AIMS Mathematics, 11(2): 4082–4097.
DOI:10.3934/math.2026164
Received: 18 November 2025
Revised: 17 January 2026
Accepted: 28 January 2026
Published: 10 February 2026

https://www.aimspress.com/journal/Math

Research article

Entropic inversion of Fourier transforms with incomplete data

Cécile Gauthier-Umaña[1], Valérie Gauthier-Umaña[2], Henryk Gzyl[3] and Enrique ter Horst[4,*]

[1] Department of Systems Engineering, Pontificia Universidad Javeriana, Bogota, Colombia
[2] Systems and Computing Engineering Department, Universidad de los Andes, Bogotá, Colombia
[3] Center for Finance, IESA School of Business, Caracas, Venezuela
[4] School of Management, Universidad de los Andes, Bogotá, Colombia

* **Correspondence:** Email: ve.gauthier@uniandes.edu.co.

Abstract: Here we provide an entropic method to invert the Fourier transform of a bounded function defined on an interval, when the data consists of its sine and cosine transforms, but not necessarily of consecutive frequencies. The classical direct approach consists in just forming the linear combination of the trigonometric functions of the given frequencies multiplied by their respective coefficients. The problem is that this approach does not yield any information about the projection of the function on the space spanned by the missing frequencies. Our approach consists of regarding the Fourier inversion as an ill-posed linear inverse problem with box constraints, consisting of finding a function given a few of its sine and cosine transforms. To solve this problem, we propose a non-linear approach, consisting of minimizing an entropy function subject to the Fourier data as constraints. This approach provides us with a solution that has a non-vanishing projection on the space spanned by the Fourier coefficients in the original data set, from which a better approximation to the unknown function can be recovered. In addition to obtaining an explicit representation of the solution, we prove that the solution converges to the unknown function as the number of data points increases. Even though the reconstruction procedure is non-linear in the data, there is some quasi-linearity in the procedure.

Keywords: Fourier inversion from real data; ill-posed linear problem with convex constraints; entropy minimization
Mathematics Subject Classification: 45Q05, 65N21, 65R32, 65T50

1. Introduction and preliminaries

Fourier analysis is one of the most widely used methods in signal analysis, allowing the decomposition of signals into their constituent frequency components. It is used in a wide range of areas such as signal and image processing, communications, spectral analysis, and the numerical

solution of partial differential equations [18, 20]. Some references covering a wide range of aspects are [4, 5, 10, 21, 23].

The application of Fourier analysis is subject to several limitations. The Nyquist criterion must be satisfied to prevent aliasing, while spectral leakage can appear when dealing with finite signals, and dependencies between time and frequency resolution influence the analysis of transient signals [1, 7, 14]. Furthermore, Gibbs artifacts and loss of reconstruction accuracy due to abrupt truncation or missing frequency components can be observed [6, 14]. Issues related to computational complexity are also well documented in the literature. For example, the Fast Fourier Transform (FFT) algorithm decreased complexity from $O(n^2)$ to $O(n \log n)$, allowing efficient analysis of both periodic and non-periodic signals [8, 20].

To address these limitations, other representations such as short-time Fourier analysis, filter banks, and wavelets have been developed to achieve better time-frequency representations [7, 19]. Additionally, the recovery performance for incomplete or noisy data is improved based on the regularization methods, such as maximum entropy spectral estimation and compressed sensing [1, 18].

The classical Fourier inversion has been well studied in classical analysis, see [10, 21]. Given a function $X : [0, 1] \to \mathbb{R}$, the conditions upon which the function can be approximated by the trigonometric series:

$$a_0 + \sqrt{2} \sum_{n=1}^{\infty} \Big(a_n \cos(2\pi n s) + b_n \sin(2\pi n s) \Big),$$

where

$$a_0 = \int_0^1 X(s)ds. \;\; a_n = \sqrt{2} \int_0^1 \cos(2\pi n s) X(s) ds,$$

$$b_n = \sqrt{2} \int_0^1 \sin(2\pi n s) X(s) ds, \quad n \geq 1$$

are quite well understood. Certainly, that is a solution to the problem. Geometrically speaking, the problem is that this is just the projection of the unknown function on the subspace of $L_2[0, 1]$ generated by $\{ \sqrt{2}cos(\omega_k s), \sqrt{2}sin(\omega_k s) : k = 1, ..., K\}$, and no information about the projection of the unknown function is available from that representation. In applications to image compression, say, it is supposed that the information at the missing frequencies is not important.

Our aim is to propose an alternative techniques that enhance signal recovery from partial data. For this, we restate the Fourier inversion problem as an inverse problem (or as a generalized moment problem), subject to box constraints upon the solution $X(s)$, for which the data (the generalized moments) are the given Fourier coefficients, but not necessarily at consecutive frequencies.

The entropic approach that we develop below will provide us with a solution that has the given Fourier coefficients, and has a non-zero projection on the space orthogonal to the one generated by the vectors $\{ \sqrt{2}cos(\omega_k s), \sqrt{2}sin(\omega_k s) : k = 1, ..., K\}$. This is an essential feature of our proposal.

To summarize the nature of previous work on the problem, let us rewrite the problem as consisting of: determine a vector X in some convex subset $\mathcal{K}$ of a Banach or Hilbert space $\boldsymbol{V}$ satisfying:

$$\mathfrak{A}X = \boldsymbol{y} \quad X \in \mathcal{K}, \tag{1.1}$$

where $\mathfrak{A} : \boldsymbol{V} \to \mathbb{R}^K$ is a given linear operator, and we suppose that the data vector $\boldsymbol{y}$ is in some K-dimensional linear space, and to be consistent, it is required of $\boldsymbol{y}$ to be in the range of $\mathfrak{A}$.

Problem (1.1) is a standard ill-posed problem with convex constraints. Historically, there are two basic types of solutions. One type proposes the following approach: Find X^* such that

$$X^* = arginf\{||X||^2 : X \in \mathcal{K};\ \mathfrak{A}X = \boldsymbol{y}\}. \tag{1.2}$$

Instead of using the data as a constraint, the other class of approaches restate the problem as: Find X^* such that

$$X^* = arginf\{||\mathfrak{A}X - \boldsymbol{y}||^2 : X \in \mathcal{K}\}. \tag{1.3}$$

Actually, since operator $\mathfrak{A}$ involves projections on an orthonormal family, (1.3) yields the classical solution, namely, the minimizer is the linear combination of the data points multiplied by the corresponding trigonometric functions. A list of references for these approaches is [11, 15, 22] and for emphasis on regularization [2, 16]. Useful as these approaches are in general, in our current case, in which $\mathfrak{A}$ involves an orthogonal projection, the approaches in these references are not very useful.

