A convex approach to super-resolution and regularization of lines in images

Kévin Polisano, Laurent Condat, Marianne Clausel, Valérie Perrier

A CONVEX APPROACH TO SUPER-RESOLUTION AND REGULARIZATION OF LINES IN IMAGES*

KÉVIN POLISANO†, LAURENT CONDAT‡, MARIANNE CLAUSEL†, AND VALÉRIE PERRIER†

Abstract. We present a new convex formulation for the problem of recovering lines in degraded images. Following the recent paradigm of super-resolution, we formulate a dedicated atomic norm penalty and we solve this optimization problem by means of a primal–dual algorithm. This parsimonious model enables the reconstruction of lines from lowpass measurements, even in presence of a large amount of noise or blur. Furthermore, a Prony method performed on rows and columns of the restored image, provides a spectral estimation of the line parameters, with subpixel accuracy.

Key words. Super-resolution, sparse recovery, convex optimization, line detection, splitting method, spectral estimation

AMS subject classifications. 94A08, 94A12, 94A20, 47N10, 90C22, 90C25

1. Introduction.

Many restoration or reconstruction imaging problems are ill-posed and must be regularized. So, they can be formulated as convex optimization problems formed by the combination of a data fidelity term with a norm-based regularizer. Typically, given the data $\mathbf{y} = \mathbf{A}\mathbf{x}^\sharp + \boldsymbol{\epsilon}$, for some unknown image $\mathbf{x}^\sharp$ to estimate, known observation operator $\mathbf{A}$ and some noise $\boldsymbol{\epsilon}$, one aims at solving a problem like

$$\text{Find } \tilde{\mathbf{x}} \in \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda R(\mathbf{x}) \,, \tag{1}$$

where λ controls the tradeoff between data fidelity and regularization and R is a convex regularization functional. R can be chosen to promote some kind of smoothness. The classical Tikhonov regularizer $R(\mathbf{x}) = \|\nabla\mathbf{x}\|_2^2$ generally makes the problem easy to solve, but yields over-smoothing of the textures and edges in the recovered image $\tilde{\mathbf{x}}$. A popular and better regularizer is the total variation $R(\mathbf{x}) = \|\nabla\mathbf{x}\|_1$, see e.g [11]; it yields images with sharp edges, but the textures are still over-smoothed, there are staircasing effects and the pixel values tend to be clustered in piecewise constant areas. To overcome these drawbacks, one can penalize higher order derivatives [29] or make use of non-local penalties [37, 18, 15]. Another approach, which is at the heart of the recent paradigm of sparse recovery [56, 33] and compressed sensing [23, 51], is to choose R to favor some notion of low complexity. Indeed, many phenomena, when observed by instruments, yield data living in high dimensional spaces, but inherently governed by a small number of degrees of freedom. One early choice was to set R as the ℓ_1 norm of wavelet coefficients of the image. But the signals encountered in applications like radar, array processing, communication, seismology, or remote sensing, are usually specified by parameters in a *continuous* domain, from which they depend nonlinearly. So, modern sampling theory has widened its scope to a broader class of signals, with so-called finite rate of innovation, i.e. ruled by parsimonious models [31, 24, 5,

*Preprint

Funding: The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-13-BS03-0002-01 (ASTRES).

†Univ. Grenoble Alpes, Laboratoire Jean Kuntzmann, F-38000, Grenoble, France (kevin.polisano@imag.fr).

‡Univ. Grenoble Alpes, GIPSA-lab, F-38000, Grenoble, France (laurent.condat@gipsa-lab.grenoble-inp.fr).

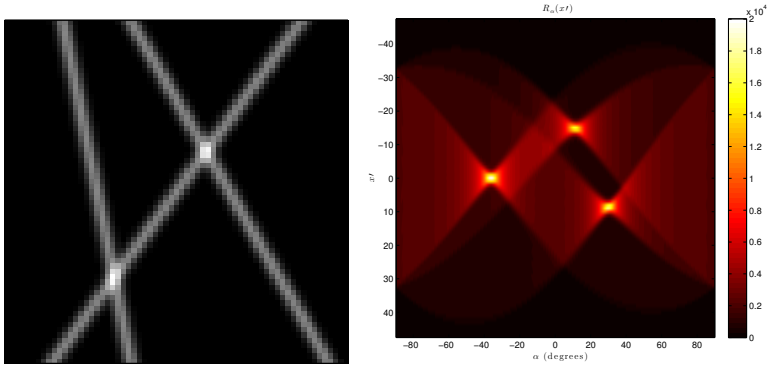
57]. This encompasses reconstruction of pulses from lowpass measurements [21] and spectral estimation, which is the reconstruction of sinusoids from point samples [48, 49], with many applications [13, 28, 55, 7, 50, 30, 40, 47, 20]. The knowledge of the kind of elements we want to promote in the image makes it possible to estimate them from coarse-scale measurements, even with infinite precision if there is no noise. Methods achieving this goal are qualified as *super-resolution* methods, because they uncover fine scale information, which was lost in the data, beyond the Rayleigh or Nyquist resolution limit of the acquisition system [25, 8]. However, in this context, maximum likelihood estimation amounts to structured low rank approximation, which forms nonconvex and very difficult, even NP-hard in general, problems [32]. An elegant and unifying formulation, which yields convex problems, is based on the *atomic norm* [3, 19]. We place ourselves in this general framework of atomic norm minimization: the sought-after image $\mathbf{x}^\sharp$ is supposed to be a sparse positive combination of the elements of an infinite dictionary $\mathcal{A}$, indexed by continuously varying parameters. Then, one can choose R as the atomic norm $\|\mathbf{x}\|_{\mathcal{A}}$ of the image $\mathbf{x}$, which can be viewed as the ℓ_1 norm of the coefficients, when the image is expressed in terms of the elements of $\mathcal{A}$, called *atoms* and and having unit norm:

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf\{t > 0 : \mathbf{x} \in t\,\mathrm{conv}(\mathcal{A})\}\ , \tag{2}$$

where $\mathrm{conv}(\mathcal{A})$ is the convex hull of the atoms. In this paper, we consider the setting, which is new to our knowledge, where the atoms are *lines*. Expressed in the Fourier domain, these atoms can be characterized with respect to their rows and columns, and the problem can be reduced to a dictionary of 1-D complex exponential samples, indexed by their frequency and phase, and the atomic norm can be computed via semidefinite programming [58]. This formulation enables us to derive a convex optimization problem under constraints, solved by mean of a splitting primal–dual algorithm [17]. Then, performing a Prony-like method [42] onto the solution of the algorithm allows us to extract the parameters of the lines. This approach provides a very high accuracy for the line estimation, where the Hough [59, 27, 34] and the Radon [41, 22, 35] transforms fail, due to their discrete nature. Our motivation stems from the frequent presence in biomedical images, e.g. in microscopy, of elongated structures like filaments, neurons, veins, which are deteriorated when reconstructed with classical penalties. For instance the super-resolution detection of lines, can be applied on an diffracted image of tubulins, which are curved structures, blurred due to the diffraction through the microscope, and behaving locally like straight lines. Thus one could apply the method set out in this paper on patches, to beat the diffraction and find their exact positions.

The paper is organized as follows. The model is exposed in section 2, the framework of atomic norm minimization underlying the super-resolution principle is introduced in section 3, the algorithms we derive are in section 4. Then a Prony-like method is developed in section 5 as a way to perform spectral estimation of the line parameters. Finally some experimental results are shown in section 6 with comparison with the classic Hough and Radon procedure of lines detection. Part of this work has been published in the preliminary paper [38]. In this paper, we present another algorithm, a new procedure estimation of the line parameters, an extension for the whole range of line angles with no more restriction, an application to inpainting problems and many other numerical experiments.

2. An image model of blurred lines. Our aim is to restore a blurred image $\mathbf{b}^\sharp$ containing lines, and to estimate the parameters—angle, offset, amplitude—of the



(a) (b)

FIG. 1. *(a) The image* $\mathbf{b}^\sharp$ *of three blurred lines with* $\kappa = 1$ *and (b) the Radon transform of* $\mathbf{b}^\sharp$.

lines, given degraded data $\mathbf{y}$. In this section, we formulate what we precisely mean by an image containing lines. In short, $\mathbf{b}^\sharp$ is a sum of perfect lines which have been blurred and then sampled. Both processes are detailed in the following.

2.1. The Ideal Continuous Model and the Objectives. We place ourselves in the quotient space $\mathbb{P} = \mathbb{R}/(W\mathbb{Z}) \times \mathbb{R}$, corresponding to the 2-D plane with horizontal $W-$periodicity, for some integer $W \geq 1$. To simplify the notations, we suppose that W is odd and we set $M = (W-1)/2$.

A line of infinite length, with angle $\theta \in (-\pi/2, \pi/2]$ with respect to verticality, amplitude $\alpha > 0$, and offset $\eta \in \mathbb{R}$ from the origin on x–axis, is defined as the distribution

$$(t_1, t_2) \in \mathbb{P} \mapsto \alpha\delta\big(\cos\theta(t_1 - \eta) + \sin\theta\, t_2\big)\,, \tag{3}$$

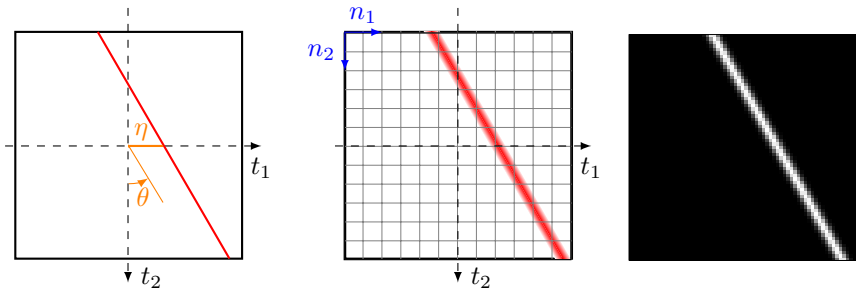
where δ is the Dirac distribution. We define the distribution $x^\sharp$, which is a sum of K different such perfect lines, for some integer $K \geq 1$, as

$$x^\sharp : (t_1, t_2) \in \mathbb{P} \mapsto \sum_{k=1}^{K} \alpha_k \delta\big(\cos\theta_k(t_1 - \eta_k) + \sin\theta_k\, t_2\big)\,. \tag{4}$$

At this time, we suppose that the lines are rather vertical; that is, for every $k = 1, \ldots, K$, $\theta_k \in (-\pi/4, \pi/4]$. We will treat the general case in subsection 4.3. As illustrated in Figure 2, we adopt the following representation: the horizontal (resp. vertical) axis corresponds to t_2 (resp. t_1) fixed. Since the ideal model $x^\sharp$ is made up of Diracs, the horizontal Fourier transform $\hat{x}^\sharp = \mathcal{F}_1 x^\sharp$, is composed of a sum of exponentials. Our goal will be to reconstruct $\hat{x}^\sharp$ by a super-resolution method, from its observations through a known degradation operator $\mathbf{A}$ and some noise, which is an ill-posed problem. Then, spectral estimation of these exponentials will allow us to recover the line parameters. Let us first characterize the blur operator $\mathbf{A}$.

2.2. A Blur Model for an Exact Sampling Process. The image observed $\mathbf{b}^\sharp$ of size $W \times H$ is obtained by the convolution of the distribution $x^\sharp$ with a blur function ϕ, followed by a sampling with unit step denoted by the operator Δ:

$$\mathbf{b}^\sharp[n_1, n_2] = (x^\sharp * \phi)(n_1, n_2), \quad \forall n_1 = 0, \ldots, W-1, \quad n_2 = 0, \ldots, H-1\,, \tag{5}$$



(a) (b) (c)

FIG. 2. *(a) Parameters (θ, η) characterizing the position of a line in the 2-D plane, (b) the matrix convention we use to display the image obtained by applying the sampling operator with unit step Δ onto the blurred line $x^\sharp * \phi$, and (c) the resulting discrete image $\mathbf{b}^\sharp[n_1, n_2] = (x^\sharp * \phi)(n_1, n_2)$.*

We also consider that the point spread function ϕ is separable; that is the function $x^\sharp * \phi$ can be obtained by a first horizontal convolution with φ_1 and then a second vertical convolution with φ_2. Formally, $x^\sharp * \phi = (x^\sharp * \phi_1) * \phi_2$ with $\phi_1(t_1, t_2) = \varphi_1(t_1)\delta(t_2)$ and $\phi_2(t_1, t_2) = \delta(t_1)\varphi_2(t_2)$.

In order to avoid any approximation when passing from the continuous to the discrete formulation, we assume that ϕ has the following properties:

- the function $\varphi_1 \in L^1(0, W)$ is W–periodic, bounded, such that $\int_0^W \varphi_1 = 1$, and bandlimited; that is, its Fourier coefficients $\frac{1}{W}\int_0^W \varphi_1(t_1)\mathrm{e}^{-\mathrm{j}2\pi m t_1/W}\,\mathrm{d}t_1$ are zero for every $m \in \mathbb{Z}$ with $|m| \geq (W+1)/2 = M + 1$. The discrete filter $\boldsymbol{g}[n] = \varphi_1(n)$, with these assumptions on φ_1, has discrete Fourier coefficients which correspond to Fourier coefficients of φ_1.
- $\varphi_2 \in L^1(\mathbb{R})$, with $\int_{\mathbb{R}} \varphi_2 = 1$. In addition, the discrete filter $\big(\boldsymbol{h}[n] = (\varphi_2 * \mathrm{sinc})(n)\big)_{n\in\mathbb{Z}}$, where $\mathrm{sinc}(t_2) = \sin(\pi t_2)/(\pi t_2)$, has compact support of length $2S+1$, for some $S \in \mathbb{N}$, i.e. $\boldsymbol{h}[n] = 0$ if $|n| \geq S+1$. Note that this assumption is not restrictive, and that if φ_2 is bandlimited, we simply have $\boldsymbol{h}[n] = \varphi_2(n)$.

So, after the first horizontal convolution, using the fact that $\delta(at) = \delta(t)/|a|$ for any $a \neq 0$, we obtain the function

$$\Psi = x^\sharp * \phi_1 : (t_1, t_2) \mapsto \sum_{k=1}^{K} \frac{\alpha_k}{\cos\theta_k} \varphi_1\Big(t_1 + \tan\theta_k\, t_2 + \eta_k\Big)\,. \tag{6}$$

We can show that, after the second vertical convolution, we get the function

$$x^\sharp * \phi : (t_1, t_2) \in \mathbb{P} \mapsto \sum_{k=1}^{K} \alpha_k \psi_k\big(\cos(\theta_k)(t_1 - \eta_k) + \sin(\theta_k) t_2\big)\,, \tag{7}$$

where

$$\psi_k = \left(\frac{1}{\cos(\theta_k)}\varphi_1\Big(\frac{\cdot}{\cos(\theta_k)}\Big)\right) * \left(\frac{1}{\sin(\theta_k)}\varphi_2\Big(\frac{\cdot}{\sin(\theta_k)}\Big)\right)\,, \tag{8}$$

if $\theta_k \neq 0$ and $\psi_k = \varphi_1$ else.

We can notice that (7) can also be interpreted as follows: every line has undergone a 1-D convolution with ψ_k in the direction transverse to it. We can also notice that

if φ_1 and φ_2 are Gaussian and have same variance κ^2, it follows from (8) that ψ_k has variance $\kappa^2\big(\cos(\theta)^2+\sin(\theta)^2\big)=\kappa^2$ as well.

PROPOSITION 1 (Nyquist-Whittaker-Shannon). *The function φ_1, which is periodic and bandlimited, is determined by W degrees of freedom only. That is, with the coefficients $\boldsymbol{g}[n]=\varphi_1(n)$, $n=0,\dots,W-1$, the function φ_1 is a linear combination of shifted Dirichlet kernels $D_M(t)=\sum_{m=-M}^{M}\mathrm{e}^{\mathrm{j}mt}$:*

$$\varphi_1(t)=\frac{1}{W}\sum_{n=0}^{W-1}\boldsymbol{g}[n]D_M\left(\frac{2\pi(t-n)}{W}\right),\quad\forall t_1\in\mathbb{R}\,. \tag{9}$$

Proof. Given in Appendix A. ∎

Since the Dirichlet kernel is a normalized function, a change of variable leads to $\frac{1}{W}\int_0^W D_M\left(\frac{2\pi(t-n)}{W}\right)\mathrm{d}t=1$ and then

$$\int_0^W\varphi_1(t)\,\mathrm{d}t=\sum_{n=-M}^{M}\boldsymbol{g}[n]\,, \tag{10}$$

that is, the filter $\boldsymbol{g}$ has to be normalized as well.
We can easily compute that $\frac{1}{W^2}\int_0^W D_M\left(\frac{2\pi(t-n)}{W}\right)D_M\left(\frac{2\pi(t-m)}{W}\right)\mathrm{d}t=\delta_{m,n}$, which proves that

$$\int_0^W\varphi_1(t)^2\,\mathrm{d}t=\sum_{n=-M}^{M}\boldsymbol{g}[n]^2=\sum_{n=-M}^{M}\varphi_1(n)^2\,, \tag{11}$$

and by Parseval

$$\int_{\mathbb{R}}|\hat{\varphi}_1(\xi)|^2\,\mathrm{d}\xi=\sum_{m\in\mathbb{Z}}|c_m(\varphi_1)|^2=\sum_{m=-M}^{M}|\hat{\boldsymbol{g}}[m]|^2=\frac{1}{W}\int_0^W|\varphi_1(t)|^2\,\mathrm{d}t\,. \tag{12}$$

Now, for every $k=1,\dots,K$, the assumption $\theta_k\in(-\pi/4,\pi/4]$ yields $|\tan(\theta_k)|\leq 1$. So, the function $\Psi=x^\sharp*\phi_1$ given in (6), as a function of t_2 at fixed t_1, is bandlimited: for every $t_1\in[0,W[$, the Fourier transform $\mathcal{F}_2\Psi:\omega_2\mapsto\int_{\mathbb{R}}(x^\sharp*\phi_1)(t_1,t_2)\mathrm{e}^{-\mathrm{j}\omega_2 t_2}\,\mathrm{d}t_2$, which is a distribution (sum of K Dirac combs), is zero for every $|\omega_2|\geq\pi$. Indeed, we have

$$[\mathcal{F}_2\Psi](\omega_2)=\sum_{k=1}^{K}\frac{\alpha_k}{\cos\theta_k}\hat{\varphi}_1\left(\frac{\omega_2}{\tan\theta_k}\right)\exp\Big(\mathrm{j}2\pi(t_1+\eta_k)\omega_2\Big)\,, \tag{13}$$

and by considering $\theta^\star$ the maximum of $|\tan\theta_k|$ achieved on $(-\pi/4,\pi/4]$, we obtain

$$1\leqslant\left|\frac{1}{\tan\theta^\star}\right|\leqslant\left|\frac{1}{\tan\theta_k}\right|\quad\Rightarrow\quad|\omega_2|\leqslant\left|\frac{\omega_2}{\tan\theta^\star}\right|\leqslant\left|\frac{\omega_2}{\tan\theta_k}\right|\,. \tag{14}$$

Consequently, since φ_1 is bandlimited; that is, $\hat{\varphi}_1(\omega)=0$ for $|\omega|\geqslant\pi$, we also have $[\mathcal{F}_2\Psi](\omega_2)=0$ for $|\omega_2|\geqslant\pi$. Then, $\mathcal{F}_2\Psi=\mathcal{F}_2\Psi\cdot\Pi_{[-\pi,\pi]}\Leftrightarrow\Psi=\Psi*\mathrm{sinc}$, and furthermore $\Psi*\varphi_2=\Psi*(\varphi_2*\mathrm{sinc})$. In the Fourier domain, the function $h=\varphi_2*\mathrm{sinc}$ is bandlimited, so $[\mathcal{F}_2\Psi]\hat{h}=[\mathcal{F}_2\Psi]\hat{h}_{\mathrm{per}}$ where $\hat{h}_{\mathrm{per}}$ corresponds to the periodization of

the spectral $\hat{h}$, which amounts to saying that $\Psi * h = \Psi * (\sum_n h[n]\delta(\cdot - n))$. Hence, it is equivalent to perform the vertical convolution of $x^\sharp * \phi_1$ with φ_2, with $\varphi_2 * \mathrm{sinc}$, or with the Dirac comb $\gamma : t_2 \mapsto \sum_{n=-S}^{S} \boldsymbol{h}[n]\delta(t_2 - n)$, where $\boldsymbol{h}[n] = (\varphi_2 * \mathrm{sinc})(n)$. We have made the assumption that the filter $(\boldsymbol{h}[n])_n$ has a compact support, but notice that the function $h = \varphi_2 * \mathrm{sinc}$ does not have a compact support, since it is bandlimited, which means that the continuous function h has to vanish at the integer points $t = n$ for $|n| > S$. Given such a compact filter $(\boldsymbol{h}[n])_{n=-S}^{S}$, the unique bandlimited function h satisfying these conditions is obtained by the Shannon interpolation formula:

$$h(t) = \sum_{n=-S}^{S} \boldsymbol{h}[n] \,\mathrm{sinc}(t - n) \;. \tag{15}$$

By unicity, we necessarily have $\varphi_2 * \mathrm{sinc} = h$, and we can notice that there is always exists a bandlimited solution φ_2 of this equation, which is simply $\varphi_2 = h$; that is why we argued that the compact support assumption is not restrictive. Of course, if we are looking for φ_2 in a wider class function, there does not exist a general method to reconstruct φ_2 from the samples of the low-resolution signal $\varphi_2 * \mathrm{sinc}$. It is typically what the super-resolution enables in the specific case where φ_2 is suppose to be a spike train.

