intersection $\bigcap_E C \cap B_E(f^\infty)$ is convex. By Lemma 2.16, the restriction of $f^\infty$ to Euclidean space $E$ is lower semicontinuous and positively homogeneous convex, and must be the support function of $\text{dom} f_{E_\ast}^\infty$. Hence we have $\text{dom} f_{E_\ast}^\infty = B_E(f^\infty)$, and the second equality.

In the view of (2.24) and (2.26), the closure of $\text{dom} f_{E_\ast}^\infty$ seems very close to $B(f^\infty)$, although we do not know whether they really differ. Under non-degeneracy assumptions, three spaces $\nabla^\infty f(M)$, $\text{dom} f_{E_\ast}^\infty$, and $B(f^\infty)$ are equal, as in Euclidean case.

**Proposition 2.28.** For a Weyl chamber $C$, if $C \cap B(f^\infty)$ has nonempty interior, then $C \cap \nabla^\infty f(M) = C \cap \text{dom} f_{E_\ast}^\infty = C \cap B(f^\infty)$.

**Proof.** By (2.21), it holds

$$C \cap \text{int} B(f^\infty) \subseteq C \cap \nabla^\infty f(M) \subseteq C \cap \text{dom} f_{E_\ast}^\infty \subseteq C \cap B(f^\infty).$$  \ (2.27)

In Euclidean space, for any closed convex set $D$ having nonempty interior, it holds $\text{int} D = D$. This can be applied to $C \cap B(f^\infty)$, which is viewed as a Euclidean convex set. From $C \cap B(f^\infty) = \text{int}(C \cap B(f^\infty)) \subseteq C \cap \text{int} B(f^\infty) \subseteq C \cap B(f^\infty)$, by taking the closure in (2.27), we have the claim.

**Proposition 2.29.** Suppose that $f$ is strictly convex. Then the gradient map $x \mapsto \nabla^\infty f(x)$ is a bijection from $M$ to $\text{int} B(f^\infty)$. For a Weyl chamber $C$, if $C \cap \nabla^\infty f(M) \neq \emptyset$, then $C \cap \nabla^\infty f(M) = C \cap \text{dom} f_{E_\ast}^\infty = C \cap B(f^\infty)$.

**Proof.** We first verify $\nabla^\infty f : M \to CM^\infty$ is injective. Indeed, if $\nabla^\infty f(x) = \nabla^\infty f(y) = p$, then $x$ and $y$ are minimizers of $f + b_p$, and it must hold $x = y$ by strict convexity of $f + b_p$. For surjectivity, from Proposition 2.22 it suffices to show that $\nabla^\infty f(x)$ for any $x \in M$ belongs to $\text{int} B(f^\infty)$. We utilize the following property, where $C$ is a Weyl chamber.

$$b_p + b_q = b_{p+q} \ (p, q \in C).$$  \ (2.28)
Indeed, consider a maximal flat $F$ containing $x_0$ and $C$ at infinity, and consider horospherical subgroup $N$ of $C$ and Iwasawa decomposition $M = N \times F$. If $x = n(y)$ for $n \in N$ and $y \in F$, then $b_p(x) + b_q(x) = b_p(y) + b_q(y) = -\langle p, y - x_0 \rangle - \langle q, y - x_0 \rangle = -\langle p + q, y - x_0 \rangle = b_{p+q}(y) = b_{p+q}(x)$, where $p + q \in C$ and $F$ is viewed as $\mathbb{R}^d$ (see Example 2.7).

Let $p := \nabla^\infty f(x)$ for $x \in M$. Suppose that $0 \neq p = \|p\|\xi$ for $\xi \in M^\infty$. Consider sufficiently small $\epsilon > 0$. Let $B_\epsilon$ be the set of all points $q$ represented as $q = (\|p\| - t)\xi + 2t\eta$, where $t \leq \epsilon$ and $\xi, \eta \in CM^\infty$ belong to the same Weyl chamber. Then $B_\epsilon$ contains $p$ in its interior.

We show that $B_\epsilon \subseteq \nabla^\infty f(M)$ for small $\epsilon > 0$. By strict convexity of $f$, $x$ is a unique minimizer of $f + b_p$. Let $\alpha > 0$ be the minimum of $(f + b_p)(\exp_x v) - (f + b_p)(x)$ over all $v \in T_x$ with $\|v\|_x = 1$, which exists by compactness. For $q = (\|p\| - t)\xi + 2t\eta \in B_\epsilon$, it holds $f + b_q = f + b_p - tb_\xi + 2tb_\eta$ by (2.28). Here the additional term $-tb_\xi + 2tb_\eta$ is $3\epsilon$-Lipschitz (Lemma 2.5). For $3\epsilon \leq \alpha$, a minimizer of $f + b_q$ exists in the unit ball around $x$. Thus $q \in \nabla^\infty f(M)$, as required. The proof of the case $p = 0$ is similar; omit $-t\xi$ above.

If $\nabla^\infty f(M) \cap C \neq 0$, then $\nabla^\infty f(M) \cap C$ has nonempty interior, and Proposition 2.28 is applicable to obtain the latter statement.

Note that there is a possibility that $C \cap \nabla^\infty f(M)$ is empty but $C \cap B(f^\infty)$ is nonempty and has no interior. Further refined study on such a degenerate situation is left for future research.

Via the inverse of $x \mapsto \nabla^\infty f(x)$, the symmetric space $M$ is coordinatized by the asymptotic gradient space $\nabla^\infty f(M)$. This can be viewed as a generalization of the dual coordinate of dually-flat manifolds in information geometry [3]. It is an interesting research direction to develop an information-geometrical theory based on this idea.

In the case of $\text{dom} f^* = B(f)$ which we will face in Section 3, the boundedness of $\inf_{x \in M} (f + b_p)$ is verified by convex optimization on Euclidean building $CM^\infty$:

$$\inf \ f^\infty(u) - \langle u, p \rangle \quad \text{s.t.} \quad u \in U,$$

where $U \subseteq CM^\infty$ is any convex neighborhood of the origin, such as a ball.

### 2.4 Symmetric space $P_n = \text{GL}(n, \mathbb{C})/U(n)$

Here we consider the symmetric space $P_n = \text{GL}(n, \mathbb{C})/U(n)$ of positive definite Hermitian $n \times n$ matrices, and present concrete descriptions and specializations of several concepts introduced above. Arguments regarding the PSD-cone as a symmetric space are found in [11] Chapter II.10 and [17] 2.2.13, where they consider $\text{GL}(n, \mathbb{R})/O(n)$ or $\text{SL}(n, \mathbb{R})/SO(n)$ but the arguments are analogous.

When regarding $P_n$ as a Riemannian manifold, the tangent space $T_x$ at $x \in P_n$ is identified with the space $S_n$ of $n \times n$ Hermitian matrices and the inner product is given by $\langle H, H' \rangle_x = \text{tr} x^{-1}Hx^{-1}H'$. The cotangent space $T^*_x$ is also identified with $S_n$ by $H(H') := \text{tr} HH'$ for $H, H' \in S_n$. Let $d$ denote the corresponding distance function on $M$. The exponential map $\exp_x : T_x \to P_n$ at $x$ is given by $H \mapsto x^{1/2}e^xHx^{-1/2}x^1/2$. In particular, any constant-speed ray is written as $t \mapsto g e^{t \text{diag} \lambda} g^\dagger$ for $\lambda \in \mathbb{R}^n$ and $g \in \text{GL}(n, \mathbb{C})$, where $\text{diag} \lambda$ denotes the diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ in order, and $(\cdot)^\dagger$ denotes the complex conjugate. By $(g, x) \mapsto gxg^\dagger$, $\text{GL}(n, \mathbb{C})$ acts
isometrically and transitively on $P_n$. The isotropy group at $I$ is the group $U(n)$ of unitary matrices. So $P_n$ is a symmetric space $GL(n, \mathbb{C})/U(n) = \mathbb{R} \times (SL(n, \mathbb{C})/SU(n))$, where the geodesic symmetry at $x \in P_n$ is given by $y \mapsto xy^{-1}x$. The identity matrix $I$ is naturally chosen as the base point $x_0$ of $P_n$.

Any maximal flat is the set of matrices of form

$$F(g) := \{ge^{\text{diag}\lambda}g^\dagger \mid \lambda \in \mathbb{R}^n\}$$

for $g \in GL(n, \mathbb{C})$, where $\lambda \mapsto ge^{\text{diag}\lambda}g^\dagger$ is an isometry from $\mathbb{R}^n$ to $F(g)$. From this, we see that a geodesic ray $t \mapsto ge^{t\text{diag}\lambda}g^\dagger$ is regular if and only if all values $\lambda_i$ are different. A vector $\lambda \in \mathbb{R}^n$ is said to be arranged if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The next lemma characterizes asymptotic classes of geodesic rays.

**Lemma 2.30** (See [11 II.10.64]). For arranged vectors $\lambda, \mu \in \mathbb{R}^n$ with $\|\lambda\|_2 = \|\mu\|_2 = 1$ and $g, h \in GL(n, \mathbb{C})$, two geodesic rays $t \mapsto ge^{t\text{diag}\lambda}g^\dagger$ and $t \mapsto he^{t\text{diag}\mu}h^\dagger$ are asymptotic if and only if $\lambda = \mu$ and $(h^{-1}g)_{ij} = 0$ for all $i \geq j$ with $\lambda_i < \lambda_j$.

**Proof.** Let $b(t) := e^{-t\text{diag}\lambda/2}hge^{t\text{diag}\lambda/2}$, where $(b(t))_{ij} = e^{-t(\lambda_i-\lambda_j)/2}(h^{-1}g)_{ij}$. Then we have $d(ge^{t\text{diag}\lambda}g^\dagger, he^{t\text{diag}\mu}h^\dagger) = d(b(t)b(t)^\dagger, I)$. Consider the smallest $i$ with $\lambda_i \neq \mu_i$ if it exists. Say $\lambda_i < \mu_i$. Then $\lim_{t \to \infty} b(t)$ has a zero block of $i$ rows and $n-i+1$ columns, and is singular (and/or has an infinite entry). Necessarily $d(b(t)b(t)^\dagger, I) \to \infty$. If $\lambda = \mu$ and $(h^{-1}g)_{ij} \neq 0$ for some $i \geq j$ with $\lambda_i < \lambda_j$, then $(b(t)b(t)^\dagger)_{ii} \to \infty$, implying $d(b(t)b(t)^\dagger, I) \to \infty$. Conversely, suppose that the condition is satisfied. Then $b(t)$ converges to a constant matrix. This means that $d(b(t)b(t)^\dagger, I)$ is bounded, and the two rays are asymptotic.