Distant cousins that use maximum entropy techniques are [9, 13] for mathematical details, [17, 25] for potential applications, and [24] for the use of entropy in statistics. And for a finite-dimensional, discrete version of the technique developed below, applied to a problem in decoding see [12].

2. Problem statement

In order to render the Fourier inversion problem into the form (1.1), let us first introduce the class of functions in which we seek the solution to our problem. Consider, to begin with, two continuous functions $l(s) < u(s)$ defined on $[0, 1]$. Put

$$\mathcal{K} = \{X : [0, 1] \to \mathbb{R} | l(s) \le X(s) \le u(s)\}, \tag{2.1}$$

for the class of measurable functions whose transform we are interested in. Let us introduce the following vector-valued transforms. Let $\boldsymbol{C} : \mathcal{K} \to \mathbb{R}^{K_1}$ and $\boldsymbol{S} : \mathcal{K} \to \mathbb{R}^{K_2}$ be defined by

$$\begin{gathered}\boldsymbol{C}X(k) \equiv \boldsymbol{\phi}_1(k) = \sqrt{2}\int_0^1 cos(\omega_k(1))s)X(s)ds, \\ \omega_1(1) < \omega_2(1) < \cdots < \omega_{K_1}(1).\end{gathered} \tag{2.2}$$

$$\begin{gathered}\boldsymbol{S}X(k) \equiv \boldsymbol{\phi}_2(k) = \sqrt{2}\int_0^1 sin(\omega_k(2)s)X(s)ds, \\ \omega_1(2) < \omega_2(2) < \cdots < \omega_{K_2}(2).\end{gathered} \tag{2.3}$$

The simplest, most common case, would be $\omega_1(1) = 0$ and $\omega_{k-1}(1) = \omega_k(2)$ for $k = 2, ..., K_2 = K_1 - 1$. For $k_1(1) = 0$, the factor $\sqrt{2}$ does not appear.

Let us set $K = K_1 + K_2$. Let us now introduce the symbols

$$\boldsymbol{A} = \begin{pmatrix}\boldsymbol{C} \\ \boldsymbol{S}\end{pmatrix},\ \ \boldsymbol{y} = \begin{pmatrix}\boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2\end{pmatrix}. \tag{2.4}$$

Here $\boldsymbol{A} = (\boldsymbol{C}, \boldsymbol{S})$ is the $K \times 1$ column vector of functions:

$$\boldsymbol{C}^t = \sqrt{2}(cos(\omega_1(1)s), ..., cos(\omega_{K_1}(1)s),$$

$$\boldsymbol{S}^t = \sqrt{2}(sin(\omega_1(2)s), ..., sin(\omega_{K_2}(2)s)),$$

or of you prefer a $K \times \infty$ matrix. It acts row wise on functions $X(s)$ the obvious way

$$\mathfrak{A}X = \int_0^1 \boldsymbol{A}(s)X(s)ds = \int_0^1 \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{S} \end{pmatrix} X(s)ds, \tag{2.5}$$

when it acts on a function $X(s)$ (or on a $\infty \times 1$ vector), it produces a $K \times 1$ vector whose components are given by (2.2) and (2.3).

With these notations, our problem consists of

$$\text{Find} \quad X \in \mathcal{K} \quad \text{such that} \quad \mathfrak{A}X = \boldsymbol{y}. \tag{2.6}$$

Having established the necessary notations, instead of the customary methods mentioned in (1.2) and (1.3), we will define a strong convex function Ψ on $\mathcal{K}$ and consider the following constrained minimization problems: Find $X^* \in \mathcal{K}$ such that

$$X^* = argin f\{\Psi(X)|\mathfrak{A}X - \boldsymbol{y}\}. \tag{2.7}$$

Note that from a general perspective of view, the standard square-norm minimization in (1.3) is just another version of the same technique. As mentioned above, (1.3) yields the classical solution, which is orthogonal to the space generated by the remaining trigonometric functions functions that are not included in the data set. In contrast, proposal in (2.6) yields a solution not orthogonal to the remaining trigonometric functions. This property is exploited to improve the reconstruction of the function below.

At this point, we outline the remainder of this paper. In Section 3, we describe the objective function that we minimize as an alternative to the traditional approach. That function is similar to the entropy of the Fermi-Dirac type.

Next, we establish the properties of the Bregman divergence derived from the objective function. These will allow us to prove that as the amount of data increases, the solutions tend to the true solution of the problem. This is an analogue of the classical speed of convergence property.

In Section 4, we show how to represent the solution explicitly in terms of the Lagrange multipliers that come up in the variational problem. The procedure is vaguely reminiscent of the method of maximum entropy in the mean. See [9, 13] for details. However, it differs from these approaches in several essential ways. The essential difference is that the algebraic problem does not have to be transformed into a problem consisting of determining an auxiliary probability estimation problem, which is solved by the usual method of maximum entropy. In our approach, the entropy is defined directly on the space of solutions

In Section 5, we prove that as the size of the data increases, the solutions tend to the unknown function. We use the results in Section 3 to prove that convergence in the Fermi-Dirac entropy implies convergence in the $L_1[0, 1]$ norm as the amount of data increases. The numerical experiments are carried out in Section 6, including a quasi-linearity in the Bregman divergence test. We close with a few remarks in Section 7.

3. The properties of the objective function

Let us begin by collecting the basic properties of the entropy-like function $\Psi(X)$. The class $\mathcal{K}$ introduced above can be regarded as a convex set in Banach space $\mathcal{B}$ of all bounded functions on $[a, b]$. Define $\Psi(X) : \mathcal{K} \to \mathbb{R}$ as follows:

$$\Psi(X) = \int_0^1 \left[\left(\frac{X(s)-l(s)}{(u(s)-l(s)}\right)\ln\left(\frac{X(s)-l(s)}{u(s)-l(s)}\right) + \left(\frac{u(s)-X(s)}{u(s)-l(s)}\right)\ln\left(\frac{u(s)-X(s)}{u(s)-l(s)}\right)\right]ds. \tag{3.1}$$

The following properties drop out from the similar properties for the one-variable function:

$$\psi(x) = \frac{x-l}{u-l}\ln\frac{x-l}{u-l} + \frac{u-x}{u-l}\ln\frac{u-x}{u-l}$$

defined on the interval $[l, u]$. Note that we can interpret $p = (x-l)/(u-l)$ as the probability of choosing a point at random from $[l, u]$, therefore $\psi(x) = p\ln p + (1-p)\ln(1-p)$ can be thought of as the entropy of a binary random variable.

