Explicit expressions and computational methods for the Fortet–Mourier distance of positive measures to finite weighted sums of Dirac measures

Hille, Sander C.; Theewis, Esmée S.

DOI
10.1016/j.jat.2023.105947

Publication date
2023

Document Version
Final published version

Published in
Journal of Approximation Theory

Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 294 (2023) 105947

Journal of Approximation Theory

www.elsevier.com/locate/jat

Full Length Article

Explicit expressions and computational methods for the Fortet–Mourier distance of positive measures to finite weighted sums of Dirac measures

Sander C. Hille[a,*], Esmée S. Theewis[b]

[a] Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands
[b] Delft Institute of Applied Mathematics, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

Received 4 November 2022; received in revised form 9 June 2023; accepted 11 July 2023
Available online 16 July 2023

Communicated by H. Mhaskar

Abstract

Explicit expressions and computational approaches are given for the Fortet–Mourier distance between a positively weighted sum of Dirac measures on a metric space and a positive finite Borel measure. Explicit expressions are given for the distance to a single Dirac measure. For the case of a sum of several Dirac measures one needs to resort to a computational approach. In particular, two algorithms are given to compute the Fortet–Mourier norm of a molecular measure, i.e. a finite weighted sum of Dirac measures. It is discussed how one of these can be modified to allow computation of the dual bounded Lipschitz (or Dudley) norm of such measures.

MSC: 28A33; 46E27; 90C05; 90C25

Keywords: Fortet–Mourier norm; Borel measure; Metric space; Linear and convex optimization; Fermat–Weber problem

1. Introduction

Let (S, d) be a metric space, equipped with its Borel σ-algebra $\mathcal{B}(S)$. We denote by $\mathrm{BL}(S)$ the real vector space of bounded Lipschitz functions on (S, d). The Lipschitz constant of $f \in \mathrm{BL}(S)$ is written as $|f|_L$. Following Lasota, Szarek and co-workers (e.g. [28,29]) we define the *Fortet–Mourier norm* on the finite signed Borel measures $\mathcal{M}(S)$ on S by

$$\|\mu\|_{\mathrm{FM}}^* := \sup\big\{\langle\mu, f\rangle : f \in \mathrm{BL}(S),\ \|f\|_\infty \le 1,\ |f|_L \le 1\big\}, \qquad \mu \in \mathcal{M}(S), \tag{1}$$

* Corresponding author.
E-mail addresses: shille@math.leidenuniv.nl (S.C. Hille), e.s.theewis@tudelft.nl (E.S. Theewis).

https://doi.org/10.1016/j.jat.2023.105947

where the indicated pairing is given by integration: $\langle\mu, f\rangle := \int_S f\,d\mu$. In this paper we provide explicit expressions and computational methods for Fortet–Mourier norms of the form

$$\Big\|\sum_{i=1}^{N}\alpha_i\delta_{x_i} - \mu\Big\|_{\mathrm{FM}}^*, \qquad N\in\mathbb{N},\ \alpha_i>0,\ x_i\in S \text{ and } \mu\in\mathcal{M}^+(S). \tag{2}$$

Here, δ_x denotes the Dirac (or point) measure located at $x\in S$ and $\mathcal{M}^+(S)$ is the convex cone of positive measures in $\mathcal{M}(S)$.

This norm, or the equivalent dual bounded Lipschitz norm (also called Dudley norm or flat metric — for the derived metric), is used much in the study of dynamical systems in spaces of measures. For example, one encounters these norms in the context of Markov operators and semigroups on (probability) measures [13,25,29], like those defined by Iterated Function Systems [28] or Piecewise Deterministic Markov Processes [4,13,23]. Deterministic systems in spaces of measures appeared e.g. in models for population dynamics and biological systems [1,2,10], transport equations [3,21,34] and interacting particle systems or crowd dynamics [20,35].

Many authors have used the extensively studied family of Wasserstein distances, from optimal transport theory [39,40], in settings where the dynamics conserve total mass (e.g. [5,33]), or where the associated optimal transport scheme provides useful additional information, like in image processing and analysis (see e.g. [30,37], and references found there). These metrics were originally defined for measures of positive equal mass only. Moreover, theoretically, on spaces with infinite diameter, the measures must also have finite first moment. Because of both the desire to apply these to measures with unequal mass and application in image analysis, extensions of the Wasserstein metrics have been developed and explored in such settings, see [30,34,37]. Because of Kantorovich duality and the Kantorovich–Rubinstein formula (see e.g. [40], Remark 5.16 or [8], Theorem 8.10.45, p. 235), the Wasserstein-1-type metrics are well-amenable to extension. Moreover, they have better computational properties than other metrics in the Wasserstein family [37]. The extensions provided in [37] include the family of metrics

$$W_{\mathrm{KR},\lambda_1,\lambda_2}(\mu,\nu) \;:=\; \sup\big\{\langle\mu-\nu, f\rangle : f\in C(S),\ |f|_L\le\lambda_1,\ \|f\|_\infty\le\lambda_2\big\}$$

defined in [30] of which the metric derived from the Fortet–Mourier norm is a member.

The 'classical' Wasserstein-1-metric, $W_1(\mu,\nu) = W_{\mathrm{KR},1,\infty}(\mu,\nu)$, has the property that the embedding $x\mapsto\delta_x$ is isometric, i.e. $W_1(\delta_x,\delta_y) = d(x,y)$. The Fortet–Mourier distance has a typical saturation if $d(x,y)\ge 2$, see Eq. (5). This may limit its applicability in the context of image processing and analysis (though, see [30] for application of the related metrics $W_{\mathrm{KR},\lambda_1,\lambda_2}$ in this setting). For theoretical purposes however, the Fortet–Mourier norm is appreciated for its ease of use in proofs (see e.g. [41], Chapter 4, p. 101). This is illustrated too by the results presented in this paper. Similar results for the equivalent Dudley norm

$$\|\mu\|_{\mathrm{BL}}^* \;:=\; \sup\big\{\langle\mu, f\rangle : f\in \mathrm{BL}(S),\ \|f\|_\infty + |f|_L\le 1\big\} \tag{3}$$

cannot be easily obtained. See [24] for first steps in this direction. Some researchers mix the use of the Wasserstein-1 and Dudley metrics, see [5]. In applications with parameter estimation, where differentiability in parameter is needed, a metric may be needed that is defined by duality with a Hölder class of functions [21].

More specifically – in settings with varying total mass – norms of the form (2) or (3) are of interest for several reasons. In a measure framework, continuum models and discrete interacting particle descriptions can be framed within one functional analytic setting. A weighted sum of Dirac measures then represents the particle model, while a measure μ that is absolutely continuous with respect to Lebesgue measure represents the other. Norm (2) then quantifies the deviation between these two descriptions. For example, in so-called Patlak–Keller–Segel type chemotactic models it has been shown that the continuum solution converges to sums of Dirac measures in finite time [16,22], yielding blow-up in the used L^p-norm ($p > 1$). Expressions like (2) may trace such 'concentration of mass' and express a rate of convergence.

In numerical analysis of particular continuum models it may be advantageous to simulate a well-chosen interacting particle system instead of simulating the partial differential equations. See e.g. [12] where this was advocated for the two-dimensional Navier–Stokes problem, to limit numerical diffusion. Estimates of norms of the form (2) then appear naturally in error estimates.

Within the single setting of a particle model, a question is to quantify deviation between two instances of the model, with different particle number. Then, μ is also a weighted sum of Dirac measures. Expression (2) then reduces to computing norms of the form

$$\|\tau\|_{\mathrm{FM}}^*, \qquad \text{with } \tau \in \mathcal{M}ol(S) := \operatorname{span}_{\mathbb{R}}\{\delta_x : x \in S\}, \tag{4}$$

the subspace of so-called *molecular measures* [32].

