Atomic Norm Minimization for Decomposition into Complex Exponentials and Optimal Transport in Fourier Domain

Laurent Condat

Atomic Norm Minimization for Decomposition into Complex Exponentials and Optimal Transport in Fourier Domain

Laurent Condat

Visual Computing Center
King Abdullah University of Science and Technology (KAUST)
Thuwal, Saudi Arabia

July 9, 2020. Final author's version. To appear in *Journal of Approximation Theory*

Abstract

This paper is devoted to the decomposition of vectors into sampled complex exponentials; or, equivalently, to the information over discrete measures captured in a finite sequence of their Fourier coefficients. We study existence, uniqueness, and cardinality properties, as well as computational aspects of estimation using convex semidefinite programs. We then explore optimal transport between measures, of which only a finite sequence of Fourier coefficients is known.

keywords: atomic norm, infinite dictionary, truncated moment problem, trigonometric moments, total variation norm, super-resolution, optimal transport

1 Introduction

Mathematical analysis aims at understanding complex objects by expressing them in terms of elementary objects, the *atoms*. The discrete Fourier transform is a typical example, which allows us to express a vector of size N as a unique linear combination of N orthogonal atoms, which are the sampled complex exponentials of frequencies the multiples of $1/N$. This framework is restrictive, however, and one may want to remove the constraint that the frequencies lie on this discrete grid. The modern paradigm of signal processing integrates the idea that a signal can be represented in a *dictionary* containing much more than N atoms, by choosing, among the multitude of possible representations offered by the redundancy, the *simplest* one [1]. Here, we study the representation of vectors as linear combinations of sampled complex exponentials, without any constraint on the frequencies. Thus, the dictionary contains the continuous infinity of sampled complex exponentials, with frequencies in $[0, 1)$.

We define the unit circle $\mathbb{T} = \{z \in \mathbb{C} \ : \ |z| = 1\}$ of the complex plane $\mathbb{C}$. For every $z \in \mathbb{C}$, $\arg(z) \in [0, 2\pi)$ is the argument of z, such that $z = |z|\mathrm{e}^{\mathrm{j}\arg(z)}$, where $\mathrm{j} = \sqrt{-1}$. We endow $\mathbb{T}$ with the distance $d : (z_1, z_2) \mapsto \min\big(|\arg(z_1) - \arg(z_2)|/(2\pi), 1 - |\arg(z_1) - \arg(z_2)|/(2\pi)\big)$.

Let $M \geq 1$ be an integer. We define the *real* Hilbert space

$$\mathbb{V} = \left\{ v = (v_m)_{m=-M}^{M} \in \mathbb{C}^{2M+1} \ : \ (\forall m)\ v_{-m} = v_m^* \right\}, \tag{1}$$

where $\cdot^*$ indicates complex conjugation, endowed with the usual inner product

$$\langle v, v' \rangle = \sum_{m=-M}^{M} v_m v'^*_m = v_0 v'_0 + 2\,\mathrm{Re}\Big(\sum_{m=1}^{M} v_m v'^*_m \Big) \in \mathbb{R}. \tag{2}$$

We consider the infinite family $\mathcal{A} = \big(\mathrm{a}(z)\big)_{z \in \mathbb{T}}$ of elements of $\mathbb{V}$, defined by

$$\mathrm{a}(z)_m = z^{-m} = \mathrm{e}^{-\mathrm{j}m \arg(z)}, \quad m = -M, \ldots, M. \tag{3}$$

We call $\mathcal{A}$ the *dictionary* and its elements $\mathrm{a}(z)$, which are sampled complex exponentials, the *atoms* [2–7]. The frequency of $\mathrm{a}(z)$ is defined as $\arg(z)/(2\pi) \in [0, 1)$.

In the following, v is an arbitrary nonzero element of $\mathbb{V}$.

We focus on the decomposition of v into atoms. For this, we need a parameterization of $\mathcal{A}$; that is, a way to express v as a linear combination of atoms. Since $\mathcal{A}$ contains a non-denumerable infinity of atoms, it is natural to work with finite signed Borel measures on $\mathbb{T}$, the set of which is denoted by $\mathcal{M}$. So, the analysis of v consists in determining a measure $\mu \in \mathcal{M}$, such that

$$v_m = \int_{\mathbb{T}} z^{-m} \mathrm{d}\mu(z) = \int_0^1 \mathrm{e}^{-\mathrm{j}2\pi f m} \mathrm{d}\mu(\mathrm{e}^{\mathrm{j}2\pi f}), \quad m = -M, \ldots, M. \tag{4}$$

Thus, the v_m are trigonometric moments, or Fourier coefficients, of μ. Therefore, we write

$$v = \mathcal{F}\mu \tag{5}$$

if the relation in (4) is satisfied, and we say that μ *explains* v.

So, one can view the problem as the identification of a measure from a truncated sequence of its trigonometric moments, the so-called *truncated moment problem* [8]. If one finds μ from $v = (v_m)_{m=-M}^{M}$, one can calculate all its moments $(v_m)_{m \in \mathbb{Z}}$ and so extrapolate v, hence the term *super-resolution* [9–11]: one recovers high-frequency information of μ, given only low-frequency information. The two "mathematical" and "signal processing" views of the problem consist in reasoning in terms of measures with moments v, or in terms of decompositions of v into atoms (viewing μ as the spectrum of v), respectively; this is just a matter of defining which are the time and frequency domains. We can note that the more general setting of the recovery of a measure or a spline function from linear measurements has been considered in the literature [12, 13], with several recent contributions to this topic [14–16]. In a nutshell, regularizing an inverse problem with a convex function yields solutions that are convex combinations of a small number of atoms, which depend on the geometry of the regularizer level sets. In this paper, we focus on the specific properties of recovering a measure, when the measurements are Fourier coefficients.

In this study, we are interested in atomic decompositions of v; that is, we look at questions of existence, uniqueness, number of atoms, and more generally at the information contained in v. Equivalently, we look at what can be said about a discrete measure given a finite number of its Fourier coefficients. We do not consider approximation, which consists in estimating a simple approximate atomic decomposition of v, e.g. when v is corrupted by noise; there is a vast literature on approximation, see [17–22] and references therein.

The paper is organized as follows: In Sections 2–5, we review the properties of atomic decompositions, the close relationship between positive semidefinite Toeplitz matrices and them, and the Prony method to extract the parameters from a mixture of atoms. Many results, known but scattered in the literature or unpublished because they are straightforward, are stated as remarks. In Section 6, we establish a precise characterization of the existence and uniqueness of a measure explaining v with minimal total variation norm. We explain how to determine this minimal measure numerically, by a two-step procedure: first solving a convex semidefinite program, for which we propose an iterative algorithm in Section 7; and then applying Prony's estimation method. In Section 8, we exhibit a new construction of an atomic decomposition of v with $2M$ atoms, uniformly located on $\mathbb{T}$. This construction plays an important role in Section 9, where the minimal atomic decomposition of a mixture of two atoms with opposite amplitudes is characterized. Finally, in Section 10, we define new functionals to do optimal transport between two elements of $\mathbb{V}$, by doing optimal transport of the atoms composing them. Equivalently, we perform optimal transport of measures, given only a finite number of their Fourier coefficients.

2 Atomic decompositions and atomic norm

We will see that we can reason with discrete measures only; that is, express v with a finite number of atoms. So, by anticipation, we call an *atomic decomposition* of v, the parameters $K \in \mathbb{N}\backslash\{0\}$, $z_k \in \mathbb{T}$ all distinct, $c_k \in \mathbb{R}\backslash\{0\}$, such that

$$v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k). \tag{6}$$

If v has such an atomic decomposition, the measure $\mu = \sum_{k=1}^{K} c_k \delta_{z_k}$ explains v, where δ_z is the Dirac measure located in $z \in \mathbb{T}$.

Remark 1. The dimension of $\mathbb{V}$ is $2M+1$. The rank of $\mathcal{A}$, which is the largest number of linearly independent atoms, is $2M+1$ as well. Indeed, the invertibility of the Vandermonde matrices

$$\begin{bmatrix} 1 & \mathrm{e}^{-\mathrm{j}\omega_1} & \dots & \mathrm{e}^{-\mathrm{j}(2M+1)\omega_1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \mathrm{e}^{-\mathrm{j}\omega_{2M+1}} & \dots & \mathrm{e}^{-\mathrm{j}(2M+1)\omega_{2M+1}} \end{bmatrix}, \tag{7}$$

for every set $\{\omega_1, \dots, \omega_{2M+1}\}$ of distinct elements of $[0, 2\pi)$, implies the invertibility of the matrices

$$\begin{bmatrix} \mathrm{e}^{\mathrm{j}M\omega_1} & \dots & \mathrm{e}^{\mathrm{j}\omega_1} & 1 & \mathrm{e}^{-\mathrm{j}\omega_1} & \dots & \mathrm{e}^{-\mathrm{j}M\omega_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{e}^{\mathrm{j}M\omega_{2M+1}} & \dots & \mathrm{e}^{\mathrm{j}\omega_{2M+1}} & 1 & \mathrm{e}^{-\mathrm{j}\omega_{2M+1}} & \dots & \mathrm{e}^{-\mathrm{j}M\omega_{2M+1}} \end{bmatrix}, \tag{8}$$

which yields the given value of the rank.

So, every family $\big(\mathrm{a}(z_1), \dots, \mathrm{a}(z_{2M+1})\big)$ of $2M+1$ distinct atoms is a basis of $\mathbb{V}$, i.e. is linearly independent and spans $\mathbb{V}$ by linear combinations with real weights.

Remark 2. The invertibility of Vandermonde matrices also implies that the *spark* of $\mathcal{A}$, which is the smallest number of linearly dependent atoms [4], is $2M+2$; $\mathcal{A}$ has *full spark* [23].

A direct consequence is the following [4, Theorem 3]: if v has two different atomic decompositions, they involve at least $\operatorname{spark}(\mathcal{A}) = 2M + 2$ different atoms.

Therefore, if v has an atomic decomposition with $1 \leq K \leq M$ atoms, there is no other atomic decomposition of v with less or as many atoms. Moreover, we will see in Section 4 that it is easy to find this decomposition. It is in the case where v does not have a decomposition with $1 \leq K \leq M$ atoms, a property which is easy to test, that it is very difficult to find the decomposition of v with the smallest number of atoms (this number being at most $2M$, see Remark 13).

Remark 3. There is an infinity of atomic decompositions of v with at most $2M + 1$ atoms. Indeed, according to Remark 1, given $2M + 1$ arbitrary distinct atoms, one can express v as a linear combination of them, with real weights c_k, obtained by solving a linear system. Moreover, if v has a decomposition with more than $2M+1$ atoms, one can re-express all the atoms in terms of a subset of size $2M + 1$ of them, to obtain a decomposition of v with at most $2M + 1$ atoms. Other said, there is an infinity of discrete measures, made of at most $2M + 1$ Dirac measures, explaining v.

Remark 4. We can exhibit the "classical" uniform atomic decomposition[1]

$$v = \sum_{k=1}^{2M+1} c_k \mathrm{a}(\mathrm{e}^{\mathrm{j}2\pi k/(2M+1)}), \tag{9}$$

where the sequence $(c_k)_{k=1}^{2M+1}$ is simply the inverse discrete Fourier transform of v:

$$c_k = \frac{1}{2M+1} \sum_{m=-M}^{M} v_m \mathrm{e}^{\mathrm{j}2\pi km/(2M+1)}, \quad k = 1, \ldots, 2M+1. \tag{10}$$

However, in most applications, v is a mixture of unknown atoms, which must be identified, with frequencies living in the continuum $[0, 1)$. Enforcing the frequencies to lie on the grid of the multiples of $1/(2M+1)$, or actually on any predetermined finite set of atoms, yields undesirable effects: the reconstruction or approximation of the estimated signal is not parsimonious, with spurious oscillations; this is called spectral leakage, basis mismatch, or gridding error [24–26]. This phenomenon is somewhat attenuated by taking a finer and finer grid, while in doing so, with more and more coherent atoms, the estimation problems become increasingly badly conditioned and numerically unstable. This justifies to place ourselves in a continuous framework, as we do in this study, instead of in a discrete framework with a finite, even if redundant, dictionary of predetermined atoms [4, 27, 28]. Thus, we are interested in finding, among the infinite multitude of possible atomic decompositions of v, the "simplest" ones, with respect to some notion of parsimony. Finding the atomic decomposition with the smallest number of atoms is a very difficult nonconvex problem. That is why we focus on its convex relaxation and consider the form of optimality defined in the following.

