LINEAR INVERSE PROBLEMS WITH NONNEGATIVITY CONSTRAINTS THROUGH THE $\beta$-DIVERGENCES: SPARSITY OF OPTIMISERS

CAMILLE POUCHOL AND OLIVIER VERDIER

Abstract. We pass to continuum in optimisation problems associated to linear inverse problems $y = Ax$ with non-negativity constraint $x \geq 0$. We focus on the case where the noise model leads to maximum likelihood estimation through the so-called $\beta$-divergences, which cover several of the most common noise statistics such as Gaussian, Poisson and multiplicative Gamma. Considering $x$ as a Radon measure over the domain on which the reconstruction is taking place, we show a general sparsity result. In the high noise regime corresponding to $y \notin \{ Ax \mid x \geq 0 \}$, optimisers are typically sparse in the form of sums of Dirac measures. We hence provide an explanation as to why any possible algorithm successfully solving the optimisation problem will lead to undesirably spiky-looking images when the image resolution gets finer, a phenomenon well documented in the literature. We illustrate these results with several numerical examples inspired by medical imaging.

1. Introduction

Linear inverse problems $Ax = y$ often come with natural constraints, one of the most common being nonnegativity of the unknown, i.e., $x \geq 0$. This happens in various applications, in particular in the imaging sciences. In this setting, $x \in \mathbb{R}^r$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times r}$, where $m$ is the number of data points, $r$ is the number of voxels. One important example is that of Positron Emission Tomography (PET) where the sought-for image is the activity, which must be nonnegative [19]. Deconvolution problems often also incorporate such constraints [13].

Depending on the noise model, the corresponding (negative) log-likelihood problem typically writes

$$\min_{x \geq 0} D(y|Ax),$$

where $D$ is some kind of divergence functional. If the noise model is Gaussian, for instance, then $D$ is simply the Euclidean distance, whereas if the noise model is Poisson, $D$ is the Kullback–Leibler divergence.

Analysing the effect of increased resolution leads to considering $x$ as a function (henceforth denoted $\mu$) in some functional space $X$, and $A$ now stands for some linear mapping from $X$ onto $\mathbb{R}^m$, leading to the optimisation problem

$$(1) \quad \min_{\mu \geq 0} D(y|A\mu).$$

The non-negativity constraint has been proved to provoke sparsity in various contexts in optimisation and optimal control [7]. As a result, the right functional space $X$ to be considered appears to be that of Radon measures, with discrete measures considered as sparse. Examples of this phenomenon can be seen in optimal control [8, 15], and the same goes for the optimisation problem (1) when $D$ is the Kullback–Leibler divergence [17].
In the examples above, sparsity (which can arise in the form of Dirac masses) is undesirable as the sought-for image is expected to be at least piecewise smooth. In other contexts, sparsity of the signal must be enforced, as is the case for instance in sparse super-resolution [10]. In the latter case, the situation is completely different as the unknown signal is known to be a sum of (nonnegative Dirac masses) and the aim is to recover their support.

The goal of the present work is to generalise the sparsity results of [17] to the optimisation problem (1) when \( D \) is any \( \beta \)-divergence for \( \beta \in [0, 2] \). The \( \beta \)-divergences have attracted interest recently in non-negative matrix factorisation [11], and are now also advocated for in some medical imaging contexts, such as in PET [4]. They indeed have the appealing property of interpolating between three common divergences, namely the Itakura–Saito divergence (\( \beta = 0 \)), the Kullback–Leibler divergence (\( \beta = 1 \)) and the Euclidean distance (\( \beta = 2 \)) [6], which correspond to different noise models.

Some intermediate values of \( \beta \) have shown to yield interesting results in instances where real data may have complex noise statistics, as is the case in PET [4]. The variable \( \beta \) then adds flexibility to treating the inverse problem.

In this article, our main result may informally be stated as follows (see Theorem 3.4 for a precise statement), with mild assumptions on the operator \( A \).

**Theorem 1.1.** If \( y \notin \{ A\mu, \mu \geq 0 \} \), then any optimal solution \( \mu^* \) to (1) posed with a \( \beta \)-divergence is sparse.

In other words, if the data is not in the image of \( \{ \mu \geq 0 \} \) under the operator \( A \), any optimiser \( \mu^* \) will be sparse, typically a sum of Dirac masses. The condition \( y \notin \{ A\mu, \mu \geq 0 \} \) should be interpreted as a condition on the level of noise: the more noise there is, the more likely it is that this condition be fulfilled. Sparsity will hence arise in the high-noise regime.

Our results show that the optimisation problem itself leads to sparse results. Thus, any algorithm successfully solving (1) will inevitably lead to undesirably spiky-looking images as one keeps iterating. In the context of medical imaging, this has been observed when using the Maximum-Likelihood-Expectation-Maximisation (ML-EM, also called the Richardson–Lucy algorithm) for solving (1), and has been referred to in the literature as the “night-sky” or the “draughtsboard” effect [21].

Consequently, the same kind of artefacts will be observed for other likelihoods, hinting at the necessity of either early stopping when solving (1) (see [18]), or adding appropriate regularisation terms in the form

\[
\min_{\mu \geq 0} \ D(y|A\mu) + \lambda R(\mu).
\]

with \( \lambda \) big enough in order to alleviate the issue [4].

In this work, we also provide examples where, solving (1) for some values of \( \beta \in [0, 2] \), reconstructions exhibit the night-sky effect for inverse problems with sufficient noise. In agreement with our theoretical results, this is the case with enough iterations of either a convergent algorithm for solving (1), or a convergent algorithm for solving (1) with \( \lambda \) small.

**Outline of the paper.** The paper is organised as follows. In §2 we define the inverse problem by setting the functional analytic framework as well as the noise model through \( \beta \)-divergences, leading to the corresponding maximum likelihood problem. Then, in §3 we analyse the resulting optimisation problem and prove Theorem 1.1. Numerical simulations are presented in §4 in different contexts taken from medical imaging, confirming our results about sparsity.
2. Setting the inverse problem

2.1. Linear inverse problem. We aim at reconstructing an image \( \mu \) defined on a compact \( K \subset \mathbb{R}^p, p \geq 1 \). The sought-for image is taken to be an element of the space of Radon measures, denoted \( \mathcal{M}(K) \), which is the topological dual space of continuous functions over the compact, denoted \( C(K) \). We endow \( \mathcal{M}(K) \) with the weak-* topology, making \( C(K) \) its dual space. The dual pairing between a function \( \mu \in \mathcal{M}(K) \) and a function \( f \in C(K) \) will be denoted \( \langle \mu, f \rangle \).

Finally, \( \mathcal{M}_+(K) \) stands for the set of nonnegative Radon measures.

The data is made of \( m \) measurements, stacked into a nonnegative vector \( y \in \mathbb{R}_m^+ \). We shall use the notation 

\[ I := \{ i \in \{1, \ldots, m\} \mid y_i > 0 \}. \]

This vector itself is the realisation of some random variable with mean \( A\mu \), where \( \mu \in \mathcal{M}_+(K) \) is the image to be reconstructed, and \( A \) is a linear mapping \( A : \mathcal{M}(K) \to \mathbb{R}^m \).

We make the only regularity assumption that \( A \) is continuous in the weak-* topology. From [3, Proposition 3.14], this forces \( A \) to be of the form 

\[ (A\mu)_i = \langle \mu, a_i \rangle, \quad i = 1, \ldots, m, \]

where the \( a_i \) are elements of \( C(K) \). This covers the case of PET [16, 17] and more generally the setting of Hilbert–Schmidt operators: if the underlying operator in infinite dimension is of the form

\[ \mu \mapsto \int_K k(\cdot, y) \, d\mu(y), \]

for some smooth kernel \( k \in C(K \times K) \), the operator \( A \) typically is obtained from a sampling for \( m \) points \( x_i \in K \) or integrating the kernel over some subdomains \( \Omega_i \subset K \), namely

\[ a_i = k(x_i, \cdot), \text{ or } a_i = \int_{\Omega_i} k(x, \cdot) \, dx, \quad i = 1, \ldots, m. \]

We will make the assumption that \( A \) maps \( \mathcal{M}_+(K) \) into the set of (componentwise) nonnegative vectors denoted \( \mathbb{R}_m^+ \), and that it is non-trivial, i.e., \( A\mu > 0 \) componentwise for \( \mu \) the uniform measure over \( K \). These two assumptions may be written in the terms of the functions \( a_i \) as

\[ a_i \geq 0, \quad a_i \neq 0, \quad i = 1, \ldots, m. \]

Another consequence of the simple continuity assumption on \( A \) is that its adjoint \( A^* : \mathbb{R}^m \to C(K) \) is simply defined as

\[ A^* \lambda = \sum_{i=1}^m \lambda_i a_i, \quad \lambda \in \mathbb{R}^m. \]