There exist a few results that provide exact algorithms to compute norms of molecular measures. Jabłoński and Marciniak-Czochra [26] provided an algorithm to compute $\|\tau\|_{\mathrm{FM}}^*$ with $\tau \in \mathcal{M}ol(S)$ and $S = \mathbb{R}$ with the Euclidean metric or a bounded closed interval therein (see also [19] Appendix, for a description and application of their algorithm). Their approach depends heavily on the total ordering that is available on $\mathbb{R}$. Generalization of this approach to higher dimension or to any Polish space is therefore inhibited. Sriperumbudur et al. [38] provided an algorithm for computing the (equivalent) Dudley norm of a difference of two empirical measures. That is, $\tau = \nu - \mu$ with $\nu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and a similar expression for μ, possibly with a different number of point measures (see [38] Theorem 2.3, p. 1557). The state space S can be any metric space.

Up till now, to our knowledge, neither for a specific choice of S, nor in the generality of an arbitrary metric space (S, d), there are hardly any explicit expressions for (2), except for the well-known

$$\left\|\delta_x - \delta_y\right\|_{\mathrm{FM}}^* = 2 \wedge d(x, y), \qquad x, y \in S \tag{5}$$

(see e.g. [25,32]). Our main results, Theorems 2.1 and 3.1, will allow to extend this e.g. to the expression

$$\left\|\delta_x - \mu\right\|_{\mathrm{FM}}^* = \langle \mu, 2 \wedge d(x, \cdot)\rangle, \qquad x \in S,\ \mu \in \mathcal{P}(S), \tag{6}$$

where $\mathcal{P}(S)$ is the subset of probability measures in $\mathcal{M}^+(S)$ (see Proposition 3.1 and various corollaries of Theorem 3.1 in Section 3), or the expression

$$\left\|\alpha\delta_x - \beta\delta_y\right\|_{\mathrm{FM}}^* = |\alpha - \beta| \ + \ (\alpha \wedge \beta)(2 \wedge d(x, y)), \qquad \alpha, \beta > 0,\ x, y \in S. \tag{7}$$

(see Corollary 3.4). Such explicit expressions may be useful in obtaining (better) estimates of Fortet–Mourier norms of the indicated form. Moreover, expression (6) enables the explicit computation of e.g. the Fortet–Mourier distance of a Dirac measure to a measure that is absolutely continuous with respect to Lebesgue measure, which is a novel result.

Another motivation for determining expressions for norms like (2) comes from approximation theory. The mathematical question in which these explicit formulae may be of help is in that of existence and computation of *best approximation* of $\mu \in \mathcal{M}^+(S)$ by a positive sum of at most N Dirac measures in Fortet–Mourier distance, where N is fixed *a priori*. This is e.g. relevant for an interacting particle approach to solving a continuum model. The continuum initial condition must then be replaced by a number of particles. How can these be 'best' distributed over space, such that the error caused by the approximation of the initial condition is minimal? Is there such a best approximation? Can it be found computationally? Moreover, this question is also relevant for efficiently computing norms $\|\mu\|_{\mathrm{FM}}^*$ for general finite signed measures μ, see Remark 5.1 in Section 5.

General results for the existence of a best approximation have been known for a long time, e.g. for reflexive Banach spaces and closed convex sets therein, see e.g. [15]. Although the indicated set of sums of N Dirac measures is closed, it is not convex. Moreover, the completion of the space $\mathcal{M}(S)$ for the $\|\cdot\|_{\mathrm{FM}}^*$-norm is hardly ever reflexive. (In fact, it is isometrically isomorphic to $\mathrm{BL}(S)$ with norm $\|\cdot\|_{\mathrm{FM}} := \max(\|f\|_\infty, |f|_L)$, which can be proven in similar way as [25], Theorem 3.7, p.360). Nevertheless, a best approximation can be shown to exist on compact and complete metric spaces, essentially by exploiting the compactness of the space of probability measures that is provided by Prokhorov's Theorem.

On non-compact spaces, the situation is much more delicate, as can be illustrated by the following particular case. Expression (6) allows to reformulate the special case $N = 1$ and μ a *convex* sum of finitely many Dirac measures located at x_j to the problem of minimizing over $x \in S$ the expression

$$\Big\|\delta_x - \sum_{j=1}^{n} \alpha_j \delta_{x_j}\Big\|_{\mathrm{FM}}^* = \sum_{j=1}^{n} \alpha_j\,(2 \wedge d(x, x_j)), \qquad (\alpha_j > 0,\ \textstyle\sum_j \alpha_j = 1). \tag{8}$$

This minimization problem is the Fermat–Weber problem for the metric $d' := 2 \wedge d$ on S, with weights α_j. In economic terms, a solution to Weber's problem provides an optimal location x for a production site such that products produced there, can be distributed to the distribution sites at x_j with minimal cost, when the α_j represent the transport cost per item per unit distance [17]. Fermat's original problem was to construct a point x for which the sum of the distances to three given points is minimal. For $n > 3$ and equal weights, in Euclidean space, there does not exist a geometric construction for the best point. Existence of a minimizer to the general Fermat–Weber problem is guaranteed for so-called Hadamard spaces (also called complete CAT(0) spaces) by [6], Lemma 2.2.19. Various numerical schemes have been developed to determine a minimizer, e.g. [17,27]. Research on the Fermat–Weber problem continues to this date [11].

In Section 2 and Section 3 we present our main results on explicit expressions, Theorem 2.1 and Theorem 3.1, and various consequences derived from these. We shall present an algorithmic approach to computing Fortet–Mourier norms of the form (2) with ν a positive molecular measure in Section 4. Section 5 is concerned with algorithms for computing $\|\tau\|_{\mathrm{FM}}^*$ for any $\tau \in \mathcal{M}ol(S)$. In both sections, S is assumed to be a metric space, without additional constraints, like separability or completeness. This substantially generalizes [26]. Section 6 discusses how the results of Section 5 can be modified to compute the Dudley norm of any molecular measure. This generalizes the result in [38] on this topic.

1.1. Preliminary results and notation

For a metric space (S, d) we let $\mathrm{BL}(S)$ denote the ordered vector space of real-valued bounded Lipschitz functions on S, with point-wise partial order. We suppress the metric in notation, because there will be no need to consider multiple metrics on the same space. For $f \in \mathrm{BL}(S)$,

$$|f|_L := \sup\Big\{\frac{|f(x) - f(y)|}{d(x, y)} : x, y \in S,\ x \neq y\Big\}$$

denotes the Lipschitz constant of f. Occasionally, we shall write $|f|_{L,S}$ if we wish to stress the underlying metric space. $\mathrm{BL}(S)$ is an algebra for point-wise multiplication, with unit $\mathbb{1}$ — the constant function that is 1 on S. It is also an ordered vector space for point-wise ordering, even a vector lattice (Riesz space) with supremum $f \vee g$ and infimum $f \wedge g$ of two elements $f, g \in \mathrm{BL}(S)$ given by point-wise maximum and minimum:

$$(f \vee g)(x) = \max\big(f(x), g(x)\big), \qquad (f \wedge g)(x) = \min\big(f(x), g(x)\big), \qquad x \in S.$$

One has

$$|f \vee g|_L \leq \max(|f|_L, |g|_L) \quad \text{and} \quad |f \wedge g|_L \leq \max(|f|_L, |g|_L),$$

see e.g. [18]. We consider the norm on $\|f\|_{\mathrm{FM}} := \max\big(\|f\|_\infty, |f|_L\big)$ on $\mathrm{BL}(S)$, which turns $\mathrm{BL}(S)$ into a Banach space. The unit ball in $\mathrm{BL}(S)$ for this norm is denoted by B^S_{FM}. Accordingly, for a subset P of S with the metric induced by that on S, B^P_{FM} will denote the $\|\cdot\|_{\mathrm{FM}}$-unit ball in $\mathrm{BL}(P)$.

The following lemma yields a result on a recurring construction related to B^S_{FM} that will appear in several proofs later.

Lemma 1.1. *Let (S, d) be a metric space. The following statements hold:*

(i) Let $g \in B^S_{\mathrm{FM}}$, $N \in \mathbb{N}$ and let $x_i \in S$ $(i = 1, \ldots, N)$. Define

$$h := (-\mathbb{1}) \vee \bigvee_{i=1}^{N} (g(x_i) - d(x_i, \cdot)). \tag{9}$$

Then $h \in B^S_{\mathrm{FM}}$, $h \leq g$ and $h(x_i) = g(x_i)$ for all $i \in \{1, \ldots, N\}$.

