ABSTRACT

Title of dissertation: An optimal transport approach to some problems in frame theory

Clare G. Wickman, Doctor of Philosophy, 2014

Dissertation directed by: Kasso A. Okoudjou
Department of Mathematics

A probabilistic frame is a probability measure on $\mathbb{R}^d$ which has finite second moment and support spanning $\mathbb{R}^d$. These objects generalize finite frames for $\mathbb{R}^d$, which are redundant spanning sets. Working in the Wasserstein space $P_2(\mathbb{R}^d)$, we investigate the properties of these measures, finding geodesics of frames in the Wasserstein space and using machinery from probability theory to define more general concepts of duality, analysis, and synthesis. We then use the Otto calculus to construct gradient flows for the probabilistic p-frame potential and a related potential which we term the (p-)tightness potential, the minimizers of which are the tight probabilistic p-frames. We demonstrate the well-posedness of the minimization problem via the minimizing movement scheme, with a focus on the case $p = 2$. We link this result to earlier approaches to solving the Paulsen Problem for finite frames which involved differential calculus.

An Optimal Transport Approach to Some Problems in Frame Theory

by

Clare Georgianna Wickman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Kasso Okoudjou, Chair/Advisor
John Benedetto
Wojciech Czaja
Radu Balan
Alexander Barg

Soli Deo gloria.

Acknowledgments

I owe the greatest gratitude to Dr. Kasso A. Okoudjou for his guidance and almost endless patience with the research and writing process and for the inspiration and brilliant insight he gave so generously. Many thanks, as well, to everyone at the Norbert Wiener Center, especially Dr. John Benedetto, Dr. Wojciech Czaja, and Dr. Radu Balan. I also thank my Dean's representative, Dr. Alexander Barg for his engaged and kind questioning. I am grateful to Dr. Konstantina Trivisa for her example and encouragement along the way, and to Dr. Muruhan Rathinam of the University of Maryland, Baltimore County, for introducing me to analysis and being an understanding mentor and faithful friend. I am also so grateful to the wonderful people who work behind the scenes, making the math department run smoothly and helping students file endless waivers, especially Celeste Regalado.

I also owe thanks to the Johns Hopkins University Applied Physics Laboratory, which supported me in the second half of my graduate studies, and most especially to my supervisors Dean Fisher, Leslie Hedderly, Anthony Webber, and Rochele Llanso for always encouraging me and smoothing the path.

Finally, I thank my roommate, mentor, and friend Karamatou Yacoubou Djima, for her inspiring devotion to math. I thank my "math ladies," Marie Chau, Hana Ueda, and Jennifer Clarkson-Greene; we kept each other sane! And I thank my other colleagues and friends in the math department, especially Mark Lai, Wei-Hsuan Yu, Joshua Ballew, Victoria Taroudaki, Alex Cloninger, Nate Strawn, Emily King, Shelby Wilson, Amanda Galante, Jinxin Xue, and Sean Rostami. I am so lucky in my fiance Timothy Lau, who assured me endlessly that it would all be alright, and listened patiently to my presentation until he knew enough to ask questions. I thank my darling parents Richard and Posey Wickman, and my grandmother, Georganne Johnson, who understood why I couldn't be around as much as any of us would have liked.

Table of Contents

List of Notations

$\mathbb{R}$	the field of real numbers
$\mathbb{C}$	the field of complex numbers
$\#\lvert A\rvert$	the cardinality of the set A
CBS	the Cauchy-Bunyakowskii-Schwarz inequality
S_N	the set of permutations of N elements
Φ	the analysis operator for the finite frame $\{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$
$A^\dagger$	the Moore-Penrose pseudoinverse of the matrix A.
S_μ	the probabilistic frame operator for μ
G_μ	the probabilistic Grammian for μ
$M_2(\mu)$	the second moment of μ
$Cov(\mu)$	the covariance of μ
$\overline{\mu}$	the mean of μ
$\text{supp}(\mu)$	the support of a measure μ
$B(\mathbb{R}^d)$	the Borel sets in $\mathbb{R}^d$
$C_b^0(\mathbb{R}^d)$	the space of bounded, continuous functions on $\mathbb{R}^d$
$L^p(\mu;\mathbb{R}^d)$	L^p space of μ-measurable $\mathbb{R}^d$-valued maps
$\mathcal{L}^1$-a.e.	almost everywhere with respect to Lebesgue measure
$P(\mathbb{R}^d)$,	the space of probability measures on $\mathbb{R}^d$
$P_p(\mathbb{R}^d)$	the space of probability measures with finite p-th moments
$W_p(\mu,\nu)$	the p-Wasserstein distance between μ and ν
$\pi^i, \pi^{i,j}$	projection operators from a product space such as $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ to its i^{th} or i^{th} and j^{th} components
ι	the identity map
$\Gamma(\mu,\nu)$	the set of joint probability distributions with marginals μ and ν
$\Gamma_0(\mu,\nu)$	the set of joint probability distributions with marginals μ and ν which are optimal for the (2)-Wasserstein distance
$DS(M,N)$	the set of $M \times N$ nonnegative matrices with entries summing to unity
$T_\# \mu$	the push-forward of μ by T
D_μ	the set of transport duals to μ
ΓD_μ	the set of joint distributions inducing duality
$B_\delta(x)$	the ball of radius δ around a point x in a metric space
$Tan_\mu P_2(\mathbb{R}^d)$	the tangent bundle to $P_2(\mathbb{R}^d)$ at μ
$\partial\phi(\mu)$	the Fréchet subdifferential of ϕ at μ in the Wasserstein space
$\partial\phi(\mu)$	the extended Fréchet subdifferential of ϕ at μ
$\partial\phi(\mu)$	the strong (extended) Fréchet subdifferential of ϕ at μ

Chapter 1

Introduction

1.1 Background

In this thesis we bring together some of the key ideas and methods of two very lively fields of mathematical research, frame theory and optimal transport, using the methods of the second to answer questions posed in the first.

1.1.1 Frames

Introduced in 1952 by Duffin and Schaeffer [27] in their paper on nonharmonic Fourier series, frames are redundant spanning sets of vectors or functions that can be used to represent signals in various spaces in a faithful but nonunique way. It is this very nonuniqueness which guarantees that the frame expansion of a signal may be more stable and robust to noise-induced errors than its expansion in any orthonormal basis. In finite-dimensional settings, because they provide an intuitive framework for describing and solving problems in coding theory, analog-to-digital quantization theory, sparse representation, and compressive sensing, certain classes of frames have proven useful in work on signal processing for telecommunications and other applications. This utility was not fully appreciated until the renaissance of interest in frame theory in infinite-dimensional settings in the late 1980s due of the work of Daubechies, Meyer, and Grossman on the construction of wavelet frames with tractable reconstruction properties [23]. These and other frames have now become some of the standard tools of image processing: Gabor frames, Fourier frames, shearlets, curvelets, wavelets, and multiresolution analyses.

Briefly, a **frame** for $\mathbb{R}^d$ is a set $\{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$, $N \geqslant d$ for which there exist constants $0 < A \leqslant B < \infty$ such that for all $x \in \mathbb{R}^d$,

$$A\|x\|^2 \leqslant \sum_{i=1}^{N}\langle \varphi_i\, , x\rangle^2 \leqslant B\|x\|^2.$$

A frame $\{\psi_i\}_{i=1}^N \subset \mathbb{R}^d$ is said to be **dual** to $\{\varphi_i\}_{i=1}^N$ if for all $x \in \mathbb{R}^d$,

$$x = \sum_{i=1}^{N}\langle \varphi_i\, , x\rangle \psi_i = \sum_{i=1}^{N}\langle \psi_i\, , x\rangle \varphi_i.$$

Tight frames, that is frames for which $A = B$ in the above definition, are particularly useful because they have a basis-like reconstruction property that is useful in applicationsthey are self-dual up to a constant. In geometry, they are also known as eutactic stars. It is a corollary of Naimark's theorem that finite tight frames are the projection of an orthonormal basis onto a lower-dimensional space [6, 22]; consequently, in principle, it is easy to construct a tight frame. However, there are subclasses of tight frames, such as finite unit-norm tight frames (FUNTFs) and equal-norm Parseval frames, equiangular tight frames, and Grassmannian frames, which are interesting, as well as desirable from a coding theory perspective, and which are not always so simple to construct.

Indeed, a number of methods of building tight frames exist for specific applications [16]. Of particular interest are FUNTFs, which are tight frames all of whose elements have norm one. These frames combine the stability properties of tight frames with the control of frames of uniform norm, and they are connected to problems of equidistribution on the sphere. In [5], Benedetto and Fickus show that FUNTFs are minimizers of a functional related to this equidistribution problem called the frame potential. Equiangular tight frames, those for which the mutual coherence between distinct frame elements is of constant magnitude, are another class of FUNTF that proves elusive, and constructions of them for higher dimensions are scarce. In [31, 45], it is shown that the minimizers of another functional called the p-frame potential are precisely the equiangular FUNTFs.

Other approaches exist to constructing FUNTFs. In [13], the authors give an algorithm for the construction of all frames with a given spectrum and compatible set of lengths, as defined by the Schur-Horn Theorem. They thus improve on the state of the art for generating FUNTFs, namely spectral tetris and truncations of the Discrete Fourier Transform (DFT) matrix. However, this algorithm requires two challenging decision steps which must be made somewhat blindly. The existence of the potentials mentioned above suggests that variational methods for construction of tight frames and FUNTFs might complement these algebraic methods.

Moreover, there are more questions to answer than simply how to construct classes of tight frames. For instance, **Parseval frames** are tight frames for which the frame constant is one, and equal-norm Parseval frames, when used to encode and decode a signal, are optimally robust to one erasure [8]. The Paulsen problem asks the distance to the closest equal-norm Parseval frame from a given almost-equal norm, almost-Parseval frame (see Definition 4.1). This is a question that constructions *à la* [13] may not be able to answer. Indeed, while partial results exist, this problem remains open; in [8, 17], Bodmann and Casazza and Fickus, Mixon, and Casazza give two distinct differential calculus approaches to answering it. It is one of the aims of this thesis to present a framework for addressing the Paulsen problem related to the approach of [17].

It should also be noted that since in some applications one cannot always choose the frame used for encoding, let alone require that it be tight, others have explored optimal dual frames for various error types, such as erasure, quantization, and noise, with various probabilistic distributions [38, 46]. Additionally, there are generalizations of finite frames, termed **fusion frames** which can be used to mimic the distributed processing of sensor networks; construction of these frames poses additional interesting questions. We present a new version of frame duality which can be viewed through the fusion lens.

In line with the development of precise estimates for random frames and optimal frames for probabilistic erasures, the idea of probabilistic frames was developed in a series of papers ([29–31]). Simply put, a **probabilistic frame** μ for $\mathbb{R}^d$ is a probability measure on $\mathbb{R}^d$ for which there exist constants $0 < A \leqslant B < \infty$ such that for all $x \in \mathbb{R}^d$,

$$A\|x\|^2 \leqslant \int_{\mathbb{R}^d} \langle x , y \rangle^2 d\mu(y) \leqslant B\|x\|^2.$$

Importantly for our purposes, the ideas of tightness and equiangularity can be extended to these objects. Probabilistic frames are related to statistical shape analysis, as detailed in [31], and they are linked to the classical problem of estimating the population covariance from a sample [29, 51]. However, the true strength of probabilistic frames lies in the fact that they embed the space of finite frames for

Euclidean space in the space of probability measures with finite second moments, a metric space with distance defined by the concept of optimal transport. Much effort has been expended over the past 25 years to define a calculus for this space, and it is this calculus which will allow us to rigorously construct gradient flows for the potentials mentioned above in order to identify tight probabilistic frames of various types.

1.1.2 Optimal Transport

The most natural space to explore probabilistic frames is the Wasserstein space of probability measures with finite second moments, or, more generally, with finite p-th moments. This space is the realm of optimal transport theory, an area going back to the work of Monge in the 1780s. The classical question in optimal transport, the Monge-Kantorovich problem, is to find a joint measure γ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, measures on $\mathbb{R}^d$, which minimizes the cost functional

$$C_\gamma(\mu, \nu) := \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma(x, y)$$

among all such joint measures, where $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a lower-semicontinuous infinitesimal cost function, integrable with respect to γ. This problem was studied by Kantorovich in the 1930s in both the continuous and discrete settings because of its many applications in logistics and economics. Today, its methods are commonly used in a multitude of applications, from radar design to image processing [3,43,48].

In [44], Monge specifically sought a deterministic map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that T is a change of variables pushing the measure μ to the measure ν, $T_\# \mu = \nu$, on whose graph in $\mathbb{R}^d \times \mathbb{R}^d$ a joint measure γ would be concentrated which would be optimal for the cost $c(x, y) = \| x-y \|$, as will be explained in more detail in Section 1.4. Proofs of the existence of this map were very difficult, and it was not until the late 1980s that a number of people working independently discovered connections between the Monge problem and PDE theory which broadened its appeal. Yann Brenier, independently of Cuesta-Albertos and Matrán and Rachev and Rüschendorf, proved that for the quadratic cost function, given an absolutely continuous source measure, a unique solution existed which would be the pushforward of the source by the gradient of a convex function [10, 21, 47].

Approaching from the PDE perspective, Evans and Gangbo worked out an alternative proof, and they were followed by a number of other mathematicians, including Caffarelli, Ambrosio, and McCann, who worked out many more details relating to more general cost functions and questions of regularity [11, 32, 37]. Otto, in a paper with Jordan and Kinderlehrer, worked out a metric calculus approach which allowed a much finer exploration of the geometry of the space of probability measures and a formal calculus for the optimal transport problem [39]. Over the past 20 years, many more people have contributed to the development of methods for solving problems in optimal transport; Villani gives an excellent history of the field in [53]. Still cited by almost every new paper in the field is the survey by Ambrosio, Gigli, and Savaré [2], upon which we shall call many times. Recasting some of the above finite frame theory problems as problems for probabilistic frames, we will use elements of this calculus to establish existence of solutions and then to construct them.

1.2 Summary of Results

Motivated in part by the Paulsen problem, we study the space of probabilistic frames from the optimal transport perspective. We consider constructions of geodesic paths of frames and of paths of frames along gradient flows for various potentials.

In Chapter 2, we briefly review the basic tools of optimal transport and then use them to construct paths of frames along geodesics in the Wasserstein space. We prove structural results about the space of probabilistic frames and identify conditions under which geodesics will remain in that space. We give examples of both discrete and continuous probabilistic frames which meet these conditions. On the discrete side, we connect them to basic results on ranks of convex combinations of matrices; on the absolutely continuous side, we connect them to deep results about regularity for optimal transport maps.

In Chapter 3, we reconsider the idea of duality and define transport duals in the space of probabilistic frames, which generalize the idea of duality in the finite-frame case. We connect this construction to fusion frames. We also generalize the operations of analysis and synthesis using decompositions of probability measures via the disintegration theorem.

In Chapter 4, we use gradient flows in the space of probability measures to find tight frames. We define a tightness potential related to the frame potential and show that gradient flow solutions exist to the corresponding minimization problem. This generalizes a result of Casazza and Fickus ([17]), which shows that FUNTFs can be found as the solution of a system of nonlinear ODEs. We also give preliminary results indicating that similar problems involving the p-frame potential are also well-posed.

1.3 Notation

Let $\mathbb{R}^d$ denote d-dimensional Euclidean space, and let $\langle \cdot\, , \cdot \rangle$ denote the inner product on this space. For any $x \in \mathbb{R}^d$, let $\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}}$. When p is not specified, it can be assumed to be 2. Let $S^{d-1} = \{x \in \mathbb{R}^d : \ \|x\| = 1\}$ denote the unit ball with respect to the 2-norm in $\mathbb{R}^d$. Let $\mathbb{R}^{m\times n}$ denote the set of $m \times n$ matrices with real entries, and given $A \in \mathbb{R}^{m\times n}$, let $A^\top$ denote its transpose and, if it is a square matrix, $tr(A)$ its trace. We will sometimes write the inner product $\langle x\, , y \rangle$ as $x^\top y$ and the outer product as $xy^\top$. As above, we say that a set of vectors $\{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$ is a **frame** if there exist constants $0 < A \leqslant B < \infty$ such that for all $x \in \mathbb{R}^d$,

$$A\|x\|^2 \leqslant \sum_{i=1}^{N} \langle x\, , \varphi_i \rangle^2 \leqslant B\|x\|^2.$$

We take A and B to be the **frame bounds**, the sharpest such values for the frame. Again, a frame is **tight** if $A = B$ and **Parseval** if $A = B = 1$. We define the **analysis operator** for a frame $\Phi = \{\varphi_i\}_{i=1}^N$ with the overloaded notation $\Phi \in \mathbb{R}^{N\times d}$,

$$\Phi = \begin{bmatrix} \varphi_1^\top \\ \vdots \\ \varphi_N^\top \end{bmatrix}.$$

Similarly, we define its adjoint, the **synthesis operator**, as

$$\Phi^\top = [\varphi_1 \cdots \varphi_N] \in \mathbb{R}^{d\times N}.$$

We define the **frame operator** $\Phi^\top\Phi : \mathbb{R}^d \to \mathbb{R}^d$, and note that

$$\Phi^\top\Phi x = \sum_{i=1}^{N} \langle \varphi_i \, , x \rangle \varphi_i.$$

We also define the **Grammian**, $\Phi\Phi^\top$, where

$$(\Phi\Phi^\top)_{i,j} = \langle \phi_i \, , \phi_j \rangle.$$

We similarly define a **probabilistic p-frame** as a probability measure μ on $\mathbb{R}^d$ for which there exist frame bounds $0 < A \leqslant B < \infty$ such that for all $x \in \mathbb{R}^d$,

$$A\|x\|^p \leqslant \int_{\mathbb{R}^d} \langle x \, , y \rangle^p d\mu(y) \leqslant B\|x\|^p.$$

When we use the term **probabilistic frame**, we mean a probabilistic 2-frame and its associated frame bounds. Each probabilistic p-frame, $p \geqslant 2$ is also a probabilistic 2-frame. Given a finite frame $\Phi = \{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$, we define the canonical probabilistic frame for Φ as the uniformly-weighted sum of delta-masses, $\mu_\Phi = \frac{1}{N}\sum_{i=1}^{N} \delta_{\varphi_i}$. Other terms related to probabilistic frames will be defined in the following preliminaries.

1.4 Preliminaries

To begin the discussion of probabilistic frames, a few definitions are needed.

Definition 1.1. A probability measure μ on $\mathbb{R}^d$ is an element of $P_p(\mathbb{R}^d)$, the space of probability measures with **finite p-th moment**, if it satisfies:

$$M_p^p(\mu) := \int_{\mathbb{R}^d} \|x\|^p d\mu(x) < \infty$$

Definition 1.2. The **support** of a probability measure μ on $\mathbb{R}^d$ is the set:

$$\operatorname{supp}(\mu) := \left\{x \in \mathbb{R}^d \text{ s.t. for all open sets } U_x \text{ containing } x, \mu(U_x) > 0\right\}.$$

Finally, we define a natural metric on $P_p(\mathbb{R}^d)$, the **(p-)Wasserstein distance**.

Definition 1.3. The **p-Wasserstein distance** between two probability measures μ and ν on $\mathbb{R}^d$ is:

$$W_p^p(\mu,\nu) := \inf_{\gamma}\left\{\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \| x-y\|^p d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\right\},$$

where $\Gamma(\mu,\nu)$ is the set of all joint probability measures γ on $\mathbb{R}^d \times \mathbb{R}^d$ such that for all $A, B \subset \mathcal{B}(\mathbb{R}^d)$, $\gamma(A \times \mathbb{R}^d) = \mu(A)$ and $\gamma(\mathbb{R}^d \times B) = \nu(B)$.

The search for the set of joint measures which induce the infimum is a variant of the **Monge-Kantorovich problem**. A joint distribution γ_0 which induces this infimum is called an **optimal transport plan**. In the quadratic case, when μ and ν do not assign positive measure to isolated points, then

$$W_2^2(\mu,\nu) := \inf_{T}\left\{\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \| x-T(x)\|^2 d\mu(x) : T_{\#}\mu = \nu\right\},$$

where T is a **deterministic transport map** (or **deterministic coupling**): i.e., for all ν-integrable functions ϕ,

$$\int_{\mathbb{R}^d} \phi(y)d\nu(y) = \int_{\mathbb{R}^d} \phi(T(x))d\mu(x).$$

When the search for the minimizing joint distributions of the Monge-Kantorovich problem is limited to deterministic transport plans, we have the original **Monge problem**. Equipped with the 2-Wasserstein distance, $P_2(\mathbb{R}^d)$ is a complete, separable metric space. In fact, the set of measures with discrete, finite support is dense in $P_2(\mathbb{R}^d)$. Convergence in the space has several equivalent formulations. We will make use of the following notions of convergence:

Definition 1.4. ([53, Definition 6.8]) A sequence of measures $\{\mu_n\} \subset P_p(\mathbb{R}^d)$ is said to **converge weakly** to $\mu \in P_p(\mathbb{R}^d)$ if the following two conditions are met:

- $\mu_n \to \mu$ **weakly** or **narrowly**, i.e.:

$$\forall f \in C_b(\mathbb{R}^d) \quad \int_{\mathbb{R}^d} f(x) d\mu_n(x) \to \int_{\mathbb{R}^d} f(x) d\mu(x)$$

- For any (and therefore every) x_0 in $\mathbb{R}^d$, $\int_{\mathbb{R}^d} \|x - x_0\|^p d\mu_n(x) \to \int_{\mathbb{R}^d} \|x - x_0\|^p d\mu(x)$.

A second, equivalent definition is:

Definition 1.5. ([53, Definition 6.8]) A sequence of measures $\{\mu_n\} \subset P_p(\mathbb{R}^d)$ is said to **converge weakly** to μ in $P_p(\mathbb{R}^d)$ if for all continuous functions ϕ with

$$|\phi(x)| \leqslant C(1 + \|x - x_0\|^p),$$

for some $C > 0$ and some $x_0 \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \phi(x) d\mu_n(x) \to \int_{\mathbb{R}^d} \phi(x) d\mu(x).$$

1.5 Probabilistic Frames as a Subset of $P_2(\mathbb{R}^d)$

With the space above in mind, we give the following definition:

Definition 1.6. A probability measure μ on $\mathbb{R}^d$ is a **probabilistic frame** if and only if there exist positive constants A and B such that for all $y \in \mathbb{R}^d$,

$$A\|y\|^2 \leqslant \int_{\mathbb{R}^d} |\langle x , y \rangle|^2 d\mu(x) \leqslant B\|y\|^2.$$

A probabilistic frame is said to be tight if $A = B$.

By [30, Theorem 5], a probability measure μ on $\mathbb{R}^d$ is a probabilistic frame if and only if it has finite second moment, and the linear span of its support is $\mathbb{R}^d$. This result may be stated in terms of the **probabilistic frame operator**, which is defined thus:

Definition 1.7. Given a measure $\mu \in P_2(\mathbb{R}^d$, its **(probabilistic) frame operator** is S_μ, which for all $y \in \mathbb{R}^d$ satisfies:

$$S_\mu y = \int_{\mathbb{R}^d} \langle x \ , y \rangle \, x \, d\mu(x).$$

Clearly, S_μ may be equated with its matrix representation $\int_{\mathbb{R}^d} xx^\top d\mu(x)$, and then the requirement that the support of μ span $\mathbb{R}^d$ is the same as requiring that this matrix be positive definite. Equivalently, the probabilistic frame definition translates into a requirement on the covariance matrix $\mathrm{Cov}(\mu)$ and mean $\overline{\mu}$ of μ, with:

$$\overline{\mu} := \int_{\mathbb{R}^d} x d\mu(x) \text{ and } \mathrm{Cov}(\mu) := \int_{\mathbb{R}^d} (x - \overline{\mu})(x - \overline{\mu})^\top d\mu(x)$$

First, the mean and the covariance matrix must be well-defined since μ has finite second moment. Second, there must exist $A > 0$ s.t for all $y \in \mathbb{R}^d$,

$$\langle y \ , \mathrm{Cov}(\mu) y \rangle \geqslant A \|y\|^2 - |\langle y \ , \overline{\mu} \rangle|^2.$$

If $\overline{\mu} = 0$, then this second condition is equivalent to requiring that $\mathrm{Cov}(\mu)$ be positive definite. Probabilistic frames for $\mathbb{R}^d$ are clearly a subset of $P_2(\mathbb{R}^d)$ because of the upper frame bound. Let us denote the probabilistic frames for $\mathbb{R}^d$ by $\mathrm{PF}(\mathbb{R}^d)$. Let $\mathrm{PF}(A, B, \mathbb{R}^d)$ denote the set of probabilistic frames in $\mathrm{PF}(\mathbb{R}^d)$ with upper frame bound less than or equal to B and lower frame bound greater than or equal to A. Let $\mathrm{PF}(A, \mathbb{R}^d)$ denote the set of tight frames with frame constant A. Let $\mathrm{DPF}(\mathbb{R}^d)$, $\mathrm{DPF}(A, B, N, \mathbb{R}^d)$, and $\mathrm{DPF}(A, N, \mathbb{R}^d)$ denote the corresponding sets of probabilistic frames with finite support containing at most N elements.

Proposition 1.8. *Given finite $A, B > 0$, $PF(A, B, \mathbb{R}^d)$ and $PF(A, \mathbb{R}^d)$ are nonempty, convex, closed subsets of $P_2(\mathbb{R}^d)$.*

Proof. The nonemptiness is clear: consider the space of nondegenerate, zero-mean Gaussian measures on $\mathbb{R}^d$ whose covariance matrices have maximum eigenvalue B and minimum eigenvalue A. For the convexity: consider $\mu, \nu \in \mathrm{PF}(A, B, \mathbb{R}^d)$, $\lambda \in [0, 1]$. Define $\mu_\lambda = (1 - \lambda)\mu + \lambda\nu$. Given $y \in \mathbb{R}^d$,

$$\begin{aligned}\int_{\mathbb{R}^d} \langle x\ , y\rangle^2 d\mu_\lambda(x) &= (1-\lambda)\int_{\mathbb{R}^d} \langle x\ , y\rangle^2 d\mu(x) + \lambda \int_{\mathbb{R}^d} \langle x\ , y\rangle^2 d\nu(x) \\ &\geqslant (1-\lambda)A\|y\|^2 + \lambda A\|y\|^2 \\ &= A\|y\|^2\end{aligned}$$

The upper bound follows similarly, and the result is clear. Finally, for the closedness, let $\{\mu_n\}$ be a sequence in $\mathrm{PF}(A, B, \mathbb{R}^d)$ converging to $\mu \in P_2(\mathbb{R}^d)$. Since $\int_{\mathbb{R}^d}\langle x\ , y\rangle^2 d\mu(x)$ is a continuous function of $y \in \mathbb{R}^d$, we can define

$$y_0 = \mathrm{argmin}_{y \in S^{d-1}} \int_{\mathbb{R}^d} \langle x\ , y\rangle^2 d\mu(x).$$

Since

$$\langle x\ , y_0\rangle^2 \leqslant \|x\|^2 \|y_0\|^2 \leqslant \|y_0\|^2(1 + \|x\|^2),$$

by the second definition of weak convergence in $P_2(\mathbb{R}^d)$ given in Definition 1.5, $\int_{\mathbb{R}^d}\langle x\ , y_0\rangle^2 d\mu_n(x) \to \int_{\mathbb{R}^d}\langle x\ , y_0\rangle^2 d\mu(x)$. Since for all n, the values of $\int_{\mathbb{R}^d}\langle x\ , y_0\rangle^2 d\mu_n(x)$

are bounded above and below by B and A, respectively, μ is an element of $\mathrm{PF}(A, B, \mathbb{R}^d)$. Taking $A = B$, we also have the closedness of $\mathrm{PF}(A, \mathbb{R}^d)$. $\square$

Remark 1.9. Note that $\mathrm{PF}(\mathbb{R}^d)$ itself is not closed, since one can construct a sequence of probabilistic frames whose lower frame bounds converge to zero: for example, a sequence of zero-mean, Gaussian measures with covariances $\frac{1}{n}I, n \in \mathbb{N}$.

Proposition 1.10. *Given finite $A, B > 0$, DPF$(A, B, N, \mathbb{R}^d)$ and DPF$(A, N, \mathbb{R}^d)$ are closed subsets of $P_2(\mathbb{R}^d)$.*

Proof. Consider a sequence $\{\mu_n\}$ in $\mathrm{DPF}(A, B, N, \mathbb{R}^d)$ converging weakly to μ in $P_2(\mathbb{R}^d)$. By the result above, $\mu \in \mathrm{PF}(A, B, \mathbb{R}^d)$. Thus, it remains to show that the support of μ is discrete and finite, containing at most N elements.

Suppose $\#|\mathrm{supp}(\mu)| = M > N$, possible infinite. Then there exists $\{y_n\}_{n=1}^M \subset \mathrm{supp}(\mu)$ such that for all open subsets U which contain some y_n, $\mu(U) > 0$. Fix $T = N + 1$. Then we have $\epsilon > 0$ such that $\|y_i - y_j\| > 3\epsilon$ for all $i, j \leqslant T$, $i \neq j$.

Define the disjoint open balls $\{B_\epsilon(y_k)\}_{k=1}^T$, ordered such that $\mu(B_\epsilon(y_1)) \geqslant \mu(B_\epsilon(y_2)) \geqslant \cdots \geqslant \mu(B_\epsilon(y_T)) > 0$. Let $\delta = \mu(B_\epsilon(y_T))$. Now, for any $n \in \mathbf{N}$, $\mathrm{supp}(\mu_n)$ contains at most N elements. Therefore, by the pigeonhole principle, for each n there exists a subset $I_n \subset \{1, 2, ..., T\}$ such that $\#|I_n| \geqslant 1$ and $\mathrm{supp}(\mu_n) \bigcap (\cup_{k \in I_n} B_\epsilon(y_k)) = \varnothing$. In particular, for all $x \in \mathrm{supp}(\mu_n)$, $\|x - y_k\| > \epsilon$ for all $k \in I_n$.

Then for all n,

$$\begin{aligned} W_2^2(\mu_n, \mu) &= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 d\gamma_0(x, y) \\ &\geqslant \sum_{k \in I_n} \int_{B_\epsilon(y_k)} \int_{\mathbb{R}^d} \|x - y\|^2 d\gamma_0(x, y) \end{aligned}$$

$$
\begin{aligned}
&\geqslant \sum_{k \in I_n} \int_{B_\epsilon(y_k)} \int_{\mathbb{R}^d} \epsilon^2 d\gamma_0(x, y) \\
&= \epsilon^2 \sum_{k \in I_n} \gamma_0(\mathbb{R}^d \times B_\epsilon(y_k)) \\
&= \epsilon^2 \sum_{k \in I_n} \mu(B_\epsilon(y_k)) \\
&\geqslant \epsilon^2 \cdot \delta, \text{ independent of N}
\end{aligned}
$$

This contradicts the convergence of the sequence, and our result follows. □

1.6 Connection between Probabilistic and Continuous Frames

As detailed in [34], the idea of discrete frames was generalized by Ali, Antoine, and Gazeau to encompass families of elements in some locally compact space possessing a Radon measure, the so-called continuous frames. Square-integrable representations of groups can generate continuous frames by acting on a fixed mother element, and in mathematical physics, these frames are called coherent states and can be carefully chosen to simplify certain problems. Rank-one positive operator valued measures (POVMs) can be written as continuous frames.