Theorem 3.1. *The function $\Psi(X)$ is strongly convex, infinitely differentiable, with Frechet derivative*

$$\frac{\delta\Psi}{\delta X}(s) = \frac{1}{u(s)-l(s)}\ln\left(\frac{X(s)-l(s)}{b(s)-X(s)}\right). \tag{3.2}$$

Furthermore, let $\tau(s) : [a, b] \to \mathbb{R}$ be (for the time being) a measurable function. Then the equation

$$\frac{\delta\Psi}{\delta X}(s) = \tau(s)$$

is invertible, with inverse

$$X(s) = \left(\frac{\delta\Psi}{\delta X}\right)^{-1}(\tau(s)) = \frac{l(s)e^{l(s)\tau(s)} + u(s)e^{u(s)\tau(s)}}{e^{l(s)\tau(s)} + e^{u(s)\tau(s)}}. \tag{3.3}$$

For the optimization part, it will prove convenient to know that the Lagrange -Fenchel conjugate (dual) of Ψ, can be explicitly calculated. Let $\tau(s) : [a, b] \to \mathbb{R}$ be an integrable function. Then, the Lagrange-Fenchel dual $M(\tau)$ of Ψ is defined by

$$M(\tau) = \int_0^1 \ln\left(e^{l(s)\tau(s)} + e^{u(s)\tau(s)}\right)ds. \tag{3.4}$$

We have

Theorem 3.2. *The function $M(\tau)$ is strongly convex in τ, infinitely differentiable, and*

$$\frac{\delta M}{\delta\tau}(s) = \frac{l(s)e^{l(s)\tau(s)} + u(s)e^{u(s)\tau(s)}}{e^{l(s)\tau(s)} + e^{u(s)\tau(s)}} = \left(\frac{\delta\Psi}{\delta X}\right)^{-1}(\tau(s)). \tag{3.5}$$

That is $\frac{\delta\Psi}{\delta X}$ and $\frac{\delta M}{\delta\tau}$ are compositional inverses of each other.

For this and for details about Bregman divergences mentioned next, the reader may consult Borwein and Lewis [3].

Definition 3.1. *Let $X, Y \in \mathcal{K}$, and let $\Psi(X)$ be defined as in* (3.1)*. The Ψ-Bregman divergence $D_\Psi(X, Y)$ between X and Y is defined by*

$$D_\Psi(X, Y) = \Psi(X) - \Psi(Y) - \int_0^1 (X(s) - Y(s))\frac{\delta\Psi}{\delta Y}(s)ds. \tag{3.6}$$

Again, a computation based on the one-variable case yields:

$$D_\Psi(X,Y) = \int_0^1 \Big(\frac{X(s)-l(s)}{u(s)-l(s)} \ln \frac{X(s)-l(s)}{Y(s)-l(s)} + \frac{u(s)-X(s)}{u(s)-l(s)} \ln \frac{u(s)-X(s)}{u(s)-Y(s)} \Big) ds. \tag{3.7}$$

Below, we make use of

Theorem 3.3. *Let $X, Y \in \mathcal{K}$, and $D_\Psi(X,Y)$ as in* (3.6)*. Then*

$$D_\Psi(X,Y) \geq 0. \quad \textit{and} \quad D_\Psi(X,Y) = 0 \quad \Leftrightarrow \quad X = Y. \tag{3.8}$$

$$D_\Psi(X,Y) \geq C\|X-Y\|_1^2. \tag{3.9}$$

Here $\|X-Y\|_1$ is the L_1-norm of $X-Y$, and c is a positive constant which is specified below.

Proof. Consider the convex function

$$\psi(x) : [l,u] \to \mathbb{R} :$$

$$\psi(x) = \frac{x-l}{u-l} \ln \frac{x-l}{u-l} + \frac{u-x}{u-l} \ln \frac{u-x}{u-l}.$$

The following identity holds for any twice continuously differentiable function, and any $x, y \in [l,u]$:

$$\phi(x) - \phi(y) - (x-y)\phi'(y) = \int_y^x (x-t)\phi''(t)dt.$$

Note that the left-hand side gives rise to D_Ψ when x, y are replaced by $x(s), Y(s)$, and integrated over $[a,b]$. Since $\phi''(t) \geq 0$ due to the convexity of ϕ, assertion (3.8) is clear. Observe that

$$\phi''(t) = 1/(t-l)(u-t) \geq 4/(u-l)^2.$$

Inserting in the identity two lines above, we obtain

$$\int_y^x (x-t)\phi''(t)dt \geq \int_y^x \frac{x-t}{(u-l)^2} dt = \frac{2}{(u-l)^2}(x-y)^2.$$

Now, replace, respectively, $x, y, \phi'(t)$ by $X(s), Y(s), \frac{\delta\Psi}{\delta Y}(s)$ and u, l by $u(s), l(s)$ and integrate over $s \in [a,b]$ to obtain

$$D_\Psi(X,Y) \geq 2 \int_0^1 \frac{\big(X(s)-Y(s)\big)^2}{\big(u(s)-l(s)\big)^2} ds \geq \frac{2}{B^2} \int_0^1 \big(X(s)-Y(s)\big)^2 ds. \tag{3.10}$$

Here we introduced

$$(u(s)-l(s)) \leq B \equiv \sup(u(s)-b(s)).$$

Now invoke Cauchy-Schwartz inequality to obtain

$$\int_0^1 \big(X(s)-Y(s)\big)^2 ds \geq \Big(\int_0^1 |X(s)-Y(s)| ds \Big)^2 = \|X-Y\|_1^2. \tag{3.11}$$

This is (3.9) after identifying $c = 2/B^2$. □

4. The entropic solution of the inverse problem

We have already put in place all the ingredients to invert the Fourier transform from partial data, that is, to solve (2.6), that is

$$\text{Find} \quad X \in \mathcal{K} \quad \text{such that} \quad \mathfrak{A}X = \boldsymbol{y}.$$

Instead of the square norm minimization approaches like (1.2) and (1.3), we propose the following variational approach:

$$\text{Find} \quad X^* = argin f\{\Psi(X) | X \in \mathcal{K};\ \mathfrak{A}X = \boldsymbol{y}\}. \tag{4.1}$$

To take care of the constraints, we introduce the auxiliary Lagrangian function:

$$\mathcal{L}(X, \boldsymbol{\lambda}) : \mathcal{K} \times \mathbb{R}^K \to \mathbb{R}$$

defined as follows:

$$\mathcal{L}(X, \boldsymbol{\lambda}) = \Psi(X) - \langle \boldsymbol{\lambda}, (\mathfrak{A}X - \boldsymbol{y})\rangle. \tag{4.2}$$