We define the *atomic norm* of v as

$$\|v\|_{\mathrm{a}} := \inf \left\{ \|\mu\|_{\mathrm{TV}} : \ \mu \in \mathcal{M}, \ v = \mathcal{F}\mu \right\}, \tag{11}$$

[1]Strictly speaking, we should remove from this sum the indices k, for which $c_k = 0$.

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm of a measure (not to be confused with the total variation of an image [29]). In that respect, let us recall the following property [30]: every measure $\mu \in \mathscr{M}$ has a unique Jordan decomposition (μ^+, μ^-) of two finite positive measures μ^+ and μ^-, such that $\mu = \mu^+ - \mu^-$, satisfying the following optimality property: for every pair of positive measures (ν^+, ν^-) with $\mu = \nu^+ - \nu^-$, then $\nu^+ \geq \mu^+$ and $\nu^- \geq \mu^-$. Hence, we can define $\|\mu\|_{\mathrm{TV}} = \mu^+(\mathbb{T}) + \mu^-(\mathbb{T})$.

As easily checked, the atomic norm is indeed a norm on $\mathbb{V}$.

We call a *minimal atomic decomposition* of v, an atomic decomposition of v, such that $\sum_{k=1}^{K} |c_k| = \|v\|_{\mathrm{a}}$.

Remark 5. The atomic norm is defined in several papers [5–7] as

$$\|v\|_{\mathrm{a}} = \inf\left\{t \geq 0 \ : \ v \in t\,\mathrm{conv}(\mathscr{A} \cup -\mathscr{A})\right\}, \tag{12}$$

where conv denotes the convex hull, which, since $\mathbb{V}$ is of dimension $2M+1$, is the set of convex combinations of $2M+1$ signed atoms. Therefore, we can equivalently write (12) as

$$\|v\|_{\mathrm{a}} = \inf\Bigg\{ \sum_{k=1}^{K} |c_k| \ : \ v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k), \text{ for some } K \leq 2M+1, \text{ distinct } z_k \in \mathbb{T},\ c_k \in \mathbb{R}\backslash\{0\}\Bigg\}. \tag{13}$$

We will see in Proposition 2 that these two definitions (11) and (12)–(13) coincide, and that the infimum in (11), (12), (13) is attained; that is, it is a minimum. In other words, a minimal atomic decomposition of v exists.

3 Relation to Toeplitz matrices

In order to characterize more precisely the atomic norm and the atomic decompositions of v minimizing it, we introduce the linear operator T, which maps v to the Hermitian Toeplitz matrix

$$\mathrm{T}(v) := \begin{bmatrix} v_0 & v_1 & \cdots & v_M \\ v_{-1} & v_0 & \cdots & v_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{-M} & v_{-M+1} & \cdots & v_0 \end{bmatrix}. \tag{14}$$

We can note that T defines a bijection between $\mathbb{V}$ and the set of $(M+1)\times(M+1)$ Hermitian Toeplitz matrices, thereafter denoted by $\mathrm{T}(\mathbb{V})$. So, we will talk equally about an atomic decomposition of v or of $\mathrm{T}(v)$.

$\mathrm{T}(v)$ is said to be indefinite if it is neither positive semidefinite nor negative semidefinite; that is, if it has at least one positive and one negative eigenvalues. $\mathrm{T}(v)$ is said to be singular if it is not invertible; that is, if it has at least one eigenvalue equal to zero.

We remark that, for every $z \in \mathbb{T}$,

$$\mathrm{T}(\mathrm{a}(z)) = \begin{bmatrix} 1 & z^{-1} & \cdots & z^{-M} \\ z & 1 & \cdots & z^{-M+1} \\ \vdots & \vdots & \ddots & \vdots \\ z^{M} & z^{M-1} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{M} \end{bmatrix} \begin{bmatrix} 1 & z^{-1} & \cdots & z^{-M} \end{bmatrix} \tag{15}$$

is positive semidefinite and of rank 1.

Remark 6. If v has an atomic decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$ with $1 \leq K \leq M+1$ atoms, the rank of $\mathrm{T}(v)$ is K. Indeed,

$$\mathrm{T}(v) = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_K \\ \vdots & \vdots & \vdots \\ z_1^M & \cdots & z_K^M \end{bmatrix} \times \begin{bmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_K \end{bmatrix} \times \begin{bmatrix} 1 & z_1^{-1} & \cdots & z_1^{-M} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_K^{-1} & \cdots & z_K^{-M} \end{bmatrix} \tag{16}$$

and the rank of $A^{\mathrm{H}}BA$, where A is a matrix with full column rank and $\cdot^{\mathrm{H}}$ denotes the Hermitian transpose, is equal to the rank of B [31, Lemma 1].

As we see, the rank of $\mathrm{T}(v)$ will play an important role in the analysis. At this stage, one can wonder if the converse of the last property is true; that is, if there always exists an atomic decomposition of v with $K = \mathrm{rank}(\mathrm{T}(v))$ atoms. Unfortunately, this is not the case, as we will see in Section 4.

Remark 7. One can think that if v has an atomic decomposition with $M+1 < K \leq 2M+1$ atoms, the rank of $\mathrm{T}(v)$ is $M+1$. This is not true in general and we can construct examples where the rank is as low as $2M+2-K$. Indeed, if $\sum_{k=1}^{2M+2} c_k \mathrm{a}(z_k) = 0$ (given distinct z_k, such $c_k \neq 0$ exist, since the $\mathrm{a}(z_k)$ are linearly dependent), then for every $M+1 < K \leq 2M+1$, $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k) = \sum_{k=K+1}^{2M+2} (-c_k)\mathrm{a}(z_k)$, and there are two different atomic decompositions of v, the second one having less atoms. The rank of $\mathrm{T}(v)$ is then $2M+2-K$.

Remark 8. The trace of $\mathrm{T}(v)$, which is also the sum of its eigenvalues, is equal to $(M+1)v_0$. So, it is equal to $(M+1)\mu(\mathbb{T})$, for every measure μ explaining v, and to $(M+1)\sum_{k=1}^{K} c_k$, for every atomic decomposition of v.

Remark 9. Let $\mathrm{n}^+(\mathrm{T}(v))$ and $\mathrm{n}^-(\mathrm{T}(v))$ be the number of positive and negative eigenvalues of $\mathrm{T}(v)$, respectively. Then every atomic decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$ has at least $\mathrm{n}^+(\mathrm{T}(v))$ positive c_k and at least $\mathrm{n}^-(\mathrm{T}(v))$ negative c_k.

As we see, decomposition into exponentials is intimately related to the spectral properties of Hermitian Toeplitz matrices.

4 Prony's estimation method

The following estimation method, already known by G. R. de Prony in the 18th century [32], allows us to know, in a simple way, whether v has an atomic decomposition with $K \leq M$ atoms, and if so, to find the corresponding parameters. We outline the method in this section, see [33–35] for more details.

First, let us suppose that there exists a decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$ with $1 \leq K \leq M$. We define the *annihilating polynomial* $H(z) = \sum_{k=0}^{K} h_k z^k := \prod_{k=1}^{K}(z - z_k)$. This polynomial owes its name to the fact that it cancels the sequence v by convolution [36, 37]:

$$\sum_{k=0}^{K} h_k v_{m-k} = \sum_{k=0}^{K} h_k \sum_{l=1}^{K} c_l z_l^{k-m} = \sum_{l=1}^{K} c_l z_l^{-m} H(z_l) = 0,$$
$$m = -M+K, \ldots, M. \tag{17}$$

Let us form the rectangular Toeplitz matrix of size $(2M+1-K)\times(K+1)$,

$$\mathrm{R}_K(v) := \begin{bmatrix} v_{M-K} & \cdots & v_M \\ v_{M-K-1} & \cdots & v_{M-1} \\ \vdots & \ddots & \vdots \\ v_{-M} & \cdots & v_{-M+K} \end{bmatrix}. \tag{18}$$

Then, the equations (17) can be rewritten as

$$\mathrm{R}_K(v)[h_K \ \cdots \ h_0]^{\mathrm{T}} = [0 \ \cdots \ 0]^{\mathrm{T}}. \tag{19}$$

Thus, the vector $[h_K \ \cdots \ h_0]^{\mathrm{T}}$ is in the kernel of $\mathrm{R}_K(v)$. More precisely, $\mathrm{R}_K(v)$ is of rank K and its kernel, of dimension 1, is spanned by $[h_K \ \cdots \ h_0]^{\mathrm{T}}$.

Now, we consider the converse of these properties. Let K be the rank of $\mathrm{T}(v)$, with $K \leq M$ (if $\mathrm{T}(v)$ is positive definite, see Remark 11, whereas if $\mathrm{T}(v)$ is of full rank and is not positive definite, the method cannot be applied to find an atomic decomposition of v). Let us form the matrix $\mathrm{R}_K(v)$ as in (18). It is of rank K as well. Let $[h_K \ \cdots \ h_0]^{\mathrm{T}}$ be the unique, up to a constant, nonzero vector in the kernel of $\mathrm{R}_K(v)$. For a numerically robust estimation, we compute the singular value decomposition of $\mathrm{R}_K(v)$ and we take $[h_K \ \cdots \ h_0]^{\mathrm{T}}$ as the right singular vector corresponding to the smallest singular value. Then, we compute the roots z_k of the polynomial $H(z) := \sum_{k=0}^{K} h_k z^k$. If we find K distinct roots belonging to $\mathbb{T}$, then v has an atomic decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$, where the amplitudes c_k are obtained by solving the least-squares linear system

$$Q^{\mathrm{H}}Q[c_1 \ \cdots \ c_K]^{\mathrm{T}} = Q^{\mathrm{H}}[v_{-M} \ \cdots \ v_M]^{\mathrm{T}}, \tag{20}$$

$$\text{where } Q := \begin{bmatrix} z_1^{-M} & \cdots & z_K^{-M} \\ \vdots & \vdots & \vdots \\ z_1^{M} & \cdots & z_K^{M} \end{bmatrix}. \tag{21}$$

In the other cases, v does not have an atomic decomposition with at most M atoms.

Note that the annihilating polynomial can have less than K nonzero roots. For instance, for $v = (2,1,1,1,1,1,2)$,

$$\mathrm{T}(v) = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \end{bmatrix} \tag{22}$$

is of rank $K = 3$, so $\mathrm{R}_K(v) = \mathrm{T}(v)$ and the annihilating polynomial is $H(z) = z(z-1)$, which has only one nonzero root.

In the case where $H(z)$ has K distinct nonzero roots z_k, but not all in $\mathbb{T}$, we can write $v_m = \sum_{k=1}^{K} c_k z_k^{-m}$, but this amounts to decomposing v into other objects than the atoms of $\mathcal{A}$. This is linked to the fact that the Hermitian Toeplitz matrices of rank 1 are of the form $\mathrm{T}((z^{-M},\ldots,z^{M}))$, but for an arbitrary $z \in \mathbb{C}\backslash\{0\}$, not necessarily in $\mathbb{T}$.

We will see in the following that if $\mathrm{T}(v)$ is positive semidefinite, Prony's method always yields K distinct roots in $\mathbb{T}$. If $\mathrm{T}(v)$ is indefinite, the key for decomposing v will consist in splitting $\mathrm{T}(v)$ into the sum of two positive semidefinite and negative semidefinite Toeplitz matrices.