(ii) Let $P = \{x_1, \ldots, x_N\}$ be a subset of S of distinct points, equipped with the restriction of the metric on S. If $g \in B^P_{\mathrm{FM}}$, then h defined by (9) is in B^S_{FM} and satisfies $h(x_i) = g(x_i)$ for all i. Moreover, $|h|_{L,S} \leq |g|_{L,P}$.

Proof. (i). The functions, $x \mapsto g(x_i) - d(x_i, x)$ are Lipschitz on S with Lipschitz constant at most 1 and bounded from above by 1, since $\|g\|_\infty \leq 1$. Hence $-\mathbb{1} \leq h \leq \mathbb{1}$ and $|h|_L \leq 1$. Thus, $h \in B^S_{\mathrm{FM}}$. First we show that $h \leq g$. Take $x \in S$. Then either $h(x) = -1$ or $h(x) = g(x_i) - d(x_i, x)$ for some $i \in \{1, \ldots, N\}$. In the first case, one trivially has $h(x) = -1 \leq -\|g\|_\infty \leq g(x)$. In the other case, one has $g(x_i) - g(x) \leq |g|_L d(x_i, x) \leq d(x_i, x)$. So $h(x) = g(x_i) - d(x_i, x) \leq g(x)$. Next, by construction of h it holds for every $i \in \{1, \ldots, N\}$ that $h(x_i) \geq g(x_i) - d(x_i, x_i) = g(x_i)$. Thus $h(x_i) = g(x_i)$.

(ii). Follows from part *(i)* and the McShane Extension Theorem, [31] Theorem 1. □

The following lemma gives two elementary properties of the maximum operator, which will be useful in Section 4.

Lemma 1.2. *Let $n \in \mathbb{N}$ and $a_i, b_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Then*

(i) $\max_i(a_i + b_i) \le \max_i a_i + \max_i b_i$,
(ii) $\big|\max_i a_i - \max_i b_i\big| \le \max_i |a_i - b_i|$.

Proof. Part *(i)* is obvious. For part *(ii)*, without loss of generality, assume that $\max_i a_i \ge \max_i b_i$. let $j \in \{1, \ldots, n\}$ be such that $a_j = \max_i a_i$. Then

$$\big|\max_i a_i - \max_i b_i\big| \;=\; a_j - \max_i b_i \;\le\; a_j - b_j \;\le\; |a_j - b_j| \;\le\; \max_i |a_i - b_i|. \quad \square$$

For a metric space (S, d), $\mathcal{M}(S)$ embeds naturally into the dual space $\mathrm{BL}(S)^*$ of continuous linear functionals on $\mathrm{BL}(S)$ by means of the 'integration functional with respect to μ', I_μ:

$$I_\mu(f) := \int_S f\, d\mu =: \langle \mu, f \rangle, \qquad \mu \in \mathcal{M}(S),\;\; f \in \mathrm{BL}(S).$$

The map $\mu \mapsto I_\mu$ is injective, because the indicator function of any closed set $C \subset S$ can be approximated point-wise by a decreasing sequence of functions in $\mathrm{BL}(S)$, namely $f_n := [1 - nd(\cdot, C)]^+$, and any $\mu \in \mathcal{M}(S)$ is regular, because S is a metric space (cf [8]. Theorem 7.17). By means of the mentioned embedding, the norm $\|\cdot\|_{\mathrm{FM}}$ introduces a norm on $\mathcal{M}(S)$ through the dual space $\mathrm{BL}(S)^*$, which is precisely given by (1) and which can be found in part of the literature under the name 'Fortet–Mourier norm'.

2. Dimensional reduction for determining the defining supremum

Let (S, d) be a metric space. Neither completeness, nor separability is required. The defining expression for the Fortet–Mourier norm (1) cannot be conveniently used for the computation of this norm in practice. The main issue is, that there is no convenient method (yet) to determine the supremum over the full unit ball B^S_{FM}.

The following key result allows to substantially reduce the dimension of the set over which to take the supremum, provided one of the two measures is molecular.

Theorem 2.1. *Let $\nu \in \mathcal{M}ol^+(S)$ with $P := \mathrm{supp}(\nu) = \{x_1, \ldots, x_N\}$, with the x_i all distinct. Let $\mu \in \mathcal{M}^+(S)$. Then*

$$\|\nu - \mu\|^*_{\mathrm{FM}} \;=\; \sup_{f \in B^P_{\mathrm{FM}}} \Big\langle \nu - \mu, (-\mathbb{1}) \vee \bigvee_{i=1}^N \big(f(x_i) - d(x_i, \cdot)\big)\Big\rangle \tag{10}$$

$$\;=\; \sup_{\theta \in [-1,1]^N} \Big\langle \nu - \mu, (-\mathbb{1}) \vee \bigvee_{i=1}^N \big(\theta_i - d(x_i, \cdot)\big)\Big\rangle. \tag{11}$$

Proof. First of all, $\|\nu - \mu\|^*_{\mathrm{FM}} = \sup_{g \in B^S_{\mathrm{FM}}} \langle \nu - \mu, g\rangle$. Moreover, $\nu = \sum_{i=1}^N \alpha_i \delta_{x_i}$, with $\alpha_i > 0$. Let $g \in B^S_{\mathrm{FM}}$. Note that its restriction to P, $g|_P$, is in B^P_{FM} and $g(x_i) = g|_P(x_i)$ for all $i \in \{1, \ldots, N\}$. Define

$$h := (-\mathbb{1}) \vee \bigvee_{i=1}^N \big(g(x_i) - d(x_i, \cdot)\big). \tag{12}$$

According to Lemma 1.1, $h \in B^S_{\mathrm{FM}}$, $h \le g$ and $h(x_i) = g(x_i)$ for all i. Therefore,

$$\langle \nu - \mu, h\rangle \;=\; \sum_{i=1}^{N} \alpha_i g(x_i) - \langle \mu, h\rangle \;\ge\; \sum_{i=1}^{N} \alpha_i g(x_i) - \langle \mu, g\rangle \;=\; \langle \nu - \mu, g\rangle. \tag{13}$$

This proves inequality '$\le$' in (10). The other inequality in this equation is an immediate consequence of the observation that for any $f \in B^P_{\mathrm{FM}}$ the function $(-\mathbb{1}) \vee \bigvee_{i=1}^{N}(f(x_i) - d(x_i, \cdot))$ is in B^S_{FM} (using Lemma 1.1 *(ii)*).

For proving equality (11), let $\theta \in [-1, 1]^N$ and define

$$h_\theta(x) := (-\mathbb{1}) \vee \bigvee_{i=1}^{N}(\theta_i - d(x_i, x)), \qquad x \in S. \tag{14}$$

By definition, $-1 \le h_\theta \le 1$. Moreover, $|h_\theta|_L \le \max_{1\le i\le N}\big(|-\mathbb{1}|_L, |\theta_i - d(x_i, \cdot)|_L\big) = 1$. Thus, $h_\theta \in B^S_{\mathrm{FM}}$. So '$\ge$' holds in (11). For '$\le$': let $f \in B^P_{\mathrm{FM}}$ and define

$$g(x) := (-\mathbb{1}) \vee \bigvee_{i=1}^{N}(f(x_i) - d(x_i, x))$$

According to Lemma 1.1 *(ii)*, $g \in B^S_{\mathrm{FM}}$ and $g(x_i) = f(x_i)$ for all i. Thus, if $\theta_i := f(x_i)$, $i = 1, \dots, N$, then $\theta_i \in [-1, 1]$ and $h_\theta = g$. So, the supremum in (11) is over a larger set than that in (10). Hence, '$\le$' holds. □

The difference between the similarly looking functions in (10) and (11) is, that $h_\theta(x_i) \ge \theta_i$ for all i in (11), but equality need not hold, while for $g \in B^P_{\mathrm{FM}}$, the function h defined by (12), which is used in (10), does satisfy $h(x_i) = g(x_i)$ for all i.

If we restrict our attention to ν and μ being probability measures, the dimension of the set over which one takes the supremum in (11) can be further reduced by one, as the following result ascertains.