In [1], we have the following definition of **continuous frame**:

Definition 1.11. Let X be a metrizable, locally compact space. Let ν be positive, inner regular Borel measure for X supported on all of X. Let H be a Hilbert space. Then a set of vectors $\{\eta_x^i, i \in \{1, \cdots, n\}, x \in X\}$ is a **rank-n (continuous) frame** if, for each $x \in X$, the vectors $\{\eta_x^i, i \in \{1, \cdots, n\}\}$ are linearly independent, and if there exist constants $0 < A \leqslant B < \infty$ such that $\forall f \in H$,

$$A\|f\|^2 \leqslant \sum_{i=1}^{n} \int_X |\langle \eta_x^i \ , f \rangle|^2 d\nu(x) \leqslant B\|f\|^2.$$

With this definition in mind, we can detail the following simple relationship between continuous frames and probabilistic frames.

Proposition 1.12. *Any probabilistic frame can be written as a rank-one continuous frame.*

Proof. Let $\mu \in P_2(\mathbb{R}^d)$ be a probabilistic frame. The support of μ is a closed subset of $\mathbb{R}^d$, so that we can take $X = \text{supp}(\mu)$ in the above definition. Then, clearly, with $n = 1$ and $\eta_x = x$, $\{x\}$, $x \in \text{supp}(\mu) \subset \mathbb{R}^d$ is trivially a continuous frame. $\square$

Remark 1.13. Conversely, let $\{\eta_x\}$ be a rank-one continuous frame for $(X = \nu)$, where X is some metrizable, locally compact space, and ν is a finite, positive, regular Borel measure. Let $\beta = \nu(X)$. Take $H = \mathbb{R}^d$ and consider $T : \mathcal{X} \to \mathbb{R}^d$, $T(x) := \eta_x$. $\mu := T_{\#}(\frac{1}{\beta}\nu)$ is then a probabilistic frame for $\mathbb{R}^d$, since for any $A \in B(\mathbb{R}^d$,

$$0 \leqslant \mu(A) = \int_{\mathbb{R}^d} \chi_A(y) d\mu(y) = \frac{1}{\beta} \int_X \chi_A(\eta_x) d\nu(x) \leqslant 1,$$

and for any $z \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \langle y \ , z \rangle^2 d\mu(y) = \frac{1}{\beta} \int_X \langle \eta_x \ , z \rangle^2 d\nu(x),$$

This equivalence is not particularly interesting, and, as we shall see in the following chapters, much more can be learned by examining the measure μ and working in the Wasserstein space.

Chapter 2

Elementary Paths in the Space of Probabilistic Frames

2.1 Geodesics for the Wasserstein Space

To investigate the distances between probabilistic frames, we consider geodesics in the Wasserstein space $P_2(\mathbb{R}^d)$; this notion will be crucial later on when we build gradient flows in this space. We identify conditions under which every measure on the geodesic between two probabilistic frames is itself a probabilistic frame, showing that for the case of discrete probabilistic frames, this question can be reduced to one of ranks of convex combinations of matrices. For probabilistic frames with density, we show that continuity of the optimal deterministic coupling is sufficient for geodesic measures to be probabilistic frames. The key results may be found in Theorems 2.13 and 2.30.

2.1.1 Wasserstein Geodesics

To begin, we work with general geodesics in the Wasserstein space. The method, taken from [35], is as follows:

Definition 2.1. Let μ_0 and μ_1 be measures in $P_2(\mathbb{R}^d)$. Define the map $\Pi^t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$\Pi^t(x, y) = (x, (1-t)x + ty) \quad \text{for } t \in [0, 1].$$

Let $\gamma_0 \in \Gamma(\mu_0, \mu_1)$ be an optimal transport plan for μ_0 and μ_1 with respect to the 2-Wasserstein distance.

Define a probability measure γ^t on $\mathbb{R}^d \times \mathbb{R}^d$ by :

$$\iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} F(x,y) d\gamma^t(x,y) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} F(\Pi^t(x,y)) d\gamma_0(x,y)$$

for all $F \in C_b(\mathbb{R}^d \times \mathbb{R}^d)$. Note that $\forall F \in C_b(\mathbb{R}^d)$,

$$\iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} F(x) d\gamma^t(x,y) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} F(x) d\gamma_0(x,y) = \int_{\mathbb{R}^d} F(x) d\mu_0(x).$$

Then, defining μ_t, $t \in [0,1]$, to be the probability measure such that for all $G \in C_b(\mathbb{R}^d)$:

$$\int_{\mathbb{R}^d} G(y) d\mu_t(y) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} G(y) d\gamma^t(x,y) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} G((1-t)x + ty) d\gamma_0(x,y), \qquad (2.1)$$

we call μ_t a measure on a **geodesic** between μ_0 and μ_1, or a **geodesic measure** when the endpoints of the path are clear from the context. We call γ_t a **lifting** of μ_t relative to γ_0.

Proposition 2.2. *Given $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$, the mapping $[0,1] \to P_2(\mathbb{R}^d)$ given by $t \mapsto \mu_t$, as defined in (2.1), is Lipschitz in t.*

Proof. Define $f_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by $f_t(x,y) = (1-t)x + ty$. Take $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ with optimal plan $\gamma_0 \in \Gamma_0(\mu_0, \mu_1)$ and $\mu_t = (f_t)_{\#}\gamma_0$. Then

$$\begin{aligned} W_2^2(\mu_{t_1}, \mu_{t_2}) &= \inf_{\pi \in \Gamma(\mu_{t_1}, \mu_{t_2}} \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \|u - v\|^2 d\pi(u,v) \\ &\leqslant \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \|u - v\|^2 d(f_{t_1}, f_{t_2})_{\#}\gamma_0(u,v) \end{aligned}$$

$$= \iint_{\mathbb{R}^d \times \mathbb{R}^d} \|((1-t_1)x + t_1 y) - (1-t_2)x + t_2 y))\|^2 d\gamma_0(x,y)$$

$$= \iint_{\mathbb{R}^d \times \mathbb{R}^d} (t_2 - t_1)^2 \|x - y\|^2 d\gamma_0(x,y)$$

$$= (t_2 - t_1)^2 W_2^2(\mu_0, \mu_1)$$

where the last equality comes from the fact that γ is an optimal plan for the 2-Wasserstein distance between μ_0 and μ_1. □

Moreover, from [35], we have the following lemma, which justifies our use of the term "geodesic."

Lemma 2.3. *The mapping $t \to \mu_t$ is a geodesic of the 2-Wasserstein distance in the sense that*

$$W_2(\mu_0, \mu_t) + W_2(\mu_t, \mu_1) = W_2(\mu_0, \mu_1).$$

In the rest of this chapter, we consider under what conditions we can construct a "path of frames"–i.e., when are the measures on the geodesic between two probabilistic frames themselves probabilistic frames? Recall that proving that a probability measure μ on $\mathbb{R}^d$ is a probabilistic frame requires showing that it is an element of $P_2(\mathbb{R}^d)$ and that $S_\mu := \mathrm{Cov}(\mu) + \overline{\mu\mu}^\top$ is positive definite. It is easy to show that μ_t, as constructed by the method above, always meets the first requirement.

Lemma 2.4. *For any measure μ_t, $t \in [0,1]$, on the geodesic between two probabilistic frames μ_0 and μ_1 with lifting γ_t relative to an optimal plan γ_0, $M_2^2(\mu_t) < \infty$.*

Proof. Given μ_t as above,

$$
\begin{aligned}
M_2^2(\mu_t) = \int_{\mathbb{R}^d} \|y\|^2 d\mu_t(y) \quad &= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \|y\|^2 d\gamma^t(x,y) \\
&= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \|(1-t)x + ty\|^2 d\gamma_0(x,y) \\
&= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} (1-t)^2\|x\|^2 + t^2\|y\|^2 + 2t(1-t)\langle x\,, y\rangle) d\gamma_0(x,y) \\
&= (1-t)^2 M_2^2(\mu_0) + t^2 M_2^2(\mu_1) + 2t(1-t) \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\,, y\rangle d\gamma_0(x,y)
\end{aligned}
$$

Now,

$$
\int_{\mathbb{R}^d} \|x\| d\mu_0(x) = \left(\int_{\mathbb{R}^d} \|x\|^2 d\mu_0(x)\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} 1 d\mu_0(x)\right)^{\frac{1}{2}},
$$

so

$$
\int_{\mathbb{R}^d} \|x\| d\mu_0(x) \leqslant M_2(\mu_0).
$$

Therefore,

$$
\begin{aligned}
M_2^2(\mu_t) &\leqslant (1-t)^2 M_2^2(\mu_0) + t^2 M_2^2(\mu_1) + 2t(1-t) M_2(\mu_0) M_2(\mu_1) \\
&= ((1-t)M_2(\mu_0) + tM_2(\mu_1))^2 \\
&< \infty.
\end{aligned}
$$

□

The question which remains is showing that $S_{\mu_t} := \int_{\mathbb{R}^d} yy^\top d\mu_t(y)$, the frame operator of μ_t, is positive definite (or, equivalently, that the support of μ_t spans $\mathbb{R}^d$). Different conclusions can be drawn about the lifting of the geodesic depending on the characteristics of the support of the measures at the endpoints. For this reason, we divide much of the remaining analysis into two parts: the discrete case and the absolutely continuous case. In both, we will make use of a monotonicity property (Definition 2.7) that characterizes optimal transport plans. We first address this question for the canonical probabilistic frames associated with finite frames.

2.2 Paths for Discrete Probabilistic Frames

2.2.1 Probabilistic Frames with Discrete Support

To give the most general statement of the discrete case, we give the following definition:

Definition 2.5. Let $\{\alpha_i\}_{i=1}^N$ be a set of nonnegative real numbers satisfying $\sum_{i=1}^N \alpha_i = 1$, and let $\Phi = \{\varphi_i\}_{i=1}^N$ be a probabilistic frame. Then the **canonical** α**-weighted probabilistic frame** for Φ is $\mu_{\Phi,\alpha}$ given by $d\mu_{\Phi,\alpha}(x) = \sum_{i=1}^N \alpha_i \delta_{\varphi_i}(x)$

Now suppose we have two frames $\Phi = \{\varphi_i\}_{i=1}^N$ and $\Psi = \{\psi_j\}_{j=1}^M$, and two sets of nonnegative weights, $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$, summing to unity. Let $\mu_0 = \mu_{\Phi,\alpha}$, and let $\mu_1 = \mu_{\Psi,\beta}$. In this case, any joint distribution γ for μ_0 and μ_1 satisfies

$$d\gamma(x, y) = [\delta_{\varphi_1}(x) \dots \delta_{\varphi_N}(x)]^\top A[\delta_{\psi_1}(y) \dots \delta_{\psi_M}(y)],$$

where $A \in \mathbb{R}^{N\times M}$ with

$$\sum_{i=1}^N A_{i,j} = \beta_j, \quad \sum_{j=1}^M A_{i,j} = \alpha_i, \quad A_{i,j} \geqslant 0 \quad \forall i, j,$$

and $\sum_{i=1}^N \sum_{j=1}^N A_{i,j} = 1$. That is, there is a one-to-one correspondence between $\Gamma(\mu_0, \mu_1)$ and a subset of the $N \times M$ nonnegative matrices whose entries sum to one.

In particular, we have:

Lemma 2.6. [2, Theorem 6.0.1] *Given μ_0 and μ_1 as above, if $M = N$, and if $\alpha_i = \beta_j = \frac{1}{N}$ for all $i, j \in \{1, ..., N\}$, then the Monge-Kantorovich problem becomes the Birkhoff problem, and denoting by $\Gamma(\frac{1}{N})$ the set of matrices with row and column sums identically $\frac{1}{N}$:*

$$W_2^2(\mu_0, \mu_1) = \min_{A \in \Gamma(\frac{1}{N})} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j} \| \varphi_i - \psi_j \|^2$$

and, by the Birkhoff-von Neumann Theorem, the optimal transport matrix A is a permutation matrix corresponding to some $\sigma \in S_N$, i.e.:

$$W_2^2(\mu_0, \mu_1) = \min_{\sigma \in S_N} \frac{1}{N} \sum_{i=1}^{N} \| \varphi_i - \psi_{\sigma(i)} \|^2$$

In this case, for some optimal $\sigma \in S_N$,

$$S_{\mu_t} := \frac{1}{N} \sum_{i=1}^{N} [(1-t)\varphi_i + t\psi_{\sigma(i)}][(1-t)\varphi_i + t\psi_{\sigma(i)}]^\top. \tag{2.2}$$

Note that the optimality of σ implies that σ maximizes $\sum_{i=1}^{N} \langle \varphi_i \ , \psi_{\sigma(i)} \rangle$ among all elements of S_N. This crucial fact motivates the following definition.

Definition 2.7. A set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be **cyclically monotone** if, given any finite subset $\{(x_1, y_1), ..., (x_N, y_N)\} \subset S$, for every $\sigma \in S_N$ holds the inequality:

$$\sum_{i=1}^{N} \langle x_i \ , y_i \rangle \geqslant \sum_{i=1}^{N} \langle x_i \ , y_{\sigma(i)} \rangle.$$

Having defined cyclical monotonicity, it will be useful to note that there are several examples of pairs of frames whose canonical probabilistic frames meet this requirement. First, however, we recall a result of [20], restated for Euclidean space, which gives a useful characterization of frames and their duals:

Lemma 2.8. *[20, Theorem 5.6.5] Let $\{\varphi_i\}_{i=1}^N$ be a frame for $\mathbb{R}^d$ with frame operator S. The dual frames of $\{\varphi_i\}_{i=1}^N$ are precisely the families:*

$$\{\psi_i\}_{i=1}^N = \left\{ S^{-1}\varphi_i + h_i - \sum_{j=1}^N \langle S^{-1}\varphi_i\ , \varphi_j \rangle h_j \right\}_{i=1}^N , \tag{2.3}$$

where $\{h_i\}_{i=1}^N$ is some subset of $\mathbb{R}^d$.

Now, we can proceed to discuss cyclical monotonicity of certain frame pairings:

Lemma 2.9. *If $\{\varphi_i\}_{i=1}^N$ is the canonical dual frame to $\{\psi_i\}_{i=1}^N$, then $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone.*

Proof. Let $S = \Psi^\top \Psi$. Then suppose that $\Phi^\top = S^{-1}\Psi^\top$. For any permutation $\sigma \in S_N$, let P_σ denote the matrix such that for

$$\forall x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N, \quad P_\sigma x = \begin{bmatrix} x_{\sigma(1)} \\ \vdots \\ x_{\sigma(N)} \end{bmatrix}.$$

Then

$$\begin{aligned}
\sum_{i=1}^N \langle \varphi_i\ , \psi_i - \psi_{\sigma(i)} \rangle &= \sum_{i=1}^N \langle S^{-1}\psi_i\ , \psi_i - \psi_{\sigma(i)} \rangle \\
&= \sum_{i=1}^N (\psi_i - \psi_{\sigma(i)})^\top S^{-1}\psi_i \\
&= \mathrm{Tr}((\Psi - P_\sigma \Psi) S^{-1} \Psi^\top) \\
&= \mathrm{Tr}((I_N - P_\sigma)\Psi S^{-1}\Psi^\top) \\
&= \mathrm{Tr}((I_N - P_\sigma) I_N^d) \\
&\geqslant 0
\end{aligned}$$

We also use the fact that, denoting by I_N^d the $N \times N$ diagonal matrix with d leading ones on the diagonal and zeros else, $\Psi S^{-1}\Psi^\top = I_N^d$ because $S^{-1}\Psi^\top$ is the Moore-Penrose pseudoinverse of Ψ. Therefore, the identity is an optimal permutation, i.e., the set $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone. $\square$

Lemma 2.10. *Let $\{\varphi_i\}_{i=1}^N$ be one of the dual frames to $\{\psi_i\}_{i=1}^N$, as given in Lemma 2.8. Assume that the set $\{h_i\}_{i=1}^N$ is ordered so that $\{(h_i, \psi_i)\}_{i=d+1}^N$ is cyclically monotone. Then $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone.*

Proof. Take $\{\varphi_i\}_{i=1}^N$ to be a dual of the form given in Lemma 2.8. Let W be the matrix whose rows are the $\{h_i\}_{i=1}^N$. Then, noting that $\Phi^\top = (S^{-1}\Psi^\top + W^\top(I_N - \Psi S^{-1}\Psi^\top))$,

$$\begin{aligned}
\sum_{i=1}^N \langle \psi_i - \psi_{\sigma(i)}\ , \varphi_i \rangle &= \mathrm{Tr}((I_N - P_\sigma)\Psi\Phi^\top) \\
&= \mathrm{Tr}((I_N - P_\sigma)\Psi(S^{-1}\Psi^\top + W^\top(I_N - \Psi S^{-1}\Psi^\top))) \\
&= \mathrm{Tr}((I_N - P_\sigma)I_N^d + (I_N - P_\sigma)\Psi W^\top(I_N - I_N^d)) \\
&= \mathrm{Tr}((I_N - P_\sigma)I_N^d) + \sum_{i=d+1}^N \langle \psi_i - \psi_{\sigma(i)}\ , h_i \rangle \\
&\geqslant 0
\end{aligned}$$

Therefore, under these conditions, $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone. $\square$

Finally, we state this last critical lemma before laying out the main results of this section.

Lemma 2.11. *[49, Theorem 2] Let A and B be $m \times n$ complex matrices, $m \geqslant n$. Let $rank(A) = rank(B) = n$. If $B^\dagger A$ has no nonnegative eigenvalues, then every* *matrix in*

$$h(A,B) := \{C : \quad C = (1-t)A + tB, \quad t \in [0,1]\}$$

has rank n. Similarly, if A and B are $n \times n$ complex matrices with rank n, we can define in

$$r(A,B) := \{C : \quad C = (I-T)A + TB\},$$

where T is a real diagonal matrix with diagonal entries in $[0,1]$. Then, if $B^{-1}A$ is a P-matrix–that is, having all principal minors positive–then every matrix in $r(A,B)$ will have rank n.

Given this lemma and the cyclical monotonicity condition, we can state the following proposition which gives sufficient conditions for a geodesic between discrete probability measures in $P_2(\mathbb{R}^d)$ to be a path of frames.

Proposition 2.12. *Let $\{\varphi_i\}_{i=1}^N$ and $\{\psi_i\}_{i=1}^N$ be frames for $\mathbb{R}^d$ with analysis operators Φ and Ψ. If $\Psi^\dagger\Phi$ has no negative eigenvalues, and if $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is a cyclically monotone set, then every measure on the geodesic between μ_Φ and μ_Ψ has support which spans $\mathbb{R}^d$.*

Proof. Note that S_{μ_t}, as defined in equation (2.2), is the frame operator for a new set of vectors, namely $\{(1-t)\varphi_i + t\psi_{\sigma(i)}\}_{i=1}^N$. Therefore, the support of μ_t will span $\mathbb{R}^d$ (equivalently, S_{μ_t} will be positive definite) provided this set of vectors spans $\mathbb{R}^d$. Now, let Φ be the matrix whose rows are the frame vectors $\{\varphi_i^\top\}$, and let Ψ be the matrix whose rows are the frame vectors $\{\psi_j^\top\}$. As was done in Lemma 2.9, define P_σ to be the $N \times N$ permutation matrix corresponding to $\sigma \in S_N$, where now σ is the optimal permutation for the Wasserstein distance. Let Ψ_σ be $P_\sigma \Psi$. In a slightly more concise way, we can write:

$$S_{\mu_t} = \frac{1}{N}\left((1-t)\Phi^\top + t\Psi_\sigma^\top\right)\left((1-t)\Phi + t\Psi_\sigma\right).$$

Ψ and Ψ_σ have rank d, and to show that S_{μ_t} is positive definite, we must prove that every matrix in the set $h(\Phi, \Psi_\sigma) := \{(1-t)\Phi + t\Psi_\sigma\}_{t\in[0,1]}$ has rank d. By Lemma 2.11, a sufficient condition for this to be true is that $\Psi_\sigma^\dagger \Phi$ be positive semi-definite, where $\Psi_\sigma^\dagger$ is the Moore-Penrose pseudoinverse of Ψ_σ. Finally, we note that if $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is a cyclically monotone set, then $P_\sigma = I$, the identity, is an optimal permutation, and then $\Psi_\sigma^\dagger \Phi = \Psi^\dagger \Phi$ is positive definite by assumption. □

By combination of Lemma 2.4 and Proposition 2.12, we have this result:

Theorem 2.13. *Let $\{\varphi_i\}_{i=1}^N$ and $\{\psi_i\}_{i=1}^N$ be frames for $\mathbb{R}^d$. If $\Psi^\dagger\Phi$ has no negative eigenvalues and $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone, then every measure on the geodesic between the canonical probabilistic frames μ_Φ and μ_Ψ is a probabilistic frame.*

These conditions hold for certain dual frame pairs, as described in the next proposition.

Proposition 2.14. *If $\{\varphi_i\}_{i=1}^N$ is the canonical dual frame to $\{\psi_i\}_{i=1}^N$, or if $\{\varphi_i\}_{i=1}^N$ is a dual frame to $\{\psi_i\}_{i=1}^N$ of the form given in (2.3), such that the $\{h_i\}_{i=1}^N$ is ordered so that $\{(h_i, \psi_i)\}_{i=d+1}^N$ is cyclically monotone, then $\Psi_\sigma^\dagger\Phi$ is positive definite, where σ is the optimal permutation.*

Proof. By definition,

$$\Psi_\sigma^\dagger = (P_\sigma \Psi)^\dagger = (\Psi^\top P_\sigma^\top P_\sigma \Psi)^{-1}\Psi^\top P_\sigma^\top = (\Psi^\top \Psi)^{-1}\Psi^\top P_\sigma^\top$$

Note that this is a permutation of the matrix whose columns are the elements of the canonical dual frame of the rows of Ψ_σ. If $\{\varphi_i\}_{i=1}^N$ is any dual of $\{\psi_i\}_{i=1}^N$, then $\Psi^\top \Phi = I_d$, and therefore, if σ is the identity, then $\Psi_\sigma^\dagger \Phi = (\Psi^\top\Psi)^{-1}$, which is positive definite. It remains to show that the optimal permutation is the identity. Lemma 2.9 shows that if $\{\varphi_i\}_{i=1}^N$ is the canonical dual to $\{\psi_i\}_{i=1}^N$, then $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone; Lemma 2.10 shows that if $\{\varphi_i\}_{i=1}^N$ is any dual to $\{\psi_i\}_{i=1}^N$ which meets the above condition, then $\{(\varphi_i, \psi_i)\}_{i=1}^N$ is cyclically monotone. $\square$

Given the preceding results involving the support of the lifting of the geodesic, we note that it may be profitable to consider the frame operator for the optimal transport plan γ_0 between two discrete probabilistic frames μ_Φ and μ_Ψ:

$$S_{\gamma_0} := \iint_{\mathbb{R}^d \times \mathbb{R}^d} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x^\top y^\top \end{bmatrix} d\gamma_0(x, y).$$

This operator has the form:

$$S_{\gamma_0} = \begin{bmatrix} \Phi^\top \Phi & \Phi^\top \Psi_\sigma \\ \Psi_\sigma^\top \Phi & \Psi^\top \Psi \end{bmatrix}.$$

Moreover, it has the property that for all $x \in \mathbb{R}^d$,

$$\langle x\ , S_{\mu_t} x\rangle = [(1-t)x^\top\ tx^\top] S_{\gamma_0} \begin{bmatrix} (1-t)x \\ tx \end{bmatrix}. \tag{2.4}$$

Thus, given $t \in [0,1]$, S_{μ_t} is positive definite on $\mathbb{R}^d$ if S_{γ_0} is positive definite on the subspace

$$W_t := \{[(1-t)x \; tx]^\top | x \in \mathbb{R}^d\} \subset \mathbb{R}^d \times \mathbb{R}^d.$$

For a class of special discrete probabilistic frames for $\mathbb{R}^d$ with $N \geqslant 2d$ which meet the condition defined below, we can show this is the case.

Definition 2.15. A frame $\{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$ is **full-spark** if every d-element subset contained in it is linearly independent.

Proposition 2.16. *Let $N \geqslant 2d$. Let $\{\varphi_i\}_{i=1}^N$ be a frame for $\mathbb{R}^d$, ordered such that $\{\varphi_i\}_{i=1}^d$ is linearly independent. Let $\{\psi_j\}_{j=1}^N$ be a full-spark frame for $\mathbb{R}^d$. Moreover, let these two sets have the property that for all subsets $J = \{j_1, ..., j_d\} \subset \{1, ..., N\}$, if there exist $\{\alpha_i\}_{i=1}^N$, not all zero, such that $\sum_{i=1}^d \alpha_i \varphi_i = \varphi_l$ for some $l > d$, then $\sum_{i=1}^d \alpha_i \psi_{j_i} \neq \psi_k$ for all $k \in \{1, ..., N\} \backslash J$. Then every measure on the geodesic between μ_Φ and μ_Ψ is a probabilistic frame.*

Proof. Given $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^N$ as above, supporting canonical discrete probabilistic frames μ_Φ and μ_Ψ, let $\gamma_0 \in P_2(\mathbb{R}^d \times \mathbb{R}^d)$ be the optimal transport plan for the Wasserstein distance. Then we can write

$$S_{\gamma_0} = \begin{bmatrix} \Phi^\top \\ \\ \Psi_\sigma^\top \end{bmatrix} [\Phi \; \Psi_\sigma],$$

which is the frame operator for the set $\{\gamma_i^\sigma\}_{i=1}^N := \left\{ \begin{bmatrix} \varphi_i \\ \\ \psi_{\sigma(i)} \end{bmatrix} \right\}_{i=1}^N$. If this set is a frame for W_t, or more generally for $\mathbf{R}^{2d}$, then the result follows. Recall that the

$\{\varphi_i\}_{i=1}^N$ are ordered such that the $\{\varphi_i\}_{i=1}^d$ are linearly independent. Therefore, the set $\{\gamma_i^\sigma\}_{i=1}^d$ spans a d-dimensional subspace of $\mathbb{R}^d \times \mathbb{R}^d$.

Since $\{\psi_j\}_{j=1}^N$ is full spark, it is guaranteed that $\{\psi_{\sigma(i)}\}_{i=d+1}^{2d}$ is linearly independent, and it follows that $\{\gamma_i^\sigma\}_{i=d+1}^{2d}$ spans a d-dimensional subspace for any $\sigma \in S_N$. Thus, it remains to show that for any σ, $\operatorname{span}\{\gamma_i^\sigma\}_{i=1}^d \bigcap \operatorname{span}\{\gamma_i^\sigma\}_{i=d+1}^{2d} = \{0\}$. To do so, define $J = \{\sigma(i)\}_{i=d+1}^{2d}$. It will be sufficient to show that for all $j \in \{d+1, \cdots, 2d\}$, $\gamma_j^\sigma \notin \operatorname{span}\{\gamma_i^\sigma\}_{i=1}^d$.

Since $\{\varphi_i\}_{i=1}^d$ is a basis for $\mathbb{R}^d$, for each φ_l, $l > d$, there exists a unique set of coefficients $\{\alpha_i^l\}_{i=1}^d$ such that $\sum_{i=1}^d \alpha_i^l \varphi_i = \varphi_l$. By assumption, since $\sigma(l) \in J$, it cannot be the case that $\sum_{i=1}^d \alpha_i^l \psi_{j_i} = \psi_{\sigma(l)}$. Hence, there does not exist a set of coefficents $\{\alpha_i^l\}_{i=1}^d$ for any $l > d$ such that $\sum_{i=1}^d \alpha_i^l \gamma_i^\sigma = \gamma_l^\sigma$. Hence $\gamma_l^\sigma \notin \operatorname{span}\{\gamma_i^\sigma\}_{i=1}^d$ for any $l > d$, and our result is proven. $\square$

2.2.2 Examples for Discrete Probabilistic Frames

To construct some simple examples, we shall call upon the following lemma from [41]:

Lemma 2.17. *[41, Proposition 6.4] Let $\{\varphi_i\}_{i=1}^N$ be a frame for a Hilbert Space H, and let S_Φ be its frame operator. Denote by Φ the analysis operator for this frame. Then $\{\psi_i\}_{i=1}^N$ is a dual frame of $\{\varphi_i\}_{i=1}^N$ if and only if there exists a sequence $\{\zeta_i\}_{i=1}^N$ with analysis operator Z such that for each i, $\psi_i = S_\Phi^{-1}\varphi_i + \zeta_i$ and for which $\Phi(H) \perp Z(H)$–that is, for all $u, v \in H$, $\langle \Phi u\ , Zv \rangle = \sum_{i=1}^N \langle u\ , \varphi_i \rangle \langle v\ , \zeta_i \rangle = 0$.*

We also define a type of dual-frame pairing:

Definition 2.18. Let H and K be Hilbert Spaces, and consider the finite sequences $\{\varphi_i\}_{i=1}^N \subset H$ and $\{\psi_i\}_{i=1}^N \subset K$ with respective analysis operators Φ and Ψ. Then the finite sequences are **disjoint** if $\Phi(H) \bigcap \Psi(K) = \{0\}$. They are **orthogonal** if $\Phi^*\Psi = 0$.

Proposition 2.19. *If $\{\varphi_i\}_{i=1}^N$ and $\{\psi_i\}_{i=1}^N$ are disjoint frames for $\mathbb{R}^d$, associated canonically with the probabilistic frames μ_Φ and μ_Ψ, then every measure on the geodesic between μ_Φ and μ_Ψ is a probabilistic frame.*

Proof. Given $v \in \mathbb{R}^d$, consider:

$$\begin{aligned}
\sum_{i=1}^{N} \langle v\ , (1-t)\psi_i + t\psi_i \rangle^2 &= \sum_{i=1}^{N} \langle v\ , (1-t)\Psi^* e_i + t\Psi^* e_i \rangle^2 \\
&= \sum_{i=1}^{N} \langle (1-t)\Psi v + t\Psi v\ , e_i \rangle^2 \\
&= \|(1-t)\Psi v + t\Psi v\,\|_{\mathbf{R}^N}^2 \\
&\geqslant C\big[(1-t)^2\|\Psi v\,\|^2 + t^2\|\Psi v\,\|^2\big]
\end{aligned}$$

for some $C > 0$. Since the two sequences in question are finite frames, choosing the minimum of the two lower frame bounds, say A_0, we can bound the last quantity below by $(1-2t+2t^2)C \cdot A_0\|v\,\|^2$ and obtain our result. □

2.2.3 Nongeodesic Paths between Discrete Probabilistic Frames

Given two frames $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^M$ with analysis operators Φ and Ψ, we wish to characterize paths between the canonical probabilistic frames supported on them. In this section we consider the equal-cardinality case.

Definition 2.20. Let $L(t)$, $t \in [0,1]$ be an $N \times N$ diagonal matrix with diagonal entries $d_i(t)$ satisfying $L(0) = I$, $L(1) = 0$, and $d_i(t) \in (0,1)$ $\forall t \in (0,1)$ and $i \in \{1, \cdots, N\}$. Then the interpolating frame is $\Theta(\Phi, \Psi, L(t)) = \{\theta_i^t\}_{i=1}^N$ by $\theta_i^t = (1 - d_i(t))\psi_i + d_i(t)\varphi_i$.