5 Case of positive decompositions

For every measure μ explaining v, $\|\mu\|_{\mathrm{TV}} \geq \mu^+(\mathbb{T}) \geq \mu^+(\mathbb{T}) - \mu^-(\mathbb{T}) = v_0$, so we always have $\|v\|_{\mathrm{a}} \geq v_0$. Moreover, if a positive measure μ explains v, we have $\|\mu\|_{\mathrm{TV}} = \mu(\mathbb{T}) = v_0$, so that the infimum in (11) is attained for this measure, with $\|v\|_{\mathrm{a}} = \|\mu\|_{\mathrm{TV}} = v_0$. So, it is interesting to address, in the first place, the case where v is explained by a positive measure.

Remark 10. If a discrete measure $\sum_{k=1}^{K} c_k \delta_{z_k}$ explains v, with $1 \leq K \leq M+1$ and $c_k > 0$, then $\mathrm{T}(v)$ is positive semidefinite and of rank K. Indeed, $\forall z \in \mathbb{T}$, $\mathrm{T}(\mathrm{a}(z))$ is positive semidefinite and the set of positive semidefinite matrices is a closed convex cone. More precisely, for every vector $x \in \mathbb{C}^{M+1}$, $x^{\mathrm{H}}\mathrm{T}(v)x = \int_{\mathbb{T}} x^{\mathrm{H}}\mathrm{T}(\mathrm{a}(z))x \,\mathrm{d}\mu(z) = \int_{\mathbb{T}} |\sum_{m=1}^{M+1} z^{1-m} x_m|^2 \,\mathrm{d}\mu(z) \geq 0$. Moreover, we have seen in Remark 6 that $\mathrm{T}(v)$ is of rank exactly K.

A major result, which implies the converse of the property in Remark 10, dates back to the work of Carathéodory [38, 39]. It can be formulated as follows:

Carathéodory's Theorem. *$\mathrm{T}(v)$ is positive and of rank K, with $1 \leq K \leq M+1$, if and only if there exist distinct elements $z_k \in \mathbb{T}$ and positive reals c_k, such that $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$. Moreover, if $K \leq M$, the discrete measure $\sum_{k=1}^{K} c_k \delta_{z_k}$ is the unique positive measure explaining v.*

Thus, if $\mathrm{T}(v)$ is positive semidefinite, a positive discrete measure $\mu = \sum_{k=1}^{K} c_k \delta_{z_k}$, with $K \leq M+1$, explains v, and we have $\|\mu\|_{\mathrm{TV}} = \sum_{k=1}^{K} c_k = v_0 = \|v\|_{\mathrm{a}}$. In other words, the Theorem states the existence of a minimal atomic decomposition of v, if $\mathrm{T}(v)$ is positive semidefinite. Moreover, according to Remark 2, there exists no decomposition of v with less atoms, so the obtained atomic decomposition is doubly optimal.

Remark 11. If $\mathrm{T}(v)$ is positive definite, with rank $K = M+1$, there exists an infinity of minimal atomic decompositions of v with $M+1$ atoms. We can construct all of them as follows [40, Remark 2.1]: we choose $z_{M+1} \in \mathbb{T}$ and we set

$$c_{M+1} := \left([1 \; z_{M+1}^{-1} \; \cdots \; z_{M+1}^{-M}]\mathrm{T}(v)^{-1}[1 \; z_{M+1} \; \cdots \; z_{M+1}^{M}]^{\mathrm{T}}\right)^{-1}. \tag{23}$$

Then $c_{M+1} > 0$ and

$$\mathrm{T}\big(v - c_{M+1}\mathrm{a}(z_{M+1})\big) = \mathrm{T}(v) - [1 \; z_{M+1} \; \cdots \; z_{M+1}^{M}]^{\mathrm{T}}[1 \; z_{M+1}^{-1} \; \cdots \; z_{M+1}^{-M}] \tag{24}$$

is positive semidefinite of rank M. Carathéodory's Theorem applied to this residual yields $v - c_{M+1}\mathrm{a}(z_{M+1}) = \sum_{k=1}^{M} c_k \mathrm{a}(z_k)$. All in all, we end up with the positive minimal atomic decomposition $v = \sum_{k=1}^{M+1} c_k \mathrm{a}(z_k)$.

For instance, for $v = (0, \ldots, 0, 1, 0, \ldots, 0)$, i.e. $\mathrm{T}(v) = \mathrm{Id}$, the identity matrix, we get the atomic decompositions, for every $\varphi \in [0, 2\pi)$,

$$v = \frac{1}{M+1} \sum_{k=1}^{M+1} \mathrm{a}\big(\mathrm{e}^{\mathrm{j}(2\pi k + \varphi)/(M+1)}\big). \tag{25}$$

Remark 12. There is a complete proof of Carathéodory's Theorem by Curto and Fialkow [8, section 6]. Given Remarks 10 and 11, there only remains to prove that, if $\mathrm{T}(v)$ is positive semidefinite of rank $1 \leq K \leq M$, then a positive measure explaining v exists, is unique,

and is concentrated at K points of $\mathbb{T}$. The difficult part, which is not detailed here, is to show the existence. The rest of the proof is simple: we consider the annihilating polynomial $H(z) = \sum_{k=0}^{K} h_k z^k = \prod_{k=1}^{K}(z - z_k)$ defined in Section 4 and we define the vector of size $M+1$, $\tilde{h} = [h_K \ \cdots \ h_0 \ 0 \ \cdots \ 0]^{\mathrm{T}}$. Let $\mu \in \mathscr{M}$ be a positive measure explaining v. We have

$$0 = \tilde{h}^{\mathrm{H}} \mathrm{T}(v) \tilde{h} = \sum_{0 \leq k,l \leq K} h_k^* h_l v_{k-l} = \int_{\mathbb{T}} \sum_{0 \leq k,l \leq K} h_k^* h_l z^{l-k} \mathrm{d}\mu(z) \tag{26}$$
$$= \int_{\mathbb{T}} |H(z)|^2 \mathrm{d}\mu(z).$$

Therefore, μ is a discrete measure concentrated at the roots z_k of $H(z)$ belonging to $\mathbb{T}$. Since μ cannot be made of less than K Diracs measures (else the rank of $\mathrm{T}(v)$ would be less than K, according to Remark 10), all the z_k, $k = 1, \ldots, K$, are in $\mathbb{T}$. Finally, the unique coefficients c_k such that $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$, defined in (20), are positive, according to Remark 9. □

Corollary 1. If a positive measure, which is not a discrete measure concentrated in at most M points, explains v, then $\mathrm{T}(v)$ is of rank $M+1$. This is specific to the positive case, see Remark 7 for the general case.

Corollary 2. An atom cannot be expressed as a linear combination of other atoms with positive coefficients. So, the atoms are the extremal points of the convex hull of $\mathscr{A}$.

6 General case

Remark 13. In the previous section, we have seen that if $\mathrm{T}(v)$ is positive semidefinite, v has a decomposition with at most $M+1$ atoms. The analysis is of course the same when $\mathrm{T}(v)$ is negative semidefinite. So, let us direct our attention to the case where $\mathrm{T}(v)$ is indefinite. We have seen in Remark 3 that there are atomic decompositions of v with $2M+1$ atoms. We can now improve this bound and construct decompositions of v with at most $2M$ atoms. We will see in Proposition 3 one such decomposition. Another construction is the following. Let $i^- < 0$ be the smallest eigenvalue of $\mathrm{T}(v)$. $\mathrm{T}(v) - i^-\mathrm{Id}$ is positive semidefinite and singular, so by Carathéodory's Theorem, it can be decomposed into at most M atoms. As in Remark 11, we can decompose $i^-\mathrm{Id}$ into $M+1$ atoms, one of which being an atom of $\mathrm{T}(v) - i^-\mathrm{Id}$. By combining them, $\mathrm{T}(v) = (\mathrm{T}(v) - i^-\mathrm{Id}) + i^-\mathrm{Id}$ can indeed be decomposed into at most $2M$ atoms.

Remark 14. The bound above of $2M$ atoms is optimal: it is possible that v has a decomposition with $2M$ atoms and not less. We can give the example [41, Example 3.4], for $M = 2$, of

$$v = (3, 1, 1, 1, 3) = \frac{3}{2}\mathrm{a}(1) - \frac{1}{2}\mathrm{a}(\mathrm{j}) + \frac{1}{2}\mathrm{a}(-1) - \frac{1}{2}\mathrm{a}(-\mathrm{j}), \tag{27}$$

for which it is shown that no decomposition with less than four atoms exists [41].

We now focus on minimal atomic decompositions of v. For this purpose, let us consider the three equivalent formulations of the same convex optimization problem, where tr denotes the

trace, $\succcurlyeq 0$ denotes positive semidefiniteness, and $V := \mathrm{T}(v)$:

$$V^+ := \underset{X}{\operatorname{argmin}}\ \mathrm{tr}(X) \quad \text{s.t.} \quad X \in \mathrm{T}(\mathbb{V}) \quad \text{and} \quad X \succcurlyeq 0 \quad \text{and} \quad X - V \succcurlyeq 0, \tag{28}$$

$$\equiv\ V^+ := \underset{X}{\operatorname{argmin}}\ \tfrac{1}{M+1}\big(\mathrm{tr}(X) + \mathrm{tr}(X - V)\big) \quad \text{s.t.}\ X \in \mathrm{T}(\mathbb{V}) \quad \text{and} \quad X \succcurlyeq 0 \quad \text{and} \quad X - V \succcurlyeq 0, \tag{29}$$

$$\equiv\ V^+ := \mathrm{T}(v^+), \ \text{where}\ \ v^+ := \underset{x \in \mathbb{V}}{\operatorname{argmin}}\ (2x_0 - v_0) \quad \text{s.t.} \quad \mathrm{T}(x) \succcurlyeq 0 \quad \text{and} \quad \mathrm{T}(x - v) \succcurlyeq 0. \tag{30}$$

Proposition 1. *The solution V^+ of the problem* (28) *exists and is unique. Accordingly, let us define $V^- := V^+ - V$. Moreover, if V is positive semidefinite, $V^+ = V$ and $V^- = 0$; if V is negative semidefinite, $V^+ = 0$ and $V^- = -V$; else, V^+ and V^- are singular.*

Proof: the case where V is positive semidefinite or negative semidefinite is obvious, so let us suppose that V is indefinite. We first remark that the feasible set of the problem (28) is not empty: let $i^- < 0$ be the smallest eigenvalue of V; $X = V - i^-\mathrm{Id}$ is positive semidefinite, Hermitian and Toeplitz, as well as $X - V$.

Then the trace defines a norm for the positive semidefinite matrices, and, as a consequence of the Weierstrass theorem [42, Theorem 1.28], a norm attains its minimum on a closed set in finite dimension, so a solution V^+ exists.

Let V^+ be a solution of (28). Set $V^- := V^+ - V$. Suppose that V^+ is positive definite. Then there exists $0 < a < 1$, such that $X := V^+ - aV^-$ is positive semidefinite. Moreover, $X \in \mathrm{T}(\mathbb{V})$ and $X - V = (1 - a)V^-$ is positive semidefinite. But $\mathrm{tr}(X) < \mathrm{tr}(V^+)$, which contradicts the optimality of V^+. Hence, V^+ is singular. By the same reasoning, V^- is singular as well.