Proposition 2.1. *Let $\nu \in \mathcal{M}ol^+(S) \cap \mathcal{P}(S)$, with* $\mathrm{supp}(\nu) = \{x_1, \dots, x_N\}$, *all $x_i \in S$ distinct, and let $\mu \in \mathcal{P}(S)$. Then*

$$\|\nu - \mu\|^*_{\mathrm{FM}} = \sup\left\{ \Big\langle \nu - \mu, (-\mathbb{1}) \vee \bigvee_{i=1}^{N}(\theta_i - d(x_i, \cdot))\Big\rangle :\; \theta_1, \dots, \theta_N \in [-1, 1],\; \max_i(\theta_i) = 1 \right\}. \tag{15}$$

Proof. The inequality '$\ge$' follows immediately from (11), since the supremum in (15) is taken over a subset of that in (11). For the other inequality, let $g \in B^S_{\mathrm{FM}}$. Put $\varepsilon := \min_{1\le i\le N}(1 - g(x_i)) \ge 0$ and $\theta_i := g(x_i) + \varepsilon$ for $i = 1, \dots, N$. Note that $\theta_i \in [-1, 1]$. Let $k \in \{1, \dots, N\}$ be such that $\varepsilon = 1 - g(x_k)$. Then $\theta_k = 1$, so $\max_i(\theta_i) = 1$. With $\theta = (\theta_1, \dots, \theta_N)$ and h_θ as defined in (14),

$$h_\theta(x_i) \ge \theta_i - d(x_i, x_i) = g(x_i) + \varepsilon. \tag{16}$$

We claim that $h_\theta \le g + \varepsilon$. Indeed, if $x \in S$ is such that $h_\theta(x) = -1$, then the inequality holds trivially, since $g \ge -1$ and $\varepsilon \ge 0$. If $h_\theta(x) > -1$, then for some $j \in \{1, \dots, N\}$,

$$\begin{aligned} h_\theta(x) \;&=\; \theta_j - d(x_j, x) \;=\; g(x_j) + \varepsilon - d(x_j, x) \;=\; g(x) + \varepsilon - d(x_j, x) + g(x_j) - g(x) \\ &\le\; g(x) + \varepsilon - d(x_j, x) + d(x_j, x) \;=\; g(x) + \varepsilon, \end{aligned} \tag{17}$$

because $g \in B^S_{\mathrm{FM}}$ and consequently, $|g|_L \le 1$. Write $\nu = \sum_{i=1}^N \alpha_i \delta_{x_i}$, with $\alpha_i > 0$, $\sum_{i=1}^N \alpha_i = 1$. The two bounds (16) and (17) yield

$$\begin{aligned}\langle \nu - \mu, h_\theta \rangle &= \sum_{i=1}^N \alpha_i h(x_i) - \int_S h_\theta \, d\mu \ \ge\ \sum_{i=1}^N \alpha_i \big(g(x_i) + \varepsilon\big) - \int_S (g + \varepsilon)\, d\mu \\ &= \sum_{i=1}^N \alpha_i g(x_i) + \varepsilon \sum_{i=1}^N \alpha_i - \int_S g \, d\mu - \varepsilon \mu(S) \ =\ \langle \nu - \mu, g \rangle.\end{aligned}$$

Since $\|\nu - \mu\|^*_{\mathrm{FM}} = \sup_{g \in B^S_{\mathrm{FM}}} \langle \nu - \mu, g \rangle$, we obtain inequality '$\le$' in (15). □

These results give rise to novel explicit expressions for the Fortet–Mourier distance to a single point mass.

3. Explicit expressions for the distance to a single point mass

Theorem 2.1 and Proposition 2.1 reduce the supremum expression for the distance to a positive molecular measure to a maximization problem of a suitable continuous function over a particular compact set. In this section we show, that if ν is a single (weighted) Dirac measure, a location where this maximum is attained can be explicitly determined. Moreover, it will become clear, that this location need not be unique. It results into various novel explicit expressions for norms of the form $\|\alpha \delta_x - \mu\|^*_{\mathrm{FM}}$, for $x \in S$, $\alpha > 0$ and $\mu \in \mathcal{M}^+(S)$.

Proposition 3.1. *Let $x \in S$ and $\mu \in \mathcal{P}(S)$. Then*

$$\|\delta_x - \mu\|^*_{\mathrm{FM}} = \big\langle \delta_x - \mu, (-\mathbb{1}) \vee (1 - d(x,\cdot)) \big\rangle = \big\langle \mu, 2 \wedge d(x,\cdot) \big\rangle.$$

Proof. Specifying the result of Proposition 2.1 to the case $N = 1$ yields the first equality. Then observe that

$$(-\mathbb{1}) \vee (1 - d(x,\cdot)) = -\big(\mathbb{1} \wedge (d(x,\cdot) - 1)\big) = -\big(2 \wedge d(x,\cdot)\big) + \mathbb{1}. \tag{18}$$

Applying the measure $\delta_x - \mu$ to the latter function gives the result, since $\mu \in \mathcal{P}(S)$. □

Example 3.1. (1) Applying Proposition 3.1 in the special case $\mu = \delta_y$ yields the well-known expression $\|\delta_x - \delta_y\|^*_{\mathrm{FM}} = 2 \wedge d(x, y)$, shown in (5).

(2) If $\mu = \alpha \delta_{y_1} + (1-\alpha)\delta_{y_2}$ with $y_1, y_2 \in S$ and $0 \le \alpha \le 1$, then Proposition 3.1 gives the explicit expression

$$\big\|\delta_x - \alpha\delta_{y_1} - (1-\alpha)\delta_{y_2}\big\|^*_{\mathrm{FM}} = \alpha(2 \wedge d(x, y_1)) + (1-\alpha)(2 \wedge d(x, y_2)). \tag{19}$$

(3) Let $S = [0, 1]$, equipped with the Euclidean metric and let λ be the Borel–Lebesgue measure on S, $\lambda([0, 1]) = 1$. Then for any $x \in S$,

$$\big\|\delta_x - \lambda\big\|^*_{\mathrm{FM}} = \int_0^1 2 \wedge |x - y| \, dy = \int_0^x (x-y)dy + \int_x^1 (y-x)dy = \tfrac{1}{2} - x + x^2. \tag{20}$$

Notice that expression (20) is minimal for $x = 1/2$ with value $1/4$. Thus, there exists a (unique) best approximation in $\mathcal{P}(S)$ of λ by a single Dirac measure in Fortet–Mourier:

$$\inf_{x\in S} \big\|\delta_x - \lambda\big\|^*_{\mathrm{FM}} = \big\|\delta_{1/2} - \lambda\big\|^*_{\mathrm{FM}} = \frac{1}{4}. \tag{21}$$

The location of best approximation is in this case the median of the uniform distribution on $[0, 1]$, which is given by λ.

(4) Let S be as in part (3) and $f \geq 0$ a probability distribution function. Then

$$\left\|\delta_x - f\,d\lambda\right\|^*_{\mathrm{FM}} = \int_0^1 |x-y| f(y)dy = \int_0^x (x-y)f(y)dy - \int_x^1 (y-x)f(y)dy.$$

For specific f the latter expression can be computed, in principle.

The following result is an immediate corollary of Theorem 2.1. It should be compared with Proposition 3.1 for the case $\mu \in \mathcal{P}(S)$:

Corollary 3.1. *Let $x \in S$ and $\mu \in \mathcal{M}^+(S)$. Then*

$$\left\|\delta_x - \mu\right\|^*_{\mathrm{FM}} = \sup_{\theta\in[-1,1]} \left\langle \delta_x - \mu, (-\mathbb{1}) \vee (\theta - d(x,\cdot))\right\rangle.$$

Proposition 3.1 states that for $\mu \in \mathcal{P}(S)$ the supremum above is attained at the value $\theta = 1$. Such a stronger result can be obtained for general positive measures too.

Let $B(x,r) := \{y \in S : d(x,y) < r\}$ be the open ball in (S,d) of radius r, centred at x. In the following result the function $(x,r) \mapsto \mu(B(x,r))$ plays a key role. It 'measures' in a way the mass distribution of μ over space.