Proposition 2.21. *If $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^N$ are dual frames, then $\Theta(\Phi, \Psi, L(t))$ will be a frame for all $t \in (0,1)$.*

Proof. The sets $\{\theta_i^t\}_{i=1}^N$ will be frames provided that their analysis operator, the matrix $(I - L(t))\Psi + L(t)\Phi$,is full rank. By Lemma 2.11, this holds provided that $\Phi^\dagger\Psi$ is a P-matrix, which is certainly true if the two frames are dual to one another. □

In the same spirit as Proposition 2.19, we can state the following proposition:

Proposition 2.22. *Let $L(t) = (1-t)I$. Given orthogonal frames $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^N$, and their canonical duals $\{\tilde{\varphi}_i\}_{i=1}^N$ and $\{\tilde{\psi}_j\}_{j=1}^N$, for each $t \in (0,1)$, $\Theta(\Phi, \Psi, L(t))$ and $\Theta(\tilde{\Phi}, \tilde{\Psi}, \hat{L}(t))$ will be dual to each other. Here, $\hat{L}(t) = \frac{1}{\sqrt{1-2t+2t^2}} L(t)$.*

Proof. Let Θ_t be the frame operator for $\Theta(\Phi, \Psi, L(t))$, and let $\tilde{\Theta}_t$ be the analysis operator for $\Theta(\tilde{\Phi}, \tilde{\Psi}, \hat{L}(t))$. Denoting by S_Φ the frame operator of $\{\varphi_i\}$, the synthesis operator of any dual to $\{\varphi_i\}$ can be written: $\hat{\Phi}_W^\top = S_\Phi^{-1}\Phi^\top + W(I - \Phi S_\Phi^{-1}\Phi^\top)$, where $W \in \mathbf{R}^{d\times N}$. In the case that $\hat{\Phi}$ is the canonical dual to Φ, $W = 0$, and $\tilde{\Phi}^\top\Psi = S_\Phi^{-1}\Phi^\top\Psi = 0$. Then:

$$\begin{aligned}
\tilde{\Theta}_t^\top\Theta_t &= \frac{1}{1-2t+2t^2}(t\Psi + (1-t)\Phi)^\top(t\tilde{\Psi} + (1-t)\tilde{\Phi}) \\
&= \frac{1}{1-2t+2t^2}((t^2\Psi^\top\tilde{\Psi}) + (1-t)^2(\Phi^\top\tilde{\Phi}) + t(1-t)(\Psi^\top\tilde{\Phi} + \Phi^\top\tilde{\Psi}) \\
&= \frac{1}{1-2t+2t^2}(t^2 + (1-t)^2)I
\end{aligned}$$

$$= I$$

$\square$

Remark 2.23. In fact, if instead of the canonical duals we chose $\hat{\Phi}_{W_1}$ and $\hat{\Psi}_{W_2}$ with $W_1\Psi = 0$ and $W_2\Phi = 0$, then the result would still hold. Additionally, note that two orthogonal frames always share the dual frame $\{S_\Phi^{-1}\varphi_i + S_\Psi^{-1}\psi_i\}_{i=1}^N$, and this frame will be dual to $\Theta(\Phi, \Psi, L(t))$ for all $t \in (0, 1)$.

2.3 Paths for Probabilistic Frames with Density

2.3.1 Absolutely Continuous Probabilistic Frames

The question of the nature of the optimal transport plan for the 2-Wasserstein distance is simpler for absolutely continuous measures. From [2, Theorem 6.2.10 and Proposition 6.2.13], which gather together a long list of characteristics, we can extract two key facts about this plan, which we collect in the following lemma. First, a definition:

Definition 2.24. $\mu \in \mathcal{P}(X)$ is a Gaussian regular measure, written $\mu \in \mathcal{P}^r(X)$, if $\mu(B) = 0$ for any Gaussian null set B. When $X = \mathbb{R}^d$, these coincide with the sets of Lebesgue-measure zero.

Lemma 2.25. *[2, Chapter 6.2.3] If μ_0 and μ_1 are Gaussian regular measures in $P_2(\mathbb{R}^d)$, then there exists a unique optimal transport plan for the 2-Wasserstein distance which is induced by a transport map r. This transport map is defined* *(and injective) μ_0-a.e. Indeed, there exists a μ_0-negligible set $N \subset \mathbb{R}^d$ such that $\langle r(x_1) - r(x_2)\ , x_1 - x_2 \rangle > 0$ for all $x_1, x_2 \in \mathbb{R}^d \backslash N$.*

Then we have the following result for absolutely continuous probabilistic frames:

Proposition 2.26. *If μ_0 and μ_1 are absolutely continuous (with respect to Lebesgue measure) probabilistic frames for which there exists a linear, positive semi-definite deterministic coupling which minimizes the Wasserstein distance, then all measures on the geodesic between these frames have support which spans $\mathbb{R}^d$ and will therefore be probabilistic frames.*

Proof. Given the assumptions, let $r(x)$ denote the linear transformation which induces the coupling $\mu_1 = r_{\#}\mu_0$. Defining $h_t(x) = (1-t)x + tr(x)$ μ_0-a.e., the geodesic measure is given by

$$\mu_t := h_{t\#}\mu_0. \tag{2.5}$$

Then $S_{\mu_t} = \int_{\mathbb{R}^d} h_t(x)h_t(x)^\top d\mu_0(x)$. If $r(x) = Ax$ for some $A \in \mathbf{A}^{d\times d}$, then:

$$\begin{aligned} S_{\mu_t} &= \int_{\mathbb{R}^d} ((1-t)Ix + tAx)((1-t)Ix + tAx)^\top d\mu_0(x) \\ &= ((1-t)I + tA)S_{\mu_0}((1-t)I + tA)^\top \end{aligned}$$

Since A must be nonsingular–recall that $S_{\mu_1} = AS_{\mu_0}A^\top$, which is certainly of rank d–by Lemma 2.11, $(1-t)I + tA$ will also nonsingular for all $t \in [0,1]$ provided that A has no negative eigenvalues, as we assumed. □

Example 2.1. *An example in which the assumptions of the above proposition hold is the case of nondegenerate Gaussian measures on $\mathbb{R}^d$. Let μ_0 and μ_1 be zero-mean Gaussians. Let $r(x) = S_{\mu_1}^{\frac{1}{2}} S_{\mu_0}^{-\frac{1}{2}} x$. r is a positive definite linear deterministic* *coupling of μ_0 and μ_1. According to a result in [26], if X and Y are two zero-mean random vectors with covariances Σ_X and Σ_Y, respectively, then a lower bound for $E(\|X-Y\|^2)$ is $Tr[\Sigma_X+\Sigma_Y-2(\Sigma_X\Sigma_Y)^{\frac{1}{2}}]$, and the bound is attained, for nonsingular Σ_X, when $Y = \Sigma_X^{-\frac{1}{2}}\Sigma_Y^{\frac{1}{2}}X$. But this simply states that a general lower bound exists for the square of the 2-Wasserstein distance between two probability measures, and that is obtained when at least one is nonsingular and they are related by the linear deterministic coupling given above. Therefore, the coupling given for the Gaussian measures is optimal.*

2.3.2 Injectivity of Transport for Probabilistic Frames with Density

Now, given absolutely continuous probabilistic frames μ, ν for $\mathbb{R}^d$, take $r(x)$ to be the optimal transport map pushing μ to ν guaranteed by Lemma 2.25. Define

$$h_t(x) = (1-t)x + tr(x) \quad \text{for } t \in [0,1];$$

then $S_{\mu_t} = \int h_t(x) \otimes h_t(x) d\mu(x)$, with $\mu_t = (h_t)_{\#}\mu$.

Proposition 2.27. *Given two such probabilistic frames, there exists a set N with $\mu(N) = 0$ such that h_t is injective for all $t \in [0,1]$ on supp$(\mu)\backslash N$.*

Proof. Given $x, y \in \mathbb{R}^d\backslash N$, with N as defined in Lemma 2.25, suppose $h_t(x) = h_t(y)$ for some $t \in [0,1]$. Then, since:

$$\begin{aligned} 0 &= \langle h_t(x) - h_t(y) \ , x-y \rangle \\ &= \langle (1-t)(x-y) + t(r(x)-r(y)) \ , x-y \rangle \\ &= (1-t)\|x-y\|^2 + t\langle r(x)-r(y) \ , x-y \rangle \end{aligned}$$

it follows that

$$\langle r(x) - r(y)\ , x - y \rangle = \frac{t-1}{t} \|x - y\|^2.$$

This implies that $\langle r(x)-r(y)\ , x-y \rangle \leqslant 0$. However, from the proposition above, we also know that $\langle r(x) - r(y)\ , x - y \rangle \geqslant 0$. Therefore $\|x - y\| = 0$, and h_t is injective on N. $\square$

For the next result, we shall need the following lemma from [24, Theorem 3.3], which builds on the results in [11].

Lemma 2.28. Regularity Result, [24, Theorem 3.3] *Let $\mu, \nu \in P_2^r(\mathbb{R}^d)$, and let r be the unique optimal transport map relative to the cost $c(x, y) = \frac{\|x-y\|^2}{2}$. Define densities such that $d\mu(x) = f(x)dx$, $d\nu(x) = g(x)dx$. Let*

$$X = \{x \in \mathbb{R}^d : f(x) > 0\}, \quad Y = \{x \in \mathbb{R}^d : g(x) > 0\}$$

be two bounded open sets. Then if f and g are bounded away from zero and infinity on X and Y, and Y is convex, it follows that r is continuous.

Corollary 2.29. *Let $\mu, \nu \in P_2^r(\mathbb{R}^d)$, and let r be the unique optimal transport map relative to the cost $c(x, y) = \frac{\|x-y\|^2}{2}$. Then if μ and ν are supported on bounded convex subsets of $\mathbb{R}^d$, r is continuous.*

In general, regularity results swiftly become more complicated as the underlying space changes or the cost functional become less friendly. We note that if we relax the convexity requirement on Y, then we obtain regularity up to sets of measure zero in X and Y ([25], Theorem 1.3). The purpose of the inclusion of this information is simply to show that the conditions of the following theorem can be met.

Theorem 2.30. *Let $\mu, \nu \in P_2^r(\mathbb{R}^d)$, and let r be the unique optimal transport map relative to the cost $c(x,y) = \frac{\|x-y\|^2}{2}$. Let N be the set of measure zero define in Proposition 2.27 If r is continuous, and if supp$(\mu)\backslash N$ contains an open set, then every geodesic measure μ_t is a probabilistic frame.*

Proof. Now, since r is continuous and, by proposition 2.27, monotone outside a set N of measure zero, so is h_t for each t. Let $x_0 \in \text{supp}(\mu)\backslash N$. First, we show that for any $\epsilon > 0$, $h_t^{-1}(B_\epsilon(h_t(x_0)))$ contains an open set containing x_0.

Since h_t is continuous at any such x_0, given $\epsilon > 0$, there exists $\delta > 0$ such that $\forall x \in B_\delta(x_0)$, $\|h_t(x) - h_t(x_0)\| < \epsilon$. Hence for any $x \in B_\delta(x_0)$, $x \in h_t^{-1}(B_\epsilon(h_t(x_0)))$– i.e., $B_\delta(x_0) \subset h_t^{-1}(B_\epsilon(h_t(x_0)))$.

Then $\forall x_0 \in \text{supp}(\mu)\backslash N$, consider $B_{\frac{1}{k}}(h_t(x_0))$:

$$\begin{aligned}
\mu_t(B_{\frac{1}{k}}(h_t(x_0))) &= \int \mathbb{1}_{\left[B_{\frac{1}{k}}(h_t(x_0))\right]}(h_t(y))d\mu(y) \\
&= \int \mathbb{1}_{\left[h_t^{-1}(B_{\frac{1}{k}}(h_t(x_0)))\right]}d\mu(y) \\
&= \mu(h_t^{-1}(B_{\frac{1}{k}}(h_t(x_0)))) \\
&> 0
\end{aligned}$$

where the last inequality holds since $x_0 \in \text{supp}(\mu)$ and, as shown above, $h_t^{-1}(B_{\frac{1}{k}}(h_t(x0))))$ contains an open set containing x_0. Thus, we have shown that for any $k \in \mathbb{N}$, the open ball of radius $\frac{1}{k}$ around $h_t(x_0)$ has positive μ_t-measure, and therefore $h_t(x_0)$ lies in supp(μ_t). Thus $h_t(\text{supp}(\mu)\backslash N) \subset \text{supp}(\mu_t)$.

Therefore, since h_t is injective and continuous on $\text{supp}(\mu)\backslash N$ and by assumption, there exists open set $U \subset \text{supp}(\mu)\backslash N$, by invariance of domain, $h_t(U) \subset \text{supp}(\mu_t)$ is open, and therefore $h_{t\#}\mu$ has support which spans $\mathbb{R}^d$. $\square$

Chapter 3

Duality, Analysis, and Synthesis

3.1 Duality

In this chapter, we explore the familiar concept of duality, analysis, and synthesis that are well-understood in finite frame theory, using the extra flexibility of the probabilistic setting to extend their definitions. The key ideas, results, and examples may be found in Definition 3.1, Proposition 3.16, Definition 3.27, and the examples.

3.1.1 Definition of Duality and Properties

Definition 3.1. Given a probabilistic frame μ for $\mathbb{R}^d$, we define the set of **transport duals** to μ to be

$$D_\mu := \left\{ \nu \in P_2(\mathbb{R}^d) \quad \middle| \quad \exists \gamma \in \Gamma(\mu, \nu) \text{ with } \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} xy^\top d\gamma(x, y) = I \right\}.$$

We denote the set of joint distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ ($\pi^1_\# \gamma = \mu$) for which $\iint_{\mathbb{R}^d \times \mathbb{R}^d} xy^\top d\gamma(x, y) = I$ by ΓD_μ.

The restriction of the set of transport duals D_μ to lie inside $P_2(\mathbb{R}^d)$ is necessary, unlike in the finite frame case. One might consider the following simple example:

Example 3.1. *Let* $\{e_i\}_{i=1}^d \subset \mathbb{R}^d$ *denote the standard orthonormal basis. Let* $\{\varphi_i\}_{i=1}^{d+1}$ *be given by* $\varphi_i = \sqrt{i2^i} e_i$, $i \in \{1, \cdots, d\}$, *and let* $\varphi_{d+1} = 0$. *Take the weights* $\alpha_i =$

$\frac{1}{2^i}, i \in \mathbb{N}$, and let $\alpha_0 = 1 - \sum\limits_{i=1}^{d} \frac{1}{2^i}$. Define

$$\mu_1 = \alpha_0\delta_0 + \sum_{i=1}^{d} \alpha_i \delta_{\varphi_i}.$$

Let $\{\psi_i\}_{i=1}^{\infty}$ be given by $\psi_i = \sqrt{\frac{2^i}{i}} e_{[(i-1) \mod d)+1]}, \quad i \in \mathbb{N}$. Let

$$\mu_2 = \sum_{i=1}^{\infty} \alpha_i \delta_{\psi_i}.$$

Then $\mu_1 \in P_2(\mathbb{R}^d)$, but

$$M_2^2(\mu_2) = \sum_{i=1}^{\infty} \frac{1}{2^i} \|\psi_i\|^2 = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{2^i}{i} = \infty.$$

Hence, $\mu_2 \notin P_2(\mathbb{R}^d)$. However, letting $\gamma \in P(\mathbb{R}^d \times \mathbb{R}^d)$ be given by

$$\gamma = \sum_{i=1}^{d} \alpha_i \delta_{(\varphi_i, \psi_i)} + \sum_{i=d+1}^{\infty} \alpha_i \delta_{(\varphi_i, 0)},$$

it is clear that $\gamma \in \Gamma(\mu_1, \mu_2)$, and

$$\iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} xy^\top d\gamma(x,y) = \sum_{i=1}^{d} \frac{1}{2^i} \sqrt{i2^i} \sqrt{\frac{2^i}{i}} e_i e_i^\top = I.$$

However, once we have this Bessel-like restriction on the class of transport duals, we can assert the following proposition:

Proposition 3.2. *Let μ be a probabilistic frame, and take $\nu \in D_\mu$. Then ν is also a probabilistic frame.*

Proof. Since $D_\mu \subset P_2(\mathbb{R}^d)$ by definition, it is sufficient to show that *supp*(ν) spans $\mathbb{R}^d$. Let us assume the contrary. There exists some $\gamma \in \Gamma(\mu, \nu)$ such that $\iint xy^\top d\gamma(x,y) = I$. Suppose there exists $z \in \mathbb{R}^d$, $z \neq 0$, such that $z \perp w$ for all $w \in \text{span}(\text{supp}(\nu))$. Then for all $x \in \text{supp}(\nu)$, $z^\top x = 0$. Then

$$\|z\|^2 = \iint \langle z\ , x \rangle \langle z\ , y \rangle d\gamma(x,y)$$

$$= \iint \langle z\ , x\rangle\langle z\ , y\rangle \mathbb{1}_{[\mathrm{supp}(\nu)\times\mathbb{R}^d]}(x,y)d\gamma(x,y)$$

$$= 0$$

Thus, by contradiction we have our result. □

Remark 3.3. The transport plan for the canonical dual to a probabilistic frame is

$$\gamma = (\iota \times S_\mu^{-1})_\# \mu,$$

i.e.,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} xy^\top d\gamma(x,y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} x(S_\mu^{-1}x)^\top d\mu(x) = S_\mu S_\mu^{-1} = I.$$

This was the only type of duality defined in [29–31].

Proposition 3.4. *Given $\mu \in PF(\mathbb{R}^d)$, D_μ is a closed subset of $P_2(\mathbb{R}^d)$ with respect to the weak topology.*

Proof. Let $\mu \in \mathrm{PF}(\mathbb{R}^d)$. Suppose ν_n is a sequence of duals to μ converging weakly to some ν in $P_2(\mathbb{R}^d)$. Let $P := \{\nu_n\}_{n\in\mathbb{N}} \cup \{\nu\}$ and $Q := \{\mu\}$. Then P and Q are tight, so $\Gamma(P,Q)$ is tight in $P_2(\mathbb{R}^d\times\mathbb{R}^d)$ and therefore precompact for the weak topology. Let $\{\gamma_n\}$ be a sequence of joint measures yielding the duality. Since $\{\gamma_n\} \subset \Gamma(P,Q)$, there exists a subsequence $\{\gamma_{n_k}\}$ converging weakly in $P_2(\mathbb{R}^d\times\mathbb{R}^d)$ to some γ. First, we show that $\gamma \in \Gamma(\mu,\nu)$: For all $\varphi \in C_b(\mathbb{R}^d\times\mathbb{R}^d)$,

$$\iint \varphi(x,y)d\gamma_{n_k}(x,y) \longrightarrow \iint \varphi(x,y)d\gamma(x,y).$$

In particular, for all $\psi \in C_b(\mathbb{R}^d)$,

$$\iint \psi(x)d\gamma_{n_k}(x,y) \longrightarrow \iint \psi(x)d\gamma(x,y)$$

and

$$\iint \psi(y)d\gamma_{n_k}(x,y) \longrightarrow \iint \psi(y)d\gamma(x,y).$$

But since for all k, $\iint \psi(x)d\gamma_{n_k}(x,y) = \int \psi(x)d\mu(x)$, it follows that

$$\iint \psi(x)d\gamma(x,y) = \int \psi(x)d\mu(x),$$

so $\pi_{1\#}\gamma = \mu$. Similarly, for all k, $\pi_{2\#}\gamma_{n_k} = \nu_{n_k}$, and

$$\lim_k \iint \psi(y)d\gamma_{n_k}(x,y) = \lim_k \int \psi(y)d\nu_{n_k}(y)$$

and since $\{\nu_n\}$ is a weakly convergent sequence, $\nu_{n_k} \xrightarrow{w} \nu$, passing to the limit on both sides,

$$\iint \psi(y)d\gamma(x,y) = \int \psi(y)d\nu(y).$$

Thus, $\gamma \in \Gamma(\mu,\nu)$

Then, for all $\varphi \in C(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying for some $C > 0$ $|\varphi(x,y)| \leqslant C(1 + \|x\|^2 + \|y\|^2)$,

$$\iint \varphi(x,y)d\gamma_{n_k}(x,y) \longrightarrow \iint \varphi(x,y)d\gamma(x,y).$$

Since $|x_i y_j| \leqslant \frac{1}{2}(\|x\|^2 + \|y\|^2)$, it follows that

$$\iint x_i y_j d\gamma_{n_k}(x,y) \longrightarrow \iint x_i y_j d\gamma(x,y).$$

Then, since for each n_k, $\iint x_i y_j d\gamma_{n_k}(x,y) \equiv \delta_{i,j}$, it follows that $\iint x_i y_j d\gamma(x,y) = \delta_{i,j}$, and therefore $\nu \in D_\mu$. $\square$

As a corollary we have:

Proposition 3.5. *Given $\mu \in PF(\mathbb{R}^d)$, D_μ is a closed subset of $PF(\mathbb{R}^d)$ with respect to the weak topology on $P_2(\mathbb{R}^d)$.*

Proof. Given the above result about the closedness of D_μ in $P_2(\mathbb{R}^d)$, this follows fom Proposition 3.2. $\square$

Using Proposition 3.1.1, we can then prove weak compactness of the set of transport duals.

Theorem 3.6. *Given $\mu \in PF(\mathbb{R}^d)$, D_μ is a compact subset of $P_2(\mathbb{R}^d)$ with respect to the weak topology.*

Proof. Consider the lifting of the dual set, $\Gamma D\mu := \{\gamma \in \Gamma(\mu, \nu)$ s.t. $\iint xy^\top d\gamma(x, y) = I\}$. Since $\{\mu\}$ is tight, given $\epsilon > 0$, there exists a compact set $K_\epsilon \subset \mathbb{R}^d$ such that $\int_{K_\epsilon} d\mu < \epsilon$. Then, given any compact set $L \subset \mathbb{R}^d$, $K_\epsilon \times L$ is compact, and for all $\gamma \in \Gamma D_\mu$,

$$\iint\limits_{K_\epsilon \times L} d\gamma \leqslant \iint\limits_{K_\epsilon \times \mathbb{R}^d} d\gamma = \int_{K_\epsilon} d\mu < \epsilon.$$

Therefore, ΓD_μ is tight and hence by Prokhorov is precompact. That is, given $\{\gamma_n\} \subset \Gamma D_\mu$, there exists a subsequence $\{\gamma_{n_k}\}$ converging weakly to a limit $\gamma \in P_2(\mathbb{R}^d \times \mathbb{R}^d)$. With this in mind, if $\{\nu_n\}$ is a sequence in D_μ, choose the corresponding $\{\gamma_n\}$, and let $\nu_{n_k} = \pi^2_\# \gamma_{n_k}$. For all $\varphi \in C_b(\mathbb{R}^d \times \mathbb{R}^d)$, $\iint \varphi(x, y) d\gamma_{n_k} \longrightarrow \iint \varphi(x, y) d\gamma(x, y)$. In particular, for all $\varphi \in C_b(\mathbb{R}^d)$,

$$\iint \varphi(x) d\gamma_{n_k}(x, y) = \int \varphi(x) d\nu_{n_k}(x) \longrightarrow \iint \varphi(x) d\gamma(x, y) = \int \varphi(x) d(\pi^2_\# \gamma)(x).$$

Thus $\nu_{n_k} \xrightarrow{w} \pi^2_\# \gamma$, so that $\{\nu_n\}$ contains a weakly convergent subsequence. Therefore D_μ is precompact, and since it is also closed, it follows that it is compact.

$\square$

3.1.2 Deterministic Couplings for Duality

We recall from Chapter 1 that probabilistic frames form a subclass of the set of continuous frames for $\mathbb{R}^d$. However, as we have seen above, we can broaden our approach to how duality is induced. In some cases, there exists a clear deterministic coupling which induces duality. Generalizing the set of duals for discrete frames outlined in [20, Theorem 5.6.5], we have the following construction:

Theorem 3.7. *Let μ be a probabilistic frame for $\mathbb{R}^d$, and let $h : \mathbb{R}^d \to \mathbb{R}^d$ be any function in $L^2(\mu, \mathbb{R}^d)$. Define $\psi_h : \mathbb{R}^d \to \mathbb{R}^d$ by $\psi_h(x) = x + h(x) - \int_{\mathbb{R}^d} \langle S_\mu^{-1} x \,, y \rangle h(y) d\mu(y)$. Then $\psi_{h\#}\mu \in D_\mu$.*

Proof. Consider $\mu, \psi_{h\#}\mu$ as above. Define $\gamma := (\iota, \psi_h)_\# \mu \in \Gamma(\mu, \psi_{h\#}\mu)$. Then

$$\begin{aligned}
\iint_{\mathbb{R}^d \times \mathbb{R}^d} xy^\top d\gamma(x,y) &= \int_{\mathbb{R}^d} x \left[x + h(x) - \int_{\mathbb{R}^d} \langle S_\mu^{-1} x \,, z \rangle h(z) d\mu(z) \right]^\top d\mu(x) \\
&= I + \int_{\mathbb{R}^d} xh(x)^\top - \iint_{\mathbb{R}^d \times \mathbb{R}^d} x(S_\mu^{-1}x)^\top z h(z)^\top d\mu(x) d\mu(z) \\
&= I
\end{aligned}$$

□

Remark 3.8. However, all transport duals cannot be constructed this way. Let $\mu \in P_2(\mathbb{R}^d)$ be a probabilistic frame which is the first marginal of the standard normal probability measure η on $\mathbb{R}^d \times \mathbb{R}^d$. Let ν be the second marginal of η, so that $\nu \in D_\mu$. Then the support of η is all of $\mathbb{R}^d \times \mathbb{R}^d$; in particular, η is not supported on a curve in $\mathbb{R}^d \times \mathbb{R}^d$, so that there does not exist a mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $(\iota, T)_\# \mu = \eta$, even though, clearly, $\eta \in \Gamma D_\mu$.

3.2 Constructions of Discrete Transport Duals and Fusion Frames

In what follows, we shall construct transport duals for discrete probabilistic frames which generalize the case of finite frame theory. From Definition 3.1, it is clear that the construction of a transport dual depends on the construction of a probability distribution on the product space with predetermined second-moments matrix and first and second marginals. In the finite case, these joint distributions will correspond to the set of matrices defined in the next section.

3.2.1 Doubly-Stochastic Matrices

Definition 3.9. Let $DS(M, N)$ denote the set of matrices $A \in \mathbf{R}^{M\times N}$ satisfying

$$\begin{cases} a_{i,j} \geqslant 0 \quad \forall i, j \\ \sum\limits_{i=1}^{M} a_{i,j} = \frac{1}{N} \quad \forall j \\ \sum\limits_{j=1}^{N} a_{i,j} = \frac{1}{M} \quad \forall i. \end{cases}$$

Remark 3.10. Given $A \in DS(M, N)$, we have $N + M - 1$ linear constraints on the entries of A, yielding an affine subspace of dimension $MN - N - M + 1 = (N - 1)(M - 1)$.

Remark 3.11. At times, we may explicitly relax Definition 3.9 to allow $\sum\limits_{i=1}^{M} a_{i,j} = \alpha_j$ and $\sum\limits_{j=1}^{N} a_{i,j} = \beta_i$ where $\sum\limits_{j=1}^{N} \alpha_j = \sum\limits_{i=1}^{M} \beta_i = 1$. In what follows, we choose the stricter definition unless otherwise noted.

Consider $A \in DS(M,N)$. Let

$$A_0 = \begin{bmatrix} \frac{1}{MN} & \cdots & \frac{1}{MN} \\ \vdots & \ddots & \vdots \\ \frac{1}{MN} & \cdots & \frac{1}{MN} \end{bmatrix}$$

and let $\{E_{i,j}\}_{i,j=1}^{M,N}$ denote the set of elementary $M \times N$ matrices which have $e_{i,j} = 1$ and zero in all other places. Then A can be decomposed as

$$A = A_0 + \sum_{i=1}^{M} \sum_{j=1}^{N} \lambda_{i,j} E_{i,j},$$

where

$$\lambda_{i,j} \geqslant -\frac{1}{MN} \text{ and } \sum_{i=1}^{M} \lambda_{i,j} = \sum_{j=1}^{N} \lambda_{i,j} = 0.$$

Together, these constraints imply that $\lambda_{i,j} \in [-\frac{1}{MN}, \frac{1}{MN} \min\{N-1, M-1\}]$.

Then given two frames $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^M$ with analysis operators Φ and Ψ, to show that $\mu_\Psi \in D_{\mu_\Phi}$, one must construct a matrix $A \in DS(M,N)$ solving:

$$\begin{aligned} (\Psi^\top A \Phi)_{k,l} &= \sum_{j=1}^{N} \sum_{i=1}^{M} \Psi_{i,k} \Phi_{j,l} \left(\frac{1}{MN} + \lambda_{i,j}\right) \\ &= \delta_{k,l} \end{aligned}$$

3.2.2 Construction of Transport Duals

The previous section begs two questions:

A Given frames $\{\varphi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^M$ for $\mathbb{R}^d$ with analysis operators Φ and Ψ, under what conditions on Φ and Ψ can we construct $A \in DS(M,N)$ with $\Psi^\top A \Phi = I$.

B Given a frame Φ, what conditions on $A \in DS(M,N)$ guarantee that there exists a frame Ψ such that $\Psi^\top A\Phi = I$.

Remark 3.12. Clearly, a necessary and sufficient condition to answer question B is that $rank(A\Phi) = d$, but this is very general. If $M \geqslant N$, then a sufficient condition is simply that $rank(A) = N$; in this case $rank(A\Phi)$ is guaranteed to be d [42]. To guarantee the existence of such a Ψ, it is sufficient to require that some subset of the rows of $A\Phi$ of cardinality d is linearly independent. That is, for some $I = \{i_1, ..., i_d\} \subset \{1, ..., M\}$, given any $\{\lambda_k\}_{k=1}^d$ not identically zero,

$$\sum_{k=1}^{d} \lambda_k \sum_{j=1}^{N} a_{i_k,j}\varphi_j^\top = \sum_{j=1}^{N} \varphi_j^\top \sum_{k=1}^{d} \lambda_k a_{i_k,j} \neq 0.$$

Thus, if we choose a subset $J = \{j_1, ..., j_d\} \subset \{1, ..., N\}$ such that $\{\varphi_j\}_{j\in J}$ is linearly independent and then choose a set I of row indices as above, any $A \in DS(M,N)$ satisfying $a_{i,j} = 0$ for all $i \in I$, $j \in \{1, ..., N\}\backslash J$ will be a transport plan inducing duality between $\{\varphi_i\}_{i=1}^N$ and the columns of any generalized inverse of $A\Phi$.

Given such a frame $\{\varphi_i\}_{i=1}^N$, let N_Φ denote the number of distinct linearly independent subsets of the frame vectors of cardinality d. (If the frame is full spark, then $N_\Phi = \binom{N}{d}$.) There will then be $\binom{M}{d} N_\Phi$ sets of entries of A to zero out in order to guarantee that A is a duality-inducing transport plan.

Theorem 3.13. *If $\{\psi_i\}_{i=1}^N \subset \mathbb{R}^d$ has centroid zero, then it has no transport dual of cardinality d.*

Proof. Suppose that a frame $\{\psi_i\}_{i=1}^N \subset \mathbb{R}^d$ has centroid zero. Recall that, given $\{u_i\}_{i=1}^d$, $\{v_i\}_{i=1}^d \subset \mathbf{R}^N$, $\langle \bigwedge_{i=1}^d u_i \ , \bigwedge_{j=1}^d v_j \rangle := \det([\langle u_i \ , v_j \rangle])$.