Let us finally show that the solution of (28) is unique. Let X_1 and X_2 be two solutions. Then $X = (X_1 + X_2)/2$ is a solution as well, so X is singular. By Carathéodory's Theorem, X has an atomic decomposition $\sum_{k=1}^{K} c_k \mathrm{a}(z_k)$, where $1 \leq K \leq M$ is the rank of X. Let us define the annihilating polynomial $H(z) = \sum_{k=0}^{K} h_k z^k := \prod_{k=1}^{K}(z - z_k)$ and the vector of size $M+1$, $\tilde{h} = [h_K \ \cdots \ h_0 \ 0 \ \cdots \ 0]^{\mathrm{T}}$. We have $\tilde{h}^{\mathrm{H}} X \tilde{h} = 0$, so $\tilde{h}^{\mathrm{H}} X_1 \tilde{h} + \tilde{h}^{\mathrm{H}} X_2 \tilde{h} = 0$, and since X_1 and X_2 are positive semidefinite, $\tilde{h}^{\mathrm{H}} X_1 \tilde{h} = \tilde{h}^{\mathrm{H}} X_2 \tilde{h} = 0$. As in (26), this implies that the positive atomic decompositions of X_1 and X_2 are of the form $\sum_{k=1}^{K} c_{1,k} \mathrm{a}(z_k)$ and $\sum_{k=1}^{K} c_{2,k} \mathrm{a}(z_k)$, respectively, for some $c_{1,k} \geq 0$ and $c_{2,k} \geq 0$. By the same reasoning, $X_1 - V$ and $X_2 - V$ have atomic decompositions of same support, of the form $\sum_{k=1}^{K'} c'_{1,k} \mathrm{a}(z'_k)$ and $\sum_{k=1}^{K'} c'_{2,k} \mathrm{a}(z'_k)$, respectively, with $1 \leq K' \leq M$, some $z'_k \in \mathbb{T}$ all distinct and different from the z_k, and some $c'_{1,k} \geq 0$ and $c'_{2,k} \geq 0$. Therefore, $v = \sum_{k=1}^{K} c_{1,k} \mathrm{a}(z_k) - \sum_{k=1}^{K'} c'_{1,k} \mathrm{a}(z'_k) = \sum_{k=1}^{K} c_{2,k} \mathrm{a}(z_k) - \sum_{k=1}^{K'} c'_{2,k} \mathrm{a}(z'_k)$. By subtraction, $\sum_{k=1}^{K}(c_{1,k} - c_{2,k})\mathrm{a}(z_k) + \sum_{k=1}^{K'}(c'_{2,k} - c'_{1,k})\mathrm{a}(z'_k) = 0_{\mathbb{V}}$, the zero sequence of $\mathbb{V}$. Since the atoms are linearly independent, for every k, $c_{1,k} = c_{2,k}$ and $c'_{1,k} = c'_{2,k}$. Hence, $X_1 = X_2$ and the solution, call it V^+, is unique. □

Consequently, to every $V \in \mathrm{T}(\mathbb{V})$, we can assign a unique pair (V^+, V^-) of Hermitian Toeplitz positive semidefinite matrices, with V^+ solution of (28) and $V^- = V^+ - V$. By analogy with the

Jordan decomposition of measures, we call (V^+, V^-) the *Jordan decomposition* of V. Equivalently, to every $v \in \mathbb{V}$, we can assign a unique pair $(v^+, v^-) \in \mathbb{V}^2$, with v^+ solution of (30) and $v^- = v^+ - v$. We call (v^+, v^-) the Jordan decomposition of v.

Proposition 2.
(i) *The problem*

$$\underset{\mu \in \mathcal{M}}{\text{minimize}} \ \|\mu\|_{\text{TV}} \quad s.t. \quad v_m = \int_{\mathbb{T}} z^{-m} \mathrm{d}\mu(z), \quad m = -M, \ldots, M, \tag{31}$$

has a solution, concentrated at a finite number $K \leq 2M$ of points of $\mathbb{T}$, i.e. of the form $\mu^\star = \sum_{k=1}^{K} c_k \delta_{z_k}$, for distinct $z_k \in \mathbb{T}$ and nonzero reals c_k. Equivalently, a minimal atomic decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$ exists. We obtain the parameters z_k and c_k by combining those of v^+ and v^-; that is $v^+ = \sum_{k=1}^{K^+} c_k \mathrm{a}(z_k)$ and $v^- = \sum_{k=K^++1}^{K^++K^-} (-c_k)\mathrm{a}(z_k)$, where $K = K^+ + K^-$, $K^+ := \mathrm{rank}(V^+)$, $K^- := \mathrm{rank}(V^-)$, $V := \mathrm{T}(v)$. The existence of these two atomic decompositions is guaranteed by Carathéodory's Theorem, since $V^+ \succcurlyeq 0$ and $V^- \succcurlyeq 0$. Thus,

$$\|v\|_{\mathrm{a}} = \|\mu^\star\|_{\text{TV}} = \sum_{k=1}^{K} |c_k| = \tfrac{1}{M+1}\big(\mathrm{tr}(V^+) + \mathrm{tr}(V^-)\big) = v_0^+ + v_0^- = 2v_0^+ - v_0. \tag{32}$$

(ii) *The solution $\mu^\star$ exhibited in* (i) *is the unique solution of the problem* (31) *if and only if* $\mathrm{T}(v)$ *is neither positive definite, nor negative definite. More precisely, if* $\mathrm{T}(v)$ *is neither positive definite, nor negative definite, the measure $\mu^\star = \sum_{k=1}^{K} c_k \delta_{z_k}$ exhibited in* (i) *is the unique solution of the problem* (31). *Then we have $0 \leq K^+ \leq M$ and $0 \leq K^- \leq M$. On the contrary, if* $\mathrm{T}(v)$ *is positive definite (resp. negative definite), there is an infinity of measures, necessarily positive (resp. negative), solution of* (31)*, among which an infinity of discrete measures concentrated at exactly $K = M + 1$ points of $\mathbb{T}$; we can construct them explicitly, according to Remark 11.*

Proof: (i) let $\mu \in \mathcal{M}$, with Jordan decomposition (μ^+, μ^-), explaining v. Let $(x^+, x^-) \in \mathbb{V}^2$, such that μ^+ explains x^+ and μ^- explains x^-. We have $x^+ - x^- = v$ and $\|\mu\|_{\text{TV}} = x_0^+ + x_0^- = 2x_0^+ - v_0$. Since x^+ is feasible for the problem (30), the solution v^+ of which exists and is unique, μ is a solution of (31) if and only if $x^+ = v^+$ (and $x^- = v^-$). The rest of the statement follows.
(ii) Suppose that $V = \mathrm{T}(v)$ is neither positive definite nor negative definite. Then V^+ and V^- are singular, according to Proposition 1, so $K^+ = \mathrm{rank}(V^+) \leq M$ and $K^- = \mathrm{rank}(V^-) \leq M$. Let μ be a solution of (30). μ^+ and μ^- explain v^+ and v^-, respectively, as shown above. But according to Carathéodory's Theorem, such positive measures are unique. So, the construction in (i) yields the unique solution. □

So, we have shown that the infimum is attained in (11), (12), (13). Thus, it is licit to replace the infimum by a minimum in these definitions of the atomic norm. In addition, we can strengthen the constraint $K \leq 2M + 1$ to $K \leq 2M$ in (13).

Finally, we can write, for every $v \in \mathbb{V}$,

$$\begin{aligned} \|v\|_{\mathrm{a}} = \min_{X} \ & \left(\tfrac{2}{M+1}\mathrm{tr}(X) - v_0\right) \quad \text{s.t.} \quad X \in \mathrm{T}(\mathbb{V}) \\ & \text{and} \quad X \succcurlyeq 0 \quad \text{and} \quad X - \mathrm{T}(v) \succcurlyeq 0. \end{aligned} \tag{33}$$

We can note that this semidefinite characterization of the atomic norm was given in [43, eq. 12], without any proof.

If we go back to the problem of finding a decomposition of v with the smallest number of atoms, it amounts to:

$$\begin{aligned} \underset{X}{\text{minimize}} \ & \big(\operatorname{rank}(X) + \operatorname{rank}(X - V)\big) \quad \text{s.t.} \quad X \in \mathrm{T}(\mathbb{V}) \\ & \text{and} \quad X \succcurlyeq 0 \quad \text{and} \quad X - V \succcurlyeq 0. \end{aligned} \tag{34}$$

Such nonconvex problems of rank minimization are notoriously difficult, and atomic norm minimization is a convex and efficient alternative.

Given an atomic decomposition $v = \sum_{k=1}^{K} c_k \mathrm{a}(z_k)$, we define its *separation* $D(\{z_k\}) := \min_{k \neq l} d(z_k, z_l)$, as the minimal distance between its atoms. A remarkable result by Candès and Fernandez-Granda [9] is the following: if v has an atomic decomposition with large enough separation, then it is both the minimal atomic decomposition of v and the atomic decomposition of v with the fewest atoms; that is, the solutions of (29) and (34) are the same and correspond to this decomposition. A sufficient condition for this is $D(\{z_k\}) \geq 1.87/M$ (and $M \geq 128$) [9]. This condition was further strengthened to $D(\{z_k\}) \geq 1.26/M$ (and $M \geq 1000$) [44]. Some authors conjecture that $D(\{z_k\}) \geq 1/M$ is a sufficient condition [22], see also [45]. Note that every minimal atomic decomposition has a distance $d(z_k, z_l) \geq 1/(2M)$ between two atoms, if c_k is positive and c_l is negative [22, Corollary 2]. Of course, these sufficient conditions imply that the number of atoms is $K \leq M$, and we have seen that in that case, Prony's method yields this optimal decomposition directly. But we cannot hope for a stable method to recover atomic decompositions with $K \geq M + 1$ atoms, characterized by $2K \geq 2M + 2$ real parameters, from v, characterized by $2M + 1$ real degrees of freedom. More importantly, the fact that separated enough atomic decompositions are represented by solutions of convex optimization problems, opens the door to their robust approximate estimation by convex minimization, from inexact or noisy coefficients $y_m \approx v_m$. For instance, given $y \in \mathbb{V}$, we can

$$\underset{x \in \mathbb{V}}{\text{minimize}} \ \|y - x\|_2^2 + \lambda \|x\|_{\mathrm{a}}, \tag{35}$$

$$\begin{aligned} \equiv \underset{(x^+, x^-) \in \mathbb{V}^2}{\text{minimize}} \ & \|y - x^+ + x^-\|_2^2 + \lambda(x_0^+ + x_0^-) \\ \text{s.t.} \quad & \mathrm{T}(x^+) \succcurlyeq 0 \quad \text{and} \quad \mathrm{T}(x^-) \succcurlyeq 0, \end{aligned} \tag{36}$$

for some regularization parameter $\lambda > 0$. Approximation goes beyond the scope of this paper and we refer to [46] and [47] for some results on the topic.

Remark 15. A measure $\mu \in \mathcal{M}$ is said to be *identifiable* if, given its sequence of Fourier coefficients $v = (v_m)_{m=-M}^{M}$, it is the unique solution of the problem (31). That is, an identifiable measure can be perfectly recovered from $v = \mathcal{F}\mu$. If $\mathrm{T}(v)$ is indefinite, we have seen above that well separated measures are identifiable, but this is only a sufficient condition: there are measures which are not well separated but are identifiable. The picture is different if we restrict ourselves to positive measures; as a consequence of Proposition 2, we have the property: *a positive measure is identifiable if and only if it is concentrated at $K \leq M$ points of $\mathbb{T}$*. It is remarkable that all the information about a positive discrete measure concentrated in at most M points is encoded in its coefficients $(v_m)_{m=1}^{M}$ (since v_0 can be recovered as the opposite of the smallest eigenvalue of $\mathrm{T}((v_M^*, \ldots, v_1^*, 0, v_1, \ldots, v_M))$), characterized by the critical number $2M$ of real degrees of freedom, without any constraint.

7 Numerical resolution of (30)

Since the equivalent problems (28) and (30) are semidefinite programs, many efficient algorithmic strategies exist to solve them to an arbitrary precision. It is beyond the scope of this paper to compare them. We present a first order iterative algorithm, which is simple to implement and turns out to be fast. It is an instance of the overrelaxed version [48, 49] of the Chambolle–Pock algorithm [50]. The algorithm is the following, where T^* is the adjoint operator of T and $P_{\preccurlyeq 0}$ denotes the projection onto the convex cone of negative semidefinite matrices, by computing the eigendecomposition and setting the positive eigenvalues to zero.