We shall sometimes need to know when \( A^* \) is injective. This is of course equivalent to the linear independence of the family \( (a_i)_{i=1, \ldots, m} \). Note that since the codomain of \( A \) is finite dimensional, we have

\[ A^* \text{ is injective } \iff A \text{ is surjective}. \]

In order to solve the inverse problem (written informally as \( A\mu = y \)), we aim at solving the optimisation problem

\[ \min_{\mu \in \mathcal{M}_+(K)} D_{\beta}(y | A\mu), \]

2. Linear inverse problem. We aim at reconstructing an image \( \mu \) defined on a compact \( K \subset \mathbb{R}^p, p \geq 1 \). The sought-for image is taken to be an element of the space of Radon measures, denoted \( \mathcal{M}(K) \), which is the topological dual space of continuous functions over the compact, denoted \( C(K) \). We endow \( \mathcal{M}(K) \) with the weak-* topology, making \( C(K) \) its dual space. The dual pairing between a function \( \mu \in \mathcal{M}(K) \) and a function \( f \in C(K) \) will be denoted \( \langle \mu, f \rangle \). Finally, \( \mathcal{M}_+(K) \) stands for the set of nonnegative Radon measures.

The data is made of \( m \) measurements, stacked into a nonnegative vector \( y \in \mathbb{R}_m^+ \). We shall use the notation 

\[ I := \{ i \in \{1, \ldots, m\} \mid y_i > 0 \}. \]

This vector itself is the realisation of some random variable with mean \( A\mu \), where \( \mu \in \mathcal{M}_+(K) \) is the image to be reconstructed, and \( A \) is a linear mapping \( A : \mathcal{M}(K) \to \mathbb{R}^m \).

We make the only regularity assumption that \( A \) is continuous in the weak-* topology. From [3, Proposition 3.14], this forces \( A \) to be of the form 

\[ (A\mu)_i = \langle \mu, a_i \rangle, \quad i = 1, \ldots, m, \]

where the \( a_i \) are elements of \( C(K) \). This covers the case of PET [16, 17] and more generally the setting of Hilbert–Schmidt operators: if the underlying operator in infinite dimension is of the form

\[ \mu \mapsto \int_K k(\cdot, y) \, d\mu(y), \]

for some smooth kernel \( k \in C(K \times K) \), the operator \( A \) typically is obtained from a sampling for \( m \) points \( x_i \in K \) or integrating the kernel over some subdomains \( \Omega_i \subset K \), namely

\[ a_i = k(x_i, \cdot), \text{ or } a_i = \int_{\Omega_i} k(x, \cdot) \, dx, \quad i = 1, \ldots, m. \]

We will make the assumption that \( A \) maps \( \mathcal{M}_+(K) \) into the set of (componentwise) nonnegative vectors denoted \( \mathbb{R}_m^+ \), and that it is non-trivial, i.e., \( A\mu > 0 \) componentwise for \( \mu \) the uniform measure over \( K \). These two assumptions may be written in the terms of the functions \( a_i \) as

\[ a_i \geq 0, \quad a_i \neq 0, \quad i = 1, \ldots, m. \]

Another consequence of the simple continuity assumption on \( A \) is that its adjoint \( A^* : \mathbb{R}^m \to C(K) \) is simply defined as

\[ A^* \lambda = \sum_{i=1}^m \lambda_i a_i, \quad \lambda \in \mathbb{R}^m. \]

We shall sometimes need to know when \( A^* \) is injective. This is of course equivalent to the linear independence of the family \( (a_i)_{i=1, \ldots, m} \). Note that since the codomain of \( A \) is finite dimensional, we have

\[ A^* \text{ is injective } \iff A \text{ is surjective}. \]

In order to solve the inverse problem (written informally as \( A\mu = y \)), we aim at solving the optimisation problem

\[ \min_{\mu \in \mathcal{M}_+(K)} D_{\beta}(y | A\mu), \]
where $D_\beta$ is the $\beta$-divergence for $\beta \in [0, 2]$, see below for the definition. We will use the notation
$$\ell(\mu) := D_\beta(y | A\mu).$$
Finally, let us define the notion of support for the various relevant cases (all these cases can be covered in one single definition, but we prefer separating them for clarity):

- for a vector $w \in \mathbb{R}^m_+$, 
  $$\text{supp}(w) = \{ i \in \{1, \ldots, m\} \mid w_i > 0 \}$$
- for a nonnegative function $f \in \mathcal{C}(K)$, 
  $$\text{supp}(f) = \{ x \in K \mid f(x) > 0 \}$$
- for a nonnegative measure $\mu \in \mathcal{M}_+(K)$, 
  $$\text{supp}(\mu) := \{ x \in K \mid \mu(N) > 0, \forall N \in N(x) \}.$$ 
  where $N(x)$ is the set of all open neighbourhoods of $x$.

2.2. $\beta$-divergences.

2.2.1. For scalar positive variables. For $u > 0$, $v > 0$ scalar variables, $\beta \in [0, 2]$, we define
$$d_\beta(u|v) := \frac{1}{\beta (\beta - 1)} \left( u^\beta + (\beta - 1)v^\beta - \beta uv^{\beta - 1} \right),$$
which for $\beta = 2$ gives the Euclidean distance
$$d_2(u|v) = \frac{1}{2} (u - v)^2,$$
and by continuity for $\beta = 1$ the Kullback–Leibler divergence
$$d_1(u|v) = u \log \left( \frac{u}{v} \right) - u + v,$$
and for $\beta = 0$ the Itakura–Saito divergence
$$d_0(u|v) = \frac{u}{v} - \log \left( \frac{u}{v} \right) - 1.$$
More precisely, using the convention $0/0 = 0$, $0 \log 0 = 0$, $d_\beta$ is defined for nonnegative scalars $u \geq 0$, $v \geq 0$ as follows.

Case $1 < \beta \leq 2$:
$$d_\beta(u|v) = \frac{1}{\beta (\beta - 1)} \left( u^\beta + (\beta - 1)v^\beta - \beta uv^{\beta - 1} \right)$$

Case $\beta = 1$:
$$d_1(u|v) = \begin{cases} +\infty & \text{if } v = 0, u > 0 \\ v - u - u \log \left( \frac{v}{u} \right) & \text{otherwise} \end{cases}$$

Case $0 < \beta < 1$:
$$d_\beta(u|v) = \begin{cases} +\infty & \text{if } v = 0, u > 0 \\ \frac{1}{\beta (\beta - 1)} \left( u^\beta + (\beta - 1)v^\beta - \beta uv^{\beta - 1} \right) & \text{otherwise} \end{cases}$$

Case $\beta = 0$:
$$d_0(u|v) = \begin{cases} +\infty & \text{if } u = 0 \text{ or } v = 0 \\ \frac{u}{v} - \log \left( \frac{u}{v} \right) - 1 & \text{otherwise} \end{cases}$$

Let us review some of the most important properties of the $\beta$-divergences, which will be used throughout.
Separation: for all $\beta \in [0, 2],$
\[ d_\beta(u|v) = 0 \iff u = v. \]

Convexity: if (and only if) $\beta \in [1, 2]$ and for any $u \in \mathbb{R}_+, v \mapsto d_\beta(u|v)$ is convex on $\mathbb{R}_+$.

Lower semicontinuity: for all $\beta \in [0, 2]$ and $u \in \mathbb{R}_+$, the mapping $v \mapsto d_\beta(u|v)$ is lower-semicontinuous on $\mathbb{R}_+$.

Finally, it will be convenient to analyse when, for a given $u \in \mathbb{R}_+$, $v \mapsto d_\beta(u|v)$ is differentiable. More precisely, we will insist on situations when there is a point $v$, then necessarily 0 in view of the definition, at which $v \mapsto d_\beta(u|v)$ is non-differentiable.

This happens in the two following regimes:
- if $\beta \in (1, 2)$, and $u > 0$, the function $v \mapsto d_\beta(u|v)$ is not differentiable at $v = 0$.
- if $\beta \in (0, 1)$, the function $v \mapsto d_\beta(u|v)$ is not differentiable at $v = 0$.

Note that the function is not subdifferentiable either at such points.

2.2.2. For nonnegative vectors. The $\beta$-divergence for vectors $y, w \in \mathbb{R}_{+}^n$ is then defined as
\[ D_\beta(y|w) := \sum_{i=1}^{n} d_\beta(y_i|w_i), \]
and inherits the smoothness properties of $d_\beta$.