Thus, the minimal parabolic subgroup for Weyl chamber $C_0 := \{e^{\text{diag}\lambda} \mid \lambda : \text{arranged}\}$ is the group $B$ of upper triangular matrices. Then $GL(n, \mathbb{C})/B$ is viewed as the space of complete flags $U_1 \subset U_2 \subset \cdots \subset U_n = \mathbb{C}^n$ of vector subspaces $U_i$ of $\mathbb{C}^n$. As an abstract simplicial complex, the building $CP_n^\infty$ is the order complex of the lattice of all (nonzero) vector subspaces of $\mathbb{C}^n$. As consistent with (2.22), any point $p$ in $CP_n^\infty$ is represented as $p = \lambda \cdot \mathcal{U}$ for an arranged vector $\lambda$ of complete flag $\mathcal{U}$. Specifically, if $p$ is written as $p = ge^{\text{diag}\lambda}g^\dagger$ for an arranged vector $\lambda$ and $g \in GL(n, \mathbb{C})$, then $\mathcal{U}$ is the complete flag consisting of vector subspaces spanned by the first $i$ columns of $g$ for $i = 1, 2, \ldots, n$. The $i$-th vector subspaces of complete flags $\mathcal{U}$ and $\mathcal{V}$ are denoted by $U_i$ and $V_i$, respectively. Lemma 2.30 rephrases as: $\lambda \cdot \mathcal{U} = \mu \cdot \mathcal{V}$ if and only if $\lambda = \mu$ and $U_i = V_i$ for each $i \in [n-1]$ with $\lambda_i > \lambda_{i+1}$. By associating $\lambda \cdot \mathcal{U}$ with formal sum $\sum_{i=1}^n(\lambda_i - \lambda_{i+1})U_i$ with $\lambda_{n+1} := 0$, the boundary $CP_n^\infty$ is also identified with the set of all formal sums

$$p = \sum_i \alpha_i U_i,$$

where vector subspaces $U_i$ form a (partial) flag and nonzero coefficients $\alpha_i \in \mathbb{R}$ are positive if $U_i \neq \mathbb{C}^n$. Note that this expression (2.31) is unique. In this way, any single vector subspace $X(= 1X)$ is viewed as a point of $CP_n^\infty$.

A Weyl chamber $C \subseteq CP_n^\infty$ consists of $\lambda \cdot \mathcal{U}$ over all arranged vectors $\lambda$ for a (fixed) flag $\mathcal{U}$. The distance $d^{CP}(p, q)$ of two points $\lambda \cdot \mathcal{U}$ and $\mu \cdot \mathcal{U}$ in the same chamber is given by $\|\lambda - \mu\|_2$. Accordingly, the distance of any two points in $CP_n^\infty$ is given by the length metric of the two points (the infimum of the length of a path connecting them). The whole space $CP_n^\infty$ is a polyhedral cone complex obtained by gluing these
Euclidean polyhedral cones. An apartment \( E(g) := F(g) = \{ g e^{\lambda \text{diag}} \mid \lambda \in \mathbb{R}^n \} \) for \( g \in GL(n, \mathbb{C}) \) consists of all points of form \( \lambda \cdot U \) such that each subspace \( U_i \) in flag \( U \) is spanned by column vectors of \( g \). \( GL(n, \mathbb{C}) \) acts isometrically on \( CP_n^{\infty} \) by \( (g, \lambda \cdot U) \mapsto \lambda \cdot gU \).

**Example 2.31.** Consider the case of \( n = 2 \). Any complete flag \( U \) is uniquely determined by its 1-dimensional subspace \( U = U_1 \). The corresponding Weyl chamber is a form of \( \{ \lambda \cdot U \mid \lambda \in \mathbb{R}^2, \lambda_1 \geq \lambda_2 \} \), and is isometric to half-plane \( \{ x \in \mathbb{R}^2 \mid x_1 \geq x_2 \} := C_U \subseteq \mathbb{R}^2 \). Then \( CP_2^{\infty} \) is obtained by gluing \( C_U \) for all 1-dimensional subspaces \( U \), along the line of \( x_1 = x_2 \). Specifically, it is the disjoint union \( \bigsqcup U C_U \) over all 1-dimensional subspaces \( U \) modulo the equivalence relation: \( (U, x) \sim (U', x') \) if and only if \( x_1 = x_2 = x'_1 = x'_2 \). Subspaces \( U \) and \( \mathbb{C}^2 \) are the points of \( CP_2^{\infty} \) that are the images of \( (U, (1, 0)) \) and \( (U, (1, 1)) \), respectively. See Figure 2. This shape of \( CP_2^{\infty} \) can be directly seen from the expression \( CP_2^{\infty} = \mathbb{R} \times C(SL(2, \mathbb{C})/SU(2))^{\infty} \). Here \( SL(2, \mathbb{C})/SU(2) \) is a 3-dimensional hyperbolic space, and \( C(SL(2, \mathbb{C})/SU(2))^{\infty} \) is an infinite star (Example 2.15).

In this setting, we present explicit descriptions of Busemann functions, asymptotic gradients, and inner product \( \langle \cdot, \cdot \rangle \) on \( CP_n^{\infty} \). For \( g \in GL(n, \mathbb{C}) \), let \( [g] \) denote the image in \( G/B \), which is the complete flag of vector subspaces spanned by the first \( i \) columns of \( g \) for \( i = 1, 2, \ldots, n \). For a Hermitian matrix \( H \) and subset \( I \subseteq [n] := \{1, 2, \ldots, n\} \), let \( H[I] \) denote the principal matrix of \( H \) consisting of row/column indices in \( I \), where \( H[0] := 1 \).

**Lemma 2.32** (see [11, 11.10.69]). For \( p = \lambda \cdot [u] \in CP_n^{\infty} \) with \( u \in U(n) \), it holds

\[
\begin{align*}
\nu_p(x) &= -\sum_{i=1}^{n} \lambda_i \log \frac{\det(u^t xu)[\{i, \ldots, n\}]}{\det(u^t xu)[\{i+1, \ldots, n\}]} , \\
\nabla \nu_p(x) &= -ub \text{diag } \lambda b_u^t ,
\end{align*}
\]

where \( u^t x^{1/2} = bk \) for upper triangular matrix \( b \) and unitary matrix \( k \) (Gram–Schmidt orthonormalization).

**Proof.** It suffices to consider \( p \in P_n^{\infty} \). The geodesic ray \( c \) issuing from \( I \) with \( c(\infty) = p \) is written as \( \nu(t) = uc^t \text{diag } \lambda u^t \). Therefore, we have \( \nu_p(x) = \lim_{t \to \infty} d(uc^t \text{diag } \lambda u^t, x) = t = \)
lim_{t \to \infty} d(e^{\lambda}u^t, u^t xu) - t. Suppose that all \( \lambda_i \) are different. Decompose \( u^t xu = n \text{diag} \, r n^t \), where \( n \) is an upper triangular matrix having 1 on each diagonal and \( r \in \mathbb{R}^n \) is a positive vector with \( r_i \) written as

\[
    r_i = \frac{\det(u^t xu)[\{i, \ldots, n\}]}{\det(u^t xu)[\{i + 1, \ldots, n\}].}
\]

As in the proof of Lemma 2.30, it holds \( \lim_{t \to \infty} d(e^{\lambda}u^t, e^{\lambda}u^t) = 0 \), and we have \( b_p(x) = \lim_{t \to \infty} d(e^{\lambda}u^t, \text{diag} \, r) - t = \lim_{t \to \infty} \| t \text{diag} \, \lambda - \log r \|_2 - t = - \sum_{i=1}^n \lambda_i \log r_i \), (see Example 2.7) where \( \log r \in \mathbb{R}^n \) is defined by \( (\log r)_i := \log r_i \). Then we obtain (2.32) from (2.34). If some of \( \lambda_i \) are equal, we decompose \( u^t xu \) to \( n y n^t \), where \( n \) is an upper triangular matrix satisfying \( n_{ii} = 1 \) and \( n_{ij} = n_{ji} = 0 \) if \( \lambda_i = \lambda_j \) and \( i \neq j \), and \( y \) is a block diagonal matrix with \( y_{ij} = y_{ji} = 0 \) if \( \lambda_i \neq \lambda_j \). As above, it holds \( b_p(x) = \lim_{t \to \infty} d(e^{\lambda}u^t, y) - t \). Diagonizing \( y \) in each block by unitary matrices, we obtain the same formula.

Next we verify the second equation. For \( H := ub \text{diag} \, \lambda b^t u \), consider the geodesic \( t \mapsto x^{1/2} e^{t/2} x^{-1/2} H x^{-1/2} x^{1/2} \) (issuing from \( x \)). By \( x^{1/2} = ubk = k^t b^t u \) and \( x^{-1/2} = k^t b^t u \), we have \( x^{1/2} e^{t/2} x^{-1/2} H x^{-1/2} x^{1/2} = ub e^{t \text{diag} \, \lambda} b^t u \). Hence, by Lemma 2.30, this geodesic is asymptotic to \( c \), i.e., \( t \mapsto u e^{t \text{diag} \, \lambda} u^t \). From Lemma 2.18, we have \( \nabla b_p(x) = -H \).

Let \( \| \cdot \|_F \) denote the Frobenius norm; then \( \| H \|_x = \| x^{-1/2} H x^{-1/2} \|_F \) for \( H \in T_x \).

**Proposition 2.33.** Let \( f : P_n \to \mathbb{R} \) be a smooth convex function.

1. For \( p = \lambda \cdot [u] \) with unitary matrix \( u \), it holds

\[
    \| \nabla(f + b_p)(x) \|_x = \| b^t u^t df(x) ub - \text{diag} \, \lambda \|_F, \tag{2.35}
\]

where \( u^t x^{1/2} = bk \) for upper triangular matrix \( b \) and unitary matrix \( k \). In particular, if \( \nabla \infty(f)(x) = p \), then \( x^{1/2} df(x)x^{1/2} = k^t \text{diag} \, \lambda k \).

2. Let \( s \) denote the projection \( \lambda \cdot \mathcal{U} \mapsto \lambda \). Then it holds

\[
    s \nabla \infty(f(P_n)) = s \text{dom} \, f^* = s \mathcal{B}(f^\infty) = \bigcup_C s(C \cap \mathcal{B}(f^\infty)), \tag{2.36}
\]

where \( C \) ranges over all Weyl chambers.

We will see in Section 3.2 that \( s \nabla \infty(f(P_n)) \) is viewed as an analogue of the moment polytope.