Algorithm 1. Input: $v \in \mathbb{V}$. Output: estimate $x^{(i)}$ of v^+, the solution of (30). Choose the proximal parameter $\tau > 0$ and the relaxation parameter $\rho \in [1, 2)$. Choose the initial estimate $x^{(0)} \in \mathbb{V}$ of v^+. Set $\sigma := 1/\tau/(2\|\mathrm{T}\|^2) = 1/\tau/(2M+2)$. Set $U_1^{(0)} := U_2^{(0)} := \mathrm{T}(0_{\mathbb{V}})$, the zero matrix. Set $V := \mathrm{T}(v)$. Then iterate, for $i = 0, 1, \ldots$

$$
\left|\begin{array}{ll}
1. & \tilde{x}^{(i+1)} := x^{(i)} - \tau \mathrm{T}^*(U_1^{(i)} + U_2^{(i)}) - \tau(1, 0, \ldots, 0), \\
2. & \widetilde{U}_1^{(i+1)} := P_{\preccurlyeq 0}\left(U_1^{(i)} + \sigma \mathrm{T}(2\tilde{x}^{(i+1)} - x^{(i)})\right), \\
3. & \widetilde{U}_2^{(i+1)} := P_{\preccurlyeq 0}\left(U_2^{(i)} - \sigma V + \sigma \mathrm{T}(2\tilde{x}^{(i+1)} - x^{(i)})\right), \\
4. & (x^{(i+1)}, U_1^{(i+1)}, U_2^{(i+1)}) := \rho(\tilde{x}^{(i+1)}, \widetilde{U}_1^{(i+1)}, \widetilde{U}_2^{(i+1)}) \\
 & \quad + (1-\rho)(x^{(i)}, U_1^{(i)}, U_2^{(i)}).
\end{array}\right.
$$

The algorithm converges to the solution v^+ of (30) [48, Theorem 5.3]. Moreover, the number of positive eigenvalues of $U_1^{(i)} + \sigma \mathrm{T}(2\tilde{x}^{(i+1)} - x^{(i)})$ and $U_2^{(i)} - \sigma V + \sigma \mathrm{T}(2\tilde{x}^{(i+1)} - x^{(i)})$ are robust estimates of $K^+ = \mathrm{rank}(V^+)$ and $K^- = \mathrm{rank}(V^-)$, the number of positive and negative atoms in the minimal atomic decomposition of v, respectively. The complexity of every iteration is dominated by the computation of two eigendecompositions, so the algorithm has complexity $O(M^3)$ per iteration. In practice, we initialize the algorithm with the positive part of the decomposition given in Proposition 3; that is, given the parameters $\{z_k\}$, $\{c_k\}$ of this decomposition, we set $x^{(0)} = \sum_{k=1}^{2M} \max(c_k, 0)\mathrm{a}(z_k)$. So, $x^{(0)}$ is feasible for the problem (30); that is, $\mathrm{T}(x^{(0)}) \succcurlyeq 0$ and $\mathrm{T}(x^{(0)} - v) \succcurlyeq 0$.

We can note that the overrelaxed Chambolle–Pock algorithm can also be used to solve (36).

An alternative way to compute the atomic norm of v is to solve the semidefinite program [6, 7, 9]

$$
\begin{aligned}
&\underset{q \in \mathbb{C}^{4M+1} \text{ s.t. } q_{-m} = q_m^*,\, t \in \mathbb{R}}{\text{minimize}} \quad \tfrac{1}{2}(q_0 + t) \\
&\text{s.t.} \quad \left[\begin{array}{ccc|c} & \mathrm{T}(q) & & \begin{array}{c} v_{-M} \\ \vdots \\ v_M \end{array} \\ \hline v_M & \cdots & v_{-M} & t \end{array}\right] \succcurlyeq 0,
\end{aligned}
\tag{37}
$$

the optimal value of which is equal to $\|v\|_{\mathrm{a}}$. Since the constraint in (37) implies $\mathrm{T}(q) \succcurlyeq 0$, Carathéodory's Theorem is applicable and Prony's method applied to $\mathrm{T}(q)$ allows us to recover the parameters z_k of the minimal atomic decomposition of v. It is easy to show that at optimality,

$t = q_0 = \|v\|_{\mathrm{a}}$, so that it is better, for faster convergence, to remove the unnecessary variable t and to consider the problem

$$\begin{aligned} &\underset{q\in\mathbb{C}^{4M+1} \text{ s.t. } q_{-m}=q_m^*}{\text{minimize}} \quad q_0 \\ &\text{s.t.} \quad \left[\begin{array}{ccc|c} & \mathrm{T}(q) & & \begin{array}{c} v_{-M} \\ \vdots \\ v_M \end{array} \\ \hline v_M & \cdots & v_{-M} & q_0 \end{array}\right] \succcurlyeq 0. \end{aligned} \tag{38}$$

A typical algorithm to solve (37) or (38) needs one eigendecomposition per iteration, instead of two with the proposed Algorithm 1, but it involves a matrix twice larger. Since the complexity of the eigendecomposition scales with the cube of the matrix size, we expect Algorithm 1 to be faster. We leave for future work a numerical comparison of different formulations and algorithms to compute the atomic norm of v.

8 Uniform decomposition with at most $2M$ atoms

In this section, we present a new construction of a *uniform* atomic decomposition of v, with at most $2M$ atoms.

Proposition 3. *There exists an atomic decomposition of v, with at most $2M$ atoms located uniformly on $\mathbb{T}$, with an explicit form:*

$$v = \sum_{k=1}^{2M} c_k \mathrm{a}(z_k), \tag{39}$$

$$\textit{where} \quad z_k := \mathrm{e}^{\mathrm{j}(2\pi k/(2M)-\phi)} \quad \textit{and} \quad \phi := \arg(v_M)/M \tag{40}$$

(if $v_M = 0$, every ϕ is suitable), and the c_k are obtained by inverse discrete Fourier transform from the sequence $(v_m \mathrm{e}^{-\mathrm{j}m\phi})_{m=-M}^{M-1}$:

$$c_k := \frac{1}{2M} \sum_{m=-M}^{M-1} (v_m \mathrm{e}^{-\mathrm{j}m\phi}) \mathrm{e}^{\mathrm{j}2\pi mk/(2M)}, \quad k = 1, \ldots, 2M. \tag{41}$$

Proof: we first remark that the c_k in (41) are real, since $v_{-m}\mathrm{e}^{-\mathrm{j}(-m)\phi} = (v_m \mathrm{e}^{-\mathrm{j}m\phi})^*$ and this value is real for $m = 0$ and $m = -M$. Then it is sufficient to notice that

$$v_m \mathrm{e}^{-\mathrm{j}m\phi} = \sum_{k=1}^{2M} c_k \mathrm{e}^{-\mathrm{j}2\pi mk/(2M)}, \quad m = -M, \ldots, M-1. \tag{42}$$

So (we check that the formula is also valid for $m = M$),

$$v_m = \sum_{k=1}^{2M} c_k \mathrm{e}^{-\mathrm{j}m(2\pi k/(2M)-\phi)}, \quad m = -M, \ldots, M. \qquad \square \tag{43}$$

Note that the number of atoms in the proposition is at most $2M$, and not exactly $2M$, because some c_k can be zero in (39) and (41).

In terms of extrapolation of $v = (v_m)_{m=-M}^{M}$ to the infinite sequence $(v_m)_{m\in\mathbb{Z}}$ of Fourier coefficients of $\mu = \sum_{k=1}^{2M} c_k \delta_{\mathrm{e}^{\mathrm{j}(2\pi k/(2M)-\phi)}}$, we get

$$v_m = \mathrm{e}^{\mathrm{j}(m-[m])\phi} v_{[m]}, \quad m \in \mathbb{Z}, \tag{44}$$

where $[m] = \mathrm{rem}(m + M, 2M) - M$ and $\mathrm{rem}(n, N) \in \{0, \ldots, N-1\}$ is the remainder of the Euclidean division of $n \in \mathbb{Z}$ by N.

Remark 16. Duval and Peyré showed that there exists a measure with separation $1/(2M)$, which is not identifiable, by exhibiting $\delta_{\mathrm{e}^{-\mathrm{j}2\pi/(2M)}} + \delta_{\mathrm{e}^{0}} - \delta_{\mathrm{e}^{\mathrm{j}2\pi/(2M)}}$ and proving that it is not the solution of (31) [22, Corollary 1]. According to Proposition 3, measures with separation $1/(2M)$, which are not identifiable, are the rule rather than the exception. Indeed, every $v \in \mathbb{V}$ is explained by the measure with separation $1/(2M)$ constructed in Proposition 3, and there is no reason that this measure has minimal total variation norm, in general.

Now, let us consider the case where v is made of a single atom: $v = c\mathrm{a}(z)$, for some $c > 0$ and $z \in \mathbb{T}$ (the case $c < 0$ is addressed the same way). $\mathrm{T}(v)$ is positive definite, of rank 1, so the analysis of Section 5 shows that $v = c\mathrm{a}(z)$ is the unique positive atomic decomposition of v and the unique minimal atomic decomposition of v. Moreover, this single-atom decomposition coincides with the one of Proposition 3: the c_k in (41) are all zero, except $c_{k_0} = c$ for $k_0 = \{2M \text{ if } \omega = 0,\ 2M(\phi+\omega)/(2\pi) \text{ else}\}$, where $\omega = \arg(z)$. Indeed, if $\omega > 0$, $\phi = \arg(\mathrm{e}^{-\mathrm{j}M\omega})/M$, so $\phi = -\omega + 2\pi k_0/(2M)$ for some even $k_0 \in \{2, 4, \ldots, 2M\}$. Since the sequence $(v_m \mathrm{e}^{-\mathrm{j}m\phi} = c\mathrm{e}^{-\mathrm{j}2\pi m k_0/(2M)})_{m=-M}^{M-1}$ is a sampled complex exponential of frequency multiple of $1/(2M)$, its inverse discrete Fourier transforms gives c_k all equal to zero, except $c_{k_0} = c$.

After this simple case of a single atom, we direct our attention to the case of two atoms with opposite amplitudes, in the next section.

9 Case of 2 atoms with opposite amplitudes

Let us now consider that v is the sum of two atoms with opposite amplitudes. So, let z_1 and z_2 be two distinct elements of $\mathbb{T}$ and let c be a positive real. We set

$$v = c\mathrm{a}(z_1) - c\mathrm{a}(z_2). \tag{45}$$

Set $\omega_1 := \arg(z_1)$, $\omega_2 := \arg(z_2)$. Then we have

$$v_m = c\mathrm{e}^{-\mathrm{j}m\omega_1} - c\mathrm{e}^{-\mathrm{j}m\omega_2}, \quad m = -M, \ldots, M. \tag{46}$$

We introduce the 2π-periodic function

$$\Upsilon : \omega \in \mathbb{R} \mapsto \frac{\sin\big(\omega(2M+1)/2\big) + \sin\big(\omega(2M-1)/2\big)}{4M\sin(\omega/2)} \tag{47}$$

$$= \frac{1}{4M}\mathrm{e}^{-\mathrm{j}M\omega} + \frac{1}{2M}\sum_{m=-M+1}^{M-1} \mathrm{e}^{\mathrm{j}m\omega} + \frac{1}{4M}\mathrm{e}^{\mathrm{j}M\omega}, \tag{48}$$

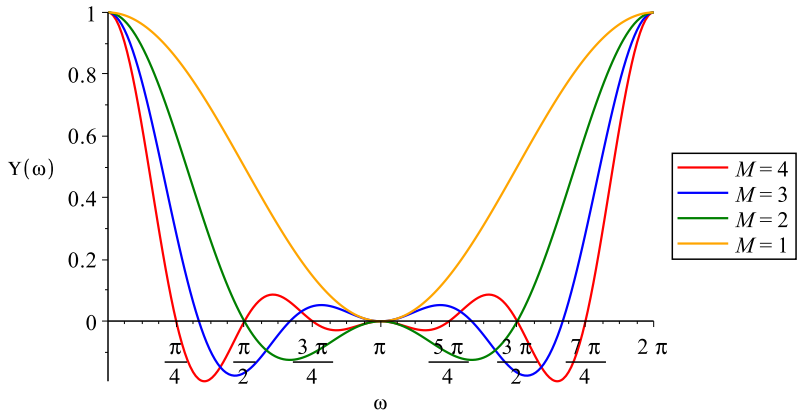


Figure 1: Plot of the function $\Upsilon(\omega)$, defined in (47), for different values of M.

extended by continuity to $\Upsilon(0) = 1$. Note that $\Upsilon(2\pi k/(2M)) = 0$ for every $k = 1, \ldots, 2M-1$. The function Υ is plotted in Figure 1 for the first values of M.