2.3. The noise model. Let us review the underlying statistical model, of which the minimisation problem (2.1) is (up to constants) the (negative-log) likelihood maximum problem, denoting $w = A\mu$, and $\phi$ a dispersion parameter. We write the statistical model for $y$ and $w$ as scalar variables, the full statistical model is defined component by component by independent draws of the same form.

A general way to write the noise model giving rise to $\beta$-divergences is to use the so-called Tweedie distributions. The $\beta$-divergences are a special case of such distributions, as the corresponding Tweedie distribution is given by
\[ y \mapsto h(x, \phi) \exp \left( -\frac{1}{\phi} d_\beta(y|w) \right), \]
so that minimising the negative log-likelihood problem indeed is equivalent to minimising $w \mapsto d_\beta(y|w)$. We refer to [20] for more details.

The underlying density is not always tractable (this is the case if $1 < \beta < 2$), making the noise model unclear. In other cases, the noise model can be further identified as follows.

Case $\beta = 2$: the noise model is Gaussian:
\[ y = \mathcal{N}(w, \phi). \]

Case $1 < \beta < 2$: no explicit model is known.

Case $\beta = 1$: the noise model is Poisson:
\[ y = \phi P\left( \frac{1}{\phi} w \right). \]

Case $0 < \beta < 1$: the noise model is compound Poisson, whose distribution is the sum of a singular measure at 0 and an absolutely continuous measure on the positive reals. More precisely:
\[ y = \sum_{i=1}^{n} g_i. \]
where \( n = \mathcal{P}(\lambda), q_i = \mathcal{G}(a,b) \) (here, \( \mathcal{G} \) stands for the Gamma distribution, of density \( x \mapsto \frac{b^a e^{-bx}}{\Gamma(a)} \) over \((0,+\infty))\), and the parameters \( \lambda, a \) and \( b \) are given as functions of \( \phi \) and \( \beta \) through

\[
\lambda = \frac{1}{\phi} \beta, \quad a = \frac{\beta}{1 - \beta}, \quad b = \frac{1}{\phi} \frac{w}{1 - \beta}.
\]

Case \( \beta = 0 \): the noise model is multiplicative Gamma:

\[
y = w \mathcal{G}(a,b),
\]

with

\[
a = \frac{1}{\phi}, \quad b = \frac{1}{\phi}.
\]

A common property of these noise models is that, the random variable \( y \) tends to \( w \) as the dispersion parameter \( \phi \) goes to zero:

\[
y \to w \quad \text{a.s., as } \phi \to 0.
\]

3. Optimisation problem

We now investigate the optimisation problem \( (2.1) \), starting with a related optimisation problem and its dual.

3.1. Related optimisation problem and its dual. Let us define the cone

\[
C := A(M_+) = \{ A\mu \mid \mu \in M_+ \},
\]

which is closed and convex. The original optimisation problem \( (2.1) \) is related to the following optimisation problem:

\[
(5) \min_{w \in A(M_+)} D_\beta(y|w).
\]

More precisely, for any \( w^* \) optimal for the above problem, any measure \( \mu^* \in M_+(K) \) such that \( A\mu^* = w^* \) is optimal for the original problem.

**Lemma 3.1.** The minimum in \( (2.1) \) is attained.

**Proof.** To prove the claim, we prove that the optimal value in the minimisation problem \( (3.1) \) is attained. For \( \beta > 0 \), the function \( w \mapsto D_\beta(y|w) \) behaves at infinity like \( w \mapsto \frac{1}{\beta} |w|_w^\beta \) and hence is coercive. For \( \beta = 0 \), it is also coercive since it behaves live \( w \mapsto -\sum_{i=1}^m \log(w_i) \) at infinity. It is also lower semi-continuous, and since \( A(M_+) \) is closed, there exists an optimal \( w^* \in A(M_+) \) for the problem \( (3.1) \). Any \( \mu^* \) such that \( A\mu^* = w^* \) then provides a minimiser for the original problem \( (2.1) \). \( \square \)

We now compute the dual problem to the original one. To be more precise, we compute the dual to \( (3.1) \). Finding the dual will be instrumental in proving our main result, **Theorem 3.4** via **Lemma 3.2**. These computations will also happen to be crucial in analysing results from numerical simulations, determining whether we should expect sparsity or not for some given \( y \in \mathbb{R}^m_+ \).

We define the cone \( A(M_+)^* := \{ \lambda \in \mathbb{R}^m \mid \langle \lambda, w \rangle \geq 0, \forall w \in A(M_+) \} \) dual to \( A(M_+) \), which can be characterised as in [12] by

\[
A(M_+)^* = \{ \lambda \in \mathbb{R}^m \mid A^* \lambda \geq 0 \text{ on } K \}.
\]

The dual problem writes

\[
\max_{\lambda \in A(M_+)^*} g(\lambda),
\]

where the function \( g: \mathbb{R}^m \to \mathbb{R} \) is defined for \( \lambda \in A(M_+)^* \) by

\[
g(\lambda) := \min_{w \in \mathbb{R}^m} D_\beta(y|w) - \langle \lambda, w \rangle.
\]
The point of taking the dual in this form is that, since $D_\beta$ decomposes, so does $g$ and we find
\[
g(\lambda) = \sum_{i=1}^m \min_{w_i \in \mathbb{R}} (d_\beta(y_i | w_i) - \lambda_i w_i).
\]
As a result, we focus on computing
\[
h(y, \lambda) := \min_{w \in \mathbb{R}} (d_\beta(y | w) - \lambda w) = \min_{w \geq 0} (d_\beta(y | w) - \lambda w)
\]
for a scalar $y \geq 0$. The resulting dual function will take the form $g(\lambda) = \sum_{i=1}^m h(y_i, \lambda_i)$.

As the explicit computation of the functions $h$ in (3.1) is quite involved, we postpone it to Appendix A.

**Lemma 3.2.** Let $w^*$ be an optimal vector for the optimisation problem (3.1). Then we have
\[
supp(y) \subset supp(w^*).
\]
**Proof.** Let us argue depending on the value of $\beta$. For any $0 \leq \beta \leq 1$, the objective function would be infinite at any $w$ such that $w_i = 0$ with $y_i > 0$, which obviously contradicts optimality.

When $1 < \beta \leq 2$, we shall make use of duality. The lemma is equivalent to proving that any $w^*$ optimal for (3.1) will satisfy $w^*_i > 0$ whenever $i$ is such that $y_i > 0$. Since $1 < \beta \leq 2$, the problem at hand is convex. Slater’s condition is satisfied since the cone $A(M_+)$ then has non-empty relative interior, as $A$ is non-trivial. Thus, strong duality holds and the dual optimal value is attained.

Denoting $\lambda^*$ a dual optimal variable, this implies that any primal optimal variable $w^*$ must minimise $w \mapsto D_\beta(y | w) - \langle \lambda^*, w \rangle$ over $\mathbb{R}^m$. Thus, for each $i \in \{1, \ldots, m\}$, $w^*_i$ minimises $w \mapsto d_\beta(y_i | w) - \lambda^*_i w$ over $\mathbb{R}$. Computations made to evaluate the dual problem impose that when $y_i > 0$, we must have
\[
(w^*_i)^{\beta-2}(w^*_i - y_i) = \lambda^*_i
\]
for such an $i$. Consequently, $w^*_i > 0$ and the proof is finished. \(\square\)

We also gather some results about optimal solutions to (3.1) in Appendix B.

### 3.2. **KKT conditions.**
Let us compute the KKT conditions for problem (2.1). We will do so at points at which $\ell$ is differentiable (endowing $M_+(K)$ with its strong topology), namely at measures $\mu \in M_+(K)$ such that
- when $\beta \in [1,2)$, supp$(\mu) \subset$ supp$(A\mu)$,
- when $\beta \in [0,1)$, we have $A\mu > 0$.

We shall denote $S$ the set of such measures $\mu$.

For $\mu \in S$, we define
\[
\lambda(\mu) := (A\mu)^{\beta-2}(A\mu - y),
\]
where multiplications and divisions are to be understood componentwise. This formula is well-defined when $\mu \in S$, recalling the convention $0/0 = 0$.

**Proposition 3.3.** Let $\bar{\mu} \in S$. Then the KKT conditions write
\[
A^* \lambda(\bar{\mu}) \geq 0 \text{ on } K, \quad A^* \lambda(\bar{\mu}) = 0 \text{ on supp}(\bar{\mu}).
\]

---

1This stems from the usual computations leading to the KKT conditions, see [2]: by strong duality, letting $\lambda^*$ be dual optimal, we have for any $w^*$ primal optimal, $D_\beta(y | w^*) = g(\lambda^*) = \min_{w \in \mathbb{R}^m} (D_\beta(y | w^*) - \langle \lambda^*, w \rangle) \leq D_\beta(y | w^*) - \langle \lambda^*, w \rangle \leq D_\beta(y | w^*)$. All the inequalities must be equalities and in particular, $w^*$ must minimise $w \mapsto D_\beta(y | w) - \langle \lambda^*, w \rangle$ over $\mathbb{R}^m$. 