**Proof.** (1). From \( \langle \nabla f(x), H \rangle_x = \text{tr} (d f(x) H) \), we have \( \nabla f(x) = x df(x) x \). Then \( \| \nabla(f + b_p)(x) \|_x = \| x^{-1/2} (x df(x)x - ub \text{diag} \, \lambda b^t u) x^{-1/2} \|_F = \| k^t (b^t u^t df(x) ub - \text{diag} \, \lambda) k \|_F = \| b^t u^t df(x) ub - \text{diag} \, \lambda \|_F. \)

(2). By (2.21), the inclusion \( s \nabla \infty(f(P_n)) \subseteq s \text{dom} \, f^* \subseteq s \mathcal{B}(f^\infty) \) is clear. Take \( p = \lambda \cdot [u] \in \mathcal{B}(f^\infty) \). By Theorem 2.24, there is a sequence \( (x_i) \) in \( P_n \) such that \( \lim_{i \to \infty} \| \nabla(f + b_p)(x_i) \|_{x_i} = 0 \). Suppose that \( \nabla \infty(f(x_i)) = \lambda_i \cdot [u_i] \). Via the decomposition \( u^t x_i^{1/2} = b_k h_i \) and \( u^t x_i^{1/2} = b_i^t h_i \), it holds \( x_i^{1/2} df(x)x_i^{1/2} = h_i \text{diag} \, \lambda_i h_i \), and \( s \nabla \infty(f(x_i)) = \lambda_i \). By the above calculation, we have \( \| \nabla(f + b_p)(x_i) \|_{x_i} = \| h_i^t \text{diag} \, \lambda_i h_i - k^t \|_F \to 0 \) \((i \to \infty)\). This implies that \( s \nabla \infty(f(x_i)) = \lambda_i \to \lambda \), and \( \lambda \in s \nabla \infty(f(P_n)) \).
Lemma 2.34. For two points \( p = \lambda \cdot U, \ q = \mu \cdot V \) in \( CP_n^\infty \), it holds
\[
\langle p, q \rangle = \sum_{1 \leq i,j \leq n} (\lambda_i - \lambda_{i+1})(\mu_j - \mu_{j+1}) \dim U_i \cap V_j.
\] (2.37)

Proof. It is well-known (see [11 II. 10.80]) that there are \( g \in GL(n, \mathbb{C}) \) and a permutation matrix \( \sigma \) such that \( U = [g] \) and \( V = [g\sigma] \); this is nothing but an axiom of building. In particular, both \( p = ge^\infty_\sigma \text{diag} \lambda g^\dagger \) and \( q = ge^\infty_\sigma \text{diag} \mu \sigma^\dagger g^\dagger \) belong to the apartment \( F(g)^\infty \). They are regarded as points in \( \mathbb{R}^n \): \( p = \lambda = \sum_i(\lambda_i - \lambda_{i+1})1_{[i]} \) and \( q = \sum_i(\mu_i - \mu_{i+1})1_{\sigma[i]} \), where \( 1_J \) denotes the \( n \)-dimensional 0,1-vector taking \( 1 \) only on indices in \( J \subseteq [n] \). Then \( \langle p, q \rangle = \sum_{i,j}(\lambda_i - \lambda_{i+1})(\mu_j - \mu_{j+1})|[i] \cap \sigma[j]| \). Here \( |[i] \cap \sigma[j]| = \dim U_i \cap V_j \). \( \square \)

Kapovich, Leeb, and Millson [31 Lemma 6.1] gives the corresponding formula of the angle of two vector subspaces regarded as points in \((SL(n, \mathbb{C})/SU(n))^\infty\).

Connections to submodular functions. Let \( S(\mathbb{C}^n) \) denote the family of all vector subspaces of \( \mathbb{C}^n \). A function \( \rho : S(\mathbb{C}^n) \to \mathbb{R} \cup \{\infty\} \) is called submodular if it satisfies
\[
\rho(X) + \rho(Y) \geq \rho(X \cap Y) + \rho(X + Y) \quad (X, Y \in S(\mathbb{C}^n)).
\] (2.38)

This extends the classical submodular functions, which are functions \( \kappa \) on \( 2^{[n]} \) satisfying \( \kappa(X) + \kappa(Y) \geq \kappa(X \cap Y) + \kappa(X \cup Y) \); see e.g., [21] [38] [43]. A submodular function \( \rho \) with \( \rho(\{0\}) = 0 \) gives rise to a positively homogeneous function \( \overline{\rho} : CP_n^\infty \to \mathbb{R} \cup \{\infty\} \) via piecewise linear extension
\[
\overline{\rho}(p) := \sum_{i=1}^n (\lambda_i - \lambda_{i+1})\rho(U_i) \quad (p = \lambda \cdot U \in CP_n^\infty).
\] (2.39)

This is an analogy of the Lovász extension [37] in the classical setting. So we call \( \overline{\rho} \) the Lovász extension of \( \rho \). This is equivalent to the one considered in [26] [27], where \( \lambda \) is restricted to \( \lambda \in [0,1]^n \).

Proposition 2.35 ([27 Theorem 3.9]). For a positively homogeneous function \( h : CP_n^\infty \to \mathbb{R} \cup \{\infty\} \), the following are equivalent:

(i) \( h \) is a convex function that is affine on each Weyl chamber.

(ii) \( h \) is the Lovász extension of a submodular function on \( S(\mathbb{C}^n) \).

The proof reduces to the classical convexity characterization [37] by restricting \( h \) to each apartment.

Lemma 2.36. Let \( q = \mu \cdot V \in CP_n^\infty \). Then \( p \mapsto -\langle q, p \rangle \) is the Lovász extension of submodular function
\[
X \mapsto -\langle q, X \rangle = -\sum_{i=1}^n (\mu_i - \mu_{i+1}) \dim V_i \cap X.
\] (2.40)
Proof. By Lemma 2.12, \( p \mapsto -\langle q, p \rangle \) is convex. Also it is an affine function on any apartment containing \( q \). Necessarily, it is affine on every Weyl chamber. Therefore, if suffices to show that (2.40) is submodular. We first show that \( X \mapsto -\dim V \cap X \) is submodular. This follows from \( \dim V \cap X + \dim V \cap Y = \dim V \cap X \cap Y + \dim (V \cap Y) \) and \( V \cap X + V \cap Y \subseteq V \cap (X + Y) \), we have submodularity of each summand in (2.40) with \( i \neq n \). For \( i = n \) (\( V_n = C \)), the equality \( \dim X \cap V_n + \dim X \cap V_n = \dim(X \cap Y) \cap V_n + \dim(X \cup Y) \cap V_n \) holds, and hence (2.40) is submodular (with taking zero on \( \{0\} \)). \( \square \)

For a submodular function \( \rho : \mathcal{S}(\mathbb{C}^n) \to \mathbb{R} \cup \{\infty\} \) with \( \rho(\{0\}) = 0 \), the subset \( B(\mathcal{P}) \) is described by fewer inequalities indexed by vector subspaces: It equals the set of points suffices to show that (2.40) is submodular. We first show that\( X \mapsto -\dim V \cap X \) is convex, where:

\[
\rho(X) = \dim V \cap X \leq \langle x, p \rangle \quad (X \in \mathcal{S}(\mathbb{C}^n)),
\]

Indeed, if \( p \) satisfies (2.41), then for \( q = \mu \cdot V \in CP_n^\infty \), it holds \( \langle q, p \rangle = \sum_i (\mu_i - \mu_{i+1}) \langle V_i, p \rangle \leq \sum_i (\mu_i - \mu_{i+1}) h(V_i) = h(q) \).

Then \( B(\mathcal{P}) \) is called the base polyhedron of \( \rho \), which is clearly an analogue of the classical one. A convex function \( f \) on \( P_n \) is said to be asymptotically submodular if \( f^\infty \) is the Lovász extension of a submodular function. In this case, the condition (c) in Theorem 3.15 can be replaced by (2.41) with \( \rho = f^\infty \). Accordingly, the convex optimization problem (2.29) becomes "discrete" convex optimization (submodular function minimization) over the lattice of vector subspaces:

\[
\inf \ f^\infty(X) - \langle p, X \rangle \quad \text{s.t.} \quad X \in \mathcal{S}(\mathbb{C}^n).
\]

## 3 Scaling problems

In this section, we explain how the results in the previous sections are applied to operator scaling and its generalizations. In our argument, the following convex function on \( \mathbb{R}^n \) plays important roles. This function appears in proving (semi)stability results (Kempf-Ness theorem, Hilbert-Mumford criterion) in invariant theory; see [32, 45].

**Lemma 3.1.** For \( a_i > 0 \) and \( w_i \in \mathbb{R}^n \) \((i = 1, 2, \ldots, m)\), define \( f : \mathbb{R}^n \to \mathbb{R} \) by

\[
f(x) := \log \sum_{i=1}^{m} a_i e^{\langle w_i, x \rangle} \quad (x \in \mathbb{R}^n).
\]

Then \( f \) is convex, where:

\[
f^\infty(p) = \max_{i=1,2,\ldots,m} \langle w_i, p \rangle \quad (p \in \mathbb{R}^n),
\]

\[
dom f^* = B(f^\infty) = \text{the convex hull of } w_1, w_2, \ldots, w_m.
\]

**Proof.** The convexity of \( f \) is well-known. Computing \( f^\infty \) is also a little exercise [40] p. 68: Letting \( v^* := \max_i \langle w_i, p \rangle \), we have \( f^\infty(p) = \lim_{t \to \infty} \frac{1}{t} \log \sum_{i=1}^{m} a_i e^{\langle w_i, p \rangle} = v^* + \lim_{t \to \infty} \frac{1}{t} \log \sum_{i=1}^{m} a_i e^{\langle w_i, p \rangle - v^*} = v^* \).

Thus \( B(f^\infty) \) equals the convex hull of \( w_i \). We show that it equals \( \text{dom } f^* \). Suppose \( p = \sum_i \lambda_i w_i \in B(f^\infty) \) with \( \lambda_i \geq 0 \) and \( \sum_i \lambda_i = 1 \). Then, for small \( c > 0 \), holds \( c\lambda_i \leq a_i \) for all \( i \). From \( \log \sum_{i=1}^{m} a_i e^{\langle w_i, x \rangle} \geq \log \sum_{i=1}^{m} c\lambda_i e^{\langle w_i, x \rangle} \geq c + c \log \sum_{i=1}^{m} \lambda_i e^{\langle w_i, x \rangle} = \log c + \langle p, x \rangle \), we have \( f(x) - \langle p, x \rangle \geq \log c \) for all \( x \in \mathbb{R}^n \). Hence \( B(f^\infty) \subseteq \text{dom } f^* \); the reverse inclusion is generally true. \( \square \)
3.1 Operator scaling with specified marginals

Let \( A = (A_1, A_2, \ldots, A_m) \) be an \( m \) tuple of nonzero complex \( n \times n \) matrices. Let \( \lambda, \mu \in \mathbb{R}^n \) be nonnegative arranged vectors with the same sum \( \sum_i \lambda_i = \sum_i \mu_i = n \) (say). The \emph{operator scaling problem with marginal} \( \lambda, \mu \), introduced by Franks \cite{19}, is to find a pair of nonsingular matrices \( g, h \in GL(n, \mathbb{C}) \) such that

\[
\sum_{k=1}^m g_A^k h_\lambda^k A_k^\dagger g = \text{diag} \lambda, \quad \sum_{k=1}^m h_A^k g_\mu^k A_k^\dagger h = \text{diag} \mu. \tag{3.2}
\]

If such \( g, h \) exist, then \( A \) is said to be \((\lambda, \mu)\)-\emph{scalable}. If for every \( \epsilon > 0 \) there are \( g, h \) such that

\[
\left\| \sum_{k=1}^m g_A^k h_\lambda^k A_k^\dagger g - \text{diag} \lambda \right\|_F < \epsilon, \quad \left\| \sum_{k=1}^m h_A^k g_\mu^k A_k^\dagger h - \text{diag} \mu \right\|_F < \epsilon, \tag{3.3}
\]

then \( A \) said to be \emph{approximately} \((\lambda, \mu)\)-\emph{scalable}. If \( \lambda = \mu = 1 \), it is the original operator scaling problem by Gurvits \cite{25}. In this case, the approximate scalability is equivalent to the \emph{noncommutative nonsingularity} of symbolic matrix \( \sum_k A_k x_k \) \cite{18, 30}. See \cite{19, 20, 22} for further applications of operator scaling.

