WASSERSTEIN GEOMETRY OF GAUSSIAN MEASURES

ASUKA TAKATSU

Abstract. This paper concerns the Riemannian/Alexandrov geometry of Gaussian measures, from the view point of the L^2-Wasserstein geometry. The space of Gaussian measures is of finite dimension, which allows to write down the explicit Riemannian metric which in turn induces the L^2-Wasserstein distance. Moreover, its completion as a metric space provides a complete picture of the singular behavior of the L^2-Wasserstein geometry. In particular, the singular set is stratified according to the dimension of the support of the Gaussian measures, providing an explicit nontrivial example of Alexandrov space with extremal sets.

1. Introduction

The universal importance of the Gaussian measures is evident in various fields. It sets a starting point of partial differential equations of parabolic type, and needless to say, it is central in the field of probability. The wide recognition of Wasserstein geometry defined on the space of Borel measures on a given metric space, on the other hand, is a more recent phenomenon, excited by some important advances made by Brenier [4] and McCann [16] in the 90's, and also implicit in Perelman's work [24] on Ricci flows (see also [13],[18],[29]).

This paper concerns the geometric structures of Gaussian measures, from the view point of the L^2-Wasserstein distance function W_2. The space of Gaussian measures is of finite dimension, which allows to write down the explicit Riemannian metric which in turn induces the L^2-Wasserstein distance. This situation is rare since the L^2-Wasserstein geometry is derived from the global data of the underlying space (in the case of Gaussian measures, $\mathbb{R}^d$), which makes it difficult to transcribe into the local geometry. The fact that the Wasserstein geometry involves some singular aspects is not surprising. Actually, a measure whose support has zero Lebesgue measure (for example a Dirac measure), which in turn causes singularities in the Wasserstein sense. Hence it is expected that the language of Alexandrov geometry is suited, and indeed much investigation has been done on the subject, which includes the recent works of Lott–Villani [14, 15] and Sturm [26, 27]. However the nature of singularity can be quite complicated, and

2000 *Mathematics Subject Classification.* 60D05,53C23.

Key words and phrases. Wasserstein space, Gaussian measures, Alexandrov space.

This work was partially supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

often the description of the complexity is incomplete. Restricting one's attention to the space of Gaussian measures once again provides a complete picture of the singular behavior of the L^2-Wasserstein geometry, due to the finite dimensional space of parameterization. In particular, the singular set is stratified according to the dimension of the support of the Gaussian measures, providing an explicit nontrivial example of Alexandrov space with extremal sets first introduced by Perelman–Petrunin [23]. Also one notes that the singularity is closely related to the dissipative behavior of the heat equation, which in this context can be regarded as a gradient flow of a suitable functional (see [2],[25]). In this sense, the space of Gaussian measures is a good testing ground for the L^2-Wasserstein geometry, where the fields of probability, partial differential equations, and Riemannian/Alexandrov geometry meet and interact.

The probability measure $N(m, V)$ on $\mathbb{R}^d$ is called the Gaussian measure with mean m and covariance matrix V if its density with respect to the Lebesgue density dx is given by

$$\frac{dN(m,V)}{dx} = \frac{1}{\sqrt{\det(2\pi V)}} \exp\left[-\frac{1}{2}\langle x-m, V^{-1}(x-m)\rangle\right],$$

where m is a vector in $\mathbb{R}^d$ and V is the symmetric positive definite matrix of size d. We denote the space of Gaussian measures on $\mathbb{R}^d$ by $\mathcal{N}^d$. Let $\mathrm{Sym}^+(d,\mathbb{R})$ be the space consisting of symmetric positive definite matrices of size d. The map

$$\mathcal{N}^d \to \mathbb{R}^d \times \mathrm{Sym}^+(d,\mathbb{R}), \quad N(m,V) \mapsto (m,V) \tag{1.1}$$

gives the differential structure on $\mathcal{N}^d$.

Otto [21] showed that the heat equation and the porous medium equation can be considered to be gradient flows in the L^2-Wasserstein space of probability measures on $\mathbb{R}^d$ equipped with the L^2-Wasserstein metric W_2. He introduced a formal Riemannian metric on the L^2-Wasserstein space. In addition, he obtained a formal expression of its sectional curvatures, which implies that the L^2-Wasserstein space has non-negative sectional curvature. (We refer to [12], where Lott computed the Levi-Civita connection and curvature on the L^2-Wasserstein space over a smooth compact Riemannian manifold. See also Sturm's result [26], which states that an underlying space is an Alexandrov space of non-negative curvature if and only if so is its L^2-Wasserstein space.)

McCann [16] showed that $\mathcal{N}^d$ is a totally geodesic submanifold in the L^2-Wasserstein space. Then it is natural to expect that $\mathcal{N}^d$ admits a Riemannian metric whose Riemannian distance coincides with the L^2-Wasserstein distance. We call such a Riemannian metric L^2-Wasserstein metric. We confirm in Section 2 that $(\mathcal{N}^d, W_2)$ is a product metric space of the Euclidean space $\mathbb{R}^d$ and $(\mathcal{N}_0^d, W_2)$, where $\mathcal{N}_0^d$ stands for the space of Gaussian measures with mean 0. Since the geometry of the Euclidean space $\mathbb{R}^d$ is trivial, we fix a mean $m = 0$ and analyze the geometry $\mathcal{N}_0^d$. We shall denote $N(0, V)$ simply by $N(V)$. According to the differential structure on $\mathcal{N}_0^d$, the tangent space of $\mathcal{N}_0^d$ at each point can be identified with the space $\mathrm{Sym}(d,\mathbb{R})$ of all symmetric matrices of size d. However,

the natural coordinate (1.1) is somewhat unadapted to the L^2-Wasserstein geometry. We give the adapted one and analyze the value of the L^2-Wasserstein metric in the coordinate.

Proposition A. *On the space of Gaussian measures, the Riemannian metric g given by*

$$g_{N(V)}(X,Y) = \mathrm{tr}(XVY)$$

for any tangent vectors X, Y in $T_{N(V)}\mathcal{N}_0^d = \mathrm{Sym}(d,\mathbb{R})$ induces the L^2-Wasserstein distance.

We mention that the L^2-Wasserstein metric is different from the Fisher metric. For example, for $d = 1$, the space of Gaussian measures with the Fisher metric can be regarded as an upper half plane with the hyperbolic metric (see [1]). Meanwhile, the space of Gaussian measures with the L^2-Wasserstein metric has non-negative sectional curvature. This follows from the following formula for the sectional curvatures on $\mathcal{N}_0^d$ with the L^2-Wasserstein metric. Namely, we verify Otto's formula by restricting to a finite dimensional, totally geodesic submanifold $\mathcal{N}_0^d$ of the L^2-Wasserstein space. We denote the transpose of a matrix A by ${}^{\mathsf{T}}A$. Let $[\cdot,\cdot]$ be the bracket product, that is, $[X,Y] = XY - YX$ for matrices X, Y. Tangent vectors X, Y at $N(V)$ are said to be linearly independent if

$$q_{N(V)}(X,Y) = g_{N(V)}(X,X)g_{N(V)}(Y,Y) - g_{N(V)}(X,Y)^2 \neq 0.$$

Theorem B. *Let g be the L^2-Wasserstein metric on $\mathcal{N}_0^d$, expressed by*

$$g_{N(V)}(X,Y) = \mathrm{tr}(XVY)$$

for any tangent vectors X, Y in $T_{N(V)}\mathcal{N}_0^d = \mathrm{Sym}(d,\mathbb{R})$. For linearly independent tangent vectors X, Y at $N(V)$, the sectional curvature $K_{N(V)}(X,Y)$ is given by

$$K_{N(V)}(X,Y)q_{N(V)}(X,Y) = \frac{3}{4}\mathrm{tr}\left(([Y,X]-S)V\,{}^{\mathsf{T}}([Y,X]-S)\right),$$

where the matrix S is a symmetric matrix so that $([Y,X]-S)V$ is anti-symmetric.