---

1This stems from the usual computations leading to the KKT conditions, see [2]: by strong duality, letting $\lambda^*$ be dual optimal, we have for any $w^*$ primal optimal, $D_\beta(y | w^*) = g(\lambda^*) = \min_{w \in \mathbb{R}^m} (D_\beta(y | w^*) - \langle \lambda^*, w \rangle) \leq D_\beta(y | w^*) - \langle \lambda^*, w \rangle \leq D_\beta(y | w^*)$. All the inequalities must be equalities and in particular, $w^*$ must minimise $w \mapsto D_\beta(y | w) - \langle \lambda^*, w \rangle$ over $\mathbb{R}^m$. 

---
Proof. The KKT conditions here write
\[ \nabla \ell(\bar{\mu}) \in - N_{M_+}(\bar{\mu}), \]
where \( N_{M_+}(K)(\mu) \) is the normal cone of \( M_+(K) \) at \( \bar{\mu} \), defined by
\[ N_{M_+}(K)(\mu) := \{ f \in C(K) \mid \forall \nu \in M_+(K), \langle \mu - \nu, f \rangle \geq 0 \}. \]
On the one hand, the normal cone can be identified as
\[ N_{M_+}(K)(\mu) = \{ f \in C(K) \mid f \leq 0 \text{ on } K, f = 0 \text{ on } \text{supp}(\mu) \}, \]
see [17], and on the other hand the gradient of \( \ell \) at points of differentiability is readily computed as
\[ \nabla l(\mu) = A^* \lambda(\mu). \]
The combination of these two results exactly leads to (3.3). \( \square \)

Since \( \ell \) is smooth at points of regularity and the constraint \( \{ \mu \geq 0 \} \) is convex, we have the following dichotomy for a point \( \bar{\mu} \) satisfying the KKT conditions
\[ \text{• for any } \beta \in [0, 2], \quad \bar{\mu} \text{ is optimal } \implies \bar{\mu} \text{ satisfies (3.3)}. \]
\[ \text{• for any } \beta \in [1, 2], \quad \bar{\mu} \text{ satisfies (3.3) } \implies \bar{\mu} \text{ is optimal.} \]

3.3. Sparsity theorem. We are now in a position to prove the sparsity theorem. We will say that a measure \( \mu \) is sparse (with respect to an operator \( A : M \to \mathbb{R}^m \) as described in §2.1) if there exists \( \varphi \in A^* \mathbb{R}^m, \varphi \geq 0, \varphi \neq 0 \) such that
\[ \text{supp}(\mu) \subset \arg \min(\varphi). \]

**Theorem 3.4.** For any \( \mu^* \) optimal for (2.1),
\[ \text{supp}(\mu^*) \subset \arg \min(\varphi^*), \quad \varphi^* = A^* \lambda(\mu^*). \]
In particular, if \( A \) is surjective and \( y \notin A(M_+) \), any optimal measure \( \mu^* \) is sparse.

**Proof.** Let \( \mu^* \) be optimal for (2.1).

**First case, \( \beta \in \{0\} \cup [1, 2] \):** we know from [Lemma 3.2] that the function \( \ell \) is differentiable at \( \mu^* \). We may thus write the KKT conditions which are necessary (and sufficient for \( \beta \in [1, 2] \)). Hence, we find that the support of \( \mu^* \) satisfies
\[ \text{supp}(\mu^*) \subset \arg \min(\varphi^*), \quad \varphi^* = A^* \lambda(\mu^*). \]

**Second case, \( \beta \in (0, 1) \):** let us denote
\[ I_{\mu^*} := \{ i \in \{1, \ldots, m\} \mid (A\mu^*)_i > 0 \} = \text{supp}(A\mu^*). \]
First, we notice that [Lemma 3.2] ensures that for any \( i \notin I_{\mu^*} \), we have \( y_i = 0 \). Thus we have
\[ D_{\beta}(y|A\mu^*) = \sum_{i \in I_{\mu^*}} d_{\beta}(y_i|(A\mu^*)_i). \]
We define a reduced compact
\[ \tilde{K} = \bigcup_{i \in I_{\mu^*}} \text{supp}(a_i), \]
and, correspondingly, a reduced operator \( \tilde{A} : M_+(\tilde{K}) \to \mathbb{R}^{\#I_{\mu^*}} \) by
\[ \tilde{A}\mu := (A\mu)_{i \in I_{\mu^*}}. \]
Since \((A\mu^*)_i = 0\) for all \(i \notin I_{\mu^*}\), \(\text{supp}(\mu^*) \subset \tilde{K}\). Let us prove that \(\mu^*\), regarded as an element of \(\mathcal{M}_+(\tilde{K})\) is optimal for the problem

\[
\min_{\mu \in \mathcal{M}_+(\tilde{K})} \ell(\mu), \quad \ell(\mu) := D_\beta(y\mid A\mu).
\]

If not, and arguing by contradiction, we may find \(\tilde{\mu} \in \mathcal{M}_+(\tilde{K})\) such that \(\ell(\tilde{\mu}) < \ell(\mu^*)\). Let us extend \(\tilde{\mu}\) by 0 outside of \(\tilde{K}\), still abusively denoting the corresponding measure of \(\mathcal{M}_+(K)\) as \(\tilde{\mu}\). Given that \((A\tilde{\mu})_i = 0\), and \(y_i = 0\) for all \(i \notin I_{\mu^*}\), we find

\[
\ell(\mu^*) = \ell(y) > \ell(\tilde{\mu}) = \ell(\tilde{\mu}),
\]

which contradicts the optimality of \(\mu^*\) for the original problem.

Considering that \(\mu^*\) is optimal for the optimisation problem \([3.3]\), and that \(\ell\) is differentiable at \(\mu^*\), we may this time use the necessity of KKT conditions for this problem, which leads to

\[
\text{supp}(\mu^*) \subset \text{arg min} \min (\tilde{A}^*\lambda(\mu^*)), \quad \lambda^*_i = (A\mu^*)^{\beta-2}(A\mu^* - y_i), \quad i \in I_{\mu^*},
\]

where \(\mu^*\) is regarded as a measure in \(\mathcal{M}_+(\tilde{K})\). With the convention \(0/0 = 0\) and from the fact that the original measure \(\mu^* \in \mathcal{M}_+(\tilde{K})\) has support included in \(\tilde{K}\), the result above is equivalent to

\[
\text{supp}(\mu^*) \subset \text{arg min}(A^*\lambda(\mu^*)),
\]

as claimed.

Let us now assume that \(A\) is surjective and \(y \notin A(\mathcal{M}_+)\) and we shall prove that sparsity holds. We already know that \(\varphi^* = A^*\lambda(\mu^*) \geq 0\). It remains to show that \(\varphi^* \neq 0\). In order to do so, let us recall that the surjectivity of \(A\) is equivalent to the injectivity of \(A^*\) (itself equivalent to the linear independence of the \(a_i\)'s in \(\mathcal{C}(K)\)). Since \(A^*\) is injective and \(\varphi^* = A^*\lambda(\mu^*)\), all we need to prove is that \(\lambda^* \neq 0\). Given that \(\lambda(\mu^*) = (A\mu^*)^{\beta-2}(A\mu^* - y)\), this vector cannot vanish as otherwise we would get \(A\mu^* = y\), contradicting our initial assumption. 

For a discussion of why we may speak of a sparse measure when its support is included in \(\text{arg min}(\varphi)\) for some non trivial function \(\varphi \in A^*\mathbb{R}^m\) satisfying \(\varphi \geq 0\), we refer to [17]. For example, if \(A\) is surjective and if the \(a_i\) functions defining \(A\) are analytic on \(K\), any sparse measure in the sense defined here will be a sum of Dirac masses.

**Remark 3.5.** In the case \(\beta \in [1,2]\), since \(w \mapsto D_\beta(y|w)\) is strictly convex, we may say the following:

- if \(\mu_1\) and \(\mu_2\) are minimisers of \([2.1]\), then \(\lambda(\mu_1) = \lambda(\mu_2)\), where \(\lambda\) is given by \([3.2]\). Let us denote this unique value by \(\lambda^*\). In particular, this means that all the minimisers have a support included in the same set \(\text{arg min}(A^*\lambda^*)\),
- the set \(\{A\mu^* \mid \mu^* \text{ optimal for } [2.1]\}\) is reduced to a singleton,
- the optimisation problem \([3.1]\) has a single minimum \(w^*\).