So, let us define $\mathbf{u}^\sharp$ by sampling $x^\sharp * \phi_1$ with unit step

$$\mathbf{u}^\sharp[n_1, n_2] = (x^\sharp * \phi_1)(n_1, n_2), \quad \forall n_1 = 0, \dots, W-1, \quad n_2 = -S, \dots, H-1+S \;, \tag{16}$$

With the above assumptions, we can express $\mathbf{b}^\sharp$ from $\mathbf{u}^\sharp$ using a discrete vertical convolution with the filter $\boldsymbol{h}$:

$$\mathbf{b}^\sharp[n_1, n_2] = \sum_{p=-S}^{S} \mathbf{u}^\sharp[n_1, n_2 - p]\boldsymbol{h}[p], \quad \forall n_1 = 0, \dots, W-1, \quad n_2 = 0, \dots, H-1 \;. \tag{17}$$

Altogether, we completely and exactly characterized the sampling process, which involves a continuous blur, using the discrete and finite filters $(\boldsymbol{g}[n])_{n=0}^{W-1}$ and $(\boldsymbol{h}[n])_{n=-S}^{S}$. We insist on the fact that no discrete approximation is made during this sampling process, due to the assumptions. An example of three blurred lines is depicted in Figure 1, with the normalized filter $\boldsymbol{h}$ approximating a Gaussian function of variance κ; that is, $t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$, on the compact set $[-S, S]$ with $S = \lceil 4\kappa \rceil - 1$; and the normalized filter $\boldsymbol{g} = [\mathbf{0}_{M-S}, \boldsymbol{h}, \mathbf{0}_{M-S}]$ whose DFT is an interpolation of $\hat{h}$, which approach the continous Fourier transform $\widehat{\mathcal{G}} : \nu \mapsto \mathrm{e}^{-2\pi^2\sigma^2\nu^2}$. So, note that $\|\hat{\boldsymbol{g}}\|_\infty = \|\boldsymbol{h}\|_\infty = 1$.

2.3. Toward an Inverse Problem in Fourier Domain. Let us further characterize the image $\mathbf{b}^\sharp$ in Fourier domain. We consider the image $\hat{\mathbf{u}}^\sharp$ obtained by applying the 1-D Discrete Fourier Transform (DFT) on every row of $\mathbf{u}^\sharp$:

$$\hat{\mathbf{u}}^\sharp[m, n_2] = \frac{1}{W} \sum_{n_1=0}^{W-1} \mathbf{u}^\sharp[n_1, n_2] \mathrm{e}^{-\mathrm{j}\frac{2\pi m}{W} n_1}, \quad \forall m = -M, \dots, M, \quad n_2 \in \mathbb{Z} \;. \tag{18}$$

which coincide with the exact Fourier coefficients of the function $t \mapsto (x^\sharp * \phi_1)(t, n_2)$. Hence, from (6) and computation of $\hat{\mathbf{u}}^\sharp[m, n_2] = \frac{1}{W} \int_0^W (x^\sharp * \phi_1)(t, n_2) \mathrm{e}^{-\mathrm{j}\frac{2\pi m}{W} t} \,\mathrm{d}t$, we obtain

$$\hat{\mathbf{u}}^\sharp[m, n_2] = \hat{\boldsymbol{g}}[m] \hat{\mathbf{x}}^\sharp[m, n_2], \quad \forall m = -M, \dots, M, \quad n_2 \in \mathbb{Z} \;, \tag{19}$$

$$\hat{\mathbf{x}}^\sharp[m,n_2] = \sum_{k=1}^{K} \frac{\alpha_k}{\cos\theta_k} \mathrm{e}^{\mathrm{j}2\pi(\tan\theta_k\, n_2+\eta_k)\frac{m}{W}} \ . \tag{20}$$

Applying a 1-D DFT on the first component of $\mathbf{b}^\sharp[n_1,n_2] = \mathbf{u}^\sharp[n_1,:] * \boldsymbol{h}$, leads to the elements $\hat{\mathbf{b}}^\sharp[m,n_2] = \hat{\mathbf{u}}^\sharp[m,:] * \boldsymbol{h}$. Since the image $\mathbf{u}^\sharp$ is real, then $\hat{\mathbf{x}}^\sharp$ is Hermitian, so we can only deal with the right part $\hat{\mathbf{x}}^\sharp[0:M,:]$ and notice that the column corresponding to $m=0$ is real and equal to $\sum_{k=1}^{K} \frac{\alpha_k}{\cos\theta_k}$. We consider in the following the image $\hat{\mathbf{x}}^\sharp[m,n_2]$ of size $(M+1)\times H_S$, with $H_S = H+2S$, due to the addition of S pixels beyond the borders for the convolution by the filter $\boldsymbol{h}$. More precisely, $\hat{\mathbf{x}}^\sharp \in \mathcal{X}$, where $\mathcal{X} = \{\hat{\mathbf{x}} \in \mathcal{M}_{M+1,H_S}(\mathbb{C}) : \mathrm{Im}(\hat{\mathbf{x}}[0,:]) = 0\}$, endowed with the following inner product, and $\cdot^*$ is complex conjugation:

$$\langle \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle = \sum_{n_2=0}^{H_S-1} \hat{\mathbf{x}}_1[0,n_2]\hat{\mathbf{x}}_2[0,n_2]^* + 2\mathrm{Re}\left(\sum_{m=1}^{M}\sum_{n_2=0}^{H_S-1} \hat{\mathbf{x}}_1[m,n_2]\hat{\mathbf{x}}_2[m,n_2]^* \right) . \tag{21}$$

Let $\mathbf{A}$ denote the operator which multiplies each row vector $\hat{\mathbf{x}}^\sharp[m,:]$ by the corresponding Fourier coefficient $\hat{\boldsymbol{g}}[m]$ and convolves it with the filter $\boldsymbol{h}$. Thus, we have $\mathbf{A}\hat{\mathbf{x}}^\sharp = \hat{\mathbf{b}}^\sharp$. The image $\mathbf{b}^\sharp$ of the blurred lines is affected by some noise ε, so that we observe the degraded image $\mathbf{y} = \mathbf{b}^\sharp + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\zeta^2)$ and ζ is the noise level. Our notations are explained in more details in Figure 3, illustrating the relation between all continuous and discrete variables. The problem is revert to an inverse problem we need to solve. In absence of noise, one can find a theoritical solution of this ill-posed problem, by solving on each column m the following reduced problems: $\hat{\boldsymbol{g}}[m]\check{\mathbf{H}}\hat{\mathbf{x}}[m,:] = \hat{\mathbf{b}}^\sharp[m,:]$ where $\check{\mathbf{H}}$ is the convolution matrix (54) corresponding to the vertical convolution by filter $\boldsymbol{h}$. In order to treat the noisy case, we will need to derive an optimization problem of this inverse problem, under constraints which will exploit the sparse structure of the signal we are looking for, namely a combination of lines. In both cases, the super-resolution consists to recover the high frequencies lost because of the blur operator, so it can be viewed as a spectral extrapolation. Our goal is to recover the parameters $(\theta_k,\eta_k,\alpha_k)$ of these lines from the degraded image $\mathbf{y}$, first by solving the optimization problem we derive by a primal–dual algorithm in the Fourier domain; that is, to go back on the bottom line of the diagram of Figure 3 from $\hat{\mathbf{y}}$ to $\hat{\mathbf{x}}^\sharp$, and then by performing a Prony-like method to estimate the parameters involving in $\hat{\mathbf{x}}^\sharp$. These two steps are summarized in Figure 4. Note that this work also covers the case where a mask is applied, in which we have no observation; that is, can encompass inpainting problems. In the next section, we present the framework of atomic norm from which will be derived the optimization problem.

3. Super-Resolution Detection of Lines.

3.1. Atomic Norm and Semidefinite Characterizations. Consider a complex signal $\boldsymbol{z} \in \mathbb{C}^N$ represented as a K–sparse mixture of atoms from the set

$$\mathcal{A} = \left\{ \boldsymbol{a}(\omega) \in \mathbb{C}^N \ : \ \omega \in \Omega \right\} \ , \tag{22}$$

that is,

$$\boldsymbol{z} = \sum_{k=1}^{K} c_k \boldsymbol{a}(\omega_k), \quad c_k \in \mathbb{C}, \quad \omega_k \in \Omega \ . \tag{23}$$

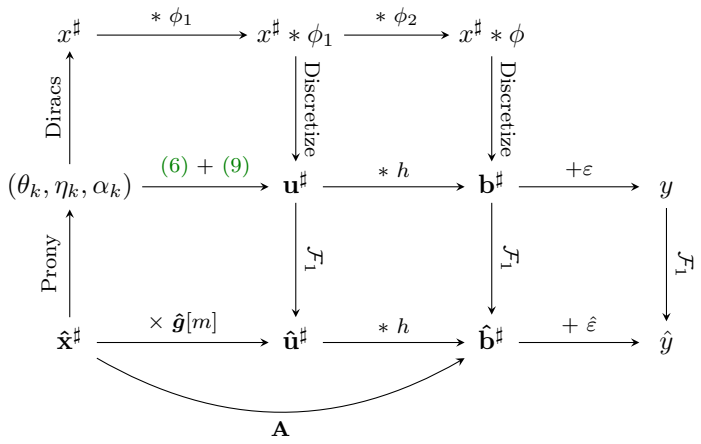


FIG. 3. *Relation between variables.*

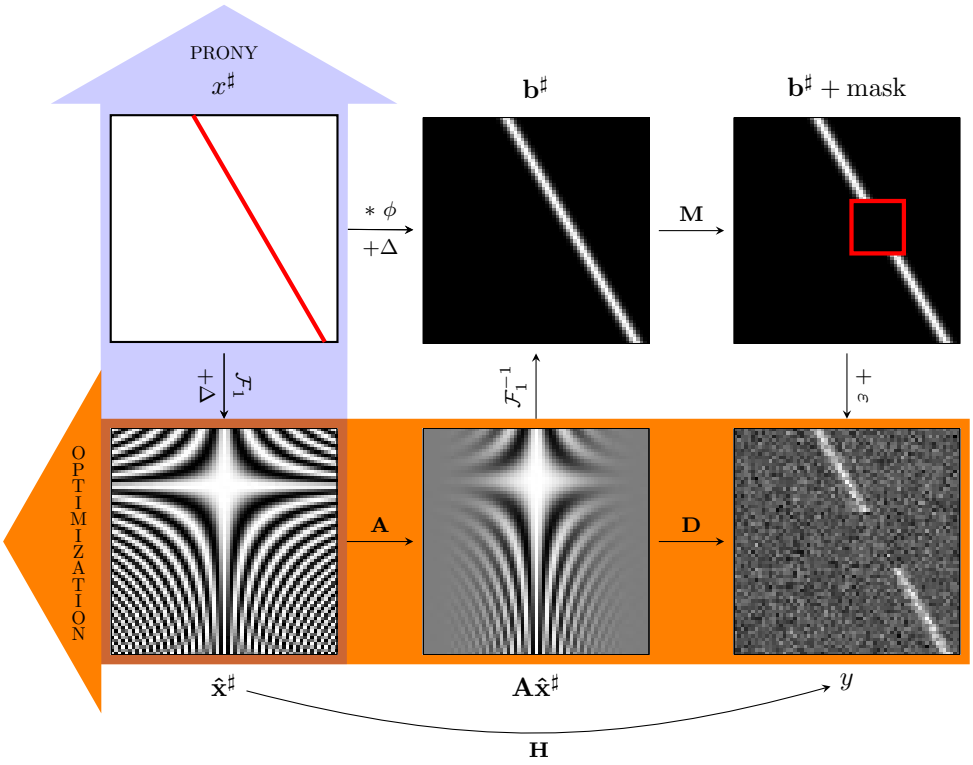


FIG. 4. *Relation between images.*

We consider atoms $\boldsymbol{a}(\omega) \in \mathbb{C}^N$ that are continuously indexed in the dictionary $\mathcal{A}$ by the parameter ω in a compact set Ω. The atomic norm, first introduced in [14], is defined as

$$\|\boldsymbol{z}\|_{\mathcal{A}} = \inf_{c'_k, \omega'_k} \left\{ \sum_k |c'_k| : \boldsymbol{z} = \sum_k c'_k \boldsymbol{a}(\omega'_k) \right\} , \tag{24}$$

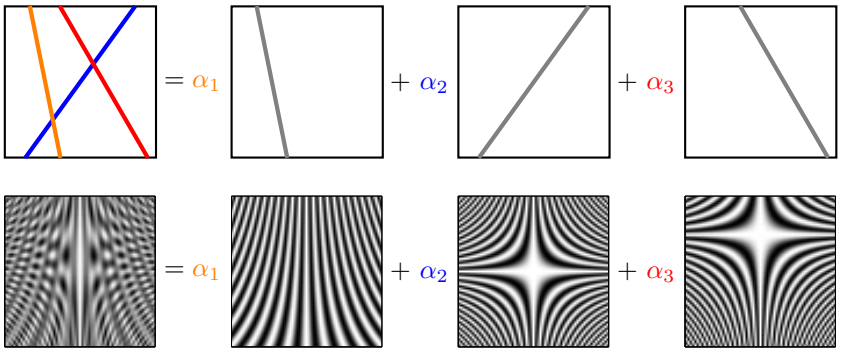


FIG. 5. *Illustration with a signal made of a weighted combination of three lines atoms (in gray) $\boldsymbol{a}(\theta_k, \eta_k)$. In the Fourier domain, we have the same kind of combination but with 2-D exponentials atoms $\hat{\boldsymbol{a}}(\theta_k, \eta_k)$. In both cases the sum of the weights $\alpha_1 + \alpha_2 + \alpha_3$, where α_i are the amplitudes of the lines, corresponds to the atomic norm on the appropriate dictionary.*

enforcing sparsity with respect to a general atomic set $\mathcal{A}$. Chandrasekaran *et al.* [14] argue that the atomic norm is the best convex heuristic for underdetermined, structured linear inverse problems, and it generalizes the norm for sparse recovery and the nuclear norm for low-rank matrix completion.

In our problem, the atoms are *lines*, so one can considered the dictionary $\mathcal{A}_{2D}$ indexed by the angle and the offset; that is, composed by the line atoms $\boldsymbol{a}(\theta_k, \eta_k) = \delta\big(\cos(\theta_k)(t_1 - \eta_k) + \sin(\theta_k)t_2\big)$. Or alternatively in the Fourier domain, the dictionary $\hat{\mathcal{A}}_{2D}$, composed by the 2-D exponentials atoms $\hat{\boldsymbol{a}}(\theta_k, \eta_k) = \frac{1}{\cos\theta_k} \mathrm{e}^{\mathrm{j}2\pi(\tan(\theta_k)n_2 - \eta_k)m/W}$ as illustrated in Figure 5. The problem is that the atomic norm computation in these 2-D dictionaries does not correspond to any known procedure to our knowledge. However, in the case of 1-D complex exponentials, there is a way to compute the atomic norm via semidefinite programming. So, the trick is to reformulate the problem using the simplified 1-D case. From now on, we consider the dictionary

$$\mathcal{A} = \left\{ \boldsymbol{a}(f, \phi) \in \mathbb{C}^{|I|}, \quad f \in [0, 1], \quad \phi \in [0, 2\pi) \right\}, \tag{25}$$

in which the *atoms* are the vectors of components $[\boldsymbol{a}(f, \phi)]_i = \mathrm{e}^{\mathrm{j}(2\pi f i + \phi)}$, $i \in I$, and simply $[\boldsymbol{a}(f)]_i = \mathrm{e}^{\mathrm{j}2\pi f i}$, $i \in I$, if $\phi = 0$. The atomic norm writes:

$$\|\boldsymbol{z}\|_{\mathcal{A}} = \inf_{\substack{c'_k \geqslant 0 \\ f'_k \in [0,1] \\ \phi'_k \in [0, 2\pi)}} \left\{ \sum_k c'_k : \boldsymbol{z} = \sum_k c'_k \boldsymbol{a}(f'_k, \phi'_k) \right\}. \tag{26}$$

THEOREM 2 (Caratheodory). *A vector $\boldsymbol{z} = (z^*_{N-1}, \ldots, z^*_1, z_0, z_1, \ldots, z_{N-1})$ of length $2N + 1$, with $z_0 \in \mathbb{R}$, is a positive combination of $K \leqslant N + 1$ atoms $\boldsymbol{a}(f_k)$ if and only if $\mathbf{T}_N(\boldsymbol{z}_+) \succcurlyeq 0$ and is of rank K, where $\boldsymbol{z}_+ = (z_0, \ldots, z_{N-1})$ is of length N and $\mathbf{T}_N : \mathbb{C}^N \to \mathcal{M}_N(\mathbb{C})$ is the Toeplitz operator*

$$\mathbf{T}_N : \boldsymbol{z}_+ = (z_0, \ldots, z_{N-1}) \mapsto \begin{pmatrix} z_0 & z_1^* & \cdots & z_{N-1}^* \\ z_1 & z_0 & \cdots & z_{N-2}^* \\ \vdots & \vdots & \ddots & \vdots \\ z_{N-1} & z_{N-2} & \cdots & z_0 \end{pmatrix}, \tag{27}$$

and $\succcurlyeq 0$ *denotes positive semidefiniteness. Moreover, this decomposition is unique, if* $K \leqslant N$.