For simplicity, we assume that at least one of \( \bigcap_{k=1}^m \ker A_k \) and \( \bigcap_{k=1}^m \ker A_k^\dagger \) is trivial \( \{0\} \). Otherwise, by coordinate change, we can make \( A \) satisfy \((A_k)_{in} = (A_k)_{nj} = 0 \) for \( i, j, k \). Then the problem reduces to the upper left \((n-1) \times (n-1)\) submatrices.

The operator scaling problem is viewed as the problem of finding a point \((x, y)\) in \( P_n \times P_n \) at which the following convex function \( f_A : P_n \times P_n \to \mathbb{R} \) has a specified asymptotic gradient:

\[
f_A(x, y) := n \log \sum_{k=1}^m \text{tr} x A_k y A_k^\dagger. \tag{3.4}
\]

This function is known to be \( \text{(geodesically)} \) convex.

**Lemma 3.2** \( \text{(See e.g.,} \cite{2}) \). \( f_A \) is \emph{convex}.

Indeed, on a maximal flat \( F = F(g) \times F(h) = \{ g e^{\text{diag } \alpha} g^\dagger \}_{\alpha \in \mathbb{R}^n} \times \{ h e^{\text{diag } \beta} h^\dagger \}_{\beta \in \mathbb{R}^n} \), \( f_A \) is written as

\[
f_A(g e^{\text{diag } \alpha} g^\dagger, h e^{\text{diag } \beta} h^\dagger) = n \log \sum_{1 \leq i, j \leq n} a_{ij}(g, h) e^{\alpha_i + \beta_j} \quad (\alpha, \beta \in \mathbb{R}^n), \tag{3.5}
\]

where \( a_{ij}(g, h) := \sum_{k=1}^m |(g A_k h)_{ij}|^2 \). By Lemma 3.1 \( f_A \) is convex in every flat.

In addition to the \((\lambda, \mu)\)-scalability, we consider a sharper scalability concept. Let \( \mathcal{U}, \mathcal{V} \) be complete flags, and consider points \((\lambda \cdot \mathcal{U}, \lambda \cdot \mathcal{V})\) in the boundary \( C(P_n \times P_n)^\infty = C P_n^\infty \times C P_n^\infty \). We say that \( A \) is \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable if there are \( g, h \in GL(n, \mathbb{C}) \) such that \([(g), [h]] = ([g], [h]) = ([\mathcal{U}], [\mathcal{V}]) \) and \( (3.2) \) hold. Accordingly, we say that \( A \) is \emph{approximately} \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable if for every \( \epsilon > 0 \) there are \( g, h \in GL(n, \mathbb{C}) \) such that \([(g), [h]] = ([\mathcal{U}], [\mathcal{V}]) \) and \( (3.3) \) hold. By definition, \( A \) is (approximately) \((\lambda, \mu)\)-scalable if and only if \( A \) is (approximately) \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable for some flags \( \mathcal{U}, \mathcal{V} \).

When \( \mathcal{U} \) and \( \mathcal{V} \) are standard flag \( \mathcal{E} := [[I]] \), scaling matrices \( g, h \) are upper triangular, and hence the (approximate) \((\lambda \cdot \mathcal{E}, \mu \cdot \mathcal{E})\)-scalability is equivalent to (approximate) \((\lambda, \mu)\)-scalability by triangular matrices in the sense of Franks \cite{19}. Note that the
(\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalability reduces to the triangular scalability, since \( A \) is \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable if and only if \( g^\dagger A h \) is \((\lambda \cdot \mathcal{E}, \mu \cdot \mathcal{E})\)-scalable for \([(g), [h]) = (\mathcal{U}, \mathcal{V})\).

The \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalability is rephrased by using asymptotic gradient \(\nabla^\infty\) and Busemann functions.

**Proposition 3.3.** (1) \( A \) is \((\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable if and only if there are points \(x, y \in \mathcal{P}_n\) such that \(\nabla^\infty f_A(x, y) = (\lambda \cdot \mathcal{U}, \mu \cdot \mathcal{V})\).

(2) \( A \) is approximately \((\lambda \mathcal{U}, \mu \cdot \mathcal{V})\)-scalable if and only if \(\inf_{x,y \in \mathcal{P}_n} \|\nabla (f_A + b \mathcal{U}, \mu \cdot \mathcal{V})(x, y)\|_{x,y} = 0\).

**Proof.** From \(d f_A(x, y)(H, G) = \frac{d}{dt} |_{t=0} n \log \sum_{k=1}^m \text{tr}(x + tH)A_k(x + tG)A_k^\dagger\), we have

\[ df_A(x, y) = C_{x,y} \left( \sum_k A_k y A_k^\dagger, \sum_k A_k^\dagger x A_k \right), \quad (3.6) \]

where \(C_{x,y} := n/\sum_{k=1}^m \text{tr} x A_k y A_k^\dagger\). Let \(\mathcal{U} = [u]\) and \(\mathcal{V} = [v]\) for \(u, v \in U(n)\). By Proposition 2.33 we have

\[ \|\nabla (f_A + b \mathcal{U}, \mu \cdot \mathcal{V})(x, y)\|_{x,y}^2 = \left\| C_{x,y} \left( \sum_k b_k u^\dagger A_k v c c^\dagger v^\dagger A_k^\dagger u - \text{diag} \lambda \right) \right\|_F^2 + \left\| C_{x,y} \left( \sum_k c^\dagger v^\dagger A_k^\dagger u b b^\dagger u^\dagger A_k v c - \text{diag} \mu \right) \right\|_F^2, \quad (3.7) \]

where \(u^\dagger x^\frac{1}{2} = bk\) and \(v^\dagger y^\frac{1}{2} = c k'\) for \(k, k' \in U(n)\) and upper-triangular matrices \(b, c\). From this, we have the claims, where required scaling matrices \(g, h\) are given as \(g = C_{x,y}^{1/4} u b\) and \(h = C_{x,y}^{1/4} v c\) with \([g] = [u] = \mathcal{U}\) and \([h] = [v] = \mathcal{V}\). \(\square\)

For the function \(f_A\), the inclusion \(\text{dom} f_A^* \subseteq \text{dom} f_A^\infty \subseteq B(f_A^\infty)\) becomes equality.

**Proposition 3.4.** \(\text{dom} f_A^* = B(f_A^\infty)\).

We will prove a general version (Proposition 3.14) in Section 3.2. Thus, the \((p, q)\)-scalability with \(p = \lambda \cdot \mathcal{U}\) and \(q = \mu \cdot \mathcal{V}\) can be decided by the boundeness of convex optimization:

\[ \inf \ (f_A + b_{p,q})(x, y) = f_A(x, y) + b_p(x) + b_q(y) \text{ s.t. } (x, y) \in \mathcal{P}_n \times \mathcal{P}_n, \]

where Busemann functions \(b_p\) and \(b_q\) are explicitly given by Lemma 2.32. By optimizing \(y\) under a fixed \(x \in \mathcal{P}_n\), we may minimize function \(g_{A,q} : P_n \to \mathbb{R}\):

\[ g_{A,q}(x) := \inf_{y \in \mathcal{P}_n} f_A(x, y) + b_q(y) \quad (x \in \mathcal{P}_n). \quad (3.8) \]

One can see from (3.6) and (3.7) that optimal \(y\) is obtained by \(y = h h^\dagger\) for \(h \in GL(n, \mathbb{C})\) with \(\mathcal{V} = [h]\) and \(h^\dagger (\sum_k A_k x A_k^\dagger) h = \text{diag} \mu\). When \(\mathcal{U} = \mathcal{V} = \mathcal{E}\), the infimum of \(f_A + b_p + b_q\) (or \(g_{A,q} + b_p\)) equals (up to constant) the logarithm of the capacity of specified marginal in Franks [19].

We compute explicit descriptions of the recession functions of \(f_A\) and the associated subset \(B(f_A^\infty)\). See Remark 3.3(1) for \(g_{A,q}\) and \(B(g_A^\infty)\). Let \(\mathcal{S}_A\) be the family of all pairs \((X, Y)\) of vector subspaces in \(\mathbb{C}^n\) such that \(u^\dagger A_k v = 0\) for all \(u \in X, v \in Y, k \in [m]\).

\(^1\)To see the consistency with his formulation, use the relation \(b_{\lambda, \mathcal{E}}(gg^\dagger) = -\log \det(\text{diag} \lambda, g^\dagger g)\) for any upper-triangular matrix \(g\), where \(\det(\text{diag} \lambda, g^\dagger g)\) is the relative determinant in the sense of [19].
Proposition 3.5. (1) The recession function \( f_A^\infty \) is given by
\[
f_A^\infty(p, q) = n \max\{\alpha_i + \beta_j \mid i, j \in [n]: (g^\dagger A_k h)_{ij} \neq 0 \ (\exists k \in [m])\},
\]
where \( p = \alpha \cdot [g], q = \beta \cdot [h] \in CP_n^\infty \).