We do not know if these observations still hold for \(\beta \in [0,1]\).

### 3.4. Low-noise and high noise regimes.

For high values of the dispersion parameter \(\phi\), i.e., when the noise level is high, we expect that the random variable \(y\) may be outside of the cone \(A(\mathcal{M}_+)\), leading to sparse measures as optimal solution to the optimisation problem \([2.1]\). We may prove the surprising result that there will be a single optimal measure to \([2.1]\) under quite general conditions, which require \(y \notin A(\mathcal{M}_+)\) and \(\beta \in [1,2]\).

Let us start with a formula for optimisers \(\lambda(\mu^*) = (A\mu^*)^{\beta-2}(A\mu^* - y)\) for an optimal measure \(\mu^*\).
Lemma 3.6. Let \( \mu^* \) be an optimal measure. Then
\[
\sum_{i=1}^{m} y_i (A\mu^*)_i^{\beta-1} = \sum_{i=1}^{m} (A\mu^*)_i^\beta. \tag{11}
\]
In particular, if \( y \notin A(\mathcal{M}_+) \), we find that \( \lambda(\mu^*) \) must have at least one positive and one negative component.

Proof. Proving the formula requires establishing that
\[
\sum_{i=1}^{m} y_i (w^*)_i^{\beta-1} = \sum_{i=1}^{m} (w^*)_i^\beta, \tag{12}
\]
for any \( w^* \) optimal form for (3.1), which is nothing but the result of Lemma B.2 in Appendix B.

Now assume that \( y \notin A(\mathcal{M}_+) \) and by contradiction that all components of \( \lambda^* \) are nonnegative. Then \( y_i \leq (A\mu^*_i) \) for all \( i \in \{1, \ldots, m\} \), and at least one of inequalities must be strict since \( y \notin A(\mathcal{M}_+) \), leading to
\[
\sum_{i=1}^{m} (A\mu^*)_i^\beta = \sum_{i=1}^{m} y_i (A\mu^*)_i^{\beta-1} < \sum_{i=1}^{m} (A\mu^*)_i^\beta, \]
a contradiction with (3.4).

Corollary 3.7. Assume that \( \beta \in [1, 2] \) and \( y \notin A(\mathcal{M}_+) \). Then, if the function \( \varphi^* := A^\dagger \lambda^* \neq 0 \), where \( \lambda^* \) is defined in Remark 3.3, has a single minimum \( x^* \), there is a single optimiser solving (2.1), given by
\[
\mu^* = \xi \delta_{x^*}, \quad \xi := \frac{\sum_{i=1}^{m} y_i a_i^{\beta-1}(x^*)}{\sum_{i=1}^{m} a_i^\beta(x^*)},
\]
with \( \delta_{x^*} \) the Dirac mass at \( x^* \).

Proof. Let \( \tilde{\mu}^* \) be optimal. We know that \( \text{supp}(\mu^*) \subset \text{arg min}(\varphi^*) \), with \( \varphi^* \neq 0 \) since \( y \notin A(\mathcal{M}_+) \). The assumption imposes \( \mu^* = \xi \delta_{x^*} \) for some \( \xi > 0 \). In virtue of the formula (3.4), we must have
\[
\sum_{i=1}^{m} y_i \xi^{\beta-1} (A\delta_{x^*})_i^{\beta-1} = \sum_{i=1}^{m} \xi^\beta (A\delta_{x^*})_i^\beta,
\]
equation with a unique solution for \( \xi \), as given in the statement.

On the other hand, in the low-noise regime and in view of (2.3), we expect \( y \in A(\mathcal{M}_+) \). In this more favourable situation, a natural question is the regularity of optimisers, hoping that some measures solving of \( A\mu = y \) will be absolutely continuous with respect to the Lebesgue measure on \( K \) (in the sequel, absolutely continuous will mean with respect to the Lebesgue measure). This question is tackled in [17], based on the results of [12], we end up with the following result where we use the notation
\[
\tilde{K} := \bigcup_{i \in I} \text{supp}(a_i), \quad \tilde{A}: \mathcal{M}_+(\tilde{K}) \to \mathbb{R}^{|I|}, \quad \tilde{\mu} := (A\mu)_{i \in I}.
\]

Proposition 3.8 ([17 Proposition 3.13]). Assume that \( y \in A(\mathcal{M}_+) \), and further that
\[
\tilde{y} := (y_i)_{i \in I} \in \text{int}(\overline{A(\mathcal{M}_+)}) \subseteq \overline{A(\mathcal{M}_+)} := \left\{ \tilde{A}\mu \mid \mu \in \mathcal{M}_+(\tilde{K}) \right\}. \tag{13}
\]
Then there exists an absolutely continuous measure \( \mu^* \) solving (2.1), i.e., such that \( A\mu^* = y \).
Quantifying the probability to have \( y \in A(M_+) \) is outside of the scope of this paper, we refer to \([17]\) for some results in that direction in the case of Poisson noise, namely for \( \beta = 1 \). Also outside the scope of this paper is to know whether a given algorithm will pick an absolutely continuous measure at the limit, rather than a sparse one, in the low-noise regime where \( y \in A(M_+) \) with \((3.8)\) satisfied.

\[ \square \]

4. Numerical experiments

We here present some simulations of algorithms solving \((2.1)\) in different contexts, where the results exhibit sparsity as expected from the theoretical results. All simulations are run using Python.

4.1. Sparsity Certificates. As evidenced by our results, the relevant criterion for sparsity of optimisers for the optimisation problem \((2.1)\) is independent of \( \beta \), as the question reduces to:

\[
\text{do we have } y \in A(M_+)?
\]

In practice, as one wants to solve \((2.1)\) (or possibly a regularised version thereof) in the form of some iterative algorithm defined by iterates of the form

\[
\mu_{k+1} = G_k(\mu_k),
\]

we are looking for methods allowing us to certify that \( y \notin A(M_+) \) along iterates, i.e., a method to prove that \( y \notin A(M_+) \) which writes as a function of \( \mu_k \) and that gets better as \( k \to +\infty \).

One approach towards this is to make use of duality: by weak duality (and recalling the definition of \( g \) in \((3.1)\)) we always have

\[
\forall \mu \in M_+(K), \, \forall \lambda \in A(M_+)^*, \, \ell(\mu) \geq g(\lambda).
\]

Since \( y \in A(M_+) \iff \ell(\mu) = 0 \), this entails the following straightforward result:

\[
(\exists \lambda \in A(M_+)^*, g(\lambda) > 0) \implies y \notin A(M_+).
\]

We will call a vector \( \lambda \in A(M_+)^* \) such that \( g(\lambda) > 0 \) a dual certificate of sparsity.

This provides a natural method when it comes to establishing sparsity: assume we have a convergent algorithm for solving \((2.1)\), in the sense that all subsequences of \((\mu_k)_{k \in \mathbb{N}}\) converge (in the weak-* sense) to some minimiser \( \mu^* \) of \((2.1)\).

Then, a candidate of choice for a dual certificate is given by

\[
\lambda_k := (A\mu_k)^{\beta - 2}(A\mu_k - y).
\]

Indeed, at least for \( \beta \in [1, 2] \), strong duality holds and the (unique) optimal dual variable \( \lambda^* \in A(M_+)^* \) is related to the (unique) primal variable \( w^* \) by \( \lambda^* := (w^*)^{\beta - 2}(w^* - y) \).

In particular, \( \lambda_k \) will converge to \( \lambda^* \) along subsequences, whence the convergence of \( g(\lambda_k) \) to \( \max_{\lambda \in A(M_+)^*} g(\lambda) = \min_{\mu \geq 0} \ell(\mu) \). Hence, if \( y \notin A(M_+) \), we should have \( \lim g(\lambda_k) > 0 \) as \( k \to +\infty \).

We will use this choice of \( \lambda_k \) even for \( \beta \notin [1, 2] \), although no theoretical guarantee is of avail when it comes to the convergence to a certificate.

A caveat with our choice is that we should only expect \( \lambda_k \in A(M_+)^* \) at the limit \( k \to +\infty \), and not for a fixed iteration number \( k \). In practice, if \( \lambda_k \notin A(M_+)^* \), we set \( \tilde{\lambda}_k := \lambda_k + c \) where \( c > 0 \) is a small constant, large enough to ensure that \( A^*\tilde{\lambda}_k \geq 0 \), i.e., \( \tilde{\lambda}_k \in A(M_+)^* \).
4.2. First example. We first look at an example from PET, where the aim is to solve (2.1) with $y$. A common way to do so is to use the following iterates, called multiplicative [13] [11]. Starting from some $\mu_0 \in \mathcal{M}_+(K)$ (typically a positive constant over the domain), the iterates write

$$ \mu_{k+1} = \mu_k \frac{A^*((A\mu_k)^{\beta-2}y)}{A^*((A\mu_k)^{\beta-1})}. $$

These iterates have the ML-EM algorithm ($\beta = 1$) and the Iterative Image Space Reconstruction ($\beta = 2$) as particular cases [9], and proofs of convergence for these algorithms with any $\beta \in [0, 2]$ can be found in [22], in the finite-dimensional case.