Proof. See references [9, 10, 54] and thesis of the first author. ∎

We also have the above result improved from [52, Proposition II.1]:

PROPOSITION 3. *The atomic norm* $\|\boldsymbol{z}\|_{\mathcal{A}}$ *can be characterized by the following semidefinite program* $\mathrm{SDP}(\boldsymbol{z})$*:*

$$\|\boldsymbol{z}\|_{\mathcal{A}} = \min_{\boldsymbol{q}\in\mathbb{C}^N} \left\{ q_0 : \mathbf{T}'_N(\boldsymbol{z},\boldsymbol{q}) = \begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix} \succcurlyeq 0 \right\} . \tag{28}$$

with $\mathbf{T}'_N : \mathbb{C}^{2N} \to \mathcal{M}_{N+1}(\mathbb{C})$.

Proof. Given in Appendix B. ∎

Notice that the matrix $\mathbf{T}'_N(\boldsymbol{z},\boldsymbol{q})$ is Hermitian; that is, its eigenvalues $(\lambda_i)_{0\leqslant i\leqslant N}$ are positive reals, since the matrix is positive semidefinite. Consequently, we get $q_0 = \frac{1}{N+1}\mathrm{Trace}(\mathbf{T}'_N(\boldsymbol{z},\boldsymbol{q})) = \frac{1}{N+1}\sum_{i=0}^{N}\lambda_i$, which is real and positive.

We also define, for the upcoming convex optimization problem, the following set of complex matrices $\mathcal{Q} = \{\boldsymbol{q} \in \mathcal{M}_{M+1,H_S}(\mathbb{C}) : \mathrm{Im}(\boldsymbol{q}[:,0]) = 0\}$, endowed with the inner product defined in (21).

3.2. Properties of the Model $\hat{\mathbf{x}}^\sharp$ with respect to the Atomic Norm . From (20), the rows $\boldsymbol{l}_{n_2}$ (resp. columns $\boldsymbol{t}_m$) of the matrix $\hat{\mathbf{x}}^\sharp$, with $I = \{0,\dots,H_S-1\}$ (resp. $I = \{-M,\dots,M\}$), can be viewed as a sum of atoms

$$\boldsymbol{l}^\sharp_{n_2} = \hat{\mathbf{x}}^\sharp[:,n_2] = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_{n_2,k}) , \tag{29a}$$

$$\left(\text{resp. } \boldsymbol{t}^\sharp_m = \hat{\mathbf{x}}^\sharp[m,:] = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_{m,k},\phi_{m,k})^\mathsf{T} \right) , \tag{29b}$$

with

$$c_k = \frac{\alpha_k}{\cos\theta_k}, \quad f_{n_2,k} = \frac{\tan\theta_k\, n_2 + \eta_k}{W} , \tag{30}$$
$$\phi_{m,k} = \frac{2\pi\eta_k m}{W}, \quad f_{m,k} = \frac{\tan\theta_k\, m}{W} .$$

We define for later use, the frequency $\nu_k = \eta_k/W$ and the coefficients $d_{m,k} = c_k \mathrm{e}^{\mathrm{j}\phi_{m,k}}$, $e_{m,k} = \mathrm{e}^{\mathrm{j}\phi_{m,k}}$. The vectors $\boldsymbol{l}^\sharp_{n_2}$ of size $W = 2M+1$ are positive combinations of K atoms $\boldsymbol{a}(f_{n_2,k})$, with $K \leqslant M$ since we can reasonably assume that the number of lines K is smaller than half the number of pixels M. Thus, Theorem 2 ensures that the decomposition (29a) is unique, hence

$$\|\boldsymbol{l}^\sharp_{n_2}\|_{\mathcal{A}} = \sum_{k=1}^{K} c_k = \hat{\mathbf{x}}^\sharp[0,n_2], \quad \forall n_2 = 0,\dots,H_S-1 , \tag{31}$$

whereas, since the $d_{m,k}$ are complex in (29b), Theorem 2 no longer holds, we simply have from Proposition 3:

$$\|\boldsymbol{t}^\sharp_m\|_{\mathcal{A}} = \mathrm{SDP}(\boldsymbol{t}^\sharp_m) \leqslant \sum_{k=1}^{K} c_k, \quad \forall m = -M,\dots,M . \tag{32}$$

Let's take a closer look at one line; that is, $K = 1$, characterized by parameters (θ, η, α). We recall by (20) that $\hat{\mathbf{x}}^\sharp$ can be written as:

$$\hat{\mathbf{x}}^\sharp[m, n_2] = c_0 \mathrm{e}^{\mathrm{j}2\pi((f_1-f_0)n_2+f_0)m}, \quad c_0 = \frac{\alpha}{\cos\theta}, \quad f_0 = \frac{\eta}{W}, \quad f_1 = \frac{\tan\theta + \eta}{W} \ . \tag{33}$$

Let $\boldsymbol{z} = (z_0, \ldots, z_{|I|-1})$ be a complex vector, whose elements z_i are rearrange in a Toeplitz matrix $\mathbf{P}_K(\boldsymbol{z})$ of size $(|I| - K) \times (K + 1)$ and rank K as follows

$$\mathbf{P}_K(\boldsymbol{z}) = \begin{pmatrix} z_K & \cdots & z_0 \\ \vdots & \ddots & \vdots \\ z_{|I|-1} & \cdots & z_{|I|-K-1} \end{pmatrix} \ . \tag{34}$$

We get the following characterization of one line in Fourier:

PROPOSITION 4. *An image $\hat{\mathbf{x}}$ is of the form $\hat{\mathbf{x}}[m, n] = c_0 \mathrm{e}^{\mathrm{j}2\pi((f_1-f_0)n+f_0)m)}$ if and only if rows $\boldsymbol{l}_n$ and columns $\boldsymbol{t}_m$ of $\hat{\mathbf{x}}$ satisfy $\mathbf{T}_M(\boldsymbol{l}_n) \succcurlyeq 0$ and of rank one, $\mathbf{P}_1(\boldsymbol{t}_m)$ is of rank one, and $\hat{\mathbf{x}}[0, n] = \hat{\mathbf{x}}[0, 0]$ for all m and n.*

Proof. See Appendix C. ∎

Besides, with $\mathbf{D} = \mathrm{diag}(c_1, \ldots, c_K)$ and $\mathbf{V}_n = \begin{bmatrix} \boldsymbol{a}(f_{n_2,1}) & \cdots & \boldsymbol{a}(f_{n_2,K}) \end{bmatrix}$, remark that

$$\mathbf{T}_M(\boldsymbol{l}_{n_2}) = \sum_{k=1}^{K} c_k \mathbf{T}_M(\boldsymbol{a}(f_{n_2,k})) = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_{n_2,k})\boldsymbol{a}(f_{n_2,k})^* = \mathbf{V}_{n_2}\mathbf{D}\mathbf{V}_{n_2}^* \ , \tag{35}$$

hence, the nuclear norm of $\mathbf{T}_M(\boldsymbol{l}_{n_2})$ is

$$\|\mathbf{T}_M(\boldsymbol{l}_{n_2})\|_* = \mathrm{Tr}(|\mathbf{D}|) = \sum_{k=1}^{K} c_k = \|\boldsymbol{l}_{n_2}^\sharp\|_{\mathcal{A}} \ . \tag{36}$$

The first equality is explained by the SVD decomposition of a matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$. The nuclear norm is defined by $\|\mathbf{X}\|_* = \mathrm{Tr}(\sqrt{\mathbf{X}^\mathsf{T}\mathbf{X}}) = \mathrm{Tr}(\sqrt{\mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{U}\Sigma\mathbf{V}^\mathsf{T}})$. $\mathbf{U}$ and $\mathbf{V}$ are orthogonal; that is, $\mathbf{U}^\mathsf{T}\mathbf{U} = 1$ and $\mathbf{V}^\mathsf{T}\mathbf{V} = 1$, so $\|\mathbf{X}\|_* = \mathrm{Tr}(\sqrt{\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^\mathsf{T}})$, and then by circularity of the trace we obtain $\|\mathbf{X}\|_* = \mathrm{Tr}(\sqrt{\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{\Sigma}^2}) = \mathrm{Tr}(|\mathbf{\Sigma}|)$, that's the point.

The nuclear norm of a matrix corresponds to the sum of the absolute value of its singular values decomposition, and it is the best convex approximation of the rank of this matrix [44, 43]. Consequently, in the following, we consider a convex relaxation of the line characterization given in Proposition 4, in which the rank constraint on $\mathbf{T}_M(\boldsymbol{l}_{n_2})$ is replaced by an atomic norm constraint on $\|\boldsymbol{l}_{n_2}^\sharp\|_{\mathcal{A}}$. Since the minimum value to achieve is $c^\sharp = \sum_{k=1}^{K} c_k$, and that the atomic norm lie on the first column $\hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0]$, we impose the constraint $\hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0] \leqslant c^\sharp$. Moreover, $\left\{\boldsymbol{z} \in \mathbb{C}^N : \|\boldsymbol{z}\|_{\mathcal{A}} \leqslant 1\right\}$ is the smallest convex set containing $\left\{\boldsymbol{z} \in \mathbb{C}^N : \mathbf{P}_1(\boldsymbol{z}) \text{ of rank 1 and } |z_0| = 1\right\}$. Indeed, let $\boldsymbol{z} = (z_0, \ldots, z_{N-1})$ be a complex vector such that $\mathbf{P}_1(\boldsymbol{z})$ of rank 1, it implies that $z_i = z_0 \mathrm{e}^{\mathrm{j}2\pi f_0 i} = |z_0|\mathrm{e}^{\mathrm{j}(2\pi f_0 i + \phi_0)}$ (see subsection 2.1), hence $\boldsymbol{z} = |z_0| a(f_0, \phi_0)$ and if $|z_0| = 1$ we have $\boldsymbol{z} = a(f_0, \phi_0)$ and then $\left\{\boldsymbol{z} \in \mathbb{C}^N : \mathbf{P}_1(\boldsymbol{z}) \text{ of rank 1 and } |z_0| = 1\right\} = \mathcal{A}$. Besides, $\|\boldsymbol{z}\|_{\mathcal{A}}$ is a Minkowski functional associated to the close convex set $\mathrm{Conv}(A)$, and we have that the unit ball $\left\{\boldsymbol{z} \in \mathbb{C}^N : \|\boldsymbol{z}\|_{\mathcal{A}} \leqslant 1\right\} = \mathrm{Conv}(A)$, which is the smallest convex set containing $\mathcal{A} = \left\{\boldsymbol{z} \in \mathbb{C}^N : \mathbf{P}_1(\boldsymbol{z}) \text{ of rank 1 and } |z_0| = 1\right\}$. So one can replace the rank constraint on $\mathbf{P}_1(\boldsymbol{t}_m)$ by an atomic norm constraint on $\|\boldsymbol{t}_m\|_{\mathcal{A}}$.

4. Minimization Problem with Atomic Norm Regularization. Given $\hat{\mathbf{y}} = \hat{\mathbf{b}}^\sharp + \hat{\boldsymbol{\epsilon}}$ and the filters $\boldsymbol{g}$ and $\boldsymbol{h}$, we are looking for an image $\hat{\mathbf{x}} \in \mathcal{X}$ which minimizes $\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|$, for the norm derived from the inner product (21), and satisfies properties (31) and (32). We fixed a constant c greater than the oracle $c^\sharp = \sum_{k=1}^{K} c_k$. Consequently, the following optimization problem provides an estimator of (20):

$$\tilde{\mathbf{x}} \in \underset{\hat{\mathbf{x}},\mathbf{q}\in\mathcal{X}\times\mathcal{Q}}{\arg\min} \frac{1}{2}\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2 \,, \tag{37}$$

$$s.t \quad \begin{cases} \forall n_2 = 0, ..., H_S - 1,\ \forall m = 0, ..., M \,, & \\ \hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0] \leqslant c \,, & \text{(38a)} \\ \mathbf{q}[m, 0] \leqslant c \,, & \text{(38b)} \\ \mathbf{T}'_{H_S}(\hat{\mathbf{x}}[m, :], \mathbf{q}[m, :]) \succcurlyeq 0 \,, & \text{(38c)} \\ \mathbf{T}_{M+1}(\hat{\mathbf{x}}[:, n_2]) \succcurlyeq 0 \,, & \text{(38d)} \end{cases}$$

Note that this optimization problem could be rewritten in a Lagrangian form, but it would involve a parameter λ to tune, what is not any better than tuning parameter c which has the merit of having a physical meaning. We keep this constraint formulation and write it in a more suitable way as follows. Let denote $\mathcal{H} = \mathcal{X} \times \mathcal{Q}$ the Hilbert space in which the variable of optimization $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$ lies. Let us define $\mathrm{L}_m^{(1)}(\mathbf{X}) = \mathbf{T}'_{H_S}(\hat{\mathbf{x}}[m, :], \mathbf{q}[m, :])$ and $\mathrm{L}_{n_2}^{(2)}(\mathbf{X}) = \mathbf{T}_{M+1}(\hat{\mathbf{x}}[:, n_2])$, ι_C be the indicator function of a convex set C defined by

$$\iota_C : x \mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases} \,, \tag{39}$$

$\mathcal{B} \subset \mathcal{H}$ the set corresponding to the boundary constraints (38)–(38), and $\mathcal{C}$ the cone of positive semidefinite matrices. Then, the optimization problem **??** 37–38 can be rewritten as follows:

$$\tilde{\mathbf{X}} = \underset{\mathbf{X}=(\hat{\mathbf{x}},\mathbf{q})\in\mathcal{H}}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2 + \iota_{\mathcal{B}}(\mathbf{X}) + \sum_{m=0}^{M} \iota_{\mathcal{C}}(\mathrm{L}_m^{(1)}(\mathbf{X})) + \sum_{n_2=0}^{H_S-1} \iota_{\mathcal{C}}(\mathrm{L}_{n_2}^{(2)}(\mathbf{X})) \right\} \,. \tag{40}$$

4.1. First Algorithm Design. The optimization problem (40) can be viewed in the framework above, involving Lipschitzian, proximable and linear composite terms [17]:

$$\tilde{\mathbf{X}} = \underset{\mathbf{X}\in\mathcal{H}}{\arg\min} \left\{ F(\mathbf{X}) + G(\mathbf{X}) + \sum_{i=0}^{N-1} H_i(\mathrm{L}_i(\mathbf{X})) \right\} \,, \tag{41}$$

with $F(\mathbf{X}) = \frac{1}{2}\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2$, $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$, ∇F a β–Lipschitz gradient (with $\beta = \|\mathbf{A}\|^2 = 1$, see below), $G = \iota_{\mathcal{B}}$, which is proximable, and $N = M + 1 + H_S$ linear composite terms where $H_i = \iota_{\mathcal{C}}$ and $\mathrm{L}_i \in \{\mathrm{L}_i^{(1)}, \mathrm{L}_i^{(2)}\}$. We define $\mathbf{H}\boldsymbol{x} = \sum_{i=0}^{N-1} H_i x_i$ and $\mathbf{L}^{(1)}(\mathbf{X}) = (\mathrm{L}_0^{(1)}(\mathbf{X}), \ldots, \mathrm{L}_m^{(1)}(\mathbf{X}))$, $\mathbf{L}^{(2)}(\mathbf{X}) = (\mathrm{L}_0^{(2)}(\mathbf{X}), \ldots, \mathrm{L}_{H_S-1}^{(1)}(\mathbf{X}))$ and $\mathbf{L} = (\mathbf{L}^{(1)}, \mathbf{L}^{(2)})$. We get

$$\|\mathbf{L}\|^2 = \|\mathbf{L}^{(1)}\|^2 + \|\mathbf{L}^{(2)}\|^2 = \|\mathbf{T}'_{H_S}\|^2 + \|\mathbf{T}_{M+1}\|^2 = (H_S - 1) + (M + 1) \,. \tag{42}$$

Let $\tau > 0$ and $\sigma > 0$ such that

$$\frac{1}{\tau} - \sigma\|\mathbf{L}\|^2 > \frac{\beta}{2} \,. \tag{43}$$

Then the primal–dual Algorithm 1 converges to a solution $(\tilde{\mathbf{X}}, \tilde{\boldsymbol{\xi}}_0, ..., \tilde{\boldsymbol{\xi}}_{N-1})$ of the problem (41) [17, Theorem 5.1].

