RAIRO-Oper. Res. 55 (2021) 355–393
https://doi.org/10.1051/ro/2021005

RAIRO Operations Research
www.rairo-ro.org

MAXIMUM ENTROPY ON THE MEAN APPROACH TO SOLVE GENERALIZED INVERSE PROBLEMS WITH AN APPLICATION IN COMPUTATIONAL THERMODYNAMICS

Fabrice Gamboa[1,2], Christine Guéneau[3], Thierry Klein[1,2,4] and Eva Lawrence[1,2,3,*]

Abstract. In this paper, we study entropy maximisation problems in order to reconstruct functions or measures subject to very general integral constraints. Our work has a twofold purpose. We first make a global synthesis of entropy maximisation problems in the case of a single reconstruction (measure or function) from the convex analysis point of view, as well as in the framework of the embedding into the Maximum Entropy on the Mean (MEM) setting. We further propose an extension of the entropy methods for a multidimensional case.

Mathematics Subject Classification. 90C25, 90C46, 62G07, 62P30.

Received July 20, 2020. Accepted January 14, 2021.

1. Introduction

In some problems coming from applied physics, a multidimensional function f taking values in $\mathbb{R}^p$ ought to be reconstructed given a set of observations. In thermodynamics, information on the function of interest, namely the p components of the function f we wish to reconstruct, are indirectly available. In general the available information consists in the value of integrals that involves the unknown function f and some known weights $(\lambda^i)_{i=1,\dots,p}$. For example, one can consider an interpolation problem when the integration measure consists in Dirac masses. In this case we give at known locations the value of a scalar product between f and λ, see expression (1.2). In the present work, we need to consider more general constraints. Therefore we study a reconstruction problem in which constraints are defined as integrals involving the unknown function f and the weight function λ against suitable measures Φ, see expressions (1.1) and (1.3).

In the sequel we provide a general method for the reconstruction of a p-real valued function from partial knowledge submitted to the general constraints previously discussed. We refer the interested reader to Chapter

Keywords. Entropy maximisation problems, Bayesian statistics, application in engineering.

[1] Institut de Mathématiques de Toulouse, Toulouse, France.

[2] ANITI Toulouse.

[3] ISAS – Service de la Corrosion et du Comportement des Matériaux dans leur Environnement (SCCME), CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France.

[4] Ecole Nationale d'Aviation Civile, Toulouse, France.

*Corresponding author: eva.lawrence@math.univ-toulouse.fr

2 of [22] for the basic rules of thermodynamics and Chapter 5 of [22] for the description of functions which are ordinarily considered for the reconstruction of thermodynamic quantities.

To be more precise, we consider a $\mathbb{R}^p$ valued function $f(x) = (f^1(x), \dots, f^p(x))$ defined for all x in the compact set $U \subset \mathbb{R}^d$ (we assume the interior of U to be non-empty). We set our work in a probability space $(U, \mathcal{B}(U), P_U)$ where $\mathcal{B}(U)$ is the Borel σ-algebra and P_U is the given reference measure. In such framework, we wish to reconstruct f over U and such that the reconstructed quantity satisfies the N following integral constraints

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) = z_l \qquad 1 \leqslant l \leqslant N, \tag{1.1}$$

where Φ_l are N positive (known) finite measures on $(U, \mathcal{B}(U))$ and λ^i are known continuous weight functions.

The expression of integral constraints as in (1.1) allows to express a wide range of problems. For example, one can consider some pairs $(x_l, z_l) \in U \times \mathbb{R}$ for $l = 1, \dots, N$, and wish to solve the following interpolation equations

$$\sum_{i=1}^{p} \lambda^i(x_l) f^i(x_l) = z_l \qquad 1 \leqslant l \leqslant N. \tag{1.2}$$

Expression (1.2) can be obtained from (1.1) by choosing $\mathrm{d}\Phi_l(x) = \delta_{x_l}(\mathrm{d}x)$ the Dirac measure located at x_l for all $l = 1, \dots, N$. Therefore, the integral constraints (1.1) become interpolation constraints.

When U is a subset of $\mathbb{R}$, one can also involves the l first moments of f^i by taking $\mathrm{d}P_U(x) = \mathrm{d}x$ and $\mathrm{d}\Phi_l(x) = x^l \mathrm{d}P_U(x)$. Namely, integral constraints (1.1) become in this case

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) x^l \mathrm{d}P_U(x) = z_l \qquad 1 \leqslant l \leqslant N.$$

In our work, the z_l represent N ideal real-valued measurements. In the case of noisy observations, a relaxed version of problem (1.1) can be considered. The aim is then to reconstruct p real-valued functions on U such that

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N, \tag{1.3}$$

where $\mathcal{K}_l$ is an interval in $\mathbb{R}$. In the sequel $\mathcal{K}$ will denote the product of the N intervals $\mathcal{K}_l$.

In the general case, problem (1.3) is ill-posed and has many solutions. In our work, we propose to choose among the solutions the function f that maximises I_γ the γ-entropy of the function f defined by

$$I_\gamma(f) = -\int_U \gamma\left(f^1(x), \dots, f^p(x)\right) \mathrm{d}P_U(x) \tag{1.4}$$

where γ is a strictly convex function from $\mathbb{R}^p$ to $\mathbb{R}$. In this framework, the reconstruction problem we consider can be rephrased as

$$\begin{aligned} &\max \ I_\gamma(f) \\ &\text{s.t.} \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi(x) \in \mathcal{K}. \end{aligned} \tag{$\mathcal{F}^p_{\Phi,\gamma}$}$$

The resolution of problem ($\mathcal{F}^p_{\Phi,\gamma}$) is conducted in two steps. We consider a dual problem on (signed) measures as a first step. The second step consists in solving a discrete approximation of the dual problem. This approach is summarised in Figures 1 and 2 (see on the next page). The Figure 1 presents the approach in the case $p = 1$ and $\lambda(x) = 1$, $\forall x \in U$, which has already been treated in [10]. Figure 2 presents the case $p \neq 1$, which is the extension of [10] treated in this work. The resolution we propose involves the embedding into a more complicated framework. Let us sketch the description of this framework. Let V be a Polish space and P_V be

Case $p = 1$.

- **Problem on function.**

$$\boxed{\begin{array}{l}\max\ I_\gamma(f)\\ f\ :\ \displaystyle\int_U f(x)d\Phi(x) \in \mathcal{K}.\end{array}} \qquad (\mathcal{F}^1_{\Phi,\gamma})$$

$\Updownarrow$ Transfer principle [8].

- **Problem on signed measure** [9], [6], [2].

$$\boxed{\begin{array}{l}\min\ D_\gamma(F, P_V)^a\\ F\ :\ \displaystyle\int_V \varphi(t)dF(t) \in \mathcal{K}.\end{array}} \qquad (\mathcal{M}^1_{\varphi,\gamma})$$

$\Uparrow\Downarrow$ Point-measure approximation of F $\sum_{i=1}^n \frac{1}{n} Y_i \delta_{t_i}$.

- **Maximum Entropy on the Mean** (MEM) [13].

$$\boxed{\begin{array}{l}\min\ D_{KL}(Q, \Pi^{\otimes n})^b\\ Q\ :\ \mathbb{E}_Q\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^n \varphi(t_i)Y_i\right] \in \mathcal{K}.\end{array}} \qquad (\mathcal{M}^{1,n}_{\varphi,\Pi})$$

[a] Integral criterion D_γ is defined by (7).
[b] Criterion D_{KL} is defined by (6). Problem is explained in (8-11) and section 2.

FIGURE 1. Problems raised in the sequel, case $p = 1$. Such problems have already been studied, see Section 2 for more details.

the reference measure on V. Unless it is specified, V is a compact space. The resolution we propose involves a transfer between U and V. A more precise description of such transfer will be given in Section 2 (unidimensional case) and Section 3 (multidimensional case).

We first recall how the Maximum Entropy (ME) method is put in action. Originally, the ME method aims at the reconstruction of a probability measure P when dealing with information on the expectation under P of some random variables. We give below a first example.

Example 1.1. When $V = \mathbb{R}$, one may want to reconstruct a probability measure P such that the quantity $\int t^k \mathrm{d}P(t)$ for some $k \in \mathbb{N}^\star$ is equal to given values m_k.

More precisely, define the entropy of a probability measure P with respect to the measure P_V as

$$S(P) = \begin{cases} -\displaystyle\int_V \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right)\mathrm{d}P & \text{if } P \ll P_V \text{ and } \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right) \in L^1(P) \\ -\infty & \text{otherwise,}\end{cases} \tag{1.5}$$

where $P \ll P_V$ means that P is absolutely continuous with respect to P_V. ME method derives as solution the probability P^{ME} which maximises the entropy, provided that the information for the reconstructed probability measure meets the information asked.

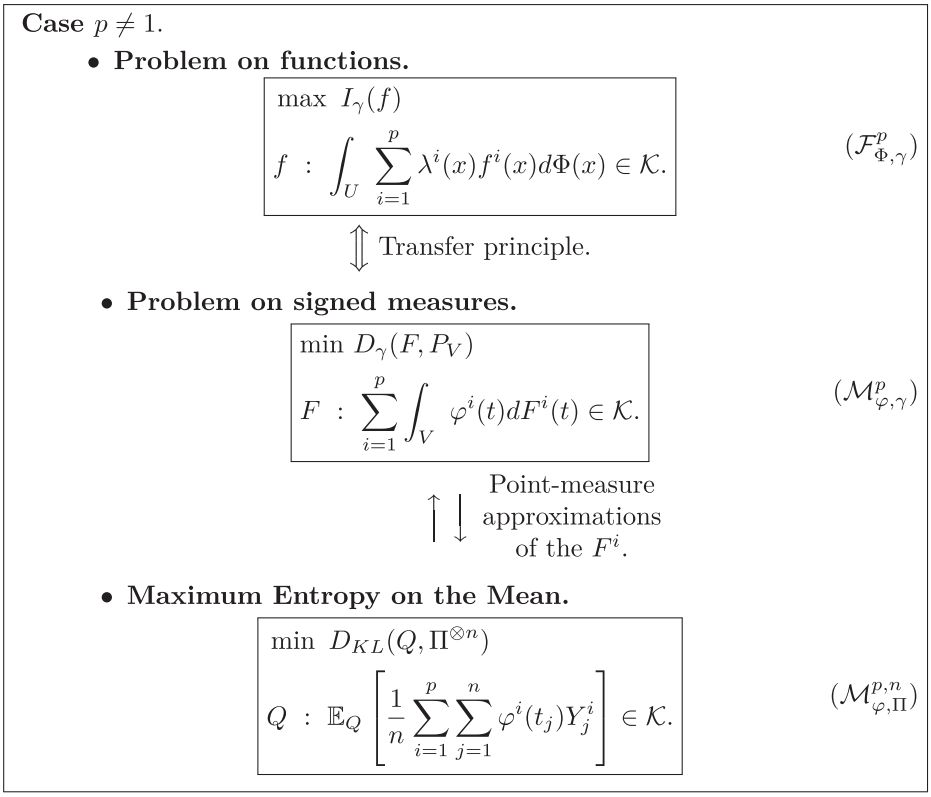

FIGURE 2. Problems raised in the sequel, $p \neq 1$. See Section 3 for the extension of the method provided by [10] to solve inverse problems.

In information theory and statistics, one usually considers the opposite of the entropy, that is the so-called Kullback–Leibler divergence of P with respect to P_V which is defined by

$$D_{\mathrm{KL}}(P, P_V) = \begin{cases} \displaystyle\int_V \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right) \mathrm{d}P & \text{if } P \ll P_V \text{ and } \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right) \in L^1(P) \\ +\infty & \text{otherwise.} \end{cases} \tag{1.6}$$

Equivalently the ME method derives as a solution the probability measure P^{ME} which minimises the Kullback–Leibler divergence from the reference measure P_V under the constraints. Reference measure P_V can be interpreted as a *prior* measure.

The Kullback–Leibler divergence defined in (1.6) is called the I-divergence in [6, 7]. The author also calls I-projection the probability measure that maximises the entropy (1.5) on a convex set of probability measures. Further in [8] an axiomatic justification for the use of the ME method is provided.

In a more general case, the entropy problem can target a reconstruction of a signed measure. In this case, the γ-divergence D_γ, defined in the expression (1.7), is considered instead of the Kullback–Leibler divergence D_{KL} (1.6). The authors in [2,3] have studied the minimisation of the γ-divergence D_γ under linear constraints. Let F be a signed measure defined on V. The classical Lebesgue decomposition of F with respect to P_V is

$$F = F^a + F^s$$

with $F^a \ll P_V$ the absolutely continuous part and F^s the singular part. F^s is singular with respect to P_V means that it is concentrated on a set $\tilde{V}$ such that $P_V(\tilde{V}) = 0$. We recall as well the Jordan decomposition of

measure F^s

$$F^s = F^{s,+} - F^{s,-}$$

with $F^{s,+}$ and $F^{s,-}$ two positive measures mutually singular. The γ-divergence D_γ is then defined as follows

$$D_\gamma(F, P_V) = \int_V \gamma\left(\frac{\mathrm{d}F^a}{\mathrm{d}P_V}\right)\mathrm{d}P_V + b_\psi F^{s,+}(V) - a_\psi F^{s,-}(V) \tag{1.7}$$

where the integrand γ is a convex function and F^a, $F^{s,+}$ and $F^{s,-}$ are as defined previously. The scalar quantities b_ψ and a_ψ, with $a_\psi < b_\psi$, are the endpoints of ψ domain with ψ the convex conjugate of γ defined by

$$\psi(t) = \sup_y \{ty - \gamma(y)\}.$$

Taking back the Example 1.1, the reconstruction problem can be put in the frame of an optimisation problem as $(\mathcal{M}^1_{\varphi,\gamma})$ defined in Figure 1. In this example, the objective function is the Kullback–Leibler divergence D_{KL}, that is the criterion D_γ when γ is the convex function defined on $\mathbb{R}^\star_+$ by $y \to y\log(y) - y - 1$. The scalar quantities a_ψ and b_ψ are respectively equal to $-\infty$ and $+\infty$ which leads to a reconstruction with no singular parts. The moment constraint can be written as $\int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}$ by taking $\mathcal{K} = \{m_k\}$ and $\varphi :\ t \to t^k$. Finally, one has to add the constraint $\int_V \mathrm{d}F(t) = 1$ to ensure that the reconstructed measure is a probability measure.

Notice that the expression in (1.7) contains terms depending on the singular part F^s of measure F. Those terms may not be considered in the γ-divergence (1.7) depending on the convex function γ used, see [5] and Example 1.1.

More generally the author in [18,19] studies the characterization of the optimal signed measure which minimises (1.7) under linear constraints. Integral functionals with normal convex integrand, *i.e.* integrals for which the integrated function is strictly convex with respect to one of its variable, are studied. See more comments on normal convex integrand in Chapter 14 of [28]. See also [20] for a systematic treatment of convex distance minimisation.

We now recall that many usual optimisation problems $(\mathcal{M}^1_{\varphi,\gamma})$ can be set in an entropy maximisation problem frame, as proposed in the early paper [23]. This general embedding in ME is called the Maximum Entropy on the Mean (MEM) method and has been developed in [11,13]. The method is based on a suitable discretisation of the working set V. The reference measure P_V is approximated by a point-measure supported by n pixels, $t_1, \ldots, t_n$ which are deterministic points in V such that $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{t_i} \to P_V$. By associated to each pixel t_i a real random amplitude Y_i, we defined the random point-measure F_n by

$$F_n = \frac{1}{n}\sum_{i=1}^n Y_i \delta_{t_i}. \tag{1.8}$$

By construction $F_n \ll P_n$. Notice that we choose to present the simple case of real random amplitudes Y_i for this introduction, but one can consider more complicated constructions. We will do so in the Section 3.3 where the amplitudes Y_i will be vectors in $\mathbb{R}^p$.

In the MEM problem, one wants to determine the "optimal" distribution Q to generate vectors of n real random amplitudes $(Y_1, \ldots, Y_n)$. Optimal distribution Q must be such that the constraints considered in problem $(\mathcal{M}^1_{\varphi,\gamma})$ applied to the random point-measure F_n is met on average, that is that

$$\mathbb{E}_Q\left[\int_V \varphi(t)\mathrm{d}F_n(t)\right] = \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n \varphi(t_i)Y_i\right] \in \mathcal{K}. \tag{1.9}$$

Taking back Example 1.1, the constraints applied to the point-measure approximation become

$$\mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] = 1 \quad \text{and} \quad \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n (t_i)^k Y_i\right] = m_k. \tag{1.10}$$

We assume that the random amplitudes are independent. Let Π denote their *prior* distribution so that the reference distribution for $(Y_1, \dots, Y_n)$ is the tensor measure $\Pi^{\otimes n}$. In order to build the optimal distribution, we minimise the Kullback–Leibler divergence (1.6) with respect to the prior $\Pi^{\otimes n}$ under the constraints (1.9). When such "optimal" distribution exists, we will denote the solution by Q_n^{MEM}. Then let F_n^{MEM} be defined as follows

$$F_n^{\mathrm{MEM}} = \mathbb{E}_{Q_n^{\mathrm{MEM}}}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i \delta_{t_i}\right]. \tag{1.11}$$

Notice that unlike F_n, the quantity F_n^{MEM} is no longer random. Let the log Laplace transform of probability measure Π be denoted by ψ_Π

$$\psi_\Pi(\tau) = \log\left(\int_{\mathbb{R}} \exp(\tau y)\, \mathrm{d}\Pi(y)\right) \qquad \text{for all } \tau \in \mathcal{D}_{\psi_\Pi}, \tag{1.12}$$

with $\mathcal{D}_{\psi_\Pi}$ the domain of ψ_Π. We denote by γ_Π the convex conjugate of the log-Laplace transform ψ_Π[1]. We hope that the reconstruction F_n^{MEM} is a good approximation (in a sense we will further specify) of the solution of the corresponding continuous problem $(\mathcal{M}^1_{\varphi,\gamma})$, for which convex function γ is the function γ_Π. The properties of the minimising sequence $(F_n^{\mathrm{MEM}})_n$ have been studied in [13]. The authors in [14] deal with a multidimensional case, that is estimating a vector of reconstructions when dealing with information on generalised moments of each components. In [15], Bayesian and MEM methods are proposed to solve inverse problems on measures. More details about the MEM method will be provided in Section 2 for the case $p = 1$ and Section 3.3 for the case $p \neq 1$. In particular we explain how to choose reference probability measure Π so that the criterion $D_{\mathrm{KL}}(\cdot, \Pi^{\otimes n})$ for discrete problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$ is a good alternative to criterion $D_\gamma(\cdot, P_V)$ of continuous problem $(\mathcal{M}^1_{\varphi,\gamma})$.