Proposition 4. *In the present case of two atoms with opposite amplitudes, the decomposition of Proposition 3 can be rewritten as (up to a circular permutation on the indices):*

$$v = \sum_{k=1}^{2M} c_k \mathrm{a}(\mathrm{e}^{\mathrm{j}\omega_k}), \tag{49}$$

$$\text{where } \omega_k = \frac{2\pi k}{2M} - \frac{2\pi}{4M} + \frac{\omega_1 + \omega_2}{2}, \tag{50}$$

$$c_k = c\Upsilon\left(\frac{2\pi k}{2M} - \frac{2\pi}{4M} + \frac{\omega_2 - \omega_1}{2}\right) - c\Upsilon\left(\frac{2\pi k}{2M} - \frac{2\pi}{4M} + \frac{\omega_1 - \omega_2}{2}\right). \tag{51}$$

Proof: By combining (46), (41) and the fact that $v_{-M}\mathrm{e}^{\mathrm{j}M\phi} = v_M \mathrm{e}^{-\mathrm{j}M\phi} \in \mathbb{R}$, we obtain

$$c_k = c\Upsilon\left(\tfrac{2\pi k}{2M} - \phi - \omega_1\right) - c\Upsilon\left(\tfrac{2\pi k}{2M} - \phi - \omega_2\right). \tag{52}$$

Moreover, we have $\phi = \arg(\mathrm{e}^{-\mathrm{j}M\omega_1} + \mathrm{e}^{-\mathrm{j}(M\omega_2+\pi)})/M$, so $\phi = -(\omega_1+\omega_2)/2 + 2\pi/(4M) + 2\pi k_0/(2M)$, for some $k_0 \in \mathbb{Z}$. □

An example is illustrated in Figure 2. In Figure 3, we show the behavior of the atomic norm of v, as a function of $d(z_1, z_2)$. We observe, without proof, that if this distance is larger than or equal to $1/(2M)$, the measure $\delta_{z_1} - \delta_{z_2}$ is identifiable.

10 Optimal transport controlled in Fourier domain

Building upon the analysis in the previous section, we now focus on optimal transport between two measures, given only finite sequences of their Fourier coefficients. Optimal transport of

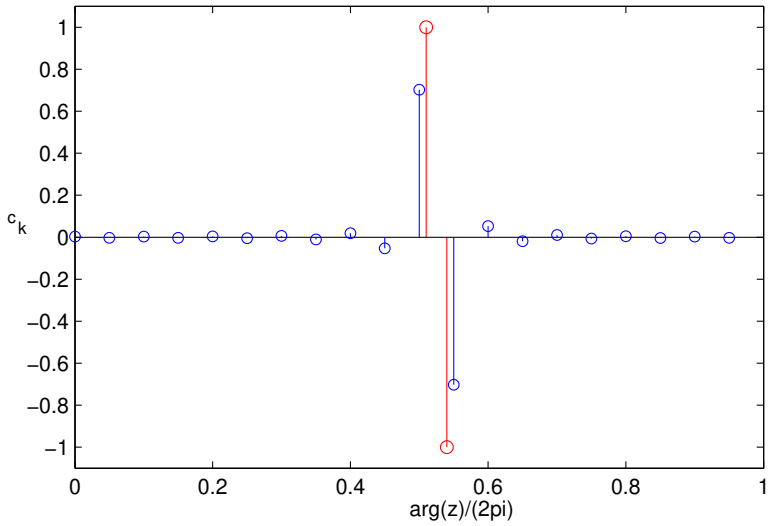


Figure 2: The two measures $\mu^\sharp = \delta_{\mathrm{e}^{\mathrm{j}2\pi\,0.51}} - \delta_{\mathrm{e}^{\mathrm{j}2\pi\,0.54}}$, in red, and $\mu^\star = \sum_{k=1}^{2M} c_k \delta_{z_k}$ given by Propositions 3 and 4, in blue, for $M = 10$, have the same sequence of Fourier coefficients $v = (v_m)_{m=-M}^{M}$. These two measures have total variation norm $\|\mu^\sharp\|_{\mathrm{TV}} = 2$ and $\|\mu^\star\|_{\mathrm{TV}} = \sum_{k=1}^{2M} |c_k| \approx 1.6$, respectively. Numerical minimization of the atomic norm of v yields the minimal atomic decomposition $v = \sum_{k=1}^{2M} c_k \mathrm{a}(z_k)$, so $\|v\|_{\mathrm{a}} = \|\mu^\star\|_{\mathrm{TV}}$ and $\mu^\star$ is the measure of minimal total variation norm explaining v.

measures [51, 52] has grown significantly in recent years, due to a wide range of applications, from image processing [53] to speech processing [54].

Given a nonnegative cost function f on $\mathbb{T}^2$, the optimal transport cost $\mathcal{T}_f(\mu, \mu')$ between two positive measures μ and μ' of $\mathcal{M}^2$, with same mass $\mu(\mathbb{T}) = \mu'(\mathbb{T})$, is the infimum, over all positive measures ν on $\mathbb{T}^2$ with marginals μ and μ' on the first and second factors, respectively, of the cost $\int_{\mathbb{T}^2} f(z, z')\mathrm{d}\nu(z, z')$ [51, 55].

Let us extend this definition to signed measures. We define the optimal transport cost $\mathcal{T}_f$ between a pair of signed measures $(\mu, \mu') \in \mathcal{M}^2$ with same mass as the infimum, over all finite signed Borel measures ν on $\mathbb{T}^2$ with marginals μ and μ', of $\int_{\mathbb{T}^2} f(z, z')\mathrm{d}|\nu|(z, z')$, where $|\nu| = \nu^+ + \nu^-$. This definition, however, is meaningful only for concave costs, since it vanishes for convex costs: if the cost is concave, i.e. $f = \phi \circ d$ for some increasing and concave function ϕ with $\phi(0) = 0$, then $\mathcal{T}_f(\mu, \mu') = \mathcal{T}_f(\mu^+ + \mu'^-, \mu'^+ + \mu^-)$.

Some classical cost functions f are the following. First, if we consider the binary distance $f(z, z') = \{0 \text{ if } z = z', 1 \text{ else}\}$, then $\mathcal{T}_f(\mu, \mu')$ is equal to the scaled Radon distance $d_{\mathcal{R}}(\mu, \mu')$ defined as

$$d_{\mathcal{R}}(\mu, \mu') = \tfrac{1}{2}\|\mu - \mu'\|_{\mathrm{TV}}, \tag{53}$$

for every μ and μ' in $\mathcal{M}$ with same mass. Second, if $f = d$, for positive measures μ and μ' with same mass, then $\mathcal{T}_f$ becomes the *Wasserstein-1 distance*, denoted by $d_{\mathcal{W}1}$. Since f is a concave function of d, $\mathcal{T}_f$ can be extended to signed measures, not only positive ones. Finally, if $f = d^2$, for positive measures with same mass, the square root of $\mathcal{T}_f$ is the *Wasserstein-2 distance* denoted by $d_{\mathcal{W}2}$. We are not able to extend this functional to signed measures, however.

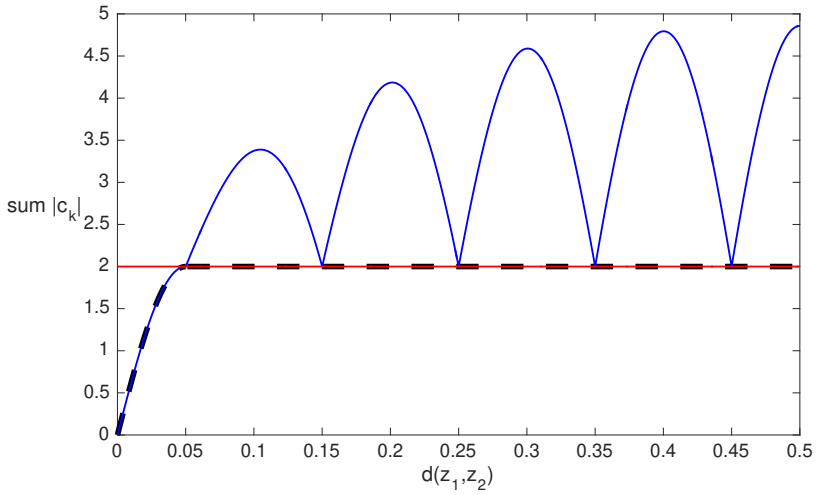


Figure 3: Value of the total variation norm $\|\mu^\sharp\|_{\mathrm{TV}} = \{0 \text{ if } z_1 = z_2,\ 2 \text{ else}\}$, in red, and $\|\mu^\star\|_{\mathrm{TV}} = \sum_{k=1}^{2M} |c_k|$, in blue, as a function of $d(z_1, z_2)$, for $M = 10$, where the two measures $\mu^\sharp = \delta_{z_1} - \delta_{z_2}$ and $\mu^\star = \sum_{k=1}^{2M} c_k \delta_{z_k}$ given by Proposition 4, have the same sequence of Fourier coefficients $v = (v_m)_{m=-M}^{M}$. In black, the atomic norm of v. We observe that when $d(z_1, z_2) < 1/(2M)$, $\mu^\star$ is the measure of minimal total variation norm explaining v, so $\|v\|_{\mathrm{a}} = \|\mu^\star\|_{\mathrm{TV}} = 2\sin\big(\pi M d(z_1, z_2)\big)$. Else, when $d(z_1, z_2) \geq 1/(2M)$, $\mu^\sharp$ is the measure of minimal total variation norm explaining v, so $\|v\|_{\mathrm{a}} = \|\mu^\sharp\|_{\mathrm{TV}} = 2$.

Since we want to do optimal transport between measures, given only finite sequences of their Fourier coefficients, we now introduce the *atomic transport cost*, denoted by $\mathscr{T}_{\mathrm{a},f}$, associated to any nonnegative cost function f on $\mathbb{T}^2$. We restrict ourselves to the concave case; that is, $f = \phi \circ d$ for some increasing and concave function ϕ with $\phi(0) = 0$. So, for every $(v, v') \in \mathbb{V}^2$ with $v_0 = v'_0$,

$$\mathscr{T}_{\mathrm{a},f}(v, v') := \min\left\{\mathscr{T}_f(\mu, \mu') \ : \ (\mu, \mu') \in \mathscr{M}^2, \quad \mathscr{F}\mu = v,\ \mathscr{F}\mu' = v'\right\}. \tag{54}$$

The difficulty is of course to exhibit computable expressions of the atomic transport cost.

In the simple case of the binary distance $f(z, z') = \{0 \text{ if } z = z',\ 1 \text{ else}\}$, we call the atomic transport cost the *atomic Radon distance*, denoted $d_{\mathrm{a}0}$, and we have, for every $(v, v') \in \mathbb{V}^2$ with $v_0 = v'_0$,

$$\begin{aligned}
d_{\mathrm{a}0}(v, v') &:= \min\left\{d_{\mathscr{R}}(\mu, \mu') \ : \ (\mu, \mu') \in \mathscr{M}^2,\ \mathscr{F}\mu = v,\ \mathscr{F}\mu' = v'\right\} && (55)\\
&= \min\left\{\tfrac{1}{2}\|\mu - \mu'\|_{\mathrm{TV}} \ : \ (\mu, \mu') \in \mathscr{M}^2,\ \mathscr{F}\mu = v,\ \mathscr{F}\mu' = v'\right\} && (56)\\
&= \min\left\{\tfrac{1}{2}\|\eta\|_{\mathrm{TV}} \ : \ \eta \in \mathscr{M},\ \mathscr{F}\eta = v - v'\right\} && (57)\\
&= \tfrac{1}{2}\|v - v'\|_{\mathrm{a}}. && (58)
\end{aligned}$$

This distance, which is convex with respect to the pair (v, v'), allows us to quantify whether v and v' are different. Its precision naturally depends on M. We show in Figure 4 the value of $d_{\mathrm{a}0}$ taken at a pair of atoms, as a function of the distance of their frequencies, which is explained by the analysis carried out in Section 9.