Algorithm 1 Primal–dual splitting algorithm for (41)

Input: $\hat{\mathbf{y}}$ 1D FFT of the blurred and noisy data image y
Output: $\tilde{x}$ solution of the optimization problem **??** 37–38

1: Initialize all primal and dual variables to zero
2: **for** $n = 1$ **to** Number of iterations **do**
3: $\quad \mathbf{X}_{n+1} = \mathrm{prox}_{\tau G}(\mathbf{X}_n - \tau \nabla F(\mathbf{X}_n) - \tau \sum_{i=0}^{N-1} \mathrm{L}_i^* \boldsymbol{\xi}_{i,n})$,
4: $\quad$ **for** $i = 0$ **to** $N-1$ **do**
5: $\quad\quad \boldsymbol{\xi}_{i,n+1} = \mathrm{prox}_{\sigma H_i^*}(\boldsymbol{\xi}_{i,n} + \sigma \mathrm{L}_i(2\mathbf{X}_{n+1} - \mathbf{X}_n))$,
6: $\quad$ **end for**
7: **end for**

We detail below the terms in step 3 and 5.
For $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$, the gradient of F is

$$\nabla F(\mathbf{X}) = (\mathbf{A}^*(\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}), \mathbf{0})^\mathsf{T}. \tag{44}$$

with $\mathbf{A}\hat{\mathbf{x}} = (\hat{\mathbf{x}}\hat{\mathbf{G}}) * \boldsymbol{h}$, where $\hat{\mathbf{G}} = \mathrm{diag}(\hat{g}_{M+1}, \ldots, \hat{g}_W)$, and then $\mathbf{A}^*\boldsymbol{z} = (\boldsymbol{z}\overline{\hat{\mathbf{G}}}) * \bar{\boldsymbol{h}}'$ where $\boldsymbol{h}'$ is the flipped vector from $\boldsymbol{h}$. Let compute the norm of this operator. First remark that if we denote by $\hat{\mathbf{x}}_k$ the k-th column of $\hat{\mathbf{x}}$, then we have by the Parseval equality $\|\hat{\mathbf{x}}_k * \boldsymbol{h}\|_2^2 = \|\hat{\mathbf{x}}_k \hat{\boldsymbol{h}}\|_2^2 \leqslant \|\hat{\boldsymbol{h}}\|_\infty^2 \|\hat{\mathbf{x}}_k\|_2^2$. With the norm derived from (21) we get

$$\begin{aligned} \|\mathbf{A}\hat{\mathbf{x}}\|^2 &= |\hat{g}_{M+1}|^2 \|\hat{\mathbf{x}}_0 * \boldsymbol{h}\|_2^2 + 2|\hat{g}_{M+1}|^2 \|\hat{\mathbf{x}}_1 * \boldsymbol{h}\|_2^2 + \cdots + 2|\hat{g}_W|^2 \|\hat{\mathbf{x}}_M * \boldsymbol{h}\|_2^2 \,, \\ &\leqslant \|\hat{\boldsymbol{g}}\|_\infty^2 \|\hat{\boldsymbol{h}}\|_\infty^2 (\|\hat{\mathbf{x}}_0\|_2^2 + 2\|\hat{\mathbf{x}}_1\|_2^2 + \cdots + 2\|\hat{\mathbf{x}}_M\|_2^2) \,, \\ &\leqslant \|\hat{\boldsymbol{g}}\|_\infty^2 \|\hat{\boldsymbol{h}}\|_\infty^2 \|\hat{\mathbf{x}}\|^2 \,, \end{aligned} \tag{45}$$

which proves that $\|\mathbf{A}\| = \|\hat{\boldsymbol{g}}\|_\infty \|\hat{\boldsymbol{h}}\|_\infty = 1$. Consequently,

$$\|\nabla F(\mathbf{X}) - \nabla F(\mathbf{X}')\| \leqslant \|\mathbf{A}^*\mathbf{A}(\hat{\mathbf{x}} - \hat{\mathbf{x}}')\| \leqslant \|\mathbf{A}^*\mathbf{A}\| \|\hat{\mathbf{x}} - \hat{\mathbf{x}}'\| \leqslant \|\mathbf{A}\|^2 \|\mathbf{X} - \mathbf{X}'\| \,. \tag{46}$$

Set $\overline{x_0} = \frac{1}{H_S} \sum_{n_2=0}^{H_S-1} \hat{\mathbf{x}}[0, n_2]$, we get $\forall m, n_2$:

$$\mathrm{prox}_{\tau G}(\hat{\mathbf{x}}, \mathbf{q}) = \begin{cases} \hat{\mathbf{x}}[0, n_2] = \overline{x_0} & \text{if } \overline{x_0} \leqslant c \\ \hat{\mathbf{x}}[0, n_2] = c & \text{otherwise} \\ \mathbf{q}[m, 0] = c & \text{if } \mathbf{q}[m, 0] > c \end{cases} . \tag{47}$$

Let be $M^{(1)} \in \mathcal{M}_{H_S+1}(\mathbb{C})$ and $M^{(2)} \in \mathcal{M}_{M+1}(\mathbb{C})$. We give the expression of the vectors resulting from these adjoints $\mathbf{T}'^*_{H_S+1} M^{(1)} = (\hat{\mathbf{x}}_1, \mathbf{q}_1) \in \mathbb{C}^{2H_S}$, $\mathbf{T}^*_{M+1} M^{(2)} = \hat{\mathbf{x}}_2 \in \mathbb{C}^{M+1}$ with the appropriated inner product on rows and columns herited from the inner product (21).

$$\hat{\mathbf{x}}_1[k] = \frac{1}{2}(M^{(1)}_{k,H_S+1} + {M^{(1)}_{H_S+1,k}}^*) \,, \tag{48}$$

$$\mathbf{q}_1 = \mathbf{T}^*_{H_S+1} M^{(1)}, \quad \hat{\mathbf{x}}_2 = \mathbf{T}^*_{M+1} M^{(2)} \,, \tag{49}$$

where $\mathbf{T}_N^*$ the adjoint of $\mathbf{T}_N$, applied to $M^{(\ell)}$, $\ell \in \{1,2\}$, is

$$(\mathbf{T}_N^* M^{(\ell)})[k] = \begin{cases} \frac{1}{2}\mathrm{Re}\left\{\sum_{i=1}^{N} M_{ii}^{(\ell)}\right\} & \text{if k=1 ,} \\ \frac{1}{2}\sum_{i=1}^{N-k}(M_{i,k+i-1}^{(\ell)} + M_{k+i-1,i}^{(\ell)}{}^{*}) & \text{if k¿1 .} \end{cases}$$

Let $P_{\mathcal{C}}$ be the projection operator onto $\mathcal{C}$, by Moreau identity:

$$\mathrm{prox}_{\sigma H_i^*}(v) = v - \sigma \mathrm{prox}_{\frac{H_i}{\sigma}}\left(\frac{v}{\sigma}\right) = v - \sigma P_{\mathcal{C}}\left(\frac{v}{\sigma}\right). \tag{50}$$

For more details on convex analysis, monotone operator theory and proximal splitting methods see [6, 2, 16, 36].

4.2. Second Algorithm Design. Notice that in the Algorithm 1, τ must be smaller than $2/\beta$, which is a limitation in terms of convergence speed. To overcome this issue, we subsequently developed a second algorithm, similar to Algorithm 1, but with the data fidelity term $\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|$ activated through its proximity operator, instead of its gradient. We consider the optimization problem as a relaxed version of the Chambolle–Pock algorithm [12]:

$$\mathbf{X}^\star = \arg\min_{\mathbf{X}\in\mathcal{H}} \{G(\mathbf{X}) + \mathbf{H}(\mathbf{L}(\mathbf{X}))\}, \tag{51}$$

with $G = \frac{1}{2}\|\mathbf{A}\cdot - \hat{\mathbf{y}}\|^2$ which is proximable, $\mathbf{H}\boldsymbol{x} = \sum_{i=0}^{N} H_i x_i$ with $H_i = \iota_{\mathcal{C}}$ and $\mathrm{L}_i \in \{\mathrm{L}_i^{(1)}, \mathrm{L}_i^{(2)}\}$ for $i < N$, $H_N = \iota_{\mathcal{B}}$ and $L_N = \mathrm{Id}$. So now, $\|\mathbf{L}\| = H_S + M + 1$. Since we have

$$\begin{aligned} \boldsymbol{p} = \mathrm{prox}_{\tau G}(\hat{\mathbf{x}}) &\Leftrightarrow \hat{\mathbf{x}} - \boldsymbol{p} = \nabla(\tau G)(\boldsymbol{p}) , \\ &\Leftrightarrow \hat{\mathbf{x}} - \boldsymbol{p} = \tau\mathbf{A}^*(\mathbf{A}\boldsymbol{p} - \hat{\mathbf{y}}) , \\ &\Leftrightarrow \hat{\mathbf{x}} + \tau\mathbf{A}^*\hat{\mathbf{y}} = (\mathbf{I} + \tau\mathbf{A}^*\mathbf{A})\boldsymbol{p} , \end{aligned} \tag{52}$$

then the proximal operator has the following expression:

$$\mathrm{prox}_{\tau G}(\hat{\mathbf{x}}) = (\mathbf{I} + \tau\mathbf{A}^*\mathbf{A})^{-1}(\hat{\mathbf{x}} + \tau\mathbf{A}^*\hat{\mathbf{y}}) . \tag{53}$$

Let explicit the operator $\mathbf{A}$ from a matricial point of view. The vertical convolution with the filter $\boldsymbol{h} = [h_{-S}, \ldots, h_0, \ldots, h_S]$ and the column multiplication by the Fourier coefficient $[\hat{g}_{M+1}, \ldots, \hat{g}_W]$ are respectively equivalent to a right and left matrix multiplication with the matrices $\hat{\mathbf{G}}$ of size $(M+1)\times(M+1)$ and $\check{\mathbf{H}}$ of size $H \times H_S$ defined by:

$$\check{\mathbf{H}} = \begin{pmatrix} h_{-S} & \cdots & h_S & 0 & 0 & \cdots & 0 \\ 0 & h_{-S} & \cdots & h_S & 0 & \cdots & 0 \\ \vdots & \ldots & \ddots & \ddots & \ddots & \ldots & \vdots \\ 0 & \cdots & 0 & h_{-S} & \cdots & h_S & 0 \\ 0 & \cdots & 0 & 0 & h_{-S} & \cdots & h_S \end{pmatrix}, \tag{54}$$

that is, $\mathbf{A}\hat{\mathbf{x}} = \check{\mathbf{H}}\hat{\mathbf{x}}\hat{\mathbf{G}}$. We prove in a similar way that $\mathbf{A}^*\hat{\mathbf{y}} = \check{\mathbf{H}}^{\mathsf{T}}\hat{\mathbf{y}}\overline{\hat{\mathbf{G}}}$ and then:

$$(\mathbf{I} + \tau\mathbf{A}^*\mathbf{A})\hat{\mathbf{x}} = \hat{\mathbf{x}} + \mathbf{P}\hat{\mathbf{x}}\mathbf{Q}, \quad \mathbf{P} = \tau\check{\mathbf{H}}^{\mathsf{T}}\check{\mathbf{H}}, \quad \mathbf{Q} = \hat{\mathbf{G}}\overline{\hat{\mathbf{G}}} . \tag{55}$$

The square matrix $\mathbf{P}$ and $\mathbf{Q}$ are of size $p = H_S$ and $q = M + 1$. We have to solve $(\mathbf{I} + \tau \mathbf{A}^* \mathbf{A})\hat{\mathbf{x}} = \mathbf{z}$; that is, $\hat{\mathbf{x}} + \mathbf{P}\hat{\mathbf{x}}\mathbf{Q} = \mathbf{z}$. This kind of system can be solved by the mean of the Kronecker Product as:

$$\hat{\mathbf{x}} + \mathbf{P}\hat{\mathbf{x}}\mathbf{Q} = \mathbf{z} \quad \Longleftrightarrow \quad (\mathbf{I}_{\mathbf{pq},\mathbf{pq}} + \mathbf{Q} \otimes \mathbf{P}^{\mathsf{T}})\mathbf{Vec}(\hat{\mathbf{x}}) = \mathbf{Vec}(\mathbf{z}). \tag{56}$$

where $\mathbf{Vec}(\hat{\mathbf{x}})$ denotes the vectorization of the matrix $\hat{\mathbf{x}}$ formed by stacking the columns of $\hat{\mathbf{x}}$ into a single column vector, and $\mathbf{I}_{\mathbf{pq},\mathbf{pq}} + \mathbf{Q} \otimes \mathbf{P}^{\mathsf{T}}$ is a matrix of size $pq \times pq$ which can be inverted, given access to $\mathbf{Vec}(\hat{\mathbf{x}})$ and then to $\hat{\mathbf{x}}$. Finally, the operator $\mathrm{prox}_{\tau G}$ is no more than a big matrix-vector product.

Another option consists in working on the columns $\hat{\mathbf{x}}_m$ of $\hat{\mathbf{x}}$, since the operator $\mathbf{A}$ act on them.

$$(\mathbf{I} + \tau \mathbf{A}^* \mathbf{A})\hat{\mathbf{x}} = \boldsymbol{z} \iff (\mathbf{I} + |\hat{\boldsymbol{g}}[m]|^2 \mathbf{P})\hat{\mathbf{x}}_m = \mathbf{z}_m, \quad \forall m = 0, \dots, M\ . \tag{57}$$

This time, the operator $\mathrm{prox}_{\tau G}$ suppose to perform $M + 1$ matrix-vector products of size $p \times p$, which appears to be more efficient in practice.

Let $\tau > 0$ and $\sigma > 0$ such that $\tau\sigma\|\mathbf{L}\| = 1$, then the primal-dual Algorithm 2, with $F = 0$ and weight $\rho = 2$, which is a relaxed version of the Chambolle–Pock algorithm, converges to a solution $(\tilde{\mathbf{X}}, \tilde{\boldsymbol{\xi}}_0, ..., \tilde{\boldsymbol{\xi}}_{N-1})$ of the problem (41) [17, Theorem 5.1].

Algorithm 2 Primal–dual splitting algorithm for (51)

Input: $\hat{\boldsymbol{y}}$ 1D FFT of the blurred and noisy data image y
Output: $\tilde{x}$ solution of the optimization problem **??** 37–38)

Initialize all primal and dual variables to zero
for $n = 1$ **to** Number of iterations **do**
  $\tilde{\mathbf{X}}_{n+1} = \mathrm{prox}_{\tau G}(\mathbf{X}_n - \tau \sum_{i=0}^{N-1} \mathrm{L}_i^* \boldsymbol{\xi}_{i,n})$,
  $\mathbf{X}_{n+1} = 2\tilde{\mathbf{X}}_{n+1} - X_n$
  for $i = 0$ **to** $N - 1$ **do**
    $\tilde{\boldsymbol{\xi}}_{i,n+1} = \mathrm{prox}_{\sigma H_i^*}(\boldsymbol{\xi}_{i,n} + \sigma \mathrm{L}_i(2X_{n+1} - X_n))$,
    $\boldsymbol{\xi}_{i,n+1} = 2\tilde{\boldsymbol{\xi}}_{i,n+1} - \boldsymbol{\xi}_{i,n}$
  end for
end for

4.3. Extended Problem Formulation. We now consider a data image $\mathbf{b}^\sharp$ containing lines with no angle restriction, which extends our previous work [38]. We can decompose this image into the sum of two images $\mathbf{b}^\sharp = \mathbf{b}_1^\sharp + \mathbf{b}_2^\sharp$, with $\mathbf{b}_1^\sharp$ (resp. $\mathbf{b}_2^\sharp$) containing vertical (resp. horizontal) lines. We can also define $\hat{\mathbf{x}}_1^\sharp$ of size $(M+1) \times H_S$ and $\hat{\mathbf{x}}_2^\sharp$ of size $W_S \times (P + 1)$ with $W_S = W + 2S$ and $P = (H - 1)/2$ such as $\mathbf{A}\hat{\mathbf{x}}_1^\sharp = \hat{\mathbf{b}}_1^\sharp$ and $\tilde{\mathbf{A}}\hat{\mathbf{x}}_2^\sharp = \hat{\mathbf{b}}_2^\sharp$, where $\boldsymbol{g}_2 = [\mathbf{0}_{P-S}, \boldsymbol{h}, \mathbf{0}_{P-S}]$ and $\tilde{\mathbf{A}}$ denotes the operator which multiplies each row vector $\hat{\mathbf{x}}_2^\sharp[:, n_2]$ by the corresponding Fourier coefficient $\hat{\boldsymbol{g}}_2[n_2]$ and convolves it with the filter $\boldsymbol{h}$; that is, $\tilde{\mathbf{A}}\hat{\mathbf{x}}_2 = (\hat{\mathbf{G}}_2\hat{\mathbf{x}}_2) * \boldsymbol{h}$ with $\hat{\mathbf{G}}_2 = \mathrm{diag}(\hat{\boldsymbol{g}}_2[P + 1], \dots, \hat{\boldsymbol{g}}_2[H])$. We finally define the Hermitian symmetry operator $\mathbf{S}_1$ (resp. $\mathbf{S}_2$) which from each row $\boldsymbol{v} = [v_0, v_1, \dots, v_M]$ (resp. column $\boldsymbol{v} = [v_0, v_1, \dots, v_P]$) associates the symmetric extension $[\overline{v_M}, \dots, v_0, \dots, v_M]$ (resp. $[\overline{v_P}, \dots, v_0, \dots, v_P]$). Let $\mathbf{X}_1 = (\hat{\mathbf{x}}_1, \mathbf{q}_1)$ and $\mathbf{X}_2 = (\hat{\mathbf{x}}_2, \mathbf{q}_2)$ be the optimization variables. Then, the data

fidelity term is now:

$$F(\mathbf{X}_1,\mathbf{X}_2)=\frac{1}{2}\|\mathcal{F}_1^{-1}\mathbf{S}_1\mathbf{A}\hat{\mathbf{x}}_1+\mathcal{F}_2^{-1}\mathbf{S}_2\tilde{\mathbf{A}}\hat{\mathbf{x}}_2-\mathbf{y}\|_F^2=\frac{1}{2}\|\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y}\|_F^2\,, \tag{58}$$

with $\mathbf{A_1}=\mathcal{F}_1^{-1}\mathbf{S}_1\mathbf{A}$ and $\mathbf{A_2}=\mathcal{F}_2^{-1}\mathbf{S}_2\tilde{\mathbf{A}}$. Let compute the differential:

$$\begin{aligned}
&F(\mathbf{X}_1+\mathbf{h}_1,\mathbf{X}_2+\mathbf{h}_2)\\
&=\frac{1}{2}\langle\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2+\mathbf{A_1}\mathbf{h}_1+\mathbf{A_2}\mathbf{h}_2-\mathbf{y},\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2+\mathbf{A_1}\mathbf{h}_1+\mathbf{A_2}\mathbf{h}_2-\mathbf{y}\rangle\,,\\
&=F(\mathbf{X}_1,\mathbf{X}_2)+\frac{1}{2}\langle\mathbf{A_1}\mathbf{h}_1,\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y}\rangle+\frac{1}{2}\langle\mathbf{A_1}\mathbf{h}_1,\mathbf{A_2}\mathbf{h}_2\rangle+\frac{1}{2}\langle\mathbf{A_1}\mathbf{h}_1,\mathbf{A_1}\mathbf{h}_1\rangle\\
&\quad+\frac{1}{2}\langle\mathbf{A_2}\mathbf{h}_2,\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y}\rangle+\frac{1}{2}\langle\mathbf{A_2}\mathbf{h}_2,\mathbf{A_1}\mathbf{h}_1\rangle+\frac{1}{2}\langle\mathbf{A_2}\mathbf{h}_2,\mathbf{A_2}\mathbf{h}_2\rangle\,,
\end{aligned}$$

and

$$|\langle\mathbf{A_1}\mathbf{h}_1,\mathbf{A_2}\mathbf{h}_2\rangle|\leqslant\|\mathbf{A_1}\|\|\mathbf{A_2}\|\|\mathbf{h}_1\|\|\mathbf{h}_2\|, \tag{59}$$

and with the inner product on the product space we have $\|(\mathbf{h}_1,\mathbf{h}_2)\|^2=\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2$ and then

$$\frac{\|\mathbf{h}_1\|\|\mathbf{h}_2\|}{\|(\mathbf{h}_1,\mathbf{h}_2)\|}=\frac{\|\mathbf{h}_1\|\|\mathbf{h}_2\|}{\sqrt{\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2}}\leqslant\frac{\|\mathbf{h}_1\|\|\mathbf{h}_2\|}{\|\mathbf{h}_1\|}=\|\mathbf{h}_2\|\longrightarrow 0\,, \tag{60}$$

so it follows that $|\langle\mathbf{A_1}\mathbf{h}_1,\mathbf{A_2}\mathbf{h}_2\rangle|=o\left(\|(\mathbf{h}_1,\mathbf{h}_2)\|\right)$, and we deduce that

$$\nabla F(\mathbf{X}_1,\mathbf{X}_2)=\frac{1}{2}\begin{pmatrix}\mathbf{A_1}^*(\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y})\\ \mathbf{A_2}^*(\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y})\end{pmatrix}\,. \tag{61}$$

The adjoints are $\tilde{\mathbf{A}}^*\boldsymbol{z}=(\overline{\hat{\mathbf{G}}_2}\boldsymbol{z})*\bar{\boldsymbol{h}}'$, $(\mathcal{F}_1^{-1})^*=\frac{1}{W}\mathcal{F}_1$ and $(\mathcal{F}_2^{-1})^*=\frac{1}{H}\mathcal{F}_2$, and $\mathbf{S}_1^*(v_{-M},\ldots,v_0,\ldots,v_P)=(v_0,\ldots,v_M)$ and $\mathbf{S}_2^*(v_{-P},\ldots,v_0,\ldots,v_P)=(v_0,\ldots,v_P)$.