Back to the function reconstruction problem, an extension of the MEM method to solve generalised moment problems for function reconstruction as in $(\mathcal{F}^1_{\Phi,\gamma})$ is proposed by [10] in the case $p = 1$ and $\lambda(x) = 1$, $\forall x \in U$. The method uses a transfer principle which links the function to reconstruct to a corresponding measure. The transfer relies on the use of suitable kernels. Such transfer is particularly useful when considering measures Φ in the constraints equation (1.3) that might not all be absolutely continuous with respect to the reference measure P_U.

Our work has a twofold purpose. We first make a summary of entropy maximisation problems in order to reconstruct a single measure and, by extension with the linear transfer, entropy maximisation problems in order to reconstruct a single function. We propose a global synthesis of the entropy maximisation methods for such reconstruction problems from the convex analysis point of view, as well as in the framework of the embedding into the MEM setting. We then propose an extension of the entropy methods for a multidimensional case. Such extension is the main contribution of this work. We study the MEM embedding for the function reconstruction problem that is proposed in [10] in the extended case of inverse problems, that is when $p \neq 1$ and the λ^i are any known bounded continuous functions. We provide a general method of reconstruction based on the γ-entropy maximisation for functions submitted to generalised moment and interpolation constraints as in (1.3).

This paper is organized as follows.

In Section 2, we recall in a global synthesis of entropy maximisation methods some results for the specific case of a single function reconstruction problem $(\mathcal{F}^1_{\Phi,\gamma})$ or a single measure reconstruction problem $(\mathcal{M}^1_{\varphi,\gamma})$. First we describe how the transfer principle works, that is how a function problem $(\mathcal{F}^1_{\Phi,\gamma})$ can be linked to the measure problem $(\mathcal{M}^1_{\varphi,\gamma})$ in Section 2.1. Then we recall some results about the resolution of the γ-divergence minimisation problem $(\mathcal{M}^1_{\varphi,\gamma})$ in Section 2.2. In Section 2.3, we take a specific look at problem $(\mathcal{M}^1_{\varphi,\gamma})$ when the convex function γ is the function $y \to y\log(y) - y - 1$. This specific problem is the ME problem which will be denoted by $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$. We then extend the class of studied optimisation problems by giving the construction of the MEM problem setting and we provide some properties of the MEM reconstruction in Section 2.4. Finally in Section 2.5, the setting of some usual optimisation problems into the entropy maximisation frame is reminded.

[1]Notice that as ψ_Π is a convex function, ψ_Π is also the convex conjugate of γ_Π.

The main contributions of this paper are the results presented in Section 3. It consists in the study of the entropy methods for the reconstruction of a multidimensional function submitted to very general constraints such as an integral inverse problem as in (1.3). We extend the approach of [10] for the case $p \neq 1$ and any known bounded continuous functions λ^i. We study the embedding of the functions reconstruction problem $(\mathcal{F}^p_{\Phi,\gamma})$ into the MEM problem $(\mathcal{M}^{p,n}_{\varphi,\Pi})$ framework. Such study is stepped in three independent parts. First in Section 3.1, we study the problem $(\mathcal{F}^p_{\Phi,\gamma})$ in a convex analysis framework. We express the optimal solution thanks to Fenchel duality theorem. This first approach lacks to give a suitable reconstruction when constraint measures Φ are not absolutely continuous with respect to P_U. To remedy this issue, we propose in Section 3.2 to transfer the functions reconstruction problem $(\mathcal{F}^p_{\Phi,\gamma})$ to a corresponding measures reconstruction problem $(\mathcal{M}^p_{\varphi,\gamma})$. The transfer is performed using suitable continuous kernels. Finally in Section 3.3, we set problem $(\mathcal{M}^p_{\varphi,\gamma})$ obtained by the transfer into a MEM problem framework and we study the reconstructions given by problem $(\mathcal{M}^{p,n}_{\varphi,\Pi})$.

Applications will be presented in Section 4. We consider first some simple examples of single function reconstructions and then a two-functions case study inspired by computational thermodynamics.

2. The specific γ-entropy maximisation problem for a single reconstruction

We give in this section some details about the γ-entropy maximisation problem in the case of a single reconstruction, that is that we are interested in a single function or a single measure reconstruction. Those problems have already been studied, see for example in [10] for problem $(\mathcal{F}^1_{\Phi,\gamma})$, in [3] for problem $(\mathcal{M}^1_{\varphi,\gamma})$ and in [13] for problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$. Let us recall some results of these authors.

In Section 2.1 we are interested in the link between a function reconstruction problem $(\mathcal{F}^1_{\Phi,\gamma})$ and a measure reconstruction problem $(\mathcal{M}^1_{\varphi,\gamma})$. The function reconstruction problem is set as

$$\begin{aligned} &\max\ I_\gamma(f) \\ &f:\ \int_U f(x)\mathrm{d}\Phi(x) \in \mathcal{K} \end{aligned} \tag{$\mathcal{F}^1_{\Phi,\gamma}$}$$

and the measure reconstruction problem as

$$\begin{aligned} &\min\ D_\gamma(F, P_V) \\ &F:\ \int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}. \end{aligned} \tag{$\mathcal{M}^1_{\varphi,\gamma}$}$$

The idea is to set a transformation from measures on V to functions on U. Such transformation is the linear transfer we will further describe in the Section 2.1.

We remind the reader that we consider a number N of constraints. Therefore Φ and φ take values in $\mathbb{R}^N$. In addition we will consider that φ is continuous.

In Section 2.2 we study the γ-divergence minimisation problem under constraints, that is problem $(\mathcal{M}^1_{\varphi,\gamma})$. We recall the results of [3] for the existence of an optimal solution to problem $(\mathcal{M}^1_{\varphi,\gamma})$.

We take a closer look in Section 2.3 at the Maximum Entropy (ME) problem. This problem corresponds to problem $(\mathcal{M}^1_{\varphi,\gamma})$ in the special setting when γ is the function $y \to y\log(y) - y - 1$ and the γ-divergence coincides with the Kullback–Leibler divergence. We recall results on the existence of the optimum and its expression when such minimiser exists.

In Section 2.4, we recall the setting of the nth MEM problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$

$$\begin{aligned} &\min\ D_{\mathrm{KL}}(Q, \Pi^{\otimes n}) \\ &Q:\ \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n \varphi(t_i)Y_i\right] \in \mathcal{K}. \end{aligned} \tag{$\mathcal{M}^{1,n}_{\varphi,\Pi}$}$$

We give, when it exists, an expression for the minimiser Q_n^{ME} of the nth problem. A function g_n^{ME}, related to the expectation of the random amplitudes under Q_n^{ME}, can be defined. We will see that under some assumptions, it exists one particular $v \in \mathbb{R}^N$ such that the sequence $(g_n^{\mathrm{ME}})_n$ converges to a function $g_\infty^{\mathrm{ME}}(t) = \psi'_\Pi(\langle v, \varphi(t)\rangle)$. This limit will be expressed.

Finally in Section 2.5, we give examples of some classical optimisation problems embedding into the MEM framework.

2.1. Transfer principle

We briefly recall the idea of the transfer principle developed in [10], in order to put to work the γ-entropy methods in the case of a function reconstruction problem. Let us denote by $(U, \mathcal{B}(U), P_U)$ the probability space where $U \subset \mathbb{R}^d$ is compact (non empty), $\mathcal{B}(U)$ is the associated Borel set and P_U is the reference measure. In the case of the reconstruction problem for a single function, one wants to reconstruct over U a function f taking values in $\mathbb{R}$ such that f satisfies the integral constraints

$$\int_U f(x)\mathrm{d}\Phi(x) \in \mathcal{K}. \tag{2.1}$$

Let γ be a given convex function taking values in $\mathbb{R}$ and $\mathcal{D}_\gamma \subset \mathbb{R}$ its domain. Then the γ-entropy is defined by

$$I_\gamma(f) = -\int_U \gamma\left(f(x)\right) \mathrm{d}P_U(x),$$

for a function f defined on U and taking values in $\mathcal{D}_\gamma$.

In order to chose among the functions that satisfy (2.1), we propose as a selection criterion to maximise the γ-entropy. This means that we consider the optimisation problem $(\mathcal{F}^1_{\Phi,\gamma})$

$$\begin{aligned} &\max\ I_\gamma(f) \\ &f:\ \int_U f(x)\mathrm{d}\Phi(x) \in \mathcal{K}. \end{aligned} \tag{$\mathcal{F}^1_{\Phi,\gamma}$}$$

The method proposed by [10] is to transfer the function reconstruction problem to a measure reconstruction problem thanks to some continuous kernel K. Such kernel links the function to reconstruct to a corresponding signed measure. Recall that V is a Polish space and P_V the reference probability measure on V. The idea is that if one can reconstruct over V a signed measure F such that

$$\int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}, \tag{2.2}$$

then a regularized function f_K can be reconstructed linked to the measure F. To do so, we proceed as follows.

We denote by K a continuous kernel defined over $U \times V$ taking values in $\mathbb{R}$. The kernel K is such that measure Φ of the integral constraint (2.1) is linked to a regularized function φ_K involved in an integral constraint as in (2.2). The relation linking Φ to φ_K is given by

$$\varphi_K(t) = \int_U K(x,t)\mathrm{d}\Phi(x), \qquad t \in V.$$

Therefore, for any continuous kernel K, one can reconstruct the regularized function f_K associated to F by defining

$$f_K(x) = \int_V K(x,t)\mathrm{d}F(t), \qquad x \in U.$$

Hence as a consequence of Fubini theorem, if the measure F satifies (2.2), the regularized function f_K defined above satisfies (2.1).

2.2. γ-divergence minimisation problem

In this section, we recall the results provided by [3] for the γ-divergence minimisation problem for signed measure reconstruction. We set our work on V. Let F be a signed measure defined on V. The classical Lebesgue decomposition of F with respect to P_V is

$$F = F^a + F^s$$

with $F^a \ll P_V$ the absolutely continuous part and F^s the singular part. F^s singular with respect to P_V means that it is concentrated on a set $\tilde{V}$ such that $P_V(\tilde{V}) = 0$. Notice that F^a and F^s are still signed measures. Recall the Jordan decomposition of measure F^s

$$F^s = F^{s,+} - F^{s,-}$$

with $F^{s,+}$ and $F^{s,-}$ mutually singular. Let γ be essentially strictly convex. We denote by ψ its convex conjugate, that is

$$\psi(t) = \sup_{y \in \mathcal{D}_\gamma} \{ty - \gamma(y)\}.$$

Considering a_ψ and b_ψ, with $a_\psi < b_\psi$, the endpoints of ψ domain, we define the γ-divergence D_γ by

$$D_\gamma(F, P_V) = \int_V \gamma\left(\frac{\mathrm{d}F^a}{\mathrm{d}P_V}\right) \mathrm{d}P_V + b_\psi F^{s,+}(V) - a_\psi F^{s,-}(V).$$

The problem we consider is the following

$$\begin{aligned} &\min \ D_\gamma(F, P_V) \\ &F : \int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}. \end{aligned} \tag{$\mathcal{M}^1_{\varphi,\gamma}$}$$

The authors in [3] study the existence conditions of an optimal solution using convex analysis tools. Their first result is to consider the following dual problem of $(\mathcal{M}^1_{\varphi,\gamma})$ which relies on a Lagrange multiplier $v \in \mathbb{R}^N$

$$\sup_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \mathcal{K}} \langle v, c\rangle - \int_V \psi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \right\}. \tag{$\mathcal{M}^{1,\star}_{\varphi,\gamma}$}$$

In addition, they give conditions for problems $(\mathcal{M}^1_{\varphi,\gamma})$ and $(\mathcal{M}^{1,\star}_{\varphi,\gamma})$ to have solutions. These results are recalled in Theorem 2.1 below.

Theorem 2.1 ([3], Thm. 3.4)**.**

(1) *If there exists $v' \in \mathbb{R}^N$ such that $\langle v', \varphi(t)\rangle \in \mathcal{D}_\psi$ then*

$$\inf_{F:\ \int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}} D_\gamma(F, P_V) \ = \ \sup_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \mathcal{K}} \langle v, c\rangle - \int_V \psi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \right\}.$$

(2) *If, in addition, $\int_V \psi\left(\langle v', \varphi(t)\rangle\right) \mathrm{d}P_V(t)$ is finite and if it exists a signed measure F' such that $\int_V \varphi(t)\mathrm{d}F'(t) \in \mathcal{K}$, $\frac{\mathrm{d}F'}{\mathrm{d}P_V}$ is in the relative interior of $\mathcal{D}_\gamma$ and $D_\gamma(F', P_V)$ is finite, then*

$$\inf_{F:\ \int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}} D_\gamma(F, P_V) \ = \ \min_{F:\ \int_V \varphi(t)\mathrm{d}F(t) \in \mathcal{K}} D_\gamma(F, P_V)$$

and

$$\sup_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \mathcal{K}} \langle v, c\rangle - \int_V \psi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \right\} = \max_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \mathcal{K}} \langle v, c\rangle - \int_V \psi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \right\}.$$

The next theorem proposes a more precise characterisation of the solution for problem $(\mathcal{M}^1_{\varphi,\gamma})$ under the same conditions specified in Theorem 2.1.

Theorem 2.2 ([3], Thm. 4.1)**.** *Under the assumptions of Theorem 2.1.*

(1) *The absolutely continuous part with respect to P_V of solution F^o of $(\mathcal{M}^1_{\varphi,\gamma})$ is given by*

$$\frac{\mathrm{d}F^o}{\mathrm{d}P_V}(t) = \psi'\left(\langle v^\star, \varphi(t)\rangle\right)$$

where $v^\star$ is the solution of

$$\max_{v\in\mathbb{R}^N}\left\{\inf_{c\in\mathcal{K}}\langle v,c\rangle - \int_V \psi\left(\langle v,\varphi(t)\rangle\right)\mathrm{d}P_V(t)\right\}.$$

(2) *If, in addition, for all $v\in\mathbb{R}^N$ and for all $t\in V$, $\langle v,\varphi(t)\rangle$ is in the interior of $\mathcal{D}_\psi$, the singular part vanishes.*

It can be noted that when $\mathcal{D}_\psi = \mathbb{R}$, Theorem 2.2 always gives solutions that are absolutely continuous with respect to P_V. The condition to have $\mathcal{D}_\psi = \mathbb{R}$, is to consider a function γ that is such that the ratio $|\frac{\gamma(y)}{y}|$ is equal to ∞ on the edges of $\mathcal{D}_\gamma$, see Lemma 2.1 of [3].

We will see in Section 2.4 that the approach proposed by the embedding in the MEM framework boils down to the same results provided by Theorem 2.2.

2.3. Maximum entropy problem

In this section we take a better look at the problem of maximising the entropy of a probability measure under generalised moments constraints, that is the ME problem

$$\begin{aligned}&\min\ D_{\mathrm{KL}}(P,P_V)\\ &P:\ \int_V \varphi(t)\mathrm{d}P(t)\in\mathcal{K}\end{aligned}\qquad(\mathcal{M}^1_{\varphi,\mathrm{ME}})$$

with φ defined on V and taking values in $\mathbb{R}^N$. We remind the reader that in this section, the first component of function φ is the constant 1 and that $\mathcal{K}$ is the product $\{1\}\times\mathcal{K}_1\times\cdots\times\mathcal{K}_{N-1}$. Notice that for the results recalled in this section only, V does not need to be compact and φ has not to be continuous. The definition of the Kullback–Leibler divergence D_{KL} is recalled below

$$D_{\mathrm{KL}}(P,P_V) = \begin{cases}\displaystyle\int_V \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right)\mathrm{d}P & \text{if } P\ll P_V \text{ and } \log\left(\frac{\mathrm{d}P}{\mathrm{d}P_V}\right)\in L^1(P)\\ +\infty & \text{otherwise.}\end{cases}$$

The problem proposed in Section 2.2 is a more general setting of the original ME problem. The reconstruction provided by the ME method satisfies the following properties. Those are showed by Shore and Johnson in [29].

Proposition 2.3.

(1) ***Uniqueness:*** *If the solution of ME problem exists, it is unique.*
(2) ***Coordinate independence:*** *The reconstruction is independent of coordinate system choice.*
(3) ***System independence:*** *If the probability space $(V,\mathcal{B}(V),P_V)$ consists in the product space of m probability spaces, the reconstruction over the whole probability space is the tensor product of the reconstructions on each probability space.*
In other words, if $P_V = \otimes_{i=1}^m P_{V_i}$ where P_{V_i} is a reference probability measure on $(V_i,\mathcal{B}(V_i))$ and $(V,\mathcal{B}(V),P_V)$ is the product space of all $(V_i,\mathcal{B}(V_i),P_{V_i})$, then the reconstruction P on $(V,\mathcal{B}(V))$ is given by $\otimes_{i=1}^m P_i$ where P_i is the reconstruction on $(V_i,\mathcal{B}(V_i))$.

(4) ***Subset independence:*** *If probability space $(V, \mathcal{B}(V), P_V)$ consists in the union of m probability spaces, the reconstruction over the global space leads to the same measure than the reconstruction problem conditioning on each probability space.*
In other words, it does not matter whether one treats the information as a subset V_j of whole set V in a conditional constraint or in the full system.

We recall in this section some results of [9,13] in order to solve problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$. Results are stated without any proof for the setting of our example.

We give first a definition of the generalised solution for problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$. This definition requires the use of a minimising sequence of $D_{\mathrm{KL}}(\cdot, P_V)$, defined as follows. Let (P_n) be a sequence of probability measures. (P_n) is a minimising sequence of $D_{\mathrm{KL}}(\cdot, P_V)$ if we have

$$\lim_{n\to\infty} D_{\mathrm{KL}}(P_n, P_V) = \inf_P D_{\mathrm{KL}}(P, P_V). \tag{2.3}$$

Definition 2.4 ([9])**.** Let us consider a sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ defined on $(V, \mathcal{B}(V))$ such that (P_n) converges and is a minimising sequence of $D_{\mathrm{KL}}(\cdot, P_V)$ and such that for all n, probability measure P_n satisfies

$$\int_V \varphi(t)\mathrm{d}P_n(t) \in \mathcal{K}.$$

Then we call generalised maximal entropy solution the measure P^{MEG} that is such that

$$P^{\mathrm{MEG}} = \lim_{n\to\infty} P_n, \qquad \text{in total variation.}$$

If P^{MEG} also meets the constraint, then it is called the maximal entropy solution of problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$ denoted by P^{ME}.