This function is strongly convex, differentiable in X, and linear (thus convex and concave) in $\boldsymbol{\lambda}$. Thus, we consider the following optimization problem: Find $(X^*, \boldsymbol{\lambda}^*)$ at which

$$\sup_{\boldsymbol{\lambda}} \inf_{X} \{\mathcal{L}(X, \boldsymbol{\lambda}) | (X, \boldsymbol{\lambda}) : \mathcal{K} \times \mathbb{R}^K \to \mathbb{R}\} \tag{4.3}$$

is achieved. We begin by establishing the first-order conditions for $(X^*, \boldsymbol{\lambda}^*)$ to be a minimizer. Since X is a composite variable, we actually have a composite system of equations to solve. With an eye on (3.2), we have

$$\frac{\delta \mathcal{L}}{\delta X}(s) = \frac{\delta \Psi}{\delta X}(s) = \frac{1}{u(s) - l(s)} \ln\left(\frac{X^*(s) - l(s)}{u(s) - X^*(s)}\right) = \boldsymbol{A}^t \boldsymbol{\lambda}^*(s). \tag{4.4}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = (\boldsymbol{A}X^*)_j - y_j = 0. \tag{4.5}$$

Once the Lagrange multiplier $\boldsymbol{\lambda}^*$ the first two equations in the system yield (recall (3.3):

$$X^*(s) = \frac{l(s)e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}^*(s)} + u(s)e^{u(s)\boldsymbol{A}^t\boldsymbol{\lambda}^*(s)}}{e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}^*(s)} + e^{u(s)\boldsymbol{A}^t\boldsymbol{\lambda}^*(s)}}. \tag{4.6}$$

Here, we stress that $\boldsymbol{A}^t\boldsymbol{\lambda}^*(s)$ is a linear combination of functions evaluated at $s \in [a, b]$. To determine $\boldsymbol{\lambda}^*$ let us go back to (4.3). Notice that

$$\begin{aligned}\sup_{\boldsymbol{\lambda}} \inf_{X} \{\mathcal{L}(X, \boldsymbol{\lambda}) | (X, \boldsymbol{\lambda}) : \mathcal{K} \times \mathbb{R}^K \to \mathbb{R}\} &= \sup_{\boldsymbol{\lambda}} \{\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle - \sup_{X} \{\langle \boldsymbol{A}^t\boldsymbol{\lambda}, X\rangle - \Psi(X)\}\} \\ &= \sup_{\boldsymbol{\lambda}} \{\langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle - M(\boldsymbol{A}^t\boldsymbol{\lambda})\} - \inf_{\boldsymbol{\lambda}} \{M(\boldsymbol{A}^t\boldsymbol{\lambda}) - \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle\}.\end{aligned}$$

Observe that

$$\langle \boldsymbol{A}^t\boldsymbol{\lambda}, X\rangle - \Psi(X) = \langle \boldsymbol{A}^t\boldsymbol{\lambda}, X\rangle - \Psi(X).$$

Taking the sup, we obtain the quantity $M(\boldsymbol{A}^t\boldsymbol{\lambda})$, after the obvious identifications.

In order to complete the solution process, we need to minimize $\Sigma(\boldsymbol{\lambda}, \boldsymbol{y}) = M_e(\boldsymbol{A}^t\boldsymbol{\lambda}) - \langle \boldsymbol{\lambda}, \boldsymbol{y}\rangle$ over $\boldsymbol{\lambda} \in \mathbb{R}^K$. Depending on the minimization software used, it may prove convenient to have the components of the gradient of that convex function. of the gradient. Below we denote by $A_{,0}(s)$ the i-th entry of the matrix $\boldsymbol{A}_0$:

$$\frac{\partial \Sigma}{\partial \lambda_i} = \int_0^1 A_i(s) \frac{l(s)e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)} + u(s)e^{u(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)}}{e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)} + e^{u(s)\boldsymbol{A}_0^t\boldsymbol{\lambda}(s)}} ds - y_i. \tag{4.7}$$

5. Dependence of the solution on the size of the data set

To make the analysis simpler, we suppose that we have an increasing collection of sets of labels $\mathcal{J}_K = \{n_1 < n_2 < ... < n_K\}$ such that $\{0\} \cup \bigcup_{K\geq 1} \mathcal{J}_K = \mathbb{N}$. In this case, according to (2.2) and (2.3), the size N of the data sets consists of K-sine transforms at the frequencies $\omega_k = 2\pi n_k/(b-a)$, while the $K+1$-cosine transforms also include the transform at $\omega_0 = 0$.

Regarding the notations introduced in (2.4), we denote, respectively, by $\boldsymbol{A}_0^N$, $\boldsymbol{A}^N$ and $\boldsymbol{y}^N$, the matrices denoting the transform at the frequencies specified in $\mathcal{J}_K$ (with $N = 2K+1$), and respectively, the augmented matrix and the data point. With this notational convention, it is clear that for $M > N$, the matrix $\boldsymbol{A}_0^N$ is a sub-matrix of $\boldsymbol{A}_0^M$, and similarly for the other objects.

Also, for $N < M$ let X^N and X^M be the respective solutions to (4.3) satisfying $\mathfrak{A}^N X^N = \boldsymbol{y}^N$, respectively, by $\mathfrak{A}^M X^M = \boldsymbol{y}^M$. By our notational assumptions, we also have $\mathfrak{A}^N X^M = \boldsymbol{y}^N$. That is the solution to the problem with a larger number of constraints satisfies the same constraints as the solution with a subset of the same constraints. According to (3.6) and (4.4), when $M > N$, we have

$$D_\Psi(X^M, X^N) = \Psi(X^N) - \Psi(X^N) - \int_0^1 \langle (X^M(s) - X^N(s)), \frac{\delta \Psi}{\delta X}(s)\rangle ds. \tag{5.1}$$

Making use of (4.4) and the consistency condition, we obtain

$$\int_0^1 \langle (X^M(s) - X^N(s))\boldsymbol{A}^t \lambda^N\rangle ds = \int_0^1 \langle \mathfrak{A}^N(X^M(s) - X^N(s)), \lambda^N\rangle ds = 0. \tag{5.2}$$

That is

$$D_\Psi(X^M, X^N) = \Psi(X^N) - \Psi(X^N) \geq 0. \tag{5.3}$$

By X_t being a true signal that it is a function $X_t : [a,b] \to \mathcal{K}$, such that for every N it satisfies $\mathfrak{A}^N X_t = \boldsymbol{y}^N$. Thus, if X_t is a true signal, then for every $N \geq 1$, the same reasoning as above yields:

$$D_\Psi(X_t, X^N) = \Psi(X_t) - \Psi(X^N) \geq 0. \tag{5.4}$$

These remarks add up to the proof of the following statement.