Let $\{v_j\}_{j=1}^d \subset \mathbf{R}^N$ denote the columns of Ψ, the analysis operator for our frame, and let $\{u_i\}_{i=1}^d \subset \mathbf{R}^N$ denote the rows of A, where $A \in DS(d, N)$ in the strict sense of Definition 3.9. $\{\psi_i\}_{i=1}^N$ will have a transport dual of cardinality d if and only if for some A, $A\Psi = [[\langle u_i \, , v_j \rangle]]$ is invertible. Recall that each $u_i = a_0 + \lambda^i$, where $a_0 = [\frac{1}{Nd} \cdots \frac{1}{Nd}]^\top$,

$$\sum_{k=1}^{N} \lambda_k^i = 0 \text{ for each } i \in \{1, ..., d\}$$
$$\sum_{i=1}^{d} \lambda_k^i = 0 \text{ for each } k \in \{1, ..., N\} \tag{3.1}$$

Then

$$\begin{aligned}\bigwedge_{i=1}^{d} u_i &= (a_0 + \lambda^1) \wedge (a_0 + \lambda^2) \wedge \cdots \wedge (a_0 + \lambda^d) \\ &= a_0 \wedge \lambda^2 \wedge \cdots \wedge \lambda^d + \lambda^1 \wedge a_0 \wedge \lambda^3 \wedge \cdots \wedge \lambda^d + \cdots \\ &+ \lambda^1 \wedge \cdots \wedge \lambda^{d-1} \wedge a_0 + \lambda^1 \wedge \cdots \wedge \lambda^d.\end{aligned}$$

Because of the zero-centroid condition, $\forall j \in \{1, ..., d\}$ $\sum_{k=1}^{N} v_j^k = 0$, and it follows that $\langle a_0 \, , v_j \rangle = 0$ for each $j \in \{1, ..., d\}$. Therefore,

$$\begin{aligned}\det([[\langle u_i \, , v_j \rangle]]) &= \langle \bigwedge_{i=1}^{d} u_i \, , \bigwedge_{j=1}^{d} v_j \rangle \\ &= \langle \lambda^1 \wedge \cdots \wedge \lambda^d \, , v_i \wedge ... \wedge v_d \rangle \\ &= 0\end{aligned}$$

where the last equality follows from equation (3.1)–i.e., the fact that the $\{\lambda^i\}_{i=1}^d$ are linearly dependent. $\square$

Corollary 3.14. *In particular, Theorem 3.13 implies that no equiangular tight frame in* $\mathbf{R}^2$ *has a transport dual of cardinality 2.*

3.2.3 Relationship to Fusion Frames

Consider the following generalization of finite frames:

Definition 3.15. [9, Definition 2.2] Let I be an index set, let $\{W_i\}_{i\in I}$ be a family of closed subspaces of $\mathbb{R}^d$, and let $\{v_i\}_{i\in I}$ be a family of positive weights. Let P_i denote the orthogonal projection onto W_i. Then $\{(W_i, v_i)\}_{i\in I}$ is a **fusion frame** for $\mathbb{R}^d$ if there exist $0 < C \leqslant D < \infty$ such that for all $x \in \mathbb{R}^d$,

$$C\|x\|^2 \leqslant \sum_{i\in I} v_i^2 \|P_i x\|^2 \leqslant D\|x\|^2$$

First introduced as "frames of subspaces" in [19], fusion frames are designed to formalize a signal processing or measurement scheme in which the analysis of a signal must be performed in a distributed way, either because of the dimension of the problem or because the measurement system is not centralized, as is the case for a wireless sensor network. Some reconstruction can and must be done locally, and the results of that distributed processing, which will be of smaller dimension than the original signal, pieced back together in a meaningful way at the end of the process. From a frame theory perspective, one considers the projections of the signal onto a series of overlapping, possibly nonorthogonal ([12]) subspaces and formulates sufficient conditions on a recombination scheme for perfect reconstruction from the sets of coefficients derived from those projections to be achievable. It turns out that by simply using transport plans between discrete probabilistic frames with supports of different cardinalities, we can construct objects similar to fusion frames.

In general, if Ψ is a transport dual to Φ via $A \in DS(M, N)$, one can decompose the set $\{1, ..., M\}$ into disjoint subsets $I_1, ..., I_r$ and $\{1, ..., N\}$ into $J_1, ..., J_s$. Then one may write

$$I = \sum_{k=1}^{r} \sum_{l=1}^{s} \sum_{i \in I_k} \sum_{j \in J_l} a_{i,j} \psi_i \varphi_j^\top.$$

If A is block diagonal (i.e., $r = s$ and $a_{i,j} = 0$ for $i \in I_k$, $j \in J_l$ if $k \neq l$), then one has

$$I = \sum_{k=1}^{r} \sum_{i \in I_k} \sum_{j \in J_k} a_{i,j} \psi_i \varphi_j^\top \tag{3.2}$$

This leads us to the following decomposition:

Proposition 3.16. *If we allow the relaxed definition of $DS(M, N)$, as in Remark 3.11, then the frame operator of any fusion frame can be decomposed as a coupling of two discrete probabilistic frames.*

Proof. Given a fusion frame $\{(W_k, v_k)\}_{k=1}^r$ for $\mathbb{R}^d$, its analysis operator is $T : \mathbb{R}^d \to \sum_{k=1}^{r} \oplus W_k$ given by $T(x) = \{v_k P_k(x)\}_{k=1}^r$, where P_k is an orthogonal projector onto W_k. Its synthesis operator is $T(\{x_k\}_{k=1}^r) = \sum_{k=1}^{r} v_k x_k$, and its frame operator is given by $S(x) = \sum_{k=1}^{r} v_k^2 P_k(x)$. Let $\hat{\Psi}_k^\top \hat{\Phi}_k = P_k$ be a decomposition for each projection operator, so that $\{\hat{\varphi}_j^k, \hat{\psi}_j^k\}_{j=1}^{n_k}$ is a frame/dual-frame pair for W_k. Let $\sum_{k=1}^{r} w_r = 1$ for some arbitrary positive sequence. Then define $\varphi_j^k = v_k \sqrt{\frac{n_k}{w_k}} \hat{\varphi}_j^k$ and $\psi_j^k = v_k \sqrt{\frac{n_k}{w_k}} \hat{\psi}_j^k$ for all $j \in \{1, ..., n_k\}$,

$k \in \{1, ..., r\}$. Let $N = \sum_{k=1}^{r} n_r$, and define $A \in \mathbf{R}^{N \times N}$ by

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{bmatrix}$$

where $A^k_{i,j} = \frac{w_k}{n_k}\delta_{i,j}$. Note that A is doubly stochastic in the relaxed sense, and if we choose $w_r = \frac{n_r}{N}$, then it is doubly stochastic in the stricter sense, as well. Then with

$$\Psi^\top = \left[\begin{array}{c|c|c|c|c|c|c|c|c|c} \psi^1_1 & \dots & \psi^1_{n_1} & \psi^2_1 & \dots & \psi^2_{n_2} & \dots & \psi^r_1 & \dots & \psi^r_{n_r} \end{array} \right]$$

and Φ defined similarly,

$$\begin{aligned} \Psi^\top A \Phi &= \sum_{k=1}^{r} \Psi_k^\top A_k \Phi_k \\ &= \sum_{k=1}^{r} \frac{w_k}{n_k} \left(v_k \sqrt{\frac{n_k}{w_k}} \right)^2 \hat{\Psi}_k^\top \hat{\Phi}_k \\ &= \sum_{k=1}^{r} v_k^2 P_k \end{aligned}$$

□

Remark 3.17. To do reconstruction with fusion frames, one would in general still have to invert the fusion frame operator and apply the inversion to each subspace.

We can speculate, however, that an efficient reconstruction scheme for distributed processing could be devised using a cleverly constructed fusion-like transport duals. Given a set of subspaces $\{W_k\}_{k=1}^r$ of $\mathbb{R}^d$, one would choose frames for those subspaces $\{\{\varphi_j^k\}_{j=1}^{n_k}\}_{k=1}^r$. Letting $N = \sum_{k=1}^r n_k$, one would choose a set $\{m_k\}_{k=1}^r$, with $0 < m_k \leqslant n_k$ for each k, and a positive sequence w_k with $\sum\limits_{k=1}^{r} w_k = 1$.

Then, one would define a block matrix A as above with $\sum\limits_{i=1}^{m_k} a^k_{i,j} = \frac{w_r}{n_r}$ and $\sum\limits_{j=1}^{n_k} a^k_{i,j} = \frac{w_r}{m_r}$ for each k. One could use the constraints outlined in Remark 3.12 to guarantee that $A\Phi$ would have a generalized inverse $\Phi^\top$ and form the columns of such an inverse into dual frames for the subspaces $\{\{\psi_i^k\}_{i=1}^{m_k}\}_{k=1}^r$.

If $M < rd$, this could model a reconstruction algorithm for a distributed sensor network, doing a partial local reconstruction if a signal x at each of r sensors using

$A_r \Phi_r x$ and transmitting the result to reconstruct fully with Ψ, which might require sending fewer bits than doing a full reconstruction on each subspace $v_k^2 S^{-1} P_k(x)$ and transmitting the result.

3.2.4 Decomposition of Full Rank Matrices

Finally, we have the following interesting result about decompositions of full-rank matrices in $\mathbb{R}^{d \times N}$ in terms of doubly-stochastic matrices. This resulted from an attempt to answer question A, which still remains open.

Proposition 3.18. *Given a full-rank matrix $B \in \mathbf{R}^{d \times N}$, where $d \leqslant N$, B can always be decomposed as $B = UAF$, where $U \in \mathbf{R}^{d \times d}$ is a unitary matrix, $A \in DS(d, N)$, and $F \in \mathbf{R}^{N \times N}$ is nonsingular.*

Proof. Take B as above, and consider its singular value decomposition $B = UDV^\top$. Since $rk(B) = d$, $D \in \mathbf{R}^{d \times N}$ is of the form:

$$D = \begin{bmatrix} \Lambda & 0 \end{bmatrix}$$

where

$$\Lambda_{i,j} = \begin{cases} \sigma_i & \text{for } i = j \\ 0 & \text{otherwise} \end{cases}$$

where $\{\sigma_i\}_{i=1}^d$ are the singular values of B. Then let

$$T = \begin{bmatrix} & \frac{1}{N\sigma_1} & \cdots & \frac{1}{N\sigma_1} \\ \frac{d}{N}\Lambda^{-1} & \vdots & \ddots & \vdots \\ & \frac{1}{N\sigma_d} & \cdots & \frac{1}{N\sigma_d} \\ 0 & & T_0^{-1} & \end{bmatrix}$$

with T_0^{-1} the inverse of some $(N-d) \times (N-d)$ real matrix T_0. Then, letting R_0 denote the upper-right block of T, i.e.,

$$R_0 = \begin{bmatrix} \frac{1}{N\sigma_1} & \cdots & \frac{1}{N\sigma_1} \\ \vdots & \ddots & \vdots \\ \frac{1}{N\sigma_d} & \cdots & \frac{1}{N\sigma_d} \end{bmatrix}$$

we construct:

$$T^{-1} = \begin{bmatrix} \frac{N}{d}\Lambda & -\frac{N}{d}\Lambda R_0 T_0 \\ 0 & T_0 \end{bmatrix}$$

Letting $A = DT$, we note that $A \in DS(d, N)$, and letting $F = T^{-1}V^\top$, we note that F is nonsingular. Therefore,

$$B = UDV^\top = UDTT^{-1}V^\top = UAF.$$

□

3.3 Analysis and Synthesis

By now, we have made use many times of the analysis and synthesis operators which are the backbone of finite frame theory. However, our construction of transport duals suggests that for probabilistic frames, a more probability-theoretic definition of analysis and synthesis may be called for.

In [29–31], the analysis and synthesis operators are defined in a manner similar to that of continuous frames. To wit, we quote:

Definition 3.19. *Analysis and Synthesis, [30, 2.2]* Given a probabilistic frame μ, its analysis operator is $A_\mu : \mathbb{R}^d \to L^2(\mu, \mathbb{R}^d)$ given by $x \mapsto \langle x\ , \cdot \rangle$. Its synthesis operator is $A_\mu^* : L^2(\mu, \mathbb{R}^d) \to \mathbb{R}^d$ given by $f \mapsto \int_{\mathbb{R}^d} x f(x) d\mu(x)$.

As defined here, the analysis operator A_μ is independent of the measure μ. Indeed, it is not clear from this definition how one could do "analysis" with one probabilistic frame and "synthesis" with another. However, finite frame theory itself gives us a clue about how to think about analysis and synthesis in the probabilistic context.

Example 3.2. *Consider two frames for $\mathbb{R}^d$, $\{\varphi_i\}_{i=1}^N$ and $\{\psi_i\}_{i=1}^N$. Let $\{e_i\}_{i=1}^N \subset \mathbb{R}^N$ be an orthonormal basis for $\mathbb{R}^N$. Then the analysis operator for Φ, $A_\Phi : \mathbb{R}^d \to \mathbb{R}^N$ given by*

$$A_\Phi(x) = \Phi x = \sum_{i=1}^N \langle \varphi_i \, , x \rangle e_i \quad \text{for } x \in \mathbb{R}^d.$$

The synthesis operator for Ψ, $A_\Psi^* : \mathbb{R}^N \to \mathbb{R}^d$, is given by

$$A_\Psi^*(y) = \Psi^\top y = \sum_{i=1}^N \langle y \, , e_i \rangle \psi_i \quad \text{for } y \in \mathbb{R}^N.$$

Then we can compose the operators simply by writing $A_\Psi^* A_\Phi(x) = \sum_{i=1}^N \langle \varphi_i \, , x \rangle \psi_i$. If, however, we choose some σ and π in S_N, and instead choose to do analysis and synthesis with the two frames as

$$A_\Psi^* A_\Phi(x) = \sum_{i=1}^N \langle \varphi_{\sigma(i)} \, , x \rangle \psi_{\pi(i)},$$

then it will be as if we had chosen two different finite frames to work with. This is because the ordering of the frame vectors is implicitly tied to the ordering of the reference basis $\{e_i\}_{i=1}^N$.

In what follows, we shall generalize this idea of a reference ordering through the use of disintegration of measure–the construction of conditional probabilities with respect to some reference measure. The orthogonality of the reference basis in the above example will turn out not to be crucial; its function is to match up frame coefficients with the appropriate vectors. What will be crucial is that transport plans exist between the probabilistic frame and the reference measure and that no information be lost in the encoding. For this reason, we will use reference measures that are absolutely continuous probabilistic frames, except in the discrete case, where we will simply need a reference measure with enough elements in its support to define a transport plan between finite frames of interest.

3.3.1 Measure-Valued Maps and Disintegration

To make this idea of coefficient-matching rigorous, we shall use some ideas from machinery from probability theory.

Remark 3.20. First, we note that for brevity we will sometimes use the expected value notation in place of integral notation in what follows, i.e., for a measure η and a function $f \in \mathcal{L}^1(\eta)$, we will write:

$$E_\eta(f) = \int f(w) d\eta.$$

We start with conditional probabilities. Let $\mathcal{X}, \mathcal{Y}$ be separable metric spaces; following [2, Section 5.3], we define:

Definition 3.21. Let $x \in \mathcal{X} \mapsto \mu_x \in P(\mathcal{Y})$ be a measure-valued map. Then μ_x is Borel if $x \mapsto \mu_x(B)$ is a Borel map for any Borel set $B \in \mathcal{B}(\mathcal{Y})$.

With this in hand, we recall the following key result on disintegration, originally attributed to Rokhlin:

Lemma 3.22. *[2, Theorem 5.3.1] Let $\boldsymbol{\mathcal{X}}$, $\mathcal{X}$ be Radon separable metric spaces (i.e., having every Borel probability measure inner regular), $\boldsymbol{\mu} \in P(\boldsymbol{\mathcal{X}})$, and $\pi : \boldsymbol{\mathcal{X}} \to \mathcal{X}$ a Borel map.*

Let $\nu = \pi_{\#}\boldsymbol{\mu} \in P(\mathcal{X})$. Then there exists a ν-a.e. uniquely determined Borel family of probability measures $\{\mu_x\}_{x\in\mathcal{X}} \subset P(\mathcal{X})$ such that

$$\mu_x(\boldsymbol{X}\backslash\pi^{-1}(x)) = 0$$

for ν-a.e. $x \in \mathcal{X}$, and for every Borel map $f : \boldsymbol{\mathcal{X}} \to [0, +\infty]$,

$$\int_{\boldsymbol{\mathcal{X}}} f(\boldsymbol{x}) d\boldsymbol{\mu}(\boldsymbol{x}) = \int_{\mathcal{X}} \left(\int_{\pi^{-1}(x)} f(\boldsymbol{x}) d\mu_x(\boldsymbol{x}) \right) d\nu(x).$$

Remark 3.23. [2, p.122] In particular, if $\boldsymbol{\mathcal{X}} = \mathcal{X} \times \mathcal{Y}$, $\boldsymbol{\mu} \in P(\mathcal{X} \times \mathcal{Y})$, $\nu = \mu^1 = \pi^1_{\#}\boldsymbol{\mu}$, then one can canonically identify each fiber $(\pi^1)^{-1}(x)$ with $\mathcal{Y}$ and find a Borel family of probability measures $\{\mu_x\}_{x\in\mathcal{X}} \subset P(\mathcal{Y})$ which is μ^1-a.e. uniquely determined such that $\boldsymbol{\mu} = \int_{\mathcal{X}} \mu_x d\mu^1(x)$.

That is, for any $f \in C_b(\mathcal{X} \times \mathcal{Y})$, we can write

$$\iint_{\mathcal{X}\times\mathcal{Y}} f(x, y) d\boldsymbol{\mu}(x, y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x, y) d\boldsymbol{\mu}(y|x) d\mu^1(x).$$

Secondly, we have the following result about gluings, which we state in its full generality. We note that a Radon space is a separable metric space on which every Borel probability measure is inner regular, so that $\mathbb{R}^d$ is certainly within its purview:

Lemma 3.24. Gluing Lemma [2, Lemma 5.3.2] *Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ be Radon separable metric spaces and let $\gamma^{12} \in P(\mathcal{X}_1 \times \mathcal{X}_2)$, $\gamma^{13} \in P(\mathcal{X}_1 \times \mathcal{X}_3)$ such that $\pi^1_{\#}\gamma^{12} = \pi^1_{\#}\gamma^{13} = \mu^1$. Then there exists $\boldsymbol{\mu} \in P(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ such that $\pi^{1,2}_{\#}\boldsymbol{\mu} = \gamma^{12}$ and $\pi^{1,3}_{\#}\boldsymbol{\mu} = \gamma^{13}$.*

Moreover, if $\gamma^{12} = \int \gamma^{12}_{x_1} d\mu^1$, $\gamma^{13} = \int \gamma^{13}_{x_1} d\mu^1$, *and* $\boldsymbol{\mu} = \int \boldsymbol{\mu}_{x_1} d\mu^1$ *are the disintegrations of* γ^{12}, γ^{13}, *and* $\boldsymbol{\mu}$ *with respect to* μ^1, *then the first statement is equivalent to* $\boldsymbol{\mu}_{x_1} \in \Gamma(\gamma^{12}_{x_1}, \gamma^{13}_{x_1}) \subset P(\mathcal{X}_2, \mathcal{X}_3)$ *for* μ^1*-a.e.* $x_1 \in \mathcal{X}_1$.

3.3.2 Construction of Analysis and Synthesis Operators

From Lemma 3.24, we know that given $\mu, \eta \in \mathrm{PF}(\mathbb{R}^d)$ and $\gamma \in \Gamma(\mu, \eta)$, we have set of conditional probability measures $\{\gamma(\cdot|w)\}_{w \in \mathbb{R}^d}$ that are uniquely defined η-a.e. such that for any test function $f \in C_b(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} f(y,w) d\gamma(y,w) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y,w) d\gamma(y|w) \right) d\eta(w).$$

Proposition 3.25. *If* $f \in L^2(\mathbb{R}^d \times \mathbb{R}^d, \gamma)$, *it follows that* $g(w) := \int_{\mathbb{R}^d} f(y,w) d\gamma(y|w)$ *is in* $L^2(\mathbb{R}^d, \eta)$.

Proof. By conditional Jensen's inequality,

$$\begin{aligned} \int \left(\int f(y,w) d\gamma(y|w) \right)^2 d\eta(w) &\leqslant \iint f^2(y,w) d\gamma(y|w) d\eta(w) \\ &= \iint f^2(y,w) d\gamma(y,w) \\ &= \| f \|^2_{L^2(\mathbb{R}^d \times \mathbb{R}^d, \gamma)} \\ &< \infty. \end{aligned}$$

□

Remark 3.26. In particular, if $f(y,w) = \langle x \,, y \rangle$ for some $x \in \mathbb{R}^d$, then

$$\| f \|_{L^2(\mathbb{R}^d \times \mathbb{R}^d, \gamma)} \leqslant \| x \|^2 M_2^2(\mu).$$

Now we are ready to construct analysis and synthesis operators which are truly tied to their probabilistic frames, as suggested in the introduction to this section.

Definition 3.27. Given $\mu \in \mathrm{PF}(\mathbb{R}^d)$, choose a reference measure $\eta \in \mathrm{PF}(\mathbb{R}^d)$ which is absolutely continuous with respect to Lebesgue measure. Then we define the **analysis and synthesis operators** for μ with respect to η.

The **analysis operator**, $A_\mu : \mathbb{R}^d \times \Gamma(\mu, \eta) \to L^2(\mathbb{R}^d, \eta)$, is given by

$$A_\mu(x, \gamma)(w) = \int_{\mathbb{R}^d} \langle x\ , y \rangle d\gamma(y|w).$$

Noting that $h(z, w) := \|z\| \in L^2(\gamma)$ for any $\gamma \in \Gamma(\mu, \eta)$ provided that $\mu \in P_2(\mathbb{R}^d)$, the vector-valued function $\int z d\gamma(z|w)$ lies in $L^2(\eta)$. Therefore, we can define the **synthesis operator**, $Z_\mu : L^2((\mathbb{R}^d), \eta) \times \Gamma(\mu, \eta) \to \mathbb{R}^d$, is given by

$$\begin{aligned} Z_\mu(f, \gamma) &= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} z f(w) d\gamma(z|w) d\eta(w) \\ &= E_\eta\left[f(w) \int_{\mathbb{R}^d} z d\gamma(z|w) \right] \\ &= \left\langle f(w)\ , \int_{\mathbb{R}^d} z d\gamma(z|w) \right\rangle_{L^2(\eta)} \end{aligned}$$

3.3.3 Adjoints and Composition

Given $\mu, \nu \in \mathrm{PF}(\mathbb{R}^d)$, and a fixed reference measure η as above, we write, somewhat formally,

$$Z_\nu^*(T_\mu(x, \gamma), \xi) = \iiint \langle x\ , y \rangle z d\gamma(y|w) d\xi(z|w) d\eta(w),$$

knowing that a gluing $\zeta \in P_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ exists with the marginals satisfying $\pi_\#^{1,2}\zeta = \gamma$ and $\pi_\#^{3,2}\zeta = \xi$ and marginal conditional probabilities agreeing with the versions we chose η-a.e.

Then, given $v, x \in \mathbb{R}^d$,

$$\begin{aligned}
\langle v\ , Z_\nu(A_\mu(x,\gamma),\xi)\,\rangle &= \iiint\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x\ , y\,\rangle\langle z\ , v\,\rangle d\gamma(y|w)d\xi(z|w)d\eta(w) \\
&= \langle Z_\mu(A_\nu(v,\xi),\gamma)\ , x\,\rangle \\
&= \langle A_\nu(v,\xi)\ , A_\mu(x,\gamma)\,\rangle_{L^2(\eta)}
\end{aligned}$$

Similarly, we can define for $f \in L^2(\mathbb{R}^d, \eta)$,

$$\begin{aligned}
A_\mu(Z_\nu(f,\xi),\gamma) &= \iiint\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle y\ , z\,\rangle f(u)d\xi(z|u)d\eta(u)d\gamma(y|w) \\
&= \iiint\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle y\ , z\,\rangle f(u)d\xi(z,w)d\gamma(y|w)
\end{aligned}$$

Then, given $f, g \in L^2(\mathbb{R}^d, \eta)$,

$$\begin{aligned}
\langle g\ , A_\mu(Z_\nu(f,\xi),\gamma)\,\rangle_{L^2(\eta)} &= \iiint\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle y\ , z\,\rangle f(u)g(w)d\xi(z|u)d\eta(u)d\gamma(y|w)d\eta(w) \\
&= \langle Z_\mu(g,\gamma)\ , Z_\nu(f,\xi)\,\rangle \\
&= \langle A_\nu(Z_\mu(g,\gamma),\xi)\ , f\,\rangle_{L^2(\eta)}
\end{aligned}$$

We have the next result concerning the continuity of this construction:

Proposition 3.28. *Let $\mu \in PF(\mathbb{R}^d)$ and fix a reference measure η and $\gamma \in \Gamma(\mu, \eta)$. Then $x \mapsto A_\mu(\cdot, \gamma)$ is continuous in its first argument; in fact, it is Lipschitz.*

Proof. The key point in the proof is again conditional Jensen.

$$\begin{aligned}
\|A_\mu(x_1,\gamma) - A_\mu(x_2,\gamma)\,\|^2_{L^2(\eta)} &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\langle x_1 - x_2\ , y\,\rangle d\gamma(y|w)\right)^2 d\eta(w) \\
&\leqslant \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle x_1 - x_2\ , y\,\rangle^2 d\gamma(y,w)
\end{aligned}$$

$$= \int_{\mathbb{R}^d} \langle x_1 - x_2\ , y \rangle^2 d\mu(y)$$

$$\leqslant \| x_1 - x_2 \|^2 M_2^2(\mu)$$

□

3.3.4 Examples of Analysis/Synthesis Pairings

As a preliminary, we discuss the transport of an absolutely continuous measure to a discrete measure using power (Voronoi) cells, following [43].

Definition 3.29. Given a probability measure μ on $\mathbb{R}^d$, a finite set P of points in $\mathbb{R}^d$ and $w : P \to \mathbb{R}_+$ a weight vector, the **power diagram** or **weighted Voronoi diagram** of (P, w) is a decomposition of $\mathbb{R}^d$ into cells corresponding to each member of P. A point x belongs to $\mathrm{Vor}_P^w(p)$ if and only if for every $q \in P$,

$$\| x - p \|^2 - w(p) \leqslant \| x - q \|^2 - w(q).$$

Definition 3.30. Call the application T_P^w which maps every point x in a power cell $\mathrm{Vor}_P^w(p)$ to the "center" of that power cell p, the weighted Voronoi mapping. Then

$$T_P^w|_{\#}\mu = \sum_{p \in P} \mu(\mathrm{Vor}_P^w(p))\delta_p.$$

It is a consequence of some of Brenier's work, cited in [43, Theorem 1], that T_P^w is an optimal transport map between μ and $T_P^w|_{\#}\mu$ for the Wasserstein distance when μ is absolutely continuous with respect to Lebesgue measure.

Definition 3.31. Let η be an absolutely continuous measure in $P_2(\mathbb{R}^d)$, and let ν be a discrete measure in $P_2(\mathbb{R}^d)$ supported on a finite set of points P with weights

$\{\lambda_p\}$ summing to unity. Then we say that a vector weight $w : P \to \mathbb{R}_+$ is **adapted** to (η, ν) if for all $p \in P$, $\lambda_p = \eta(\mathrm{Vor}_S^w(p)) = \int_{\mathrm{Vor}_S^w(p)} d\eta(x)$.

Example 3.3. *Now given discrete frames* $\{\varphi_i\}_{i=1}^M$ *and* $\{\psi_j\}_{j=1}^N$ *for* $\mathbb{R}^d$*, and* η *a reference measure in Definition 3.27, choose* $\gamma_1 = (\iota, T_\Phi^{w_1})_\# \eta$ *and* $\gamma_2 = (\iota, T_\Psi^{w_2})_\# \eta$*, where* w_1 *and* w_2 *are adapted to* (μ_Φ, η) *and* (μ_Ψ, η)*, respectively. Then*

$$Z_{\mu_\Psi}(A_{\mu_\Phi}(x, \gamma_1), \gamma_2) = \int \langle x \ , T_\Phi^{w_1}(y) \rangle T_\Psi^{w_2}(y) d\eta(y).$$

Example 3.4. Recovering the old definitions of analysis and synthesis

In the special case $M = N$*, we could choose* $\{\alpha_i\}_{i=1}^N \subset \mathbb{R}^d$ *and* w_0 *adapted to* (μ_α, η)*. Then let* $f_\Psi : \alpha \to \Psi$ *be given by* $f_\Psi(\alpha_i) = \psi_i$*, and let* $f_\Phi : \alpha \to \Phi$ *be similarly defined. Then if* $\gamma_1 = (\iota, f_\Phi \circ T_\alpha^{w_0})_\# \eta$ *and* $\gamma_2 = (\iota, f_\Psi \circ T_\alpha^{w_0})_\# \eta$*, it follows that*

$$Z_{\mu_\Psi}(A_{\mu_\Phi}(x, \gamma_1), \gamma_2) = \int \langle x \ , f_\Phi \circ T_\alpha^{w_0}(y) \rangle f_\Psi \circ T_\alpha^{w_0}(y) d\eta(y) = \sum_{i=1}^N \langle x \ , \varphi_i \rangle \psi_i.$$

Hence, we have recovered the analysis and synthesis operation of finite frames.

Example 3.5. Discrete dual to absolutely continuous probabilistic frame

Finally, let us imagine that η *is an absolutely continuous probabilistic frame for* $\mathbb{R}^d$ *and choose a frame contained in its support, say* $\{\psi_i\}_{i=1}^N$*. Let* T_Ψ^w *be the transport map between* η *and* μ_Ψ*, as constructed above. Choose* $\{\varphi_i\}_{i=1}^N$ *to be any dual to* $\{\psi_i\}_{i=1}^N$*, and let* $f : \Psi \to \Phi$ *be given by* $f(\psi_i) = \varphi_i$*. Then* $\gamma = (\iota, f \circ T_\Psi^w)_\# \eta \in P_2(\mathbb{R}^d \times \mathbb{R}^d)$ *is a joint transport plan in* $\Gamma(\eta, \mu_\Psi)$ *such that* $\iint x y^\top d\gamma(x, y) = \int x T_\Psi^w(x) d\eta(x) = I$*, so that* η *and* μ_Ψ *are dual to one another in* $PF(\mathbb{R}^d)$*.*

Chapter 4

Frame Forces and Gradient Flows

4.1 Introduction

While spanning sets are a dime a dozen, certain frames with specified structure are in high demand. First among these are, of course, the FUNTFs and the other equal-norm, tight frames. Examples of this sort of frame for $\mathbb{R}^d$ can be constructed easily, as remarked in [17], by the technique of majorization of matrices, by spectral tetris methods [18], or by simply using submatrices of the DFT matrix of the correct dimension. However, these methods produce only a few samples from the nontrivial manifolds contained in the set of all FUNTFs of sufficiently high cardinality modulo rotations [28]. For that reason, it might be useful to find methods to "traverse" the set of frames in a continuous manner in order to find approximations to tight frames.