We will now recall some results on the existence of a generalised solution for problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$ and the shape of the solution when it exists.

Let us first define $\mathcal{P}_{\mathcal{K}}$ the subset of probability measures that satisfy the constraints of the ME problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$, that is

$$\mathcal{P}_{\mathcal{K}} = \left\{ P \text{ probability measure on } V, \text{ such that } \int_V \varphi(t)\mathrm{d}P(t) \in \mathcal{K} \right\}.$$

With the previous definition of P^{MEG}, we recall a result of [9] for the existence in our framework of the generalised solution of problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$.

Lemma 2.5 ([9])**.** *If $\inf_{P\in\mathcal{P}_{\mathcal{K}}} D_{\mathrm{KL}}(P, P_V) < \infty$, then P^{MEG} exists.*

Let us now introduce several definitions involved in the characterisation of the problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$ solution. For all $v \in \mathbb{R}^N$, we define the quantity

$$Z_{P_V,\varphi}(v) = \int_V \exp\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \tag{2.4}$$

where $\langle\cdot,\cdot\rangle$ is the usual scalar product in $\mathbb{R}^N$. We denote the domain of $Z_{P_V,\varphi}$ by $D_{P_V,\varphi}$, that is the subset of vectors in $\mathbb{R}^N$ that are such that $Z_{P_V,\varphi}(\cdot)$ is finite

$$D_{P_V,\varphi} = \left\{ v \in \mathbb{R}^N,\ Z_{P_V,\varphi}(v) < \infty \right\}. \tag{2.5}$$

The following definition describes the so-called exponential familly with respect to probability measure P_V. The interested reader may be referred to [1] for more details about exponential models.

Definition 2.6. The φ-Hellinger arc of P_V is a family of measures P_v that are defined by

$$\mathrm{d}P_v(t) = \left(Z_{P_V,\varphi}(v)\right)^{-1} \exp\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t) \tag{2.6}$$

for all $v \in D_{P_V,\varphi}$.

The family of measures P_v defined as in (2.6) for all $v \in D_{P_V,\varphi}$ may also be called the exponential model with respect to P_V.

We recall below the Theorem 4 from [7] that characterises the generalised reconstruction P^{MEG}. This theorem describes the reconstruction P^{MEG} as an element of the φ-Hellinger arc of P_V. More important, the reconstruction problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$, which is infinite dimensional, is transformed into the finite dimension problem (2.7) that considers the vectors v in $D_{P_V,\varphi}$, see expression (2.5).

Theorem 2.7 ([7], Thm. 4)**.** *The reconstruction P^{MEG} belongs to the φ-Hellinger arc of P_V if and only if it exists one measure P in $\mathcal{P}_\mathcal{K}$ such that $P \ll P_V$.*

Then, defines

$$H_{P_V,\varphi}(v,\mathcal{K}) = \inf_{c\in\mathcal{K}} \langle v, c\rangle - \log\left(Z_{P_V,\varphi}(v)\right), \qquad \forall v \in D_{P_V,\varphi}$$

the reconstruction P^{MEG} is obtained by determining $v^\star \in D_{P_V,\varphi}$ such that

$$H_{P_V,\varphi}(v^\star,\mathcal{K}) = \sup_{v\in D_{P_V,\varphi}} H_{P_V,\varphi}(v,\mathcal{K}). \tag{2.7}$$

The previous theorem does not ensure that the reconstruction P^{MEG} will satisfy the constraints of problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$. Of course, the reconstruction P^{MEG} is more interesting when P^{MEG} belongs to $\mathcal{P}_\mathcal{K}$. We then have the following corollary for a reconstruction that satisfies the constraints. Such reconstruction is then denoted by P^{ME}.

Corollary 2.8 ([13])**.** *If there exists a measure $P \in \mathcal{P}_\mathcal{K}$ such that $P \ll P_V$ and if $D_{P_V,\varphi}$ is an open set, then P^{MEG} is the reconstruction P^{ME} in $\mathcal{P}_\mathcal{K}$ and P^{ME} belongs to the φ-Hellinger arc of P_V.*

As an illustration, we propose the following simple example of a probability measure reconstruction which maximises the entropy with respect to the standard Gaussian distribution $\mathcal{N}(0,1)$. The added constraint is a fixed valued for the first order moment. By giving the first order moment equal to m, the reconstruction we obtain is, as one can expect, the Gaussian distribution centred in m and with unit variance.

Example 2.9. The working probability space is $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathcal{N})$ where $\mathcal{N}$ is the standard Gaussian distribution $\mathcal{N}(0,1)$. We wish to reconstruct the probability measure with given first order moment which minimises the Kullback–Leibler divergence. Our problem is rewritten

$$\begin{aligned} &\min\ D_{\mathrm{KL}}(P,\mathcal{N}) \\ &\text{s.t.} \int_{\mathbb{R}} \mathrm{d}P(t) = 1 \\ &\qquad \int_{\mathbb{R}} t\mathrm{d}P(t) = m. \end{aligned}$$

Using the Theorem 2.7 recalled previously the problem becomes

$$\max_{v\in\mathbb{R}} vm - \log\left(\int_{\mathbb{R}} \exp(vt)\mathrm{d}\mathcal{N}(t)\right). \tag{2.8}$$

Classical optimality criterion applied to the maximisation problem (2.8) gives $v = m$. Then, the Radon–Nikodym derivative, see [24, 26], with respect to $\mathcal{N}$ for the reconstructed probability measure P^{ME} is equal to

$$\frac{\mathrm{d}P^{\mathrm{ME}}}{\mathrm{d}\mathcal{N}}(t) = \exp\left(-\frac{m^2}{2} + mt\right).$$

P^{ME} is therefore the Gaussian distribution $\mathcal{N}(m,1)$.

2.4. Maximum Entropy on the Mean

We recall in this section how the MEM method works. Let us first recall the problem studied. We assume here that V is compact. V is discretised with a suitable deterministic sequence $t_1, \dots, t_n$ such that point probability measure $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{t_i}$ approximates well the probability measure P_V. We denote by Q a distribution that generates a vector $(Y_1, \dots, Y_n)$ of n real random amplitudes and by F_n the random point measure defined by

$$F_n = \frac{1}{n}\sum_{i=1}^n Y_i \delta_{t_i}. \tag{1.8}$$

One wants to determine the "optimal" distribution Q to generate $Y_1, \dots, Y_n$ such that point-measure F_n meets on average the constraints defined as follows

$$\mathbb{E}_Q\left[\int_V \varphi(t)\mathrm{d}F_n(t)\right] = \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n \varphi(t_i)Y_i\right] \in \mathcal{K},$$

with φ a continuous function taking values in $\mathbb{R}^N$.

We set Π a given reference distribution on $\mathbb{R}$ and we study the Kullback–Leibler divergence between the joint distribution Q and the tensor distribution $\Pi^{\otimes n}$ under some constraints. This is summed up in a more concise way by the following problem

$$\begin{aligned} &\min\ D_{\mathrm{KL}}(Q, \Pi^{\otimes n}) \\ &Q:\ \mathbb{E}_Q\left[\frac{1}{n}\sum_{i=1}^n \varphi(t_i)Y_i\right] \in \mathcal{K}. \end{aligned} \tag{$\mathcal{M}^{1,n}_{\varphi,\Pi}$}$$

We assume the support of Π to be $]a;b[$ with $-\infty \leqslant a < b \leqslant \infty$. The domain of the moment generating function of Π is denoted D_Π

$$D_\Pi = \left\{\tau \in \mathbb{R}:\ \int_a^b \exp(\tau y)\mathrm{d}\Pi(y) < \infty\right\}.$$

We denote by ψ_Π the logarithm of the moment generating function of Π

$$\psi_\Pi(\tau) = \log\int_a^b \exp(\tau y)\mathrm{d}\Pi(y), \qquad \forall \tau \in D_\Pi. \tag{2.9}$$

We recall below some results of [13] for the resolution of the nth MEM problem and the convergence of the obtained solution. First Section III.3, Lemma 3.1 of [13] recalled below gives sufficient conditions for the existence of a solution to the MEM problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$.

Lemma 2.10 ([13], Sect. III.3, Lem. 3.1)**.** *Let us assume the following*

(H1): *It exists at least one $c \in \mathcal{K}$ for which one can find $y_1, \dots, y_n$ with $y_i \in]a;b[$ for $i = 1, \dots, n$, that are such that $\frac{1}{n}\sum_{i=1}^n \varphi(t_i)y_i = c$.*

Then, for n sufficiently large, it always exists a solution to the MEM problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$.

The Corollary 3.1 of Section III.3 from [13] describes the solution of the MEM problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$ when it exists. As for problem $(\mathcal{M}^1_{\varphi,\mathrm{ME}})$, solving the reconstruction problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$ consists in solving a finite dimension problem that considers the vectors v in $D_{\Pi,\varphi}$, with $D_{\Pi,\varphi}$ the subset of $\mathbb{R}^N$ defined in (H3).

Corollary 2.11 ([13], Sect. III.3, Cor. 3.1)**.** *Let us assume that assumption (H1) is satisfied and the following*

(H2): *The closed convex hull of Π support is $[a;b]$.*

(H3): *Let the set $D_{\Pi,\varphi}$ be a non empty open set where $D_{\Pi,\varphi}$ is defined by*

$$D_{\Pi,\varphi} = \left\{ v \in \mathbb{R}^N,\ Z_{\Pi,\varphi}(v) = \int_{[a;b]^n} \prod_{i=1}^{n} \exp\left(\langle v, \varphi(t_i)\rangle y_i\right) \mathrm{d}\Pi^{\otimes n}(y_1,\ldots,y_n) < \infty \right\}.$$

Then, for n sufficiently large, the solution Q_n^{ME} to the MEM problem $(\mathcal{M}_{\varphi,\Pi}^{1,n})$ is

$$Q_n^{\mathrm{ME}}(y_1,\ldots,y_n) = \left(Z_{\Pi,\varphi}(v^\star)\right)^{-1} \exp\left(\sum_{i=1}^{n} \langle v^\star, \varphi(t_i)\rangle y_i\right) \Pi^{\otimes n}(y_1,\ldots,y_n)$$

where $v^\star \in D_{\Pi,\varphi} \subset \mathbb{R}^N$ is the unique maximiser of

$$H_{\Pi,\varphi}^{n}(v,\mathcal{K}) = \inf_{c\in\mathcal{K}} \langle v, c\rangle - \log\left(Z_{\Pi,\varphi}(v)\right).$$

For the convergence result of the MEM reconstruction, we define the function g_n^{ME} by

$$g_n^{\mathrm{ME}}(t) = \frac{1}{\sharp M_n(t)} \sum_{i\in M_n(t)} \mathbb{E}_{Q_n^{\mathrm{ME}}}[Y_i]$$

with $M_n(t)$ the subset of indices in $[|1;n|]$ defined by

$$M_n(t) = \left\{ j \in [|1;n|] :\ ||t - t_j|| = \min_{i\in[|1;n|]} ||t - t_i|| \right\}$$

and $\sharp M_n(t)$ is the number of elements in $M_n(t)$.

The next theorem requires the strong assumption denoted by (H6). The notation ∂ of assumption $(H6)$ refers to the edge of the set.

Theorem 2.12 ([13], Sect. III.4, Thm. 3.1)**.** *Under assumptions (H1), (H2) and*

(H4): *D_Π is a non empty open set and it exists $v \in \mathbb{R}^k$ such that for all $t \in V$, $\langle v, \varphi(t)\rangle \in D_\Pi$.*
(H5): *$D(\Pi,\varphi) = \left\{ v \in \mathbb{R}^N, \forall t \in V,\ \langle v, \varphi(t)\rangle \in D_\Pi \right\}$ is non empty.*
(H6): *$\forall v \in \partial D(\Pi,\varphi)$, we have $\lim_{u\in D(\Pi,\varphi),\ u\to v} \left| \int_V \varphi(t)\psi'_\Pi\left(\langle u, \varphi(t)\rangle\right) \mathrm{d}P_V(t)\right| = +\infty$.*
Then g_n^{ME} converges to g_∞^{ME}

$$g_\infty^{\mathrm{ME}}(t) = \psi'_\Pi\left(\langle v^\star, \varphi(t)\rangle\right)$$

where $v^\star$ maximises

$$H_{\Pi,\varphi}^{\infty}(v,\mathcal{K}) = \inf_{c\in\mathcal{K}} \langle v, c\rangle - \int_V \psi_\Pi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t).$$

Remark 2.13. Let the real random sequence (X_n) be defined by $X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i\varphi(t_i)$ and let us denote by Q_n the law of X_n. Under the assumptions (H1) and (H2) and provided that ψ_Π is sufficiently regular, one can characterise the asymptotic behaviour of Q_n. As n tends to infinity, Q_n tends to concentrate on the events that belong to the compact set $\mathcal{K}$. That is, let $x \in \mathcal{K}$, we have that

$$Q_n(X_{n,l} > x_l, l = 1,\ldots,N) \approx \exp\left(-nI(x)\right) \tag{2.10}$$

where the rate function is $I(x) = \sup_{v\in\mathbb{R}^N}\left\{\langle v, x\rangle - \int_V \psi_\Pi\left(\langle v, \varphi(t)\rangle\right) \mathrm{d}P_V(t)\right\}$. Here, the approximation is related to the classical large deviations property, see [12].

This remark is more formally given in Section III.4, Corollary 3.5 of [13] as the large deviation property of the sequence (Q_n).

Thereafter in the Section 3, we aim to reconstruct the p components of a vectorial measure F subject to the following integral constraints

$$\sum_{i=1}^{p} \int_V \varphi_l^i(t) \mathrm{d}F^i(t) \in \mathcal{K}_l, \qquad 1 \leqslant l \leqslant N.$$

We will follow the MEM construction given in the present section. For the multidimensional case, the random measure F_n will then be vectorial and the sequence $(Y_i)_{i=1,\dots,n}$ will be a sequence of vectorial amplitudes in $\mathbb{R}^p$.

2.5. Connection with classical minimisation problems

One can notice that the link between the γ-entropy maximisation problem $(\mathcal{M}^1_{\varphi,\gamma})$ and MEM problem $(\mathcal{M}^{1,n}_{\varphi,\Pi})$. Indeed if one chooses the convex function γ involved in problem $(\mathcal{M}^1_{\varphi,\gamma})$ to be the convex conjugate of ψ_Π, we have that

$$\max_{v \in D_{\Pi,\varphi,\psi_\Pi}} H^\infty_{\Pi,\varphi}(v, \mathcal{K}) = \inf_{F \in \mathcal{F}_\mathcal{K}} D_{\gamma_\Pi}(F, P_V)$$

where $H^\infty_{\Pi,\varphi}$ is the function defined in the Theorem 2.12.

In this section we detail some classical minimisation problems set in the MEM embedding. We set our work under the assumptions of Theorem 2.12.

Poisson distribution and Kullback–Leibler divergence minimisation

Let the reference distribution Π be a Poisson distribution $\mathcal{P}(\theta)$ with parameter $\theta \in]0; +\infty[$. The support of Π is $\mathbb{N}$. The log-Laplace transform ψ of a Poisson distribution $\mathcal{P}(\theta)$ is

$$\psi_{\mathcal{P}(\theta)}(\tau) = \theta(e^\tau - 1)$$

with domain $D_\Pi = \mathbb{R}$.

Its convex conjugate is the following function

$$\gamma_{\mathcal{P}(\theta)}(y) = y \log(y) - y(1 + \log(\theta)) + \theta, \quad \text{for } y \in \mathbb{R}_+.$$

The associated convex criterion to minimise for problem $(\mathcal{M}^1_{\varphi,\gamma})$ becomes

$$D_{\gamma_{\mathcal{P}(\theta)}}(F, P_V) = \begin{cases} \displaystyle\int_V \log\left(\frac{\mathrm{d}F}{\mathrm{d}P_V}\right) \mathrm{d}F - (1 + \log(\theta))\, F(V) + \theta, & \text{if } F \ll P_V \text{ and } \frac{\mathrm{d}F}{\mathrm{d}P_V} \geqslant 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Such criterion gives the Kullback–Leibler divergence when $\theta = 1$

$$D_{\mathrm{KL}}(F, P_V) = \begin{cases} \displaystyle\int_V \log\left(\frac{\mathrm{d}F}{\mathrm{d}P_V}\right) \mathrm{d}F - F(V) + 1, & \text{if } F \ll P_V \text{ and } \frac{\mathrm{d}F}{\mathrm{d}P_V} \geqslant 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Gaussian distribution and least squares minimisation

Let the reference distribution Π be a Gaussian distribution $\mathcal{N}(m, \sigma^2)$, $\sigma^2 > 0$. The support of Π is $]a; b[=]-\infty; +\infty[$. The log-Laplace transform ψ of a Gaussian distribution $\mathcal{N}(m, \sigma^2)$ is

$$\psi_{\mathcal{N}(m,\sigma^2)}(\tau) = \frac{\tau^2\sigma^2}{2} + \tau m$$

with domain $D_\Pi = \mathbb{R}$.

Its convex conjugate is the function

$$\gamma_{\mathcal{N}(m,\sigma^2)}(y) = \frac{(y-m)^2}{2\sigma^2}, \quad \text{for } y \in \mathbb{R}.$$

The associated convex criterion to minimise for problem $(\mathcal{M}^1_{\varphi,\gamma})$ becomes

$$D_{\gamma_{\mathcal{P}(\theta)}}(F,P_V)=\begin{cases}\displaystyle\int_V \frac{1}{2\sigma^2}\left(\frac{\mathrm{d}F}{\mathrm{d}P_V}-m\right)^2\mathrm{d}P_V, & \text{if } F\ll P_V\\ +\infty & \text{otherwise}\end{cases}$$

which gives the minimisation problem consisting in finding the least squares deviation of Radon–Nikodym derivative $\frac{\mathrm{d}F}{\mathrm{d}P_V}$ from constant m.

Exponential distribution and Burg entropy minimisation

Let the reference distribution Π be an exponential distribution $\mathcal{E}(\theta)$ with mean θ, (where $\theta>0$). The support of Π is $[0;+\infty[$. The log-Laplace transform ψ of the exponential distribution $\mathcal{E}(\theta)$ is

$$\psi_{\mathcal{E}(\theta)}(\tau)=-\log(1-\tau\theta)$$

with domain $D_\Pi=\left\{\tau<\frac{1}{\theta}\right\}$.