(2) \( B(f_A^\infty) \) is the set of \((p, q) \in CP_n^\infty \times CP_n^\infty\) satisfying
\[
\langle X, p \rangle + \langle Y, q \rangle \leq n \quad ((X, Y) \in S_A).
\]

Proof. (1) follows from Lemma 3.1 and the expression (3.5). (2). Let \((p, q) \in CP_n^\infty \times CP_n^\infty\). For \((X, Y) \in S_A\), choose \(g, h \in GL(n, \mathbb{C})\) such that \(g\) and \(h\) span \(X\) and \(Y\) in the first \(k\) and \(l\) column subsets, respectively. Then, each \(g^\dagger A_k h\) has a \(k \times l\) zero block in the upper left corner. If \((X, Y)\) are viewed as points in \(CP_n^\infty\) by (2.31), then \(X = 1_{[k]} \cdot [g]\) and \(Y = 1_{[l]} \cdot [h]\). By the assumption that \(\bigcap_{k=1}^m \ker A_k = \{0\}\) or \(\bigcap_{k=1}^m \ker A_k^\dagger = \{0\}\), the maximum in (3.9) is attained by \(i \in [k], j \in [n] \setminus [l] \) or \(i \in [n] \setminus [k], j \in [l]\), and we have \(f_A^\infty(X, Y) = n\). Thus (3.10) is a necessary condition for \((p, q) \in B(f_A^\infty)\). Consider an apartment \(E(g) \times E(h) = F(g)^\infty \times F(h)^\infty\) containing \((p, q)\) and identify it with \(\mathbb{R}^n \times \mathbb{R}^n\) by \((ge^\infty \gamma^1, he^\infty \beta^1) \mapsto (\alpha, \beta)\). From Lemma 3.1 and (3.5), we have
\[
B_{E(g) \times E(h)}(f^\infty) = n \text{ the convex hull of } e_i + f_j \text{ for all } i, j \text{ with } a_{ij}(g, h) \neq 0,
\]
where \(e_i\) and \(f_i\) denote the \(i\)-th unit vectors of \(E(g)\) and \(E(h)\), respectively. This is \(n\) times the clique polytope of the bipartite graph \(G\) with vertex set \([n] \sqcup [n]\) and edge set \(\{ij \mid a_{ij}(g, h) \neq 0\}\); see [13, Section 65.4]. By a standard network flow argument, we obtain the inequality description of \(B_{E(g) \times E(h)}\) as
\[
\sum_{i \in S} \alpha_i + \sum_{j \in T} \beta_j \leq n \quad ((S, T) \subseteq [n] \times [n]: a_{ij}(g, h) = 0 \ (i \in S, j \in T)),
\]
where \(S \sqcup T\) is nothing but a stable set of the graph \(G\). Notice that (3.12) is the subsystem for (3.10) such that vector subspace \(X\) and \(Y\) are spanned by columns vectors of \(g\) and \(h\). By Proposition 2.27 satisfying all such inequalities is also sufficient for \((p, q) \in B(f_A^\infty)\).

Thus, by Theorem 2.24, Propositions 3.3, 3.4, 3.5, and Lemma 2.34, we have:

Theorem 3.6 ([19]). The following conditions are equivalent:

(a) \(A\) is approximately \((\lambda \cdot U, \mu \cdot V)\)-scalable.

(b) \(\inf_{x,y \in P_n} f_A(x, y) + b_{\lambda \cdot U}(x) + b_{\mu \cdot V}(y) > -\infty\).

(c) For all \((X, Y) \in S_A\), it holds
\[
\sum_{i=1}^n (\lambda_i - \lambda_{i+1}) \dim U_i \cap X + \sum_{i=1}^n (\mu_i - \mu_{i+1}) \dim V_i \cap Y \leq n.
\]

Franks [19] showed that the approximate scalability reduces to the triangular scalability in the generic case.
Theorem 3.7 (19). A is approximately \((\lambda, \mu)\)-scalable if and only if \(g^1Ah\) is approximately \((\lambda \cdot \mathcal{E}, \mu \cdot \mathcal{E})\)-scalable for generic \(g, h \in GL(n, \mathbb{C})\).

Here “generic” means that there is an affine variety \(V \subseteq GL(n, \mathbb{C})^2\) such that the latter property holds for all \((g, h) \in GL(n, \mathbb{C})^2 \setminus V\). We will verify this theorem for a general setting of the moment polytope membership in the next section.

Remark 3.8. \(1\) One can show that the recession function \(g^\infty_{A,q}\) of \(g_{A,q}\) is the Lovász extension of submodular function

\[
X \mapsto n - \langle X^\perp, q \rangle,
\]

where \(X^\perp\) denotes the maximum subspace \(Y\) with \((X, Y) \in \mathcal{S}_A\). In particular, \(g_{A,q}\) is asymptotically submodular, and \(B(g^\infty_{A,q})\) coincides with the base polyhedron of \(g^\infty_{A,q}\).

\(2\) Computation of the constant (BL-constant) of the Brascamp-Lieb inequality [10] is formulated as the same type of convex optimization over the product of PSD-cones (over \(\mathbb{R}\)) [23]. The objective function is also asymptotically submodular. A finiteness characterization of the BL-constant by [8] can be deduced by the same way as for (3.13) above.

\(3\) Since \((X, Y), (X', Y') \in \mathcal{S}_A\) implies \((X \cap X', Y + Y'), (X + X', Y \cap Y') \in \mathcal{S}_A\), the function \((X, Y) \mapsto -\langle X, p \rangle - \langle Y, q \rangle\) also admits a submodular function structure on the lattice \(\mathcal{S}_A\) with \(\wedge = (\cap, +), \vee = (+, \cap)\). Its Lovász extension (in the sense of [26, 27]) coincides with a part of \(f^\infty_{A,p,q}\). The nc-rank computation algorithm in [26] is interpreted as minimizing \(f^\infty_{A,p,q}\) (with \(\mu = \lambda = 1\)) over a convex neighborhood of 0.

3.2 Optimization on group orbits

The operator scaling and its generalizations (e.g., tensor scaling [13, 14]) can be formulated as optimization over an orbit of a group action. We finally consider the generalized scaling problems formulated by Bürgisser, Franks, Garg, Oliveira, Walter, and Wigderson [15]. Let \(G \subseteq GL(n, \mathbb{C})\) be a reductive algebraic group over \(\mathbb{C}\), i.e., \(G\) is defined by the zero set of a finite number of polynomials with complex coefficients, and \(g \in G\) implies \(g^1 \in G\). We assume that \(G\) is connected. Since \(G\) is a closed subgroup of Lie group \(GL(n, \mathbb{C})\), it is also a Lie group. Let \(K := G \cap U(n)\) be a maximal compact subgroup of \(G\). Let \(g\) and \(u\) denote the Lie algebras of \(G\) and \(K\), respectively, where \(g = u + iu\) is the complexification of \(u\) (or Cartan decomposition of involution \(X \mapsto -X^\dagger\)). This is a situation of [34, VII. 2. Example (2)].

Let \(\pi : G \to GL(N, \mathbb{C})\) be a rational representation, i.e., each entry of matrix \(\pi(g)\) is a polynomial of \(g_{ij}\) and \((\det g)^{-1}\). Let \(\langle \cdot, \cdot \rangle\) denote a \(K\)-invariant inner product on \(\mathbb{C}^N\), i.e., it satisfies \(\langle \pi(k)u, \pi(k)v \rangle = \langle u, v \rangle\) for all \(k \in K\). Let \(\Pi := d\pi(I)\) be the Lie algebra representation of \(\pi\). Then \(\pi(e^H) = e^{\Pi(H)}\) holds for \(H \in g\). The conjugate of \(g \in GL(N, \mathbb{C})\) with respect to \(\langle \cdot, \cdot \rangle\) is denoted by \(g^\dagger\) (the matrix satisfying \(\langle g^\dagger U, V \rangle = \langle U, gV \rangle\) for all \(U, V \in \mathbb{C}^N\)). Then \(\Pi(H^\dagger) = \Pi(H)\) holds for \(H \in g^\dagger\).

\footnote{From the \(\frac{\partial}{\partial \theta} \mid_{\theta=0} \langle \pi(e^{i\theta H_0})u, \pi(e^{i\theta H_0})v \rangle = 0\) for \(H_0 \in u\), we have \(\Pi(H_0)u, v + \langle u, \Pi(H_0)v \rangle = 0\). Thus \(\Pi(H_0)^\dagger = -\Pi(H_0)\). For \(H = H_0 + iH_1\) with \(H_0, H_1 \in u\), we have \(\Pi(H)^\dagger = \Pi(H_0)^\dagger - i\Pi(H_1)^\dagger = -\Pi(H_0) + i\Pi(H_1) = \Pi(-H_0 + iH_1) = \Pi(H^\dagger)\). Also}
\( \pi(g) \dagger = \pi(g) \dagger \) holds for \( g \in G \).

Given a vector \( v \in \mathbb{C}^N \), consider minimization of \( \log \| \pi(g)v \|^2 \) (twice of the Kempf-Ness function in [15]) over the \( G \)-orbit of \( v \):

\[
\inf \, \log \| \pi(g)v \|^2 \quad \text{s.t.} \quad g \in G.
\]  

(3.15)

This optimization can decide whether \( 0 \in \pi(G)v \) (the closure of orbit \( \pi(G)v \)), via the unboundedness. This is equivalent to the membership of \( v \) in the null-cone of the invariant ring of \( \pi \). The operator scaling in the previous section corresponds to the left-right action \( (g, h) \mapsto g \dagger Ah \), where \( f_A \) is a constant multiple of the Kempf-Ness function.

Since the norm is \( K \)-invariant, the optimization problem (3.15) is viewed as that the quotient space \( G/K \), which turns out to be a symmetric space of nonpositive curvature. We formulate the optimization problem (3.15) more explicitly, as in [15, Remark 3.4]. Note that \( \| \pi(g)v \|^2 = \langle v, \pi(g)g \rangle v \), where \( g \dagger g \in G \cap P_n \). Since \( G \) is algebraic, \( x \in G \cap P_n \) implies \( x^{1/2} \in G \cap P_n \); see [11, II.10.59]. Therefore (3.15) is also written as

\[
\inf \, f_\nu(x) := \log \langle v, \pi(x)v \rangle \quad \text{s.t.} \quad x \in M := G \cap P_n.
\]  

(3.16)

Here \( M = G \cap P_n \) is a totally geodesic subspace of \( P_n \), and hence is a symmetric space of nonpositive curvature (see [11, II. 10. 50]). By polar decomposition \( G = Ke^u \), we have \( M = e^{iu} \simeq G/K \), where \( iu \) is viewed as the tangent space at \( I \) with inner product \( \langle \lambda, \nu \rangle \mapsto \tr \lambda \nu \).

Then, \( M \) is decomposed as a Euclidean space and symmetric space of noncompact type as follows. Since \( G \) is reductive, the Lie algebra \( g \) is the direct sum of the center \( \mathfrak{z} \) and semisimple Lie algebra \( g_1 := [g, g] \), where \( \mathfrak{z} \) and \( g_1 \) are orthogonal in the inner product \( X, Y \mapsto \Re \tr(XY^\dagger) \); see [34, Proposition 1.59]. Now \( G \) is commuting product \( ZG_1 \) of the center \( Z \) for \( \mathfrak{z} \) and semisimple Lie group \( G_1 \) for \( g_1 \), where \( iu = \mathfrak{z} \cap iu + g_1 \cap iu \); see [34, Proposition 7.19 (e)]. Thus \( M = G \cap P_n \) is Riemannian product of Euclidean space \( \mathbb{R}^k \simeq e^{iu} \mathbb{R} \) and symmetric space \( M_1 := G_1 \cap P_n = e^{iu} \mathbb{R} \) of noncompact type. If \( g = zg_1 \) for \( z \in Z \) and \( g_1 \in G_1 \), then the action of \( g \) on \( M = \mathbb{R}^k \times M_1 \) is given so that \( g_1 \) acts on \( M_1 \) as \( x \mapsto g_1 x g_1^\dagger \) (as before) and \( z \) acts on \( \mathbb{R}^k \) as translation. Particularly, \( z \) acts trivially on the boundary \( CM = \mathbb{R}^k \times CM_1 \).