In particular, we might also ask: "How close is the nearest FUNTF to a given frame which is almost tight and almost unit norm?" We state this more precisely as:

Definition 4.1. *The Paulsen Problem*

Given a frame $\Phi = \{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$ and $\epsilon > 0$, Φ is ϵ**-almost unit norm** if

$$\|\varphi_i\| \in (1-\epsilon, 1+\epsilon) \quad \forall i \in \{1, \cdots, N\}$$

and ϵ**-almost tight** if

$$(1-\epsilon)A \leqslant S_\Phi \leqslant (1+\epsilon)A$$

in the operator sense for some $A > 0$. Then the **Paulsen Problem** is, given $\delta > 0$, N, and d, to find the largest $\epsilon > 0$ such that whenever $\Phi = \{\varphi_i\}_{i=1}^N$ is ϵ-almost tight and ϵ-almost unit norm, there is a FUNTF $\{\psi_i\}_{i=1}^N$ such that

$$\sum_{i=1}^N \|\varphi_i - \psi_i\|^2 \leqslant \epsilon^2.$$

There are multiple approaches to this two-sided problem: identifying the closest FUNTF and calculating a minimum distance to that FUNTF. In [8], the approach is to start with a tight frame which is almost unit-norm and to solve a system of ODEs based on a quantity termed the "frame energy." The solution maintains the tightness of the starting frame and solves the Paulsen problem in the case that the number of frame vectors and the dimension of the space are relatively prime (RP). In [17], an alternate approach is taken; the starting frame is assumed to be unit-norm, and a discretized gradient descent for the frame potential of [5] is constructed which maintains the norm of the frame vectors while pushing the frame toward a tight frame. In [17], the authors can guarantee linear convergence of their method to a FUNTF provided that either the RP condition holds or that the frames along the descent are not almost orthogonally partitionable. In [6], the authors considered a related frame optimization problem based on minimizing a potential tied to the probability of error in quantum detection. To do so, they constructed a flow over the set of orthonormal bases in a higher-dimensional space which converged to a minimum for this quantity and then used Naimark's theorem to obtain a tight frame from this solution. Thus, the idea of using differential calculus to find useful frames is not new. However, the setting of probabilistic frames in the Wasserstein space allows the construction of much more general gradient flows for frame potentials because of the sophisticated machinery which has been developed for this space, which is outlined in brief in the following section. The following sections, beginning with Section 4.3 then explain its application to probabilstic frames, with the main results in Theorems 4.29, 4.38, and 4.41.

4.2 Gradient Flows

4.2.1 Transport Equation

The connection between the transport equation and the 2-Wasserstein distance has been studied for years ([2,10,32,37]). Indeed, as noted in [4], in Monge's original problem ([44]), there was already an implicit continuum mechanics formulation, and what is now considered Monge's problem is the result of a clever elimination of the time variable. Reintroducing the time variable, as in [4], allows one to use methods from numerical PDEs to find solutions to the Monge-Kantorovich. However, this reintroduction of the time variable has much larger implications because the space $P_2(\mathbb{R}^d)$ is a Polish (separable, complete, metric) space. As a result, much effort has gone into developing a rich theory of gradient flows on this space, with weak solutions to flows based on the theory of 2-absolutely continuous curves (e.g. [2, 36]). Tangent spaces can be defined and with them a formal calculus. Many PDEs can be reformulated as energy minimization problems in this space (e.g., [14, 39]).

The authors of [36] had in mind the goal of viewing gradient flows on $P_2(\mathbb{R}^d)$ as Hamiltonian flows and therefore, of necessity, developing a symplectic formalism for the space. However, a great deal of technical effort is required, in particular because $P_2(\mathbb{R}^d)$ is a stratified, rather than smooth, manifold [36, Chapter 6]. For our purposes, the technical basis for weak solutions provided by [2] will be enough, although we will refer to intuitions and certain reformulations provided by [36].

Definition 4.2. *[36, 2.10, Absolutely continuous curves in $P_2(\mathbb{R}^d)$]*

A curve $\sigma_t : (a, b) \to P_2(\mathbb{R}^d)$ is 2-**absolutely continuous** if $\exists\ \beta \in L^2((a, b))$ such that

$$W_2(\sigma_t, \sigma_s) \leqslant \int_s^t \beta(\tau) d\tau \quad \text{for all } a < s < t < b.$$

For such $\sigma \in AC_2(a, b; P_2(\mathbb{R}^d))$, the metric derivative $|\sigma'|(t) := \lim_{s\to t} \frac{W_2(\sigma_t,\sigma_s)}{|t-s|}$ exists for $\mathcal{L}^1$-a.e. $t \in (a, b)$.

Definition 4.3. *[2, p.169]* Let μ_t be a family of Borel probability measures on $\mathbb{R}^d$ for $t \in (0, T)$ and $v : (x, t) \to v_t(x) \in \mathbb{R}^d$ a Borel velocity field satisfying

$$\int_0^T \int_{\mathbb{R}^d} \| v_t(x) \| d\mu_t(x) dt < \infty.$$

Then the **continuity equation**

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \tag{4.1}$$

is interpreted in the sense of distributions, i.e. $\forall \varphi \in C_c^\infty(\mathbb{R}^d \times (0, T))$,

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi(x, t) + \langle v_t(x)\ , \nabla_x \varphi(x, t) \rangle d\mu_t(x) dt = 0. \tag{4.2}$$

In order to discuss representation of solutions to the continuity equation, we require the following technical lemma on characteristics which provides us with the definition of flow.

Lemma 4.4. [2, Lemma 8.1.4] *Let $v_t : (x) = v(x,t)$ be a Borel vector field such that for every compactly subset $B \subset \mathbb{R}^d$,*

$$\int_0^T \left(\sup_B \|v_t\| + Lip(v_t, B) \right) dt < \infty \tag{4.3}$$

Then, for every $x \in \mathbb{R}^d$ and $s \in [0,T]$, we let $\varphi_t(x,s)$ denote the location in $\mathbb{R}^d$ at time t of a trajectory passing through a point x at time s which satisfies the ODE:

$$\varphi_s(x,s) = x, \quad \frac{d}{dt}\varphi_t(x,s) = v_t(\varphi_t(x,s)). \tag{4.4}$$

*This ODE admits a unique maximal solution defined on an interval $I(x,s)$ relatively open in $[0,T]$ and containing s as a point in its relative interior. We say that φ_t is the **flow** of v_t.*

Furthermore, if $t \mapsto |\varphi_t(x,s)|$ is bounded on the interior of $I(x,s)$, then $I(x,s) = [0,T]$; finally, if v satisfies

$$S := \int_0^T (\sup_{\mathbb{R}^d} \|v_t\| + Lip(v_t, \mathbb{R}^d))dt < \infty$$

then the flow map φ_t satisfies

$$\int_0^T \sup_{x \in \mathbb{R}^d} |\partial_t \varphi_t(x,s)| dt \leqslant S \tag{4.5}$$

and

$$\sup_{t,s \in [0,T]} Lip(\varphi_t(\cdot, s), \mathbb{R}^d) \leqslant e^S \tag{4.6}$$

When $s = 0$, write $\varphi_t(x) := \varphi_t(x,0)$.

Having defined a flow for a vector field in terms of characteristics, we can now address solutions to the continuity equation.

Lemma 4.5. Representation formula for continuity equation, [2, Proposition 8.1.8] *Let μ_t, $t \in [0,T]$ be a narrowly continuous (i.e. continuous in the weak topology) family of Borel probability measures solving the continuity equation* (4.1) *with respect to a Borel vector field v_t satisfying equations* (4.2) *and* (4.3). *Then for μ_0-a.e. $x \in \mathbb{R}^d$, the characteristic system* (4.4) *admits a globally-defined solution $\varphi_t(x)$ in $[0,T]$, and*

$$\mu_t = (\varphi_t)_{\#}\mu_0 \quad \forall t \in [0,T] \tag{4.7}$$

Moreover, if for some $p > 1$,

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p d\mu_t(x) dt < \infty$$

then the velocity field v_t is the time derivative of φ_t in the $\mathcal{L}^p$-sense, i.e.

$$\lim_{h \downarrow 0} \int_0^{T-h} \int_{\mathbb{R}^d} \Big| \frac{\varphi_{t+h}(x) - \varphi_t(x)}{h} - v_t(\varphi_t(x)) \Big|^p d\mu_0(x) dt = 0 \tag{4.8}$$

and

$$\lim_{h \to 0} \frac{\varphi_{t+h}(x,t) - x}{h} = v_t(x) \text{ in } \mathcal{L}^p(\mu_t; \mathbb{R}^d) \text{ for } \mathcal{L}^1 - a.e. \ t \in (0,T) \tag{4.9}$$

Lemma 4.6. Absolutely continuous curves and the continuity equation, [2, Theorem 8.3.1] *Let I be an open interval in $\mathbb{R}$, let $\mu_t : I \to P_2(\mathbb{R}^d)$ be an absolutely continuous curve, and let $|\mu'| \in L^1(I)$ be its metric derivative, i.e.*

$$|\mu'|(t) := \lim_{s \to t} \frac{W_2(\mu_s, \mu_t)}{|s-t|}.$$

Then there exists a Borel vector field $v : (x,t) \mapsto v_t(x)$ *such that* $v_t \in \mathcal{L}^2(\mu_t; \mathbb{R}^d)$ *with*

$$\|v_t\|_{L^1(\mu_t;\mathbb{R}^d)} \leqslant |\mu'|(t) \text{ for } \mathcal{L}^1 - a.e.\ t \in I \tag{4.10}$$

and the continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \text{ in } \mathbb{R}^d \times I \tag{4.11}$$

holds in the sense of distributions, i.e.

$$\int_I \int_{\mathbb{R}^d} (\partial_t \psi(x,t) + \langle v_t(x)\ , \nabla_x \psi(x,t) \rangle) d\mu_t(x) dt = 0 \quad \forall \psi \in C_c^\infty(\mathbb{R}^d \times I) \tag{4.12}$$

Conversely, if a narrowly continuous curve $\mu_t : I \to P_2(\mathbb{R}^d)$ *satisfies the continuity equation for some Borel velocity field* v_t *with* $\|v_t\|_{L^2(\mu_t)\mathbb{R}^d} \in L^1(I)$*, then* $\mu_t : I \to P_2(\mathbb{R}^d)$ *is absolutely continuous and* $|\mu'|(t) \leqslant \|v_t\|_{L^2(\mu_t,\mathbb{R}^d)}$ *for* $\mathcal{L}^1$*−a.e.* $t \in I$*.*

The above lemma is also formulated in [36, Proposition 2.12].

4.2.2 Calculus on $P_2(\mathbb{R}^d)$

4.2.2.1 Tangent Spaces

Following [36, Section 2.3], let $\mathcal{X}_c$ denote the space of compactly-supported, smooth vector fields on $\mathbb{R}^d$. Let $\nabla C_c^\infty := \{\nabla f : f \in C_c^\infty\} \subset \mathcal{X}_c$. For $\mu \in P_2(\mathbb{R}^d)$, let $L^2(\mu, \mathbb{R}^d)$ denote the set of Borel maps $X : \mathbb{R}^d \to \mathbb{R}^d$ such that $\|X\|_\mu^2 = \int_{\mathbb{R}^d} \|X\|^2 d\mu < \infty$.

Definition 4.7. *[2, Definition 8.4.1]* Given $\mu \in P_2(\mathbb{R}^d)$, let $T_\mu P_2(\mathbb{R}^d)$ denote the closure of ∇C_c^∞ in $L^2(\mu)$, the **tangent space** of $P_2(\mathbb{R}^d)$ at μ. The **tangent bundle** $\mathcal{TP}_2(\mathbb{R}^d)$ is defined as the union of all such tangent spaces.

Definition 4.8. *[36, Definition 2.6]* Given $\mu \in P_2(\mathbb{R}^d)$, we define the divergence operator

$$div_\mu : \mathcal{X}_c \to (C_c^\infty)^*$$

by $\langle div_\mu(X)\ , f\,\rangle := \int_{\mathbb{R}^d} df(X)d\mu$

Lemma 4.9. [2, Lemma 8.4.2] *A vector* $f \in L^2(\mu;\mathbb{R}^d)$ *belongs to the tangent cone* $T_\mu P_2(\mathbb{R}^d)$ *if and only if*

$$\|f+g\,\|_{L^2(\mu)} \geqslant \|f\,\|_{L^2(\mu)} \quad \forall g \in L^2(\mu) \ s.t. \ \nabla \cdot (g\mu) = 0.$$

In particular, for every $f \in L^2(\mu)$*, there exists a unique* $\pi_\mu f \in T_\mu P_2(\mathbb{R}^d)$ *in the equivalence class of* f *modulo divergence-free vector fields which is the element of minimal* L^2*-norm in this class, and*

$$\int_{\mathbb{R}^d} \langle f\ , g - \pi_\mu g\,\rangle d\mu(x) = 0 \quad \forall f \in T_\mu P_2(\mathbb{R}^d), g \in L^2(\mu).$$

It is proved in [36], by Lemma 4.9, that one obtains the orthogonal decomposition:

$$L^2(\mu) = \overline{\nabla C_c^\infty}^{\mu} \oplus Ker(div_\mu)$$

so that one can define the projection $\pi_\mu : L^2(\mu) \to \overline{\nabla C_c^\infty}^{\mu}$.

4.2.2.2 Functionals and Their Subdifferentials

In what follows, we shall explain how the ideas above can be used to create a calculus for the Wasserstein space. This is the subject of [2, 53].

Definition 4.10. Let $F : P_2(\mathbb{R}^d) \to (-\infty, \infty]$ be a functional on the (2-)Wasserstein space. The the **domain** of F is $D(F) = \{\mu \in P_2(\mathbb{R}^d) : \quad F(\mu) < \infty\}$. A functional is **proper** if its domain is nonempty.

Definition 4.11. *[36, Definition 4.9]* If $F : P_2(\mathbb{R}^d) \to \mathbf{R}$ is a functional on $P_2(\mathbb{R}^d)$, then a function $\xi \in L^2(\mu)$ belongs to the **subdifferential** $\partial_- F(\mu)$, which we will also write as $\partial F(\mu)$, if

$$F(\nu) \geqslant F(\mu) + \sup_{\gamma \in \Gamma_0(\mu,\nu)} \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle \xi(x) \ , y - x \rangle d\gamma(x,y) + o(W_2(\mu,\nu))$$

as $\nu \to \mu$. Similarly, ξ belongs to the **superdifferential** $\partial^+ F(\mu)$ if $-\xi \in \partial(-F)(\mu)$. If $\exists\ \xi \in \partial_- F(\mu) \bigcap \partial^+ F(\mu)$ then for any $\gamma \in \Gamma_0(\mu,\nu))$ we have:

$$F(\nu) = F(\mu) + \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle \xi(x) \ , y - x \rangle d\gamma(x,y) + o(W_2(\mu,\nu). \tag{4.13}$$

In this case, F is differentiable at μ, and its **gradient vector** is $\nabla_\mu F := \pi_\mu(\xi)$.

Remark 4.12. To give a concrete example of the meaning of this differential in $P_2(\mathbb{R}^d)$, we consider the following two examples of the utility of the gradient given in [36]: For a differentiable functional $F : P_2(\mathbb{R}^d) \to \mathbf{R}$ and a compactly supported, smooth vector field $X \in \nabla C_c^\infty(\mathbb{R}^d)$, with flow φ_t,

a. If $\nu_t := (\iota + tX)_\# \mu$, then

$$F(\nu_t) = F(\mu) + t \int_{\mathbb{R}^d} \langle \nabla_\mu F \ , X \rangle d\mu + o(t).$$

b. If $\mu_t := \varphi_{t\#}\mu$ and $\| \nabla_\mu F \|$ is bounded on compact subsets of $P_2(\mathbb{R}^d)$, then

$$F(\mu_t) = F(\mu) + t \int_{\mathbb{R}^d} \langle \nabla_\mu F \ , X \rangle d\mu + o(t).$$

That is, the functions $t \mapsto F(\nu_t)$ and $t \mapsto F(\mu_t)$ are differentiable.

The above definition gives an intuition into the nature of these subdifferentials, but it is technically only correct for absolutely continuous measures; for this reason, we will give the following more technical definition which holds for a much more general class of measures and, indeed, can be extended to p-Wasserstein spaces with $p > 2$.

Definition 4.13. *[2, The strong subdifferential, Definition 10.3.1]* Let $\phi : P_2(\mathbb{R}^d) \to (-\infty, \infty]$ be a proper and lower semi-continuous functional, and let $\mu^1 \in D(\phi)$. Then $\gamma \in P_2(\mathbb{R}^d \times \mathbb{R}^d)$ belongs to the **extended Fréchet subdifferential** $\partial\phi(\mu^1)$ if $\pi^1_\# \gamma = \mu^1$ and

$$\phi(\mu^3) - \phi(\mu^1) \geqslant \inf_{\nu \in \Gamma_0(\gamma,\mu^3)} \iiint \langle x_2 \, , x_3 - x_1 \rangle d\nu + o(W_2(\mu^1, \mu^3)).$$

We say that $\gamma \in \partial\phi(\mu^1)$ is a **strong Fréchet subdifferential** if for every $\nu \in \Gamma(\gamma, \mu^3)$, it satisfies

$$\phi(\mu^3) - \phi(\mu^1) \geqslant \iiint \langle x_2 \, , x_3 - x_1 \rangle d\nu + o(C_{2,\nu}(\mu^1, \mu^3)), \tag{4.14}$$

where $C_{2,\nu}(\mu^1, \mu^3)$ is the pseudo-distance given by the cost

$$C^2_{2,\nu}(\mu^1, \mu^3) = \iiint \| x_1 - x_3 \|^2 d\nu(x_1, x_2, x_3).$$

The following definition was given for functionals on general metric spaces, but can be made specific to the Wasserstein space:

Definition 4.14. [2, Definition 1.2.4] The **metric slope** $|\partial\phi|(\mu)$ of a functional $\phi : P_2(\mathbb{R}^d) : (-\infty, \infty]$ at μ is given by

$$|\partial\phi|(\mu) = \limsup_{W_2(\mu,\nu)\to 0} \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\mu, \nu)}, \tag{4.15}$$

where $u+ = \max(0, u)$.

Definition 4.15. *[2, Regular functionals, Definition 10.3.9]* A proper, lower semi-continuous functional $\phi : P_2(\mathbb{R}^d) \to (-\infty, \infty]$ is regular if whenever the strong subdifferentials $\gamma_n \in \partial\phi(\mu_n)$ satisfy:

$$\phi(\mu_n) \to \varphi \in \mathbb{R}, \quad \mu_n \to \mu \quad \text{in } P_2(\mathbb{R}^d),$$

$$\sup_n M_2(\gamma_n) < \infty, \quad \gamma_n \to \gamma \quad \text{in } P(\mathbb{R}^d \times \mathbb{R}^d)$$

then $\gamma \in \partial\phi(\mu)$, and $\varphi = \phi(\mu)$.

4.2.2.3 Gradient Flows and the Variational Method

Now, the subdifferential $\partial\phi(\mu)$ of a functional ϕ at μ in $P_2(\mathbb{R}^d)$ may be multi-valued. Thus, we define a gradient flow in terms of a differential inclusion:

Definition 4.16. *[2, Definition 11.1.1]* Given a map $\mu_t \in AC^2_{loc}((0,\infty); P_2(\mathbb{R}^d))$ with $v_t \in Tan_{\mu_t}P_2(\mathbb{R}^d)$ the velocity vector field of μ_t, μ_t is a solution of the **gradient flow** equation

$$v_t \in -\partial\phi(\mu_t) \quad t > 0 \tag{4.16}$$

if v_t belongs to the subdifferential of ϕ at μ_t for a.e. $t > 0$, or, equivalently, $(\iota, -v_t)_{\#}\mu_t \in \partial\phi(\mu)$ for a.e. $t > 0$.

This may also be expressed as the requirement that there exist a Borel vector field v_t which that $v_t \in Tan_{\mu_t}P_2(\mathbb{R}^d)$ for a.e. $t > 0$, with $\|v_t\|_{L^2(\mu_t)} \in L^2_{loc}((0,\infty))$ satisfying the continuity equation in the sense of distributions and satisfying (4.16) for a.e. $t > 0$.

One approach to solving the gradient flow equation in the Wasserstein space is to draw an analogy with the usual setting of gradient flows on a Riemannian manifold and perform a time discretization of the steepest descent equation. This scheme was pioneered by [39], and its convergence is equivalent to the above formulation of the gradient flow, as laid out in [2, Chapter 11]. To describe this scheme, we will follow [40] and [2, Chapter 11.1.3].

Definition 4.17. *The Minimizing Movement Scheme* Assume the following:

A Let $\phi : P_2(\mathbb{R}^d) \to (-\infty, \infty]$ be a proper, lower semicontinuous functional such that

$$\nu \mapsto \Phi(\tau, \mu; \nu) := \frac{1}{2\tau} W_2^2(\mu, \nu) + \phi(\nu)$$

admits a minimum point for all $\tau \in (0, \tau^*)$ for $\mu \in P_2(\mathbb{R}^d)$ and some $\tau^* > 0$.

Fix a measure $\mu_0 \in P_2(\mathbb{R}^d)$. Given any step size $\tau > 0$, we can partition $(0, \infty]$ into $\bigcup_{n=1}^{\infty} I_n$, with $I_\tau^n := ((n-1)\tau, n\tau]$. For a given family of initial values M_τ^0 such that

$$M_\tau^n \to \mu_0 \quad \text{in } P_2(\mathbb{R}^d) \quad , \phi(M_\tau^0) \to \phi(\mu_0) \quad \text{as } \tau \downarrow 0$$

we can define for each $\tau \in (0, \tau^*)$ a family of sequences $\{M_\tau^n\}_{n=1}^{\infty}$ satisfying

$$M_\tau^n = \underset{\nu \in D(\phi)}{\arg\min}\, \Phi(\tau, M_\tau^{n-1}; \nu),$$

where the choice of M_τ^n may not be unique, but such a measure will always exist. Then the piecewise constant interpolant path in $P_2(\mathbb{R}^d)$,

$$\overline{M}_\tau(t) := M_\tau^n, \quad t \in ((n-1)\tau, n\tau],$$

is termed the **discrete solution**. A curve μ will be a **Generalized Minimizing Movement** for Φ and μ if there exists a sequence $\tau_k \downarrow 0$ such that

$$\overline{M}_{\tau_k}(t) \to \mu_t \quad \text{narrowly in } P(\mathbb{R}^d) \text{ for every } t > 0, \qquad \text{as } k \to \infty.$$

For $\mu \in D(\phi)$, by a compactness argument, this solution always exists and is an absolutely continuous curve $\mu \in AC^2_{loc}([0,\infty); P_2(\mathbb{R}^d))$.

As observed by [2] to illustrate the goal of this method, if we can restrict the domain of the functional ϕ and its gradient to the regular measures, then we can define a sequence of optimal transport maps T^n_τ pushing M^n_τ to M^{n-1}_τ. Then the discrete velocity vector can be defined as

$$V^n_\tau := \frac{T^n_\tau - \iota}{\tau} \in \partial\phi(M^n_\tau),$$

which is an implicit Euler discretization of (4.16). The piecewise constant interpolant

$$\overline{V}_\tau(t) := V^n_\tau \quad \text{for } t \in ((n-1)\tau, n\tau],$$

converges distributionally in $\mathbb{R}^d \times (0,\infty)$ up to subsequences to a vector field which solves the continuity equation. The problem which remains is proving that this vector field is also a solution of (4.16).

For regular functionals, without having to restrict ourselves to the convex case or to regular measures, it can be shown that this convergence occurs; the following lemma gives sufficient conditions for this convergence.

Before we state the key lemma, [2, Theorem 11.3.2.], we have the following important result about strong subdifferentials related to the metric slope of Definition 4.14.

Lemma 4.18. *[2, Theorem 10.3.11] Let ϕ be a regular functional on $P_2(\mathbb{R}^d)$ satisfying assumption A, and let μ be a point of strong subdifferentiability. Then there exists a unique plan $\gamma_0 \in \partial\phi(\mu)$ which attains the minimum*

$$|\gamma_0|_{2,2} = \min\{|\gamma|_{2,2} : \gamma \in \partial\phi(\mu)\}.$$

Indeed, when, for instance $\gamma_0 = (\iota, \xi)_{\#}\mu$, we can choose the **barycenter** $\xi \in L^2(()\mu)$ and denote it by the symbol $\partial^0\phi(\mu)$.

Lemma 4.19. [2, Theorem 11.3.2.] *Let $\phi : P_2(\mathbb{R}^d) \to (-\infty, \infty]$ be a proper and lower semicontinuous regular functional with relatively compact sublevel sets. Then for every initial datum $\mu_0 \in D(\phi)$, each sequence of discrete solutions $\overline{M}_{\tau_k}$ of the variational scheme admits a subsequence such that*

1. *$\overline{M}_{\tau_k}(t)$ narrowly converges in $P(\mathbb{R}^d)$ to μ_t locally uniformly in $[0, \infty)$, with $\mu_t \in AC_2^2([0, \infty); P_2(\mathbb{R}^d))$.*

2. *μ_t is a solution of the gradient flow equation*

$$v_t = -\partial^0\phi(\mu_t), \quad \|v_t\|_{L^2(\mu_t;\mathbb{R}^d)} = |\mu^{'}|(t), \text{ for a.e. } t > 0$$

   with $\mu_t \to \mu_0$ as $t \downarrow 0$, where v_t is the tangent vector to the curve μ_t.

3. *The energy inequality*

$$\int_a^b \int_{\mathbb{R}^d} |v_t(x)|^2 d\mu_t(x) dt + \phi(\mu_b) \leqslant \phi(\mu_a)$$

   holds for every $b \in [0, \infty)$ and $a \in [0, b)\backslash\mathcal{N}$, where $\mathcal{N}$ is a $\mathcal{L}^1$-negligible subset of $(0, \infty)$.

4.3 Frame Forces

With the above gradient flow framework established, we can apply it to potentials useful for characterizing probabilistic frames.

4.3.1 The Frame Potential

To begin our discussion of frame forces, we define the frame potential for finite frames and the analogous quantity for probabilistic frames.

Definition 4.20. Given a probabilistic frame μ, the **probabilistic frame potential** for μ is given by

$$PFP(\mu) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} \langle x\ , y \rangle^2 d\mu(x) d\mu(y) \tag{4.17}$$

As a special case, we define the **frame potential** for a finite frame, $\Phi = \{\varphi_i\}_{i=1}^N \subset \mathbb{R}^d$, by

$$FP(\Phi) = \sum_{i,j=1}^{N} \langle \varphi_i\ , \varphi_j \rangle^2 = N^2 PFP(\mu_\Phi) \tag{4.18}$$

Remark 4.21. The frame potential is a well-studied object. In their celebrated paper on finite unit-norm tight frames (FUNTFs), Benedetto and Fickus establish that, among all unit-norm frames, FUNTFs are the minimizers of equation 4.18 [5]. Because FUNTFs (and tight frames in general) have a multitude of uses in pure mathematics, statistics, and coding theory, this consequently made the frame potential a very useful quantity. The frame potential and related potentials are also studied in the context of spherical t-designs.

In what follows, we explore several functionals on the space $P_2(\mathbb{R}^d)$ related to questions in frame theory, starting with the probabilistic frame potential. For equal-norm frames restricted to spheres, this potential is sufficient to identify tightness. For more general probabilistic frames, we tweak this to a quantity we term the tightness potential. We also explore higher-order potentials related to other classes of tight frames.

4.3.2 Locating tight probabilistic frames

Some further analysis is needed before we can use the frame potential to find probabilistic tight frames. In the following propositions and lemmas, we narrow our search space, establish a lower bound on how close the nearest probabilistic tight frame can be, and show that, as in the finite case, the frame potential is indeed a crucial quantity in constructing gradient flows that will lead us to tight probabilistic frames.

In fact, for a given probabilistic frame μ, we have control on the spectrum of the frame operators of the measures nearby in $P_2(\mathbb{R}^d)$, as seen in the next result.

Lemma 4.22. *Suppose $\{\nu_n\}$ is a sequence converging to μ in $P_2(\mathbb{R}^d)$. Then there exists some positive constant C_μ such that $\|S_{\nu_n} - S_\mu\| \leqslant C_\mu W_2(\mu, \nu_n)$. In particular, convergence of a sequence of measures in the Wasserstein space implies the convergence of their frame operators.*

Proof. Since $\nu_n \longrightarrow \mu$ in $P_2(\mathbb{R}^d)$, for n sufficiently large, $M_2(\nu_n) \leqslant 2M_2(\mu)$. Then, for $\gamma_n \in \Gamma(\nu_n, \mu)$,

$$\begin{aligned}
\|S_{\nu_n} - S_\mu\| &= \max_{v \in S^{d-1}} v^\top (S_{\nu_n} - S_\mu) v \\
&= \max_{v \in S^{d-1}} \iint_{\mathbb{R}^d \times \mathbb{R}^d} (\langle x\ , v\rangle^2 - \langle y\ , v\rangle^2) d\gamma_n(x,y) \\
&= \max_{v \in S^{d-1}} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \langle v\ , x-y\rangle \langle x+y\ , v\rangle d\gamma_n(x,y) \\
&\leqslant \max_{v \in S^{d-1}} \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\langle v\ , x-y\rangle)^2 d\gamma_n(x,y) \right)^{\frac{1}{2}} \cdot \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\langle v\ , x+y\rangle)^2 d\gamma_n(x,y) \right)^{\frac{1}{2}} \\
&\leqslant \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\|x-y\|)^2 d\gamma_n(x,y) \right)^{\frac{1}{2}} \cdot \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\|x+y\|)^2 d\gamma_n(x,y) \right)^{\frac{1}{2}} \\
&\leqslant \sqrt{2} C_{2,\gamma_n}(\mu, \nu_n) \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} \|x\|^2 + \|y\|^2 d\gamma_n(x,y) \right)^{\frac{1}{2}} \\
&\leqslant \sqrt{2} C_{2,\gamma_n}(\mu, \nu_n) \cdot 3 M_2(\mu)
\end{aligned}$$

where the last inequality holds for n sufficiently large. In particular, if we choose $\gamma_n \in \Gamma_0(\nu_n, \mu)$, then for n sufficiently large,

$$\|S_{\nu_n} - S_\mu\| \leqslant \sqrt{2} W_2(\mu, \nu_n) \cdot 3 M_2(\mu) \tag{4.19}$$

□

This control on the spectrum of the frame operator allows us to prove the following:

Proposition 4.23. *Let $\{\nu_n\}$ be a sequence converging in $P_2(\mathbb{R}^d)$ to a probabilistic frame μ. Then there exists N sufficiently large such that $\forall n \geqslant N$, ν_n is also a probabilistic frame.*

Proof. Let $\{\nu_n\}$ and μ be as above, and let S_{ν_n}, S_μ denote the matrix representations of their respective frame operators which exist since the measures in question are in $P_2(\mathbb{R}^d)$. Let the eigenvalues of S_{ν_n} be given by $\lambda_1(S_{\nu_n}) \leqslant \cdots \leqslant \lambda_d(S_{\nu_n})$. Then

$$\begin{aligned}
\lambda_1(S_\mu) &= \min_{v \in S^{d-1}} \langle v\ , S_\mu v \rangle \\
&= \min_{v \in S^{d-1}} (\langle v\ , S_\mu v \rangle - \langle v\ , S_{\nu_n} v \rangle + \langle v\ , S_{\nu_n} v \rangle) \\
&\leqslant \langle x\ , S_\mu x \rangle - \langle x\ , S_{\nu_n} x \rangle + \langle x\ , S_{\nu_n} x \rangle \qquad \forall x \in S^{d-1} \\
&\leqslant \max_{v \in S^{d-1}} (\langle v\ , S_\mu v \rangle - \langle v\ , S_{\nu_n} v \rangle) + \langle x\ , S_{\nu_n} x \rangle \qquad \forall x \in S^{d-1} \\
&= \lambda_d(S_\mu - S_{\nu_n}) + \langle x\ , S_{\nu_n} x \rangle \qquad \forall x \in S^{d-1}
\end{aligned}$$

Since the last statement above holds for all x in S^{d-1}, it holds in particular for $x_* := \arg\min_{x \in S^{d-1}} \langle x\ , S_{\nu_n x} \rangle$. Hence

$$\lambda_1(S_\mu) \leqslant \lambda_d S_\mu - S_{\nu_n} + \lambda_1(S_{\nu_n}).$$

Therefore, since by Lemma 4.22,

$$|\lambda_d(S_\mu - S_{\nu_n})| \leqslant \|S_\mu - S_{\nu_n}\| \to 0$$

as $\nu_n \to \mu$ in $P_2(\mathbb{R}^d)$, given $\alpha \in (0,1)$, we can choose N such that $\forall n \geqslant N$,

$$|\lambda_d(S_\mu - S_{\nu_n})| < \alpha \cdot \lambda_1(S_\mu),$$

and for such n,

$$\lambda_1(S_{\nu_N}) > (1-\alpha)\lambda_1(S_\mu) > 0.$$

$\square$

As one might expect, given a probabilistic frame μ, this control also allows us to obtain a lower limit on the distance in $P_2(\mathbb{R}^d)$ to the nearest tight frame.