The formulae of sectional curvatures on the L^2-Wasserstein space which have been derived in [12],[21] are quite difficult to compute. However our formula of sectional curvatures on $\mathcal{N}^d$ is easily computable in terms of matrices. We moreover show that $\mathcal{N}^d$ has a cone structure. To prove it, we construct the completion of $\mathcal{N}_0^d$ as a metric space, denoted by $\overline{\mathcal{N}_0^d}$, in Section 4. ($\mathcal{N}_0^d$ is not complete with respect to the L^2-Wasserstein distance.) The space $\overline{\mathcal{N}_0^d}$ is identified with the set $\mathrm{Sym}(d,\mathbb{R})_{\geq 0}$ which consists of symmetric non-negative definite matrices of size d. Hence elements of $\overline{\mathcal{N}_0^d}$ can be denoted by $N(V)$ for some V in $\mathrm{Sym}(d,\mathbb{R})_{\geq 0}$. An element of $\overline{\mathcal{N}_0^d} \setminus \mathcal{N}_0^d$ is called the degenerate Gaussian measure. For example, the Dirac measure is one of degenerate Gaussian measures.

We see that in Section 5 the space $\overline{\mathcal{N}_0^d}$ is a finite dimensional Alexandrov space of non-negative curvature. We shall study the stratification and the tangent cone

of $\overline{\mathcal{N}_0^d}$. For definitions and further details, we refer to [22] for stratifications and [5] for tangent cones.

Theorem C. *For each $0 \leq k \leq d$, we set*

$$S(d,k) = \left\{ V \in \mathrm{Sym}(d,\mathbb{R}) \;\middle|\; \begin{array}{c} k \text{ eigenvalues are positive,} \\ (d-k) \text{ eigenvalues are } 0 \end{array} \right\},$$
$$\mathcal{S}_k = \{N(V) \mid V \in S(d,k)\}.$$

Then the set $\{\mathcal{S}_k\}_{k=0}^d$ forms a finite stratification of $\overline{\mathcal{N}_0^d}$ into topological manifolds.

This gives an explicit nontrivial example of extremal subset in the context of Gaussian measures, since closures of strata are extremal (see [22],[23]). The definition of extremal subset is first given in [23] for subsets of compact Alexandrov spaces of finite dimension. Petrunin [25] later showed that the compactness assumption is unnecessary.

Let $\{e_i\}_{i=1}^d$ be the canonical basis of $\mathbb{R}^d$ where e_i is the d-tuple consisting of all zeros except for a 1 in the ith spot. We denote the space of orthogonal matrices of size d by $O(d)$. For P in $O(d)$, we set the subspace $S^P(d,k)$ of $\mathrm{Sym}(d,\mathbb{R})$ as

$$S^P(d,k) = \left\{ X \in \mathrm{Sym}(d,\mathbb{R}) \;\middle|\; \langle a^i P e_i, X a^j P e_j \rangle \geq 0, \text{ for numbers } \{a^i\}_{i=k+1}^d \right\}.$$

In other words, $S^P(d,k)$ can be identified with the set of symmetric positive semidefinite bilinear forms on a $(d-k)$-dimensional subspace of $\mathbb{R}^d$ spanned by $\{Pe_i\}_{i=k+1}^d$. (In this paper, we use the Einstein summation convention.)

A stratification gives a homeomorphism between a sufficiently small neighborhood of a point and its tangent cone (see [22],[23]). Hence Theorem C implies the following corollary.

Corollary D. *For $N(V)$ in $\mathcal{S}_k$, we assume that V is decomposed as*

$$V = P\mathrm{diag}[\lambda^1, \ldots, \lambda^k, 0, \ldots, 0]^{\mathsf{T}} P,$$

where P is an orthogonal matrix and $\{\lambda^i\}_{i=1}^k$ are positive numbers. Then a tangent cone at $N(V)$ is homeomorphic to $S^P(d,k)$.

In a special case, we determine a distance function on the tangent cones. First we consider the case of points in $\mathcal{N}_0^d$. For a point of $\mathcal{N}_0^d$, its tangent cone coincides with its tangent space of $\mathcal{N}_0^d$ with the L^2-Wasserstein metric.

Proposition E. *For $N(V)$ in $\mathcal{N}_0^d$, the tangent cone at $N(V)$ is isometric to $(\mathrm{Sym}(d,\mathbb{R}), \mathsf{d}_V)$, where d_V is given by*

$$\mathsf{d}_V(X,Y) = \sqrt{\mathrm{tr}((X-Y)V(X-Y))}$$

for any X, Y in $\mathrm{Sym}(d,\mathbb{R})$.

We next consider the tangent cone at the Dirac measure $N(0)$ centered at the origin 0 in $\mathbb{R}^d$. The key fact is that dilations of covariance matrices induces dilations in the L^2-Wasserstein space.

Theorem F. *The tangent cone at $N(0)$ is isometric to $(\overline{\mathcal{N}_0^d}, W_2)$.*

Theorem F implies that $\overline{\mathcal{N}_0^d}$ has a cone structure. More generally, Yokota and the author [28] have shown that if an underlying space has a cone structure, then so does its L^2-Wasserstein space.

In the case of $d = 2$, we acquire the following equivalence;

$$X = \begin{pmatrix} x & z \\ z & y \end{pmatrix} \in S^E(2,1) \Longleftrightarrow (x, y, z) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R},$$

where E is the identity matrix. It means that $S^E(2,1)$ is homeomorphic to the upper half-space $\mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$.

Theorem G. *For a positive number λ, the tangent cone at $N(\mathrm{diag}[\lambda^2, 0])$ is isometric to the Euclidean upper half-space $\mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$.*

The organization of the paper is as follows: After introducing the L^2-Wasserstein geometry, we state preceding results of the L^2-Wasserstein geometry on $\mathcal{N}^d$ in Section 2. Then we prove Proposition A and Theorem B in Section 3. The principal objective in Section 4 is the completion of $\mathcal{N}_0^d$. We shall consider properties of $\overline{\mathcal{N}_0^d}$ as an Alexandrov space in Section 5. We first investigate the stratification, which is stated in Theorem C. The last part of Section 5 deals with the tangent cones.

Acknowledgement . This paper contains a part of the author's doctoral dissertation at Tohoku University, and she wishes here to thank her supervisor Sumio Yamada for his support and encouragement. She would also like to express her gratitude to Takashi Shioya and Shin-ichi Ohta for many valuable suggestions and discussions. She thanks the anonymous referees for valuable remarks which lead to several crucial revisions.

2. L^2-Wasserstein spaces

In this section, we shall review the L^2-Wasserstein space. It is a pair of a subset of probability measures on a complete separable metric space and a distance function derived from the Monge–Kantorovich transport problem. See [30] and [31] for general theory.

Given a complete separable metric space (M, d_M), let μ and ν be Borel probability measures on M. The set of Borel probability measures μ on M satisfying

$$\int_M \mathsf{d}_M(x,y)^2 d\mu(y) < \infty$$

for some x in M will be denoted by $\mathcal{P}_2(M)$. A transport plan π between μ and ν is a Borel probability measure on $M \times M$ with marginals μ and ν, that is,

$$\pi[B \times M] = \mu[B], \quad \pi[M \times B] = \nu[B]$$

for all measurable sets B in M. Then the L^2-Wasserstein distance between μ and ν in $\mathcal{P}_2(M)$ is defined by

$$W_2(\mu,\nu)=\left(\inf_{\pi}\int_{M\times M}\mathsf{d}_M(x,y)^2d\pi(x,y)\right)^{\frac{1}{2}}.$$

Here the infimum is taken over all the transport plans π between μ and ν. Then W_2 is a distance function on $\mathcal{P}_2(M)$ (see [31, Chapter 6] for further details). We call the pair $(\mathcal{P}_2(M),W_2)$ L^2-Wasserstein space over M. A transport plan is optimal if it achieves the infimum. Optimal transport plans on the Euclidean spaces are characterized by the push forward measures.

Let F be a measurable map on $\mathbb{R}^d$. For a Borel probability measure μ on $\mathbb{R}^d$, the push forward measure $F_\sharp\mu$ through F on $\mathbb{R}^d$ is defined by $F_\sharp\mu[B]=\mu[F^{-1}(B)]$ for all measurable sets B in $\mathbb{R}^d$. We denote the identity map on $\mathbb{R}^d$ by id.

Theorem 2.1 ([4],[9])**.** *For μ, ν in $\mathcal{P}_2(\mathbb{R}^d)$, we assume that μ is absolutely continuous with respect to the Lebesgue measure. Then we obtain the following properties;*

(1) *there exists a convex function ψ whose gradient $\nabla\psi$ pushes μ to ν.*
(2) *this gradient is uniquely determined (μ-almost everywhere).*
(3) *the joint measure $[\mathrm{id}\times\nabla\psi]_\sharp\mu$ is optimal.*
(4) *For t in $[0,1]$, $[(1-t)\mathrm{id}+t\nabla\psi]_\sharp\mu$ is a geodesic from μ to ν.*

In this paper, a geodesic is always assumed to be minimizing with constant speed. We also refer to [17], where McCann extended this theorem to a case of the connected compact Riemannian manifold without boundary in place of $\mathbb{R}^d$.