Its convex conjugate is the function

$$\gamma_{\mathcal{E}(\theta)}(y)=\log\left(\frac{\theta}{y}\right)+\frac{y}{\theta}-1,\quad \text{for } y\in\mathbb{R}^\star_+.$$

The associated convex criterion to minimise for problem $(\mathcal{M}^1_{\varphi,\gamma})$ becomes

$$D_{\gamma_{\mathcal{P}(\theta)}}(F,P_V)=\begin{cases}\log(\theta)-\displaystyle\int_V\log\left(\frac{\mathrm{d}F}{\mathrm{d}P_V}\right)\mathrm{d}P_V+\frac{F(V)}{\theta}-1, & \text{if } F\ll P_V \text{ and } F^s\geqslant 0\\ +\infty & \text{otherwise.}\end{cases}$$

Such criterion gives the Burg-entropy of F when $\theta=1$, which is the reverse Kullback–Leibler divergence of P_V with respect to F

$$D_{\mathrm{KL}}(P_V,F)=\begin{cases}-\displaystyle\int_V\log\left(\frac{\mathrm{d}F}{\mathrm{d}P_V}\right)\mathrm{d}P_V+F(V)-1, & \text{if } F\ll P_V \text{ and } F^s\geqslant 0\\ +\infty & \text{otherwise.}\end{cases}$$

3. The γ-entropy maximisation problem for the reconstruction of a multidimensional function

In this section we propose to study the γ-entropy maximisation method for the reconstruction of p real-valued functions with domain U (with U compact and non-empty) when they are subjected to very general constraints such as

$$\int_U\sum_{i=1}^p\lambda^i(x)f^i(x)\mathrm{d}\Phi_l(x)\in\mathcal{K}_l\qquad 1\leqslant l\leqslant N. \tag{3.1}$$

This study is performed in three independent parts.

First in Section 3.1, we study in the convex analysis framework the problem $(\mathcal{F}^p_{\Phi,\gamma})$ recalled below

$$\begin{aligned}&\max\ -\int_U\gamma\left(f^1,\ldots,f^p\right)\mathrm{d}P_U\\ &f:\ \int_U\sum_{i=1}^p\lambda^i(x)f^i(x)\mathrm{d}\Phi_l(x)\in\mathcal{K}_l,\quad l=1,\ldots,N.\end{aligned}\tag{$\mathcal{F}^p_{\Phi,\gamma}$}$$

We detail the construction of a dual problem of finite dimension. We are able to express an optimal solution thanks to Fenchel duality theorem when the constraint measures Φ_l are absolutely continuous with respect

to P_U. However, we show that this first approach does not give a suitable reconstruction when the constraint measures Φ_l are not absolutely continuous with respect to P_U.

To remedy this issue, we propose in Section 3.2 to transfer the functions reconstruction problem $(\mathcal{F}^p_{\Phi,\gamma})$ to a corresponding reconstruction problem $(\mathcal{M}^p_{\varphi,\gamma})$ on signed measures. Such problem is recalled below

$$\begin{aligned} &\min\ D_\gamma(F, P_V) \\ &F:\ \sum_{i=1}^{p} \int_V \varphi^i(t)\mathrm{d}F^i(t) \in \mathcal{K}. \end{aligned} \tag{$\mathcal{M}^p_{\varphi,\gamma}$}$$

The linear transfer is performed by means of suitable continuous kernels.

Finally in Section 3.3, we set the problem $(\mathcal{M}^p_{\varphi,\gamma})$ obtained with the linear transfer into a MEM problem framework. We detail the construction of a sequence of random point-measures for the multidimensional framework. We study the reconstructions given by problem $(\mathcal{M}^{p,n}_{\varphi,\Pi})$.

3.1. The γ-entropy maximisation problem for the multidimensional case in the convex analysis framework

In this section we study the way to reconstruct a multidimensional function subject to an inverse problems by a γ-entropy maximisation approach. We propose to study such approach within the framework of convex analysis. We recall some general definitions and properties further used in the Appendix A.

We frame our work in a probability space $(U, \mathcal{B}(U), P_U)$ where U is compact (non empty) and P_U is the reference probability measure. We aim at the reconstruction of p real-valued functions such that

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N. \tag{3.2}$$

As there exists many solutions fitting the previous constraints, we wish to choose among them the solution of a convex problem under constraints. Given a closed convex function $\gamma:\ \mathbb{R}^p \to [0; +\infty]$, we decide to characterize the optimal solution $f^o = (f^{1,o}, \ldots, f^{p,o})$ that maximises the γ-entropy defined by

$$-\int_U \gamma\left(f^1(x), \ldots, f^p(x)\right) \mathrm{d}P_U(x)$$

provided that f^o satisfies the constraints stated in (3.2). We denote by $I_\gamma(\cdot)$ the opposite of the γ-entropy that is

$$I_\gamma(f) = \int_U \gamma\left(f^1(x), \ldots, f^p(x)\right) \mathrm{d}P_U(x).$$

To put it in a more concise way, we study problem $(\mathcal{F}^p_{\Phi,\gamma})$

$$\begin{aligned} &\max\ -I_\gamma(f) \\ &f:\ \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l, \quad l = 1, \ldots, N \end{aligned} \tag{$\mathcal{F}^p_{\Phi,\gamma}$}$$

or equivalently

$$\begin{aligned} &\min\ I_\gamma(f) \\ &f:\ \int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l, \quad l = 1, \ldots, N. \end{aligned} \tag{$\mathcal{F}^p_{\Phi,\gamma}$}$$

We denote by ψ the convex conjugate of γ, which is defined by

$$\psi(t) = \sup_y \{ty - \gamma(y)\}.$$

The domain of γ (respectively of ψ) is denoted by $\mathcal{D}_\gamma$ (respectively $\mathcal{D}_\psi$). We will further make the following assumption denoted by $\mathcal{H}_1$.

$\mathcal{H}_1$: γ (respectively its convex conjugate ψ) is a differentiable, closed, essentially strictly convex function for all interior points on its domain $y \in \text{int } \mathcal{D}_\gamma$ (respectively in $\mathcal{D}_\gamma$).
The minimum of $\gamma(y)$ is 0 and is attained at some $y_0 = (y_0^1, \ldots, y_0^p)$ such that $y_0 \in \text{int } \mathcal{D}_\gamma$.
The convex function γ is such that the ratio $\frac{\gamma(y)}{||y||}$ tends to infinity on the edges of D_γ.

As in the case of the single function reconstruction (recalled in the Sect. 2), we will need to define the multidimensional analogue of the γ-divergence of signed measures. Such γ-divergence features some terms that depend on the singular parts. As in the case in one dimension, the singular part vanishes when $\frac{\gamma(y)}{||y||}$ tends to infinity on the edges of D_γ. The assumption on the ratio $\frac{\gamma(y)}{||y||}$ is then taken for the sake of simplicity.

Let E be a convex set of measurable $\mathbb{R}^p$-valued functions, the minimum of $I_\gamma(\cdot)$ on the convex E will be denoted as

$$I_\gamma(E) := \min_{f \in E} I_\gamma(f).$$

The first result we have is a characterisation of the minimum of $I_\gamma(\cdot)$ over E with respect to a specific convex functional. Define f_1 and f_2 two functions. Provide $\gamma(f_1)$ and $\gamma(f_2)$ are finite P_U-a.s., we define the γ-Bregman distance (see [4]) of function f_1 and f_2 on U as

$$B_\gamma(f_1, f_2) := \int_U \left[\gamma(f_1) - \gamma(f_2) - (\nabla\gamma(f_2))^T (f_1 - f_2)\right] \mathrm{d}P_U. \tag{3.3}$$

One can remark that $B_\gamma(f_1, f_2) \geqslant 0$, $\forall f_1, f_2$ with finite I_γ values by the convexity of γ. The next theorem characterises the minimum of $I_\gamma(\cdot)$ over a convex set E of functions with respect to the Bregman distance.

Theorem 3.1. *Given any convex set E of measurable functions on U, such that $I_\gamma(E)$ is finite, there exists a differentiable function f^o not necessarily in E such that for all $f \in E$ with $I_\gamma(f) < \infty$*

$$I_\gamma(f) \geqslant I_\gamma(E) + B_\gamma(f, f^o). \tag{3.4}$$

In addition, f^o is unique P_U-a.s. and any sequence of functions $f_n \in E$, for which $I_\gamma(f_n) \to I_\gamma(E)$, converges to f^o in P_U.

Proof. We adapt the proof of [9] in the multidimensional case proposed in this section. The proof relies on an identity that holds for all function $f \in E$ with finite $I_\gamma(\cdot)$ value and that a I_γ-minimising sequence of function is in some weak sense a Cauchy sequence. In the following, $||.||$ will denote the Euclidean norm on $\mathbb{R}^p$.

First notice that for all $\alpha \in]0;1[$ and all f, $f_1 \in E$ such that $I_\gamma(f)$ and $I_\gamma(f_1)$ are finite, the following equality holds

$$\begin{aligned} \alpha I_\gamma(f) + (1-\alpha) I_\gamma(f_1) = I_\gamma\left(\alpha f + (1-\alpha) f_1\right) + \alpha B_\gamma\left(f, \alpha f + (1-\alpha) f_1\right) \\ + (1-\alpha) B_\gamma\left(f_1, \alpha f + (1-\alpha) f_1\right). \end{aligned} \tag{3.5}$$

Let denote by f_2 the function $f_2 = \alpha f + (1-\alpha) f_1$. One can first notice that by developing and rearranging those terms, we have that

$$\alpha\left(\nabla\gamma(f_2)\right)^T (f - f_2) + (1-\alpha)\left(\nabla\gamma(f_2)\right)^T (f_1 - f_2) = 0.$$

Therefore, only the $I_\gamma(\cdot)$ parts remain in the sum of $\alpha B_\gamma(f, f_2)$ and $(1-\alpha) B_\gamma(f_1, f_2)$, that is

$$\alpha B_\gamma(f, f_2) + (1-\alpha) B_\gamma(f_1, f_2) = \alpha I_\gamma(f) + (1-\alpha) I_\gamma(f_1) + I_\gamma(f_2)$$

and equality (3.5) holds.

Let $(f_n) \subset E$ be a minimising sequence of I_γ and such that for all n, $I_\gamma(f_n) < \infty$. Then (f_n) is a Cauchy sequence in probability, that is

$$\lim_{n,m\to\infty} P_U\left(\{x : ||f_n(x) - f_m(x)|| > \varepsilon\}\right) = 0 \qquad \forall \varepsilon > 0.$$

See Lemma C.1 for the proof.

Then, there exists a subsequence (f_{n_k}) of (f_n) in $\mathbb{R}^p$ such that f_{n_k} converges a.s. to a function f^o in P_U measure and such f^o satisfies the inequality in (3.4). Indeed by replacing f_1 by f_{n_k} in (3.5), it becomes

$$\begin{aligned} I_\gamma(f) &= \frac{1}{\alpha} I_\gamma(\alpha f + (1-\alpha) f_{n_k}) - \frac{1-\alpha}{\alpha} I_\gamma(f_{n_k}) + B_\gamma(f, \alpha f + (1-\alpha) f_{n_k}) \\ &\quad + \frac{1-\alpha}{\alpha} B_\gamma(f_{n_k}, \alpha f + (1-\alpha) f_{n_k}) \\ &\geqslant I_\gamma(E) + B_\gamma(f, \alpha f + (1-\alpha) f_{n_k}) + \frac{1-\alpha}{\alpha}\left(I_\gamma(E) - I_\gamma(f_{n_k})\right), \end{aligned} \tag{3.6}$$

the last inequality coming from the positivity of the Bregman distance and from the fact that $\alpha f + (1-\alpha) f_{n_k} \in E$ by the convexity of E and therefore $I_\gamma(\alpha f + (1-\alpha) f_{n_k}) \geqslant I_\gamma(E)$.

As f_{n_k} is a I_γ-minimising sequence, the last term in (3.6) tends to 0 when k goes to infinity.

Finally, taking a sequence (α_k) converging slower than f_{n_k} to 0, by Fatou's lemma

$$\liminf_{k\to\infty} B_\gamma(f, \alpha_k f + (1-\alpha_k) f_{n_k}) \geqslant B_\gamma(f, f^o).$$

This proves the existence of f^o in inequality (3.4). □

Given N real-valued positive measures $\Phi_1, \ldots, \Phi_N$, their Lebesgue decompositions are given by $\Phi_l = \phi_l P_U + \Sigma_l$, for $l = 1, \ldots, N$ where their Radon–Nikodym derivatives with respects to P_U are denoted by ϕ_l. Measures Σ_l are singular with respect to P_U which means that they are concentrated on a set $\tilde{U}$ such that $P_U(\tilde{U}) = 0$. We denote $E_{\phi,\mathcal{K}}$ the subset of functions as follows

$$E_{\phi,\mathcal{K}} := \left\{ f = (f^1, \ldots, f^p), \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \phi_l(x) \mathrm{d}P_U(x) \in \mathcal{K}_l,\ l = 1, \ldots, N \right\}. \tag{3.7}$$

The next theorem leads to a description of the optimal function $(f^{1,o}, \ldots, f^{p,o})$ which is solution on $E_{\phi,\mathcal{K}}$ of $I_\gamma(\cdot)$ minimisation problem. The result is obtained thanks to Fenchel's duality theorem for convex functions and is a generalisation in higher dimension of Theorem II.2 from [10].

Theorem 3.2. *Suppose there exists a $\mathbb{R}^p$-valued function $f \in E_{\phi,\mathcal{K}}$ and such that it satisfies*

$$f(x) = (f^1, \ldots, f^p) \in \operatorname{int} \mathcal{D}_\gamma, \quad P_U\text{-a.s.}$$

Let L be the subspace of $\mathbb{R}^N$ such that for given $(\phi_1, \ldots, \phi_N)$

$$L = \left\{ v \in \mathbb{R}^N : v_l = \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \phi_l(x) \mathrm{d}P_U(x), l = 1, \ldots, N, \text{ for some } f : U \to \mathbb{R}^p \right\}$$

and m be the following application

$$m(v) := \int_U \psi\left(\tau^1(x, v), \ldots, \tau^p(x, v)\right) \mathrm{d}P_U(x), \tag{3.8}$$

with $\tau^i(x, v) = \sum_{l=1}^N \lambda^i(x) v_l \phi_l(x)$. The minimum of $I_\gamma(\cdot)$ over the set $E_{\phi,\mathcal{K}}$ can be expressed by

$$I_\gamma(E_{\phi,\mathcal{K}}) = \max_{v\in\mathbb{R}^N}\left\{\inf_{c\in\mathcal{K}\cap L}\langle v,c\rangle - m(v)\right\}. \tag{3.9}$$

Then, for $v^o\in\mathbb{R}^N$ which maximises (3.9), the minimiser of $I_\gamma(\cdot)$ over $E_{\phi,\mathcal{K}}$ is

$$f^{i,o}(x) = \frac{\partial\psi}{\partial\tau_i}\big(\tau^1(x,v^o),\dots,\tau^p(x,v^o)\big), \quad \forall i=1,\dots,p \tag{3.10}$$

and such $f^o=(f^{o,1},\dots,f^{o,p})$ satisfies (3.4).

Proof. Outline of the proof is as in [10] with a multidimensional approach:

(1) Expression of the minimum in (3.9) is given thanks to Fenchel's duality theorem.
(2) Letting $\tau^i(x,v)=\sum_{l=1}^N\lambda^i(x)v_l\phi_l(x)$ and v^o the vector of $\mathbb{R}^N$ which maximises (3.9), one must verify that $\big(\tau^1(x,v^o)\dots\tau^p(x,v^o)\big)^T$ belongs to the interior of $\mathcal{D}_\psi$.
(3) Candidate function f^o must be such that it satisfies the Bregman inequality (3.4) of the Theorem 3.1.

Let $k(z):=\inf\{I_\gamma(f):\ \int_U\ \sum_{i=1}^p\lambda^i(x)f^i(x)\phi(x)\mathrm{d}P_U(x)=z\}$. Function k is a convex function with dom $k\subset L$. Let h be such that:

$$h(z)=\begin{cases}0 & \text{if } z\in\mathcal{K}\cap L\\ +\infty & \text{else.}\end{cases} \tag{3.11}$$

Then $I_\gamma(E_{\phi,\mathcal{K}})=\inf_{z\in\mathcal{K}}k(z)=\inf_{z\in\mathbb{R}^N}\{k(z)+h(z)\}$.

In order to apply Fenchel duality theorem, we need that dom $k\cap$ dom $h\neq\emptyset$. Equivalently, that means that there exists $z_0\in$ dom k such that $z_0\in\mathcal{K}\cap L$. From Theorem 3.2 assumptions, there exists $f\in\mathcal{D}_\gamma\cap E_{\phi,\mathcal{K}}$. By applying Lemma 3.3 there exists a closed convex set $\tilde{\mathcal{D}}$ included in $\mathcal{D}_\gamma$ and a function $\tilde f$ such that it belongs to $\tilde{\mathcal{D}}\cap E_{\phi,\mathcal{K}}$. As $\tilde f\in\tilde{\mathcal{D}}$, $\gamma(\tilde f)$ is well-defined and $I_\gamma(\tilde f)<\infty$. Let $z_0\in\mathbb{R}^N$ be defined by

$$z_{0,l}=\int_U\sum_{i=1}^p\lambda^i(x)\tilde f^i(x)\phi_l(x)\mathrm{d}P_U(x),\ l=1,\dots,N. \tag{3.12}$$

By definition $z_0\in L$, as $I_\gamma(\tilde f)<\infty$, $z_0\in$ dom k and as $\tilde f\in E_{\phi,\mathcal{K}}$, $z_0\in\mathcal{K}$. Therefore dom $k\ \cap$ dom $h\neq\emptyset$ and Fenchel–Moreau duality theorem can then be applied, see [27,28].

With the superscript * denoting the convex conjugate, using Fenchel–Moreau duality theorem, we have the following equality

$$\inf_{z\in\mathbb{R}^N}\{k(z)+h(z)\}=\max_{v\in\mathbb{R}^N}\{-k^\star(v)-h^\star(-v)\} \tag{3.13}$$

and then $I_\gamma(E_{\phi,\mathcal{K}})=\max_{v\in\mathbb{R}^N}\{-k^\star(v)-h^\star(-v)\}$.