Theorem 5.1. *For any N let $\boldsymbol{A}^N$ and X_t be as introduced above. Let $\boldsymbol{y}^N = \mathfrak{A}^N X_t$ be the observed data, and X^N be the solution to problem* (4.3)*; therefore, $\mathfrak{A}^N X^N = \boldsymbol{y}^N$ holds true. Suppose that for $N < M$ the consistency assumptions stated above hold. Then* (5.3) *and* (5.4) *are satisfied, and in particular $X^N \to X_t$ in L_1.*

Proof. Since the function $\psi(x) = x\ln x + (1-x)\ln(1-x)$ is bounded, above by 0 and below by $-1/2$, then for any X the function $\Psi(X)$ is bounded. Since $\Psi(X^N)$ increases, it must have a limit. Then the sequence is Cauchy and invoking (3.9), the sequence is Cauchy. If X_∞ denotes its limit in $L_1[a,b]$), by continuity we have $\mathfrak{A}^N X_\infty = \boldsymbol{y}^N$ for any N. That is X_∞ is a solution to the Fourier inversion problem. From the uniqueness of the Fourier inversion problem, we obtain that $X_t = X_\infty$ almost surely in $[a,b]$.

□

6. Numerical experiments

In this section, we carry out two types of exercise. First, a reconstruction of a function from consecutive Fourier coefficients. We compare the simple, classical Fourier reconstruction given by

$$X_c(s) = \sum_{k=0}^{N} \phi_1(k)\,\sqrt{2}cos(\omega_k s) + \sum_{k=1}^{N} \phi_2(k)\,\sqrt{2}sin(\omega_k s) \quad s \in [0,1]. \tag{6.1}$$

Next, we apply the entropic procedure to obtain $X_N^*(s)$, and we plot it along with $X_t(s)$ and $X_c(s)$.

After that, we consider the case in which the Fourier coefficients are not consecutive. The underlying idea is to see whether the entropic reconstruction, being non-linear, can provide supplementary information about the missing coefficients.

As test cases, we use three periodic functions on $[0,1]$. First, a 3rd-degree polynomial with continuous periodic derivative, a continuous function, not differentiable at 0 and 1, and a discontinuous bump function to test the Gibbs phenomenon at the discontinuity. We refer to them as the "true" (but unknown) function in each case and denote it by $X_t(s)$. Their Fourier coefficients can be computed exactly, and we assume that they are the data in each of the cases.

The true polynomial function is

$$X_{true}(s) = s - 3s^2 + 2x^3. \tag{6.2}$$

Its Fourier coefficients are

$$\phi_n(1) = 0 \quad n \geq 1. \tag{6.3}$$

$$\phi_n(2) = \frac{3\,\sqrt{2}}{2(\pi)^3 n^3}, \quad n \geq 1. \tag{6.4}$$

As an example of a periodic, continuous function with discontinuous derivatives, we consider

$$\begin{aligned} X_{true}(s) &= cosh(t - 1/2) - (e^{1/2} - e^{-1/2}), \\ \phi_n(1) &= \sqrt{2}(e^{1/2} - e^{-1/2})\frac{1}{1 + (2\pi n)^2}. \end{aligned} \tag{6.5}$$

The corresponding Fourier coefficients for $n \geq 1$ are

$$\phi_n(1) = \sqrt{2}(e^{1/2} - e^{-1/2})\frac{1}{1 + (2\pi n)^2}. \tag{6.6}$$

$$\phi_n(2) = 0. \tag{6.7}$$

To finish, we consider a periodic discontinuous bump function.

$$X_{true}(s) = \begin{cases} -\frac{1}{2}, & 0 \leq s < \frac{1}{4}, \\ \frac{1}{2}, & \frac{1}{4} < s < \frac{3}{4}, \\ -\frac{1}{2}, & \frac{3}{4} < s \leq 1. \end{cases} \tag{6.8}$$

The corresponding Fourier coefficients are

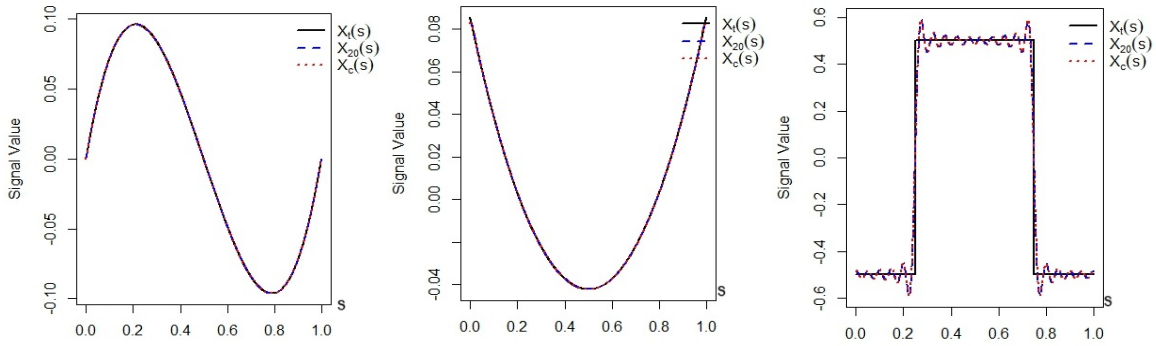
$$\phi_n(1) = \frac{\sqrt{2}}{4\pi n}(cos(\pi n) - 1)sin(\frac{n\pi}{4}), \quad n \geq 1. \tag{6.9}$$

$$\phi_n(2) = 0, \quad n \geq 1. \tag{6.10}$$

In all examples, the zero-frequency coefficient $\phi_0(1) = 0$. That is why it is not listed in any of the cases.

6.1. Simple test example

Here we carry on a comparison between the classical reconstruction and the entropic procedure. For that, we consider the first 20 consecutive Fourier coefficients of the true function. With these, we compute both the classical reconstruction X_c using (6.1) and the entropic X^*_{20} using the results in Section 3. The results are plotted in Figure 1(a)–(c).



(a) Smooth case **(b)** Continuous case **(c)** Discontinuous case

Figure 1. Comparison of approaches in three cases.