Though Theorem 2.1 characterizes optimal transport plans, it is usually difficult to obtain optimal transport plans and concrete values of the L^2-Wasserstein distance between two given Borel probability measures. However, the L^2-Wasserstein distance between Gaussian measures was explicitly computed by several authors; Dowson-Landau [7], Givens-Short [10], Knott-Smith [11] and Olkin-Pukelsheim [19]. For X in $\mathrm{Sym}^+(d,\mathbb{R})$, we define a symmetric positive definite matrix $\sqrt{X}=X^{1/2}$ so that $X^{1/2}\cdot X^{1/2}=X$.

Theorem 2.2 ([7], [10], [11], [19])**.** *The L^2-Wasserstein distance between Gaussian measures $N(m,V)$ and $N(n,U)$ is given by*

$$W_2(N(m,V),N(n,U))^2=|m-n|^2+\mathrm{tr}V+\mathrm{tr}U-2\mathrm{tr}\sqrt{U^{\frac{1}{2}}VU^{\frac{1}{2}}}. \tag{2.1}$$

As observed from this formula, $\mathcal{N}^d$ is isometric to the product metric space $\mathbb{R}^d\times\mathcal{N}_0^d$. Therefore we fix mean $m=0$ and we consider the geometry on $\mathcal{N}_0^d$. McCann [16] demonstrated that $\mathcal{N}_0^d$ is a totally geodesic submanifold of the L^2-Wasserstein space.

Lemma 2.3. ([16, Example 1.7]) *For $N(V)$ and $N(U)$, we define a symmetric positive definite matrix T and its associated linear map $\mathcal{T}$ by*

$$T=U^{\frac{1}{2}}(U^{\frac{1}{2}}VU^{\frac{1}{2}})^{-\frac{1}{2}}U^{\frac{1}{2}},\quad \mathcal{T}(x)=Tx.$$

Then, $\mathcal{T}$ pushes $N(V)$ forward to $N(U)$ and $[\mathrm{id}\times\mathcal{T}]_\sharp N(V)$ is an optimal transport plan between $N(V)$ and $N(U)$. We moreover define a matrix $W(t)$ by

$$W(t) = [(1-t)E + tT]V[(1-t)E + tT]$$

for $t \in [0,1]$. Then, $\{N(W(t))\}_{t\in[0,1]}$ is a geodesic from $N(V)$ to $N(U)$.

We call the above matrix T (unique) linear transform between $N(V)$ and $N(U)$. For $X \in \mathrm{Sym}(d,\mathbb{R})$ and $V \in \mathrm{Sym}^+(d,\mathbb{R})$, we set

$$V(t) = (E + tX)V(E + tX) \tag{2.2}$$

for sufficiently small t so that $V(t)$ is positive definite. Then $N(V(t))$ is well-defined and Lemma 2.3 guarantees that $N(V(t))$ is a geodesic in $(\mathcal{N}_0^d, W_2)$.

3. Properties of $\mathcal{N}^d$ as a Riemannian manifold

We shall prove Proposition A and Theorem B. In the end of this section, we refer to Otto's result [21]. For general theory of Riemannian geometry, see [8].

Let $\mathcal{G} = \mathrm{Gl}(d,\mathbb{R})$ denote the set of invertible matrices of size d. Then the tangent space of $\mathcal{G}$ at each point is identified with the set $\mathrm{M}(d,\mathbb{R})$ of all square matrices of size d. Define $G : \mathrm{M}(d,\mathbb{R}) \times \mathrm{M}(d,\mathbb{R}) \to \mathbb{R}$ by

$$G(Z,W) = \mathrm{tr}({}^{\mathsf{T}}ZW),$$

then G is obviously a flat Riemannian metric on $\mathcal{G}$.

We define a map $\Pi : \mathcal{G} \to \mathcal{N}_0^d$ by $\Pi(A) = N({}^{\mathsf{T}}AA)$, then Π is surjective and the differential map $d\Pi_A$ of Π at A is given by

$$d\Pi_A(Z) = {}^{\mathsf{T}}ZA + {}^{\mathsf{T}}AZ.$$

For A in $\Pi^{-1}(N(V))$, $\mathcal{V}_A$ is a subspace of $T_A\mathcal{G} = \mathrm{M}(d,\mathbb{R})$ consisting of vectors which tangent to fibers $\Pi^{-1}(N(V))$. In other words,

$$Z \in \mathcal{V}_A \iff {}^{\mathsf{T}}Z = -A^{-1}{}^{\mathsf{T}}ZA. \quad ({}^{\mathsf{T}}ZA \text{ is anti-symmetric.}) \tag{3.1}$$

Let $\mathcal{H}_A$ be the subspace of $T_A\mathcal{G} = \mathrm{M}(d,\mathbb{R})$ consisting of vectors which are normal to fibers $\Pi^{-1}(N(V))$, that is, Z satisfies $G(X,Y) = 0$ for any W in $\mathcal{V}_A$. Then we obtain the following equivalent condition;

$$Z \in \mathcal{H}_A \iff {}^{\mathsf{T}}Z = {}^{\mathsf{T}}AZA^{-1}. \quad (ZA^{-1} \text{ is symmetric.}) \tag{3.2}$$

For Z in $T_A\mathcal{G} = \mathrm{M}(d,\mathbb{R})$, $Z_\mathcal{V}$ and $Z_\mathcal{H}$ stand for the orthogonal projections of Z onto $\mathcal{V}_A$ and $\mathcal{H}_A$, respectively. We call elements of $\mathcal{V}_A$ and $\mathcal{H}_A$ vertical and horizontal vectors, respectively.

We define a Riemannian metric g on $\mathcal{N}_0^d$ by

$$g_\Pi(d\Pi(Z), d\Pi(W)) = G(Z_\mathcal{H}, W_\mathcal{H}).$$

Then Π is a Riemannian submersion.

Under a Riemannian submersion, horizontal geodesics are mapped to geodesics.

Proposition 3.1. ([8, Proposition 2.109]) *Let $\Pi : (\widetilde{M}, \widetilde{g}) \to (M, g)$ be a Riemannian submersion. For a geodesic $\widetilde{c}(t)$ in $(\widetilde{M}, \widetilde{g})$, if the vector $\widetilde{c}'(0)$ is horizontal, then $\Pi \circ \widetilde{c}$ is also geodesic of (M, g).*

Using this proposition, we construct a geodesic from $N(V)$ to $N(U)$ in $(\mathcal{N}_0^d, g)$ and prove Proposition A. To do this, let us summarize the correspondence between $T_{\Pi(A)}\mathcal{N}_0^d$ and $\mathrm{Sym}(d, \mathbb{R})$ due to (3.2), which is different from that due to the natural coordinate (1.1). The horizontal part $\mathcal{H}_A$ is canonically identified with $T_{\Pi(A)}\mathcal{N}_0^d$ and then we associate $Z \in \mathcal{H}_A$ with $ZA^{-1} \in \mathrm{Sym}(d, \mathbb{R})$. In this coordinate, we have $d\Pi_A(Z) = ZA^{-1}$ for $Z \in \mathcal{H}_A$, in other words,

$$Z \in \mathcal{H}_{V^{1/2}} \text{ and } d\Pi_{V^{1/2}}(Z) = X \Longleftrightarrow X = ZV^{-\frac{1}{2}} \in \mathrm{Sym}(d, \mathbb{R}), \tag{3.3}$$

and the geodesic $N(V(t))$ with $V(0) = 0$ and initial velocity $V'(0) = X$ coincides with the geodesic (2.2) (as we will verify below). Moreover, (3.3) rewrites g as

$$g_{N(V)}(X, Y) = G(XV^{1/2}, YV^{1/2}) = \mathrm{tr}(XVY). \tag{3.4}$$

In what follows, we use this identification (3.3) for simplicity.

Proposition A. *On the space of Gaussian measures, the Riemannian metric g given by*

$$g_{N(V)}(X, Y) = \mathrm{tr}(XVY)$$

for any tangent vectors X, Y in $T_{N(V)}\mathcal{N}_0^d = \mathrm{Sym}(d, \mathbb{R})$ induces the L^2-Wasserstein distance.