One advantage of these algorithms is the decrease of the functional $\ell$ along iterates, a result we may generalise to the infinite-dimensional setting, when $\beta \in [1, 2]$, see Appendix C, adapting a proof of [11]. Up to our knowledge, such a result remains elusive even in the finite-dimensional case when $\beta \in [0, 1)$, although simulations suggest it does hold true as well.

We now present the results of applying the algorithm in the case of a 2D PET operator $A$ with 90 views and 64 tangential positions (hence, $m = 5760$). Simulations are run using Python with the Operator Discretization Library [1]. The sought-for image is taken to be the Derenzo phantom, denoted $\mu_r$. The data is obtained by (re-scaled) Poisson draws, with a time-variable (or dose-variable) $t$ which accounts for the level of noise. In other words, $y \sim \frac{1}{t} \mathcal{P}(tA\mu_r)$, and the higher $t$, the lower the noise. In order to approach the infinite-dimensional setting of our work, we take a fine image space discretisation of $512 \times 512$ pixels.

Finally, we take $\beta = 1.2$, on purpose not quite matching the noise statistics, as Poisson noise should lead one to take $\beta = 1$. We hence mimic the situation of not knowing the exact noise statistics.

Figures 1 for $t = 1$ and 2 for $t = 10^{-1}$ both display the evolution of the loss function $\ell$ along iterates, i.e., $k \mapsto \ell(\mu_k)$, starting from $\mu_0 = 1$. As theoretically expected, the function decreases. We also plot the maximum attained for each reconstruction, namely $k \mapsto \max(\mu_k)$, which tends to increase. Finally, we show the reconstruction after $k = 100$ and $k = 1000$ iterates.

In the noisier case $t = 0$, some pixels clearly take over as one keeps iterating. Moreover, we can certify that the we should indeed expect sparsity, as we may provide a dual certificates proving that $y \notin A(\mathcal{M}_+)$. This is also suggested by the fact that $k \mapsto \ell(\mu_k)$ seems to converge to a positive value rather than to 0.

In the less noisy case $t = 10^{-1}$, it seems like the divergence $\ell(\mu_k)$ is not converging to 0, which might be a hint that we should still expect sparsity. However, we are not able to certify it. Note that if sparsity were to hold true in this case as well, it may be that many more iterates are required to see sparsity arise more clearly.

4.3. Examples with $\beta = 2$.

4.3.1. Toy Example. We illustrate the sparsity with a toy example. Here we choose $\beta = 2$, $K = [0, 1]$, $a_0 = 1$, $a_1(x) = x$, with $m = 2$. In this case, one can compute the sparse solutions explicitly depending on the parameter $y \in \mathbb{R}^2$, as shown in Figure 3.

We also look at the effect of regularisation. In this case, we use total variation regularisation, that is, we solve

$$ \min_{\mu \geq 0} \ell(\mu) + \rho \text{TV}($$

where $\text{TV}(\mu)$ is the total variation of the derivative of the measure $\mu$ and $\rho$ is a regularisation parameter. In this discretised, one-dimensional setting, this is simply
Figure 1. Case \( t = 0 \). (A) Divergence along iterates. (B) Maximum of reconstruction along iterates. (C) Reconstruction after 100 iterates. (D) Reconstruction after 1000 iterates.

\[
\text{TV}(\mu) = \sum_i |\mu^i - \mu^{i+1}|, \quad \text{where } \mu^i \text{ is the value of the discretised measure at pixel } i.
\]

We then compute the minimum using a primal-dual hybrid gradient method [5]. We plot the resulting minima for various values of the regularisation parameter \( \rho \) in Figure 4.

4.3.2. Tomography example. The setting is similar to the previous section: in Figure 5, we look at the effect of the regularisation parameter (with the same total variation regularisation), with the same optimisation algorithm (primal-dual hybrid gradient). The image resolution is 127 × 127, and there are 285 angles and 183 tangential coordinates.

REFERENCES

Figure 2. Case $t = 10^{-1}$. (A) Divergence along iterates. (B) Maximum of reconstruction along iterates. (C) Reconstruction after 100 iterates. (D) Reconstruction after 1000 iterates.


Figure 3. The sparse solutions of the problem (2.1) with two detectors, $a_0 = 1$, $a_1(x) = x$ on the interval $K = [0, 1]$. There are three regions outside the cone $\mathcal{AM}_+$. In one of them, the optimal solution is zero, otherwise it takes the form $\xi \delta_{x_0}$ with $x_0$ being zero or one depending on the region.

Figure 4. The solutions for $y_0 = 0$ and $y_1 = 1$ of the problem in Figure 3 with total variation regularisation. Each curve corresponds to a different regularisation parameter, labelled by its base-10 logarithm. We see that when the regularisation parameter goes to zero, the computed solution converges to the expected exact solution, which, we see from Figure 3 is $\mu = .5\delta_1$.

(A) The maximum of the reconstruction for various regularisation parameters $\rho$

(b) Reconstruction with no regularisation ($\rho = 0$)

(c) Reconstruction with high regularisation ($\rho = 6$)

**Figure 5.** Relation between regularisation parameter and the reconstruction of a Shepp–Logan phantom.


**Appendix A. Computation of the dual**

Let us denote

$$\psi_y(w) := d_\beta(y|w) - \lambda w.$$ 

A.1. **Case $\beta = 0$.** Recall that in this case, we always assume $y > 0$. If $\lambda > 0$, we have $\psi_y(w) \sim -\lambda w$ as $w \to +\infty$, whence $h(y, \lambda) = -\infty$. We now focus on the case $\lambda < 0$. We have $\psi_y(w) \sim -\lambda w$ as $w \to +\infty$. As $w \to 0$, we have $\psi_y(w) \sim \frac{y}{w}$, thus the function tends to $+\infty$ at both ends. Since $\psi'_y(w) = w^{-2}(w - y) + \lambda$, $\psi^{(2)}_y(w) = w^{-3}(2y - w)$ for $w > 0$, $w \mapsto \psi'_y(w)$ increases to the positive value $2y$ and then decreases to $0^+$: there is a unique $w \geq 0$ minimising $\psi_y(w)$, which we denote $w(y, \lambda)$, solving

$$w(y, \lambda)^{-2}(w(y, \lambda) - y) = \lambda.$$ 

A bit of algebra actually yields $w(y, \lambda) = \frac{\sqrt{1 - 4\lambda y} - 1}{-2\lambda}$.

If $\lambda = 0$, $w \mapsto d_\beta(y|w)$ is minimised at $w = y$, and we may gather this case with $\lambda < 0$, since $w^{-2}(w - y) = \lambda$ indeed has solution $y$ for $\lambda = 0$, which amounts to setting $w(y, 0) = y$. 
Summing up, we find
\[ h(y, \lambda) = \begin{cases} -\infty & \text{if } \lambda > 0 \\ d_\beta(y|w(y, \lambda)) - \lambda w(y, \lambda) & \text{if } \lambda \leq 0 \end{cases} \]

In the last case, further computations lead to
\[ d_\beta(y|w(y, \lambda)) - \lambda w(y, \lambda) = \sqrt{1 - 4\lambda y} - \ln\left(\frac{1}{2} (\sqrt{1 - 4\lambda y} + 1)\right) - 1. \]

**A.2. Case** \( 0 < \beta < 1 \). If \( \lambda > 0 \), we have \( \psi_y(w) \sim -\lambda w \) as \( w \to +\infty \), whence \( h(y, \lambda) = -\infty \) in this case.

We now focus on the case \( \lambda < 0 \). We still have \( \psi_y(w) \sim -\lambda w \) as \( w \to +\infty \). If \( y > 0 \), \( \psi_y(w) \sim -\frac{\beta}{\beta - 1} w^{\beta-1} \) as \( w \to 0 \), thus the function tends to \( +\infty \) at both ends.

Since \( \psi'_y(w) = w^{\beta-2}(w-y) - \lambda, \psi''_y(w) = w^{\beta-3}((2-\beta)y - (1-\beta)w) \) for \( w > 0 \), \( w \to \psi'_y(w) \) increases to the positive value \( \frac{2}{\beta - 2} y \) and then decreases to \( 0 \): there is a unique \( w \geq 0 \) minimising \( \psi_y(w) \), which we denote \( w(y, \lambda) \), solving
\[ w(y, \lambda)_{\beta-2}(w(y, \lambda) - y) = \lambda. \]

If \( y = 0 \), it is easily seen that the function \( \psi_y(w) = \frac{1}{\beta} w^\beta - \lambda w \) is minimised at \( w = 0 \), with value 0.