Let determine the Lipschitz coefficient of the gradient ∇F:

$$\begin{aligned}
\|\nabla F(\mathbf{X}_1,\mathbf{X}_2)-\nabla F(\mathbf{X}_1',\mathbf{X}_2')\|^2=&\frac{1}{4}\|\mathbf{A_1}^*(\mathbf{A_1}(\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1')+\mathbf{A_2}(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'))\|^2\\
&+\frac{1}{4}\|\mathbf{A_2}^*(\mathbf{A_1}(\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1')+\mathbf{A_2}(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'))\|^2\,.
\end{aligned} \tag{62}$$

We are looking for a majoration of each term. We treat the first one C_1, the second C_2 is obtained in the same manner.

$$\begin{aligned}
C_1&\leqslant(\|\mathbf{A_1}^*\mathbf{A_1}\|\|\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1'\|+\|\mathbf{A_1}^*\mathbf{A_2}\|\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'\|)^2\,,\\
&\leqslant\|\mathbf{A_1}^*\mathbf{A_1}\|^2\|\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1'\|^2+\|\mathbf{A_1}^*\mathbf{A_2}\|^2\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'\|^2\,,\\
&\quad+2\|\mathbf{A_1}^*\mathbf{A_1}\|\|\mathbf{A_1}^*\mathbf{A_2}\|\|\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1'\|\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'\|\,,\\
&\leqslant(\|\mathbf{A_1}^*\mathbf{A_1}\|^2+\|\mathbf{A_1}^*\mathbf{A_1}\|\|\mathbf{A_1}^*\mathbf{A_2}\|)\|\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1'\|^2\,,\\
&\quad+(\|\mathbf{A_1}^*\mathbf{A_2}\|^2+\|\mathbf{A_1}^*\mathbf{A_1}\|\|\mathbf{A_1}^*\mathbf{A_2}\|)\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'\|^2\,.
\end{aligned} \tag{63}$$

We have $\|\mathbf{A}\|=\|\mathbf{A}^*\|=1$, $\|\mathbf{S_i}\hat{\mathbf{x}}_1\|_F=\|\hat{\mathbf{x}}_i\|$, for $i\in\{1,2\}$; that is, $\|\mathbf{S_i}\|=1$, and $\|\mathcal{F}_i^{-1}\boldsymbol{v}\|_2^2=\frac{1}{N^2}\|\boldsymbol{v}\|^2$; that is, $\|\mathcal{F}_i^{-1}\|=\frac{1}{N}$. Hence, $\|\mathbf{A_1}\|\leqslant\frac{1}{W}$, $\|\mathbf{A_1}^*\|\leqslant 1$, $\|\mathbf{A_2}\|\leqslant\frac{1}{H}$ and $\|\mathbf{A_2}\|\leqslant 1$. Consequently, we get

$$C_1\leqslant\left(\frac{1}{W^2}+\frac{1}{WH}\right)\|\hat{\mathbf{x}}_1-\hat{\mathbf{x}}_1'\|^2+\left(\frac{1}{H^2}+\frac{1}{WH}\right)\|\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_2'\|^2\,, \tag{64}$$

and exactly the same majoration for C_2. Thus, we have

(65)
$$\|\nabla F(\mathbf{X}_1,\mathbf{X}_2)-\nabla F(\mathbf{X}_1',\mathbf{X}_2')\|^2 \leqslant \beta^2(\|\mathbf{X}_1-\mathbf{X}_1'\|^2+\|\mathbf{X}_2-\mathbf{X}_2'\|^2), \quad \beta=\frac{1}{\min(W,H)}\,.$$

The image $\hat{\mathbf{x}}_1^\sharp$ keep the same constraints as in the Algorithm 1, which act similarly on the image $\hat{\mathbf{x}}_2^\sharp$ in a rotated way; that is, we define $\mathrm{L}_m^{(3)}(\mathbf{X}_2)=\mathbf{T}_{P+1}(\hat{\mathbf{x}}_2[m,:])$ and $\mathrm{L}_{n_2}^{(4)}(\mathbf{X}_2)=\mathbf{T}'_{W_S}(\mathrm{fliplr}(\hat{\mathbf{x}}_2[:,n_2]),\mathrm{fliplr}(\mathbf{q}_2[:,n_2]))$. The boundary constraints on $\hat{\mathbf{x}}_1$ are given in (38) and (38), so the boundary constraints on $\hat{\mathbf{x}}_2$ are $\hat{\mathbf{x}}_2[m,0]=\hat{\mathbf{x}}_2[0,0]\leqslant c_2$ and $\mathbf{q}_2[P,:]\leqslant c_2$. We still denote by $\mathcal{B}_1$ and $\mathcal{B}_2$ both these constraints. Likewise, the inner product on the space where $\hat{\mathbf{x}}_2$ belongs is:

$$\langle \mathbf{z}_1,\mathbf{z}_2\rangle = \sum_{m=0}^{W_S-1}\mathbf{z}_1[m,0]\mathbf{z}_2[m,0]^* + 2\mathrm{Re}\left(\sum_{n_2=1}^{P}\sum_{m=0}^{W_S-1}\mathbf{z}_1[m,n_2]\mathbf{z}_2[m,n_2]^*\right). \tag{66}$$

and so the adjoint of the operators remain the same, except for the operator $\mathbf{T}_N$ which is slightly different (the first component is not divided by two anymore):

$$(\mathbf{T}_N^* M^{(\ell)})[k] = \begin{cases} \mathrm{Re}\left\{\sum_{i=1}^{N} M_{ii}^{(\ell)}\right\} \text{ if } k=1, \\ \frac{1}{2}\sum_{i=1}^{N-k}(M_{i,k+i-1}^{(\ell)} + M_{k+i-1,i}^{(\ell)}{}^*) \text{ if } k>1. \end{cases}$$

As previously, we define $\mathbf{L}^{(3)}(\mathbf{X}_2)=(\mathrm{L}_0^{(3)}(\mathbf{X}_2),\dots,\mathrm{L}_{W_S-1}^{(3)}(\mathbf{X}_2))$, also $\mathbf{L}^{(4)}(\mathbf{X}_2)=(\mathrm{L}_0^{(4)}(\mathbf{X}_2),\dots,\mathrm{L}_P^{(4)}(\mathbf{X}_2))$ and $\mathbf{L}=(\mathbf{L}^{(1)},\mathbf{L}^{(2)},\mathbf{L}^{(3)},\mathbf{L}^{(4)})$. Again, an easy computation leads to

$$\|\mathbf{L}\|^2 = \|\mathbf{L}^{(1)}\|^2+\|\mathbf{L}^{(2)}\|^2+\|\mathbf{L}^{(3)}\|^2+\|\mathbf{L}^{(4)}\|^2 \tag{67}$$
$$= (H_S-1)+(M+1)+(P+1)+(W_S-1)\,. \tag{68}$$

Finally, we have

$$\begin{aligned}(\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2) = \underset{(\mathbf{X}_1,\mathbf{X}_2)\in\mathcal{H}}{\arg\min}\Bigg\{ &\frac{1}{2}\|\mathbf{A_1}\hat{\mathbf{x}}_1+\mathbf{A_2}\hat{\mathbf{x}}_2-\mathbf{y}\|_F^2 \\ &+\iota_{\mathcal{B}_1}(\mathbf{X}_1)+\sum_{m=0}^{M}\iota_{\mathcal{C}}(\mathrm{L}_m^{(1)}(\mathbf{X}_1))+\sum_{n_2=0}^{H_S-1}\iota_{\mathcal{C}}(\mathrm{L}_{n_2}^{(2)}(\mathbf{X}_1)) \\ &+\iota_{\mathcal{B}_2}(\mathbf{X}_2)+\sum_{m=0}^{W_S-1}\iota_{\mathcal{C}}(\mathrm{L}_m^{(3)}(\mathbf{X}_2))+\sum_{n_2=0}^{P}\iota_{\mathcal{C}}(\mathrm{L}_{n_2}^{(4)}(\mathbf{X}_2))\Bigg\}\,.\end{aligned} \tag{69}$$

4.4. Inpainting problems. We now consider the case in which a binary mask $\mathbf{M}$ is applied on the data image, as in Figure 4. More precisely, the result $\mathbf{M}\cdot\mathbf{x}$ is an element-wise multiplication of the matrix $\mathbf{x}$ with the binary matrix $\mathbf{M}$, whose zero coefficients are the indices of the pixels unavailable to observation. It is clear that with the Frobenius inner product we have $\mathbf{M}^*=\mathbf{M}$. The data fidelity term becomes $F(\mathbf{X})=\frac{1}{2}\|\mathbf{M}\mathcal{F}_1^{-1}\mathbf{S}_1\mathbf{A}\hat{\mathbf{x}}-\mathbf{y}\|_F^2$, whose gradient is computed as previously, with $\beta=1/W$ (since $\|\mathcal{F}_1^{-1}\|=1/W$ and $\|\mathbf{M}\|=1$). The constraints remain the same

as in (38), and the method is also easily transposable to the extended version.

At this stage, we completed the first step; that is, we are able to restore the image $\hat{\mathbf{x}}^\sharp$ from the degraded image $\mathbf{y}$. From now, we can for instance reduce the blur by taking other filters $\boldsymbol{g}_r$ and $\boldsymbol{h}_r$ with a smaller spread, and visualize the resulting image $\boldsymbol{b}_r$ passing the solution $\hat{\mathbf{x}}$ through this new blur operator $\mathbf{A}_r$; that is, $\hat{\boldsymbol{b}}_r = \mathbf{A}_r\hat{\mathbf{x}}$. Note that for a too small variance σ, the gaussian $\mathcal{G} : t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}$ is tight, so its Fourier transform $\hat{\mathcal{G}} : \nu \mapsto e^{-2\pi^2\sigma^2\nu^2}$ is spread out. On one hand, the filter $\boldsymbol{g} = [\mathbf{0}_{M-S}, \boldsymbol{h}, \mathbf{0}_{M-S}]$ interpolates the few samples $\boldsymbol{h}$ of the Gaussian $\mathcal{G}$ in the Fourier domain, but on the other hand the resulting vector $\hat{\boldsymbol{g}}$ of size $2M+1 = W$ does not contain a sufficiently large part of $\hat{\mathcal{G}}$ due to its slight decrease. Consequently, the bandlimited nature of φ_1 cut off an effective part of the high frequencies of the signal, and thus introduce some oscillations which appear as artifacts on the blurred image.

In the next section, we present the method that underlies the second step of this work, namely the estimation of the line parameters, which is related to the spectral estimation field.

5. Recovering Line Parameters by a Prony-Like Method. Finally, the goal is to estimate the parameters $(\theta_k, \alpha_k, \eta_k)$, which characterize the K lines, from the solution of the minimization problem $\hat{\mathbf{x}}$, symmetrized to $m = -M, \dots, -1$ beforehand. This requires the use of a classical spectral estimation method [48, 49]. The recovering procedure hereafter, based on [42], is an extended method of the famous Prony method [39]. There exists plenty other methods like MUSIC [46] and root–MUSIC [1], ESPRIT [45] or matrix pencil method [26].

Let sketch this Prony–like method [42], which is based on an annihilating property [4]. Let $\boldsymbol{z} = (z_0, \dots, z_{|I|-1})$ be a complex vector, whose components are:

$$z_i = \sum_{k=1}^{K} d_k \left(\mathrm{e}^{\mathrm{j}2\pi f_k}\right)^i, \quad \forall i = 0, \dots, |I|-1 \,, \tag{70}$$

with $d_k \in \mathbb{C}$, $f_k \in [-1/2, 1/2)$ and $\zeta_k = \mathrm{e}^{\mathrm{j}2\pi f_k}$. We introduce the annihilating polynomial $H(\zeta) = \prod_{l=1}^{K}(\zeta - \zeta_l) = \sum_{l=0}^{K} h_l \zeta^{K-l}$ with $h_0 = 1$, and Prony [39] noticed that for all $r = K, \dots, |I|-1$

$$\sum_{l=0}^{K} h_l z_{r-l} = \sum_{l=0}^{K} h_l \left(\sum_{k=1}^{K} d_k \zeta_k^{r-l}\right) = \sum_{k=1}^{K} d_k \zeta_k^{r-K} \underbrace{\left(\sum_{l=0}^{K} h_l \zeta_k^{K-l}\right)}_{H(\zeta_k)=0} = 0 \,. \tag{71}$$

We rearrange the elements z_i in a Toeplitz matrix $\mathbf{P}_K(\boldsymbol{z})$ of size $(|I|-K) \times (K+1)$ and rank K as follows

$$\mathbf{P}_K(\boldsymbol{z}) = \begin{pmatrix} z_K & \cdots & z_0 \\ \vdots & \ddots & \vdots \\ z_{|I|-1} & \cdots & z_{|I|-K-1} \end{pmatrix} \,, \tag{72}$$

and (71) can be written with $\boldsymbol{h} = (h_0, \dots, h_K)$ as:

$$(\boldsymbol{z} * \boldsymbol{h})(r) = 0, \quad \forall r = K, \dots, |I|-1 \iff \mathbf{P}_K \boldsymbol{h} = 0 \,. \tag{73}$$

Consequently, the method consists in finding a right singular vector of the matrix $\mathbf{P}_K$ associated to the singular value zero, which leads to vector $\boldsymbol{h} = (h_0, \dots, h_K)$ and then to the polynomial $H(\zeta) = \sum_{l=0}^{K} h_l \zeta^{K-l}$, whose roots are the searched complex $\zeta_k = \mathrm{e}^{\mathrm{j}2\pi f_k}$ and then $f_k = \arg(\zeta_k)/(2\pi)$. The complex amplitudes can be retrieved as well by written (70) matricially $\boldsymbol{z} = \mathbf{U}\boldsymbol{d}$ where $\boldsymbol{d} = (d_1, \dots, d_K)$ and the following matrix $\mathbf{U}$ of size $|I| \times K$

$$\mathbf{U} = \begin{pmatrix} \boldsymbol{a}(f_1) & \cdots & \boldsymbol{a}(f_K) \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ \mathrm{e}^{-\mathrm{j}2\pi f_1} & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_K} \\ \mathrm{e}^{-\mathrm{j}2\pi f_1 \times 2} & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_K \times 2} \\ \vdots & \vdots & \vdots \\ \mathrm{e}^{-\mathrm{j}2\pi f_1 (|I|-1)} & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_K (|I|-1)} \end{pmatrix} . \tag{74}$$

Finally, we recover the amplitudes by a least square approximation

$$\boldsymbol{d} = (\mathbf{U}^{\mathbf{H}}\mathbf{U})^{-1}\mathbf{U}^{\mathbf{H}}\boldsymbol{z} \, . \tag{75}$$

Applied on the columns of $\hat{\mathbf{x}}^\sharp$, the procedure becomes the following:

– From $m = 1, \dots, M$,

1. Compute $\tilde{f}_{m,k} = \arg(\zeta_{m,k})/(2\pi)$, where $(\zeta_{m,k})_k$ are roots of the polynomial $\sum_{k=0}^{K} h_{m,k}\zeta^k$ with $\boldsymbol{h}_m = [h_{m,0}, \dots, h_{m,K}]^\mathsf{T}$ being the right singular vector of $\mathbf{P}_K(\tilde{\mathbf{x}}[m,:])$ with $I = \{0, \dots, H_S - 1\}$. It corresponds to the singular value zero (the smallest value in practice).
2. Form the matrix $\tilde{\mathbf{U}}_\mathbf{m} = [\boldsymbol{a}(\tilde{f}_{m,1}) \cdots \boldsymbol{a}(\tilde{f}_{m,K})]$, and compute the complex amplitudes $\tilde{\boldsymbol{d}}_m = [\tilde{d}_{m,1}, \dots, \tilde{d}_{m,K}]^\mathsf{T}$ by solving the least-squares linear system $\tilde{\mathbf{U}}_\mathbf{m}^\mathbf{H}\tilde{\mathbf{U}}_\mathbf{m}\tilde{\boldsymbol{d}}_m = \tilde{\mathbf{U}}_\mathbf{m}^\mathbf{H}\tilde{\mathbf{x}}[m,:]$.
3. Compute $\tilde{\theta}_{m,k} = \arctan(W\tilde{f}_{m,k}/m)$ from (30).
4. Compute $\tilde{\alpha}_{m,k} = |\tilde{d}_{m,k}| \cos\tilde{\theta}_{m,k}$.
5. Compute $\tilde{e}_{m,k} = \tilde{d}_{m,k}/|\tilde{d}_{m,k}|$.

– For $k = 1, \dots, K$

6. Sort the $\tilde{f}_{m,k}$ with respect to k, and apply this permutation on the other array. Compute the mean of all estimated angles $\tilde{\theta}_k = \frac{1}{M}\sum_{m=1}^{M}\tilde{\theta}_{m,k}$ and amplitudes $\tilde{\alpha}_k = \frac{1}{M}\sum_{m=1}^{M}\tilde{\alpha}_{m,k}$.
7. Compute the frequency $\tilde{\nu}_k$ as previously from $\mathbf{P}_K(\tilde{\boldsymbol{e}}_k)$ with $\tilde{\boldsymbol{e}}_k = (\tilde{e}_{m,k})_m$ and $I = \{-M, \dots, M\}$, and then the horizontal offset $\tilde{\eta}_k = W\tilde{\nu}_k/(2\pi)$.

The point 6. and 7. are possible to the extent that the sorting frequencies $\tilde{f}_{m,k}$ are always related to the same angles $\theta_1 \leqslant \cdots \leq \theta_k \leq \cdots \leq \theta_K$ for all m, which allows us to compute the mean according to m. It would not be possible to do the same with $\tilde{f}_{n_2,k} = (\tan(\tilde{\theta}_k)n_2 - \tilde{\eta}_k)/W$ because the affine relation does not preserve the order, for instance one can find n and n' such that $\tilde{f}_{n,k_1} \leq \tilde{f}_{n,k_2}$ and $\tilde{f}_{n',k_1} \geq \tilde{f}_{n',k_2}$.

Thus, the trick is to perform the Prony method on the vectors $\tilde{e}_k$ again, with this aim to preserve the order. Indeed, if we would chosen instead of step 6. to perform the Prony method on the rows, in order to extract the frequencies $\tilde{f}_{n_2,k} = (\tan(\tilde{\theta}_k)n_2 - \tilde{\eta}_k)/W$, then on one hand they are not uniquely determined since they belong to an interval of length greater than one $\tilde{f}_{n_2,k} \in [-(H_S + M)/W, (H_S + M)/W]$, and on the other hand we would lose the correspondence between the $\tilde{f}_{n_2,k}$ and the previous estimated angles $\tilde{\theta}_k$, what compromises the estimation of the η_k. The procedure is summarized in Figure 6.