Proof. For V, U in $\mathrm{Sym}^+(d, \mathbb{R})$, we define

$$A = V^{\frac{1}{2}}, \quad Z = [U^{\frac{1}{2}}(U^{\frac{1}{2}}VU^{\frac{1}{2}})^{-\frac{1}{2}}U^{\frac{1}{2}} - E]V^{\frac{1}{2}},$$

then Z lies in $\mathcal{H}_A$ due to (3.2). We set a curve $\{Z(t)\}_{t\in[0,1]}$ in $\mathcal{G}$ as

$$Z(t) = A + tZ.$$

Since $(\mathcal{G}, G)$ is flat, $\{Z(t)\}_{t\in[0,1]}$ is a geodesic whose initial velocity $Z'(0)$ is the horizontal vector Z at $Z(0) = A$. Proposition 3.1 yields that $\{\Pi \circ Z(t)\}_{t\in[0,1]}$ is a geodesic in $(\mathcal{N}_0^d, g)$ from $N(V) = \Pi(Z(0))$ to $N(U) = \Pi(Z(1))$. By definition of Riemannian distance, we obtain the following equalities;

$$\begin{aligned}
W_2(N(V), N(U))^2 &= \mathrm{tr}V + \mathrm{tr}U - 2\mathrm{tr}\sqrt{U^{\frac{1}{2}}VU^{\frac{1}{2}}} \\
&= \mathrm{tr}(Z^{\mathsf{T}}Z) \\
&= G(Z, Z) \\
&= g_\Pi(d\Pi(Z), d\Pi(Z))(A) \\
&= \mathsf{d}_g(\Pi(Z(0)), \Pi(Z(1)))^2 \\
&= \mathsf{d}_g(N(V), N(U))^2.
\end{aligned}$$

It implies that the Riemannian distance d_g of g coincides with the L^2-Wasserstein distance W_2. □

We now give an expression of sectional curvatures for the Riemannian manifold.

Theorem B. *Let g be the L^2-Wasserstein metric on $\mathcal{N}_0^d$, expressed by*

$$g_{N(V)}(X,Y) = \mathrm{tr}(XVY)$$

for any tangent vectors X, Y in $T_{N(V)}\mathcal{N}_0^d = \mathrm{Sym}(d,\mathbb{R})$. For linearly independent tangent vectors X, Y at $N(V)$, the sectional curvature $K_{N(V)}(X,Y)$ is given by

$$K_{N(V)}(X,Y)q_{N(V)}(X,Y) = \frac{3}{4}\mathrm{tr}\left(([Y,X]-S)V^{\intercal}([Y,X]-S)\right),$$

where the matrix S is a symmetric matrix so that $([Y,X]-S)V$ is anti-symmetric.

Proof. Let $\mathbf{Z}, \mathbf{W}$ be horizontal vector fields on $\mathcal{G}$ satisfying

$$Q(\mathbf{Z},\mathbf{W}) = G(\mathbf{Z},\mathbf{Z})G(\mathbf{W},\mathbf{W}) - G(\mathbf{Z},\mathbf{W})^2 \neq 0.$$

Then O'Neill's formula [20] yields

$$K(d\Pi(\mathbf{Z}), d\Pi(\mathbf{W}))Q(\mathbf{Z},\mathbf{W}) = \frac{3}{4}G([\mathbf{Z},\mathbf{W}]_{\mathcal{V}}, [\mathbf{Z},\mathbf{W}]_{\mathcal{V}}). \tag{3.5}$$

For any X and Y in $\mathrm{Sym}(d,\mathbb{R})$, there exist unique horizontal vectors Z and W at $V^{1/2}$ so that $d\Pi_{V^{1/2}}(Z) = X$ and $d\Pi_{V^{1/2}}(W) = Y$. By (3.3), we get

$$X = ZV^{-\frac{1}{2}}, \quad Y = WV^{-\frac{1}{2}}.$$

We define vector fields $\mathbf{Z}$ and $\mathbf{W}$ on $\mathcal{G}$ by

$$\mathbf{Z}(A) = ZV^{-\frac{1}{2}}A = XA, \quad \mathbf{W}(A) = WV^{-\frac{1}{2}}A = YA$$

for any A in $\mathcal{G}$. Due to (3.2), these vector fields are horizontal and we obtain

$$d\Pi(\mathbf{Z})(V^{\frac{1}{2}}) = X, \quad d\Pi(\mathbf{W})(V^{\frac{1}{2}}) = Y, \quad Q(\mathbf{Z},\mathbf{W})(V^{\frac{1}{2}}) = q_{N(V)}(X,Y). \tag{3.6}$$

In the (global) standard coordinate functions $\{x_{ij}\}_{1\leq i,j\leq d}$ of $\mathcal{G}$, these vector fields are expressed by

$$\mathbf{Z} = \left(X^{ik}x_{kj}\right)\frac{\partial}{\partial x_{ij}}, \quad \mathbf{W} = \left(Y^{ik}x_{kj}\right)\frac{\partial}{\partial x_{ij}},$$

where X^{ij} and Y^{ij} stand for the (i,j)-components of X and Y, respectively. Therefore we have

$$[\mathbf{Z},\mathbf{W}] = \left(X^{lk}x_{kj}Y^{il} - Y^{lk}x_{kj}X^{il}\right)\frac{\partial}{\partial x_{ij}}.$$

It implies that

$$[\mathbf{Z},\mathbf{W}](A) = [Y,X]\cdot A$$

for any A in $\mathcal{G}$. Let $S = d\Pi([\mathbf{Z},\mathbf{W}])(V^{1/2})$ in $\mathrm{Sym}(d,\mathbb{R})$. Then we obtain $S = [\mathbf{Z},\mathbf{W}]_{\mathcal{H}}(V^{\frac{1}{2}})\cdot V^{-1/2}$ by (3.3) and

$$[\mathbf{Z},\mathbf{W}]_{\mathcal{V}}(V^{\frac{1}{2}}) = \left([\mathbf{Z},\mathbf{W}] - [\mathbf{Z},\mathbf{W}]_{\mathcal{H}}\right)(V^{\frac{1}{2}}) = ([Y,X]-S)\cdot V^{\frac{1}{2}}.$$

The property (3.1) of vertical vectors implies that $([Y,X]-S)\,V$ is anti-symmetric. The norm of the vertical vector $[\mathbf{Z},\mathbf{W}]_{\mathcal{V}}$ at $V^{1/2}$ is as follows;

$$G([\mathbf{Z},\mathbf{W}]_{\mathcal{V}},[\mathbf{Z},\mathbf{W}]_{\mathcal{V}})(V^{\frac{1}{2}}) = G(([Y,X]-S)\cdot V^{\frac{1}{2}},([Y,X]-S)\cdot V^{\frac{1}{2}}) \\ = \operatorname{tr}\left(([Y,X]-S)V\,{}^{\mathsf{T}}([Y,X]-S)\right). \tag{3.7}$$

Substituting (3.6) and (3.7) into the O'Neill formula (3.5), we acquire the formula in Theorem B. □

Remark 3.2. Otto [21] constructed a "Riemannian" submersion π from $\mathcal{D}$ onto $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ by using push forward measure with a base measure μ, that is, $\pi(F) = F_\sharp\mu$. Here $\mathcal{D}$ denotes the set of all diffeomorphisms of $\mathbb{R}^d$ and $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ denotes the the subspace of $\mathcal{P}_2(\mathbb{R}^d)$ consisting of absolutely continuous measures with respect to the Lebesgue measure, respectively. The infinite dimensionality of $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mathcal{D}$ makes the result formal, however restrictions into totally geodesic finite dimensional subspaces are rigorous. Therefore we choose the space of Gaussian measures as the totally geodesic Riemannian submanifold and Theorem B rewrote Otto's expression in a simple explicit form on $\mathcal{N}_0^d$. See also [12], where Lott gave an expression of the sectional curvature on the L^2-Wasserstein spaces over smooth compact Riemannian manifolds.

4. Metric completion of $\mathcal{N}_0^d$

In this section, we construct the completion of $\mathcal{N}_0^d$ in terms of characteristic functions (more detailed treatment of this topic can be found for instance in [3]).