If \( \lambda = 0 \), \( w \to d_\beta(y|w) \) is minimised at \( w = y \), and we may gather this case with \( \lambda < 0 \), since if \( y = 0 \) we find \( w = 0 \) and if \( y > 0 \), \( w^{\beta-2}(-y) = 0 \) leads to \( w = y \), and we also here set \( w(y, 0) = y \).

Summing up, we find
\[ h(y, \lambda) = \begin{cases} -\infty & \text{if } \lambda > 0 \\ 0 & \text{if } \lambda \leq 0, y = 0 \\ d_\beta(y|w(y, \lambda)) - \lambda w(y, \lambda) & \text{if } \lambda \leq 0, y > 0 \end{cases} \]

**A.3. Case** \( \beta = 1 \). We have \( \psi_y(w) \sim (1-\lambda)w \) as \( w \to +\infty \), whence \( h(y, \lambda) = -\infty \) if \( \lambda > 1 \). If \( \lambda = 1 \) and \( y > 0 \), \( \psi_y(w) \sim -y \ln(w) \) as \( w \to 0 \): the function tends to \( -\infty \) as \( w \to 0 \) and \( h(y, \lambda) = -\infty \). If \( y = 0 \), the function equals 0 identically and its minimum is 0.

We now focus on the case \( \lambda < 1 \). We still have \( \psi_y(w) \sim (1-\lambda)w \) as \( w \to +\infty \). If \( y > 0 \), \( \psi_y(w) \sim -y \ln(w) \) as \( w \to 0 \), thus the function tends to \( +\infty \) at both ends.

Since \( \psi_y \) is strictly convex in this case, it has a unique minimum for \( w > 0 \), which we again denote \( w(y, \lambda) \), solving
\[ w(y, \lambda)^{\beta-1}(w(y, \lambda) - y) = \lambda \iff w(y, \lambda) = \frac{y}{1-\lambda}. \]

If \( y = 0 \), it is easily seen that the function \( \psi_y(w) = \frac{1}{\beta} w^\beta - \lambda w \) is minimised at \( w = 0 \), with value 0.

We may also gather the cases \( y = 0 \) and \( y > 0 \) whenever \( \lambda \leq 1 \), since the formula for \( w(y, \lambda) \) shows that it vanishes with \( y \).

Summing up, we find
\[ h(y, \lambda) = \begin{cases} -\infty & \text{if } \lambda > 1 \\ d_\beta(y|w(y, \lambda)) - \lambda w(y, \lambda) & \text{if } \lambda \leq 1 \end{cases} \]

In the last case, further computations lead to
\[ d_\beta(y|w(y, \lambda)) - \lambda w(y, \lambda) = y \ln(1 - \lambda). \]
A.4. Case $1 < \beta < 2$. If $y > 0$, $\psi_y(w) \sim \frac{w^\beta}{\beta}$ as $w \to +\infty$, and $\psi_y(w) \sim -\frac{w}{\beta - 1} w^{\beta - 1}$ as $w \to 0$. The derivative $\psi_y'$ satisfies $\psi_y'(w) \sim -y w^{\beta - 2}$ as $w \to 0$, and $\psi_y'(w) \sim w^{\beta - 1}$ as $w \to +\infty$. Thus the function $\psi'$ increases (as $\psi$ is convex) from $-\infty$ to $+\infty$. As a consequence, it has a unique minimum $w > 0$, which we again denote $w(y, \lambda)$, solving

$$w(y, \lambda)^{\beta - 2} (w(y, \lambda) - y) = \lambda.$$  

If $y = 0$ and $\lambda \leq 0$, it is easily seen that the function $\psi_y(w) = \frac{1}{\beta} w^\beta - \lambda w$ is minimised at $w = 0$, with value $0$, whereas if $\lambda > 0$, it has a unique minimum (also defined by the equation for $w(y, \lambda)$). Summing up, we find

$$h(y, \lambda) = \begin{cases} 
0 & \text{if } y = 0, \lambda \leq 0 \\
 d_\lambda(y|w(y, \lambda)) - \lambda w(y, \lambda) & \text{otherwise.}
\end{cases}$$

A.5. Case $\beta = 2$. In this case, $\psi_y$ has a unique minimum $w$ given by $w = \lambda + y$, which gives the explicit formula

$$h(y, \lambda) = -\frac{1}{2} (\lambda + y)^2 + \frac{1}{2} y^2.$$  

**Appendix B. Further results regarding optimality**

**Lemma B.1.** Let $w^*$ be optimal for the optimisation problem (3.1). Then there holds

$$\sum_{i=1}^m (w_i^*)^\beta = \sum_{i=1}^m y_i (w_i^*)^{\beta - 1}.$$  

**Proof.** We consider small (multiplicative) variations around $w^*$: in other words, we compute the first order expansion of the difference $\psi(\varepsilon) := D_\beta(y|w^*) - D_\beta(y|(1 - \varepsilon)w^*)$ for $\varepsilon$ small (without any sign assumption), with $(1 - \varepsilon)w^*$ an admissible choice since $C$ is a cone. We find

$$\psi(\varepsilon) = \frac{1}{\beta} \sum_{i=1}^m (w_i^*)^\beta (1 - (1 - \varepsilon)^\beta) + \frac{1}{1 - \beta} \sum_{i=1}^m y_i (w_i^*)^{\beta - 1} (1 - (1 - \varepsilon)^{\beta - 1})$$

$$= \varepsilon \left( \sum_{i=1}^m (w_i^*)^\beta - \sum_{i=1}^m y_i (w_i^*)^{\beta - 1} \right) + o(\varepsilon).$$

The first order term has to vanish for $(1 - \varepsilon)w^*$ not to violate the optimality of $w^*$, whence the claim. Note that the final expansion obtained is also valid for $\beta = 0$ and $\beta = 1$. \hfill \square

**Lemma B.2.** Let $w^*$ be optimal for the optimisation problem (3.1). Then for all $v \in C$ such that $\text{supp}(v) \subset \text{supp}(w^*)$, we have

$$\sum_{i=1}^m (w_i^*)^{\beta - 2} (w_i^* - y_i) (w_i^* - v_i) \leq 0.$$  

This inequality holds regardless of the assumption $\text{supp}(v) \subset \text{supp}(w^*)$ when $\beta \in \{0\} \cup [1, 2]$.

**Proof.** Similarly to Lemma B.2, we perturb $w^*$ but this time in the form $(1 - \varepsilon)w^* + \varepsilon v$ with $v \in C$ satisfying $\text{supp}(v) \subset \text{supp}(w^*)$, where $\varepsilon \in [0, 1]$ is small. Denoting $\psi(\varepsilon) := D_\beta(y|w^*) - D_\beta(y|(1 - \varepsilon)w^* + \varepsilon v)$, we find $\psi(\varepsilon) = \sum_{i=1}^m \psi_i(\varepsilon)$, where

$$\psi_i(\varepsilon) := \frac{1}{\beta} ((w_i^*)^\beta - ((1 - \varepsilon)w_i^* + \varepsilon v_i)^\beta) + \frac{1}{1 - \beta} y_i ((w_i^*)^{\beta - 1} - ((1 - \varepsilon)w_i^* + \varepsilon v_i)^{\beta - 1})$$

For a given $i \in \{1, \ldots, m\}$, let us argue depending on whether $i \in \text{supp}(w^*)$ or not.
If \( i \in \text{supp}(w^*) \), the functions \( x \mapsto x^\beta \) and \( x \mapsto x^{\beta-1} \) are differentiable at \( x = w_i^* \) and we find by a Taylor expansion
\[
\psi_i(\varepsilon) = (w_i^*)^{\beta-2}(w_i^* - y_i)(w_i^* - v_i)\varepsilon + o(\varepsilon),
\]
When \( i \not\in \text{supp}(w^*) \), Lemma 3.2 ensures that \( y_i = 0 \), so the only remaining contribution is explicitly given by
\[
\psi_i(\varepsilon) = -\frac{1}{\beta} v_i^2 \varepsilon^\beta,
\]
Note these two expansions also hold for \( \beta = 0 \) and \( \beta = 1 \) (the case \( i \not\in \text{supp}(w^*) \) is void when \( \beta = 0 \)).