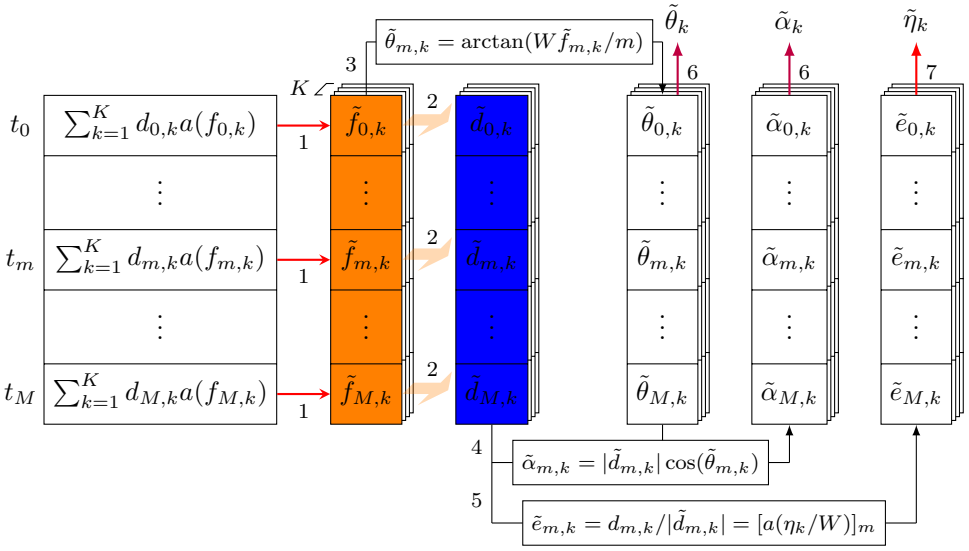


FIG. 6. *Procedure of extraction*

In the preliminary version of this work [38], we propose this simplistic method which is not sufficiently robust to noise remaining in the output solution of the algorithm, whose stopping criteria is met when this one is sufficiently close to the exact solution. In the noiseless case, the Prony method is able to recover the frequencies with an infinite precision if the number of samples $|I|$ is greater than $2K$. But in our case the estimate $\tilde{f}_{m,k}$ is affected by some uncertainty $\epsilon_{m,k}$; that is, $\tilde{f}_{m,k} = f_{m,k} + \epsilon_{m,k}$, due to the instability of roots finding in presence of noise, and then the error of $\tilde{\theta}_{m,k} = \arctan(W\tilde{f}_{m,k}/m) = \arctan(Wf_{m,k}/m + W\epsilon_{m,k}/m) \approx \theta_{m,k} + W\epsilon_{m,k}/m$, is amplified by a factor W/m and gives a bad result, in particularly for a small m. Consequently, the mean $\tilde{\theta}_k = \frac{1}{M}\sum_{m=1}^M \tilde{\theta}_{m,k}$ does not lead to a robust estimation of the angles $\tilde{\theta}_k$. We improve the robustness by applying a linear regression to the data $\{f_{m,k}\}_{1\leqslant m\leqslant M}$ since $\tilde{f}_{m,k} = \frac{\tan\theta_k}{W}m + \epsilon_{m,k}$, in order to estimate the slope $\tan\theta_k$. The errors $\epsilon_{m,k}$ committed by evaluating the frequencies $f_{m,k}$ are not of the same order of magnitude according to m. Indeed, for a small m, the frequencies $f_{m,k} = \frac{\tan\theta_k}{W}m$, are close to each other, and the Prony method fails to accurately determine the frequencies when the minimal separation is too small, especially when K and the amount of residual noise are not too. Consequently, it is preferable to start the linear regression from greater values of m, in order to space the frequencies on the unit disk. Make sure that for high values of m the two extremal frequencies, say $f_{m,1} \leq 0$ and $f_{m,K} \geq 0$, are not both close respectively to $-\pi$ and π, what would violate the separation criteria as well. The angle parameters $\tilde{\theta}_k$ are now better estimated.

We keep the same procedure of estimation for the offsets $\tilde{\eta}_k$, because it has the advantage to conserve the correspondance with the estimated angles $\tilde{\theta}_k$. Notice that we could compare the results by applying the Prony method on the middle line ($n_2 = 0$) of the image, since $\boldsymbol{l}_0[m] = \sum_{k=1}^K c_k e^{j2\pi\eta_k m/W}$ and that the argument of these exponentials are uniquely determined, giving a better estimation of the offsets $\tilde{\eta}_k$ provided that the latter are well separated, but losing the correspondance with the angles $\tilde{\theta}_k$.

Finally, as far as concern the amplitudes c_k, taking the modulus of $\tilde{d}_{m,k}$'s gives also bad results, since they are computed from the estimate $\tilde{\mathbf{x}}$ whose amplitudes have been cut down, due to the choice of a parameter $c < c^\star$ for removing noise. It is more appropriated, given the estimated 2D atoms $\frac{1}{\cos\tilde{\theta}_k}\hat{\boldsymbol{a}}(\tilde{\theta}_k, \tilde{\eta}_k)$, to evaluate the amplitudes $\tilde{\alpha}_k$ by performing a least square method with respect to the noisy data $\hat{\mathbf{y}}$:

$$(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_K) = \underset{\alpha_1,\ldots,\alpha_K}{\arg\min} \left\| \sum_{k=1}^{K} \alpha_k \hat{\mathbf{I}}_k - \hat{\mathbf{y}} \right\|^2, \quad \hat{\mathbf{I}}_k = \mathbf{A} \cdot \frac{1}{\cos\tilde{\theta}_k} \hat{\boldsymbol{a}}(\tilde{\theta}_k, \tilde{\eta}_k) \,. \tag{76}$$

The new procedure is the following:

– From $m = 1, \ldots, M$,

Compute $\tilde{f}_{m,k} = \arg(\zeta_{m,k})/(2\pi)$, where $(\zeta_{m,k})_k$ are roots of the polynomial $\sum_{k=0}^{K} h_{m,k}\zeta^k$ with $\boldsymbol{h}_m = [h_{m,0}, \ldots, h_{m,K}]^\mathsf{T}$ being the right singular vector of $\mathbf{P}_K(\tilde{\mathbf{x}}[m, :])$ with $I = \{0, \ldots, H_S - 1\}$. It corresponds to the singular value zero (the smallest value in practice).

– For $k = 1, \ldots, K$

Perform a linear regression on $\{\tilde{f}_{m,k}\}_m$ to estimate $\tan\tilde{\theta}_k$ and then $\tilde{\theta}_k$.

– From $m = 1, \ldots, M$,

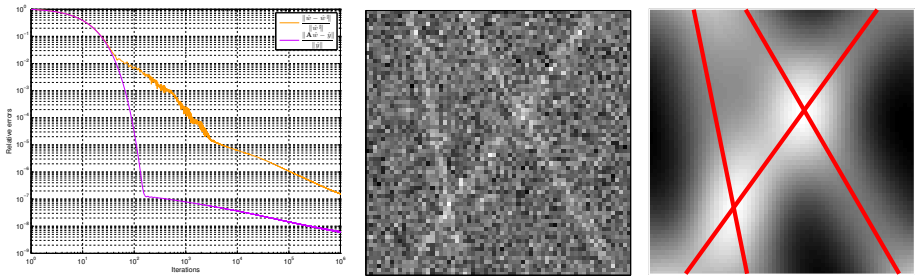
1. Form the matrix $\tilde{\mathbf{U}}_\mathbf{m} = [\boldsymbol{a}(\tan\tilde{\theta}_1 m/W) \cdots \boldsymbol{a}(\tan\tilde{\theta}_K m/W)]$, and compute the complex amplitudes $\tilde{\boldsymbol{d}}_m = [\tilde{d}_{m,1}, \ldots, \tilde{d}_{m,K}]^\mathsf{T}$ by solving the least-squares linear system $\tilde{\mathbf{U}}_\mathbf{m}^\mathbf{H}\tilde{\mathbf{U}}_\mathbf{m}\tilde{\boldsymbol{d}}_m = \tilde{\mathbf{U}}_\mathbf{m}^\mathbf{H}\tilde{\mathbf{x}}[m, :]$.
2. Compute $\tilde{e}_{m,k} = \tilde{d}_{m,k}/|\tilde{d}_{m,k}|$.

– For $k = 1, \ldots, K$

Compute the frequency $\tilde{\nu}_k$ as previously from $\mathbf{P}_K(\tilde{\boldsymbol{e}}_k)$ with $\tilde{\boldsymbol{e}}_k = (\tilde{e}_{m,k})_m$ and $I = \{-M, \ldots, M\}$, and then the horizontal offset $\tilde{\eta}_k = W\tilde{\nu}_k/(2\pi)$.

– Find the amplitudes by debiasing as explained in (76)

6. Experimental results. The reconstruction procedure described in Section 3, was implemented in Matlab code, available on the webpage of the first author. We consider an image of size $W = H = 65$, containing 3 lines of parameters: $(\theta_1, \eta_1, \alpha_1) = (-\pi/5, 0, 255)$, $(\theta_2, \eta_2, \alpha_2) = (\pi/16, -15, 255)$ and $(\theta_3, \eta_3, \alpha_3) = (\pi/6, 10, 255)$. The first experiment consists in the reconstruction of the lines from $\tilde{\mathbf{x}}$ in absence of noise, (1) by applying the operator $\mathbf{A}$ on this solution, possibly with others kernels $\boldsymbol{g}$ and $\boldsymbol{h}$, and then taking the 1D inverse Fourier transform ; and (2) by applying the Prony method to recover parameters of the lines, in the aim to display these one as vectorial lines. We run the algorithm for 10^6 iterations. Results of relative errors for the solution $\tilde{\mathbf{x}}$ and the estimated parameters are given Fig. 7 (a) and Table 1, where $\Delta_{\theta_i}/\theta_i = |\theta_i - \tilde{\theta}_i|/|\theta_i|$, $\Delta_{\alpha_i}/\alpha_i = |\alpha_i - \tilde{\alpha}_i|/|\alpha_i|$ and $\Delta_{\eta_i} = |\eta_i - \tilde{\eta}_i|$. Although the algorithm is quite slow to achieve high accuracy, we insist on the fact that convergence to the exact solution $\hat{\mathbf{x}}^\sharp$ is guaranteed, when the lines are not too close to each other. The purpose of the second experiment is to highlight the robustness of the method in presence of a strong noise level. With $c = 700$ and only 2.10^3 iterations, we are able to completely remove noise and to estimate the line parameters with an error of 10^{-2}. For both first experiments, we do not depict the estimated images, because it is strictly identical to the one in Figure 1. Finally, the last experiment for 10^5 iterations, illustrates the efficiency of the method even in presence of a large blur, yielding an error of 10^{-4}. We emphasize that our algorithm has an accuracy which could not be achieved by detecting peaks of the Hough or Radon transform (see Figure 1). These methods are relevant for giving a coarse estimation of line parameters. They are



(a) (b) (c)

FIG. 7. *(a) Decrease of the relative errors* $\frac{\|\hat{\mathbf{x}}-\hat{\mathbf{x}}^\sharp\|}{\|\hat{\mathbf{x}}^\sharp\|}$ *and* $\frac{\|\mathbf{A}\hat{\mathbf{x}}-\hat{\mathbf{y}}\|}{\|\hat{\mathbf{y}}\|}$ *for the first experiment, (b) Lines affected by a strong noise level* ($\zeta = 200$) *for the second experiment, (c) Lines degraded by a strong blur* ($\kappa = 8$) *for the third experiment. In red, the recovered lines by the Prony Method.*

TABLE 1
Errors on line parameters recovered by the proposed method.

	Experiment 1	**Experiment 2**	**Experiment 3**
Δ_θ/θ	$(10^{-7}, 3.10^{-6}, 7.10^{-7})$	$(10^{-2}, 6.10^{-2}, 9.10^{-2})$	$(6.10^{-7}, 9.10^{-5}, 8.10^{-6})$
Δ_α/α	$(10^{-7}, 10^{-7}, 10^{-7})$	$(10^{-2}, 9.10^{-2}, 2.10^{-1})$	$(4.10^{-5}, 2.10^{-5}, 2.10^{-5})$
Δ_η	$(4.10^{-6}, 7.10^{-6}, 7.10^{-6})$	$(5.10^{-2}, 4.10^{-2}, 3.10^{-2})$	$(5.10^{-5}, 10^{-4}, 3.10^{-4})$

TABLE 2
Angles, offsets and amplitudes of the seven lines.

θ_k	-0.75	-0.5	-0.25	10^{-3}	0.3	0.55	0.75
η_k	15	25	2	7	-20	-5	-10
α_k	60	80	255	100	180	120	240

robust to strong noise, but completely fail with a strong blur, which prevents peaks detection (see Appendix D).

One way to get rid off the periodicity is to work with an image four times bigger, but it is computationally prohibitive. As well, considering a Poisson noise would be more realistic, but it would involve to activate the data fidelity term through its proximal operator in $\mathbf{H} \circ \mathbf{L}$, which would slow down the algorithm further.

6.1. Close lines. For a reasonable amount of noise ($\zeta = 20$), the algorithm succeed in separating two close lines $(\theta_1, \eta_1, \alpha_1) = (-0.73, -1, 255)$ and $(\theta_2, \eta_2, \alpha_2) = (-0.75, 1, 255)$ as illustrated in Figure 8. The estimation of the parameters gives $(\tilde{\theta}_1, \tilde{\eta}_1, \tilde{\alpha}_1) = (-0.725, -0.7, 237)$ and $(\tilde{\theta}_2, \tilde{\eta}_2, \tilde{\alpha}_2) = (-0.753, -0.6, 251)$.

6.2. More lines and different amplitudes. A more complicated example is depicted in Figure 9 (a), containing seven lines whose parameters are enumerated in Table 2, and affected by some noise with variance $\zeta = 20$.

We run the algorithm with $c = 0.8c^\star$, $\tau = 1$, $\sigma = (\tau(M + H_S + 1))^{-1}$, and after only 2.10^3 iterations, we are able to denoise the image as illustrated in Figure 9 (b), and to estimate, thanks to the new Prony procedure, the line parameters as illustrated in Figure 9 (c), with an error of 10^{-2} as given in Table 3.

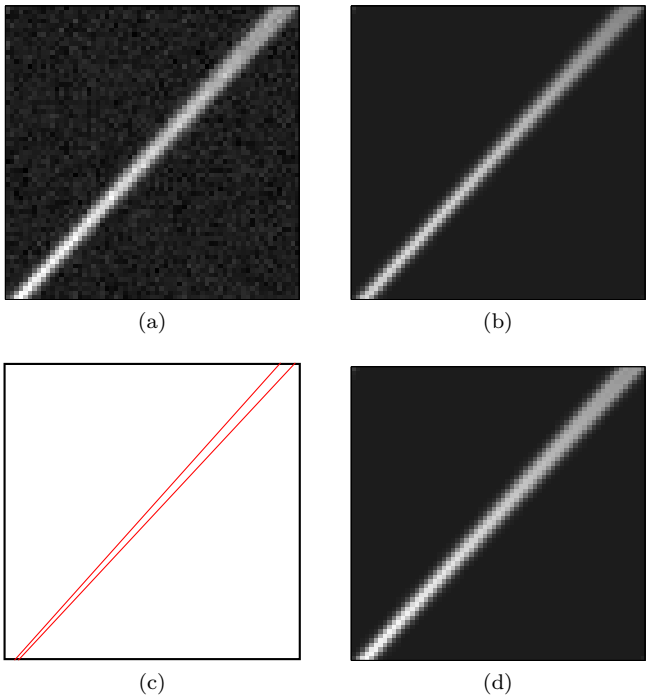


FIG. 8. *(a) Image* $\mathbf{y}$ *of seven lines affected by noise* ($\zeta = 20$)*, (b) Result* $\tilde{\mathbf{x}}$ *of the denoising step from the optimization, (c) Extraction of the line parameters from the Prony step, (d) Theoritical solution* $\hat{\mathbf{x}}^\sharp$ *to compare with.*

TABLE 3
Errors on line parameters recovered by the proposed method.

$\Delta\theta$	1.10^{-2}	2.10^{-2}	1.10^{-3}	2.10^{-3}	5.10^{-3}	5.10^{-3}	1.10^{-3}
$\Delta\eta$	5.10^{-1}	7.10^{-2}	4.10^{-2}	1.10^{-1}	1.10^{-2}	2.10^{-2}	1.10^{-2}
$\Delta\alpha/\alpha$	4.10^{-2}	5.10^{-2}	5.10^{-3}	4.10^{-2}	6.10^{-3}	1.10^{-2}	4.10^{-3}

7. Conclusion. We provided a new formulation for the problem of recovering lines in degraded images using the framework of atomic norm minimization. A primal–dual splitting algorithm has been used to solve the convex optimization problem. We applied it successfully to the deblurring of images, recovering lines parameters by the Prony method, and we showed the robustness of the method for strong blur and strong noise level. We insist on the novelty of our approach, which is to estimate lines with parameters (angle, offset, amplitude) living in a continuum, with perfect reconstruction in absence of noise, without being limited by the discrete nature of the image, nor its finite size. This work can be viewed as a proof of concept for super-resolution line detection, and invite us to revisit the Hough transform in a continuous way. Many theoretical questions remain open, e.g study the separation conditions under which perfect reconstruction can be guaranteed. We should also investigate the possibility to relax the periodicity assumption, and from a practical point of view, parallel computing would be welcome to speed up the proposed algorithm, and to

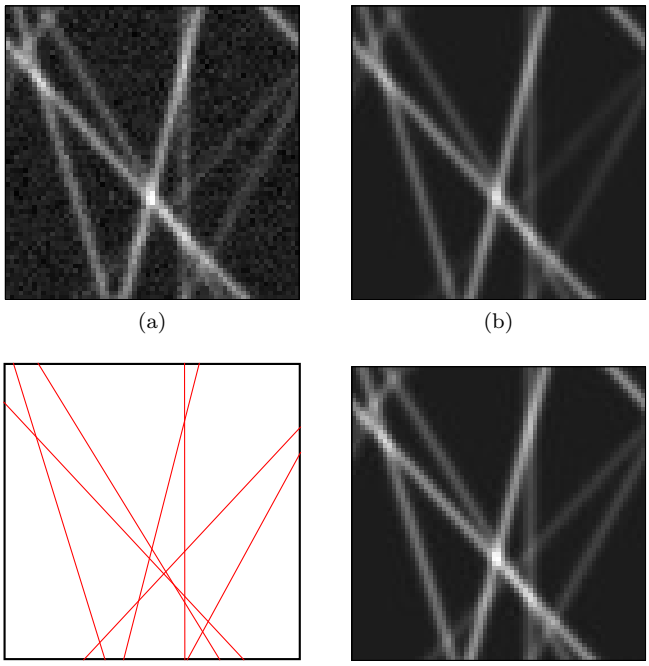


FIG. 9. *(a) Image* $\mathbf{y}$ *of seven lines affected by noise* ($\zeta = 20$)*, (b) Result* $\tilde{\mathbf{x}}$ *of the denoising step from the optimization, (c) Extraction of the line parameters from the Prony step, (d) Theoritical solution* $\hat{\mathbf{x}}^\sharp$ *to compare with.*

pretend to an industrial use.

Appendix A. Proof of Proposition 1 .