The convex conjugate of k can be expressed as follow:

$$\begin{aligned}
k^\star(v) &= \sup_{z\in\mathbb{R}^N}\{\langle v,z\rangle - k(z)\}\\
&= \sup_{z\in\mathbb{R}^N}\sup_{f\in E_{\phi,\mathcal{K}}}\left\{\langle v,z\rangle - \int_U\gamma(f)\mathrm{d}P_U\right\}\\
&= \sup\left\{\left\langle v,\int_U\sum_{i=1}^p\lambda^i(x)f^i(x)\phi(x)\mathrm{d}P_U(x)\right\rangle - \int_U\gamma(f)\mathrm{d}P_U\right\}\\
&= \sup\left\{\int_U\sum_{i=1}^p f^i(x)\left\{\sum_{l=1}^N v_l\lambda^i(x)\phi_l(x)\right\} - \gamma(f(x))\mathrm{d}P_U(x)\right\}\\
&= \int_U\psi\left(\sum_{l=1}^N v_l\lambda^1(x)\phi_l(x),\dots,\sum_{l=1}^N v_l\lambda^p(x)\phi_l(x)\right)\mathrm{d}P_U(x)\\
&= \int_U\psi\left(\tau^1(x,v),\dots,\tau^p(x,v)\right)\mathrm{d}P_U(x),
\end{aligned}$$

with $\tau^i(x,v) = \sum_{l=1}^N \lambda^i(x) v_l \phi_l(x)$.

By definition of h, its convex conjugate is:

$$h^\star(-v) = \sup_{c\in\mathcal{K}\cap L} \langle -v, c\rangle = -\inf_{c\in\mathcal{K}\cap L} \langle v, c\rangle.$$

Therefore we have equality (3.9)

$$\begin{aligned} I_\gamma(E_{\phi,\mathcal{K}}) &= \min_{f\in E_{\phi,\mathcal{K}}} \int_U \gamma(f)\mathrm{d}P_U \\ &= \max_{v\in\mathbb{R}^N} \{-k^\star(v) - h^\star(-v)\} \\ &= \max_{v\in\mathbb{R}^N} \left\{ \inf_{c\in\mathcal{K}\cap L} \langle v, c\rangle - m(v) \right\}. \end{aligned} \tag{3.14}$$

Let us denote by v^o, the vector in $\mathbb{R}^N$ which realises the maximum of $\{\inf_{c\in\mathcal{K}\cap L}\langle v,c\rangle - m(v)\}$.

We remind that

$$m(v) = \int_U \psi\left(\tau^1(x,v), \ldots, \tau^p(x,v)\right) \mathrm{d}P_U(x)$$

$$\text{with } \tau^i(x,v) = \sum_{l=1}^N \lambda^i(x) v_l \phi_l(x).$$

From the assumptions on ψ, $\tau^i(x,v^o)$ belongs to the interior of $\mathcal{D}_\psi$ and therefore one can differentiate ψ at $\tau^i(x,v^o)$.

Let us now demonstrate that $f^o = (f^{1,o}, \ldots, f^{p,o})$ with

$$f^{i,o}(x) = \frac{\partial\psi}{\partial\tau_i}\left(\tau^1(x,v^o), \ldots, \tau^p(x,v^o)\right), \quad \forall i = 1,\ldots,p,$$

satisfies the Bregman inequality in (3.4) for $E = E_{\phi,\mathcal{K}}$ and for any f such that $I_\gamma(f) < \infty$. Equivalently, it boils down to showing that for any f with $I_\gamma(f) < \infty$

$$\int_U \sum_{i=1}^p \left(\tau^i(x,v) - \frac{\partial\gamma}{\partial y_i}\left(f^o(x)\right)\right)\left(f^{o,i}(x) - f^i(x)\right)\mathrm{d}P_U(x) \geqslant 0. \tag{3.15}$$

Inequality (3.15) comes by using (3.14) and noticing that for any $\tau \in \mathcal{D}_\psi$

$$\gamma\left(\nabla\psi(\tau)\right) = \tau^T \nabla\psi(\tau) - \psi(\tau) \tag{3.16}$$

and that for any f with $I_\gamma(f) < \infty$

$$\int_U \sum_{i=1}^p \tau^i(x,v^o) f^i(x) \mathrm{d}P_U(x) \geqslant \inf_{z\in\mathcal{K}\cap L} \langle v^o, z\rangle.$$

Let us denote by $\mathcal{D}_\nabla$ the set of γ gradient

$$\mathcal{D}_\nabla = \{\nabla\gamma(y), y \in \text{int } \mathcal{D}_\gamma\}.$$

For $\tau^o = (\tau^1(x,v^o), \ldots, \tau^p(x,v^o)) \in \text{int } \mathcal{D}_\nabla$, using (3.16) we have $\nabla\gamma\left(\nabla\psi(\tau^o)\right) = \tau^o$ and then (3.15) stands with equality. When $\tau^o \in \mathbb{R}^p \setminus (\text{int } \mathcal{D}_\nabla)$, let $u \in \partial(\mathcal{D}_\gamma)$ such that $\nabla\gamma(u)$ defined below is the projection of τ^o on $\partial\mathcal{D}_\nabla$

$$\nabla\gamma(u) := \lim_{\substack{y\to u \\ y\in\text{int}\,\mathcal{D}_\gamma}} \nabla\gamma(y).$$

At such τ^o, we have $\psi(\tau^o) = u^T\tau^o - \gamma(u)$ and so $\nabla\psi(\tau^o) = u$. Therefore

$$\begin{aligned}\nabla\gamma(f^o) &= \nabla\gamma(\nabla\psi(\tau^o)) \\ &= \nabla\gamma(u)\end{aligned}$$

and

$$\begin{aligned}f^{o,i} &= (\nabla\psi(\tau^o))_i \\ &= u_i.\end{aligned}$$

If, for $i \in \{1,\dots,p\}$, we have $\tau^i(x,v^o) \leqslant (\nabla\gamma(u))_i$, then $u_i \leqslant f^i$ for any f such that $I_\gamma(f) < +\infty$. Therefore both terms of the product in (3.15) are negative. Conversely, if we have $\tau^i(x,v^o) \geqslant (\nabla\gamma(u))_i$, then $u_i \geqslant f^i$ for any f such that $I_\gamma(f) < +\infty$. Then optimal solution f^o satisfies Bregman inequality. □

To apply Fenchel duality Theorem in the previous Theorem 3.2, one needs to prove that the domains of the two convex conjugates functions k and h defined in the proof do not have an empty intersection. In order to prove so, the following lemma is required. Lemma 3.3 features a function $\tilde{f}$ which belongs to a closed convex set $\tilde{\mathcal{D}}$ in the interior of γ domain. Given a function f in the interior of $\mathcal{D}_\gamma$, we prove the existence of $\tilde{f}$ for which the constraint values $\int_U \sum_{i=1}^p \lambda^i(x)\tilde{f}^i(x)\phi_l(x)\mathrm{d}P_U(x)$ are equal to the constraints values obtained with function f.

Lemma 3.3. *Let $\phi_1,\dots,\phi_N$ be given integrable functions. Assume there exists a measurable function $f:\ U \to \mathbb{R}^p$ such that*

$$\begin{aligned}&f \in \text{int } \mathcal{D}_\gamma \\ &\sum_{i=1}^p \lambda^i(x)f^i(x)\phi_l(x) \in L^1(P_U),\ l = 1,\dots,N.\end{aligned}$$

Then, there exist $\tilde{f}$ and $\tilde{\mathcal{D}}$ such that $\tilde{f}$ is a function defined on U, $\tilde{f} \in \tilde{\mathcal{D}}$ and

$$\int_U \sum_{i=1}^p \lambda^i(x)f^i(x)\phi_l(x)\mathrm{d}P_U(x) = \int_U \sum_{i=1}^p \lambda^i(x)\tilde{f}^i(x)\phi_l(x)\mathrm{d}P_U(x),\ l = 1,\dots,N \tag{3.17}$$

with $\tilde{\mathcal{D}}$ a closed convex set of $\mathbb{R}^p$ such that $\tilde{\mathcal{D}} \subsetneq$ int $\mathcal{D}_\gamma$.

Proof. Let L be the subset of vectors of $\mathbb{R}^N$ defined by

$$L = \left\{ v:\ v_l = \int_U \sum_{i=1}^p \lambda^i(x)h^i(x)\phi_l(x)\mathrm{d}P_U(x),\ l = 1,\dots,N,\ h: U \to \mathbb{R}^p \right\}.$$

Let (D_n) be a sequence of closed convex sets such that for all n, $D_n \subsetneq$ int D_{n+1}, meaning that D_n is strictly growing to its limit D_γ. Let $T_n = \{x \in U : f(x) \in D_n\}$.

Let L_n be the subset of vector in $\mathbb{R}^N$ such that

$$L_n = \left\{ v:\ v_l = \int_U \sum_{i=1}^p \lambda^i(x)h^i(x)\phi_l(x)\mathrm{d}P_U(x),\ l = 1,\dots,N,\ h \text{ bounded on } T_n,\ h = 0 \text{ on } T_n^c \right\}$$

then there exists n_0 such that for all $n \geqslant n_0$, $L_n = L$. Indeed, if L_n is a proper subset of L, there exists v_n with $||v_n|| \neq 0$ such that

$$\begin{aligned}&\sum_{l=1}^N v_{n,l}\left(\int_U \sum_{i=1}^p \lambda^i(x)h^i(x)\phi_l(x)\mathrm{d}P_U(x)\right) = 0 \\ &\int_U \sum_{i=1}^p \lambda^i(x)h^i(x)\left(\sum_{l=1}^N v_{n,l}\phi_l(x)\right)\mathrm{d}P_U(x) = 0.\end{aligned}$$

That implies that $\sum_{l=1}^{N} v_{n,l}\phi_l(x) = 0$ P_U-a.s. on T_n. Taking a convergent subsequence v_{n_k} with limits v such that $||v|| \neq 0$, we have $\sum_{l=1}^{N} v_l\phi_l(x) = 0$ P_U-a.s. on U which goes against $||v|| \neq 0$.

For $\delta > 0$, defines C^δ by

$$C^\delta = \left\{ v : \; v_l = \int_U \sum_{i=1}^{p} \lambda^i(x) h^i(x) \phi_l(x) \mathrm{d}P_U(x), \; l = 1, \ldots, N, \; ||h|| \leqslant \delta \text{ on } T_{n_0}, \; h = 0 \text{ on } T_{n_0}^c \right\}.$$

The affine hull of C^δ equals L and $0 \in \text{int } \; C^\delta$ in the relative topology of L.

We denote by f_n the projection onto D_n of f at x. Therefore $\int_U \sum_{i=1}^{p} \lambda^i(x)(f_n^i(x) - f^i(x))\phi_l(x)\mathrm{d}P_U(x)$ approaches 0 as n tends to infinity. For $\delta > 0$ there exists then $n_1(\delta)$ such that for all $n \geqslant n_1(\delta)$, $\int_U \sum_{i=1}^{p} \lambda^i(x)(f_n^i(x) - f^i(x))\phi_l(x)\mathrm{d}P_U(x)$ belongs to C^δ. Therefore, it exists h defined on U such that $||h|| \leqslant \delta$ on T_{n_0} and $h = 0$ on $T_{n_0}^c$ and such that

$$\int_U \sum_{i=1}^{p} \lambda^i(x)(f_n^i(x) - f^i(x))\phi_l(x)\mathrm{d}P_U(x) = \int_U \sum_{i=1}^{p} \lambda^i(x) h^i(x) \phi_l(x) \mathrm{d}P_U(x).$$

For such h, we set $\tilde{f} = f_n + h$. $\tilde{f}$ belongs to $\tilde{\mathcal{D}}$ with $\tilde{\mathcal{D}} = D_n \cap D_{n_0}^\delta$ and

$$D_{n_0}^\delta = \{ f(x) : \; ||f(x) - f_{n_0}(x)|| \leqslant \delta \}.$$

□

The next proposition describes under which conditions the infimum of the Theorem 3.2 is reached for an optimal function in $E_{\phi,\mathcal{K}}$.

Proposition 3.4. *For a function m defined as in (3.8) and for v^o which maximises (3.9). If v^o is an interior point of m domain, then optimal solution f^o, with components $f^{i,o}$ as in (3.10), belongs to $E_{\phi,\mathcal{K}}$.*

Proof. Recall that

$$I_\gamma(E_{\phi,\mathcal{K}}) = \max_{v \in \mathbb{R}^N} \left(-k^\star(v) - h^\star(-v) \right)$$

with $k^\star(v) = \int_U \psi(\tau)\mathrm{d}P_U$ and $h^\star(-v) = -\inf_{z \in \mathcal{K} \cap L} \langle v, z \rangle$.

v^o belongs to the interior of dom m implies that $k^\star$ is differentiable in v^o with its gradient d with components d_l defined for all $l = 1, \ldots, N$ by

$$d_l = \int_U \sum_{i=1}^{p} \lambda^i(x) \frac{\partial \psi}{\partial \tau_i} \left(\tau^1(x, v^o), \ldots, \tau^p(x, v^o) \right) \phi_l(x) \mathrm{d}P_U(x).$$

By Corollary 23.5.3 of [27], the subgradient of $h^\star(-v)$ at $v = v^o$, denoted by $\partial h^\star$ is included in $(-\mathcal{K})$. As the relative interiors of dom $h^\star$ and dom $k^\star$ have a non-empty intersection set, Theorem 23.8 of [27] implies on the sum of convex function subgradients that $\partial g = \{d\} + \partial h^\star$. As g reaches its maximum at v^o, ∂g is a subset which contains 0, which implies that $d \in \mathcal{K}$.

Then, with $f^{i,o}(x) = \frac{\mathrm{d}\psi}{\mathrm{d}\tau_i} \left(\tau^1(x, v^o), \ldots, \tau^p(x, v^o) \right)$ for all $l = 1, \ldots, N$,

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^{i,o}(x) \phi_l(x) \mathrm{d}P_U(x) \in \mathcal{K}_l.$$

Therefore, $f^o \in E_{\phi,\mathcal{K}}$. □

The following Corollary is deduced from the Theorem 3.2 when dealing with measures Φ that are not absolutely continuous with respect to P_U.

Let us define the analogue of $E_{\phi,\mathcal{K}}$ for measures Φ that are not absolutely continuous with respect to P_U. $E_{\Phi,\mathcal{K}}$ is the set of function

$$E_{\Phi,\mathcal{K}} := \left\{ f = (f^1, \ldots, f^p) : \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l,\ l = 1, \ldots, N \right\}.$$

As a result, we see that when considering measures Φ with a singular part, the optimum function defined in the Theorem 3.2 does not meet in fact the constraints (3.2). The corollary stresses on the fact that the problem is ill-posed when dealing with measures Φ which are not absolutely continuous with respect to the reference measure P_U.

Corollary 3.5. *Suppose there exists a function $f \in E_{\Phi,\mathcal{K}}$ and such that it satisfies*

$$f(x) = (f^1(x), \ldots, f^p(x)) \in \text{int } \mathcal{D}_\gamma, \quad P_U\text{-a.s.}$$

Let L, m be as defined in the Theorem 3.2 and $\Phi = \phi P_U + \Sigma$. The minimum of $I_\gamma(\cdot)$ over $E_{\Phi,\mathcal{K}}$ can be expressed by

$$I_\gamma(E_{\Phi,\mathcal{K}}) = \max_{v \in \mathbb{R}^N} \left\{ \inf_{c \in \tilde{\mathcal{K}} \cap L} \langle v, c \rangle - m(v) \right\} \tag{3.18}$$

for some $\tilde{\mathcal{K}}$ different of $\mathcal{K}$.

Therefore the optimal solution f^o defined in (3.10) no longer meets the constraints in (3.2).

Proof. Let $k_\Phi(c) = \inf \left\{ I_\gamma(f) :\ \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \mathrm{d}\Phi(x) = c \right\}$ and $\Phi = \phi P_U + \Sigma$. Then,

$$\begin{aligned}
I_\gamma(E_{\Phi,\mathcal{K}}) &= \inf_{f \in E_{\Phi,\mathcal{K}}} \{I_\gamma(f)\} = \inf_{c \in \mathcal{K}} k_\Phi(c) \\
&= \inf \left\{ I_\gamma(f) :\ \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \phi(x) \mathrm{d}P_U(x) + \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \mathrm{d}\Sigma(x) = c,\ c \in \mathcal{K} \right\} \\
&= \inf \left\{ I_\gamma(f) :\ \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \phi(x) \mathrm{d}P_U(x) = c - \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \mathrm{d}\Sigma(x),\ c \in \mathcal{K} \right\} \\
&= \inf \left\{ I_\gamma(f) :\ \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \phi(x) \mathrm{d}P_U(x) \tilde{c},\ \tilde{c} \in \tilde{\mathcal{K}} \right\} = I_\gamma(E_{\phi,\tilde{\mathcal{K}}}),
\end{aligned}$$

where $\tilde{\mathcal{K}} = \mathcal{K} + \left\{ c_\Sigma \in \mathbb{R}^N :\ c_\Sigma = \int_U \sum_{i=1}^p \lambda^i(x) f^i(x) \mathrm{d}\Sigma(x) \right\}$.

We apply the Theorem 3.2 to get expression (3.18). □

Corollary 3.5 points out the necessity of a different approach for solving problem described in (3.2), particularly when dealing with Φ not absolutely continuous with respect to the reference measure on U, as it is the case for solving interpolation problems.

3.2. Linear transfer principle for the multidimensional case

Following equivalence of problem solutions introduced in [10], the inverse problem on functions described in (3.2) can be treated as an inverse problem on measures. Sets and generic elements will be distinguished with V and t when discussing the measure reconstruction problem. Measures considered thereafter are always finite

real-valued measures. The set of all finite measures on a set V will be denoted by $\mathcal{M}(V)$. The aim is then to reconstruct p real-valued measures $F^i \in \mathcal{M}(V)$ such that

$$\sum_{i=1}^{p} \int_V \varphi_l^i(t) \mathrm{d}F^i(t) \in \mathcal{K}_l, \qquad l = 1, \dots, N \tag{3.19}$$

with φ_l^i being given real-valued functions on V, $\mathcal{K}_l \subset \mathbb{R}$, for all $i = 1, \dots, p$ and $l = 1, \dots, N$.

Let us first denote the following assumption

$\mathcal{H}_2$: V is a compact metric space, P_V is a probability measure having full support, all φ_l^i are continuous and for each $i = 1, \dots, p$, $(\varphi_l^i)_{l=1,\dots,N}$ are linearly independent.