The ordering of the plots is the same as the labeling of the cases. We see that in all three examples, the reconstruction looks quite similar. To go beyond the visual appearance, in Table 1, we collect the $L_1[0, 1]$ distances between the true and the classical and between the true and the entropic reconstructions from the first 20 Fourier coefficients. As we have an analytical representation for the three solutions, the distances are computed using a numerical integrator. In these examples, the entropic solution is slightly closer to the true solution than the classical.

Table 1. Comparison of L_2-distances.

	Case 1	Case 2	Case 3
Classical	0.0093733	0.009373342	0.1184565
Entropic	0.0092356	0.00923568	0.1183498

6.2. Linearity test

One important feature of the Fourier transform is its linearity. Even though the entropic procedure is non-linear, using the properties (3.8) and (3.9), we can prove that approximate linearity holds. We shall prove that the sum of two entropic solutions is close in Bregman divergence to the signal reconstructed from the sum of the two data sets.

To make the statements above explicit, let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ denote two sets of Fourier coefficients and let $\boldsymbol{\zeta} = \boldsymbol{\xi} + \boldsymbol{\eta}$ denote their sum. There is no loss of generality in assuming that they are of the same length. Let X_c Y_c, and Z_c denote the classical solutions that they determine. Then $Z_c = X_c + Y_c$. If we denote by X^*_1, Y^*_{10}, and Z^*_{10} the corresponding entropic solutions, then there is a priori no reason to expect that $Z^*_{10} = X_{10^*} + Y^8_{10}$.

At this point, the divergence induced by the entropy Ψ provides a partial rescue. We know from (3.9) that

$$D_\Psi(X^*_{10} + Y_{10^*}, Z^*_{10}) \geq C\|X^*_{10} + Y^*_{10} - Z^*_{10}\|_1^2.$$

On the other hand, an argument similar to that used in (5.3) and (5.4) shows that

$$D_\Psi(X^*_{10} + Y^*_{10}, Z^*_{10}) = \Psi(X^*_{10} + Y^*_{10}) - \Psi(Z^*_{10}) \geq 0.$$

Using these facts, we obtain a notion of *epsilon*-quasilinearity that goes as follows:

Theorem 6.1. *Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be the Fourier coefficients of unknown signals, and let $\boldsymbol{\zeta} = \boldsymbol{\xi} + \boldsymbol{\eta}$. Denote respectively by X^*_{10}, Y^*_{10} and Z^*_{10} the entropic solutions that these coefficients determine. If for some $\epsilon > 0$ we have $\epsilon \geq D_\Psi(X^*_{10} + Y^*_{10}, Z^*_{10})$. Then $\|X^*_{10} + Y^*_{10} - Z^*_{10}\|_1 \leq (\epsilon/C)^{1/2}$.*

To exemplify that result, we generated three data vectors as follows. First, we generate four sets of 10 random numbers uniformly in $[-1, 1]$, and call them $u_k : k = 1, ..., 10$. Then form $u_k/k : k=1, ..., 10$ to mimic the beginning of a sequence that converges to ℓ_2 if continued indefinitely. Put these together as two vectors of 20 components $\boldsymbol{\xi} = (\boldsymbol{\xi}(1), \boldsymbol{\xi}(2))$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}(2), \boldsymbol{\eta}(2))$ and form the two classical Fourier reconstructions:

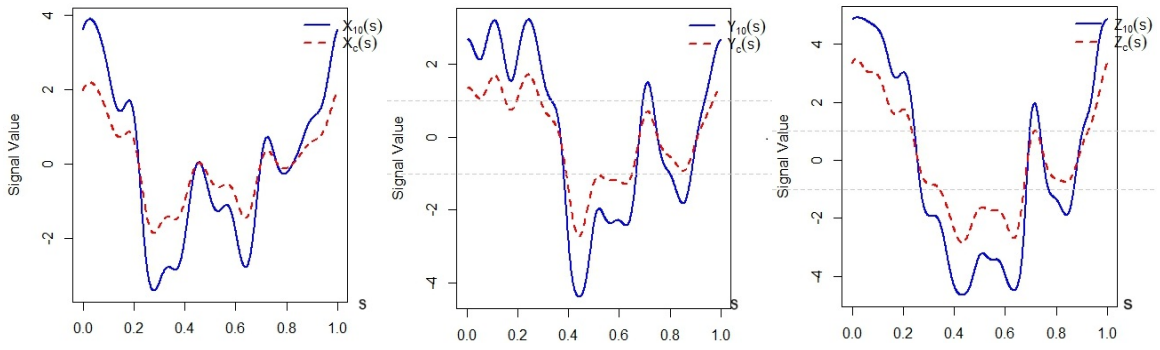
$$X_c(s) = \sum_{k=1}^{10} \Big(\sqrt{2}\xi_k(1)\cos(2\pi k) + \sqrt{2}\xi_k(2)\sin(2\pi k) \Big),$$

$$Y_c(s) = \sum_{k=1}^{10} \Big(\sqrt{2}\eta_k(1)\cos(2\pi k) + \sqrt{2}\eta_k(2)\sin(2\pi k) \Big).$$

Combine these two into $Z_c(s) = X_c(s) + Y_c(s)$, which is the Fourier reconstruction from the aggregate data $\boldsymbol{\zeta} = \boldsymbol{\xi} + \boldsymbol{\eta}$.

Next, we run the entropic procedure to recover the corresponding entropic reconstructions $X_{10}(s)$, $Y_{10}(s)$, and $Z_{10}(s)$. To see how much each of the latter resemble its classical counterpart, we plot each pair of them in Figure 2.

Keep in mind that X_c, Y_c and Z_c are the true signals, determined by the 10 sine and cosine trigonometric polynomials. Certainly, there is a resemblance. To quantify the approximation error, we compute the L_1 distance between the curves in the three plots using the exact analytic formula for each in a numerical integrator. The results are shown in Table 2.



(a) X_c and X_{10} **(b)** Y_c and Y_{10} **(c)** Z_c and Z_{10}

Figure 2. Evidence of quasi-linearity of the entropic procedure.

Table 2. L_1 distance between the signal and the entropic reconstructions.

$d(X_c, X^*_{10})$	0.814349
$d(Y_c, Y^*_{10})$	0.9598621
$d(Z_c, Z^*_{10})$	1.279874

Not only is the result consistent with the triangle inequality, but to test it, we graph $X^*_{10} + Y^*_{10}$ and Z^*_{10} in the same plot. It is displayed in Figure 3. The result is in agreement with the theoretical considerations in Section 6.2. We can confidently assert that there is some quasi-linearity in the entropic procedure.