THEOREM 5 (Nyquist Whittaker-Shannon). *Let* f *be* T*-periodic function and* $c_m(f) = 0$ *for* $|m| \geqslant M+1$*, then* f *can be reconstructed from the regular sampling* $\{f(ka), k = 0, 1, ..., 2M\}$*, with* $a = \frac{T}{2M+1}$ *the sampling rate, in this way:*

$$f(x) = \frac{1}{2M+1} \sum_{k=0}^{2M} f(ka) D_M\left(\frac{2\pi}{T}(x-ka)\right) , \tag{77}$$

where D_M *is the Dirichlet Kernel:*

$$D_M(x) = \sum_{m=-M}^{M} \mathrm{e}^{\mathrm{j}mx} = \frac{\sin\left((M+\frac{1}{2})x\right)}{\sin\frac{x}{2}} . \tag{78}$$

Proof. The function f is T–periodic and has a finite number of Fourier coefficients, hence its Fourier serie is written:

$$f(x) = \sum_{m=-M}^{M} c_m(f) \mathrm{e}^{\mathrm{j}2\pi\frac{m}{T}x} . \tag{79}$$

For $p \in \{-M, ..., M\}$, let us compute the following sum:

$$\begin{aligned}(80)\qquad \sum_{k=0}^{2M} f(ka)\mathrm{e}^{-\mathrm{j}2\pi\frac{p}{T}ka} &= \sum_{k=0}^{2M}\left[\sum_{m=-M}^{M} c_m(f)\mathrm{e}^{\mathrm{j}2\pi\frac{m}{T}ka}\right]\mathrm{e}^{-\mathrm{j}2\pi\frac{p}{T}ka}\ ,\\ &= \sum_{m=-M}^{M} c_m(f)\left[\sum_{k=0}^{2M}\exp\left(\mathrm{j}2\pi\frac{k}{2M+1}(m-p)\right)\right]\ ,\\ &= (2M+1)c_p(f)\ .\end{aligned}$$

The discrete Fourier transform of the samples corresponds to the Fourier coefficient of f. More precisely, if we define $\boldsymbol{g}[k] = f(ka)$ and the discrete Fourier transform

$$(81)\qquad \hat{\boldsymbol{g}}[k] = \frac{1}{2M+1}\sum_{i=0}^{2M} g[i]\mathrm{e}^{-\mathrm{j}2\pi\frac{ik}{2M+1}}\ .$$

then

$$(82)\qquad (\boldsymbol{g}[0], \boldsymbol{g}[1], \ldots, \boldsymbol{g}[2M]) \overset{\mathrm{DFT}}{\longleftrightarrow} (\hat{\boldsymbol{g}}[0], \hat{\boldsymbol{g}}[1], \ldots, \hat{\boldsymbol{g}}[2M]) = (c_{-M}(f), \ldots, c_M(f))\ .$$

Finally, we prove the reconstruction formula:

$$\begin{aligned}(83)\qquad f(x) &= \sum_{m=-M}^{M}\left[\frac{1}{2M+1}\sum_{k=0}^{2M} f(ka)\exp\left(-\mathrm{j}2\pi\frac{m}{T}ka\right)\right]\exp\left(\mathrm{j}2\pi\frac{m}{T}x\right)\ ,\\ &= \frac{1}{2M+1}\sum_{k=0}^{2M} f(ka)\left[\sum_{m=-M}^{M}\exp\left(\mathrm{j}\frac{2\pi}{T}(x-ka)\right)\right]\ ,\\ &= \frac{1}{2M+1}\sum_{k=0}^{2M} f(ka)D_M\left(\frac{2\pi}{T}(x-ka)\right)\ ,\end{aligned}$$

which concludes the proof. □

In the present case, φ_1 is W–periodic and $c_m(\varphi_1) = 0$ for $|m| \geqslant M+1$ with $2M+1 = W$, so $a = 1$ and applying the Nyquist theorem:

$$(84)\qquad \varphi_1(t) = \sum_{n=0}^{W-1}\varphi_1(n)\frac{\sin\pi(x-n)}{W\sin\left(\frac{\pi(x-n)}{W}\right)}\ .$$

Consequently we give this explicit formula $\boldsymbol{g}[n] = \varphi_1(n)$ and $\hat{\boldsymbol{g}}[n] = c_{n-M}$ for $n = 0, 1, ..., 2M$. In order to deal with gaussian samples of φ_1 centered at the origin, we prefer this equivalent formula:

$$(85)\qquad \varphi_1(t) = \sum_{n=-M}^{M}\varphi_1(n)\frac{\sin\pi(x-n)}{W\sin\left(\frac{\pi(x-n)}{W}\right)}\ ,$$

since a vector of samples shifted of $M \in \mathbb{Z}$ has the DFT:

$$(86)\qquad (\boldsymbol{g}[n-M])_{n=0,\ldots,2M} \quad \overset{\mathrm{DFT}}{\longleftrightarrow} \quad (\mathrm{e}^{-\mathrm{j}2\pi\frac{Mm}{2M+1}}\hat{\boldsymbol{g}}[m])_{m=0,\ldots,2M}\ ,$$

so in the second line of (83) by remplacing $c_m(f)$ by $\mathrm{e}^{+\mathrm{j}2\pi\frac{Mm}{2M+1}}c_m(f)$ which is equivalent to translate of $+M$ in the Dirichlet Kernel:

$$\varphi_1(t) = \sum_{n=0}^{2M} \varphi_1(n-M)\frac{\sin\pi(x-n+M)}{W\sin\left(\frac{\pi(x-n+M)}{W}\right)}\ , \tag{87}$$

we recover the result by the change of variable $n \leftarrow n-M$.

Appendix B. Proof of Proposition 3.
Denote the right-hand side of (28) by $\mathrm{SDP}(\boldsymbol{z})$.

• Suppose $\boldsymbol{z} = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k,\phi_k)$ with $c_k > 0$.
Defining $\boldsymbol{q} = \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k)$, then $q_0 = \sum_{k=1}^{K} c_k$. The atoms $\boldsymbol{a}(f_k)$ has for components $[\boldsymbol{a}(f_k)]_i = \mathrm{e}^{\mathrm{j}2\pi f_k i}$ for $i = 0,\dots,N-1$, hence

$$\begin{aligned}
\mathbf{T}_N(\boldsymbol{a}(f_k)) &= \begin{pmatrix} 1 & \mathrm{e}^{-\mathrm{j}2\pi f_k} & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_k(N-1)} \\ \mathrm{e}^{\mathrm{j}2\pi f_k} & 1 & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_k(N-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{e}^{\mathrm{j}2\pi f_k(N-1)} & \mathrm{e}^{\mathrm{j}2\pi f_k(N-2)} & \dots & 1 \end{pmatrix}\ , \\
&= \begin{pmatrix} 1 \\ \mathrm{e}^{\mathrm{j}2\pi f_k} \\ \vdots \\ \mathrm{e}^{\mathrm{j}2\pi f_k(N-1)} \end{pmatrix} \begin{pmatrix} 1 & \mathrm{e}^{-\mathrm{j}2\pi f_k} & \dots & \mathrm{e}^{-\mathrm{j}2\pi f_k(N-1)} \end{pmatrix}\ , \\
&= \boldsymbol{a}(f_k)\boldsymbol{a}(f_k)^*\ .
\end{aligned}$$

We deduce that

$$\begin{aligned}
\mathbf{T}_N(\boldsymbol{q}) &= \sum_{k=1}^{K} c_k \mathbf{T}(\boldsymbol{a}(f_k))\ , \\
&= \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k)\boldsymbol{a}(f_k)^*\ , \\
&= \sum_{k=1}^{K} c_k \boldsymbol{a}(f_k,\phi_k)\boldsymbol{a}(f_k,\phi_k)^*\ .
\end{aligned}$$

Therefore, the matrix

$$\begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix} = \sum_{k=1}^{K} \begin{pmatrix} \boldsymbol{a}(f_k,\phi_k) \\ 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{a}(f_k,\phi_k) \\ 1 \end{pmatrix}^*\ ,$$

is positive semidefinite. Given $q_0 = \sum_{k=1}^{K} c_k$, we get $\mathrm{SDP}(\boldsymbol{z}) \leqslant \sum_{k=1}^{K} c_k$.
Since this holds for any decomposition of $\boldsymbol{z}$, we conclude that $\mathrm{SDP}(\boldsymbol{z}) \leqslant \|\boldsymbol{z}\|_{\mathcal{A}}$.

• Conversely, let $\boldsymbol{q} \in \mathbb{C}^N$ a vector such that $q_0 \geqslant 0$ and $\begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix} \succcurlyeq 0$. In particular we have $\mathbf{T}_N(\boldsymbol{q}) \succcurlyeq 0$. We denote by r the rank of $\mathbf{T}_N(\boldsymbol{q})$. Theorem 2 insures that $\mathbf{T}_N(\boldsymbol{q}) \succcurlyeq 0$ and of rank $r \leqslant N$ if and only if there exists $d_k > 0$ and distincts f_k,

such that

$$\boldsymbol{q} = \sum_{k=1}^{r} d_k \boldsymbol{a}(f_k) \;,$$

Let us denote matrices $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_r)$ and

$$\mathbf{V} = \begin{pmatrix} \boldsymbol{a}(f_1) & \cdots & \boldsymbol{a}(f_r) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mathrm{e}^{\mathrm{j}2\pi f_1} & \mathrm{e}^{\mathrm{j}2\pi f_2} & \cdots & \mathrm{e}^{\mathrm{j}2\pi f_r} \\ \mathrm{e}^{\mathrm{j}2\pi f_1 2} & \mathrm{e}^{\mathrm{j}2\pi f_2 2} & \cdots & \mathrm{e}^{\mathrm{j}2\pi f_r 2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{e}^{\mathrm{j}2\pi f_1 (N-1)} & \mathrm{e}^{\mathrm{j}2\pi f_2 (N-1)} & \cdots & \mathrm{e}^{\mathrm{j}2\pi f_r (N-1)} \end{pmatrix} .$$

By linearity of the operator $\mathbf{T}_N$:

$$\begin{aligned} \mathbf{T}_N(\boldsymbol{q}) &= \sum_{k=1}^{r} d_k \mathbf{T}_N(\boldsymbol{a}(f_k)) \;, \\ &= \sum_{k=1}^{r} d_k \boldsymbol{a}(f_k) \boldsymbol{a}(f_k)^* \;, \\ &= \mathbf{V}\mathbf{D}\mathbf{V}^* \;. \end{aligned}$$

Since $\mathbf{T}_N(\boldsymbol{a}(f_k))$ contains only ones on the diagonal, we have

$$\frac{1}{N}\mathrm{Tr}(\mathbf{T}_N(\boldsymbol{q})) = \sum_{k=1}^{r} d_k > 0 \;.$$

Besides, $\frac{1}{N}\mathrm{Tr}(\mathbf{T}_N(\boldsymbol{q})) = q_0$, therefore $q_0 > 0$.

Let be $\mathbf{M}$ a general block matrix $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{C} \end{pmatrix}$, the Schur complement gives

$$[\, \mathbf{C} \succ 0 \Rightarrow \mathbf{M} \succcurlyeq 0 \,] \quad \Longrightarrow \quad [\, \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} \succcurlyeq 0 \,] \;.$$

We apply this lemma to $\mathbf{M} = \begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix}$, with $\mathbf{A} = \mathbf{T}_N(q)$, $\mathbf{B} = z$ and $\mathbf{C} = q_0$. The left term is verified by hypothesis, hence

$$\mathbf{T}_N(\boldsymbol{q}) - q_0^{-1}\boldsymbol{z}\boldsymbol{z}^* \succcurlyeq 0 \quad \Longleftrightarrow \quad \mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\boldsymbol{z}\boldsymbol{z}^* \succcurlyeq 0 \;.$$

We define the square matrix $\mathbf{V}_r$ by extracting the r first rows of $\mathbf{V}$, which is a Vandermonde matrix, whose determinant is $\det(\mathbf{V}_r) = \prod_{1 \leqslant k < l \leqslant r} (f_l - f_k)$. Since we assumed $f_k \neq f_l$, $\mathbf{V}_r$ is invertible, and $\mathrm{rank}(\mathbf{V}) = r$. Let define $\boldsymbol{v} : \mathbb{C}^r \to \mathbb{C}^N$ and $\boldsymbol{v}^* : \mathbb{C}^N \to \mathbb{C}^r$ the endomorphisms corresponding to matrices $\mathbf{V}$ and $\mathbf{V}^*$. We have $\mathrm{rank}(\boldsymbol{v}^*) = \mathrm{rank}(\boldsymbol{v}) = r$. By the rank-nullity theorem:

$$\mathrm{rank}(\boldsymbol{v}^*) + \dim(\mathrm{Ker}\ \boldsymbol{v}^*) = \dim(\mathbb{C}^N) \quad \Longrightarrow \quad \dim(\mathrm{Ker}\ \boldsymbol{v}^*) = N - r \;.$$

Thus, there exists a vector $\boldsymbol{p} \in \mathbb{C}^N$ such that $\mathbf{V}^*\boldsymbol{p} = 0 \Leftrightarrow \boldsymbol{p}^*\mathbf{V} = 0$. Consequently,

$$\begin{aligned}\boldsymbol{p}^*(\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\boldsymbol{z}\boldsymbol{z}^*)\boldsymbol{p} \geqslant 0 &\Leftrightarrow (\boldsymbol{p}^*\mathbf{V})\mathbf{D}(\mathbf{V}^*\boldsymbol{p}) - q_0^{-1}\boldsymbol{p}^*\boldsymbol{z}\boldsymbol{z}^*\boldsymbol{p} \geqslant 0\ ,\\ &\Leftrightarrow q_0^{-1}\|\boldsymbol{p}^*\boldsymbol{z}\|_2^2 \leqslant 0\ ,\\ &\Leftrightarrow \|\boldsymbol{p}^*\boldsymbol{z}\|_2^2 = 0\ ,\\ &\Leftrightarrow \boldsymbol{p}^*\boldsymbol{z} = 0\ ,\\ &\Leftrightarrow \boldsymbol{p} \perp \boldsymbol{z}\ .\end{aligned}$$

Since $\boldsymbol{p} \in \operatorname{Ker} \boldsymbol{v}^*$, then $\boldsymbol{z} \in (\operatorname{Ker} \boldsymbol{v}^*)^\perp = \operatorname{Im} \boldsymbol{v}$, so it exists a vector $\boldsymbol{w} \in \mathbb{C}^r$ such that $\boldsymbol{z} = \mathbf{V}\boldsymbol{w} = \sum_{k=1}^r w_k \boldsymbol{a}(f_k)$, hence

$$\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{V}\boldsymbol{w}\boldsymbol{w}^*\mathbf{V}^* \succcurlyeq 0\ .$$

Besides, $\operatorname{Im} \boldsymbol{v}^* \subset \mathbb{C}^r$ and $\dim(\operatorname{Im} \boldsymbol{v}^*) = \operatorname{rank}(\boldsymbol{v}^*) = r = \dim(\mathbb{C}^r)$, thus $\operatorname{Im} \boldsymbol{v}^* = \mathbb{C}^r$ and $\boldsymbol{v}^*$ is surjective. Consequently, it exists a vector $\boldsymbol{u} \in \mathbb{C}^N$ such that $\mathbf{V}^*\boldsymbol{u} = \operatorname{sign}(\boldsymbol{w}) = (w_1/|w_1|, \dots, w_r/|w_r|)^\mathsf{T}$, and

$$\begin{aligned}\boldsymbol{u}^*(\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{V}\boldsymbol{w}\boldsymbol{w}^*\mathbf{V}^*)\boldsymbol{u} \geqslant 0 &\Leftrightarrow (\boldsymbol{u}^*\mathbf{V})\mathbf{D}(\mathbf{V}^*\boldsymbol{u}) - q_0^{-1}(\boldsymbol{u}^*\mathbf{V})\boldsymbol{w}\boldsymbol{w}^*(\mathbf{V}^*\boldsymbol{u}) \geqslant 0\ ,\\ &\Leftrightarrow \operatorname{sign}(\boldsymbol{w})^*\mathbf{D}\operatorname{sign}(\boldsymbol{w}) - q_0^{-1}\operatorname{sign}(\boldsymbol{w})^*\boldsymbol{w}\boldsymbol{w}^*\operatorname{sign}(\boldsymbol{w}) \geqslant 0\ ,\\ &\Leftrightarrow \sum_{k=1}^r d_k \left|\frac{w_k}{|w_k|}\right|^2 - q_0^{-1}\left(\sum_{k=1}^r \frac{\bar{w_k}}{|w_k|}w_k\right)^2 \geqslant 0\ ,\\ &\Leftrightarrow q_0^2 \geqslant \left(\sum_{k=1}^r |w_k|\right)^2\ ,\\ &\Leftrightarrow q_0 \geqslant \sum_{k=1}^r |w_k| \geqslant \|\boldsymbol{z}\|_{\mathcal{A}}\ .\end{aligned}$$

Taking the infimum leads to $\mathrm{SDP}(\boldsymbol{z}) \geqslant \|\boldsymbol{z}\|_{\mathcal{A}}$.

Finally, let us show that the infimum of the linear form $\ell : \boldsymbol{q} \mapsto q_0$ is achieved on

$$A = \left\{\boldsymbol{q} \in \mathbb{C}^N : \begin{pmatrix}\mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z}\\ \boldsymbol{z}^* & q_0\end{pmatrix} \succcurlyeq 0\right\}\ .$$

Let notice that ℓ defines a norm on A, since $\boldsymbol{q} \in A$ implies $\mathbf{T}_N(\boldsymbol{q}) \succcurlyeq 0$ and then

$$q_0 = \frac{1}{N}\operatorname{Tr}(\mathbf{T}_N(\boldsymbol{q})) = \sum_{i=0}^{N-1} \lambda_i \geqslant 0\ ,$$

with λ_i the eigenvalues of $\mathbf{T}_N(\boldsymbol{q})$ which are positives.
A is non empty, since $\boldsymbol{q} = (\|\boldsymbol{z}\|_2, 0, \dots, 0)^\mathsf{T} \in A$. Indeed for a vector $\boldsymbol{z} = (z_0, \dots, z_{N-1}) \in \mathbb{C}^N$ fixed, $\boldsymbol{v} = (v_0, \dots, v_N) \in \mathbb{C}^{N+1}$ and $\boldsymbol{v}' = (v_0, \dots, v_{N-1}) \in \mathbb{C}^N$ we have for this $\boldsymbol{q}$

:

$$\begin{aligned}
\boldsymbol{v}^* \begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix} \boldsymbol{v} &= \|\boldsymbol{z}\|_2 \|\boldsymbol{v}\|_2^2 + 2\,\mathrm{Re}\left(v_N \sum_{i=0}^{N-1} z_i \bar{v}_i \right) , \\
&\geqslant \|\boldsymbol{z}\|_2 \|\boldsymbol{v}\|_2^2 - 2|v_N||\langle \boldsymbol{z}, \boldsymbol{v}' \rangle| , \\
&\geqslant \|\boldsymbol{z}\|_2 \|\boldsymbol{v}\|_2^2 - 2|v_N|\|\boldsymbol{z}\|_2 \|\boldsymbol{v}'\|_2 , \\
&\geqslant \|\boldsymbol{z}\|_2 (\|\boldsymbol{v}'\|_2^2 - 2|v_N|\|\boldsymbol{v}'\|_2 + |v_N|^2) , \\
&\geqslant \|\boldsymbol{z}\|_2 (\|\boldsymbol{v}'\|_2 - |v_N|)^2 , \\
&\geqslant 0 .
\end{aligned}$$

Then $\boldsymbol{q} = (\|\boldsymbol{z}\|_2, 0, \ldots, 0)^\mathsf{T} \in A$ and $q_0 = \|\boldsymbol{z}\|_2$, hence $\|\boldsymbol{z}\|_{\mathcal{A}} \leqslant \|\boldsymbol{z}\|_2$.
Since norms are equivalent in finite dimension, the infimum has to be reached on the ball $\mathcal{B} = B(0, \|\boldsymbol{z}\|) \subset A$ with $r > 0$ and the norm ℓ. Thus $\mathcal{B}$ is bounded.
Considering a sequence $\left\{\boldsymbol{q}^{(p)}\right\}_{p\in\mathbb{N}}$ which belongs to $\mathcal{B}$. Let suppose it converges to $\boldsymbol{q}^*$.
By linearity in finite dimension of the application $\mathcal{T} : \boldsymbol{q} \mapsto \begin{pmatrix} \mathbf{T}_N(\boldsymbol{q}) & \boldsymbol{z} \\ \boldsymbol{z}^* & q_0 \end{pmatrix}$, we have that $\mathcal{T}$ is continuous, then $\mathcal{T}(\boldsymbol{q}^{(p)}) \xrightarrow[p\to+\infty]{} \mathcal{T}(\boldsymbol{q}^*)$. Since the sequence $\left\{\mathcal{T}(\boldsymbol{q}^{(p)})\right\}_{p\in\mathbb{N}}$ lying on the cone of positive matrix, which is closed, we deduce that the limit $\mathcal{T}(\boldsymbol{q}^*)$ is still positive. Therefore $\boldsymbol{q}^* \in \mathcal{B}$ and so $\mathcal{B}$ is closed.
Consequently, $\mathcal{B}$ is a compact, and the linear form ℓ achieves its minimum on $\mathcal{B}$.