Theorem 3.1. *Let $x \in S$ and $\mu \in \mathcal{M}^+(S)$. Then*

$$\left\|\delta_x - \mu\right\|^*_{\mathrm{FM}} = \left\langle \delta_x - \mu, (-\mathbb{1}) \vee (\theta_0 - d(x,\cdot))\right\rangle,$$

where $\theta_0 = \theta_0(x) := \big(2 \wedge \inf\{r \geq 0 : \mu\big(B(x,r)\big) \geq 1\}\big) - 1 \in [-1, 1]$.

(We use the convention, that $\inf \emptyset = +\infty$).

Proof. Define $\phi(\theta) := \langle \delta_x - \mu, (-\mathbb{1}) \vee (\theta - d(x,\cdot))\rangle$. Note that ϕ is continuous. In view of Corollary 3.1 we have to maximize ϕ over $[-1, 1]$. Let $\theta, \tilde\theta \in [-1,1]$, such that $\theta > \tilde\theta$. Then

$$\begin{aligned}
\phi(\theta) - \phi(\tilde\theta) = \theta \; - \int_{B(x,\theta+1)} \theta - d(x,y)\,d\mu(y) \; - \int_{S\setminus B(x,\theta+1)} -\mathbb{1}\,d\mu \\
- \tilde\theta \; + \int_{B(x,\tilde\theta+1)} \tilde\theta - d(x,y)\,d\mu(y) \; + \int_{S\setminus B(x,\tilde\theta+1)} -\mathbb{1}\,d\mu \\
= (\theta - \tilde\theta)\big(1 - \mu\big(B(x,\tilde\theta+1)\big)\big) \; - \int_{B(x,\theta+1)\setminus B(x,\tilde\theta+1)} \theta - d(x,y) + 1\,d\mu(y).
\end{aligned}$$

Here we used that

$$\int_{S\setminus B(x,\theta+1)} \mathbb{1}\,d\mu - \int_{S\setminus B(x,\tilde\theta+1)} \mathbb{1}\,d\mu \;=\; -\big[\mu\big(B(x,\theta+1)\big) - \mu\big(B(x,\tilde\theta+1)\big)\big] \tag{22}$$

$$= -\mu\big(B(x,\theta+1) \setminus B(x,\tilde\theta+1)\big). \tag{23}$$

Therefore, $\phi(\theta) > \phi(\tilde\theta)$ if and only if

$$(\theta - \tilde\theta)\big(1 - \mu\big(B(x,\tilde\theta+1)\big)\big) \; > \int_{B(x,\theta+1)\setminus B(x,\tilde\theta+1)} \theta - d(x,y) + 1\,d\mu(y). \tag{24}$$

The function $\tilde\theta \mapsto \mu(B(x,\tilde\theta+1))$ is non-decreasing, since μ is a positive measure. According to the definition of θ_0, $\mu(B(x,\tilde\theta+1)) \geq 1$ for all $\tilde\theta > \theta_0$. In that case, inequality (24) cannot hold, because the right-hand side is non-negative. We conclude that ϕ is non-increasing on $(\theta_0, 1]$. (If $\theta_0 = 1$, then $(\theta_0, 1] = \emptyset$ and this statement is true trivially.)

We claim that ϕ is strictly increasing on $[-1, \theta_0)$. If $\theta_0 = -1$, then $[-1, \theta_0) = \emptyset$ and there is nothing to prove. So assume $\theta_0 > -1$. To prove the claim in this case, take $\theta, \tilde{\theta} \in [-1, \theta_0)$, $\theta > \tilde{\theta}$. For all $y \in B(x, \theta+1) \setminus B(x, \tilde{\theta}+1)$ one has $d(x, y) \geq \tilde{\theta}+1$, so $\theta - \tilde{\theta} \geq \theta - d(x, y)+1$. So if the condition

$$(\theta - \tilde{\theta})\big(1 - \mu\big(B(x, \tilde{\theta}+1)\big)\big) \;>\; \int_{B(x,\theta+1)\setminus B(x,\tilde{\theta}+1)} \theta - \tilde{\theta}\, d\mu(y) \tag{25}$$

is satisfied, then also condition (24). Since $\theta > \tilde{\theta}$, condition (25) holds if and only if

$$1 - \mu\big(B(x, \tilde{\theta}+1)\big) > \mu\big(B(x, \theta+1) \setminus B(x, \tilde{\theta}+1)\big),$$

which is equivalent to the condition

$$\mu\big(B(x, \theta+1)\big) < 1. \tag{26}$$

Thus, if condition (26) holds, then (24) is satisfied and $\phi(\theta) > \phi(\tilde{\theta})$. By definition of θ_0, (26) holds for all $\theta < \theta_0$. Thus, ϕ is strictly increasing on $[-1, \theta_0)$.

Because ϕ is continuous on $[-1, 1]$, strictly increasing on $[-1, \theta_0)$ and non-increasing on $(\theta_0, 1]$, ϕ attains its maximum value at θ_0. $\square$

Remark 3.1. Paradoxically, in the special case that $\mu \in \mathcal{P}(S)$, Proposition 3.1 claims that the value of the norm $\|\delta_x - \mu\|^*_{\mathrm{FM}}$ can be obtained by taking $\theta = 1$, instead of $\theta = \theta_0$, which is possibly less than 1. However, the proof of Theorem 3.1 shows that the *minimum* value for θ at which the maximum of the function ϕ is attained, which equals the stated Fortet–Mourier norm, is θ_0. Since ϕ is non-increasing on $(\theta_0, 1]$, there must exist also a maximum value at which this (same) maximum value is attained, say θ_1. In case μ is a probability measure, Proposition 3.1 shows that $\theta_1 = 1$. So there is no contradiction between the result of Theorem 3.1 and Proposition 3.1.

Let us collect some immediate consequences of Theorem 3.1.

Corollary 3.2. *Let $x \in S$ and $\mu \in \mathcal{M}^+(S)$ with $\mu(S) < 1$. Then $\theta_0 = 1$ and*

$$\big\|\delta_x - \mu\big\|^*_{\mathrm{FM}} = \big\langle \delta_x - \mu, (-\mathbb{1}) \vee (1 - d(x,\cdot))\big\rangle = 1 - \mu(S) + \big\langle \mu, 2 \wedge d(x,\cdot)\big\rangle.$$

Corollary 3.3. *Let $x \in S$, $\alpha > 0$ and $\mu \in \mathcal{M}^+(S)$. Then*

$$\big\|\alpha\delta_x - \mu\big\|^*_{\mathrm{FM}} = \big\langle \alpha\delta_x - \mu, (-\mathbb{1}) \vee (\theta_0^\alpha - d(x,\cdot))\big\rangle = \alpha\theta_0^\alpha - \big\langle \mu, (-\mathbb{1}) \vee (\theta_0^\alpha - d(x,\cdot))\big\rangle,$$

where $\theta_0^\alpha = \theta_0^\alpha(x) = 2 \wedge \inf\big\{r \geq 0 : \mu\big(B(x, r)\big) \geq \alpha\big\} - 1 \in [-1, 1]$.