The characteristic function φ_μ of a probability measure μ on $\mathbb{R}^d$ is defined as the Fourier transform of the probability measure μ:

$$\varphi_\mu(\xi) = \int_{\mathbb{R}^d} \exp\left[\sqrt{-1}\langle x,\xi\rangle\right] d\mu(x).$$

Then the characteristic function of a Gaussian measure $N(m,V)$ is given by

$$\exp\left[\sqrt{-1}\langle m,\xi\rangle - \frac{1}{2}\langle \xi, V\xi\rangle\right].$$

It is well-known that the weak convergence of probability measures is equivalent to the pointwise convergence of their characteristic functions. An element of the completion of $\mathcal{N}_0^d$ in the sense of weak convergence is called the degenerate Gaussian measure. Its characteristic function is given by

$$\exp\left[-\frac{1}{2}\langle \xi, V\xi\rangle\ \right],$$

where V is a symmetric non-negative definite matrix.

Remark 4.1. The characteristic function of the Dirac measure δ_m centered at m in $\mathbb{R}^d$ is given by $\exp[\sqrt{-1}\langle m,\xi\rangle]$.

On the other hand, the convergence in the sense of L^2-Wasserstein space implies the weak convergence (see [31, Theorem 6.9]). Therefore the completion of $\mathcal{N}_0^d$ in the sense of weak convergence contains the one in the sense of L^2-Wasserstein space, and after simple calculations we conclude that the both spaces coincide. Thus it is natural to denote any element of $\overline{\mathcal{N}_0^d}$ by $N(V)$ for some V in $\mathrm{Sym}(d,\mathbb{R})_{\geq 0}$.

Remark 4.2. We can generalize the distance formula (2.1) defined on $\mathcal{N}_0^d$ to $\overline{\mathcal{N}_0^d}$ because we can define $V^{1/2}$ for any V in $\mathrm{Sym}(d,\mathbb{R})_{\geq 0}$. Namely, $V^{1/2}$ is a symmetric non-negative definite matrix such that $V^{1/2}\cdot V^{1/2}=V$. Therefore we obtain

$$W_2(N(V),N(U))^2=\mathrm{tr}V+\mathrm{tr}U-2\mathrm{tr}\sqrt{U^{\frac{1}{2}}VU^{\frac{1}{2}}},$$

for $N(V)$ and $N(U)$ in $\overline{\mathcal{N}_0^d}$.

5. Properties of $\overline{\mathcal{N}_0^d}$ as an Alexandrov space

5.1. **Alexandrov space.** Let us summarize some definitions and basic results on the geometry of Alexandrov spaces. Standard references are [5] and [6].

Let (M,d_M) be a complete geodesic metric space, that is, each pair of points x, y is connected by a geodesic. For κ in $\mathbb{R}$, we denote a simply-connected, 2-dimensional Riemannian manifold of constant sectional curvature κ by $\mathbb{M}^2(\kappa)$. Let d_κ be a Riemannian distance of $\mathbb{M}^2(\kappa)$. For any three points x,y,z in M (provided that $[\mathsf{d}_M(x,y)+\mathsf{d}_M(y,z)+\mathsf{d}_M(z,x)]^2<4\pi^2/\kappa$ if $\kappa>0$), there exist corresponding points $\tilde{x},\tilde{y},\tilde{z}$ in $\mathbb{M}^2(\kappa)$ which are unique up to an isometry so that $\mathsf{d}_M(x,y)=\mathsf{d}_\kappa(\tilde{x},\tilde{y})$, $\mathsf{d}_M(y,z)=\mathsf{d}_\kappa(\tilde{y},\tilde{z})$ and $\mathsf{d}_M(z,x)=\mathsf{d}_\kappa(\tilde{z},\tilde{x})$. We also denote the unique geodesic from $\tilde{x}$ to $\tilde{y}$ by $\gamma_{\tilde{x}\tilde{y}}:[0,1]\to\mathbb{M}^2(\kappa)$.

Definition 5.1. Let (M,d_M) be a complete geodesic metric space. Given any κ in $\mathbb{R}$, we say that (M,d_M) is an Alexandrov space of curvature $\geq\kappa$ if for any three points x,y,z in M (provided that $[\mathsf{d}_M(x,y)+\mathsf{d}_M(y,z)+\mathsf{d}_M(z,x)]^2<4\pi^2/\kappa$ if $\kappa>0$), any geodesic $\gamma:[0,1]\to M$ from y to z and any t in $[0,1]$, we have

$$\mathsf{d}_M(x,\gamma(t))\geq\mathsf{d}_\kappa(\tilde{x},\gamma_{\tilde{y}\tilde{z}}(t)).$$

The triangle $\triangle\tilde{x}\tilde{y}\tilde{z}$ is called the comparison triangle of $\triangle xyz$. We next define a comparison angle.

Definition 5.2. Let x,y,z be three distinct points in a geodesic space (M,d_M). A comparison angle of $\angle xyz$, denoted by $\tilde{\angle}xyz$, is defined by

$$\tilde{\angle}xyz=\arccos\frac{\mathsf{d}_M(x,y)^2+\mathsf{d}_M(y,z)^2-\mathsf{d}_M(z,x)^2}{2\mathsf{d}_M(x,y)\mathsf{d}_M(y,z)}.$$

In a similar way, we define an angle between two paths starting at the same point.

Definition 5.3. Let γ and σ be two paths in a geodesic space starting at the same point p. We define an angle $\angle_p(\gamma,\sigma)$ between γ and σ as

$$\angle_p(\gamma,\sigma)=\lim_{s,t\searrow 0}\tilde{\angle}\gamma(s)p\sigma(t),$$

if the limit exists.

In the case of geodesics in Alexandrov spaces, the limit always exists and does not depend on the choices of s, t. For any two geodesics γ, σ with unit speed starting at the same point, the angle between γ and σ is equal to 0 if and only if there exists a positive number ε such that $\gamma(t) = \sigma(t)$ for all t in $[0, \varepsilon]$. Using these facts, we briefly discuss the tangent cone of M. Let (M, d_M) be an Alexandrov space of curvature $\geq \kappa$. Fix a point p in M. We define Σ'_p as the set of geodesics starting at p equipped with an equivalence relation $\parallel$, where $\gamma \parallel \sigma$ holds if $\angle_p(\gamma, \sigma) = 0$. The angle $\angle_p$ is independent of the choices of γ and σ in their equivalence classes. Then $\angle_p$ is a natural distance function on Σ'_p. We define the space of directions $(\Sigma_p, \angle_p)$ at p as the completion of $(\Sigma'_p, \angle_p)$ and the tangent cone (K_p, d_p) as a cone over $(\Sigma_p, \angle_p)$.

Definition 5.4. Let (M, d_M) be a metric space. The cone over M is a quotient space $M \times [0, \infty)/\sim$, where an equivalence relation $\sim$ is defined by $(x, s) \sim (y, t)$ if and only if $s = t = 0$. We call the equivalence class of $(\cdot, 0)$ vertex. The distance d_C on the cone is defined by

$$\mathsf{d}_C((x,s),(y,t)) = \sqrt{s^2 + t^2 - 2st\cos\min\{\mathsf{d}_M(x,y), \pi\}}.$$

We denote a vertex of a tangent cone K_p at p by o_p. For each point p in a finite dimensional Alexandrov space of non-negative curvature, there exists a homeomorphism from a neighborhood of p to the tangent cone at p sending p to o_p. Here we mean by the dimension the Hausdorff dimension. Tangent cones and stratification are useful when we analyze local structures of finite dimensional Alexandrov spaces. Roughly speaking, every finite dimensional Alexandrov space is stratified into topological manifolds.

Definition 5.5. A collection $\{\mathcal{S}_i\}_{i=0}^N$ of subsets of a topological space M forms a (finite) stratification of M into topological manifolds if

(1) the sets $\mathcal{S}_i$ are mutually disjoint, and $\bigcup_{i=0}^N \mathcal{S}_i = M$.
(2) every set $\mathcal{S}_i$ is a topological manifold.
(3) $\dim \mathcal{S}_0 < \dim \mathcal{S}_1 < \cdots < \dim \mathcal{S}_N$.
(4) for every $k = 1, \ldots, N$, the closure of $\mathcal{S}_k$ is contained in $\mathcal{S}_k^+ = \bigcup_{i=0}^k \mathcal{S}_i$.

Sturm [26] proved that if an underlying space is an Alexandrov space of non-negative curvature, then so is its L^2-Wasserstein space. It yields that $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is an Alexandrov space of non-negative curvature. Since $\overline{\mathcal{N}_0^d}$ is a complete, totally geodesic subspace of $\mathcal{P}_2(\mathbb{R}^d)$, $\overline{\mathcal{N}_0^d}$ provides an explicit nontrivial example of Alexandrov space of non-negative curvature. We consider properties of $\overline{\mathcal{N}_0^d}$ as an Alexandrov space.