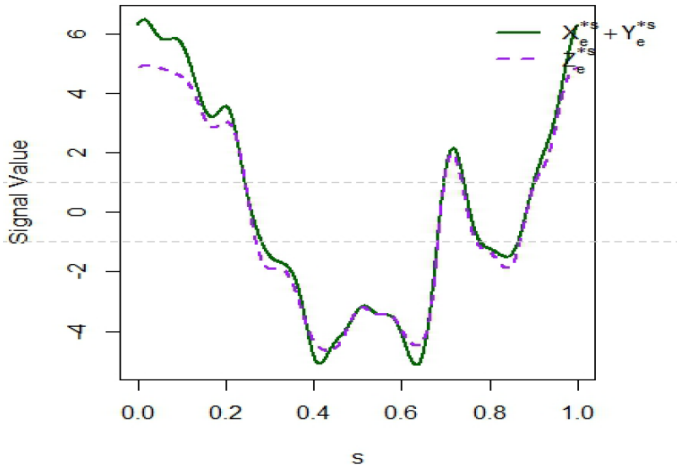


Figure 3. Plot of $X^*_{10} + Y^*_{10}$ and Z^*_{10}.

In this example, the trigonometric polynomial is the actual signal, so even though the reconstruction is not very accurate, there is some form of quasi-linearity. Notice that the L_1 distance between $X^*_{10} + Y^*_{10}$ and Z^*_{10} is 0.2239598.

6.3. The ghost in the reconstructed signal

Here we explore the issue that we mentioned above, caused by the non-linearity of the procedure. Namely, the entropic solution is not perpendicular to the space spanned by the data. It is perhaps analogous to the phenomenon called "ghosts in tomography". We mentioned above that the entropic procedure may yield information not present in the data. The example considered in the previous section is tailor made to illustrate the issue. Consider the data vector $\boldsymbol{\xi} = (\boldsymbol{\xi}(1), \boldsymbol{\xi}(2))$ used in the previous section. And consider the classical solution $X_c(s)$ and the entropic solution $X^*_{10}(s)$ that has $\boldsymbol{\xi}$ as data. Now, determine the first 40 coefficients of the entropic solution.

$$\phi_n(1) = \sqrt{2}\int_0^1 X^*_{10}(s)cos(2\pi ns)ds, \quad ; \; n = 1, 2, ..., 20.$$

$$\phi_n(2) = \sqrt{2}\int_0^1 X^*_{10}(s)sin(2\pi ns)ds, \quad ; \; n = 1, 2, ..., 20.$$

The first half of these will coincide with those of the vector $\boldsymbol{\xi}$, but the second half are coefficients of the part of X_{10}^* orthogonal to the span of the first 10 sine and cosine functions. Now define

$$X_g(s) = \sum_{n=1}^{20} \Big(\sqrt{2}\phi_k(1)\cos(2\pi n) + \sqrt{2}\phi_n(2)\sin(2\pi n) \Big).$$

Observe that the first half of the sum coincides with $X_c(s)$ because its coefficients coincide with those of $X_{10}^*(s)$, and observe that the second half is the ghost created by the entropic method. The energies in each part is

- $\sum_{n=1}^{10} ((\phi_n(1))^2 + (\phi_n(2))^2) = 1.149793717569.$
- $\sum_{n=11}^{20} ((\phi_n(1))^2 + (\phi_n(2))^2) = 0.001461941170.$
- $\sum_{n=1}^{0} ((\phi_n(1))^2 + (\phi_n(2))^2) = 1.151255658740.$

That is the ghost is energetic, but not so much. To finish we display the graphs of $X_c(s)$ and $X_{alt}(s)$ in one single plot. Their difference is the "ghost" reconstruction. The plot of the original $X_c(s)$ along with that of its ghost is displayed in Figure 4

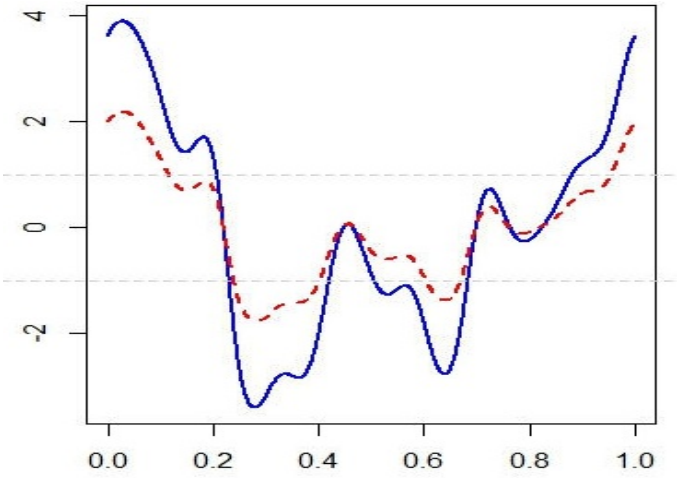


Figure 4. Plot of $X_c(s)$ (red) and $X_g(s)$(blue).

So, while the method can obtain part of the solution perpendicular to the original data, it can create non-existing data. A part that is in the kernel of the mapping $\mathfrak{A}$ introduced above, in (2.5).

7. Final remarks

To sum up. Above, we considered the problem of determining a continuous function on a bounded interval based on the knowledge of a few, not necessarily consecutive values of its sine or cosine transforms, which can be regarded as a generalized moment problem. As such it is an ill-posed problem. The boundedness of the function imposes box constraints upon the solution. The entropic procedure that we considered, naturally accounts for these constraints.

Besides solving the linear problem consisting of finding a function from the knowledge of a few of its sine and cosine coefficients, the entropic solution is non-orthogonal to the space of spanned

by the other trigonometric polynomials. This is compensates for the lack of linearity in the entropic procedure.

From the point of view of inverse problem theory, the problem that we considered is part of a large class of linear ill-posed inverse problems consisting of determining a continuous function $X : [0, 1] \to \mathbb{R}$ when the data consists of

$$\phi_k = \int_0^1 K(t_k, x)X(x)dx, \quad k = 1, ..., M.$$

Here, to keep it simple, $K(t, x) : [01] \times [0, 1] \to \mathbb{R}$ is a continuous kernel and $\{t_1, ..., t_M\}$ is a collection of distinct points in $[0, 1]$. The technique that we developed above, applies to this case, as well as to the problem of reconstructing a function from its projection onto a basis in a given Hilbert space. For example, the $\phi's$ might be the projection of X on a family of non-necessarily orthogonal functions, like, for example, a given collection of polynomials.

An important aspect of the entropic solution, is that the solutions obtained from partial data will converge to the true solution as the number of data points increases. For example, as the set of points $\{t_1, ..., t_M\}$ increases to a dense set, and the kernel defines a one-to-one-mapping.