Proof. One has $\|\alpha\delta_x - \mu\|^*_{\mathrm{FM}} = \alpha\|\delta_x - \alpha^{-1}\mu\|^*_{\mathrm{FM}}$. Now apply Theorem 3.1 to the measure $\alpha^{-1}\mu$ instead of μ. $\square$

The following corollary nicely generalizes the expression for the well-known Fortet–Mourier distance between two Dirac measures, see (5):

Corollary 3.4. *Let $x, y \in S$ and $\alpha, \beta > 0$. Then*

$$\big\|\alpha\delta_x - \beta\delta_y\big\|^*_{\mathrm{FM}} = |\alpha - \beta| + (\alpha \wedge \beta)(2 \wedge d(x, y)). \tag{27}$$

Proof. Corollary 3.3 yields

$$\left\|\alpha\delta_x - \beta\delta_y\right\|^*_{\mathrm{FM}} = \alpha\theta_0^\alpha - \beta\big((-1)\vee(\theta_0^\alpha - d(x,y))\big), \tag{28}$$

with $\theta_0^\alpha = \big(2\wedge\inf\{r\geq 0 : \delta_y\big(B(x,r)\big)\geq \alpha/\beta\}\big) - 1 \in [-1,1]$. One easily checks that

$$\left\{r\geq 0 : \delta_y\big(B(x,r)\big)\geq\alpha/\beta\right\} = \begin{cases}\{r : r > d(x,y)\}, & \text{if } \alpha/\beta\leq 1,\\ \emptyset, & \text{if } \alpha/\beta > 1.\end{cases}$$

Therefore,

$$\theta_0^\alpha = \begin{cases} d(x,y)-1, & \text{if } \alpha\leq\beta \text{ and } d(x,y)<2,\\ 1, & \text{otherwise.}\end{cases}$$

From (28) we find that

$$\begin{aligned}\left\|\alpha\delta_x - \beta\delta_y\right\|^*_{\mathrm{FM}} &= \begin{cases}\alpha(d(x,y)-1)+\beta, & \text{if } \alpha\leq\beta \text{ and } d(x,y)<2,\\ \alpha-\beta((-1)\vee 1-d(x,y)), & \text{otherwise,}\end{cases}\\ &= \begin{cases}\beta-\alpha+\alpha d(x,y), & \text{if } \alpha\leq\beta \text{ and } d(x,y)<2,\\ \alpha-\beta\big(1-2\wedge d(x,y)\big), & \text{otherwise,}\end{cases}\\ &= \begin{cases}|\alpha-\beta|+\alpha d(x,y), & \text{if } \alpha\leq\beta \text{ and } d(x,y)<2,\\ \alpha-\beta+\beta(2\wedge d(x,y)), & \text{otherwise,}\end{cases}\\ &= |\alpha-\beta| + (\alpha\wedge\beta)(2\wedge d(x,y)).\end{aligned} \tag{29}$$

Here we used (18) in the second step. To get to (29), for the case $\alpha\leq\beta$ and $d(x,y)\geq 2$, we used that $2(\alpha\wedge\beta) = \alpha+\beta-|\alpha-\beta|$. □

Remark 3.2. Without the use of the results that we presented, one could estimate as follows. Put $z := x$ if $\alpha\geq\beta$ and $z := y$ if $\alpha<\beta$. Then

$$\begin{aligned}\left\|\alpha\delta_x - \beta\delta_y\right\|^*_{\mathrm{FM}} &= \left\|(\alpha\wedge\beta)(\delta_x-\delta_y) \;+\; |\alpha-\beta|\delta_z\right\|^*_{\mathrm{FM}}\\ &\leq (\alpha\wedge\beta)\left\|\delta_x-\delta_y\right\|_F M^* + |\alpha-\beta|\|\delta_z\|^*_{\mathrm{FM}}\\ &= (\alpha\wedge\beta)(2\wedge d(x,y)) + |\alpha-\beta|.\end{aligned}$$

The point of Corollary 3.4 is, that equality holds.

We could not obtain an explicit expression for the suprema in (10) or (11), like (27), when ν is a weighted sum of two or more Dirac measures. The distance can be computed though in those cases, by algorithms that we shall exhibit in the next section.

4. Distance to positive molecular measures — an algorithmic approach

We are now concerned with computing $\|\nu-\mu\|^*_{\mathrm{FM}}$ where $\nu\in\mathcal{M}ol^+(S)$ and $\mu\in\mathcal{M}^+(S)$. Explicit expressions, like those presented for $\nu = \alpha\delta_x$ in the previous section, could not be obtained. It is possible to compute the norm in particular cases, most importantly when μ is also a positive molecular measure. Put otherwise, we shall provide an exact algorithm to compute $\|\mu\|^*_{\mathrm{FM}}$ for any $\mu\in\mathcal{M}ol(S)$. Note we assume the generality of (S,d) being a metric space. Thus, our results provide a substantial generalization of both [26,38].

Let $\nu\in\mathcal{M}ol^+(S)$ and $\mu\in\mathcal{M}^+(S)$ and write $\nu = \sum_{i=1}^N \alpha_i\delta_{x_i}$ with all x_i distinct and $\alpha_i>0$. Put $P := \mathrm{supp}(\nu) = \{x_1,\ldots,x_N\}$ and view P as a metric space for the restriction of d to P.

Define $\iota : B^P_{\mathrm{FM}} \to [-1,1]^N$ by $\iota(f) := (f(x_1), \ldots, f(x_N))$. Since for any $f, g \in B^P_{\mathrm{FM}}$ and $j \in \{1, \ldots, N\}$,

$$\left|\iota(f)_j - \iota(g)_j\right| \le \|f - g\|_\infty \le \|f - g\|_{\mathrm{FM}},$$

ι is a non-expansive map when $[-1,1]^N$ is equipped with the max-distance or the Euclidean metric. For any $\tau \in \mathcal{M}(S)$, define

$$\psi_\tau : [-1,1]^N \to \mathbb{R} : \theta \mapsto \Big\langle \tau, (-\mathbb{1}) \vee \bigvee_{i=1}^N (\theta_i - d(x_i, \cdot)) \Big\rangle. \tag{30}$$

The motivation for studying this function is given by Theorem 2.1:

$$\left\|\nu - \mu\right\|^*_{\mathrm{FM}} = \sup_{\theta \in [-1,1]^N} \psi_{\nu-\mu}(\theta) = \sup_{f \in B^P_{\mathrm{FM}}} \psi_{\nu-\mu}\big(\iota(f)\big). \tag{31}$$

Proposition 4.1. *If $[-1,1]^N$ is equipped with the* max*-distance or the Euclidean metric, then ψ_τ is Lipschitz continuous and $|\psi_\tau|_L \le \|\tau\|_{\mathrm{TV}}$. For $\tau \in \mathcal{M}^+(S)$, ψ_τ is convex on $[-1,1]^N$.*

Proof. Let $\theta, \tilde{\theta} \in [-1,1]^N$. Recall the definition of h_θ in (14). According to Lemma 1.2 *(ii)*, for every $x \in S$ one has

$$\left|h_\theta(x) - h_{\tilde{\theta}}(x)\right| \;\le\; \max_{1 \le i \le n} |\theta_i - \tilde{\theta}_i|. \tag{32}$$

This yields that for any $\tau \in \mathcal{M}(S)$,

$$\begin{aligned} \left|\psi_\tau(\theta) - \psi_\tau(\tilde{\theta})\right| \;&=\; \left|\langle \tau, h_\theta - h_{\tilde{\theta}} \rangle\right| \;\le\; \|\tau\|_{\mathrm{TV}} \|h_\theta - h_{\tilde{\theta}}\|_\infty \\ &\le\; \|\tau\|_{\mathrm{TV}} \max_{1 \le i \le n} |\theta_i - \tilde{\theta}_i| \;\le\; \|\tau\|_{\mathrm{TV}} \left(\sum_{i=1}^N |\theta_i - \tilde{\theta}_i|^2 \right)^{1/2}. \end{aligned} \tag{33}$$

So, ψ_τ is Lipschitz with $|\psi_\tau|_L \le \|\tau\|_{\mathrm{TV}}$.

Now assume that $\tau \in \mathcal{M}^+(S)$. Let $\theta, \tilde{\theta} \in [-1,1]^N$ and $0 \le t \le 1$. Then, using Lemma 1.2 *(ii)* and the positivity of τ to get to inequalities (34) and (35), we arrive at

$$\begin{aligned} \psi_\tau(t\theta + (1-t)\tilde{\theta}) \;&=\; \Big\langle \tau, (-\mathbb{1}) \vee \bigvee_{i=1}^N \big(t(\theta_i - d(x_i,\cdot)) + (1-t)(\tilde{\theta}_i - d(x_i,\cdot))\big) \Big\rangle \\ &\le\; \Big\langle \tau, \big(t(-\mathbb{1}) + (1-t)(-\mathbb{1})\big) \vee \Big(t \bigvee_{i=1}^N (\theta_i - d(x_i,\cdot)) + (1-t) \bigvee_{i=1}^N (\tilde{\theta}_i - d(x_i,\cdot)) \Big) \Big\rangle \quad (34) \\ &\le\; \Big\langle \tau, t \Big((-\mathbb{1}) \vee \bigvee_{i=1}^N (\theta_i - d(x_i,\cdot)) \Big) + (1-t) \Big((-\mathbb{1}) \vee \bigvee_{i=1}^N (\tilde{\theta}_i - d(x_i,\cdot)) \Big) \Big\rangle \quad (35) \\ &=\; t\psi_\tau(\theta) \;+\; (1-t)\psi_\tau(\tilde{\theta}). \end{aligned}$$

Thus, ψ_τ is convex on $[-1,1]^N$. □

Because $\psi_{\nu-\mu}$ is continuous, the suprema in (31) are attained on the compact sets $[-1,1]^N$ and $\iota(B^P_{\mathrm{FM}})$, respectively.