Proposition 4.24. *Suppose μ is a probabilistic frame for $\mathbb{R}^d$ which is not tight. Let*

$$\delta := \lambda_d(\mu) - \lambda_1(\mu) \in (0, \lambda_d(\mu)).$$

Then for any tight frame ν, $W_2(\mu,\nu) \geqslant \frac{\delta}{4(M_2(\mu)+M_2(\nu))}$.

Proof. From Lemma 4.22, we know that measures close to μ in $P_2(\mathbb{R}^d)$ will have frame operators whose spectra are close to that of the frame operator of μ. Let ν be a tight frame with frame constant $A := \frac{M_2^2(\nu)}{d}$. Then

$$\max_k |\lambda_k(\mu) - \lambda_k(\nu)| = \max\{|\lambda_1(\mu) - A|, |\lambda_d(\mu) - A|\} \geqslant \frac{\delta}{2}. \tag{4.20}$$

Moreover, for any $k \in \{1, ..., d\}$, $|\lambda_k(\mu) - \lambda_k(\nu)| \leqslant \|S_\nu - S_\mu\|$. Therefore, since from the proof of Lemma 4.22 we know that for any $\gamma \in \Gamma_0(\mu,\nu)$,

$$\begin{aligned}\|S_\nu - S_\mu\| &\leqslant \sqrt{2}W_2(\mu,\nu) \cdot \left(\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \|x\|^2 + \|y\|^2 d\gamma(x,y)\right)^{\frac{1}{2}} \\ &\leqslant 2W_2(\mu,\nu)\cdot(M_2(\mu)+M_2(\nu)),\end{aligned}$$

it follows from (4.20) that $W_2(\mu,\nu) \geqslant \frac{\delta}{4(M_2(\mu)+M_2(\nu))}$. □

Remark 4.25. We note that if

$$\mathrm{supp}(\mu) \subset \{x \in \mathbb{R}^d : \quad (1-\epsilon) \leqslant \|x\| \leqslant (1+\epsilon)\}$$

and

$$\forall k \in \{1,...,d\}, \quad (\frac{M_2^2(\mu)}{d} - \epsilon) \leqslant \lambda_k(\mu) \leqslant (\frac{M_2^2(\mu)}{d} + \epsilon),$$

then the lower bound on the Wasserstein distance to the nearest probabilistic tight frame ν supported on S^{d-1} can be pushed correspondingly small:

$$\frac{\lambda_d(\mu) - \lambda_1(\mu)}{4(M_2(\mu)+M_2(\nu))} \leqslant \frac{2\epsilon}{4((1-\epsilon)+1)} = \frac{\epsilon}{3-2\epsilon}.$$

This relates to the problem in finite frame theory of finding the closest unit-norm tight frame to a given ϵ-nearly unit norm, ϵ-nearly tight frame.

We also note that the identification of tight frames with minimizers of the frame potential holds in the case of probabilistic frames. The next theorem depends on a result due to [29, Theorem 4.2], a version of which we reproduce in the following lemma.

Lemma 4.26. *Let μ be a measure in $P_2(\mathbb{R}^d)$. The the following bound holds for the probabilistic frame potential: $PFP(\mu) \geqslant \frac{M_2^4(\mu)}{d}$.*

Proof. Note that, writing $m_{i,j}(\mu) = \int_{\mathbb{R}^d} x_i x_j d\mu(x)$, we have:

$$\begin{aligned} PFP(\mu) &= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle x\ , y\,\rangle^2 d\mu(x) d\mu(y) \\ &= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \sum_{i=1}^{d} \sum_{j=1}^{d} x_i y_i x_j y_j d\mu(x) d\mu(y) \\ &= \sum_{i=1}^{d} \sum_{j=1}^{d} m_{i,j}^2(\mu) \end{aligned}$$

And by Hölder,

$$\begin{aligned} M_2^2(\mu) &= \sum_{i=1}^{d} m_{i,i}(\mu) \\ &\leqslant \left(\sum_{i=1}^{d} m_{i,i}^2(\mu) \right)^{\frac{1}{2}} \left(\sum_{i=1}^{d} 1 \right)^{\frac{1}{2}} \\ &\leqslant \sqrt{d} \left(\sum_{i=1}^{d} \sum_{j=1}^{d} m_{i,j}^2(\mu) \right)^{\frac{1}{2}} \end{aligned}$$

Therefore, $\sum\limits_{i=1}^{d} \sum\limits_{j=1}^{d} m_{i,j}^2(\mu) \geqslant \frac{M_2^4(\mu)}{d}$, and the result follows. □

Remark 4.27. Clearly, minimizers exist. In particular, if μ is a tight probabilistic frame, then equality holds in the above claim, since the frame bound of a probabilistic tight frame μ is precisely $\frac{M_2^2(\mu)}{d}$, and

$$
\begin{aligned}
PFP(\mu) = \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\ ,y\rangle^2 d\mu(x)d\mu(y) \quad &= \int_{\mathbb{R}^d} \langle S_\mu y\ ,y\rangle d\mu(y) \\
= \int_{\mathbb{R}^d} \frac{M_2^2(\mu)}{d}\|y\|^2 d\mu(y) \quad &= \frac{M_2^4(\mu)}{d}
\end{aligned}
$$

Theorem 4.28. *A probabilistic frame μ with $M_2(\mu) = 1$ is tight if and only if it is a minimizer among $\{\nu \in P_2(\mathbb{R}^d) : M_2(\nu) = 1\}$ of the probabilistic frame potential.*

Proof. The necessity is clear from Remark 4.27. For the sufficiency, we consider a measure μ in $P_2(\mathbb{R}^d)$ which minimizes the probabilistic frame potential among $\{\nu \in P_2(\mathbb{R}^d) : M_2(\nu) = 1\}$. Given any $\nu \in \{\nu \in P_2(\mathbb{R}^d) : M_2(\nu) = 1\}$, and $\lambda \in [0,1]$, let $\mu_\lambda := \lambda\mu + (1-\lambda)\nu$. That is, given a test function $f(x)$ with at most quadratic growth,

$$
\int_{\mathbb{R}^d} f(x)d\mu_\lambda(x) = \lambda\int_{\mathbb{R}^d} f(x)d\mu(x) + (1-\lambda)\int_{\mathbb{R}^d} f(x)d\nu(x).
$$

Then

$$
\begin{aligned}
M_2^2(\mu_\lambda) &= \int_{\mathbb{R}^d} \|x\|^2 d\mu_\lambda(x) \\
&= (\lambda)\int_{\mathbb{R}^d} \|x\|^2 d\mu(x) + (1-\lambda)\int_{\mathbb{R}^d} \|x\|^2 d\nu(x) \\
&= \lambda M_2^2(\mu) + (1-\lambda)M_2^2(\nu) \\
&= 1
\end{aligned}
$$

Therefore, since it follows that $PFP(\mu) \leqslant PFP(\mu_\lambda)\ \forall\lambda \in [0,1]$, we obtain:

$$
0 \leqslant PFP(\mu_\lambda) - PFP(\mu)
$$

$$
\begin{aligned}
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\ ,y\rangle^2 d\mu_\lambda(x)d\mu_\lambda(y) - \iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\ ,y\rangle^2 d\mu(x)d\mu(y)\\
&= (\lambda^2-1)PFP(\mu) + (1-\lambda)^2 PFP(\nu) + 2\lambda(1-\lambda)\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\ ,y\rangle^2 d\mu(x)d\nu(y)\\
&= (\lambda^2-1)PFP(\mu) + (1-\lambda)^2 PFP(\nu) + 2\lambda(1-\lambda)\int_{\mathbb{R}^d} \langle y\ ,S_\mu y\rangle d\nu(y)\\
&= (\lambda-1)\left((\lambda+1)PFP(\mu) - (1-\lambda)PFP(\nu) - 2\lambda\int_{\mathbb{R}^d}\langle y\ ,S_\mu y\rangle d\nu(y)\right)\\
&\leqslant (\lambda-1)\left(\frac{(\lambda+1)}{d} - \frac{(1-\lambda)}{d} - 2\lambda\int_{\mathbb{R}^d}\langle y\ ,S_\mu y\rangle d\nu(y)\right)\\
&= (\lambda-1)\left(\frac{2\lambda}{d} - 2\lambda\int_{\mathbb{R}^d}\sum_{k=1}^{d}\lambda_k\langle y\ ,v_k\rangle^2 d\nu(y)\right)
\end{aligned}
$$

where the second inequality comes from the fact that $PFP(\nu) \geqslant \frac{M_2^4(\nu)}{d} = \frac{1}{d}$ and $PFP(\mu) = \frac{1}{d}$, and in the last equality, the values $\{\lambda_k\}_{k=1}^d$ are the eigenvalues of the frame operator S_μ, and the $\{v_k\}_{k=1}^d$ are the corresponding orthonormal set of eigenvectors guaranteed by the spectral theorem. From this inequality it follows that

$$
\int_{\mathbb{R}^d}\sum_{k=1}^{d}\lambda_k\langle y\ ,v_k\rangle^2 d\nu(y) \geqslant \frac{1}{d}.
$$

Let λ_1 denote the smallest eigenvalue of μ, and v_1 the corresponding eigenvector of S_μ. Since ν was chosen arbitrarily in $\{\nu \in P_2(\mathbb{R}^d) : M_2(\nu) = 1\}$, it follows that for any $\epsilon > 0$, one can choose $d\nu = (1-\epsilon)\delta_{v_1} + \frac{\epsilon}{d-1}\sum_{k=2}^{d}\delta_{v_k}$. Then

$$
\begin{aligned}
\frac{1}{d} &\leqslant \int_{\mathbb{R}^d}\sum_{k=1}^{d}\lambda_k\langle y\ ,v_k\rangle^2 d\nu(y)\\
&= (1-\epsilon)\lambda_1\|v_1\|^2 + \frac{\epsilon}{d-1}\sum_{k=2}^{d}\lambda_k\|v_k\|^2\\
&= (1-\epsilon)\lambda_1 + \frac{\epsilon}{d-1}\sum_{k=2}^{d}\lambda_k
\end{aligned}
$$

and as $\epsilon \longrightarrow 0$, we see that, in fact, $\lambda_1 \geqslant \frac{1}{d}$. Since $\lambda_1 \leqslant \frac{M_2^2(\eta)}{d}$ for any probabilistic frame η, with equality if and only if η is tight, it follows that our minimizer of the probabilistic frame potential, μ, is tight. $\square$

Moreover, we can broaden the above result to assert the following:

Theorem 4.29. *Given a measure $\mu \in P_2(\mathbb{R}^d)$, $\mu \neq \delta_{\{0\}}$, $PFP(\mu) = \frac{M_2^4(\mu)}{d}$ if and only if μ is tight or $\mu = \delta_0$.*

Proof. Again, if μ is a tight probabilistic frame, then the equality clearly holds by Remark 4.27. Suppose that μ is not tight. Then the eigenvalues of S_μ are $\lambda_1 \geqslant \cdots \geqslant \lambda_d$ with $\lambda_1 > \frac{M_2^2(\mu)}{d} > \lambda_d$ with a corresponding orthonormal basis of eigenvectors $\{v_i\}_{i=1}^d$ for $\mathbb{R}^d$. Then

$$\begin{aligned} PFP(\mu) &= \iint \langle x \,, y \rangle^2 d\mu(x) d\mu(y) \quad = \int \langle y \,, S_\mu y \rangle d\mu(y) \\ &= \int \langle y \,, \sum_{i=1}^d \lambda_i v_i v_i^\top y \rangle d\mu(y) \\ &= \sum_{i=1}^d \lambda_i \int \langle v_i \,, y \rangle^2 d\mu(y) \\ &= \sum_{i=1}^d \lambda_i \langle v_i \,, S_\mu v_i \rangle \quad = \sum_{i=1}^d \lambda_i^2 \end{aligned}$$

But, by Hölder,

$$\sum_{i=1}^d \lambda_i^2 > \frac{1}{d} \left(\sum_{i=1}^d \lambda_i \right)^2 \quad = \frac{M_2^4(\mu)}{d}$$

with equality if and only if $\lambda_1 = \cdots = \lambda_d$, that is, if and only if μ is tight. $\square$

4.3.3 The Tightness Potential

With propositions 4.26 and 4.29 in hand, we define the tightness potential and use a method outlined in [2] to show that a gradient flow solution exists for its minimization.

Definition 4.30. Given $\mu \in P_2(\mathbb{R}^d)$, we define the **tightness potential** $TP(\mu)$ by

$$\begin{aligned} TP(\mu) &= PFP(\mu) - \frac{M_2^4(\mu)}{d} \\ &= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \left[\langle x\ , y \rangle^2 - \frac{\|x\|^2 \|y\|^2}{d} \right] d\mu(x) d\mu(y) \end{aligned}$$

Definition 4.31. For $\mu \in P_2(\mathbb{R}^d)$, we also define the **tightness operator** $T_\mu : \mathbb{R}^d \to \mathbb{R}^d$ by

$$T_\mu(x) := \int_{\mathbb{R}^d} \left[\langle x\ , y \rangle y - \frac{1}{d} \|y\|^2 x \right] d\mu(y) = S_\mu x - \frac{M_2^2(\mu)}{d} x.$$

We immediately obtain:

Proposition 4.32. *For a measure* $\mu \in P_2(\mathbb{R}^d)$, $\|T_\mu\| \leqslant (TP(\mu))^{\frac{1}{2}}$.

Proof. Given $\mu \in P_2(\mathbb{R}^d)$, let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d \geqslant 0$ be the eigenvalues of S_μ. Noting that $M_2^2(\mu) = \sum\limits_{i=1}^{d} \lambda_i$, we have the following equivalence for the tightness potential:

$$\begin{aligned} TP(\mu) &= \iint \langle x\ , y \rangle^2 - \frac{1}{d} \|x\|^2 \|y\|^2 d\mu(x) d\mu(y) \\ &= Tr(S_\mu)^2 - \frac{1}{d} M_2^4(\mu) \\ &= \sum_{i=1}^{d} \lambda_i^2 - \frac{1}{d} \left(\sum_{i=1}^{d} \lambda_i \right)^2 \\ &= \frac{1}{d} \sum_{i=1}^{d} \sum_{j>i} (\lambda_i - \lambda_j)^2 \end{aligned}$$

Now, $\|T_\mu\| = \max\{\lambda_1 - \frac{1}{d}\sum \lambda_i, \frac{1}{d}\sum \lambda_i - \lambda_d\}$. Without loss of generality, let $\|T_\mu\| = \lambda_1 - \frac{1}{d}\sum \lambda_i$. Then, by Cauchy's inequality, noting that $\lambda_k - \lambda_j \geqslant 0$ if $k > j$,

$$
\begin{aligned}
(\frac{1}{d}\sum_{i=1}^{d}\sum_{j>i}(\lambda_i - \lambda_j)^2)^{\frac{1}{2}} &\geqslant \frac{1}{d}\sum_{i=1}^{d}\sum_{j>i}(\lambda_i - \lambda_j) \\
&= \frac{1}{d}[(d-1)\lambda_1 - \sum_{j>1}\lambda_j + \sum_{j>1}\sum_{k>j}(\lambda_j - \lambda_k) \\
&\geqslant \lambda_1 - \frac{1}{d}\sum_{j=1}^{d}\lambda_j
\end{aligned}
$$

From the above, we see that $TP(\mu) \geqslant \|T_\mu\|^2$, with equality if and only if $\lambda_i = \lambda_j$ $\forall i, j$. □

Corollary 4.33. *The tightness potential is zero if and only if μ is tight.*

Proof. Clearly, if μ is a tight probabilistic frame, then $TP(\mu) = 0$. If μ is not tight, then $\|T_\mu\|^2 > 0$, so that by the above, $TP(\mu) > 0$. □

4.3.4 Construction of gradient flows for the tightness potential

Most approaches to establishing the well-posedness of a gradient flow for a particular potential use the convexity or λ-convexity of the functional, if it can be established.

Definition 4.34. A function W on $\mathbb{R}^d \times \mathbb{R}^d$ is said to be **λ-convex** for some $\lambda \in \mathbb{R}$ if the function $(x, y) \mapsto W(x, y) - \frac{\lambda}{2}(\|x\|^2 + \|y\|^2)$ is convex.

For instance, [15] considers a class of potentials $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ describing the interaction of two particles of unit mass at positions x and y by the value $W(x, y)$. The total energy of a distribution under this potential is then given by the functional

$$\mathcal{W}[\mu] := \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x,y)d\mu(x)d\mu(y) \tag{4.21}$$

They assume the λ-convexity of the functional, however. While the tightness potential has a similar form it is *not* λ-convex. However, they still define well-posed gradient flow problems in the Wasserstein space.

Proposition 4.35. *The tightness potential is not λ-convex on $P_2(\mathbb{R}^d)$.*

Proof. Define the function $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$W(x,y) := \langle x\ , y\,\rangle^2 - \frac{\|x\,\|^2\|y\,\|^2}{d}.$$

Then, writing $w := \begin{bmatrix} x \\ y \end{bmatrix}$, W can be rewritten as:

$$W(x,y) = \frac{1}{4}\langle w\ , Kw\,\rangle^2 - \frac{1}{d}\langle w\ , I_1 w\,\rangle\langle w\ , I_2 w\,\rangle,$$

where K, I_1, $I_2 \in \mathbb{R}^{2d\times 2d}$ are given by

$$K = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, I_1 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \text{ and } \quad I_2 = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$$

By [2, Proposition 9.3.2, Remark 9.3.3., and Proposition 9.3.5], it is sufficient to show that W is not λ-convex on $\mathbb{R}^d \times \mathbb{R}^d$. Differentiating twice, we obtain the Hessian of $W(x,y)$:

$$\begin{aligned}\nabla^2 W(x,y) = \langle w\ , Kw\,\rangle K + 2Kww^\top K - \frac{4}{d}\left(I_1 ww^\top I_2 + I_2 ww^\top I_1\right) \\ - \frac{2}{d}\left(\langle w\ , I_2 w\,\rangle I_1 + \langle w\ , I_1 w\,\rangle I_2\right)\end{aligned}$$

$$= 2 \begin{bmatrix} yy^\top - \frac{1}{d}\|y\|^2 I & yx^\top - \frac{2}{d}xy^\top + \langle x\ , y\rangle I \\ xy^\top - \frac{2}{d}yx^\top + \langle x\ , y\rangle I & xx^\top - \frac{1}{d}\|x\|^2 \end{bmatrix}^\top$$

Therefore, given $\begin{bmatrix} u \\ v \end{bmatrix} \in S^{2d-1}$, we write

$$\frac{1}{2}\begin{bmatrix} u \\ v \end{bmatrix}^\top \nabla^2 W(x,y) \begin{bmatrix} u \\ v \end{bmatrix} = \langle \begin{bmatrix} u \\ v \end{bmatrix} , \begin{bmatrix} y \\ x \end{bmatrix} \rangle^2 + 2\langle u\ , v\rangle\langle x\ , y\rangle - \frac{4}{d}\langle u\ , x\rangle\langle v\ , y\rangle$$
$$- \frac{1}{d}(\|y\|^2\|u\|^2 + \|x\|^2\|v\|^2)$$

Thus, if we take $\begin{bmatrix} u \\ v \end{bmatrix} \in S^{2d-1}$ with $u \perp v$, $u \neq 0$, $v \neq 0$, we can find $y = R\frac{u}{\|u\|}$ and $x = R\frac{v}{\|v\|}$. Then

$$\frac{1}{2}\begin{bmatrix} u \\ v \end{bmatrix}^\top \nabla^2 W(x,y) \begin{bmatrix} u \\ v \end{bmatrix} = 0 + 0 - 4\frac{R^2\|u\|\|v\|}{d} - \frac{2R^2}{d}$$

Hence for every $\lambda \in \mathbb{R}$, taking $R = \sqrt{|\lambda| d}$, from the above we see that there exists $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ for which the minimum eigenvalue of $\nabla^2 W(x, y)$ is less than $-|\lambda|$. Thus, W is not λ-convex for any λ in $\mathbb{R}$. $\square$

Because we cannot use λ-convexity, we use the minimizing movement scheme and related existence result for regular measures. For this approach, we establish a few facts about the frame and tightness potentials.

Theorem 4.36. *The frame potential* $F(\mu) := \iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle x\ , y\rangle^2 d\mu(x)d\mu(y)$ *is a strongly differentiable function on* $P_2(\mathbb{R}^d)$.

Proof. Take $\mu, \nu \in P_2(\mathbb{R}^d)$ as above. Define $f_\mu(x) = S_\mu x$. Then $F(\mu) = \int_{\mathbb{R}^d} \langle f_\mu(x)\ , x\, \rangle d\mu(x)$, and for $\gamma \in \Gamma(\mu, \nu)$,

$$
\begin{aligned}
F(\nu) - F(\mu) &= \int_{\mathbb{R}^d} \langle S_\nu y\ , y\, \rangle d\nu(y) - \int_{\mathbb{R}^d} \langle S_\mu x\ , x\, \rangle d\mu(x) \\
&= \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} 4\langle S_\mu x\ , (y-x)\, \rangle d\gamma(x,y) + \\
&\quad \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} 3\langle S_\mu x\ , x\, \rangle - 4\langle S_\mu x\ , y\, \rangle + \langle S_\nu y\ , y\, \rangle d\gamma(x,y)
\end{aligned}
$$

Then, considering the second term in the preceding line, for ν sufficiently close to μ,

$$
\begin{aligned}
&\left| \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} 3\langle S_\mu x\ , x\, \rangle - 4\langle S_\mu x\ , y\, \rangle + \langle S_\nu y\ , y\, \rangle d\gamma(x,y) \right| \\
&\quad = \left| \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle S_\nu (y-x)\ , y-x\, \rangle + \langle S_\mu (y-x)\ , y-x\, \rangle \right. \\
&\qquad \left. + 2\langle (S_\mu - S_\nu)x\ , x-y\, \rangle + \langle S_\nu x\ , x\, \rangle - \langle S_\mu y\ , y\, \rangle d\gamma(x,y) \right| \\
&\quad = \left| \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle S_\nu (y-x)\ , y-x\, \rangle + \langle S_\mu (y-x)\ , y-x\, \rangle \right. \\
&\qquad \left. + 2\langle (S_\mu - S_\nu)x\ , x-y\, \rangle d\gamma(x,y) \right| \\
&\quad \leqslant \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} \| S_\nu \| \| y-x \|^2 + \| S_\mu \| \| y-x \|^2 + 2\| (S_\mu - S_\nu)x \| \| x-y \| d\gamma(x,y)| \\
&\quad \leqslant (\| S_\nu \| + \| S_\mu \|) C_{2,\gamma}^2(\mu,\nu) + 2\| S_\mu - S_\nu \| \cdot M_2(\mu) \cdot C_{2,\gamma}(\mu,\nu) \\
&\quad \leqslant (\| S_\nu \| + \| S_\mu \|) C_{2,\gamma}^2(\mu,\nu) + 6\sqrt{2} M_2^2(\mu) \cdot C_{2,\gamma}^2(\mu,\nu)
\end{aligned}
$$

where the second equality comes from the cancellation of the cross-frame potential, and the last inequality comes from the CBS inequality and Lemma 4.22.

Therefore, $F(\nu) - F(\mu) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} 4\langle S_\mu x\ , y-x\rangle + o(W_2(\mu,\nu))$ for ν sufficiently close to μ, and it follows that the gradient vector of $F(\mu)$ is $\nabla_\mu F = 4S_\mu(x)$. Moreover, since

$$\begin{aligned}\|\nabla_\mu F(\mu)|\,\|_\mu &= \int \|4S_\mu x\,\|^2 d\mu(x)\\ &\leqslant 16\int \|S_\mu\,\|^2\|x\,\|^2 d\mu(x)\\ &\leqslant 16M_2^4(\mu),\end{aligned}$$

it follows that $\|\nabla_\mu F(\mu)\,\|_{L^2(\mu)}$ is bounded on compact subsets of $P_2(\mathbb{R}^d)$. □

Theorem 4.37. *The square of the second moment* $M_2^2(\mu) := \iint_{\mathbb{R}^d\times\mathbb{R}^d}\|x\,\|^2 d\mu(x)$. *Furthermore, any even power of the second moment is a strongly differentiable function on* $P_2(\mathbb{R}^d)$.

Proof. Take ν and μ as above. Then

$$\begin{aligned}M_2^2(\nu) - M_2^2(\mu) &= \int_{\mathbb{R}^d}\|y\,\|^2 d\nu(y) - \int_{\mathbb{R}^d}\|x\,\|d\mu(x)\\ &= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d}\|y\,\|^2 - \|x\,\|^2 d\gamma(x,y)\\ &\quad \text{for } \gamma\in\Gamma_0(\mu,\nu)\\ &= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d}\langle y-x\ , y+x\,\rangle d\gamma(x,y)\\ &= 2\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d}\langle x\ , y-x\,\rangle d\gamma(x,y) + \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d}\|x-y\,\|^2 d\gamma(x,y)\end{aligned}$$

Therefore, $M_2^2(\nu) - M_2^2(\mu) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} 2\langle x\ , y-x\,\rangle + o(C_{2,\gamma}(\mu,\nu))$ for ν sufficiently close to μ, and it follows that the gradient vector of $M_2^2(\mu)$ is $\nabla_\mu F = 2x$.

To prove the second statement of the theorem, we will proceed by induction. Suppose that for $j\in\{1,\cdots,k\}$, $M_2^{2j}(\mu)$ is a differentiable functional with gradient

$\nabla_\mu M_2^{2j} = 2jM_2^{2(j-1)}(\mu)x$. Then, with ν, γ as above,

$$\begin{aligned}
M_2^{2k}(\nu) - M_2^{2k}(\mu) &= M_2^{2(k-1)}(\mu)\left(M_2^2(\nu) - M_2^2(\mu)\right) + M_2(\nu)\left(M_2^{2(k-1)}(\nu) - M_2^{2(k-1)}\right) \\
&= M_2^{2(k-1)}(\mu)\left(M_2^2(\nu) - M_2^2(\mu)\right) + M_2^2(\mu)\left((M_2^{2(k-1)}(\nu) - M_2^{2(k-1)}\right) \\
&+ \left(M_2^2(\nu) - M_2^2(\mu)\right)\left(M_2^{2(k-1)}(\nu) - M_2^{2(k-1)}\right) \\
&= M_2^{2(k-1)}\left(\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle 2x \ , y - x\rangle d\gamma(x,y) + o(C_{2,\gamma}(\mu,\nu))\right) + \\
& M_2^2(\mu)\left((k-1)M_2^{2(k-2)}(\mu)\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle 2x \ , y - x\rangle d\gamma(x,y) + o(C_{2,\gamma}(\mu,\nu))\right) + \\
& \left(M_2^2(\nu) - M_2^2(\mu)\right)\left(M_2^{2(k-1)}(\nu) - M_2^{2(k-1)}(\mu)\right) \\
&= kM_2^{2(k-1)}\iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle 2x \ , y - x\rangle d\gamma(x,y) + o(C_{2,\gamma}(\mu,\nu))
\end{aligned}$$

where we have used the inductive hypothesis for the second to last equality. Hence $\nabla_\mu M_2^{2k} = 2kM_2^{2(k-1)}x$. □

Theorem 4.38. *The tightness potential is differentiable, and the gradient of the tightness potential lies in its strong subdifferential.*

Proof. Given $\mu \in P_2(\mathbb{R}^d)$, take $\gamma = (\iota, 4T_\mu)_{\#}\mu$. Then by Theorems 4.36 and 4.37, γ clearly satisfies equation (4.14). □

Moreover, we have that this gradient is the minimal selection in the strong subdifferential:

Proposition 4.39. *Given $\mu \in P_2(\mathbb{R}^d)$, $\gamma := (\iota, 4T_\mu)_{\#}\mu \in \partial^0 TP(\mu)$.*

Proof. Recalling Definition 4.14 and Lemma 4.18, it is sufficient to show that $|\gamma|_{2,2}^2 = |\partial TP|(\mu)$. It is clear by definition of subdifferentiability that $|\gamma|_{2,2}^2 \geqslant$

$|\partial TP|(\mu)$. Now, we make use of the fact that the tightness potential is, in some sense, truly differentiable. Letting $g_t(x) = x + 4tT_\mu x$, and $\alpha_t \in \Gamma(\mu, (g_t)_\# \mu)$,

$$
\begin{aligned}
|\partial TP|(\mu) &= \limsup_{W_2(\mu,\nu)\to 0} \frac{(TP(\mu) - TP(\nu))^+}{W_2(\mu,\nu)} \\
&\geqslant \lim_{t\to 0} \frac{(TP(\mu) - TP((g_t)_\#\mu)^+}{W_2(\mu,(g_t)_\#\mu)} \\
&\geqslant \lim_{t\to 0} \frac{TP(\mu) - TP((g_t)_\#\mu}{C_{2,\alpha_t}(\mu,(g_t)_\#\mu)} \\
&= \lim_{t\to 0} \frac{\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle 4T_\mu x, y - x\ , d\,\rangle \alpha_t(x,y) + o(C_{2,\alpha_t}(\mu,(g_t)_\#\mu))}{C_{2,\alpha_t}(\mu,(g_t)_\#\mu)} \\
&= \lim_{t\to 0} \frac{1}{t} \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle 4T_\mu x\ , y - x\,\rangle d\alpha_t(x,y) \\
&= \int_{\mathbb{R}^d} \langle 4T_\mu x\ , 4T_\mu\,\rangle d\mu(x) \\
&= |\gamma|_{2,2}^2
\end{aligned}
$$

since $C_{2,\alpha}(\mu,(g_t)_\#\mu) = t$, and $\lim_{t\to 0} \frac{o(C_{2,\alpha_t}(\mu,(g_t)_\#\mu))}{C_{2,\alpha_t}(\mu,(g_t)_\#\mu)} = 0$. $\square$

We can also explicitly calculate the derivative of the frame potential along a flow.