Author contributions

Cécile Gauthier-Umaña, Valérie Gauthier-Umaña, Henryk Gzyl, and Enrique ter Horst: Conceptualization, Methodology, Validation, Writing-original draft, Writing-review & editing. All authors contributed equally to the different tasks of this project.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

1. J. Allen, L. Rabiner, A unified approach to short-time Fourier analysis and synthesis, *Proc. IEEE*, **65** (1977), 1558–1564. https://doi.org/10.1109/PROC.1977.10770
2. M. Benning, M. Burger, Modern regularization methods for inverse problems, *Acta Numer.*, **27** (2018), 1–111. https://doi.org/10.1017/S0962492918000016
3. J. Borwein, A. Lewis, *Convex analysis and nonlinear optimization: theory and examples*, New York: Springer Verlag, 2006. https://doi.org/10.1007/978-0-387-31256-9
4. R. Bracewell, *The Fourier transform and its applications*, 3 Eds., New York: McGraw-Hill, 2000.
5. M. Braun, *Differential equations and their applications: an introduction to applied mathematics*, New York: Srpinger-Verlag, 1993. https://doi.org/10.1007/978-1-4612-4360-1

6. J. Burg, Maximum entropy spectral analysis, Ph.D Thesis, Stanford University, 1975.
7. L. Cohen, *Time-frequency analysis: theory and applications*, New Jersey: Prentice-Hall, Inc., 1995.
8. J. Cooley, J. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comput.*, **19** (1965), 297–301. https://doi.org/10.2307/2003354
9. D. Dacunha-Castelle, F. Gamboa, Maximum d'entropie et probleme des moments, *Ann. Inst. Henri Poincaré*, **26** (1990), 567–596.
10. H. Dym, H. McKean, *Fourier series and integrals*, New York: Academic Press, 1972.
11. H. Engl, M. Hanke, A. Neubauer, *Regularization of inverse problems*, Dordrecht: Kluver Academic Publishers, 1996.
12. V. Gauthier-Umana, H. Gzyl, E. ter Horst, Decoding as a linear ill-posed problem: the entropy minimization approach, *AIMS Mathematics*, **10** (2025), 4139–4152. https://doi.org/10.3934/math.2025192
13. A. Golan, H. Gzyl, A generalized maxentropic inversion procedure for noisy data, *Appl. Math. Comput.*, **127** (2002), 249–260. https://doi.org/10.1016/S0096-3003(00)00172-7
14. F. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE*, **66** (1978), 51–83. https://doi.org/10.1109/PROC.1978.10837
15. K. Ito, B. Jin, *Inverse problems: tikhonov theory and algorithms*, Singapore: World Scientific Publishers, 2015. https://doi.org/10.1142/9120
16. B. Kaltenbacher, A. Neubauer, O. Scherzer, *Iterative regularization methods for nonlinear ill-posed problems*, Berlin: De Gruyter, 2008. https://doi.org/10.1515/9783110208276
17. G. Le Besnerais, J. Bercher, G. Demoment, A new look at entropy for solving linear inverse problems, *IEEE Trans. Inform. Theory*, **45** (1999), 1565–1578. https://doi.org/10.1109/18.771159
18. R. Lyons, *Understanding digital signal processing*, 3 Eds., New Jersey: Prentice Hall, 2010.
19. M. Stéphane, *A wavelet tour of signal processing: the sparse way*, 3 Eds., Amsterdam: Academic Press, 2009. https://doi.org/10.1016/B978-0-12-374370-1.X0001-8
20. A. Oppenheim, R. Schafer, J. Buck, *Discrete-time signal processing*, 2 Eds., New Jersey: Prentice Hall, 1999.
21. E. Titchmarsh, *The theory of functions*, Oxford: Oxford University Press, 1968.
22. A. Tikhonov, A. Goncharsky, V. Stepanov, A. Yagola, *Numerical methods for the solution to ill-posed problems*, Dordrecht: Springer, 1995. https://doi.org/10.1007/978-94-015-8480-7
23. F. Treves, *Basic linear partial differential equations*, New York: Academic Press, 1975.
24. Y. Vardi, D. Lee, From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems, *J. R. Stat. Soc. B*, **55** (1993), 569–598. https://doi.org/10.1111/j.2517-6161.1993.tb01925.x
25. G. Rioux, R. Choksi, T. Hoheisel, P. Maréchal, C. Scarvelis, The maximum entropy in the mean method for image deblurring, *Inverse Probl.*, **37** (2021), 015011. https://doi.org/10.1088/1361-6420/abc32e

Appendix. Notational details

The matrix $\boldsymbol{A}$ is a $2K \times 1$ (a column matrix) whose entries are functions.

$$\boldsymbol{A} = \sqrt{2}\begin{pmatrix} \cos(\omega_1 s) \\ \cos(\omega_2 s) \\ \vdots, \\ \cos(\omega_K s) \\ \sin(\omega_1 s) \\ \sin(\omega_1 s) \\ \vdots, \\ \sin(\omega_K s) \end{pmatrix}.$$

The quantity $\boldsymbol{A}^t\boldsymbol{\lambda}$ is

$$\boldsymbol{A}^t\boldsymbol{\lambda}(s) = \sqrt{2}\sum_{k=1}^{K}\lambda_k \cos(\omega_k s) + \sqrt{2}\sum_{i=1}^{K}\lambda_{K+1} k \sin(\omega_k s).$$

Hey, notice that there are $2K$ constraints, so $\boldsymbol{\lambda}$ is a $2K$-vector, and so the labels in the second sum start at $K + 1$.

The explicit formula for $\Sigma(\boldsymbol{\lambda}, \boldsymbol{y})M(\boldsymbol{A}^t\boldsymbol{\lambda}) - \langle \lambda, \boldsymbol{y}\rangle$ needs

$$M(\boldsymbol{A}^t\boldsymbol{\lambda}) = \int_0^1 \ln\left(e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)} + e^{u(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)}\right)ds.$$

The ith component of the gradient of $\Sigma(\boldsymbol{\lambda}, \boldsymbol{y})$.

$$\frac{\partial\Sigma}{\partial\lambda_i} = \int_0^1 A_i(s)\frac{l(s)e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)} + u(s)e^{u(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)}}{e^{l(s)\boldsymbol{A}^t\boldsymbol{\lambda}(s)} + e^{u(s)\boldsymbol{A}_0^t\boldsymbol{\lambda}(s)}} - y_i.$$

Here $A_i(s)$ denotes the ith component ($i = 1, 2, ..., K$) of the vector displayed at the beginning.