For general signed measure τ, ψ_τ is the difference of the convex functions ψ_{τ^+} and ψ_{τ^-}. Consequently, no particular 'convexity properties' of ψ_τ can be claimed. However, for $\tau = \nu - \mu$ with ν and μ positive measures as above, one can derive:

Proposition 4.2. *Let $\nu \in \mathcal{M}ol^+(S)$ with $P := \operatorname{supp}(\nu) = \{x_1, \ldots, x_N\}$ and $\mu \in \mathcal{M}^+(S)$. Then ψ_ν is 'linear' on $\iota(B^P_{\mathrm{FM}}) \subset [-1,1]^N$. In particular, $\psi_{\nu-\mu} = \psi_\nu - \psi_\mu$ is concave.*

Proof. One has $\psi_{\nu-\mu} = \psi_\nu - \psi_\mu$. ψ_μ is convex on $[-1,1]^N$, according to Proposition 4.1, so $-\psi_\mu$ is concave. We conclude by showing that ψ_ν is concave on $\iota(B^P_{\mathrm{FM}})$. For general $\theta \in [-1,1]^N$ one has $h_\theta(x_i) \geq \theta_i$ for all $i \in \{1, \ldots, N\}$. However, for $\theta \in \iota(B^P_{\mathrm{FM}})$ there is $f \in B^P_{\mathrm{FM}}$ such that $\theta = \iota(f)$. Therefore, according to Lemma 1.1 *(ii)*, $h_\theta(x_i) = f(x_i)$ for all i. Thus, for any $f \in B^P_{\mathrm{FM}}$, writing $f_i = \iota(f)_i = f(x_i)$,

$$\psi_\nu(\iota(f)) \;=\; \Big\langle \sum_{i=1}^N \alpha_i \delta_{x_i}, h_{\iota(f)} \Big\rangle \;=\; \sum_{i=1}^N \alpha_i f_i. \tag{36}$$

So ψ_ν on $\iota(B^P_{\mathrm{FM}})$ is the restriction of a linear functional on $\mathbb{R}^N$ to the convex subset $\iota(B^P_{\mathrm{FM}})$. In particular, ψ_ν is concave on $\iota(B^P_{\mathrm{FM}})$. $\square$

Thus, in view of (31) and Proposition 4.2 the problem of computing the Fortet–Mourier distance $\|\nu - \mu\|^*_{\mathrm{FM}}$ for $\nu \in \mathcal{M}ol^+(S)$ and $\mu \in \mathcal{M}^+(S)$ is equivalent to the problem of maximizing the concave and Lipschitzian function $\psi_{\nu-\mu}$ over the compact convex set $\iota(B^P_{\mathrm{FM}})$ in $\mathbb{R}^N$, where $P = \operatorname{supp}(\nu) \subset S$, consisting of N distinct points. This is again equivalent to minimizing the convex function $-\psi_{\nu-\mu}$ over $\iota(B^P_{\mathrm{FM}})$, which is a polyhedral convex set given by linear constraints that can be easily expressed explicitly (see (37)).

Minimization of convex functions has been widely studied (cf. e.g. [7,9,36]) and a wide variety of algorithms have been developed for convex minimization in the field of convex optimization. Generally, these problems can be solved highly efficiently by now, with specific algorithms for specific cases. Boyd and VandenBerghe even state ([9] p.8):

'*With only a bit of exaggeration, we can say that, if you formulate a practical problem as a convex optimization problem, then you have solved the original problem.*'

Therefore, we consider the theoretical side of computing distances of the form $\|\nu - \mu\|^*_{\mathrm{FM}}$ with $\nu \in \mathcal{M}ol^+(S)$ and $\mu \in \mathcal{M}^+(S)$ as solved.

Of course, in a practical setting, the implementation of the convex optimization algorithm of choice for a specific measure μ may require additional practical issues to be resolved. For example, one must be able to compute ψ_μ (approximately). In the following section we shall consider the special case where $\mu \in \mathcal{M}ol^+(S)$, i.e. computing the Fortet–Mourier norm of a molecular measure. But let us provide another example first.

Example 4.1. Let $S = [0, \infty)$, equipped with the Euclidean metric and take

$$\mu := e^{-x^2} dx, \qquad \nu := \tfrac{1}{2}\delta_0 + 2\delta_{\frac{1}{3}} + \tfrac{1}{5}\delta_{\frac{1}{2}} + \tfrac{1}{3}\delta_3.$$

We implemented an algorithm to minimize $-\psi_{\nu-\mu}$ over $\iota(B^P_{\mathrm{FM}})$, with $P = \{0, \frac{1}{3}, \frac{1}{2}, 3\}$, using the MATLAB '`fmincon`' function, see Appendix A.1. It resulted in

$$\|\nu - \mu\|^*_{\mathrm{FM}} = -\big(-\psi_{\nu-\mu}(f)\big) \approx 2.3921\ldots, \qquad \text{with } f = \big[1,\ 1,\ \tfrac{5}{6},\ 1\big].$$

5. Computing the Fortet–Mourier norm of a molecular measure

We conclude by specializing to the particular case where both ν and μ are positive molecular measures. That is, we show how to compute $\|\tau\|^*_{\mathrm{FM}}$ for $\tau \in \mathcal{M}ol(S)$. We present two ways to proceed: one by specializing the results of the previous section and one that is special to this

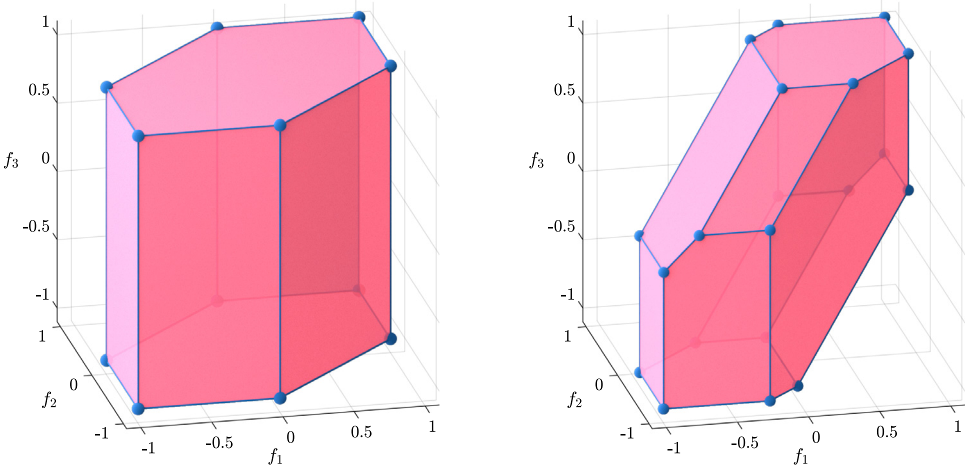

Fig. 1. The unit ball B^P_{FM} for the norm $\|\cdot\|_{\text{FM}}$ on the space $\text{BL}(P,d)$, where $P=\{x_1,x_2,x_3\}$ and $f\in\text{BL}(P,d)$ is represented by $(f_i)\in\mathbb{R}^3$ with $f_i:=f(x_i)$. The defining conditions are given by (37). The metric d is defined by $d_{ij}:=d(x_i,x_j)$ and differs for the two cases shown.
Left: $d_{12}=1,\ d_{13}=2,\ d_{23}=3$. Right: $d_{12}=0.75,\ d_{13}=1,\ d_{23}=1.25$.

specific case. Each has its benefits and drawbacks, which we shall discuss. We start with the latter method. We note that [26] (see also [19] Appendix) provides an algorithm to compute $\|\tau\|^*_{\text{FM}}$ when $S=\mathbb{R}$ or an interval therein. [38] exhibits a method that works for general space S, but τ must be the difference of two empirical measures. That is, the coefficients of the Diracs are quite specific. Before starting any further considerations, note that if τ or $-\tau$ is positive, then $\|\tau\|^*_{\text{FM}}=\|\tau\|_{\text{TV}}=\sum_i|\alpha_i|$ if $\tau=\sum_i\alpha_i\delta_{x_i}$. Thus, we shall assume $\tau^+\neq 0$ and $\tau^-\neq 0$.