Given the assumption $\mathcal{H}_1$, a solution to problem (3.19) can then be chosen by taking as optimal solution $(F^{1,o}, \dots, F^{p,o})$ the p-real valued measure which minimises the γ-divergence with respect to the reference measure P_V provided that F^o meets the constraints (3.19). The opposite of the γ-entropy I_γ and the γ-divergence are linked by the following relation ([3], Thm. 2.7)

$$D_\gamma(F, P_V) = \int_V \gamma\left(\frac{\mathrm{d}F_a^1}{\mathrm{d}P_V}, \dots, \frac{\mathrm{d}F_a^p}{\mathrm{d}P_V}\right) \mathrm{d}P_V$$

where F_a^i are measures absolutely continuous with respect to P_V.

Let us denote the set of p-real valued measure with $F^1, \dots, F^p \in \mathcal{M}(V)$ meeting the constraints described in (3.19) by the set

$$S_{\varphi,\mathcal{K}} = \left\{ (F^1, \dots, F^p) : \sum_{i=1}^{p} \int_V \varphi_l^i(t) \mathrm{d}F^i(t) \in \mathcal{K}_l,\ l = 1, \dots, N \right\}. \tag{3.20}$$

The Theorem 3.6 below describes the γ-divergence minimiser $(F^{o,1}, \dots, F^{o,p})$ such that $(F^{o,1}, \dots, F^{o,p})$ belongs to $S_{\varphi,\mathcal{K}}$. The Theorem 3.6 is the analogue of the Theorem 3.2 when dealing with measure reconstruction.

Theorem 3.6. *Under assumption $\mathcal{H}_2$, suppose we have p measures absolutely continuous with respect to P_V such that $(F^1, \dots, F^p)^T \in S_{\varphi,\mathcal{K}}$ and such that they satisfy*

$$\left(\frac{\mathrm{d}F^1}{\mathrm{d}P_V}(t), \dots, \frac{\mathrm{d}F^p}{\mathrm{d}P_V}(t)\right) \in \operatorname{int} \mathcal{D}_\psi, \quad P_V\text{-a.s.}$$

Then there exist p real-valued measures $F^{1,o}, \dots, F^{p,o}$, absolutely continuous with respect to P_V and such that $(F^{1,o}, \dots, F^{p,o})$ minimises $D_\gamma(\cdot, P_V)$ over $S_{\varphi,\mathcal{K}}$. Their Radon–Nikodym derivatives $f^{i,o}$ with respect to P_V are defined by

$$f^{i,o}(t) = \frac{\partial \psi}{\partial \tau^i}\left(\tau^1(t, v^o), \dots, \tau^p(t, v^o)\right), \quad \forall i = 1, \dots, p$$

with $\tau^i(t, v) = \sum_{l=1}^N v_l \varphi_l^i(t)$ and with v^o such that it maximises

$$D_\gamma(S_{\varphi,\mathcal{K}}, P_V) := \max_{v \in \mathbb{R}^N} \left\{ \inf_{z \in \mathcal{K}} \langle v, z \rangle - \int_V \psi\left(\tau^1(t, v), \dots, \tau^p(t, v)\right) \mathrm{d}P_V(t) \right\}.$$

Proof. Direct from the Theorem 3.2. □

Having recovered the p measures $F^{i,o}$ described in Theorem 3.6, a regularized reconstruction f^o is possible *via* the linear transfer principle. As a matter of fact, one can linearly transfer the constraints (3.2) to the

constraints (3.19) by using suitable kernel K. Let $K^i(\cdot,\cdot)$ be measurable bounded real-valued functions on $U \times V$ for $i = 1, \ldots, p$ such that

$$\begin{aligned} f_K^i(x) &= \int_V K^i(x,t)\mathrm{d}F^i(t) \qquad\qquad \forall i = 1,\ldots,p \\ \varphi_l^i(t) &= \int_U \lambda^i(x)K^i(x,t)\mathrm{d}\Phi_l(x) \qquad \forall i = 1,\ldots,p\ ,\ \forall l = 1,\ldots,N. \end{aligned} \tag{3.21}$$

Then the Fubini theorem links the two sets of constraints as it follows

$$\begin{aligned} \int_U \sum_{i=1}^p \lambda^i(x) f_K^i(x)\mathrm{d}\Phi_l(x) &= \sum_{i=1}^p \int_U \lambda^i(x) \left(\int_V K^i(x,t)\mathrm{d}F^i(t) \right) \mathrm{d}\Phi_l(x) \\ &= \sum_{i=1}^p \int_V \left(\int_U \lambda^i(x)K^i(x,t)\mathrm{d}\Phi_l(x) \right) \mathrm{d}F^i(t) \\ &= \sum_{i=1}^p \int_V \varphi_l^i(t)\mathrm{d}F^i(t). \end{aligned}$$

It is then clear that if $(F^{1,o}, \ldots, F^{p,o})$ is a solution for problem $(\mathcal{M}^p_{\varphi,\gamma})$, then $f_K^o = (f_K^{1,o}, \ldots, f_K^{p,o})$ is a regularized solution to problem $(\mathcal{F}^p_{\Phi,\gamma})$. The components of f_K^o are defined by

$$f_K^{i,o}(x) = \int_V K^i(x,t)\mathrm{d}F^{i,o}(t) \qquad \forall i = 1,\ldots,p$$

with $(F^{1,o}, \ldots, F^{p,o})$ defined as in Theorem 3.6.

The advantage of the kernel transfer method is that if it occurs that some measures Φ are not absolutely continuous with respect to the reference measure – as it is the case when considering Dirac measures for example –, the linear transfer principle provides continuous functions φ for problem (3.19) by choosing kernel K to be continuous.

The choice of K is influenced by the prior knowledge on expected properties for the regularized solution f_K, see the applications in Section 4 for some examples.

3.3. The embedding into the MEM framework for the multidimensional case

In this section we study the MEM method in the multidimensional case. We first detail the construction of the random point-measure involved in the reconstruction problem.

As before in Section 2, define a sequence of discrete probability measures $(P_n)_n$ as follows

$$P_n = \frac{1}{n}\sum_{j=1}^n \delta_{t_j} \tag{3.22}$$

with $(t_j)_{j=1,\ldots,n}$ a deterministic sequence of points in V such that the weak limit of probability measures P_n is the reference probability measure P_V. For each $i = 1, \ldots, p$, a real valued random variable Y_j^i is associated to t_j. The random variable Y_j^i can be seen as a random amplitude for a signed measure F^i at location t_j. Then let F_n^i, for $i = 1, \ldots, p$ be p random measures such that $F_n^i \ll P_n$ and such that they are defined for all $t \in V$ by

$$F_n^i(t) = \frac{1}{n}\sum_{j=1}^n Y_j^i \delta_{t_j}(t).$$

For all $j = 1, \dots, n$ the p real-valued vector of random variables $\mathbf{Y_j} = (Y_j^1, \dots, Y_j^p)^T$ is sampled from a reference distribution Π and we denote by $\Pi^{\otimes n}$ the joint distribution of the n independent, identically sampled vectors $\mathbf{Y_n}$ of dimension p. Replacing F^i by the discretised measure F_n^i, the measure constraint (3.19) can be rewritten in the following way

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\varphi_l^i(t_j)Y_j^i \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N. \tag{3.23}$$

MEM method consists then in finding the optimal joint distribution Q_n^{MEM} such that it minimises the divergence from the reference distribution $\Pi^{\otimes n}$ and such that the discrete constraints (3.23) are met on average. The MEM problem can be rewritten as

$$\min_{Q \in \mathcal{Q}_{\varphi,\mathcal{K}}^n} D_{\mathrm{KL}}(Q, \Pi^{\otimes n}). \tag{3.24}$$

where $\mathcal{Q}_{\varphi,\mathcal{K}}^n$ defines the set of distributions Q which generates $n \times p$ random amplitudes Y_j^i such, under Q, the constraints are satisfied on average, that is

$$\mathcal{Q}_{\varphi,\mathcal{K}}^n = \left\{ (\mathbf{Y_1}, \dots, \mathbf{Y_n})^T \sim Q : \ \mathbb{E}_Q\left[\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\varphi_l^i(t_j)Y_j^i\right] \in \mathcal{K}_l, \quad l = 1, \dots, N \right\},$$

where $\mathbf{Y_j}$ is the jth sample $(Y_j^1, \dots, Y_j^p)$ of amplitudes for the p random measures $F_n^1, \dots, F_n^p$.

Let us denote the following assumption

$\mathcal{H}_3$: function γ considered in the γ-divergence problem $(\mathcal{F}_{\Phi,\gamma}^p)$ is the function γ_Π that is such that its conjugate function ψ_Π has its domain equal to $\bar{\mathbb{R}^p}$ and corresponds to the logarithm of the moment generating function of Π

$$\psi_\Pi(\tau_1, \dots, \tau_p) = \log \int_{\mathbb{R}^p} \exp\left(\tau^T y\right) \mathrm{d}\Pi(y). \tag{3.25}$$

Notice that the components of $\nabla\psi_\Pi$ are then

$$\frac{\partial \psi_\Pi}{\partial \tau_i}(\tau_1, \dots, \tau_p) = \int y_i \exp\left(\tau^T y - \psi_\Pi(\tau_1, \dots, \tau_p)\right) \mathrm{d}\Pi(y). \tag{3.26}$$

Provided that it exists $y_j = \left(y_j^1 \ \dots \ y_j^p\right)^T$ in the interior of Π domain for $j = 1, \dots, n$ such that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\varphi_l^i(t_j)y_j^i \in \mathcal{K}_l, \quad l = 1, \dots, N, \tag{3.27}$$

by standard theory of the ME method the minimiser Q_n^{MEM} of $K(Q, \Pi^{\otimes n})$ exists and it belongs to the exponential family through $\Pi^{\otimes n}$ spanned by the statistics $\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{p}\varphi_l^i(t_j)y_j^i$ for $l = 1, \dots, N$. Its expression is given by

$$Q_n^{\mathrm{MEM}}(y_1, \dots, y_n) = \exp\left(\sum_{j=1}^{n}\tau_j^T y_j - \psi_\Pi\left(\tau_j\right)\right)\Pi^{\otimes n}(y_1, \dots, y_n) \tag{3.28}$$

where $\tau_j = \left(\tau^1(t_j, v_n^o), \dots, \tau^p(t_j, v_n^o)\right)$ and $\tau^i(t, v) = \sum_{l=1}^{N} v_l\varphi_l^i(t)$ and v_n^o is the maximiser of the discrete dual problem

$$H_n(v) = \inf_{c \in \mathcal{K}}\langle v, c\rangle - \frac{1}{n}\sum_{j=1}^{n}\psi_\Pi\left(\tau^1(t_j, v), \dots, \tau^p(t_j, v)\right). \tag{3.29}$$

We define the vector of measures $F_n^{\mathrm{MEM}} = (F_n^{1,\mathrm{MEM}}, \dots, F_n^{p,\mathrm{MEM}})^T$ with each component defined by

$$F_n^{i,\mathrm{MEM}} = \mathbb{E}_{Q_n^{\mathrm{MEM}}}\left[F_n^i\right]. \tag{3.30}$$

The next theorem describes the convergence of $(F_n^{\mathrm{MEM}})_n$ sequence to the solution of the γ-divergence minimisation problem on signed measures $(\mathcal{M}^p_{\varphi,\gamma})$.

Theorem 3.7. *Under assumptions $\mathcal{H}_2$ and $\mathcal{H}_3$, suppose there exists p measures $(F^1,\dots,F^p)^T \in S_{\varphi,\mathcal{K}}$ such that they satisfy*

$$\left(\frac{\mathrm{d}F^1}{\mathrm{d}P_V}(t),\dots,\frac{\mathrm{d}F^p}{\mathrm{d}P_V}(t)\right) \in \operatorname{int}\,\mathcal{D}_\Pi \qquad P_V\text{-a.s.}$$

Then the sequence (F_n^{MEM}) converges weakly to $(F^{1,o},\dots,F^{p,o})$ the minimiser of $D_\gamma(S_{\varphi,\mathcal{K}},P_V)$ which Radon–Nikodym derivatives $f^{i,o}$ with respect to P_V are defined by

$$f^{i,o}(t) = \frac{\partial\psi}{\partial\tau^i}\left(\tau^1(t,v^o),\dots,\tau^p(t,v^o)\right), \quad \forall i=1,\dots,p$$

with $\tau^i(t,v)=\sum_{l=1}^N v_l\varphi_l^i(t)$ and with v^o such that it maximises

$$D_\gamma(S_{\varphi,\mathcal{K}},P_V) := \max_{v\in\mathbb{R}^N}\left\{\inf_{z\in\mathcal{K}}\langle v,z\rangle - \int_V \psi\left(\tau^1(t,v),\dots,\tau^p(t,v)\right)\mathrm{d}P_V(t)\right\}.$$

To link the problem studied previously with the constraint of function reconstruction problem $(\mathcal{F}^p_{\Phi,\gamma})$ recalled below

$$\int_U \sum_{i=1}^p \lambda^i(x)f^i(x)\mathrm{d}\Phi_l(x) \in \mathcal{K}_l \qquad 1\leqslant l\leqslant N,$$

one can consider the analogue of constraint (3.23) for the function to reconstruct.

As problem $(\mathcal{F}^p_{\Phi,\gamma})$ and problem $(\mathcal{M}^p_{\varphi,\gamma})$ can be linked by the linear transfer studied in the Section 3.2, there exists a regularized function f_K^i, associated to the reconstructed measure F^i and chosen kernel K^i, that is defined by

$$f_K^i(x) = \int_V K^i(x,t)\mathrm{d}F^i(t) \qquad \forall i=1,\dots,p. \tag{3.31}$$

Then the function of interest f_K^i can be approximated by a random function f_n^i defined for all $x\in U$ by

$$\begin{aligned} f_n^i(x) &= \int_V K^i(x,t)F_n^i(dt) \\ &= \frac{1}{n}\sum_{j=1}^n Y_j^i K^i(x,t_j). \end{aligned} \tag{3.32}$$

For Q_n^{MEM} the solution of problem (3.24), the regularized function $f_{n,K}^{\mathrm{MEM}}$ has its components $f_{n,K}^{i,\mathrm{MEM}}$ defined by

$$f_{n,K}^{i,\mathrm{MEM}}(x) = \mathbb{E}_{Q_n^{\mathrm{MEM}}}\left[f_n^i(x)\right] = \mathbb{E}_{Q_n^{\mathrm{MEM}}}\left[\frac{1}{n}\sum_{j=1}^n Y_j^i K^i(x,t_j)\right]. \tag{3.33}$$

Then if distribution Q_n^{MEM} is a solution of problem $(\mathcal{M}^{p,n}_{\varphi,\Pi})$, the regularized solution $f_{n,K}^{\mathrm{MEM}}$ defined above meets an approximation of the constraints of problem (1.3). Such approximation is given by

$$\mathbb{E}_Q\left[\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^p \varphi_l^i(t_j)Y_j^i\right] \in \mathcal{K}_l, \quad l=1,\dots,N. \tag{3.34}$$

From previous expression, it is easy to see that properties of the solution $f_{n,K}^{i,\mathrm{MEM}}$ directly depends on the random amplitudes $(Y_j^i)_{j\in\mathbb{N}}$ properties and on the kernel K^i properties.

4. Applications

4.1. Application in the case $p = 1$

We consider in this section simple examples of functions reconstruction of one or two variables. The first example considered is the reconstruction of a real-valued convex function of one variable, $f : [-1;1] \to \mathbb{R}$ which is solution of an interpolation problem. The second example considered is the reconstruction of a polynomial function of two variables, $f : [0;1] \times [0;1] \to \mathbb{R}$ which is solution of an interpolation problem.

4.1.1. Reconstruction of an univariate convex function

The first example considered is the reconstruction of a real-valued convex function of one variable, $f : U \to \mathbb{R}$ which is solution of an interpolation problem with $U = [-1;1]$. We give $N = 3$ interpolation constraints. The function has a minimal value y_0 which is reached for a value $x_0 \in]-1;1[$. The pair (x_0, y_0) consists in the first interpolation constraint. The two other points will be denoted by (x_1, y_1) and (x_2, y_2) where x_1 and x_2 belong respectively to the interval $]-1;x_0[$ and $]x_0;1[$, where the reconstructed function will be respectively decreasing and increasing. The set of interpolation values will be denoted by $\boldsymbol{z} = (y_1 - y_0, y_2 - y_0)$. For a reason explained in the following, the interpolation constraints are expressed as the increment from the minimum value.

In this example, we consider the log Laplace transform associated with the Poisson distribution

$$\psi_{\mathcal{P}(1)}(\tau) = (e^{\tau} - 1)$$

for the convex criterion to minimise. The objective function (3.29) associated with the MEM problem becomes

$$\begin{aligned} H_n(v) &= \langle v, \boldsymbol{z} \rangle - \frac{1}{n} \sum_{j=1}^{n} \psi_{\mathcal{P}(1)} \left(\langle v, \boldsymbol{\varphi}(t_j) \rangle \right) \\ &= \langle v, \boldsymbol{z} \rangle - \frac{1}{n} \sum_{j=1}^{n} \left(\exp \left(\langle v, \boldsymbol{\varphi}(t_j) \rangle \right) - 1 \right). \end{aligned} \tag{4.1}$$

The objective function (4.1) has an analytic minimum when the number of discretisation points n is equal to 1. The analytic solution is

$$v_{o,l} = \frac{1}{\varphi_l(t)} \log \left(\frac{z_l}{\varphi_l(t)} \right), \qquad l \in \{1;2\}.$$

Otherwise, one can use the polynomial approximation of the exponential function. Solving the MEM problem is then reduced to finding the root of a polynomial.

In [17] the authors proved that

$$K_m^+(x,t) = \frac{(x-t)^{m-1}}{(m-1)!} \mathbf{1}_{[x_0;x]}(t), \quad t \in [-1;1]$$

is a kernel which leads by the linear transfer to an increasing convex function with $m-1$ derivatives and which is equal to 0 in x_0. In our frame, the kernel used for the linear transfer is

$$K(x,t) = K_m^+(x,t) \mathbf{1}_{x>x_0}(x) + K_m^-(x,t) \mathbf{1}_{x<x_0}(x), \quad t \in V = [-1;1]$$

where

$$K_m^-(x,t) = \frac{(t-x)^{m-1}}{(m-1)!} \mathbf{1}_{[x;x_0]}(t).$$

This will lead to a reconstructed function which reaches its minimum value 0 in x_0, which is decreasing on the interval $[-1;x_0]$ and increasing on the interval $[x_0;1]$.