Appendix C. Proof of Proposition 4.

The proof of the direct implication is straightforward. Let consider the converse.

By Theorem 2, since $\forall n, \mathbf{T}_M(\boldsymbol{l}_n) \succcurlyeq 0$ and of rank one, then it exists $c_n \geqslant 0$ and $f_n \in [0, 1]$ such that

$$\boldsymbol{l}_n[m] = c_n \exp(\mathrm{j}2\pi f_n m) .$$

Since we assume $\hat{\mathbf{x}}[0, n] = \hat{\mathbf{x}}[0, 0] = c_0$, then we actually have

$$\boldsymbol{l}_n[m] = c_0 \exp(\mathrm{j}2\pi f_n m) .$$

Let m be fixed. The Prony matrix $\mathbf{P}_1(\boldsymbol{t}_m)$ of size $2 \times (N-1)$

$$\mathbf{P}_1(\boldsymbol{t}_m) = \begin{pmatrix} \boldsymbol{t}_m[1] & \boldsymbol{t}_m[0] \\ \vdots & \vdots \\ \boldsymbol{t}_m[N-1] & \boldsymbol{t}_m[N-2] \end{pmatrix} ,$$

is of rank one, consequently there exists $\lambda_m \in \mathbb{C}$ such that

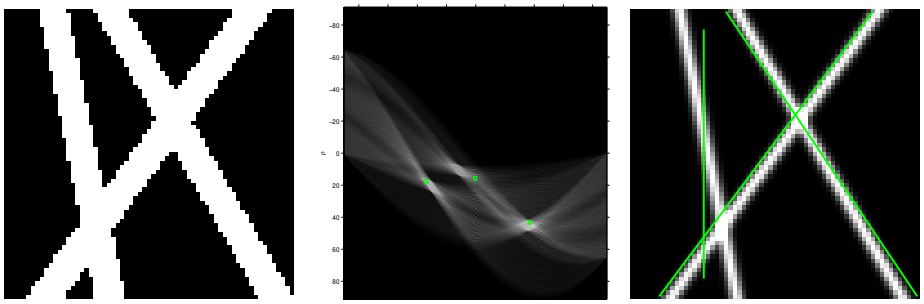
$$\boldsymbol{t}_m[n+1] = \lambda_m \boldsymbol{t}_m[n], \quad \forall\, 0 \leqslant n \leqslant N-1 .$$

Thus,

$$\boldsymbol{t}_m[n] = \lambda_m^n \boldsymbol{t}_m[0], \quad \forall\, 0 \leqslant n \leqslant N-1 .$$

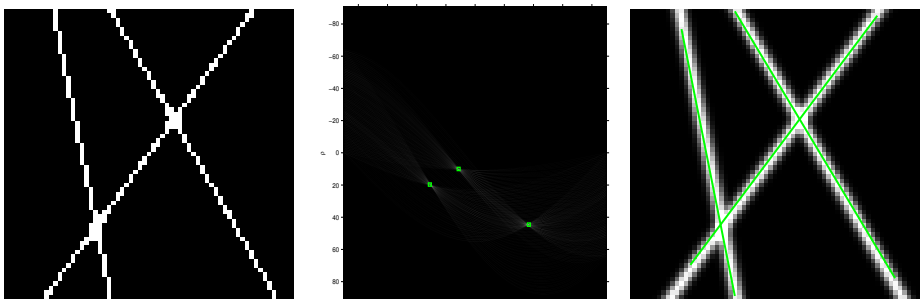
Notice that $\mathbf{P}_1(\boldsymbol{t}_m)$, being of rank one is equivalent to

$$\boldsymbol{t}_m[n] = d_m \exp(\mathrm{j}2\pi f_m n) ,$$



(a) (b) (c)

FIG. 10. *(a) The binary version of* $\mathbf{b}^\sharp$*, (b) the corresponding Hough transform with the estimaed peaks, (c) the estimated lines*



(a) (b) (c)

FIG. 11. *(a) The thresholded version of* $\mathbf{b}^\sharp$*, (b) the corresponding Hough transform with the estimaed peaks, (c) the estimated lines*

with $d_m \in \mathbb{C}$, where $d_m = \boldsymbol{t}_m[0]$ and $\lambda_m = \exp(\mathrm{j}2\pi f_m)$.
Due to the expression of $\boldsymbol{l}_n[m]$ previously obtained by Theorem 2, we get further:

$$\lambda_m = \frac{\boldsymbol{t}_m[1]}{\boldsymbol{t}_m[0]} = \frac{\ell_1[m]}{\ell_0[m]} = \exp(\mathrm{j}2\pi(f_1 - f_0)m) \ .$$

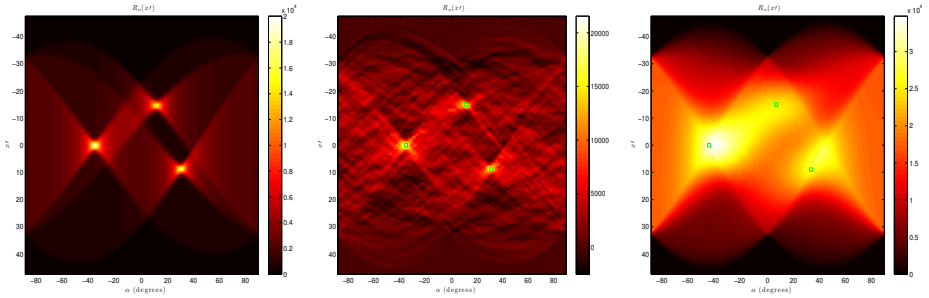
Therefore we have

$$\begin{aligned} \boldsymbol{t}_m[n] &= \lambda_m^n \boldsymbol{t}_m[0] \ , \\ &= \exp(\mathrm{j}2\pi(f_1 - f_0)m)^n c_0 \exp(\mathrm{j}2\pi f_0 m) \ , \\ &= c_0 \exp[\mathrm{j}2\pi((f_1 - f_0)n + f_0)m] \ . \end{aligned}$$

Appendix D. Comparison with the Hough and Radon transforms.

The Hough transform tackles the problem of detecting straight lines, converting this problem into peak detection in a parameter space. The transformation is done on a binary image, generally obtained after processing the original image by an edge detector. In our case, dealing with blurred image, this pre-processing is not relevant since the lines have some width. Thus, in order to convert the image to a binary one, we apply a certain threshold. We suggest these two possibilities:

• By calling the matlab routine IM2BW: This will apply a threshold on luminance of the image. We obtain the image Figure 10a, on which we perform the Hough



(a) (b) (c)

FIG. 12. *The Radon transform of the image* $\mathbf{y}$ *for experiment 1, 2 and 3. The theoretical parameters of lines are in green.*

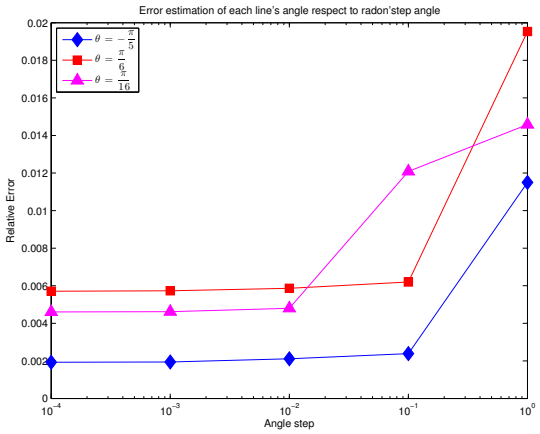


FIG. 13. *Variation of the accuracy with respect to the angle step of the Radon transform.*

transform, the resulting image is displayed in Figure 10b. The peaks detection (in green), and so the lines estimation (Figure 10c), altogether fail, due to some artifacts, which have been considered in [53], in which the authors propose a method to remove these artifacts. Nevertheless, as you can seen in Figure 10b, the peaks are in any case spread out because of the binary line's width, which provides a very coarse estimation of the parameters.

• By applying manually a threshold on the pixel values: This will keep the pixel greater than the threshold at stake. For instance, for a threshold of 230, we obtain the image Figure 11a. The peak detection is now working, and the lines estimations seems to be correct. This method provides an estimation with a mean error of 10^{-2} for θ_k, and 10^{-1} for η_k, which is not significantly improved by decreasing the discretization steps of the process. Remark that this method cannot be applied if the lines have different amplitude values, because of the threshold. For the same reason, it is not relevant for image with a large amount of blur or noise.

Therefore, we used instead the Radon transform, which appears to be more adapted for our topic.

The Radon transform is a famous mathematical tool used in tomography. The

Hough transform is actually related to the Radon transform, as a kind of discretization of the latter. The Radon transform can be done directly on grayscale images, and the peaks seem to be less spread than the peaks from the Hough transform, as we can see on Figure 12a. This approach has the advantage of achieving the same mean error of 10^{-1} for η_k, and a better rate of 10^{-3} for θ_k. Notice that even by decreasing the discretization steps of the process, we rapidly reach a plateau, as illustrated by Figure 13. Besides, the Radon transform is robust to noise, the peaks remain well detected, as you can observe in Figure 12b (b). However, this method fails for a large blur, such as presented in the third experiment, the peeks are not clearly recognizable anymore, as you observe in Figure 12c.

To sum up, the Radon transform is more suitable than the Hough transform for detecting lines in a degraded image. But, in both cases, they come up against a discrete limit as soon as we are looking for a greater accuracy. In contrary, our super-resolution method enables to achieve an infinite precision for line's parameters. Actually, our method is for the Hough transform, what the super-resolution is for the Fourier transform in 1D. Indeed, a spikes-train is a sum of exponentials in the Fourier domain; that is, the detection of spikes position is equivalent to detect the local minimum of the Fourier transform which are the frequencies of the corresponding exponentials. This latter technique applied on a degraded signal (blur and noise) provides a coarse approximation of the frequencies, where the super-resolution succeed in retrieving the true value under some conditions. Similarly, the Hough transform requires to detect the local maximum of the accumulator matrix, which provides a coarse approximation of the lines position, where our method overcomes this issue.

REFERENCES

[1] A. Barabell, *Improving the resolution performance of eigenstructure-based direction-finding algorithms*, in Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83., vol. 8, IEEE, 1983, pp. 336–339.

[2] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer Science & Business Media, 2011.

[3] B. N. Bhaskar, G. Tang, and B. Recht, *Atomic norm denoising with applications to line spectral estimation*, 61 (2013), pp. 5987–5999.

[4] T. Blu, *The generalized annihilation property: A tool for solving finite rate of innovation problems*, in SAMPTA'09, 2009, pp. Special–Session.

[5] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, *Sparse sampling of signal innovations*, IEEE Signal Processing Magazine, 25 (2008), pp. 31–40.

[6] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[7] Y. Bresler and A. Macovski, *Exact maximum likelihood parameter estimation of superimposed exponential signals in noise*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986), pp. 1081–1089.

[8] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Communications on Pure and Applied Mathematics, 67 (2014), pp. 906–956.

[9] C. Carathéodory, *Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen*, Rendiconti del Circolo Matematico di Palermo (1884-1940), 32 (1911), pp. 193–217.

[10] 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 (1884-1940), 32 (1911), pp. 218–239.

[11] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, Theoretical foundations and numerical methods for sparse recovery, 9 (2010), p. 227.

[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–

145.
[13] Y. Chan, J. Lavoie, and J. Plant, *A parameter estimation approach to estimation of frequencies of sinusoids*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 29 (1981), pp. 214–219.
[14] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, Foundations of Computational mathematics, 12 (2012), pp. 805–849.
[15] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, *A nonlocal structure tensor-based approach for multicomponent image recovery problems*, IEEE Transactions on Image Processing, 23 (2014), pp. 5531–5544.
[16] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, Springer, 2011, pp. 185–212.
[17] L. Condat, *A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.
[18] L. Condat, *Semi-local total variation for regularization of inverse problems.*, in EUSIPCO, 2014, pp. 1806–1810.
[19] L. Condat, *Atomic norm minimization for the decomposition of vectors into complex exponentials*, research report, GIPSA-Lab, 2016, Grenoble, France, (2017).
[20] L. Condat, J. Boulanger, N. Pustelnik, S. Sahnoun, and L. Sengmanivong, *A 2-d spectral analysis method to estimate the modulation parameters in structured illumination microscopy*, in 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, 2014, pp. 604–607.
[21] 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, 14 (2015), pp. p–17.
[22] S. R. Deans, *Hough transform from the radon transform*, IEEE transactions on pattern analysis and machine intelligence, (1981), pp. 185–188.
[23] D. L. Donoho, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.
[24] P. L. Dragotti, M. Vetterli, and T. Blu, *Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix*, IEEE Transactions on Signal Processing, 55 (2007), pp. 1741–1757.
[25] C. Fernandez-Granda, *Super-resolution and compressed sensing*, SIAM News, 46 (2013).
[26] Y. Hua and T. K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 38 (1990), pp. 814–824.
[27] J. Illingworth and J. Kittler, *A survey of the hough transform*, Computer vision, graphics, and image processing, 44 (1988), pp. 87–116.
[28] D. H. Johnson, *The application of spectral estimation methods to bearing estimation problems*, Proceedings of the IEEE, 70 (1982), pp. 1018–1028.
[29] S. Lefkimmiatis, J. P. Ward, and M. Unser, *Hessian schatten-norm regularization for linear inverse problems*, IEEE transactions on image processing, 22 (2013), pp. 1873–1888.
[30] M. D. Macleod, *Fast nearly ml estimation of the parameters of real or complex single tones or resolved multiple tones*, IEEE Transactions on Signal processing, 46 (1998), pp. 141–148.
[31] I. Maravic and M. Vetterli, *Sampling and reconstruction of signals with finite rate of innovation in the presence of noise*, IEEE Transactions on Signal Processing, 53 (2005), pp. 2788–2805.
[32] I. Markovsky, *Low rank approximation: algorithms, implementation, applications*, Springer Science & Business Media, 2011.
[33] F. Marvasti, A. Amini, F. Haddadi, M. Soltanolkotabi, B. H. Khalaj, A. Aldroubi, S. Sanei, and J. Chambers, *A unified approach to sparse signal processing*, EURASIP journal on advances in signal processing, 2012 (2012), p. 1.
[34] P. Mukhopadhyay and B. B. Chaudhuri, *A survey of hough transform*, Pattern Recognition, 48 (2015), pp. 993–1010.
[35] L. M. Murphy, *Linear feature detection and enhancement in noisy images via the radon transform*, Pattern Recognition Letters, 4 (1986), pp. 279–284.
[36] N. Parikh, S. P. Boyd, et al., *Proximal algorithms.*, Foundations and Trends in optimization, 1 (2014), pp. 127–239.
[37] G. Peyré, S. Bougleux, and L. Cohen, *Non-local regularization of inverse problems*, in European Conference on Computer Vision, Springer, 2008, pp. 57–68.

[38] K. Polisano, L. Condat, M. Clausel, and V. Perrier, *Convex super-resolution detection of lines in images*, (2016).
[39] R. Prony, *Essai experimental–,-*, J. de l'Ecole Polytechnique, (1795).
[40] B. G. Quinn and E. J. Hannan, *The estimation and tracking of frequency*, vol. 9, Cambridge University Press, 2001.
[41] J. Radon, *1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten*, Classic papers in modern diagnostic radiology, 5 (2005).
[42] M. Rahman and K.-B. Yu, *Total least squares approach for frequency estimation using linear prediction*, Acoustics, Speech and Signal Processing, IEEE Transactions on, 35 (1987), pp. 1440–1454.
[43] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM review, 52 (2010), pp. 471–501.
[44] B. Recht, W. Xu, and B. Hassibi, *Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization*, in Decision and Control, 2008. CDC 2008. 47th IEEE Conference on, IEEE, 2008, pp. 3065–3070.
[45] R. Roy, A. Paulraj, and T. Kailath, *Esprit–a subspace rotation approach to estimation of parameters of cisoids in noise*, IEEE transactions on acoustics, speech, and signal processing, 34 (1986), pp. 1340–1342.
[46] R. Schmidt, *Multiple emitter location and signal parameter estimation*, IEEE transactions on antennas and propagation, 34 (1986), pp. 276–280.
[47] H.-C. So, K. W. Chan, Y.-T. Chan, and K. Ho, *Linear prediction approach for efficient frequency estimation of multiple real sinusoids: algorithms and analyses*, IEEE Transactions on Signal Processing, 53 (2005), pp. 2290–2305.
[48] P. Stoica, *List of references on spectral line analysis*, Signal Processing, 31 (1993), pp. 329–340.
[49] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, NJ, 2005.
[50] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, *Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 37 (1989), pp. 378–392.
[51] T. Strohmer, *Measure what should be measured: progress and challenges in compressive sensing*, IEEE Signal Processing Letters, 19 (2012), pp. 887–893.
[52] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, *Compressed sensing off the grid*, Information Theory, IEEE Transactions on, 59 (2013), pp. 7465–7490.
[53] X. Tao and A. Eades, *Errors, artifacts, and improvements in ebsd processing and mapping*, Microscopy and Microanalysis, 11 (2005), pp. 79–87.
[54] O. Toeplitz, *Zur theorie der quadratischen und bilinearen formen von unendlichvielen veränderlichen*, Mathematische Annalen, 70 (1911), pp. 351–376.
[55] D. W. Tufts and R. Kumaresan, *Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood*, Proceedings of the IEEE, 70 (1982), pp. 975–989.
[56] M. Unser, *Sampling-50 years after shannon*, Proceedings of the IEEE, 88 (2000), pp. 569–587.
[57] J. A. Urigiien, Y. C. Eldar, P. DRAGOTYI, et al., *Sampling at the rate of innovation: Theory and applications*, Compressed Sensing: Theory and Applications, (2012), p. 148.
[58] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM review, 38 (1996), pp. 49–95.
[59] H. P. VC, *Method and means for recognizing complex patterns*, Dec. 18 1962. US Patent 3,069,654.