Maximal flats of \( M \) are the intersection of maximal flats of \( P_n \) with \( G \), and are given by \( e^{iu} \) for maximal commutative subspaces (maximal tori) \( t \) of \( u \). Fix a maximal torus \( t \) of \( u \). Then, any maximal flat is written as \( F(g) := \{ ge^{\lambda} g^\dagger \mid \lambda \in it \} \) for \( g \in G \), where \( \lambda \mapsto ge^{\lambda} g^\dagger \) is an isometry from Euclidean space \( it \) to \( F(g) \). The dimension \( d \) of \( it \) is equal to the rank of \( M \). Via \( \lambda \mapsto e^{i\lambda} \), we regard \( it \) as a subset, particularly, an apartment of \( CM \). Let \( it^+ \) be any fixed (asymptotic) Weyl chamber in \( it \), and let \( B \) denote the minimal parabolic subgroup (Borel subgroup) for \( it^+ \). Any point \( p \) in \( CM \) is written as \( \lambda \cdot F \) for \( \lambda \in it^+ \) and \( F \in G/B \).

Since \( M = G \cap P_n \), any Weyl chamber of \( M \) is written as the intersection of a Weyl chamber of \( P_n \) and a maximal flat of \( M \). Consequently, \( M^{\infty} \) is an isometric subspace of \( P_n^{\infty} \). By using the notation in Section 2.4 for some \( k_0 \in U(n) \), vectors \( \lambda \in it^+ \) are written as \( \lambda \cdot [k_0] \), where \( \lambda \) ranges over a subspace of arranged vectors.

It is known [15, 40] that the Kempf-Ness function \( f_\nu \) is convex on \( M \). Indeed, consider the expression of \( f_\nu \) in the maximal flat \( F(g) \). Since \( \{ e^{\lambda} \}_{\lambda \in H} \) is a commutative

---

\[ \pi(k^\dagger) = \pi(k) - \pi(k^-1) = \pi(k) \dagger = \pi(k) \dagger \] for \( k \in K \) and polar decomposition \( g = ke^{iH} \) for \( k \in K \) and \( H \in u \), we have

\[ \pi(g^\dagger) = e^{-\Pi(H)} \pi(k^\dagger) = e^{-\Pi(H)} \pi(k) \dagger = (\pi(k)e^{i\Pi(H)}) \dagger = \pi(g) \dagger. \]
subgroup, there is a finite set $\Omega(\pi)$ of vectors, called weights, in $\mathfrak{t}$ such that matrices $\pi(e^\lambda) = e^{\Pi(\lambda)}$ ($\lambda \in \mathfrak{t}$) are simultaneously diagonalized to a diagonal matrix of diagonals $e^{t^\omega e^\lambda}$ for $\omega \in \Omega(\pi)$. Therefore, we have

$$f_v(ge^\lambda g^\dagger) = \log \sum_{\omega \in \Omega(\pi)} \|(\pi(g^\dagger)v)_\omega\|^2 e^{t^\omega e^\lambda} \quad (\lambda \in \mathfrak{t}),$$

where $(\pi(g^\dagger)v)_\omega$ denotes the orthogonal projection of $\pi(g^\dagger)v$ to the eigenspace of $\omega$. By Lemma 3.1 and the expression (3.17), we have:

**Lemma 3.9.** $f_v$ is convex, where:

1. The recession function $f^\infty_v$ is given by
   $$f^\infty_v(p) = \max\{\text{tr} \omega \lambda \mid \omega \in \Omega(\pi) : (\pi(g^\dagger)v)_\omega \neq 0\},$$
   where $p = ge^\lambda g^\dagger$ for $g \in G$ and $\lambda \in \mathfrak{t}$.

2. $B_{CF(\pi)}(f^\infty_v)$ is the convex hull of $\omega$ over all $\omega \in \Omega(\pi)$ with $(\pi(g^\dagger)v)_\omega \neq 0$, where $\omega$ are viewed as points in $CF(\pi)$ by $\omega \mapsto ge^\omega g^\dagger$.

To study the boundedness of $f_v$, the following criterion is fundamental:

**Theorem 3.10** (Hilbert-Mumford criterion; see [45, Section 3.4.2]). If $\inf_{g \in G} \|\pi(g)v\| = 0$, then there is $u \in iu$ such that $\lim_{t \to \infty} \|\pi(e^{tu})v\| = 0$.

The reference [45, Theorem 3.23] also includes an elementary proof. As noticed in [31, 46], the nonnegativity of the asymptotic slope function of $f_v$ is equivalent to the Hilbert-Mumford criterion:

**Theorem 3.11** (see [31, 46]). The following conditions are equivalent:

1. $\inf_{x \in M} \|\nabla f_v(x)\| = 0$.
2. $\inf_{x \in M} f_v(x) > -\infty$.
3. $0 \in B(f^\infty_v)$.

The equivalence (a) $\iff$ (b) is known as the Kempf-Ness theorem [32], and is called the noncommutative duality in [13].

**Proof.** We have already seen (a) $\iff$ (c) and (b) $\Rightarrow$ (c) in general situation; see Lemma 2.14 and Theorem 2.24. We verify (c) $\Rightarrow$ (b). Suppose that $\inf_{x \in M} f_v(x) = -\infty$. By the Hilbert-Mumford criterion, there is $u \in iu$ such that $\lim_{t \to \infty} f_v(e^{tu}) = -\infty$. Consider a maximal flat $F$ containing geodesic $t \mapsto e^{tu}$. Then $f_v$ is unbounded on $F$. By Lemma 3.1, we have $0 \not\in B((f_v)_F) = B_{CF(\pi)}(f^\infty_v)$. By Proposition 2.27, we have $0 \not\in B(f^\infty_v)$. \qed

In particular, a one-parameter subgroup $t \mapsto e^{tu}$ in the Hilbert-Mumford criterion can be found by convex optimization of $f^\infty_v$ on Euclidean building $CM^\infty$:

$$\inf_u f^\infty_v(u) \quad \text{s.t.} \quad u \in U,$$

where $U$ is any convex neighborhood of the origin.
We are going to extend Theorem 3.11 for \( f_v + b_p \) with giving a whole description of \( B(f_v^\infty) \) and \( \text{dom} f_v^* \). For this, we need a representation theoretic interpretation of Busemann functions. By a weight we mean a point in \( i\mathfrak{t} \) that arises as a weight of some representation. It is known that the set of weights is a discrete subgroup (weight lattice) in \( i\mathfrak{t} \), and is generated by weights in \( i\mathfrak{t}^+ \). Any weight \( \lambda \) in \( i\mathfrak{t}^+ \) determines an irreducible representation \( \pi_\lambda \) of \( G \) such that \( \lambda \) is a highest weight. The eigenspace for \( \lambda \) is one dimensional, and the unit eigenvector is denoted by \( v_\lambda \).

**Lemma 3.12.** For a weight \( \lambda \), it holds \( b_{e^{-\lambda \infty}}(g^1g) = \log \|\pi_\lambda(g)v_\lambda\|^2 \) \( (g \in G) \).

**Proof.** Consider Iwasawa decomposition \( g = ke^{\nu}n \) for \( k \in K, \nu \in i\mathfrak{t}, n \in N \), where \( N \) is interpreted as the horospherical subgroup for \( i\mathfrak{t}^+ \) (see [17 Section 2.17]). From \( \pi_\lambda(n)v_\lambda = v_\lambda \) (see [31 Theorem 5.5]), the RHS equals \( 2 \text{tr} \nu \lambda \). On the other hand, \( n^\dagger \) is an element of the horospherical subgroup of the opposite Weyl chamber \(-i\mathfrak{t}^+\), since \( \lim_{n \to \infty} d(n^{-1}e^{-\lambda n^{-1}}, e^{-\lambda}) = \lim_{n \to \infty} d(ne^{\lambda n \dagger}, e^{\lambda}) = 0 \) for \( \lambda \in i\mathfrak{t}^+ \). Then the LHS equals \( b_{e^{-\lambda \infty}}(n^1 e^{2\nu n}) = b_{e^{-\lambda \infty}}(e^{2\nu}) = -2 \text{tr}(-\lambda \nu) \) (by Example 2.7).

Therefore, the minimization of \( f_v + b_p \) is essentially the \( p \)-scaling problem in [15]. A point \( \nu \) in \( i\mathfrak{t} \) is said to be rational if \( \alpha \nu \) is a weight for some positive integer \( \alpha \). Let \( s : CM^\infty \to i\mathfrak{t}^+ \) denote the projection \( \lambda \cdot \mathcal{F} \mapsto \lambda \).

**Proposition 3.13.** There is a finite set \( \Lambda \subseteq i\mathfrak{t}^+ \) (independent of \( v \)) such that

\[
B(f_v^\infty) = \{ p \in CM^\infty \mid \langle \nu \cdot \mathcal{F}, p \rangle \leq f_v^\infty(\nu \cdot \mathcal{F}) \; (\nu \in \Lambda, \mathcal{F} \in G/B) \}.
\]

(3.20)

For any Weyl chamber \( C \), the projection \( s(C \cap B(f_v^\infty)) \) is a rational convex polytope.

**Proof.** Take a Weyl chamber \( C \) of \( CP_n^\infty \), and consider all apartments \( E \) containing \( C \). When all \( E \) are regarded as \( \mathbb{R}^d \) with a common convex cone \( C \), by Proposition 2.27, \( C \cap B(f_v^\infty) \) is the intersection of \( C \) and finitely many (integral) polytopes \( B_E(f_v^\infty) \), which are convex hulls of finite subsets of weights in \( \Omega(\pi) \subseteq \mathbb{R}^n \). Consequently, \( s(C \cap B(f_v^\infty)) \) is a rational convex polytope.

In particular, the affine span of a facet of \( B_E(f_v^\infty) \) is spanned by a subset of \( \Omega(\pi) \), and its normal vector is chosen from it. The corresponding inequality is written as \( \langle \nu \cdot \mathcal{F}, p \rangle \leq f_v^\infty(\nu \cdot \mathcal{F}) \) for some \( \mathcal{F} \in G/B \) and \( \nu \in i\mathfrak{t}^+ \) (with \( \nu \cdot \mathcal{F} \in F \)). Thus, \( \Lambda \) can be chosen as a (finite) set of vectors arising as normal vectors of \( d-1 \)-spaces spanned by subsets of \( \Omega(\pi) \).