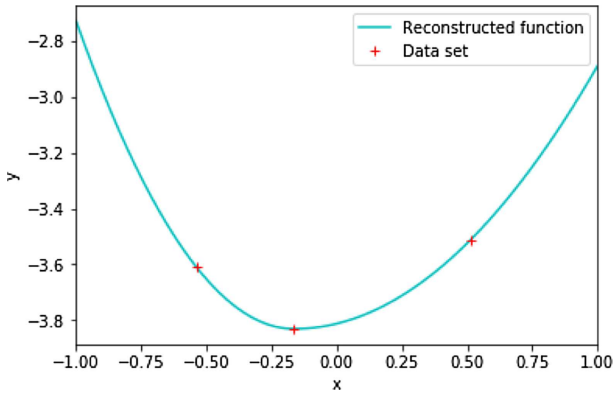

FIGURE 3. Poisson Log Laplace. Reconstruction of a convex function $f : [-1;1] \to \mathbb{R}$ twice differentiable everywhere with 3 interpolation constraints. The minimum of the function is assumed to be known.

In the end, the reconstructed function is given by

$$f_{n,K}(x) = y_0 + \frac{1}{n}\sum_{j=1}^{n} K(x,t_j)\exp\left(\sum_{l=1}^{N} v_{o,l}\varphi_l(t_j)\right).$$

The reconstructed function we obtain is displayed in Figure 3.

4.1.2. Reconstruction of a bivariate polynomial function

The second example considered is the reconstruction of a polynomial function of two variables, $f : U \to \mathbb{R}$ which is solution of an interpolation problem with $U = [0;1] \times [0;1]$. We will choose an increasing number N of interpolation points thanks to a Latin Hypercube Sampler in $[0;1] \times [0;1]$, the domain of f. We denote by $\boldsymbol{z}$ the values of function f to interpolate at the design points.

In this example, we consider the log Laplace transform associated with the standard Gaussian distribution $\mathcal{N}(0,1)$

$$\psi_{\mathcal{N}(0,1)}(\tau) = \frac{\tau^2}{2}$$

for the convex criterion to minimise. The objective function (3.29) associated with the MEM problem becomes

$$\begin{aligned} H_n(v) &= \langle v, \boldsymbol{z}\rangle - \frac{1}{n}\sum_{j=1}^{n} \psi_{\mathcal{N}(0,1)}\left(\langle v, \boldsymbol{\varphi}(t_j)\rangle\right) \\ &= \langle v, \boldsymbol{z}\rangle - \frac{1}{2n}\sum_{j=1}^{n} \left(\langle v, \boldsymbol{\varphi}(t_j)\rangle\right)^2. \end{aligned} \tag{4.2}$$

We choose $n = 100$ and the n discretising points are sampled from a Latin Hypercube Sampler in $V = [0;1] \times [0;1]$. The objective function (4.2) has an analytic solution provided that the number of discretising points n is much larger than the number of constraints N. Notice that it is also required that the N components of the moment function $\boldsymbol{\varphi}$ are linearly independent. However this condition is already an assumption in $\mathcal{H}_2$ used for the Theorem 3.6. Optimal $\boldsymbol{v}$ is solution of the linear problem

$$\mathbf{A}\boldsymbol{v} = \boldsymbol{z}$$

where matrix $\mathbf{A}$ is

$$\mathbf{A} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\varphi}(t_j)\cdot\boldsymbol{\varphi}(t_j)^T.$$

We consider the symmetric gaussian kernel

$$K_{G,\theta}(u,w) = \exp\left(-\frac{(u-w)^2}{2\theta}\right), \quad (u,w)\in[0;1]\times[0;1]$$

which is largely used as a covariance kernel in krieging problems which leads to solutions infinitely differentiable. The kernel we use for a pair of points x and t where $(x,t)\in U^2\times V^2 = [0;1]^2\times[0;1]^2$ is the following product of gaussian kernels

$$\begin{aligned} K_\theta(x,t) &= K_{G,\theta_1}(x_1,t_1)\ K_{G,\theta_2}(x_2,t_2)\\ &= \exp\left(-\frac{(x_1-t_1)^2}{2\theta_1}\right)\exp\left(-\frac{(x_2-t_2)^2}{2\theta_2}\right). \end{aligned}$$

The best parameter θ is chosen using cross-validation, namely we choose the value of θ which minimises

$$\sum_{k=1}^{N}\left(f_{K_\theta}^{(k)}(x_k)-z_k\right)^2$$

with z_k the kth value in the interpolation problem located at point x_k and $f_{K_\theta}^{(k)}$ is the reconstructed function obtained in removing (x_k,z_k) from the data set. This leads to a two step optimisation problem as matrix $\mathbf{A}$ depends on the value of θ and so does the optimal multiplier $\boldsymbol{v}$. Therefore in the following matrix $\mathbf{A}$ will be denoted with a subscript $\mathbf{A}_\theta$.

In order to determine the optimal parameter θ, we solve iteratively the following two stage problem

$$\begin{aligned} &\text{Step 1: } \boldsymbol{v}_{m+1}^{(k)}\\ &\qquad \text{solve } \mathbf{A}_{\theta_m}^{(k)}\boldsymbol{v} = \boldsymbol{z}^{(k)} \text{ for all } k=1,\dots,N.\\ &\text{Step 2: } \theta_{m+1}\\ &\qquad \text{minimise } \sum_{k=1}^{N}\left(f_{m+1,K_\theta}^{(k)}(x_k)-z_k\right)^2. \end{aligned}$$

The superscript $^{(k)}$ is used to note that the kth experiment has been removed, that is that the kth line has been removed from $\mathbf{A}_{\theta_m}$ for the matrix $\mathbf{A}_{\theta_m}^{(k)}$ and the kth value from $\boldsymbol{z}$ for the vector of observation $\boldsymbol{z}^{(k)}$. The notation $f_{m+1,K_\theta}^{(k)}$ corresponds to the reconstructed function at the mth iteration which is solution of the interpolation problem in which the kth experiment has been removed.

The reconstructed function we obtain is displayed in Figure 4.

4.2. Applications in the case $p \neq 1$

We consider thereafter some toy models inspired from computational thermodynamics. At first we explain how to compute the so-called phase diagram. Then two toy models are solved using the method described in Section 3.

4.2.1. Phase diagram description in Computational Thermodynamics

The application considered here derived from the assessment problem in Thermodynamics and the CALPHAD (CALculation of PHAse Diagram) method [16, 21, 22, 30].

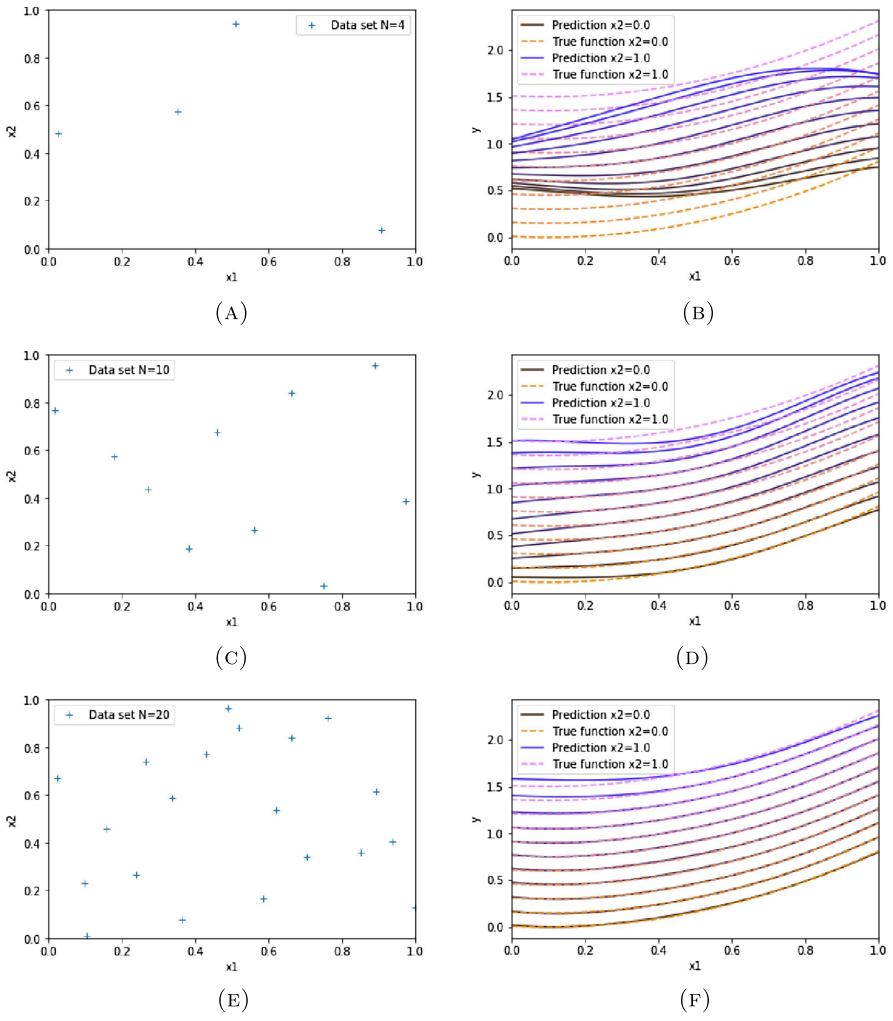

FIGURE 4. Gaussian Log Laplace. Reconstructed functions $f : [0;1] \times [0;1] \to \mathbb{R}$ with the gaussian kernel for an increasing number of interpolation constraints. Figures 4a, 4c and 4e display the design points used for the interpolation problem. Figures 4b, 4d and 4f display the evolution of the reconstructed function with respect to its first component x_1 for several values of the second component x_2. x_2 varies from 0 to 1 with a 0.1 step. (a) $N = 4$. Location of points used for training. (b) $N = 4$. Reconstructed function. (c) $N = 10$. Location of points used for training. (d) $N = 10$. Reconstructed function. (e) $N = 20$. Location of points used for training. (f) $N = 20$. Reconstructed function.

The CALPHAD method consists in the parametric reconstruction from partial information of thermodynamic quantities, which are Gibbs energy functions and their derivatives. Data at hand for the reconstruction are thermodynamic quantities and phase diagram data. The thermodynamic quantities consist in linear transformations (*e.g.* first or second derivative) of the function which is ought to be reconstructed.

The phase diagram is a map of the chemical species spatial arrangements. The internal energy of such arrangements varies with state variables, which are the chemical composition, the temperature and the pressure. Establishing the phase diagram of a chemical system means to partition the domain of permissible state variables

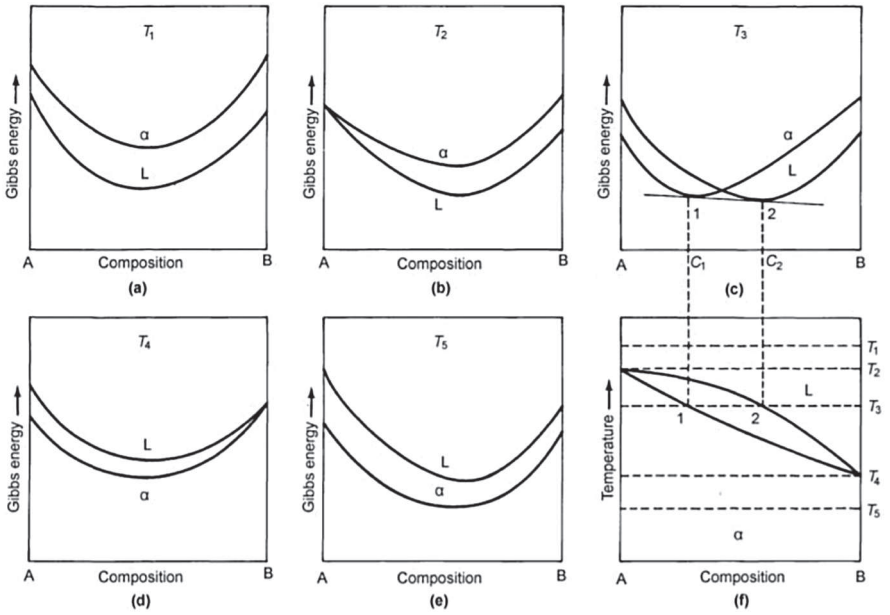

FIGURE 5. Connections between the phase diagram (f) of a system A and B with two phases (the phase α for low temperature and the phase L for high temperature) and the inner energy (a)–(e) for each phase for different temperature values. Figures (a)–(e) display the Gibbs energy for phases α and L at a given temperature with respect to the relative composition of chemical element B over the sum of chemical elements A and B. For temperature T_1 and T_2, phase L has the lowest energy whereas for temperature T_4 and T_5 phase α has the lowest energy. At temperature T_3, there exists a common tangent to the two energy functions. Therefore between composition C_1 and C_2, a combination of phases α and L is the most stable. Figure reference [25].

in several areas, each area featuring one or several stable phases. A stable phase is the phase with the lowest energy.

Figure 5 displays an example of a phase diagram. In order to determine the different divided areas of the diagram, one must compute the minimising convex hull of the list of the p functions involved in the system. This problem is performed for all temperature range. For the composition range where the minimising convex hull is confounded with a single energy function f^i, the corresponding area is the stability area of phase i. Such area is labelled by i in the phase diagram.

The phase diagram in Figure 5 is for a *binary system* A and B. Such phase diagram is called *binary phase diagram* as it features two chemical elements: A and B. In this example, the state variables which can vary are the temperature and the relative composition of B with respect to the total composition (that is of A and B), from solely element A at the left edge to solely element B at the right edge. The pressure is fixed.

The phase diagram data consist in locations in the map of stable phases, namely either temperatures or compositions where a change in the set of stable phases occurs or information on the stable phases for a given composition and a given temperature.

The novelty of the work presented here is that the reconstruction occurs in a non-parametric frame, contrarily as the usual reconstruction frame in thermodynamics.

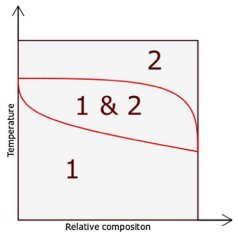

FIGURE 6. Phase diagram corresponding to an ideal solution. There are two phases: the phase 1 is the stable phase for low temperature and the phase 2 for the high temperature. Behaviour with respect to the temperature of phases 1 and 2 are known when the composition is 0 or 1 for the temperature range they are respectively the stable phase.

Thereafter, the reconstruction of thermodynamic quantities is expressed as the inverse problem $(\mathcal{F}^p_{\Phi,\gamma})$. The constraints that the solution $(f^1,\dots,f^p)$ must satisfy is recalled below

$$\int_U \sum_{i=1}^{p} \lambda^i(x) f^i(x) \mathrm{d}\Phi_l(x) \in \mathcal{K}_l \qquad 1 \leqslant l \leqslant N.$$

The functions to recover f^i, $i=1,\dots,p$ are the Gibbs energy functions of the p stable phases which occur in the phase diagram. The functions λ^i represent the phase diagram data for each phase $i=1,\dots,p$ and are supposed to be known all over the domain of permissible state variables. In the example considered thereafter, there are bounded, continuous functions but not necessary differentiable.

In this frame, x is a $(d+1)$-tuple of positive bounded quantities. First component is the temperature in Kelvin with values in $[T_0;T_{\max}]$, transposed to $[0;1]$ without loss of generality. The d following components are composition data with values in $[0;1]^d$. Kernels K^i, for all $i=1,\dots,p$, considered for the linear transfer, will be chosen as products of $d+1$ component-wise kernels. In the examples considered in the following $d=1$ or 2.

4.2.2. Phase diagram with an ideal solution

In this section we study the case of a phase diagram with an *ideal solution*. It consists in the reconstruction of two functions, *viz.* the Gibbs energy functions associated with the two phases, outside of the domain where they are known.

The inverse problem associated to the functions to reconstruct is built in accordance with the phase diagram. The phase diagram with an ideal solution is usually displayed as in Figure 6. It features two phases and three divided area. The first area at the bottom of the diagram is the area associated with phase 1 which is the stable phase at low temperature. The second area at the top of the diagram is the area associated with phase 2 which is the stable phase at high temperature. The last area between the two first areas features a mix of phase 1 and phase 2.

Such problem may not be well-posed as there can exist an infinity of pairs of functions leading to the same phase diagram. In the following the univariate case is treated, which corresponds to the case of a given temperature value. Then functions solely depend on the relative composition of element B with respect to the total composition.

In the univariate case, the pair of functions $\big(f^1(x), f^2(x)\big)$ is ought to be reconstructed for all $x \in [0;1]$. In this case, the phase diagram constraints reduce to the following definition. Given $x_1, x_2 \in [0;1]$ with $x_1 < x_2$. The interval $[x_1;x_2]$ consists in the compositions for which phase 1 and phase 2 coexist. Therefore functions λ^1 and λ^2 are defined as follows

$$\lambda^1(x) = \begin{cases} 1 & \text{if } x < x_1 \\ \dfrac{x_2 - x}{x_2 - x_1} & \text{if } x \in [x_1;x_2] \\ 0 & \text{else} \end{cases}$$

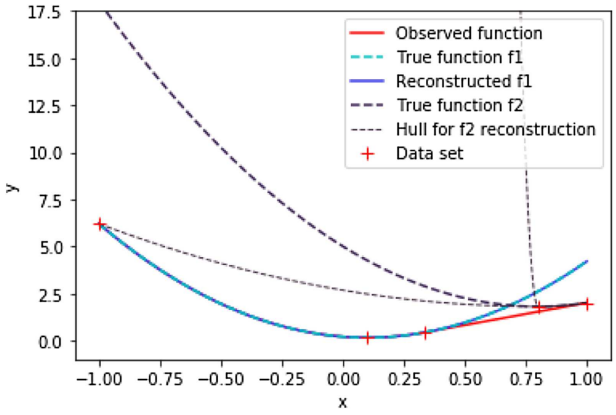

FIGURE 7. Bivariate Poisson Log Laplace. Reconstruction of two convex functions $f^1 : [-1;1] \to \mathbb{R}$ and $f^2 : [-1;1] \to \mathbb{R}$, twice differentiable everywhere. The minimum is assumed to be known for both functions.

and

$$\lambda^2(x) = \begin{cases} 0 & \text{if } x < x_1 \\ 1 - \lambda^1(x) & \text{if } x \in [x_1; x_2] \\ 1 & \text{else.} \end{cases}$$

Observations for $x < x_1$ give values of solely f^1 and for $x > x_2$ give values of solely f^2. In the interval $[x_1; x_2]$, one can observe

$$\lambda^1(x) f^1(x_1) + \lambda^2(x) f^2(x_2) \tag{4.3}$$

which does not provide direct information about the behaviour of f^1 nor f^2 on this interval.