Proposition 4.40. *Let $\phi_t : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be the flow of some compactly supported smooth vector field $X : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, i.e. $\frac{d\phi_t(x)}{dt} = X(\phi_t(x))$, $\phi_0(x) = x$, and given a probabilistic frame μ, consider $\nu_t := (\phi_t)_\#\mu$. Then the map*

$$
t \mapsto PFP(\nu_t) = \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi_t(x)\ , \phi_t(y)\,\rangle^2 d\mu(x)d\mu(y), \quad t \in [0,\infty)
$$

is differentiable.

Proof. Therefore,

$$
F(\nu_t) - F(\nu_s) = \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} [\langle \phi_t(x)\ , \phi_t(y)\,\rangle - \langle \phi_s(x)\ , \phi_s(y)\,\rangle].
$$

$$[\langle \phi_t(x)\ ,\phi_t(y)\,\rangle + \langle \phi_s(x)\ ,\phi_s(y)\,\rangle]d\mu(x)d\mu(y)$$

$$= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} [\langle \phi_t(x) - \phi_s(x)\ ,\phi_t(y)\,\rangle + \langle \phi_s(x)\ ,\phi_t(y) - \phi_s(y)\,\rangle]\cdot$$

$$\langle \phi_t(x) + \phi_s(x)\ ,\phi_t(y)\,\rangle - \langle \phi_s(x)\ ,\phi_t(y) - \phi_s(y)\,\rangle]d\mu(x)d\mu(y)$$

Hence,

$$\lim_{s\to t} \frac{F(\nu_t) - F(\nu_s)}{t-s} = \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} \frac{[\langle \phi_t(x) - \phi_s(x)\ ,\phi_t(y)\,\rangle + \langle \phi_s(x)\ ,\phi_t(y) - \phi_s(y)\,\rangle]}{t-s}\cdot$$

$$[\langle \phi_t(x) + \phi_s(x)\ ,\phi_t(y)\,\rangle - \langle \phi_s(x)\ ,\phi_t(y) - \phi_s(y)\,\rangle]d\mu(x)d\mu(y)$$

$$= \iint\limits_{\mathbb{R}^d\times\mathbb{R}^d} [\langle \nabla\phi_t(x)\ ,\phi_t(y)\,\rangle + \langle \phi_t(x)\ ,\nabla\phi_t(y)\,\rangle]\cdot$$

$$2\langle \phi_t(x)\ ,\phi_t(y)\,\rangle d\mu(x)d\mu(y)$$

$$= 4\int \langle X(\phi_t(x))\ ,S_{\nu_t}\phi_t(x)\,\rangle d\mu(x)$$

□

4.3.5 Well-posedness of the Minimization Problem

Since we could not establish the well-posedness of the problem of constructing gradient flows for the tightness potential using the standard machinery of λ-convexity, we will instead follow the approach of [2, Chapter 11.3], using in particular Lemma 4.19 introduced earlier in this chapter. This machinery does not provide a proof of uniqueness, which *a priori* seems natural, since, given a nontight probabilistic frame, there are a multitude of tight probabilistic frames outside a ball of the radius established in Proposition 4.24.

First, we state our main result:

Theorem 4.41. *Gradient flows exist for the tightness potential, i.e. for every initial datum $\mu_0 \in P_2(\mathbb{R}^d)$, each sequence of discrete solutions $\overline{M}_{\tau_k}$ of the variational scheme admits a subsequence such that*

1. $\overline{M}_{\tau_k}(t)$ *narrowly converges in $P(\mathbb{R}^d)$ to μ_t locally uniformly in $[0,\infty)$, with $\mu_t \in AC_2^2([0,\infty); P_2(\mathbb{R}^d))$.*

2. μ_t *is a solution of the gradient flow equation*

$$v_t = -\partial^0 TP(\mu_t), \quad \|v_t\|_{L^2(\mu_t;\mathbb{R}^d)} = |\mu^{'}|(t), \text{ for a.e. } t > 0$$

   with $\mu_t \to \mu_0$ as $t \downarrow 0$, where $v_t(x) = -4T_{\mu_t}(x)$ is the tangent vector to the curve μ_t.

3. *The energy inequality*

$$\int_a^b \int_{\mathbb{R}^d} |v_t(x)|^2 d\mu_t(x)dt + TP(\mu_b) \leqslant TP(\mu_a)$$

   holds for every $b \in [0,\infty)$ and $a \in [0,b)\backslash\mathcal{N}$, where $\mathcal{N}$ is a $\mathcal{L}^1$-negligible subset of $(0,\infty)$.

Proof. This will follow from Proposition 4.42 and Theorem 4.45 by Lemma 4.19, with the identification of the minimal selection with the barycenter $4T_\mu x$ coming from Proposition 4.39. □

To begin, following [2], we define the sublevel sets of a functional $\phi : P_2(\mathbb{R}^d) \to \mathbb{R}$ by

$$\Sigma_m(\phi) := \{\mu \in P_2(\mathbb{R}^d) : \quad \phi(\mu) \leqslant m, \quad M_2^2(\mu) \leqslant m\}.$$

Proposition 4.42. *The sublevels of the tightness potential are compact with respect to the narrow convergence.*

Proof. Suppose that $\{\mu_n\}$ is a sequence in $\Sigma_m(TP)$. Since the sublevels of $f(x) = \|x\|^2$ are compact in $\mathbb{R}^d$ and $\sup_{\nu \in \Sigma_m} \int \|x\|^2 d\nu(x) \leqslant m < \infty$, $\Sigma_m(TP)$ is tight ([2, Remark 5.1.5]), and therefore by Prokhorov's theorem, it is precompact for the narrow convergence. Therefore, there exists a subsequence μ_{n_k} converging weakly to some μ in $P(\mathbb{R}^d)$. It remains to show that $\mu \in \Sigma_m(TP)$.

For $R \in \mathbb{N}$, define $\eta_R : \mathbb{R}^d \to [0,1]$ such that $\eta_R \in C_c^\infty(\mathbb{R}^d)$ with

$$\eta_R(x) = \begin{cases} 1 & \text{if } \|x\| \leqslant R \\ \\ 0 & \text{if } \|x\| \geqslant R+1 \end{cases}.$$

Let $f_R(x) = \eta_R(x)\|x\|^2$. Now, f_R is an acceptable test function for the narrow convergence, so $\forall R \in \mathbb{N}$,

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} f_R(x) d\mu_{n_k}(x) = \int_{\mathbb{R}^d} f_R(x) d\mu(x).$$

Since for all k,

$$\int_{\mathbb{R}^d} f_R(x) d\mu_{n_k}(x) \leqslant m,$$

it follows that for all R,

$$\lim_{R\to\infty} \int_{\mathbb{R}^d} f_R(x) d\mu_{n_k}(x) = \int_{\mathbb{R}^d} f_R(x) d\mu(x) \leqslant m.$$

Then, since $\{f_R(x)\}$ is a nonnegative sequence of measurable functions converging to $f(x) = \|x\|^2$, by Fatou,

$$\int_{\mathbb{R}^d} \|x\|^2 d\mu(x) \leqslant \liminf_R \int_{\mathbb{R}^d} f_R(x) d\mu(x) \leqslant m.$$

Thus $M_2^2(\mu) \leqslant m$.

Second, we define $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$g(x, y) = \langle x\ , y\ \rangle^2 - \frac{1}{d}\|x\ \|^2 \|y\ \|^2.$$

Then, as above, we can define

$$g_R(x, y) = \eta_R(x)\eta_R(y)g(x, y).$$

We note that for all (x, y),

$$|g_R(x, y)| \leqslant |g(x, y)| \leqslant \frac{d+1}{d}\|x\ \|^2 \|y\ \|^2,$$

which by the above is integrable with respect to $\mu \times \mu$ in addition to $\mu_{n_k} \times \mu_{n_k}$ for all $k \in \mathbb{N}$.

Since $\mu_{n_k} \times \mu_{n_k}$ converges weakly to $\mu \times \mu$, and for each $R \in \mathbb{N}$,

$$sup_k \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} g_R(x, y) d\mu_{n_k} \times \mu_{n_k}(x, y) \leqslant m,$$

and,

$$\lim_{k\to\infty} \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} g_R(x, y) d\mu_{n_k} \times \mu_{n_k}(x, y) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} g_R(x, y) d\mu \times \mu(x, y) \leqslant m.$$

(This holds since for $\nu \in \Sigma_m(TP)$, defining ν_R by $\nu_R(A) = \frac{\int_A \eta_R(x) d\mu(x)}{\int_{\mathbb{R}^d} \eta_R(x) d\mu(x)}$ for all Borel sets $A \subset \mathbb{R}^d$, $TP(\nu_R) \geqslant 0$.)

Then, by another application of Fatou (to the sequence $g_R(x, y) + \frac{d+1}{d}\|x\ \|^2 \|y\ \|^2$, initially), since $\lim_{R\to\infty} g_R(x, y) = g(x, y)$ pointwise, we obtain

$$\iint g(x, y) d\mu(x) d\mu(y) \leqslant \liminf_{R\to\infty} \iint g_R(x, y) d\mu(x) d\mu(y) \leqslant m.$$

Thus,

$$TP(\mu) = \iint g(x,y)d\mu(x)d\mu(y) \leqslant m,$$

and it follows that $\mu \in \Sigma_m(TP)$. $\square$

To prove the regularity of the tightness potential, the following standard technical lemmas about projections and uniformly integrable moments will be needed.

Lemma 4.43. Tightness criterion ([2], Lemma 5.2.2) *Let $X, X_1, X_2, \ldots, X_N$ be separable metric spaces, and let $r^i : X \to X_i$ be continuous maps such that the product map*

$$r := r^1 \times r^2 \times ... \times r^N : X \to X_1 \times \ldots X_N$$

is proper. Let $\mathcal{K} \subset P(X)$ be such that $\mathcal{K}_i := r^i_\# \mathcal{K}$ is tight in $P(X_i)$ for $i \in \{1, \cdots, N\}$. Then $\mathcal{K}$ is also tight in $P(X)$.

Lemma 4.44. Uniform Integrability ([2], Lemma 5.2.4) *Let $\mu_n \subset P(\mathbb{R}^d \times \mathbb{R}^d)$ be a sequence narrowly converging to μ in $P(\mathbb{R}^d \times \mathbb{R}^d)$ with $\sup_n M_2(\mu_n) < \infty$. If either $\pi^1_\# \mu_n$ or $\pi^2_\# \mu_n$ has uniformly integrable second moments, then*

$$\lim_{n \to \infty} \iint \langle x_1 \ , x_2 \rangle d\mu_n = \iint \langle x_1 \ , x_2 \rangle d\mu.$$

The preceding two lemma will be needed to prove the following key result:

Theorem 4.45. *The tightness potential is a regular functional.*

Proof. Let ϕ denote the tightness potential. Suppose that $\eta_n \in \partial\phi(\mu)$ is a sequence of strong subdifferentials for a sequence of measures $\mu_n \in P_2(\mathbb{R}^d)$ satisfying:

$$\phi(\mu_n) \to \varphi \in \mathbb{R}, \quad \mu_n \to \mu \quad \text{in } P_2(\mathbb{R}^d),$$

$$\sup_n M_2(\eta_n) < \infty, \quad \eta_n \to \eta \quad \text{in } P(\mathbb{R}^d \times \mathbb{R}^d)$$

First, we show that $\phi(\mu_n) \to \phi(\mu)$. By our differentiability result of Theorem 4.36, for any $\gamma_n \in \Gamma(\mu, \mu_n)$, and in particular for $\gamma_n \in \Gamma_0 \mu, \mu_n$,

$$|\phi(\mu_n) - \phi(\mu)| = \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} \langle 4T_\mu x \ , y - x \rangle d\gamma_n(x, y) \right| + o(W_2(\mu_n, \mu))$$

$$\leqslant 4\|T_\mu \| M_2(\mu) W_2(\mu, \mu_n) + o(W_2(\mu_n, \mu))$$

Thus, as $\mu_n \to \mu$ in $P_2(\mathbb{R}^d)$, $W_2(\mu, \mu_n) \to 0$, and $\phi(\mu_n) \to \phi(\mu)$. Hence, $\varphi = \phi(\mu)$.

Second, we consider the limit of the sequence of strong subdifferentials, $\eta_n \to \eta$. Given any $\mu^0 \in P_2(\mathbb{R}^d)$ and $\nu \in \Gamma(\eta, \mu^0)$, we can choose a sequence $\nu_n \in \Gamma(\eta_n, \mu^0)$. Then we have for all $n \in \mathbb{N}$,

$$\phi(\mu^0) - \phi(\mu_n) \geqslant \iint \langle x_2 \ , x_3 - x_1 \rangle d\nu_n(x_1, x_2, x_3) + o(C_{2,\eta_n}(\mu_n, \mu^0)) \tag{4.22}$$

Then as $n \to \infty$, the left-hand side of equation(4.22) converges to $\phi(\mu^0) - \phi(\mu)$ by our first result.

As for the right-hand side, we write,

$$\iiint \langle x_2 \ , x_3 - x_1 \rangle d\nu_n = \iiint \langle x_2 \ , x_3 \rangle d\pi^{2,3}_{\#} \nu_n - \iiint \langle x_2 \ , x_1 \rangle d\pi^{1,2}_{\#} \nu_n,$$

noting that the same decomposition can be done for the integral with respect to ν, the limit point.

And, applying lemma 4.44 to $\pi^{2,3}_{\#} \nu_n$, whose second marginal, $\mu_0 \in P_2(\mathbb{R}^d)$ clearly has a [uniformly] integrable second moment, and to $\pi^{1,2}_{\#} \nu_n$, whose second marginals, μ_n are converging in $P_2(\mathbb{R}^d)$ and hence have uniformly integrable second moments, we conclude that

$$\begin{aligned}\lim_{n\to\infty}\iiint\langle x_2\ ,x_3-x_1\rangle d\nu_n &= \lim_{n\to\infty}\iint\langle x_2\ ,x_3\rangle d\pi^{2,3}_{\#}\nu_n - \lim_{n\to\infty}\iiint\langle x_2\ ,x_1\rangle d\pi^{1,2}_{\#}\nu_n\\ &= \iint\langle x_2\ ,x_3\rangle d\pi^{2,3}_{\#}\nu - \iiint\langle x_2\ ,x_1\rangle d\pi^{1,2}_{\#}\nu\\ &= \iint\langle x_2\ ,x_1-x_3\rangle d\nu\end{aligned}$$

Finally, by the lower semicontinuity property for narrowly convergenging sequences of probability measures on Hilbert spaces (c.f. [2, Lemma 7.1.4, Equation 5.1.15]),

$$C_{2,\nu}(\mu,\mu^0)\leqslant\liminf_{n\to\infty}C_{2,\nu_n}(\mu^0,\mu_n),$$

and we conclude that

$$\phi(\mu^0)-\phi(\mu)\geqslant\iint\langle x_2\ ,x_3-x_1\rangle d\nu(x_1,x_2,x_3)+o(W_2(\mu,\mu^0)),$$

so that $\eta\in\partial\phi(\mu)$. □

Remark 4.46. Let μ_0 be a probabilistic frame. By Theorem 4.41, there exists a flow ϕ_t such that $\phi_0(x)=x$ and

$$\partial_t\phi_t(x)=v_t(\phi_t(x))=-4T_{\mu_t}\phi_t(x),$$

and $\mu_t=(\phi_t)_{\#}\mu_0$ is a solution to the continuity equation with

$$\int_a^b\int_{\mathbb{R}^d}|v_t(x)|^2d\mu_t(x)dt+TP(\mu_b)\leqslant TP(\mu_a)$$

for every $b\in[0,\infty)$ and $a\in[0,b)\backslash\mathcal{N}$, where $\mathcal{N}$ is a $\mathcal{L}^1$-negligible subset of $(0,\infty)$.

Therefore, as long as the first term in the preceding equation is a.e. nonzero with respect to μ_t, then for any $t\in[a,b]$, the tightness potential is strictly decreasing on that interval. Since T_{μ_t} is nonzero unless μ_t is tight or zero, this will hold until μ_t is tight unless $\phi_t \equiv 0$ on the support of μ_t for some $t \in [a, b]$.

This is related to the question of whether μ_t remains a probabilistic frame for $t \in [0, b]$. Let $\lambda_1^t \geqslant \cdots \geqslant \lambda_d^t \geqslant 0$ denote the eigenvalues of S_{μ_t}. Clearly, $\lambda_d^0 > 0$ if μ_0 is a probabilistic frame, and for μ_t to be a probabilistic frame, we must have $\lambda_d^t > 0$. We denote the "frame gap" by:

$$\epsilon_t := \lambda_1^t - \lambda_d^t,$$

and we note that

$$\lambda_1^t \geqslant \frac{1}{d} M_2^2(\mu_t) \geqslant \lambda_d^t \quad \text{and} \quad \|T_{\mu_t}\| = \max\{\lambda_1^t - \frac{1}{d} M_2^2(\mu_t), \frac{1}{d} M_2^2(\mu_t) - \lambda_d^t\},$$

where strict inequality holds in the first statement unless μ_t is a tight frame or δ_0. From these statements and Proposition 4.32, it follows that

$$\epsilon_t \leqslant 2\|T_{\mu_t}\| \leqslant 2(TP(\mu_t))^{\frac{1}{2}}.$$

Thus, as one intuits, the frame gap is shrinking along the flows as the tightness potential decreases.

4.3.6 The Tightness Potential on the Sphere

As noted in the introduction to this chapter, the gradient flows we consider here, while developed independently, have been considered previously under more limited conditions. In [17], the authors started from a finite, unit-norm frame. They wished to push that frame to a FUNTF in an optimal way, and to do so, they constructed a system of first-order, nonlinear ODEs using the frame potential.

The metric which they used for closeness of one frame to another was the Hilbert-Schmidt norm of the difference of the frames' analysis operators, i.e., if $\Psi = \{\psi_i\}_{i=1}^N$ and $\Phi = \{\varphi_i\}_{i=1}^N$ are two finite frames for $\mathbb{R}^d$ with respective analysis operators Ψ and Φ, then

$$\begin{aligned}\|\Phi - \Psi\|_{HS}^2 &= tr\big[(\Phi - \Psi)^\top(\Phi - \Psi)\big]\\ &= \sum_{i=1}^N \|\varphi_i - \psi_i\|^2\end{aligned}$$

This can easily be superseded by the Wasserstein distance between the canonical probabilistic frames associated with the two frames in question, which corresponds to a stronger topology on the same set:

$$\begin{aligned}W_2^2(\mu_\Psi, \mu_\Phi) &= \frac{1}{N}\min_{\sigma\in S_N}\sum_{i=1}^N \|\varphi_i - \psi_{\sigma(i)}\|^2\\ &\leqslant \frac{1}{N}\sum_{i=1}^N \|\varphi_i - \psi_i\|^2\\ &= \frac{1}{N}\|\Phi - \Psi\|_{HS}^2\end{aligned}$$

The main results of [17] constitute a special case of Theorem 4.41 giving a flow on a finite unit-norm frame as a series of ODEs. Using the notation of [17], we define $\mathbb{H}_d$ to be a d-dimensional real or complex Hilbert space and $\mathbb{H}_d^N$ to be the sets of N vectors in that space. Let $\mathbb{S}_d$ be the unit sphere in $\mathbb{H}_d$, and let $\mathbb{S}_d^N$ be the N-fold product of that sphere. For simplicity, to denote the analysis and synthesis operators, we shall use our notation F and F^*, as we will use S_F for the frame operator of the frame F. The statement of the main result then comes in two parts:

Lemma 4.47. *[17, Proposition 1] For any $F = \{f_i\}_{i=1}^N \in \mathbb{S}_d^N$ and*

$$G = \{g_n\}_{n=1}^N \in \oplus_{n=1}^N f_n^\perp := \left\{\{g_n\}_{n=1}^N \in \mathbb{H}_d^N : \langle f_n\ , g_n\rangle = 0, \forall n\right\},$$

let

$$f_n(t) := cos(\|g_n\|t)f_n - sin(\|g_n\|t)\frac{g_n}{\|g_n\|}$$

whenever $g_n \neq 0$, *and let* $f_n(t) := f_n$ *otherwise. Then* $F(t) = \{f_n(t)\}_{i=1}^N \in \mathbb{S}_d^N$ *for any* $t \in \mathbb{R}$, *and the frame* F *and* $F(t)$ *with analysis operators denoted by* F *and* $F(t)$ *satisfy*

$$\|F(t) - F\|_{HS}^2 \leqslant t^2 \sum_{n=1}^N \|g_n\|^2$$

and

$$FP(F(t)) \leqslant FP(F) - 4tRe\sum_{n=1}^N \langle S_F f_n \ , g_n \rangle + 8Nt^2 \sum_{n=1}^N \|g_n\|^2 \tag{4.23}$$

Moreover,

Lemma 4.48. *[17, Theorem 2] Pick* $F = \{f_i\}_{i=1}^N \in \mathbb{S}_d^N$, *and let* P_n *denote the orthogonal projection from* $\mathbb{H}_d$ *onto the orthogonal complement of* f_n. *Then, the minimizer of the bound in* (4.23) *over all* $t \in \mathbb{R}$ *and* $\{g_n\}_{n=1}^N \in \bigoplus_{n=1}^N f_n^\perp$ *is given by* $t = \frac{1}{4N}$ *and*

$$g_n = P_n S_F f_n = S_F f_n - \langle S_F f_n \ , f_n \rangle f_n, \quad n \in \{1, \cdots, N\} \tag{4.24}$$

Moreover, for any $t \in \mathbb{R}$, *this choice for* $\{g_n\}_{n=1}^N$ *gives:*

$$\|F(t) - F\|_{HS}^2 \leqslant t^2 \sum_{n=1}^N \|P_n S_F f_n\|^2 \tag{4.25}$$

and

$$FP(F(t)) \leqslant FP(F) - 4t(1 - 2Nt)\sum_{i=1}^N \|P_n S_F f_n\|^2 \tag{4.26}$$

The authors points out that "as $t \to 0$, we expect to approach a solution to the system of nonlinear ordinary differential equations:

$$f_n'(s) = -\left(S_F(s)f_n(s) - \langle S_F(s)f_n(s) \ , f_n(s) \rangle f_n(s)\right), \quad \forall n \in \{1, \cdots, N\}$$

a matter we leave for future research." Indeed, their flow is analogous to the flow of our tightness potential. The difference lies in the fact that they constrain their frame to live on the unit sphere, whereas we allow the support of our probabilistic frame to vary. Because of this, the second moment of the canonical probabilistic frame corresponding to their frames is fixed to 1.

Observe that we can rewrite (4.24) in our language. Beginning with a probabilistic frame μ supported on the unit sphere, we can restrict our flow by reprojecting the flow of the gradient onto the sphere. Letting P_x denote the projection onto the tangent plane to the unit sphere at $x \in S^{d-1}$, we can define $\phi_t(x)$ to be the flow of the vector field $X_t(x) = -4S_{\mu_t}x$, with $\mu_t(x) = (\phi_t)_{\#}\mu$.

$P_x(-4S_{\mu_t}(x)) = (I - xx^\top)(-4S_{\mu_t}(x))$. Then $(\phi_t)_{\#}\mu$ is a flow of probabilistic frames supported on S^{d-1}, analogous to (4.24), and by Proposition 4.40,

$$\begin{aligned}
\frac{d}{dt}PFP(\mu_t) &= 4\int_{S^{d-1}} \langle X_t(\phi_t(x)) \ , S_{\mu_t}\phi_t(x) \rangle d\mu(x) \\
&= -16\int_{S^{d-1}} \langle (I - \phi_t(x)\phi_t(x)^\top) S_{\mu_t}\phi_t(x) \ , S_{\mu_t}\phi_t(x) \rangle d\mu(x) \\
&= -16\int_{S^{d-1}} \|S_{\mu_t}\phi_t(x)\|^2 - \langle \phi_t(x) \ , S_{\mu_t}\phi_t(x) \rangle^2 d\mu(x) \\
&\leqslant 0
\end{aligned}$$

with equality if and only if S_{μ_t} is a multiple of the identity, i.e. if and only if μ_t is tight because $\operatorname{supp}(\phi_t)_{\#}\mu) \subset S^{d-1}$.

4.3.7 The Fourth and Higher Potentials

In addition to the frame potential, we are interested in higher-order potentials which are only defined on $P_p(\mathbb{R}^d)$, for example:

Definition 4.49. For $\mu \in P_p(\mathbb{R}^d)$ and $p \in (0, \infty]$, we can define the p**-frame potential**, $PFP_p(\mu)$ by

$$PFP_p(\mu) = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} |\langle x \ , y \rangle|^p d\mu(x) d\mu(y).$$

It is a key result of [31] that the minimizers of this potential among probabilistic frames supported on the S^{d-1} are precisely the probabilistic tight p-frames, which we define next.

Definition 4.50. Given $p \in (0, \infty)$, a probability measure on $\mathbb{R}^d$ is a probabilistic p-frame for $\mathbb{R}^d$ if there exist $0 < A \leqslant B < 0$ such that for all $y \in \mathbb{R}^d$,

$$A \|y\|^p \leqslant \int_{\mathbb{R}^d} |\langle x \ , y \rangle|^p d\mu(x) \leqslant B \|y\|^p,$$

and μ is a tight probabilistic p-frame if $A = B$.

In this case, the Otto calculus which we have used above can be extended to $P_p(\mathbb{R}^d)$, the Wasserstein space of order p. There is a similar notion of subdifferential in this space, although the construction of gradient flows is a bit more involved.

One must first define the mixed space

$$P_{pq}(\mathbb{R}^d \times \mathbb{R}^d) := \left\{ \gamma \in P(\mathbb{R}^d \times \mathbb{R}^d) : |\gamma|_{1,p} + |\gamma|_{2,q} < \infty \right\}$$

with

$$|\gamma|_{j,p}^p = \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} |x_j|^p d\gamma(x_1, x_2), \quad j = 1, 2, \quad p > 1.$$

Then, for $p > 1$,

Definition 4.51. *[2, The strong subdifferential, Definition 10.3.1]* Let $\phi : P_p(\mathbb{R}^d) \to (-\infty, \infty]$ be a proper and lower semi-continuous functional, and let $\mu^1 \in D(\phi)$. Let $q = \frac{p}{p-1}$. Then $\gamma \in P_{pq}(\mathbb{R}^d \times \mathbb{R}^d)$ belongs to the **extended Fréchet subdifferential** $\partial\phi(\mu^1)$ if $\pi^1_\# \gamma = \mu^1$ and

$$\phi(\mu^3) - \phi(\mu^1) \geqslant \inf_{\nu \in \Gamma_0(\gamma,\mu^3)} \iiint \langle x_2 \,, x_3 - x_1 \rangle d\nu + o(W_p(\mu^1, \mu^3)).$$

We say that $\gamma \in \partial\phi(\mu^1)$ is a **strong Fréchet subdifferential** if for every $\nu \in \Gamma(\gamma, \mu^3)$, it satisfies

$$\phi(\mu^3) - \phi(\mu^1) \geqslant \iiint \langle x_2 \,, x_3 - x_1 \rangle d\nu + o(C_{p,\nu}(\mu^1, \mu^3)), \tag{4.27}$$

where $C_{p,\nu}(\mu^1, \mu^3)$ is the pseudo-distance given by the cost

$$C^p_{p,\nu}(\mu^1, \mu^3) = \iiint \| x_1 - x_3 \|^p d\nu(x_1, x_2, x_3).$$

Now we can show:

Proposition 4.52. *The p-frame potential is a differentiable function in $P_p(\mathbb{R}^d)$ for $p > 2$.*

Proof. Given μ as above, let $\nu \in P_p(\mathbb{R}^d)$. Define $g^\mu_p(x) := \int \langle x \,, z \rangle^{p-1} z d\mu(z)$. Then, letting $\gamma \in \Gamma_p(\mu, \nu)$,

$$\begin{aligned}
&PFP_p(\nu) - PFP_p(\mu) \\
&\quad = \iiiint \langle w \,, y \rangle^p - \langle x \,, z \rangle^p d\gamma(x,y) d\gamma(z,w) \\
&\quad = \iiiint (\langle w - z \,, y - x \rangle + \langle x \,, w - z \rangle + \langle z \,, y - x \rangle + \langle z \,, x \rangle)^p
\end{aligned}$$

$$- \langle x\ ,z \rangle^p d\gamma(x,y)d\gamma(z,w)$$

$$= \iiiint \underbrace{p\langle z\ ,y-x \rangle\langle z\ ,x \rangle^{p-1} + p\langle x\ ,w-z \rangle\langle x\ ,z \rangle^{p-1}}_{A} \quad +$$

$$\sum_{\substack{\{(i,j,k,l)\in\mathbb{N}^4:\quad i+j+k+l=p\}\backslash\\ \{(0,0,0,p),(0,1,0,p-1),(0,0,1,p-1)\}}} \binom{p}{i,j,k,l} \underbrace{\langle w-z\ ,y-x \rangle^i\langle x\ ,w-z \rangle^j\langle z\ ,y-x \rangle^k\langle z\ ,x \rangle^l}_{B_{i,j,k,l}} d\gamma(x,y)d\gamma(z,w)$$

$$= A + \sum \binom{p}{i,j,k,l} B_{i,j,k,l}$$

Then

$$\begin{aligned} A &= \iiiint p\langle z\ ,y-x \rangle\langle z\ ,x \rangle^{p-1} + p\langle x\ ,w-z \rangle\langle x\ ,z \rangle^{p-1} d\gamma(x,y)d\gamma(z,w) \\ &= \iiiint 2p\langle z\ ,y-x \rangle\langle z\ ,x \rangle^{p-1} d\gamma(x,y)d\gamma(z,w) \\ &= \iint 2p\langle g_p^{\mu}(x)\ ,y-x \rangle d\gamma(x,y) \end{aligned}$$

and, for $\{(i,j,k,l) \in \mathbb{N}^4 : i+j+k+l = p\}\backslash\{(0,0,0,p),(0,1,0,p-1),(0,0,1,p-1)\}$,

$$\begin{aligned} |B_{i,j,k,l}| &= \left| \iiiint \langle w-z\ ,y-x \rangle^i\langle x\ ,w-z \rangle^j\langle z\ ,y-x \rangle^k\langle z\ ,x \rangle^l d\gamma(x,y)d\gamma(z,w) \right| \\ &\leqslant \iiiint \|w-z\|^{i+j}\|y-x\|^{i+k}\|x\|^{j+l}\|z\|^{k+l} d\gamma(x,y)d\gamma(z,w) \\ &= \iint \|w-z\|^{i+j}\|z\|^{k+l} d\gamma(z,w) \cdot \iint \|y-x\|^{i+k}\|x\|^{j+l} d\gamma(x,y) \\ &\leqslant \left(\iint \|w-z\|^p d\gamma(z,w)\right)^{\frac{i+j}{p}} \left(\iint \|z\|^p d\gamma(z,w)\right)^{\frac{k+l}{p}} \cdot \\ &\left(\iint \|y-x\|^p d\gamma(x,y)\right)^{\frac{i+k}{p}} \left(\iint \|x\|^p d\gamma(x,y)\right)^{\frac{j+l}{p}} \\ &= W_p(\mu,\nu)^{2i+j+k} \cdot M_p(\mu)^{j+k+2l} \end{aligned}$$

by generalized Hölder. Therefore

$$PFP_p(\nu) - PFP_p(\mu) = \iint 2p\langle g_p^{\mu}(x)\ ,y-x \rangle d\gamma(x,y)$$

$$+\sum_{\substack{\{(i,j,k,l)\in\mathbb{N}^4:\quad i+j+k+l=p\}\setminus\\ \{(0,0,0,p),(0,1,0,p-1),(0,0,1,p-1)\}}}\binom{p}{i,j,k,l}W_p(\mu,\nu)^{2i+j+k}\cdot M_p(\mu)^{j+k+2l}$$

$$=\iint 2p\langle g_p^{\mu}(x)\ ,y-x\rangle d\gamma(x,y)+o(W_p(\mu,\nu)^{2i+j+k})$$

since $2i+j+k\geqslant 3$ for the set of admissible indices, so that PFP_p is differentiable in $P_p(\mathbb{R}^d)$. $\square$

Proposition 4.53. *The derivative of the p-th frame potential is continuous in $P_p(\mathbb{R}^d)$.*

Proof. Taking μ and ν, probabilistic frames with finite p-th moments, again, we let g_μ and g_ν denote the respective derivatives of the p-th frame potential at the each measure.