All in all, we find
\[
\psi(\varepsilon) = \left( \sum_{i \in \text{supp}(w^*)} (w_i^*)^{\beta-2}(w_i^* - y_i)(w_i^* - v_i) \right) \varepsilon + o(\varepsilon) - \frac{1}{\beta} \left( \sum_{i \not\in \text{supp}(w^*)} v_i^2 \right) \varepsilon^\beta.
\]

For \( \beta = 1 \), the terms may be regrouped since they coincide: the leading term is \( \varepsilon \sum_{i=1}^m (w_i^*)^{\beta-2}(w_i^* - y_i)(w_i^* - v_i) \), so the parenthesis has to be non-positive, as claimed. For \( \beta \in (1,2] \) and \( \beta = 0 \), the leading term is \( \varepsilon \sum_{i \in \text{supp}(w^*)} (w_i^*)^{\beta-2}(w_i^* - y_i)(w_i^* - v_i) \) and has to be non-positive.

Finally, for \( \beta \in (0,1) \), the leading term is \(-\frac{1}{\beta} \left( \sum_{i \not\in \text{supp}(w^*)} v_i^2 \right) \varepsilon^\beta \) but assumption \( \text{supp}(v) \subset \text{supp}(w^*) \) shows that it vanishes.

Thus, in all cases, we obtain the expected inequality but with a sum ranging over \( i \in \text{supp}(w^*) \). That we may take the sum to range over \( i \in \{1, \ldots, m\} \) is a consequence of Lemma 3.2, if \( i \not\in \text{supp}(w^*) \), we have \( y_i = 0 \), whence \( (w_i^*)^{\beta-2}(y_i - w_i^*)(v_i - w_i^*) = 0 \) and this term can be kept in the sum without affecting it. \( \square \)

**Appendix C. Monotonicity along multiplicative updates**

**Proposition C.1.** Assume that \( \beta \in [1,2] \). For any \( \mu_0 \in \mathcal{M}_+(K) \) with \( \tilde{K} \subset \text{supp}(\mu_0) \), the iterates
\[
\mu_{k+1} = \mu_k \frac{A^*((A\mu_k)^{\beta-2}y)}{A^*((A\mu_k)^{\beta-1})},
\]
are well-defined and they make \( \ell \) decrease, i.e.,
\[
\forall k \in \mathbb{N}, \quad D_\beta(y|A\mu_{k+1}) \leq D_\beta(y|A\mu_k).
\]

**Proof.** Let us check that the assumption \( \tilde{K} \subset \text{supp}(\mu_0) \) implies \( \text{supp}(\mu_k) = \tilde{K} \) along iterates, making all divisions are well-defined (with the convention \( 0/0 = 0 \)).

Outside of \( \text{supp}(\mu_k) \), the value of the function \( \frac{A^*((A\mu_k)^{\beta-2}y)}{A^*((A\mu_k)^{\beta-1})} \) does not matter (hence we do not care if there are divisions by 0). All we have to prove is that for all \( k \in \mathbb{N}^* \), there holds

- \( \text{supp}(y) \subset \text{supp}(A\mu_k) \),
- \( \text{supp}(\mu_k) = \tilde{K} \).

Then, the product \( (A\mu_k)^{\beta-2}y \) is indeed well-defined with the convention \( 0/0 = 0 \), and the denominator \( A^*((A\mu_k)^{\beta-1}) \) vanishes only on \( \text{int}(K \setminus \tilde{K}) \), i.e., outside of \( \text{supp}(\mu_k) = \tilde{K} \). Hence \( \mu_{k+1} \) is well-defined.

We proceed recursively. The second point is clear since we have \( \text{supp}(\mu_{k+1}) \subset \text{supp}(\mu_k) \), and \( \text{supp}(A^*((A\mu_k)^{\beta-2}y)) = \tilde{K} \). For the first point, we fix \( i \in I \) and prove that \( (A\mu_{k+1})_i > 0 \). Notice that the denominator \( A^*((A\mu_k)^{\beta-1}) \) is bounded.
from above by some positive constant $M_k$. We may write

\[
(A\mu_{k+1})_i \geq y_i(A\mu_k)_{i}^{3-2} \int_{K} \frac{a_i^2}{A^*((A\mu_k)^{3-1})} d\mu_k \geq \frac{y_i(A\mu_k)_{i}^{3-2}}{M_k} \int_{K} a_i^2 d\mu_k
\]

\[
\geq \frac{y_i(A\mu_k)_{i}^{3-2} (A\mu_k)^{i}}{M_k (K)} = \frac{y_i(A\mu_k)_{i}^{3}}{M_k \mu_k(K)} > 0,
\]

where we have used the Cauchy–Schwarz inequality from above by some positive constant $20$.

For $k \in \mathbb{N}$, and with the notation $w_k := A\mu_k$, consider the surrogate function, defined for $\mu \in \mathcal{M}_+(K)$ absolutely continuous with respect to $\mu_k$ by

\[
G_k(\mu) = \sum_{i=1}^{m} \int_{K} d_{\beta}(y_i(w_k), \frac{d\mu}{d\mu_k}) \frac{a_i}{(w_k)_i} d\mu_k.
\]

**Step 1.** Let us prove that $G_k$ is above $\ell$ and coincides with it at $\mu_k$. We compute

\[
G_k(\mu_k) = \sum_{i=1}^{m} \int_{K} d_{\beta}(y_i(w_k), \frac{d\mu}{d\mu_k}) \frac{a_i}{(w_k)_i} d\mu_k
\]

\[
= \sum_{i=1}^{m} d_{\beta}(y_i(w_k)_i, \int_{K} \frac{a_i}{(w_k)_i} d\mu_k = D_{\beta}(y(w_k)) = \ell(\mu_k)
\]

Then, for a given $\mu \in \mathcal{M}_+(K)$ absolutely continuous with respect to $\mu_k$ and by the convexity of $v \mapsto d_{\beta}(u, v)$ with the probability measure $\nu_k := \frac{a_i}{(w_k)_i} \mu_k$,

\[
\int_{K} d_{\beta}(y_i(w_k)_i, \frac{d\mu}{d\mu_k}) d\nu_k \geq d_{\beta}(y_i, \int_{K} \frac{d\mu}{d\mu_k} d\nu_k) = d_{\beta}(y_i(A\mu_k)_i).
\]

Summing on $i$ yields the result.

**Step 2.** We now prove

\[
G_k(\mu_{k+1}) \leq G_k(\mu_k).
\]

From the convexity of $v \mapsto d_{\beta}(u, v)$ and for $i \in I$, \[d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) \geq d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) \]

\[+ \partial_v d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) (\frac{d\mu_{k+1}}{d\mu_k})_i (\mu_k - \mu_{k+1}) \]

which yields

\[
G_k(\mu_{k+1}) \geq G(\mu_{k+1}) + \sum_{i=1}^{m} \int_{K} \partial_v d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) (\frac{d\mu_{k+1}}{d\mu_k})_i (\mu_k - \mu_{k+1}) \]

\[= G(\mu_{k+1}) + \sum_{i=1}^{m} a_i \partial_v d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) (\mu_k - \mu_{k+1}).
\]

We are done if we prove that

\[
\sum_{i=1}^{m} a_i \partial_v d_{\beta}(y_i(w_k)_i, \frac{d\mu_{k+1}}{d\mu_k}) = A^* \partial_v d_{\beta}(y(w_k), \frac{d\mu_{k+1}}{d\mu_k})
\]

on $K$. Since $\frac{d\mu_{k+1}}{d\mu_k} = \frac{A^*(w_{k+1}^{\frac{3}{2}-2})}{A^*(w_{k+1}^{\frac{3}{2}-2})} = f_{k+1}$, and $\partial_v d(u|v) = v^{\frac{3}{2}-2}(v - u)$, we compute

\[
A^* \partial_v d_{\beta}(y(w_k) f_{k+1}) = f_{k+1}^{\frac{3}{2}-2} A^*(w_{k+1}^{\frac{3}{2}-1} f_{k+1} - w_{k}^{\frac{3}{2}-2} y)
\]

\[= f_{k+1}^{\frac{3}{2}-2} (f_{k+1} A^*(w_{k+1}^{\frac{3}{2}-1} - A^*(w_{k+1}^{\frac{3}{2}-1} y)) = 0
\]
by the definition of $f_{k+1}$. We may now conclude as the combination of the two results entails for all $k \in \mathbb{N}$

$$\ell(\mu_{k+1}) \leq G_k(\mu_{k+1}) \leq G_k(\mu_k) = \ell(\mu_k).$$

$\square$

Université de Paris, FP2M, CNRS FR 2036, MAP5 UMR 8145, F-75006 Paris, France.
E-mail address: camille.pouchol@parisdescartes.fr

Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden.
E-mail address: olivierv@kth.se

Department of Computing, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway.
E-mail address: olivier.verdier@hvl.no