5.2. **Stratification.** We first give an expression of a stratification of $\overline{\mathcal{N}_0^d}$.

Theorem C. *For each $0 \leq k \leq d$, we set*

$$S(d,k) = \left\{ V \in \mathrm{Sym}(d,\mathbb{R}) \;\middle|\; \begin{array}{c} k \text{ eigenvalues are positive,} \\ (d-k) \text{ eigenvalues are } 0 \end{array} \right\},$$
$$\mathcal{S}_k = \{N(V) \mid V \in S(d,k)\}.$$

Then the set $\{\mathcal{S}_k\}_{k=0}^d$ forms a finite stratification of $\overline{\mathcal{N}_0^d}$ into topological manifolds.

Proof. We check that $\{\mathcal{S}_k\}_{k=0}^d$ satisfies the properties in Definition 5.5. Property (1) follows from the definition of $\mathcal{S}_k$. To show Properties (2),(3), we construct a homeomorphism from $\mathcal{S}_k$ to a topological manifold. Let $G(d,k)$ be the Grassmannian manifold of k-dimensional subspaces of $\mathbb{R}^d$. Define a set $\mathcal{M}(d,k)$ by

$$\mathcal{M}(d,k) = \{(G_k, V_k) | V_k \text{ is an inner product on } G_k \in G(d,k)\},$$

which is homeomorphic to a product topological manifold $G(d,k) \times \mathrm{Sym}^+(k,\mathbb{R})$.

For V in $S(d,k)$, there exist an orthogonal matrix P and positive numbers $\{\lambda^i\}_{i=1}^k$ so that

$$V = P\mathrm{diag}[\lambda^1, \ldots, \lambda^k, 0, \ldots, 0]\,{}^{\mathsf{T}}P.$$

We set $G_k(V)$ as a k-dimensional subspace of $\mathbb{R}^d$ spanned by $\{Pe_i\}_{i=1}^k$. Then $G_k(V)$ is independent of the choice of P and a map $V \mapsto G_k(V)$ is well-defined.

Let $f_k : \mathcal{S}_k \to \mathcal{M}(d,k)$ be defined by

$$f_k(N(V)) = (G_k(V), V|_k),$$

where $V|_k$ is a restriction of V to $G_k(V)$. We verify that the map f_k has an inverse map. For any G_k in $G(d,k)$, there exists an orthogonal k-frame $\{p_j\}_{j=1}^k$ spanning G_k. Then a map $g_k : \mathcal{M}(d,k) \to \mathcal{S}_k$ given by

$$g_k(G_k, V_k) = N({}^{\mathsf{T}}PV_kP), \quad {}^{\mathsf{T}}P = (p_1, \ldots, p_k)$$

is well-defined. It is clear that $g_k \circ f_k = \mathrm{id}_{\mathcal{S}_k}$ and $f_k \circ g_k = \mathrm{id}_{\mathcal{M}(d,k)}$. Thus g_k is the inverse map of f_k. The continuity of each component of covariance matrix is equivalent to the continuity on $\overline{\mathcal{N}_0^d}$ with respect to the L^2-Wasserstein distance. Therefore f_k and g_k are continuous and hence f_k is a homeomorphism. In turn, $\mathcal{S}_k$ is a topological manifold of dimension $k(d-k) + 2^{-1}k(k+1)$, which is the dimension of $\mathcal{M}(d,k)$. It guarantees Properties (2),(3).

Finally we confirm Property (4). In the same way as the completion of $\mathcal{N}_0^d$, the closure of $\mathcal{S}_k$ is given by

$$\{N(V) \in \mathrm{Sym}^+(d,\mathbb{R}) \mid \text{At least } (d-k) \text{ out of eigenvalues of } V \text{ are } 0\},$$

which is just $\mathcal{S}_k^+$.

Thus $\{\mathcal{S}_k\}_{k=0}^d$ forms the finite stratification of $\overline{\mathcal{N}_0^d}$ into topological manifolds. □

5.3. **Tangent cone.** In this section, we discuss the tangent cones of $(\overline{\mathcal{N}_0^d}, W_2)$. Let V be a symmetric matrix decomposed as

$$V = P\mathrm{diag}[\lambda^1, \dots, \lambda^k, 0, \dots, 0]\,{}^{\mathsf{T}}P$$

where P is an orthogonal matrix and $\{\lambda^i\}_{i=1}^k$ are positive numbers. We set $G^{d-k}(V)$ as a $(d-k)$-dimensional subspace of $\mathbb{R}^d$ spanned by $\{Pe_i\}_{i=k+1}^d$. Then a tangent cone at $N(V)$ is homeomorphic to the set of symmetric bilinear forms which are positive semidefinite on $G^{d-k}(V)$, that is

$$S^P(d,k) = \left\{ X \in \mathrm{Sym}(d,\mathbb{R}) \,\middle|\, \langle a^i Pe_i, Xa^j Pe_j\rangle \geq 0, \text{ for numbers } \{a^i\}_{i=k+1}^d \right\}.$$

Corollary D. *For $N(V)$ in $\mathcal{S}_k$, we assume that V is decomposed as*

$$V = P\mathrm{diag}[\lambda^1, \dots, \lambda^k, 0, \dots, 0]\,{}^{\mathsf{T}}P,$$

where P is an orthogonal matrix and $\{\lambda^i\}_{i=1}^k$ are positive numbers. Then a tangent cone at $N(V)$ is homeomorphic to $S^P(d,k)$.

Proof. A bijective map

$$\overline{\mathcal{N}_0^d} \to \mathrm{Sym}(d,\mathbb{R})_{\geq 0}, \quad N(X) \mapsto X^{1/2}$$

can be considered as coordinate functions due to (2.2). Let $N(V(t))$ be a geodesic starting at $N(V)$. Since any geodesic has a linear approximation, we obtain that $V^{1/2}(t) = V^{1/2}(0) + t(V^{1/2})'(0) + o(t)$ for sufficiently small t. Thus we can identify $(V^{1/2})'(0)$ with a point sufficiently close to $N(V)$. The non-negativity of $V(t)$ yields the non-negativity of $(V^{1/2})'(0)$ on $G^{d-k}(V)$. Indeed, we have

$$\left\langle a^i Pe_i, \left(V^{1/2}(t) - V^{1/2}(0)\right) a^j Pe_j \right\rangle = \left\langle a^i Pe_i, V^{1/2}(t) a^j Pe_j \right\rangle \geq 0$$

for any numbers $\{a^i\}_{i=k+1}^d$. Therefore we identify a small enough neighborhood of $N(V)$ with $S^P(d,k)$. On the other hand, Theorem 5.7 implies that $K_{N(V)}$ is homeomorphic to a neighborhood. Thus the tangent cone $K_{N(V)}$ is homeomorphic to $S^P(d,k)$. □

In a special case, we determine a distance function on the tangent cones. First we consider the case of points in $\mathcal{N}_0^d$.

Proposition E. *For $N(V)$ in $\mathcal{N}_0^d$, the tangent cone at $N(V)$ is isometric to $(\mathrm{Sym}(d,\mathbb{R}), \mathsf{d}_V)$, where d_V is given by*

$$\mathsf{d}_V(X,Y) = \sqrt{\mathrm{tr}((X-Y)V(X-Y))}$$

for any X, Y in $\mathrm{Sym}(d,\mathbb{R})$.

Proof. Because $\mathcal{N}_0^d$ are locally Euclidean, tangent cones of $(\overline{\mathcal{N}_0^d}, W_2)$ coincide with tangent spaces of $\mathcal{N}_0^d$ with the L^2-Wasserstein metric for points in $\mathcal{N}_0^d$ (see [5]). Thus the tangent cone is identified with $\mathrm{Sym}(d,\mathbb{R})$, and (3.4) implies

$$\mathsf{d}_V(X,Y)^2 = g_{N(V)}(X-Y, X-Y) = \mathrm{tr}((X-Y)V(X-Y)).$$

□

We next construct a distance function on the tangent cone at the Dirac measure $N(0)$ centered at the origin 0 in $\mathbb{R}^d$. There exists a useful theorem in understanding the metrical structure of tangent cones for finite dimensional Alexandrov spaces.