Defining

$$h_s(a,b,c):=\langle a\ ,c\rangle^{s-1}+\langle a\ ,c\rangle^{s-2}\langle b\ ,c\rangle+\cdots+\langle a\ ,c\rangle\langle b\ ,c\rangle^{s-2}+\langle b\ ,c\rangle^{s-1}.$$

Then, since for $i\in\{1,\cdots,p-1\}$,

$$\frac{1}{p}+\frac{p-i}{p}+\frac{i-1}{p}=1,$$

we have by generalized Hölder,

$$\begin{aligned}|g_p^{\mu}(y)-g_p^{\nu}(y)|&=\left|\iint\langle z\ ,y\rangle^{p-1}(z-w)+(\langle z\ ,y\rangle^{p-1}-\langle w\ ,y\rangle^{p-1})wd\gamma(z,w)\right|\\ &\leqslant\|y\|^{p-1}\iint\|z\|^{p-1}\|z-w\|d\gamma(z,w)+\iint\|w\|\|z-w\||h_{p-1}(z,w,y)|d\gamma(z,w)\\ &\leqslant\|y\|^{p-1}\Bigg[\left(\iint\|z\|^{p}d\mu(z)\right)^{\frac{p-1}{p}}\left(\int\|z-w\|^{p}d\gamma(z,w)\right)^{\frac{1}{p}}\\ &+\iint\|w\|\|z-w\|(\|w\|^{p-2}+\|w\|^{p-3}\|z\|+\cdots+\|w\|\|z\|^{p-3}+\|z\|^{p-2})d\gamma(z,w)\Bigg]\end{aligned}$$

$$= \| y \|^{p-1} \left[M_p^{p-1}(\mu) W_p(\mu,\nu) + \sum_{i=1}^{p-1} \iint \| w \|^{p-i} \| z \|^{i-1} d\gamma(z,w) \right]$$

$$\leqslant \| y \|^{p-1} \Big[M_p^{p-1}(\mu) W_p(\mu,\nu)$$

$$+ \left(\iint \| z - w \|^p d\gamma(z,w) \right)^{\frac{1}{p}} \sum_{i=1}^{p-1} \left(\int \| w \|^p d\gamma(z,w) \right)^{\frac{p-i}{p}} \left(\int \| z \|^p d\gamma(z,w) \right)^{\frac{i-1}{p}} \Big]$$

$$= \| y \|^{p-1} W_p(\mu,\nu) \left[M_p^{p-1}(\mu) + \sum_{i=1}^{p-1} M_p^{p-i}(\nu) M_p^{i-1}(\mu) \right]$$

Noting that by Minkowski, $M_p^k(\nu) \leqslant 2^k (M_p(\mu)^k + W_p^k(\mu,\nu))$, we have control over $|g_p^\mu(y) - g_p^\nu(y)|$ in terms of the p-th Wasserstein distance. □

And, as with the case $p = 2$, we have a lower bound, which generalizes the lower bound given in [31]:

Theorem 4.54. *Let μ be a probabilistic p-frame for $\mathbb{R}^d$ for $p \geqslant 2$ an even number. Then*

$$PFP_p(\mu) \geqslant \frac{(p-1)(p-3)\cdots 1}{(d+p-2)(d+p-4)\cdots d} \left(\int_{\mathbb{R}^d} \| x \|^p d\mu(x) \right)^2$$

with equality if and only if μ is tight.

Proof. Let μ be a measure in $P_2(\mathbb{R}^d)$. Let $p = 2k, k \in \mathbb{N}$. Let

$$p(y) = \int_{\mathbb{R}^d} \langle x \ , y \rangle^p d\mu(x).$$

Then $p(y)$ is a homogeneous polynomial of degree p in (the components of y.

Following [50], we have the following formal constructions for homogeneous polynomials:

1. We can construct write any homogeneous polynomial $f(x)$ in x as

$$f(x) = \sum_{|i|=p} c(i) a(i) x(i),$$

where $i = (n_1, \cdots, n_d)$ is a d-element multiindex,$c(i) = \binom{p}{n_1,\cdots,n_d}$, $a(i)$ is the coefficient corresponding to that multiindex, and $x(i)$ is the monomial corresponding to that multiindex, $x(i) = x_1^{n_1} \cdots x_d^{n_d}$.

2. We let $\rho_\alpha^m = (a_1 x_1 + \cdots + a_d x_d)^m$.

3. We define an inner product on these homogeneous polynomials: given

$$f(x) = \sum_{|i|=p} c(i)a(i)x(i),$$

$$g(x) = \sum_{|i|=p} c(i)b(i)x(i),$$

we define

$$[f, g] = \sum_{|i|=p} c(i)a(i)b(i).$$

The fact that this is an inner product on this space is validated in [50]. We can use this construction by noting that for any constant A,

$$[p(y) - A\|y\|^p, p(y) - A\|y\|^p] \geqslant 0, \tag{4.28}$$

with equality if and only if μ is a tight probabilistic p-frame.

First, suppose that p is a tight probabilistic p-frame. Then it is clear that equality holds in (4.28), and we can determine A using the following computations from [50]:

1. $\Delta_y(\langle x, y\rangle^p = p(p-1)\langle x, x\rangle\langle x, y\rangle^{p-2}$

2. $\Delta\langle y, y\rangle^k = 2k(2k + d - 2)\langle y, y\rangle^{k-1}$

3. $\Delta_x[\langle x, x\rangle^l\langle y, x\rangle^m] = 2l(2l+2m+d-2)\langle x, x\rangle^{l-1}\langle x, y\rangle^m + m(m-1)\langle y, y\rangle\langle x, x\rangle^l\langle x, y\rangle^{m-2}$

Supposing that

$$p(x) = \int_{\mathbb{R}^d} \langle x \ , y \rangle^p = c_p(\mu) \| y \|^2,$$

we can apply the above operators to each side of the equation recursively to identify the constant. In this way, we obtain

$$c_p(\mu) = \frac{(p-1)(p-3)\cdots 1}{(d+p-2)(d+p-4)\cdots d} [\int_{\mathbb{R}^d} \| x \|^p d\mu(x)]^2.$$

Suppose, conversely, that equality holds in (4.28). We will use the following relations from [50] related to homogeneous polynomials F, G of degree $p = 2k$:

1. $[\rho_z^{2k}, F] = F(z)$

2. $[F, G] = \frac{1}{(2l)!} F(\nabla) G$, where by ∇ we mean $(\partial_{x_1}, \ldots, \partial_{x_d})$.

With these in hand, we note that

$$p(y) = \int_{\mathbb{R}^d} (x_1 y_1 + \ldots x_d y_d)^p d\mu(x) = \rho_y^{\beta},$$

where $\beta_{n_1,\ldots,n_d} = \int_{\mathbb{R}^d} x_1^{n_1} \cdots x_d^{n_d} d\mu(x)$, and that $\| \nabla \|^2 = \Delta$.

Letting $c_p = \frac{(p-1)(p-3)\cdots 1}{(d+p-2)(d+p-4)\cdots d}$, we can then rewrite:

$$\begin{aligned}
\int_{\mathbb{R}^d} [p(y) - A\| y \|^p, p(y) - A\| y \|^p] d\mu(y) &= \int [p(y), p(y)] + A^2 [\| y \|^2, \| y \|^2] - 2A[p(y), \| y \|^2] d\mu(y) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \langle z \ , y \rangle^p d\mu(z), p(y) \right] d\mu(y) \\
&\quad - 2A \iint_{\mathbb{R}^d \times \mathbb{R}^d} \left[\langle x \ , y \rangle^p, \| y \|^2 \right] d\mu(x) d\mu(y) \\
&\quad + A^2 \int_{\mathbb{R}^d} [\| y \|^p, \| y \|^p] d\mu(y) \\
&= \int_{\mathbb{R}^d} p(y) d\mu(y) - 2A \int_{\mathbb{R}^d} \| y \|^p d\mu(y) + \int_{\mathbb{R}^d} \Delta^{\frac{p}{2}} \| y \|^2 d\mu(y)
\end{aligned}$$

$$= \int_{\mathbb{R}^d} p(y)d\mu(y) - 2A\int_{\mathbb{R}^d} \|y\|^p d\mu(y) + \frac{A^2}{c_p}$$

Since for all $A \in \mathbb{R}^d$, $[p(y) - A\|y\|^p, p(y) - A\|y\|^p] \geqslant 0$, we can use the discriminant of this quadratic to show that

$$c_p \left[\int_{\mathbb{R}^d} \|y\|^p d\mu(y)\right]^2 \leqslant \int_{\mathbb{R}^d} p(y)d\mu(y),$$

and, in particular, if we choose

$$A = c_p \left[\int_{\mathbb{R}^d} \|y\|^p d\mu(y)\right]^2,$$

then equality holds in (4.28). □

Remark 4.55. Future work would include constructing Wasserstein gradient flows in $P_p(\mathbb{R}^d)$ for this potential to obtain tight p-frames, which are linked to equiangular tight frames.

4.3.8 Other Potentials

Given a path of probabilistic frames in $P_2(\mathbb{R}^d)$, it might be useful to consider how frame/dual-frame pairs coevolve. Thus, we consider the following construction:

Definition 4.56. Given a probabilistic frame $\mu \in P_2(\mathbb{R}^d$ and $\gamma \in P_2(\mathbb{R}^d \times \mathbb{R}^d)$ with $\pi^1_\# \gamma = \mu$, we define the duality potential

$$G(\gamma) := \sum_{i=1}^{d}\sum_{j=1}^{d}\left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} x_i y_j - \delta_{ij} d\gamma(x,y)\right)^2$$

The motivation for the name comes from the fact that if μ is a probabilistic frame and $\gamma \in \Gamma D_\mu$, then $\iint_{\mathbb{R}^d \times \mathbb{R}^d} xy^\top d\gamma(x,y) = I$, so that for all $\gamma \in \Gamma D_\mu$, we see that $G(\gamma) = 0$. Similarly, as an alternate approach to the Paulsen problem, we can consider a support potential for a given probabilistic frame μ:

$$H(\mu) := \iint\limits_{\mathbb{R}^d \times \mathbb{R}^d} (\|x\| - \|y\|)^2 d\mu(x) d\mu(y).$$

Clearly, $H(\mu) \geqslant 0$, and $H(\mu) = 0$ if and only if $\operatorname{supp}(\mu) \subset kS^{d-1}$, where kS^{d-1} is the sphere of radius $k > 0$ centered around the origin. A better potential for the Paulsen problem could then be

$$PP(\mu) = PFP_2(\mu) + H(\mu).$$

4.4 Scaling Result

Finally, to conclude our investigation of tight frames, we end with a result about scalable frames, which we approach from the probabilistic frame perspective. We seek to scale discrete probabilistic frames by changing their weights in order to obtain tight probabilistic frames. This is a different perspective on the scalable frames problem dictated by the constraints of the probabilistic point of view; the usual approach would be equivalent to scaling the magnitudes of the vectors in the support of a probabilistic frame.

Let $\mu_0 = \sum\limits_{i=1}^{N} \delta_{\varphi_i}$ and $\mu_A = \sum\limits_{i=1}^{N} a_i \delta_{\varphi_i}$ with $\sum\limits_{i=1}^{N} a_i = 1$, $a_i \geqslant 0$ and $\|\varphi_i\| = 1$ $\forall i$. In this case,

$$PFP(\mu_A) = \sum_{i,j=1}^{N} a_i a_j \langle \varphi_i , \varphi_j \rangle^2.$$

We know that $PFP(\mu_A) \geqslant \frac{M_2^4(\mu_A}{d} = \frac{1}{d}$, with equality if and only if μ_A is tight.

Letting $Q := [[\langle \varphi_i \ , \varphi_j \rangle^2]]_{i,j}$, we see that

$$PFP(\mu_A) = a^\top Q a, \quad \text{where } a = \begin{bmatrix} a_1 & \cdots & a_N \end{bmatrix}.$$

We note that:

Lemma 4.57. *Q is positive semidefinite.*

Proof. Q is the Hadamard product of the Grammian matrix with itself, i.e., $G := [[\langle \varphi_i \ , \varphi_j \rangle^2]]_{i,j}$, and $Q = G \odot G$. Since the Grammian is positive semi-definite, Q will also be positive semi-definite. □

Because Q is symmetric positive semidefinite, letting $m = rank(Q)$, we can write

$$Q = \sum_{i=1}^{m} \lambda_i v_i v_i^\top,$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_m > 0$ are the nonzero eigenvalues, and $\{v_i\}_{i=1}^N \subset \mathbb{R}^N$ are the orthonormal eigenvectors. We can therefore express any vector $a \in \mathbb{R}^N$ as $a = \sum_{i=1}^{N} c_i v_i$ for some constants c_i. We are trying to obtain

$$a^\top Q a = \sum_{i=1}^{m} c_i^2 \lambda_i = \frac{1}{d}.$$

Since the diagonals of Q are the fourth powers of the norms of the $\{\varphi_i\}_{i=1}^N$, we know that for each $k \in \{1, \cdots, N\}$,

$$Q_{k,k} = \sum_{i=1}^{m} \lambda_i (v_i^k)^2 = 1.$$

Thus, letting $s_i = \sum_{k=1}^{N} v_i^k$, the constraints of our problem reduce to solving

$$\begin{cases} \sum_{i=1}^{N} \lambda_i c_i^2 = \frac{1}{d} & \text{constraint (1) on quadratic form} \\ \sum_{i=1}^{N} c_i v_i^k \geqslant 0 & \text{nonnegativity constraint (2) on } a \\ \sum_{i=1}^{N} c_i s_i = 1 & \text{constraint (3) on sum of entries of } a \end{cases}$$

Eliminating the free variables in constraint (1), and rewriting constraint (3) with slack variables, we obtain a revised version in $\mathbb{R}^d$:

$$\begin{cases} \sum_{i=1}^{m} \lambda_i c_i^2 = \frac{1}{d} & \text{constraint (1*) on quadratic form} \\ \sum_{i=1}^{N} c_i v_i^k \geqslant 0 & \text{nonnegativity constraint (2) on } a \\ \sum_{i=1}^{m} c_i s_i = 1 - \sum_{i=m+1}^{N} c_i s_i & \text{constraint (3*) on sum of entries of } a \end{cases}$$

Constraints (1*) and (3*) make this a problem of finding the intersection of a hyperplane H and an ellipsoid E in $\mathbb{R}^m$, where the variable is the vector $c = \begin{bmatrix} c_1 \ldots c_N \end{bmatrix}$. In particular, any intersection point y should lie between two parallel hyperplanes tangent to the ellipsoid. In particular, the coordinates of y should be bounded in magnitude by the magnitudes of the coordinates of the intersection points of the hyperplanes with the coordinate axes. That is, if z is the intersection of a tangent plane with the first coordinate axis, then $|y_1| < |z_1|$. Given a point u on E, the equation of its tangent plane is

$$2 \begin{bmatrix} \lambda_1 u_1 \\ \ldots \\ \lambda_d u_d \end{bmatrix} \cdot (x - u) = 0 \tag{4.29}$$

$$\sum_{i=1}^{m} \lambda_i u_i x_i = \sum_{i=1}^{m} \lambda_i u_i^2 \tag{4.30}$$

$$\sum_{i=1}^{m} \lambda_i u_i x_i = \frac{1}{d} \tag{4.31}$$

Thus, the i-th intercept of the tangent plane, obtained by setting $x_j = 0$ for all $j \neq i$, is $x_i = \frac{1}{d\lambda_i u_i}$. Conversely, if we have the coordinates of the intercepts of a tangent plane, we can obtain the point of tangency via $u_i = \frac{1}{d\lambda_i x_i}$.

The equation of the hyperplane H_1 whose points satisfy constraint (3*) can be written as $s \cdot (x - t)$, with $t = \sum_{i>m} s_i c_i - 1$, and if it is parallel to a tangent plane H_2 to E, then there is some $k \neq 0$ and some $u \in E$ such that $s = k \begin{bmatrix} \lambda_1 u_1 \\ \cdots \\ \lambda_m u_m \end{bmatrix}$. The point of tangency of H_2 is u, with $u_i = \frac{s_i}{k\lambda_i}$, satisfying:

$$\sum_{i=1}^{m} \lambda_i (\frac{s_i}{k\lambda_i})^2 = \frac{1}{d} \tag{4.32}$$

and the intercepts of the H_2 are

$$x_i = \frac{1}{d\lambda_i u_i} = \frac{k}{ds_i}. \tag{4.33}$$

The i-th intercept of H_1 is $x_i = -\frac{t}{s_i}$, so that from equation 4.33 and equation 4.32, we see that we must require for each $i \in \{1, \cdots, m\}$:

$$\begin{aligned} |-\frac{t}{s_i}| &< |\frac{k}{ds_i}| \\ |1 - \sum_{i>m} s_i c_i| &< \frac{|k|}{d} \\ |1 - \sum_{i>m} s_i c_i| &< \sqrt{\frac{1}{d} \sum_{i=1}^{m} \frac{s_i^2}{\lambda_i}} \end{aligned} \tag{4.34}$$

We have thus proven:

Lemma 4.58. *If $a = \sum_{i=1}^{N} c_i v_i$ with $\{c_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$ satisfying* (4.34)*, then a satisfies constraints (1*) and (3*).*

We use this result to prove the following.

Proposition 4.59. *Given* $\{\varphi_i\}_{i=1}^N \subset S^{d-1}$*, let* $Q \in \mathbb{R}^{N\times N}$ *be the matrix defined by* $Q_{i,j} = \langle \varphi_i \ , \varphi_j \rangle^2$*. Let* $z = \begin{bmatrix} \frac{1}{N} \\ \cdots \\ \frac{1}{N} \end{bmatrix} \in \mathbb{R}^N$*. Then if* $z^\top Q^\dagger z \geqslant \frac{d}{N^3}$*, there exists* $\{a_i\}_{i=1}^N$ *with* $a_i \geqslant 0$*,* $\sum_{i=1}^N a_i = 1$ *such that* $\mu := \sum_{i=1}^N a_i \delta_{\varphi_i}$ *is a tight probabilistic frame.*

Proof. Again, by Lemma 4.57, Q is symmetric, positive semi-definite. Letting $rank(Q) = m > 0$, we have m positive eigenvalues $\{\lambda_i\}_{i=1}^m$ and an orthonormal basis of eigenvectors $\{v_i\}_{i=1}^N$ and can decompose Q as $Q = \sum_{i=1}^m \lambda_i v_i v_i^\top$. Given $r \in [\frac{1}{N^2}, \frac{1}{N}]$, there exists a probability vector $a \in \mathbb{R}^N$ (i.e., $a_i \geqslant 0, \quad \in \{1, \cdots, N\}$, $\sum_{i=1}^N a_i = 1$) as above such that $\|a\|^2 = R$.

Now suppose $\frac{d}{N^3} \leqslant z^\top Q^\dagger z = \sum_{i=1}^m \frac{1}{\lambda_i}\langle v_i \ , z \rangle^2$, so that

$$\frac{N}{d}\sum_{i=1}^m \frac{1}{\lambda_i}\langle v_i \ , z \rangle^2 \geqslant \frac{1}{N^2}.$$

Then there exists some probability vector a such that

$$\|a\|^2 \leqslant \frac{N}{d}\sum_{i=1}^m \frac{1}{\lambda_i}\langle v_i \ , z \rangle^2$$

First, $\|a\|^2 = \sum_{i=1}^N \langle v_i \ , a \rangle^2 \geqslant \sum_{i=1}^m \langle v_i \ , a \rangle^2$ implies that

$$\sum_{i=1}^m \langle v_i \ , a \rangle^2 \leqslant \frac{N}{d}\sum_{i=1}^m \frac{1}{\lambda_i}\langle v_i \ , z \rangle^2 \tag{4.35}$$

Second, $\|z\|^2 = \frac{1}{N}$, so that

$$\sum_{i=1}^m \langle v_i \ , z \rangle^2 \leqslant \frac{1}{N} \tag{4.36}$$

Thus, by (4.35) and (4.36) and the CBS inequality, we have:

$$N\sum_{i=1}^{m}\langle v_i\ , z\,\rangle^2\sum_{i=1}^{m}\langle v_i\ , a\,\rangle^2 \leqslant \frac{N}{d}\sum_{i=1}^{m}\frac{1}{\lambda_i}\langle v_i\ , z\,\rangle^2$$
$$(N\sum_{i=1}^{m}\langle v_i\ , z\,\rangle\langle v_i\ , a\,\rangle)^2 \leqslant \frac{N^2}{d}\sum_{i=1}^{m}\frac{1}{\lambda_i}\langle v_i\ , z\,\rangle^2$$

Then, noting that $1-N\langle z\ , a\,\rangle = 0$, and recalling that $N\langle z\ , a\,\rangle = \sum_{i=1}^{N}\langle z\ , v_i\,\rangle\langle v_i\ , a\,\rangle$, we obtain

$$\begin{aligned}(1-N\sum_{i>m}\langle v_i\ , z\,\rangle\langle v_i\ , a\,\rangle)^2 &= (1-N\langle z\ , a\,\rangle + N\sum_{i=1}^{m}\langle v_i\ , z\,\rangle\langle v_i\ , a\,\rangle)^2 \\ &\leqslant \frac{N^2}{d}\sum_{i=1}^{m}\frac{1}{\lambda_i}\langle v_i\ , z\,\rangle^2\end{aligned}$$

But a quick calculation shows that this last inequality is equivalent to

$$(1-\sum_{i>m}s_i\langle v_i\ , a\,\rangle)^2 \leqslant \frac{1}{d}\sum_{i=1}^{m}\frac{s_i^2}{\lambda_i},$$

where, as in the lemma 4.58, $s_k = \sum_{i=1}^{N} v_i^k$. Thus, by that lemma, we have that a, in addition to satisfying constraint (2), satisfies also constraints (1*) and (3*). □

Corollary 4.60. *Let $\lambda = \lambda_{max}(Q)$. If $\lambda^2 d \leqslant \sum_{i,j=1}^{N} Q_{i,j}$, then there exists a such that $\mu := \sum_{i=1}^{N} a_i\delta_{\varphi_i}$ is a tight probabilistic frame.*

Bibliography

[1] Ali, S., Antoine, J.-P., and Gazeau, J.-P. *Coherent States, Wavelets, and Their Generalizations.* Springer, New York, 2000, ch. Positive Operator-Valued Measures and Frames.

[2] Ambrosio, L., Gigli, N., and Savare, G. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhäuser, Boston, 2005.

[3] Applegate, D., Dasu, T., Krishnan, S., and Urbanek, S. Unsupervised clustering of multidimensional distributions using Earth Mover distance. *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (2011), 636–644.

[4] Benamou, J.-D., and Brenier, Y. A computational fluid dynamics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik 84* (2000), 375–393.

[5] Benedetto, J. J., and Fickus, M. Finite normalized tight frames. *Advances in Computational Mathematics 18*, 2-4 (2003), 357–385.

[6] Benedetto, J. J., and Kebo, A. The role of frame force in quantum detection. *Journal of Fourier Analysis and Applications 14*, 3 (2008), 443–474.

[7] Bickel, P., and Levina, E. The Earth Mover's distance is the Mallows distance: Some insights from statistics. *Proceedings of ICCV 2001, Vancouver, Canada* (2001), 251–256.

[8] Bodmann, B. G., and Casazza, P. G. The road to equal-norm Parseval frames. *Journal of Functional Analysis 258* (2010), 397–420.

[9] Bodmann, B. G., Casazza, P. G., and Cahill, J. Fusion frames and the restricted isometry property. *Numerical Functional Analysis and Optimization 33*, 7-9 (2012), 770–790.

[10] Brenier, Y. Décomposition polaire et réarrangement des champs des vecteurs. *C.R. Acad. Sci. Paris Sér. I Math. 305* (1987), 805–808.

[11] Caffarelli, L. A. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society 5*, 1 (1992), 99–104.

[12] Cahill, J., Casazza, P. G., Ehler, M., and Li, S. Tight and random nonorthogonal fusion frames. *(preprint)* (2013).

[13] Cahill, J., Fickus, M., Mixon, D., Poteet, M., and Strawn, N. Constructing finite frames of a given spectrum and set of lengths. *Applied and Computational Harmonic Analysis* (2012).

[14] Carrillo, J. A., DiFrancesco, M., Figalli, A., Laurent, T., and Slepcev, D. Global-in-time weak measure solutions and finite time aggregation for nonlinear interaction equations. *Duke Mathematics Journal 156* (2011), 229–271.

[15] Carrillo, J. A., Lisini, S., and Mainini, E. Gradient flows for nonsmooth interaction potentials. *Nonlinear Analysis: Theory, Method, and Applications 100* (2014), 122–147.

[16] Casazza, P. G. *Wavelets, Frames and Operator Theory: Papers from the Focused Research Group Workshop, University of Maryland, College Park, MD, USA, 15-21 January 2003*. No. 345 in Contemporary Mathematics. American Mathematical Society, Providence, 2003, ch. Custom-building finite frames, pp. 15–21.

[17] Casazza, P. G., Fickus, M., and Mixon, D. Autotuning unit-norm frames. *Applied Computational Harmonic Analysis 32*, 1 (2012), 1–15.

[18] Casazza, P. G., Fickus, M., Mixon, D., Wang, Y., and Zhou, Z. Spectral tetris fusion frame constructions. *Journal of Fourier Analysis and Applications 18*, 4 (2012), 828–851.

[19] Casazza, P. G., and Kutyniok, G. *Wavelets, Frames and Operator Theory (College Park, MD, 2003)*. No. 345 in Contemporary Mathematics. American Mathematial Society, Providence, 2004, ch. Frames of Subspaces, pp. 87–113.

[20] Christensen, O. *An Introduction to Frames and Riesz Bases*. Birkhäuser, Boston, 2003.

[21] Cuesta-Albertos, J. A., and Matran, C. Notes on the Wasserstein metric in Hilbert spaces. *Annals of Probability 17*, 3 (1989), 1264–1276.

[22] Czaja, W. Remarks on Naĭmark's duality. *Proceedings of the American Mathematical Society 136*, 3 (2008), 867–871.

[23] Daubechies, I., Grossman, A., and Meyer, Y. Painless nonorthogonal expansions. *Journal of Mathematical Physics 27*, 5 (1986), 1271–1283.

[24] De Philippis, G., and Figalli, A. The monge-ampère equations and its links to optimal transportation. *(preprint)* (2013).

[25] De Philippis, G., and Figalli, A. Partial regularity results in optimal transportation. *(preprint)* (2014).

[26] Dowson, D. C., and Landau, B. V. The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis 12* (1982), 450–455.

[27] Duffin, R., and Schaeffer, A. A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society 72* (1952), 341–366.

[28] Dykema, K., and Strawn, N. Manifold structure of spaces of spherical tight frames. *International Journal of Pure and Applied Mathematics 28* (2006), 217–256.

[29] Ehler, M. Random tight frames. *Journal of Fourier Analysis and Applications 18*, 1 (2012), 1–20.

[30] Ehler, M., and Okoudjou, K. A. *Finite Frames, Ed. G. Kutyniok and P. G. Casazza*. Springer, New York, 2011, ch. Probabilistic frames: An overview.

[31] Ehler, M., and Okoudjou, K. A. Minimization of the probabilistic p-frame potential. *Journal of Statistical Planning and Inference 142*, 3 (2012), 645–659.

[32] Evans, L. C., and Gangbo, W. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Memoirs of the American Mathematical Society 137* (1999), viii–66.

[33] Figalli, A., Rifford, L., and Villani, C. Necessary and sufficient conditions for continuity of optimal transport maps on Riemannian manifolds. *Tohoku Mathematical Journal 63*, 4 (2011), 461–898.

[34] Fornasier, M., and Rauhut, H. Continuous frames, function spaces, and the discretization problem. *Journal of Fourier Analysis and Applications 11*, 3 (2005), 245–287.

[35] Gangbo, W. An introduction to the mass transportation theory and its applications. Lecture Notes, 2004.

[36] Gangbo, W., Kim, H. K., and Pacini, T. Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems. *Memoirs of the AMS 211*, 993 (2011).

[37] Gangbo, W., and McCann, R. J. The geometry of optimal transportation. *Acta Mathematica 177* (1996), 113–161.

[38] Han, D., Leng, J., and Huang, T. Optimal dual frames for communication coding with probabilistic erasures. *IEEE Transactions on Signal Processing 59* (2011), 5380–5389.

[39] Jordan, R., Kinderlehrer, D., and Otto, F. The variational formulation of the Fokker-Planck equation. *SIAM Journal of Mathematical Analysis 29* (1998), 1–17.

[40] Kamalinejad, E. Well-posedness of Wasserstein gradient flow solutions of higher order evolution equations. *(preprint)* (2012).

[41] Larson, D., Han, D., Kornelson, K., and Weber, E. *Frames for Undergraduates*. No. 40 in Student Mathematical Library. American Mathematical Society, Providence, 2007.

[42] Marsaglia, G., and Styan, G. Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra 2* (1974), 269–292.

[43] Merigot, Q. A multiscale approach to optimal transport. *Computer Graphics Forum 30*, 5 (2011), 1584–1592.

[44] Monge, G. Mémoire sur la théorie des déblais et des remblais. *Hist. de l'Acad. des Sciences de Paris* (1781), 666–704.

[45] Oktay, O. *Frame quantization theory and equiangular tight frames.* PhD thesis, University of Maryland, 2007.

[46] Powell, A., and Whitehouse, J. T. Error bounds for consistent reconstruction: random polytopes and coverage processes. *(preprint)* (2014).

[47] Rachev, S. T., and R uschendorf, L. *Mass Transportation Problems, Vol I: Theory, Vol II: Applications.* Probability and its applications. Springer-Verlag, New York, 1998.

[48] Solomon, J., Guibas, L., and Butscher, A. Dirichlet energy for analysis and synthesis of soft maps. *Proc. Eurographics Symposium on Geometry Processing* (2013).

[49] Szulc, T. Ranks of convex combinations of matrices. *Reliable Computing 2*, 2 (1996), 181–185.

[50] Venkov, B. *Réseaux euclideans, designs sphériques et formes modulaires.* No. 37 in Monogr. Enseign. Math. Enseignement Math, Geneva, 2001, ch. Réseaux et designs sphériques, pp. 10–86.

[51] Vershynin, R. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability 25*, 3 (2012), 655–686.

[52] Villani, C. *Topics in Optimal Transportation.* No. 58 in Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.

[53] Villani, C. *Optimal Transport, Old and New.* No. 338 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 2009.