The two functions to be reconstructed ought to be convex twice differentiable. Therefore the same assumptions are taken than in the case treated in Section 4.1.1 and the same method is applied. Figure 7 displays the results obtained for the reconstruction of the target functions.

Function f^1 is well reconstructed in this example but there exists a set of possible reconstructions for function f^2. An extra constraint is required to have a unique solution for f^2.

APPENDIX A. SOME GENERAL DEFINITIONS AND PROPERTIES

Definition A.1. General definitions.

(1) We call domain of a function γ the subset on which γ is finite, *i.e.*

$$\text{dom } \gamma = \{x : \ \gamma(x) < \infty\}.$$

(2) A function γ is said to be closed (lower semicontinuous) if for each $\alpha \in \mathbb{R}$, the sublevel set $\{x \in \text{dom } \gamma : \gamma(x) \leq \alpha\}$ is a closed set.

Definition A.2. Convex analysis definitions. See for example [27] for more details.

(1) A convex function γ from $\mathbb{R}^p$ to $\mathbb{R}$ satisfies

$$\forall x, y \in \mathbb{R}^p, \ \eta \in [0,1], \ \gamma(\eta x + (1-\eta)y) \leqslant \eta\gamma(x) + (1-\eta)\gamma(y). \tag{A.1}$$

γ is called strictly convex when equality in (A.1) occurs only when $x = y$.

(2) We call subgradient at $x \in \mathbb{R}^p$ of a convex function γ – and write $\partial\gamma_x$ – all vectors s in $\mathbb{R}^p$ which satisfy

$$\forall y \in \mathbb{R}^p,\ \gamma(y) \geqslant \gamma(x) + \langle s, y - x\rangle.$$

(3) A function γ is essentially strictly convex if γ is strictly convex on all convex subset of its subgradient domain.
(4) A function γ is essentially smooth if the interior of γ domain $\mathcal{D}_\gamma$ is not empty, γ is differentiable on the interior of $\mathcal{D}_\gamma$ and its gradient $||\nabla\gamma||$ tends to infinity when approaching to the edge of $\mathcal{D}_\gamma$.
(5) The convex conjugate ψ of a function γ is defined by

$$\psi(\tau) := \sup_{y\in\mathbb{R}^p} \{\langle y, \tau\rangle - \gamma(y)\}.$$

Proposition A.3. *Convex analysis property.*

(1) *If γ is closed convex, then its biconjugate, that is the conjugate of convex conjugate ψ, is γ itself.*
(2) *A function γ being essentially strictly convex is equivalent to its convex conjugate being essentially smooth.*
(3) *Let γ be a convex function and let its domain be not empty. If $\gamma:\ \mathbb{R}^p \to \mathbb{R}$ is such that*

$$\frac{\gamma(y)}{||y||} \underset{y\in\partial D_\gamma}{\longrightarrow} +\infty$$

then its convex conjugate has full domain, that is $\mathcal{D}_\psi = \mathbb{R}^p$.

Appendix B. Bregman distance bounds

This section generalises some results of [9] on the bounds of Bregman distance between two functions f and f_1 in the case of $\mathbb{R}^p$-valued functions.

Lemma B.1. *Let $(U, \mathcal{B}(U), P_U)$ be a probability space and E be a convex set of functions with values in $\mathbb{R}^p$. Recall the expression of $I_\gamma(f) = \int_U \gamma(f)\mathrm{d}P_U$ and one of the Bregman distance for f_1, f_2 having finite I_γ values*

$$B_\gamma(f_1, f_2) := \int_U \left[\gamma(f_1) - \gamma(f_2) - (\nabla\gamma(f_2))^T (f_1 - f_2)\right] \mathrm{d}P_U.$$

$\forall\varepsilon > 0$, $K > 0$, it exists $\iota > 0$ such that for all $f_1, f_2 \in E$ and $C \in \mathcal{B}(U)$ for which $\min_{x\in C}(||f_1(x)||, ||f_2(x)||) \leqslant K$, then

$$P_U\left(C \cap \{x : ||f_1(x) - f_2(x)|| \geqslant \varepsilon\}\right) \leqslant \iota\ B_\gamma(f_1, f_2). \tag{B.1}$$

Proof. Let f_1 and f_2 be two functions in E with finite I_γ values. Let denote by b_γ the integrand of B_γ that is

$$b_\gamma(f_1(x), f_2(x)) = \gamma(f_1(x)) - \gamma(f_2(x)) - \big(\nabla\gamma(f_2(x))\big)^T (f_1(x) - f_2(x)).$$

Let $C \in \mathcal{B}(U)$ be such that $\min_{x\in C}(||f_1(x)||, ||f_2(x)||) \leqslant K$. For $x \in C$, if $||f_1(x) - f_2(x)|| \geqslant \varepsilon$, then

$$b_\gamma(f_1(x), f_2(x)) \geqslant \min_{x\in C}\left(\min_{\text{if } ||f_1(x)||\leqslant K} b_\gamma(f_1(x), f_1(x) + \varepsilon), \min_{\text{if } ||f_2(x)||\leqslant K} b_\gamma(f_2(x) + \varepsilon, f_2(x))\right) =: \frac{1}{\iota}$$

with $\iota > 0$. Therefore, for $C \in \mathcal{B}(U)$ for which $\min_{x\in C}(||f_1(x)||, ||f_2(x)||) \leqslant K$

$$\begin{aligned}
B_\gamma(f_1, f_2) &= \int_U b_\gamma(f_1(z), f_2(z))\mathrm{d}P_U(z) \\
&\geqslant \int_{C\cap\{x:||f_1(x)-f_2(x)||\geqslant\varepsilon\}} b_\gamma(f_1(z), f_2(z))\mathrm{d}P_U(z) \\
&\geqslant \int_{C\cap\{x:||f_1(x)-f_2(x)||\geqslant\varepsilon\}} \frac{1}{\iota}\mathrm{d}P_U(z) \\
&\geqslant \frac{1}{\iota} P_U\left(C \cap \{x : ||f_1(x) - f_2(x)|| \geqslant \varepsilon\}\right)
\end{aligned}$$

which gives the result. □

Lemma B.2. *Let $(U, \mathcal{B}(U), P_U)$ be a probability space and E be a convex set of functions with values in $\mathbb{R}^p$. γ is an essentially strictly convex twice differentiable function on $\mathbb{R}^p$ and its Hessian matrix is strictly definite positive on $\mathbb{R}^p$. Recall the Bregman distance for f_1, f_2 having finite I_γ values*

$$B_\gamma(f_1, f_2) := \int_U \left[\gamma(f_1) - \gamma(f_2) - (\nabla\gamma(f_2))^T (f_1 - f_2)\right] \mathrm{d}P_U.$$

$\forall K > 0$, it exists $\beta > 0$ such that for all $f_1, f_2 \in \mathbb{R}^p$ and $C \in \mathcal{B}(U)$ with $C \subset \{x : ||f_2(x)|| \leqslant K\}$, then

$$P_U\left(C \cap \{x : ||f_1(x)|| \geqslant L\}\right) \leqslant \frac{\beta}{L^2} B_\gamma(f_1, f_2), \qquad \text{if } L \geqslant 2K. \tag{B.2}$$

Proof. Let denote by b_γ the integrand of B_γ, that is $b_\gamma(f_1, f_2) = \gamma(f_1) - \gamma(f_2) - (\nabla\gamma(f_2))^T (f_1 - f_2)$. Let $C \in \mathcal{B}(U)$ be $C \subset \{x : ||f_2(x)|| \leqslant K\}$.

Let u_K and v, two vector from $\mathbb{R}^p$ such that $||u_K|| = K$ and v belongs to the p-ball of center u_K and of radius K, denoted by $\mathcal{B}(u_K, K)$. Then, using Taylor's theorem for the decomposition of $\gamma(v)$ centered in u_K, we have

$$\gamma(v) = \gamma(u_K) + (\nabla\gamma(u_K))^T (v - u_K) + R_\gamma(v, u_K)$$

where $R_\gamma(v, u_K)$ is the remainder term of order $o(K)$.

Then for all $x \in C$

$$\begin{aligned} b_\gamma(f_1(x), f_2(x)) &\geqslant b_\gamma(2u_K, u_K) \\ &\geqslant R_\gamma(2u_K, u_K) \\ &\geqslant \frac{1}{2}||u_K||^2 \varepsilon_{\gamma,K} \\ &\geqslant \frac{1}{8} L^2 \varepsilon_{\gamma,K} \end{aligned}$$

where $\varepsilon_{\gamma,K}$ is the smallest eigen value of γ Hessian matrix located at any point of $\mathcal{B}(u_K, K)$. Given the assumptions on γ, $\varepsilon_{\gamma,K}$ is strictly positive.

Therefore, for all $x \in C$ for which $||f_2(x)|| \leqslant K$

$$\begin{aligned} B_\gamma(f_1, f_2) &= \int_U b_\gamma(f_1(z), f_2(z)) \mathrm{d}P_U(z) \\ &\geqslant \int_{C \cap \{x : ||f_1(x)|| \geqslant L\}} b_\gamma(f_1(z), f_2(z)) \mathrm{d}P_U(z) \\ &\geqslant \int_{C \cap \{x : ||f_1(x)|| \geqslant L\}} \frac{8L^2}{\varepsilon_{\gamma,K}} \mathrm{d}P_U(z) \\ &\geqslant \frac{8L^2}{\varepsilon_{\gamma,K}} P_U\left(C \cap \{x : ||f_1(x)|| \geqslant L\}\right) \\ &\geqslant \frac{L^2}{\beta} P_U\left(C \cap \{x : ||f_1(x)|| \geqslant L\}\right) \end{aligned}$$

which gives the result. □

Appendix C. Cauchy result for Bregman distance minimising sequence

Lemma C.1. *Let $(U, \mathcal{B}(U), P_U)$ be a probability space and E be a convex set of functions with values in $\mathbb{R}^p$. Recall the expression of $I_\gamma(f) = \int_U \gamma(f) \mathrm{d}P_U$ and the Bregman distance for f_1, f_2 two functions having finite I_γ values*

$$B_\gamma(f_1, f_2) := \int_U \left[\gamma(f_1) - \gamma(f_2) - (\nabla\gamma(f_2))^T (f_1 - f_2)\right] \mathrm{d}P_U.$$

Let $(f_n) \subset E$ be a I_γ-minimising sequence, then (f_n) is a Cauchy sequence in probability P_U, meaning that

$$\lim_{n,m\to\infty} P_U\left(\{x : ||f_n(x) - f_m(x)|| > \varepsilon\}\right) = 0 \qquad \forall \varepsilon > 0. \tag{C.1}$$

Proof. Let $\eta > 0$ and K such that for $f \in \mathbb{R}^p$

$$P_U(\{x : ||f(x)|| > K\}) < \eta.$$

By applying Lemma B.2 with $f_1 = \frac{f_n+f_m}{2}$, $f_2 = f$ and $C = \{x : ||f(x)|| \leqslant K\}$, we have

$$P_U\left(\{x : ||f(x)|| \leqslant K\} \cap \left\{x : \left|\left|\frac{f_n(x) + f_m(x)}{2}\right|\right| \geqslant L\right\}\right) \leqslant \frac{\beta}{L^2} B_\gamma\left(\frac{f_n + f_m}{2}, f\right),$$

if $L \geqslant 2K$. Choosing β and L such that

$$\frac{\beta}{L^2} B_\gamma\left(\frac{f_n + f_m}{2}, f\right) \leqslant \eta,$$

it follows that

$$P_U\left(\left\{x : \left|\left|\frac{f_n(x) + f_m(x)}{2}\right|\right| \geqslant L\right\}\right) \leqslant 2\eta.$$

Now applying Lemma B.1 with $f_1 = f_m$, $f_2 = \frac{f_n+f_m}{2}$ and $C = \left\{x : \left|\left|\frac{f_n(x)+f_m(x)}{2}\right|\right| < L\right\}$, there exists ι such that

$$P_U\left(\left\{x : \left|\left|\frac{f_n(x) + f_m(x)}{2}\right|\right| < L\right\} \cap \left\{x : \left|\left|f_m(x) - \frac{f_n(x) + f_m(x)}{2}\right|\right| \geqslant \varepsilon\right\}\right) \leqslant \iota\ B_\gamma\left(f_m, \frac{f_n + f_m}{2}\right)$$

meaning that

$$P_U\left(\left\{x : \left|\left|f_m(x) - \frac{f_n(x) + f_m(x)}{2}\right|\right| \geqslant \varepsilon\right\}\right) \leqslant 2\eta + \iota\ B_\gamma\left(f_m, \frac{f_n + f_m}{2}\right)$$

$$P_U\left(\{x : ||f_m(x) - f_n(x)|| \geqslant 2\varepsilon\}\right) \leqslant 2\eta + \iota\ B_\gamma\left(f_m, \frac{f_n + f_m}{2}\right)$$

By taking η as small as possible, when n and m go to infinity, it ends up that

$$\lim_{n,m\to\infty} P_U\left(\{x : ||f_m(x) - f_n(x)|| \geqslant 2\varepsilon\}\right) = \lim_{n,m\to\infty} \iota\ B_\gamma\left(f_m, \frac{f_n + f_m}{2}\right).$$

Let $(f_n) \subset E$ be a I_γ-minimising sequence with f_n having finite I_γ values. By the positivity of B_γ, we have

$$\begin{aligned} I_\gamma\left(\frac{f_n + f_m}{2}\right) &\leqslant \frac{1}{2}\left(I_\gamma(f_n) + I_\gamma(f_m)\right) \\ &\leqslant \max_k I_\gamma(f_k). \end{aligned}$$

By the convexity of E, $\frac{f_n+f_m}{2}$ belongs to E and therefore $I_\gamma\left(\frac{f_n+f_m}{2}\right) \geqslant I_\gamma(E)$.

Using the following identity when n, m tend to infinity

$$I_\gamma(f_n) + I_\gamma(f_m) = 2I_\gamma\left(\frac{f_n + f_m}{2}\right) + B_\gamma\left(f_n, \frac{f_n + f_m}{2}\right) + B_\gamma\left(f_m, \frac{f_n + f_m}{2}\right), \tag{C.2}$$

it implies that

$$\lim_{n,m\to\infty} B_\gamma\left(f_m, \frac{f_n+f_m}{2}\right) = 0$$
$$\lim_{n,m\to\infty} B_\gamma\left(f_n, \frac{f_n+f_m}{2}\right) = 0$$

which proves that (f_n) is a Cauchy sequence in probability P_U. □

Acknowledgements. Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is gratefully acknowledged. This work benefited from feedback from an anonymous reviewer.

References

[1] O. Barndorff-Nielsen, Information and Exponential Families: In Statistical Theory. John Wiley & Sons (1978).
[2] J. Borwein and A. Lewis, Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.* **29** (1991) 325–338.
[3] J. Borwein and A. Lewis, Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM J. Optim.* **3** (1993) 248–267.
[4] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **7** (1967) 200–217.
[5] M. Broniatowski and A. Keziou, Minimization of divergences on sets of signed measures. *Stud. Sci. Math. Hung.* **43** (2006) 403–442.
[6] I. Csiszár, I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* (1975) 146–158.
[7] I. Csiszár, Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* (1984) 768–793.
[8] I. Csiszár, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.* **19** (1991) 2032–2066.
[9] I. Csiszár, Generalized projections for non-negative functions. *Acta Math. Hung.* **68** (1995) 161–186.
[10] I. Csiszár, F. Gamboa and E. Gassiat, MEM pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inf. Theory* **45** (1999) 2253–2270.
[11] D. Dacunha-Castelle and F. Gamboa, Maximum d'entropie et problème des moments. *Ann. Inst. Henri Poincaré Probab. Stat.* **26** (1990) 567–596.
[12] A. Dembo and O. Zeitouni, Large deviations techniques and applications. In: Vol. 38 of *Stochastic Modelling and Applied Probability*. Corrected reprint of the second (1998) edition. Springer-Verlag, Berlin (2010).
[13] F. Gamboa, *Maximum d'entropie sur la moyenne*. Ph.D. thesis, Université d'Orsay (1989).
[14] F. Gamboa and E. Gassiat, Maximum d'entropie et probléme des moments: cas multidimensionnel. *Probab. Math. Stat.* **12** (1991) 67–83.
[15] F. Gamboa and E. Gassiat, Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Stat.* **25** (1997) 328–350.
[16] B. Jansson, *Computer operated methods for equilibrium calculations and evaluation of thermochemical model parameters*. Ph.D. thesis, Royal Institute of Technology (1984).
[17] J.H.B. Kemperman, Moment problems with convexity conditions, edited by J.S. Rustagi. In: Optimizing Methods in Statistics. Academic Press (1971).
[18] C. Leonard, Minimizers of energy functionals. *Acta Math. Hung.* **93** (2001) 281–325.
[19] C. Leonard, Minimizers of energy functional under not very integrable constraints. *J. Convex Anal.* **10** (2003) 63–88.
[20] F. Liese and I. Vajda, Convex Statistical Distances. Teubner (1987).
[21] Z.-K. Liu, First-principles calculations and CALPHAD modeling of thermodynamics. *J. Phase Equilib. Diffus.* **30** (2009) 517.
[22] H.L. Lukas, S.G. Fries and B. Sundman, Computational Thermodynamics: The CALPHAD Method. Cambridge University Press (2007) 663969016.
[23] J. Navaza, On the maximum-entropy estimate of the electron density function. *Acta Crystallogr. Sect. A* **41** (1985) 232–244.
[24] O. Nikodym, Sur une généralisation des intégrales de M. J. Radon. *Fundam. Math.* **15** (1930) 131-179.
[25] H. Okamoto, M.E. Schlesinger and E.M. Mueller, Alloy Phase Diagrams. In Vol 3 of *ASM Handbook*. ASM International (2016).
[26] J. Radon, Theorie und Anwendungen der absolut additiven Mengenfunktionen. Hölder (1913) 39514488.
[27] R.T. Rockafellar, Convex Analysis. Princeton University Press (1970).
[28] R.T. Rockafellar and R.J.-B. Wets, Variational Analysis. *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag (1998).
[29] J. Shore and R. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **26** (1980) 26–37.
[30] P.J. Spencer, A brief history of CALPHAD. *Calphad* **32** (2008) 1–8.