Definition 5.6. A sequence $\{(M_n, \mathsf{d}_n, p_n)\}_n$ of pointed metric spaces converges to a pointed metric space (M, d_M, p) in the Gromov–Hausdorff sense if the following holds: For every $r > 0$ and $\varepsilon > 0$ there exists a positive integer n_0 such that for $n > n_0$ there is a map f_n from an open ball $B_r(p_n)$, of center at p_n and radius r, to M, which satisfies the following three properties;

(1) $f_n(p_n) = p$.
(2) $\sup_{x,y \in M_n} |\mathsf{d}_n(x,y) - \mathsf{d}_M(f_n(x), f_n(y))| < \varepsilon$.
(3) the ε-neighborhood of $f_n(B_r(p_n))$ contains the open ball $B_{r-\varepsilon}(p)$.

Theorem 5.7. ([6, §7.8.1],[5, Theorem 10.9.3]) *Let (M, d_M) be a complete, finite dimensional Alexandrov space of non-negative curvature. Then, at every p in M, the tangent cone (K_p, d_p, o_p) is isometric to the limit of the scaled pointed metric space $(M, c \cdot \mathsf{d}_M, p)$ in the Gromov–Hausdorff sense as c diverges to infinity.*

We now prove that $(\overline{\mathcal{N}_0^d}, W_2)$ has a cone structure.

Theorem F. *The tangent cone at $N(0)$ is isometric to $(\overline{\mathcal{N}_0^d}, W_2)$.*

Proof. We shall prove that the scaled pointed metric space $(\overline{\mathcal{N}_0^d}, n \cdot W_2, N(0))$ converges to $(\overline{\mathcal{N}_0^d}, W_2, N(0))$ in the Gromov–Hausdorff sense as n diverges to infinity. For every $r > 0$, $\varepsilon > 0$ and a positive integer n, we shall check that a map

$$f_n : B_r(N(0)) \to \overline{\mathcal{N}_0^d}, \quad f_n(N(V)) = N(n^2 V)$$

satisfies the properties in Definition 5.6. Here $B_r(N(0))$ stands for an open ball in $(\overline{\mathcal{N}_0^d}, W_2)$. Property (1) holds since we have

$$f_n(N(0)) = N(n^2 \cdot 0) = N(0).$$

The relations $\mathrm{tr}(n^2 V) = n^2 \mathrm{tr} V$ and $(n^2 V)^{1/2} = n V^{1/2}$ imply

$$\begin{aligned} W_2(f_n(N(V)), f_n(N(U)))^2 &= n^2 \left[\mathrm{tr}V + \mathrm{tr}U - 2\mathrm{tr}\sqrt{U^{\frac{1}{2}} V U^{\frac{1}{2}}} \right] \\ &= n^2 W_2(N(V), N(U))^2 \end{aligned}$$

for any $N(V), N(U)$ in $\overline{\mathcal{N}_0^d}$. This shows that Property (2) holds.

For any $N(V)$ in $B_{r-\varepsilon}(N(0))$, $f_n(N(n^{-2}V)) = N(V)$ and $N(n^{-2}V)$ belongs to $B_{n^{-1}r}(N(0))$, proving that Property (3) holds.

Thus $(\overline{\mathcal{N}_0^d}, n \cdot W_2, N(0))$ converges to $(\overline{\mathcal{N}_0^d}, W_2, N(0))$ and Theorem 5.7 yields Theorem F. □

Finally, we state the following metrical property of the tangent cones on $\overline{\mathcal{N}_0^d}$.

Proposition 5.8. *For all V in $\mathrm{Sym}(d, \mathbb{R})_{\geq 0}$ and P in $O(d)$, the tangent cone at $N(V)$ is isometric to the tangent cone at $N({}^{\mathsf{T}}PVP)$.*

Proof. Since a pointed isometry map between $(\overline{\mathcal{N}_0^d}, N(V))$ and $(\overline{\mathcal{N}_0^d}, N({}^{\intercal}PVP))$ can be extended to an isometry map between $K_{N(V)}$ and $K_{N({}^{\intercal}PVP)}$, it is sufficient to construct such an isometry map. Let φ be a map on $\overline{\mathcal{N}_0^d}$ sending $N(X)$ to $N({}^{\intercal}PXP)$. Then we obtain

$$W_2(\varphi(N(X)), \varphi(N(Y)))^2 = \mathrm{tr}X + \mathrm{tr}Y - 2\mathrm{tr}\sqrt{X^{\frac{1}{2}}YX^{\frac{1}{2}}} = W_2(N(X), N(Y))^2$$

for all $N(X), N(Y)$ in $\overline{\mathcal{N}_0^d}$. Additionally, it is clear that $\varphi(N(V)) = N({}^{\intercal}PVP)$. Thus φ is a desired isometry map. □

5.4. **Tangent cone in the case of dimension 2.** In the previous subsection, we described the metrical structure of the tangent cones at $N(V)$ in $\mathcal{S}_k$ for $k = 0, d$. In the case of $d = 2$, we can derive a metrical structure of the tangent cones at $N(\mathrm{diag}[\lambda^2, 0])$. Applying Proposition 5.8, we obtain a metrical structure of the tangent cones at each point of $\mathcal{S}_1$. As mentioned in the introduction, we acquire the following equivalence;

$$X = \begin{pmatrix} x & z \\ z & y \end{pmatrix} \in S^E(2,1) \Longleftrightarrow (x, y, z) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}.$$

Thereby $S^E(2,1)$ is homeomorphic to the upper half-space $\mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$.

Theorem G. *For a positive number λ, the tangent cone at $N(\mathrm{diag}[\lambda^2, 0])$ is isometric to the Euclidean upper half-space $\mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$.*

Proof. Set $\Lambda = \mathrm{diag}[\lambda^2, 0]$ and $p = N(\Lambda)$. We prove Theorem G in several steps: First we construct a geodesic from p to $N(V)$ in $\mathcal{N}_0^2$. We next confirm that an inextensible geodesic starting at p meets a boundary of a ball $B_p(\lambda/2)$. Then we compute angles between geodesics starting at p. Finally we prove that $(\Sigma_p', \angle_p)$ is isometric to a semi-sphere. This isometry can be extended to an isometry between the tangent cone K_p and the Euclidean upper half-space $\mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$. In what follows, components of symmetric matrices V, X in $\mathrm{Sym}(2, \mathbb{R})_{\geq 0}$ and a vector Ξ in $\mathbb{R}^3$ are expressed by

$$V = \begin{pmatrix} u & w \\ w & v \end{pmatrix},\ X = \begin{pmatrix} x & z \\ z & y \end{pmatrix},\ \Xi = (\xi, \eta, \zeta),\ \text{respectively.}$$

For $N(V)$ in $\mathcal{N}_0^2$, u, v should be positive and a linear transform T from $N(V)$ to $N(\Lambda)$ is given by $\mathrm{diag}[\lambda u^{-\frac{1}{2}}, 0]$. In fact, let $\mathcal{T}$ be a linear map associated with T, that is, $\mathcal{T}(x) = Tx$ for x in $\mathbb{R}^2$. Then $\mathcal{T}$ pushes $N(V)$ forward to $N(\Lambda)$ and $\mathcal{T}$ is a gradient of a convex function $\langle x, Tx\rangle/2$. Theorem 2.1 yields that $[\mathrm{id} \times \mathcal{T}]_\sharp N(V)$ is an optimal transport plan between $N(V)$ and $N(\Lambda)$. Moreover, a geodesic from $N(V)$ to $N(\Lambda)$ is given by

$$[(1-t)\mathrm{id} + t\mathcal{T}]_\sharp N(V),$$

for t in $[0,1]$. Let $\{N(V(t))\}_{t\in[0,1]}$ be a geodesic from $N(\Lambda)$ to $N(V)$ (not from $N(V)$ to $N(\Lambda)$), then we acquire

$$N(V(t)) = [t\cdot \mathrm{id} + (1-t)\mathcal{T}]_\sharp N(V),$$
$$V(t) = \begin{pmatrix} (\lambda + t(\sqrt{u}-\lambda))^2 & twu^{-1/2}(\lambda + t(\sqrt{u}-\lambda)) \\ twu^{-1/2}(\lambda + t(\sqrt{u}-\lambda)) & t^2 v \end{pmatrix}. \tag{5.1}$$

This expression of geodesics can be extended to the case of $u > 0$.