**Proposition 3.14.** \( \text{dom} f_v^* = B(f_v^\infty) \).

**Proof.** By rationality and convexity (Proposition 3.13), it suffices to show that for \( p = \lambda \cdot \mathcal{F} \in CM^\infty \) with rational \( \lambda \in i\mathfrak{t}^+ \) it holds \( \inf_{x \in M} (f_v + b_p)(x) > -\infty \) if and only if \( p \in B(f_v^\infty) \). Suppose that \( \lambda = \nu/\alpha \) where \( \nu \) is a weight and \( \alpha \) is a positive integer. By Lemma 3.12, \( b_{\nu,F} \) is the Kempf-Ness function for some representation \( \pi_\nu \) and vector \( \nu \cdot \mathcal{F} \). Then \( \alpha f_v + b_{\nu,F} \) is the Kempf-Ness function for representation \( \pi \otimes \pi \otimes \cdots \otimes \pi \otimes \pi \nu \cdot \mathcal{F} \) and vector \( v \otimes v \otimes \cdots \otimes v \otimes v' \); see [15 Section 3.6]. Therefore, by Theorem 3.11, we have \( \inf_{x \in M} (f_v + b_p)(x) = \inf_{x \in M} (1/\alpha)(\alpha f_v + b_{\nu,F})(x) > -\infty \iff 0 \in B((\alpha f_v + b_{\nu,F})^\infty) \iff 0 \in B(f_v^\infty + b_p^\infty) \iff p \in B(f_v^\infty) \).
Theorem 3.15. For $p \in CM^\infty$, the following conditions are equivalent:

(a) $\inf_{x \in M} \|\nabla (f_v + b_\mu) (x)\|_v = 0$.

(b) $-f^* (p) = \inf_{x \in M} (f_v + b_\mu) (x) > -\infty$.

(c) $p \in B(f_\mu^\infty)$.

We finally consider the moment polytope membership. Here, $\overline{s\nabla^\infty f_v} (M)$ is nothing but the moment polytope for $\pi, v$ in the sense of \footnote{\[15\].}

Lemma 3.16. $\overline{s\nabla^\infty f_v} (M) = \bigcup_{g \in G} \mathrm{it}^+ \cap B(f_{\pi(g)v}^\infty)$.

Proof. Proposition \footnote{2.33} holds in this setting by replacing diag $\lambda$ with $\lambda \in \mathrm{it}^+$ and $bk$ with Iwasawa decomposition $bk$ ($b \in B, k \in K$). By \footnote{2.36}, it suffices to show that $\mathrm{it}^+ \cap B(f_{\pi(g)v}^\infty) = s(C \cap B(f_v^\infty))$ for $g \in G$ and Weyl chamber $C = g^1 (\mathrm{it}^+)$. Indeed, from $f_{\pi(g)v}^\infty (u) = f_v^\infty (g^1 u)$ (by \footnote{3.18}), we have $\lambda \in \mathrm{it}^+ \cap B(f_{\pi(g)v}^\infty) \iff \langle u, \lambda \rangle \leq f_{\pi(g)v}^\infty (u)$ ($\forall u \in M^\infty$) $\iff \langle g^1 u, g^1 \mu \rangle \leq f_v^\infty (g^1 u)$ ($\forall u \in M^\infty$) $\iff \langle u', g^1 \lambda \rangle \leq f_v^\infty (u')$ ($\forall u' \in M^\infty$) $\iff g^1 \lambda \in B(f_v^\infty) \iff \lambda \in s(C \cap B(f_v^\infty))$, where $g^1 \lambda$ is written as $\lambda \cdot F$ for $F \in G/B$ corresponding to $C$.

The convexity of the moment polytope says:

Theorem 3.17 (Convexity theorem \footnote{[24]}). The moment polytope $\overline{s\nabla^\infty f_v} (M)$ is a rational convex polytope.

The shifting trick \footnote{\footnote{[12]} [39]} reduces the membership of the moment polytope to a single optimization problem.

Theorem 3.18 (Shifting trick \footnote{[12]} [39]). A rational vector $\lambda \in \mathrm{it}^+$ belongs to $\overline{s\nabla^\infty f_v} (M)$ if and only if $\inf_{x \in M} (f_{\pi(g)v} + b_{\mu, \infty}) (x) > -\infty$ for generic $g \in G$.

We prove a slightly stronger statement from our formulation, which implies Theorems \footnote{3.17} and \footnote{3.18}.

Theorem 3.19. $\overline{s\nabla^\infty f_v} (M) = \mathrm{it}^+ \cap B(f_{\pi(g)v}^\infty)$ for generic $g \in G$.

Our proof is a direct adaptation of \footnote{[19]} Lemma 54 and Proposition 55.\footnote{\footnote{\footnote{\footnote{\footnote{\footnote{\footnote{[15]}}.}}}}}

Proof. Let $\Lambda \subseteq \mathrm{it}^+$ be a finite set in Proposition \footnote{3.13} Let $g \in G$. Then $\mathrm{it}^+ \cap B(f_{\pi(g)v}^\infty)$ is the set of $\lambda \in \mathrm{it}^+$ satisfying

$$\langle \nu \cdot F, \lambda \rangle \leq f_{\pi(g)v}^\infty (\nu \cdot F) \quad (\nu \in \Lambda, F \in G/B).$$

We use the notation $\lambda = \tilde{\lambda} \cdot [k_0]$ to deduce an explicit inequality description. Represent $\nu \cdot F$ as $\nu \cdot F = \tilde{\nu} \cdot [g^{-1} h^1 k_0]$ for $h \in G$. Then $\langle \nu \cdot F, \lambda \rangle = \langle \tilde{\nu} \cdot [g^{-1} h^1 k_0], \tilde{\lambda} \cdot [k_0] \rangle = \langle \tilde{\nu} \cdot [k_0 g^{-1} h^1 k_0], \tilde{\lambda} \cdot [h^1] \rangle$, and $f_{\pi(g)v}^\infty (\nu \cdot F) = f_{\pi(g)v}^\infty (\tilde{\nu} \cdot [g^{-1} h^1 k_0]) = f_{\pi(h) v}^\infty (\nu)$, where
\[ I = \{ E_i \} \text{ is the standard flag. Let } \mathcal{U}_{g,h} = \{ U_{g,h} \} \text{ be the flag generated by } k_0^g h^h k_0. \]

By Lemma 2.34, (3.21) is written as

\[ \sum_{i,j=1}^n (\bar{\lambda}_i - \bar{\lambda}_{i+1})(\bar{v}_j - \bar{v}_{j+1}) \dim E_i \cap U_{g,h}^{i,j} \leq f_{\pi(h)\nu}^\infty(\nu) \quad (\nu \in \Lambda, h \in G). \quad (3.22) \]

We next consider the quantities \( f_{\pi(h)\nu}^\infty(\nu) \) and \( \dim E_i \cap U_{g,h}^{i,j} \) involving \( h \). By (3.18), the former quantity \( f_{\pi(h)\nu}^\infty(\nu) \) takes a value from finite set \( A_\nu := \{ \text{tr} \nu \omega \mid \omega \in \Omega(\pi) \} \).

For \( \alpha \in A_\nu \), let \( G_{\nu,\alpha} \subseteq G \) be the affine subvariety of consisting of \( h \) with \( f_{\pi(h)\nu}^\infty(\nu) \leq \alpha \), which is defined by algebraic conditions \( (\pi(h)\nu)_\omega = 0 \) for all \( \omega \in \Omega(\pi) \) with \( \text{tr} \nu \omega > \alpha \). The latter quantity \( \dim E_i \cap U_{g,h}^{i,j} \) takes a value in \( \{ 0, 1, 2, \ldots, n \} \). Let \( D \) denote the set of all \( n \times n \) matrices \( d = (d_{ij}) \) such that each entry \( d_{ij} \) is one of \( 0, 1, 2, \ldots, n \), where the partial order \( \leq \) on \( D \) is defined by \( d \leq d' \iff d_{ij} \leq d'_{ij} \quad (\forall i, j) \). For \( \nu \in \Lambda, \alpha \in A_\nu \), let \( D_{\nu,\alpha}(g) \subseteq D \) be the set of all \( d \) such that there is \( h \in G_{\nu,\alpha} \) such that \( d_{ij} = \dim E_i \cap U_{g,h}^{i,j} \) for \( 1 \leq i, j \leq n \), where \( d_{nj} = d_{jn} = j \) holds for all \( j \in [n] \). Let \( \bar{D}_{\nu,\alpha}(g) \) denote the set of maximal members with respect to \( \leq \). Then (3.22) is written as

\[ \sum_{i,j=1}^n (\bar{\lambda}_i - \bar{\lambda}_{i+1})(\bar{v}_j - \bar{v}_{j+1}) d_{ij} \leq \alpha \quad (\nu \in \Lambda, \alpha \in A_\nu, d \in \bar{D}_{\nu,\alpha}(g)). \quad (3.23) \]

Let \( S_{\nu,\alpha,d} \) be the subvariety of \( G \times G_{\nu,\alpha} \) consisting of \( g, h \) with \( \dim E_i \cap U_{g,h}^{i,j} \geq d_{ij} \) for \( i, j \), which is defined by vanishing of subdeterminants of \( k_0^g h^h k_0 \). Indeed, \( \dim E_i \cap U_{g,h}^{i,j} \) is \( j \) minus the rank of lower left \((n-i) \times j \) submatrix \( k_0^g h^h k_0 \). Let \( \pi \) be the projection \( (g, h) \mapsto g \). We claim:

\( \ast \) \( \pi(S_{\nu,\alpha,d}) \) is an affine subvariety of \( G \).

The proof is given in the end. Let \( D^*_{\nu,\alpha} \subseteq D \) be the set of all maximal \( d \) with \( \pi(S_{\nu,\alpha,d}) = G \). Consider the set \( Q^* \) of \( \lambda \in i^+ \) satisfying

\[ \sum_{i,j=1}^n (\bar{\lambda}_i - \bar{\lambda}_{i+1})(\bar{v}_j - \bar{v}_{j+1}) d_{ij} \leq \alpha \quad (\nu \in \Lambda, \alpha \in A_\nu, d \in D^*_{\nu,\alpha}). \quad (3.24) \]

For \( d \in D^*_{\nu,\alpha} \), there is \( d' \in D_{\nu,\alpha}(g) \) with \( d \leq d' \). That is, (3.24) is looser than (3.23). Thus, \( Q^* \) contains \( i^+ \cap B(f_{\pi(h)\nu}^\infty) \) for every \( g \in G \). Consider the finite union \( H := \bigcup_{\nu \in \Lambda,\alpha \in A_\nu} D^*_{\nu,\alpha} \pi(S_{\nu,\alpha,d}) \), which is a proper subvariety of \( G \). Then we can choose a generic \( g^* \in G \setminus H \). For such \( g^* \), it must hold \( \bar{D}_{\nu,\alpha}(g^*) = D^*_{\nu,\alpha} \) for all \( \nu \in \Lambda, \alpha \in A_\nu \). This means \( Q^* = i^+ \cap B(f_{\pi(h)\nu}^\infty) = \bigcup_{g \in H} i^+ \cap B(f_{\pi(h)\nu}^\infty) = \{ \text{tr} \nu \omega \mid \omega \in \Omega(\pi) \} \).