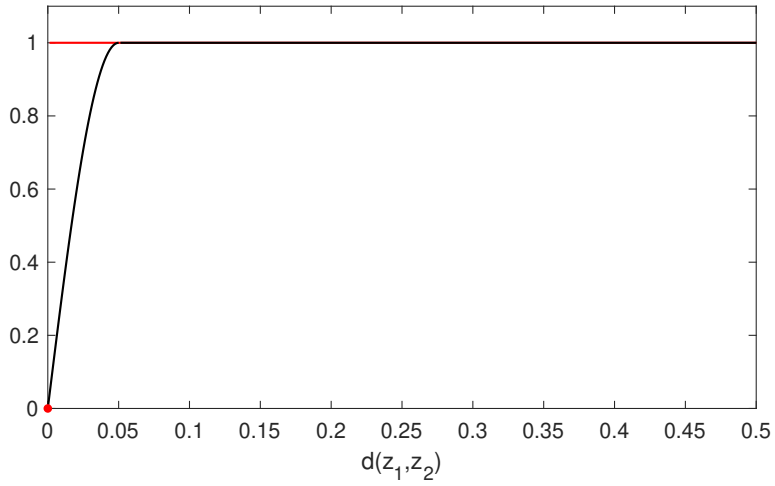


Figure 4: Value of the atomic Radon distance $d_{\mathrm{a0}}\big(\mathrm{a}(z_1), \mathrm{a}(z_2)\big)$ as a function of $d(z_1, z_2)$, in black, with $M = 10$. We observe that if $d(z_1, z_2) < 1/(2M)$, $d_{\mathrm{a0}}\big(\mathrm{a}(z_1), \mathrm{a}(z_2)\big) = \sin\big(\pi M d(z_1, z_2)\big)$. However, if $d(z_1, z_2) \geq 1/(2M)$, $d_{\mathrm{a0}}\big(\mathrm{a}(z_1), \mathrm{a}(z_2)\big) = 1$, which is the expected value, namely the Radon distance between δ_{z_1} and δ_{z_2}.

Let us now turn our attention to the 1-Wasserstein distance. When $f = d$, we call the atomic transport cost the *atomic 1-Wasserstein distance*, denoted by d_{a1}, defined as follows: for every $(v, v') \in \mathbb{V}^2$ with $v_0 = v'_0$,

$$d_{\mathrm{a1}}(v, v') := \min\big\{d_{\mathscr{W}1}(\mu, \mu') \ : \ (\mu, \mu') \in \mathscr{M}^2, \ \mathscr{F}\mu = v, \ \mathscr{F}\mu' = v'\big\}. \tag{59}$$

Like for the atomic Radon distance, we can express the atomic 1-Wasserstein distance as a semidefinite program:

Proposition 5. *Let $(v, v') \in \mathbb{V}^2$ with $v_0 = v'_0$. Set $w = \big((v_m - v'_m)/(\mathrm{j}2\pi m)\big)_{m=-M}^{M}$, with $w_0 = 0$, and set $W = \mathrm{T}(w)$. Then*

$$\begin{aligned} d_{\mathrm{a1}}(v, v') = \min\big\{&\|\eta\|_{\mathrm{TV}} \ : \ \eta \in \mathscr{M}, \ \mathscr{F}\eta = w \ \text{ with} && (60)\\ &\mathrm{j}2\pi m w_m = v_m - v'_m, \ \ m = -M, \dots, M\big\},\\ = \min_{X, \beta \in \mathbb{R}} &\big(\tfrac{2}{M+1}\mathrm{tr}(X) + \beta\big) \quad \text{s.t.} \quad X \in \mathrm{T}(\mathbb{V}) && (61)\\ &\text{and} \quad X \succcurlyeq 0 \quad \text{and} \quad X - W + \beta \mathrm{Id} \succcurlyeq 0,\\ = \min_{X} &\big(\tfrac{2}{M+1}\mathrm{tr}(X) + \mathrm{i}^+(W - X)\big) \quad \text{s.t.} && (62)\\ &X \in \mathrm{T}(\mathbb{V}) \quad \text{and} \quad X \succcurlyeq 0, \end{aligned}$$

where i^+ denotes the largest eigenvalue.

Proof: For every $(\mu, \mu') \in \mathscr{M}^2$, it is known that $d_{\mathscr{W}1}(\mu, \mu')$ only depends on $\mu - \mu'$ and has the expression

$$d_{\mathscr{W}1} = \min_{\alpha \in \mathbb{R}} \frac{1}{2\pi} \int_{\mathbb{T}} |F(z) - F'(z) - \alpha| \mathrm{d}z, \tag{63}$$

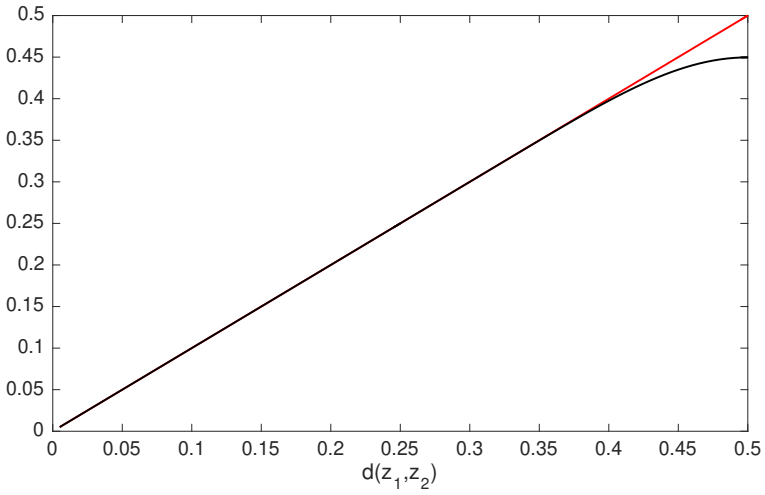


Figure 5: Value of the atomic 1-Wasserstein distance $d_{\mathrm{a}1}\big(\mathrm{a}(z_1), \mathrm{a}(z_2)\big)$ as a function of $d(z_1, z_2)$, in black, with $M = 10$. We observe that the atomic 1-Wasserstein distance is equal to $d(z_1, z_2)$, in red, except for the largest values, where this distance is underestimated.

where F and F' are the cumulative distribution functions of μ and μ', respectively [55, 56]. Let $\alpha \in \mathbb{R}$ and set the function $\eta = F - F' - \alpha$. Set $v = \mathcal{F}\mu$ and $v' = \mathcal{F}\mu'$. Let us interpret η as a measure in $\mathcal{M}$ and set $w = \mathcal{F}\eta$. We have the relationship $\mathrm{j}2\pi m w_m = v_m - v'_m$, for every $m = -M, \ldots, M$. The expression (60) follows. Then we see that $d_{\mathrm{a}1}(v, v')$ takes a form similar to the atomic norm of w, but with w_0, or equivalently the real α in (63) or the real β in (61), as an additional degree of freedom. So, (61) is nothing but the semidefinite characterization (33) of the atomic norm. Finally, for a fixed X, the real β minimizing β subject to $X - W + \beta\mathrm{Id} \succcurlyeq 0$ is the largest eigenvalue of $W - X$, so that we can remove this variable. □

In Figure 5, we can see that when v and v' are atoms, $d_{\mathrm{a}1}(v, v')$ closely matches the distance between their frequencies.

Finally, let us focus on the case of the squared 2-Wasserstein distance $d^2_{\mathcal{W}2} = \mathcal{T}_f$ with $f = d^2$. Since f is not concave with respect to d, the optimal transport cost $\mathcal{T}_f(\mu, \mu')$ does not only depend on $\mu - \mu'$ [55, 56]. Moreover, it seems difficult to extend it to a convex functional with respect to a pair of signed measures. So, we restrict ourselves to the positive case. That is, given $(v, v') \in \mathbb{V}^2$ with $v_0 = v'_0$, $\mathrm{T}(v) \succcurlyeq 0$ and $\mathrm{T}(v) \succcurlyeq 0$, we define the *atomic 2-Wasserstein distance* as

$$d_{\mathrm{a}2}(v, v') := \min\Big\{ d_{\mathcal{W}2}(\mu, \mu') \;:\; \mu \text{ and } \mu' \text{ are positive measures}, \; \mathcal{F}\mu = v, \; \mathcal{F}\mu' = v' \Big\}. \tag{64}$$

We can note that this distance is exact for atoms: for every $z \in \mathbb{T}$, there is only one positive measure explaining $\mathrm{a}(z)$, which is δ_z, and there is only one positive measure ν on $\mathbb{T}^2$ having two Diracs as marginals, which is a Dirac. So, for every $(z, z') \in \mathbb{T}^2$, $d_{\mathrm{a}2}\big(\mathrm{a}(z), \mathrm{a}(z')\big)^2 = d(z, z')^2$. However, there is little hope to express the squared atomic 2-Wasserstein distance as a finite-dimensional convex program. So, we will propose an approximation of this functional. We

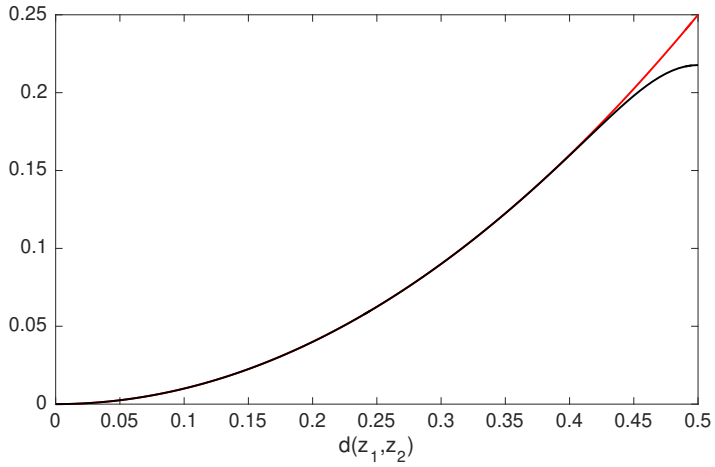


Figure 6: Value of the approximate squared atomic 2-Wasserstein distance $\tilde{d}_{\mathrm{a2},z_2}\big(\mathrm{a}(z_1)\big)^2$ as a function of $d(z_1, z_2)$, in black, with $M = 10$. We observe that the this function is very close to $d(z_1, z_2)^2$, in red, except for the largest values, where the squared distance is underestimated.

restrict ourselves to the case where v' is fixed and is an atom: $v' = \mathrm{a}(z')$ for some $z' \in \mathbb{T}$. We want to design a function $\tilde{d}_{\mathrm{a2},z'}$ of $v \in \mathbb{V}$, with $v_0 = 1$ and $\mathrm{T}(v) \geqslant 0$, whose square approximates $d_{\mathrm{a2}}(v, v')^2 = \min\{\int_{\mathbb{T}} d(z, z')^2 \mathrm{d}\mu(z) \ : \ \mu \text{ is a positive measure and } \mathcal{F}\mu = v\}$. For this, we mimic the construction of d_{a1}, which is the total variation of a measure whose 'first derivative' explains $v - v'$, up to a constant: we define $\tilde{d}_{\mathrm{a2},z'}(v)^2$ as the total variation of a measure whose 'second derivative' explains $\big(v_m - 2v'_m + (v'_m)^2 v^*_m\big)_{m=-M}^{M}$; that is,

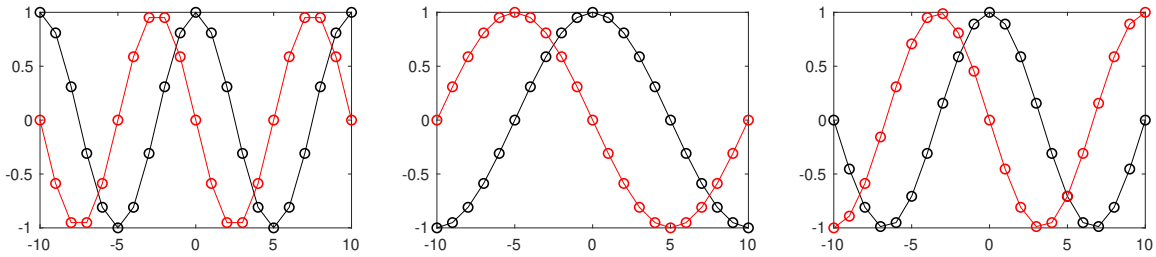
$$\begin{aligned}\tilde{d}_{\mathrm{a2},z'}(v)^2 := \min \big\{ &\eta(\mathbb{T}) \ : \ \eta \text{ is a positive measure, } \mathcal{F}\eta = w, \\ &\text{with } -4\pi^2 m^2 w_m = v_m - 2v'_m + (v'_m)^2 v^*_m, \ m = -M, \dots, M \big\}. \end{aligned} \tag{65}$$

It turns out that this function has a closed-form expression: let
$w = \big((v_m - 2v'_m + (v'_m)^2 v^*_m)/(-4\pi^2 m^2)\big)_{m=-M}^{M}$, with $w_0 = 0$, and $W = \mathrm{T}(w)$; then

$$\tilde{d}_{\mathrm{a2},z'}(v)^2 = \mathrm{i}^+(-W). \tag{66}$$

In Figure 6, we can see that when $v = \mathrm{a}(z)$ is an atom, $\tilde{d}_{\mathrm{a2},z'}(v)^2$ closely matches the squared distance $d(z, z')^2$ between the frequencies.