For $N(V)$ satisfying $W_2(p, N(V)) \geq \lambda/2$, it is trivial that a geodesic from p to $N(V)$ meets the boundary of $B_p(\lambda/2)$. We consider the case of $N(V)$ in $B_p(\lambda/2)$, that is,

$$l^2 = W_2(p, N(V))^2 = (\sqrt{u}-\lambda)^2 + v < \lambda^2/4.$$

Because V lies in $\mathrm{Sym}(2,\mathbb{R})_{\geq 0}$, v is non-negative and hence u is positive. It enables us to have a geodesic $\{N(V(t))\}_{t\in[0,1]}$ from $N(\Lambda)$ to $N(V)$ as in (5.1). For all positive numbers t, $V(t)$ are well-defined and we have

$$\det V(t) = (\det V)t^2(\lambda + t(\sqrt{u}-\lambda))^2 u^{-1} \geq 0,$$
$$\mathrm{tr} V(t) = (\lambda + t(\sqrt{u}-\lambda))^2 + t^2 v \geq 0,$$

proving $V(t)$ in $\mathrm{Sym}(d,\mathbb{R})_{\geq 0}$. For $\widetilde{V} = V(\lambda/2l)$, we have $W_2(p, N(\widetilde{V})) = \lambda/2$. Then we obtain a geodesic $\{N(\widetilde{V}(t))\}_{t\in[0,1]}$ from $N(\Lambda)$ to $N(\widetilde{V})$ as in (5.1). It is an extension of $\{N(V(t))\}_{t\in[0,1]}$ since geodesics in $\overline{\mathcal{N}_0^d}$ do not branch. Therefore inextensible geodesics starting at p meet the boundary of $B_p(\lambda/2)$.

It suggests that Σ'_p can be identified with

$$\{\gamma_V : \text{a geodesic from } p = N(\Lambda) \text{ to } N(V) \mid W_2(p, N(V)) = \lambda/2\}.$$

We compute angles between geodesics in Σ'_p. For geodesics $\gamma_V(t) = N(V(t))$ and $\gamma_X(t) = N(X(t))$ in Σ'_p, we acquire

$$V(t) = \begin{pmatrix} (\lambda + t(\sqrt{u}-\lambda))^2 & twu^{-1/2}(\lambda + t(\sqrt{u}-\lambda)) \\ twu^{-1/2}(\lambda + t(\sqrt{u}-\lambda)) & t^2 v \end{pmatrix},$$
$$X(t) = \begin{pmatrix} (\lambda + t(\sqrt{x}-\lambda))^2 & tzx^{-1/2}(\lambda + t(\sqrt{x}-\lambda)) \\ tzx^{-1/2}(\lambda + t(\sqrt{x}-\lambda)) & t^2 y \end{pmatrix}.$$

Since $N(V)$, $N(X)$ belong to the boundary of $B_p(\lambda/2)$, we have $ux \neq 0$. We compute the comparison angle of $\angle\gamma_V(t)p\gamma_X(t)$:

$$\cos\tilde{\angle}\gamma_V(t)p\gamma_X(t) = \frac{W_2(p,\gamma_V(t))^2 + W_2(p,\gamma_X(t))^2 - W_2(\gamma_V(t),\gamma_X(t))^2}{2W_2(p,\gamma_V(t))W_2(p,\gamma_X(t))}$$
$$= \frac{4}{\lambda^2}\left[(\sqrt{u}-\lambda)(\sqrt{x}-\lambda) + \frac{wz + \det(VX)^{\frac{1}{2}}}{\sqrt{ux}} + \frac{o(t^2)}{t^2}\right],$$

which converges to

$$\frac{4}{\lambda^2}\left[(\sqrt{u}-\lambda)(\sqrt{x}-\lambda) + (ux)^{-\frac{1}{2}}\left(wz + \det(VX)^{\frac{1}{2}}\right)\right]$$

as t converges to 0. Thus we have

$$\angle_p(\gamma_V,\gamma_X)=\arccos\frac{4}{\lambda^2}\left[(\sqrt{u}-\lambda)(\sqrt{x}-\lambda)+(ux)^{-\frac{1}{2}}\left(wz+\det(VX)^{\frac{1}{2}}\right)\right].$$

For the rest of the proof, we identify γ_V in Σ'_p with the matrix V. Let $\mathbb{S}^+$ be the semi-sphere defined by

$$\mathbb{S}^+=\{\Xi=(\xi,\eta,\zeta)\mid \xi^2+\eta^2+\zeta^2=1,\eta\geq 0\}.$$

It is a complete metric space with the standard Euclidean angle metric $\angle$. We define a map ψ from Σ'_p to $\mathbb{S}^+$ by

$$\psi(V)=\frac{2}{\lambda}\left(\sqrt{u}-\lambda,\left(v-w^2u^{-1}\right)^{\frac{1}{2}},wu^{-\frac{1}{2}}\right).$$

The map ψ has an inverse map φ given by

$$\varphi(\Xi)=\frac{\lambda^2}{4}\begin{pmatrix}(2+\xi)^2 & (2+\xi)\zeta\\ (2+\xi)\zeta & \eta^2+\zeta^2\end{pmatrix}.$$

For V,X in Σ'_p, we have

$$\cos\angle_p(V,X)=\langle\psi(V),\psi(X)\rangle=\cos\angle(\psi(V),\psi(X)),$$

proving that $(\Sigma',\angle_p)$ is isometric to $(\mathbb{S}^+,\angle)$.

Thus $(\Sigma',\angle_p)$ is complete and hence $(\Sigma'_p,\angle_p)=(\Sigma_p,\angle_p)$. The isometry map ψ can be extended to an isometry map Ψ from the tangent cone K_p to the Euclidean half-space $\mathbb{R}\times\mathbb{R}_{\geq 0}\times\mathbb{R}$:

$$\Psi(V,t)=t\psi(V)=\frac{2t}{\lambda}\left(\sqrt{u}-\lambda,\left(v-w^2u^{-1}\right)^{\frac{1}{2}},wu^{-\frac{1}{2}}\right).$$

This completes the proof of Theorem G. □

References

[1] S. Amari, *Differential-geometrical methods in statistics*, Lecture Notes in Statistics, vol. 28, Springer-Verlag, New York, 1985.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[3] P. Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417.

[5] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001.

[6] Y. Burago, M. Gromov, and G. Perelman, *A. D. Aleksandrov spaces with curvatures bounded below*, Uspekhi Mat. Nauk **47** (1992), no. 2(284), 3–51, 222.

[7] D.C. Dowson and B.V. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal. **12** (1982), no. 3, 450–455.

[8] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian geometry*, third ed., Universitext, Springer-Verlag, Berlin, 2004.

[9] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161.

[10] C.R. Givens and R.M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J. **31** (1984), no. 2, 231–240.
[11] M.Knott and C.S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), no. 1, 39–49.
[12] J. Lott, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys. **277** (2008), no. 2, 423–437.
[13] ———, *Optimal transport and Perelman's reduced volume*, Calc. Var. Partial Differential Equations **36** (2009), no. 1, 49–84.
[14] J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, J. Funct. Anal. **245** (2007), no. 1, 311–333.
[15] ———, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.
[16] R.J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179.
[17] R.J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), no. 3.
[18] R.J. McCann and P. Topping, *Ricci flow, entropy and optimal transportation*, Amer. J. Math(to appear).
[19] I. Olkin and F.Pukelsheim, *The distance between two random vectors with given dispersion matrices*, Linear Algebra Appl. **48** (1982), 257–263.
[20] B. O'Neill, *The fundamental equations of a submersion*, Michigan Math. J. **13** (1966), 459–469.
[21] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.
[22] G. Perelman, *Elements of Morse theory on Aleksandrov spaces*, Algebra i Analiz **5** (1993), no. 1, 232–241.
[23] G. Perelman and A.M. Petrunin, *Extremal subsets in Aleksandrov spaces and the generalized Liberman theorem*, Algebra i Analiz **5** (1993), no. 1, 242–256.
[24] G. Perelmann, *The entropy formula for the ricci flow and its geometric applications*, Preprint.
[25] A.M. Petrunin, *Semiconcave functions in Alexandrov's geometry*, Surveys in differential geometry. Vol. XI, Surv. Differ. Geom., vol. 11, Int. Press, Somerville, MA, 2007, pp. 137–201.
[26] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.
[27] K.-T. Sturm, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006), no. 1, 133–177.
[28] A. Takatsu and T. Yokota, *Cone structure of L^2-Wasserstein spaces*, Preprint.
[29] P. Topping, *$\mathcal{L}$-optimal transportation for Ricci flow*, J. Reine Angew. Math. **636** (2009), 93–122.
[30] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
[31] ———, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009.

Mathematical Institute, Tohoku University, Sendai 980-8578, Japan
E-mail address: sa6m21@math.tohoku.ac.jp