Finally we verify \( \ast \). By the closure theorem (see [16] Section 4.7, Theorem 7), \( \pi(G_{\nu,\alpha,d}) \) is a constructible set, i.e., it is an affine variety \( Z_0 \) minus an affine variety \( Z' \).

We show that \( \pi(G_{\nu,\alpha,d}) \) is a closed set in the Euclidean topology, which implies that \( \pi(G_{\nu,\alpha,d}) = Z_0 \) is an affine variety. Consider a sequence \( g_1, g_2, \ldots \in \pi(G_{\nu,\alpha,d}) \) converging to \( g \in G \). For each \( i \), there is \( h_k \in G \) with such that \( f_{\pi(h_k)\nu}^\infty(\nu) \leq \alpha \) and \( \dim E_i \cap U_{g,h_k}^{i,j} \geq d_{ij} \quad (i, j \in [n]) \). These quantities are determined by \( \nu \cdot [g^{-1} h_k^i k_0] = g^{-1} h_k^i \nu \). By Iwasawa decomposition, \( h_k \) can be chosen from the compact group \( K \). By taking a subsequence, we may assume that \( h_k \) converges to \( h \in K \). From \( h_k \in G_{\nu,\alpha} \) for each \( k \), it is clear that
As mentioned, the condition \( \dim E_i \cap U_{g_k h_k} \geq d_{ij} \) is written as vanishing of subdeterminants of \( k_0^{-1} h_k k_0 \). These subdeterminants vanish in the limit \( k_0^{-1} h_k k_0 \) as well. Then \( \dim E_i \cap U_{g_k h_k} \geq d_{ij} \). Thus, \((g, h) \in S_{\nu, \alpha, d}\), and \( \pi(S_{\nu, \alpha, d}) \) is closed.

In particular, \( B(f^\infty_v) \) contains the moment polytope in a “generic” chamber. Thus, the moment polytope membership for a given vector \( \lambda \in i \mathfrak{t}^+ \) also reduces, after taking generic \( g \in G \), to the convex optimization problem on Euclidean building \( CM^\infty \):

\[
\inf \, f^\infty_{\pi(g)v}(u) - \langle p, u \rangle \quad \text{s.t.} \quad u \in U, \tag{3.25}
\]

where \( p := e^{\infty \lambda} \in CM^\infty \) and \( U \) is any convex neighborhood of 0.

This gives rise to a challenging research problem to develop algorithms solving convex optimization problems \((3.19), (3.25)\). On a single Weyl chamber \( C \) (or an apartment), it is a usual Euclidean convex optimization. However, at a boundary point \( q \) of \( C \), one has to search a descent direction from infinitely many Weyl chambers containing \( q \). This seems impossible in principle. So one has to exploit and utilize special properties of the objective function, particularly, the recession function of the Kempf-Ness function, as in \([26]\). Moreover, to keep variable \( q = \lambda \cdot F \), one should keep basis vectors of flag \( F \) with bounded bit-length. This is also a highly nontrivial problem. A recent work \([20]\) for finding a violating vector subspace in \((3.13)\) may give hints toward this direction.

**Acknowledgments**

The author thanks Hiroyuki Ochiai for helpful discussion, and thanks for Zhiyuan Zhan for corrections. The author also thanks Harold Nieuwboer and Michael Walter for discussion on the Legendre-Fenchel duality. The work was partially supported by JST PRESTO Grant Number JPMJPR192A, Japan.

**References**

[1] P. Abramenko and K. S. Brown, *Buildings—Theory and Applications*. Springer, New York, 2008.

[2] Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, and A. Wigderson, Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. \[arXiv:1804.01076\], 2018, the conference version in STOC 2018.

[3] S. Amari and K. Nagaoka, *Methods of Information Geometry*. American Mathematical Society, Providence, RI, 2000.

[4] R. Bergmann, R. Herzog, M. S. Louzeiro, D. Tenbrinck, and J. Vidal-Núñez, Fenchel duality theory and a primal-dual algorithm on Riemannian Manifolds. *Foundations of Computational Mathematics* 2 (2021), 1465–1504, 2021.

[5] M. Bačák, *Convex Analysis and Optimization in Hadamard Spaces*. De Gruyter, Berlin, 2014.

[6] W. Ballmann, *Lectures on Spaces of Nonpositive Curvature*. Birkhäuser Verlag, Basel, 1995.
[7] W. Ballmann, M. Gromov, and V. Schroeder. *Manifolds of Nonpositive Curvature*. Birkhäuser, Boston MA, 1985.

[8] J. Bennett, A. Carbery, M. Christ, and T. Tao. The Brascamp-Lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis* 17 (2007) 1343–1415.

[9] N. Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, 2023.

[10] H. Brascamp and E. Lieb. Best constants in Young’s inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics* 20 (1976), 151–173.

[11] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-positive Curvature*. Springer-Verlag, Berlin, 1999.

[12] M. Brion. Sur l’image de l’application moment. In: M. -P. Malliavin (eds): *Séminaire d’algèbre Paul Dubreil et Marie-Paule Malliavin (Paris, 1986)*, Lecture Notes in Mathematics 1296, Springer, Berlin (1987), pp 177–192.

[13] P. Bürgisser, A. Garg, R. Oliveira, M. Walter, and A. Wigderson. Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory. *arXiv:1711.08039* 2017, the conference version in ITCS 2018.

[14] P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, and A. Wigderson. Efficient algorithms for tensor scaling, quantum marginals and moment polytopes. *arXiv:1804.04739* 2018, the conference version in FOCS 2018.

[15] P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, and A. Wigderson. Towards a theory of non-commutative optimization: geodesic first and second order methods for moment maps and polytopes. *arXiv:1910.12375* 2019, the conference version in FOCS 2019.

[16] D. A. Cox, J. Little, D. O’Shea. *Ideals, Varieties, and Algorithms, 4th edition*. Springer, Cham, 2015.

[17] P. B. Eberlein. *Geometry of Nonpositively Curved Manifolds*. University of Chicago Press, Chicago, IL, 1996.

[18] M. Fortin and C. Reutenauer. Commutative/non-commutative rank of linear matrices and subspaces of matrices of low rank. *Séminaire Lotharingien de Combinatoire* 52 (2004), B52f.

[19] C. Franks. Operator scaling with specified marginals. *arXiv:1801.01412* 2018, the conference version in STOC 2018.

[20] C. Franks, T. Soma, and M. X. Goemans. Shrunken subspaces via operator Sinkhorn iteration. *arXiv:2207.08311* 2022, the conference version in SODA 2023.

[21] S. Fujishige. *Submodular Functions and Optimization, 2nd Edition*. Elsevier, Amsterdam, 2005.

[22] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson. Operator scaling: theory and applications. *Foundations of Computational Mathematics* 20 (2020), 223–290.
[23] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via Operator Scaling. *Geometric and Functional Analysis* **28** (2018) 100–145.

[24] V. Guillemin and S. Sternberg, Convexity properties of the moment mapping, *Inventiones Mathematicae* **67** (1982) 491–513.

[25] L. Gurvits, Classical complexity and quantum entanglement, *Journal of Computer and System Sciences* **69** (2004), 448–484.

[26] M. Hamada and H. Hirai, Computing the nc-rank via discrete convex optimization on CAT(0) spaces, *SIAM Journal on Applied Geometry and Algebra* **5** (2021), 455–478.

[27] H. Hirai, L-convexity on graph structures. *Journal of the Operations Research Society of Japan* **61** (2018), 71–109.

[28] H. Hirai, H. Nieuwboer, and M. Walter, Interior-point methods on manifolds: theory and applications, [arXiv:2303.04771](https://arxiv.org/abs/2303.04771) 2023.

[29] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer-Verlag, Berlin, 2001.

[30] G. Ivanyos, Y. Qiao, and K. V. Subrahmanyam, Non-commutative Edmonds’ problem and matrix semi-invariants. *Computational Complexity* **26** (2017), 717–763.

[31] M. Kapovich, B. Leeb, and J. Millson, Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity. *Journal of Differential Geometry* **81** (2009), 297–354.

[32] G. Kempf and L. Ness, The length of vectors in representation spaces, In K. Lønsted (ed.) *Algebraic Geometry (Summer Meeting, Copenhagen, August 7–12, 1978)*, Lecture Notes in Mathematics **732**, Springer, Berlin, 1979, pp. 233–243.

[33] B. Kleiner and B. Leeb, Rigidity of invariant convex sets in symmetric spaces. *Inventiones Mathematicae* **163** (2006), 657–676.

[34] A. W. Knapp, *Lie Groups Beyond an Introduction, Second Edition*, Birkhäuser, Boston, 2002.

[35] E. Lieb, Gaussian kernels have only Gaussian maximizers. *Inventiones Mathematicae* **102** (1990), 179–208.

[36] M. S. Louzeiro, R. Bergmann, and R. Herzog, Fenchel duality and a separation theorem on Hadamard manifolds. *SIAM Journal on Optimization* **32** (2022), 854–873.

[37] L. Lovász, Submodular functions and convexity. In A. Bachem, M. Grötschel, and B. Korte (eds.): *Mathematical Programming—The State of the Art* (Springer-Verlag, Berlin, 1983), 235–257.

[38] K. Murota, *Discrete Convex Analysis*. SIAM, Philadelphia, 2004.

[39] L. Ness and D. Mumford, A stratification of the null cone via the moment map. *American Journal of Mathematics* **106** (1984), 1281–1329.

[40] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, NJ, 1970.
[41] U. G. Rothblum and H. Schneider, Scalings of matrices which have prespecified row sums and column sums via optimization. *Linear Algebra and Its Applications* **114/115** (1989), 737–764.

[42] T. Sakai, *Riemannian Geometry*. American Mathematical Society, Providence RI, 1996.

[43] A. Schrijver, *Combinatorial Optimization—Polyhedra and Efficiency*. Springer, Berlin, 2003.

[44] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematics Statistics* **35** (1964), 876–879.

[45] N. R. Wallach, *Geometric Invariant Theory*. Springer, Cham, 2017.

[46] C. Woodward, Moment maps and geometric invariant theory. [arXiv:0912.1132](https://arxiv.org/abs/0912.1132) 2009.