In Figure 7, we show a toy application of the proposed approximate atomic 2-Wasserstein distance: we compute the Wasserstein barycenter [57] $v \in \mathbb{V}$ of two atoms $\mathrm{a}(z)$ and $\mathrm{a}(z')$, with $z = \mathrm{e}^{\mathrm{j}2\pi/10}$, $z' = \mathrm{e}^{\mathrm{j}2\pi/20}$, $M = 10$. That is, v minimizes $\tilde{d}_{\mathrm{a2},z}(v)^2 + \tilde{d}_{\mathrm{a2},z'}(v)^2$. The overrelaxed Chambolle–Pock algorithm is used again to solve this convex optimization problem. The solution v turns out to be the atom $\mathrm{a}(\mathrm{e}^{\mathrm{j}3\pi/20})$, whose frequency is the mean of the frequencies of the two given atoms. So, we can interpolate between two sampled complex exponentials, by solving a convex optimization problem, without having to first identify their frequencies. This opens the door to many signal processing applications.



(a) (b) (c)

Figure 7: In (a) and (b): two atoms $\mathrm{a}(z) = \big(\mathrm{a}(z)_m\big)_{m=-M}^{M}$ and $\mathrm{a}(z') = \big(\mathrm{a}(z')_m\big)_{m=-M}^{M}$, respectively, with $z = \mathrm{e}^{\mathrm{j}2\pi/10}$, $z' = \mathrm{e}^{\mathrm{j}2\pi/20}$, $M = 10$; the index m is in abscissa, the real and imaginary parts of the values are in ordinate, with black and red colors, respectively. In (c): the atomic Wasserstein barycenter $v \in \mathbb{V}$, which minimizes $\tilde{d}_{\mathrm{a}2,z}(v)^2 + \tilde{d}_{\mathrm{a}2,z'}(v)^2$; it turns out that v is the atom $\mathrm{a}(\mathrm{e}^{\mathrm{j}3\pi/20})$, whose frequency is the mean of the frequencies of the two given atoms.

11 Conclusion

We studied the properties of decompositions of vectors into complex exponentials and we proposed convex formulations to implement these decompositions. Roughly speaking, the analysis is based on splitting the vectors (and associated Toeplitz matrices) into negative and positive (semidefinite) parts and applying the powerful characterization of Toeplitz matrices by Carathéodory to them. Besides, we have introduced functionals to perform optimal transport between two measures, given only a finite number of their Fourier coefficients. This makes it possible to deal with measures computationally, without the need to discretize them on a grid of predefined locations. Further theoretical investigations are needed and the relationship between the proposed construction and existing spectral transport approaches for audio signal processing [58, 59] should be investigated. In future work, the author will apply this framework to the approximation and reconstruction of signals and images [60–62].

References

[1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer-Verlag New York, 2010.

[2] S. Mallat and Z. Zhang, "Matching pursuit in a time–frequency dictionary," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, 1993.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[4] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[5] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, pp. 805–849, Dec. 2012.

[6] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[7] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[8] R. E. Curto and L. A. Fialkow, "Recursiveness, positivity, and truncated moment problems," *Houston J. Math*, vol. 17, no. 4, pp. 603–635, 1991.

[9] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jun. 2014.

[10] J.-M. Azaïs, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177–195, Mar. 2015.

[11] J. J. Benedetto and W. Li, "Super-resolution by means of Beurling minimal extrapolation," *Applied and Computational Harmonic Analysis*, vol. 48, no. 1, pp. 218–241, Jan. 2020.

[12] S. D. Fisher and J. W. Jerome, "Spline solutions to L1 extremal problems in one and several variables," *Journal of Approximation Theory*, vol. 13, pp. 73–83, Jan. 1975.

[13] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and Applications*, vol. 395, no. 1, pp. 336–354, Nov. 2012.

[14] M. Unser, J. Fageot, and J. P. Ward, "Splines are universal solutions of linear inverse problems with generalized TV regularization," *SIAM Review*, vol. 59, pp. 769–793, 2017.

[15] A. Flinth and P. Weiss, "Exact solutions of infinite dimensional total-variation regularized problems," *Information and Inference: A Journal of the IMA*, vol. 8, pp. 407–443, Sep. 2019.

[16] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. de Gournay, and P. Weiss, "On representer theorems and convex regularization," *SIAM Journal on Optimization*, vol. 29, pp. 1260–1281, 2019.

[17] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, NJ, 2005.

[18] V. Pereyra and G. Scherer, "Exponential data fitting," in *Exponential data fitting and its applications*, V. Pereyra and G. Scherer, Eds. Bentham Science Publ., 2010, pp. 1–26, chapter 1.

[19] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, no. 1, pp. 17–47, 2015.

[20] ——, "Super-resolution of positive spikes by Toeplitz low-rank approximation," in *Proc. of EUSIPCO*, Nice, France, Sep. 2015.

[21] F. Andersson, M. Carlsson, J.-Y. Tourneret, and H. Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5761–5774, Nov. 2014.

[22] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[23] B. Alexeev, J. Cahill, and D. G. Mixon, "Full spark frames," *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1167–1194, 2012.

[24] Y. Chi, A. Pezeshki, L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011.

[25] A. Fannjiang and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 179–202, 2012.

[26] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, Jul. 2013.

[27] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Proc. of IEEE ICASSP*, vol. 5, Mar. 2005, pp. 729–732.

[28] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 575–585, Dec. 2007.

[29] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imaging Sciences*, vol. 10, no. 3, pp. 1258–1290, 2017.

[30] T. Fisher, "Existence, uniqueness, and minimality of the Jordan measure decomposition," 2012, report arXiv:1206.5449v2.

[31] A. Satorra and H. Neudecker, "A theorem on the rank of a product of matrices with illustration of its use in goodness of fit testing," *Psychometrika*, vol. 79, 2014.

[32] G. M. Riche de Prony, "Essai expérimental et analytique : sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool à différentes températures," *Journal de l'École polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[33] J. R. Auton, "Investigation of procedures for automatic resonance extraction from noisy transient electromagnetics data," 1981, vol. III. Effects Technology Inc., Santa Barbara, CA. Report, translation of Prony's original paper and bibliography of Prony's method.

[34] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008, special issue on Compressive Sampling.

[35] D. Potts and M. Tasche, "Nonlinear approximation by sums of nonincreasing exponentials," *Applicable Analysis*, vol. 90, no. 3–4, pp. 609–626, 2011.

[36] T. Blu, "The generalized annihilation property—A tool for solving finite rate of innovation problems," in *Proc. of Int. Workshop on Sampling Theory and Appl. (SampTA)*, Marseille, France, May 2009.

[37] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, 2007.

[38] C. Carathéodory, "Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, no. 1, pp. 193–217, 1911.

[39] C. Carathéodory and L. Fejér, "Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landau'schen Satz," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, no. 1, pp. 218–239, 1911.

[40] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," May 2014, preprint arXiv:1405.6585v1.

[41] M. Bakonyi and E. V. Lopushanskaya, "Moment problems for real measures on the unit circle," in *Recent Advances in Operator Theory in Hilbert and Krein Spaces*, ser. Operator Theory: Advances and Applications, vol. 198. Birkhäuser Basel, 2010, pp. 49–60.

[42] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.

[43] Z. Yang and L. Xie, "Spectral compressed sensing of real spikes," in *Proc. of IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Oct. 2016.

[44] C. Fernandez-Granda, "Super-resolution of point sources via convex programming," *Information and Inference*, vol. 5, no. 3, pp. 251–303, 2016.

[45] M. Ferreira Da Costa and W. Dai, "A tight converse to the spectral resolution limit via convex programming," in *Proc. of IEEE International Symposium on Information Theory*, Jun. 2018.

[46] B. Hofmann, B. Kaltenbacher, C. Poeschl, and O. Scherzer, "A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators," *Inverse Problems*, vol. 23, pp. 987–1010, 2007.

[47] K. Bredies and H. K. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[48] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[49] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, "Proximal splitting algorithms: Relax them all!" 2019, preprint arXiv:1912.00137.

[50] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[51] C. Villani, *Topics in Optimal Transportation*, ser. Graduate studies in mathematics. American Mathematical Society, 2003.

[52] F. Santambrogio, *Optimal Transport for Applied Mathematicians*. Birkhauser, 2015.

[53] J. Rabin, G. Peyré, J. Delon, and M. Bernot, "Wasserstein barycenter and its application to texture mixing," in *Proc. of Scale Space and Variational Methods in Computer Vision (SSVM) 2011*, vol. Lecture Notes in Computer Science, vol 6667. Springer, Berlin, Heidelberg, 2012.

[54] J. Alaux, E. Grave, M. Cuturi, and A. Joulin, "Unsupervised hyper-alignment for multilingual word embeddings," in *Proc. of ICLR*, May 2019.

[55] J. Rabin, J. Delon, and Y. Gousseau, "Transportation distances on the circle," *Journal of Mathematical Imaging and Vision*, vol. 41, p. 147, Sep. 2011.

[56] C. A. Cabrelli and U. M. Molter, "The Kantorovich metric for probability measures on the circle," *Journal of Computational and Applied Mathematics*, vol. 57, no. 3, pp. 345–361, Feb. 1995.

[57] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM J. Math. Anal.*, vol. 43, no. 2, pp. 904–924, 2011.

[58] R. Flamary, C. Févotte, N. Courty, and V. Emiya, "Optimal spectral transportation with application to music transcription," in *Advances in Neural Information Processing Systems (Proc. of NIPS)*, 2016, pp. 703–711.

[59] A. Rolet, V. Seguy, M. Blondel, and H. Sawada, "Blind source separation with optimal transport non-negative matrix factorization," *EURASIP Journal on Advances in Signal Processing*, vol. 53, 2018.

[60] L. Condat, "A convex approach to K-means clustering and image segmentation," in *Proc. of EMMCVPR, M. Pelillo and E. Hancock eds., Lecture Notes in Computer Science vol. 10746, Springer, 2018*, Venice, Italy, 2017, pp. 220–234.

[61] K. Polisano, L. Condat, M. Clausel, and V. Perrier, "A convex approach to super-resolution and regularization of lines in images," *SIAM Journal on Imaging Sciences*, vol. 12, no. 1, pp. 211–258, 2019.

[62] L. Condat, D. Kitahara, and A. Hirabayashi, "A convex lifting approach to image phase unwrapping," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.