As before, let $(S,d$ be a metric space and $P=\{x_1,\dots x_n\}$ a set of n distinct points in S. P inherits the metric structure of S, by restriction. Put $d_{ij}:=d(x_i,x_j)$. It is readily verified that

$$\begin{aligned}\iota(B^P_{\text{FM}}) &= \bigcap_{1\le k\le n}\left\{f\in\mathbb{R}^n : |f_k|\le 1\right\} \;\cap \bigcap_{1\le i<j\le n}\left\{f\in\mathbb{R}^n : |f_i-f_j|d_{ij}^{-1}\le 1\right\}\\ &= \bigcap_{1\le k\le n}\left\{f_k\le 1\right\}\cap\left\{-f_k\le 1\right\} \;\cap \bigcap_{1\le i<j\le n}\left\{(f_i-f_j)d_{ij}^{-1}\le 1\right\}\cap\left\{(f_j-f_i)d_{ij}^{-1}\le 1\right\}. \quad (37)\end{aligned}$$

Expression (37) is in the form of the standard linear programming representation of the domain of the objective function as an intersection of finitely many half-spaces (see e.g. [14]). Fig. 1 shows two unit balls B^P_{FM}, with P consisting of three points, for two different metrics.

Write $\tau=\sum_{i=1}^n\alpha_i\delta_{x_i}$ with $0\neq\alpha_i\in\mathbb{R}$ and $x_i\in S$, and put $P:=\text{supp}(\tau)=\{x_1,\dots x_n\}$. Lemma 1.1 *(ii)* implies that restriction to P gives a surjective map from B^S_{FM} onto B^P_{FM}. Therefore,

$$\|\tau\|^*_{\text{FM}} = \sup_{g\in B^S_{\text{FM}}}\langle\tau,g,\rangle = \sup_{f\in B^P_{\text{FM}}}\langle\tau|_P,f\rangle = \max_{f\in\iota(B^P_{\text{FM}})}\sum_{i=1}^n\alpha_i f_i. \quad (38)$$

Thus, $\|\tau\|^*_{\text{FM}}$ can be computed using one of the many existing – very efficient – optimization algorithms that use linear programming, such as Gurobi and CPLEX, or the built-in '`linprog`' function in MATLAB, using the standard domain description (37). In these algorithms there is an initial step in which an extreme point of the domain is sought to start the search for the optimum. Here, $\pm(1,\dots,1)$ are always extreme points. If $\sum_i \alpha_i \geq 0$ one may start at $(1,\dots,1)$. If $\sum_i \alpha_i < 0$,one may start at $(-1,\dots,-1)$. This reduces the number of vertices of B^P_{FM} that needs to be examined by the optimization algorithm in the worst case by a factor two.

Example 5.1. Let $P = \{x_1, x_2, x_3\}$ and $d_{12} = 1$, $d_{13} = 2$ and $d_{23} = 3$. Notice that the prescribed distances are consistent with the triangle inequality. Take $\alpha_1 = 1$, $\alpha_2 = -\frac{1}{3}$ and $\alpha_3 = -\frac{2}{3}$. Eq. (19) gives an explicit result in this case:

$$\Big\|\sum_i \alpha_i \delta_{x_i}\Big\|^*_{\mathrm{FM}} = |\alpha_2|(2\wedge d_{12}) + (1-|\alpha_2|)(2\wedge d_{13}) = \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 2 = \frac{5}{3}.$$

We implemented an algorithm for computing the Fortet–Mourier norms of molecular measures in MATLAB, using the '`linprog`' algorithm, see Appendix A.2. It returned the same result for the norm, but also an f at which the optimum is attained. In this case $f = (1, 0, -1)$. This f corresponds precisely to the function $(-\mathbb{1}) \vee (1 - d(x_1,\cdot))$ that appears in the theoretical result, Proposition 3.1.

The above linear programming algorithm for computing $\|\tau\|^*_{\mathrm{FM}}$ has as domain for the objective function a polygon in $\mathbb{R}^n$, where n is the number of points in the support of $\tau \in \mathcal{M}ol(S)$. The dimensionality of the optimization problem can be reduced by halve, by resorting to the results of Section 4. The number of points in the support of either τ^+ or τ^- is less than $n/2$ or both have precisely $n/2$ points in their support. The one with the least number of points, say τ^- with $N \leq n/2$ points, can play the role of ν in Section 4, while the other takes up the role of μ, simply because $\|\tau\|^*_{\mathrm{FM}} = \|-\tau\|^*_{\mathrm{FM}}$.

Thus, with $\nu = \tau^-$ (say) and $P = \mathrm{supp}(\nu) = \{x_1,\dots,x_N\}$,

$$\|\tau\|^*_{\mathrm{FM}} = \big\|\tau^- - \tau^+\big\|^*_{\mathrm{FM}} = \max_{f\in\iota(B^P_{\mathrm{FM}})} \psi_{-\tau}(f) = -\min_{f\in\iota(B^P_{\mathrm{FM}})} \big(-\psi_{-\tau}(f)\big),$$

according to (31) and the further discussion in Section 4. The polygonal domain of optimization $\iota(B^P_{\mathrm{FM}})$ is still given by (37), but now has reduced dimension $N \leq n/2$. In this setting, ψ_{τ^-} is 'linear' on $\iota(B^P_{\mathrm{FM}})$ (Proposition 4.2). If

$$\tau = \sum_{j=1}^{n-N} \beta_j \delta_{y_j} - \sum_{i=1}^{N} \alpha_i \delta_{x_i}, \qquad \alpha_i, \beta_j > 0,$$

then for $\theta \in \iota(B^P_{\mathrm{FM}})$

$$\psi_{-\tau}(\theta) = \sum_{i=1}^{N} \alpha_i\theta_i - \sum_{j=1}^{n-N} \beta_j\Big((-1) \vee \max_{1\leq i\leq N}(\theta_i - d(x_i, y_j))\Big). \tag{39}$$

So, the reduction in dimension of the domain is to the cost of (part of the) 'linearity' of the objective function $-\psi_{-\tau}$ on $\iota(B^P_{\mathrm{FM}})$, although it is still convex. Application of one of the existing (efficient) convex optimization algorithms now yields $\|\tau\|^*_{\mathrm{FM}}$ on a lower dimensional domain.

Remark 5.1. Our proposed algorithmic approaches for computing Fortet–Mourier norms of molecular measures make it possible in principle to approximate distances $\|\mu - \nu\|^*_{\mathrm{FM}}$ for arbitrary $\mu, \nu \in \mathcal{M}^+(S)$ too. Two approaches can be envisioned, based upon the results presented so far.

First, one could approximate both μ and ν first by particular positive molecular measures, namely rescaled empirical measures μ_n and ν_n, as proposed in [38] for the Dudley distance and μ, ν probability measures. This can be done arbitrarily well, by increasing the number of points in the support. Then, one can approximate $\|\mu - \nu\|^*_{\mathrm{FM}}$ by $\|\mu_n - \nu_n\|^*_{\mathrm{FM}}$, where the latter can be computed using (38), (37) and linear programming. Convergence rates of the initial approximation of the measures by molecular measures, when S a bounded subset of $\mathbb{R}^d$ and μ, ν of equal mass, are given in [38] Corollary 3.5 for the Dudley distance.

A second – similar – approach is to approximate only one of the measures, say ν, by rescaled empirical measures ν_n and approximate $\|\mu - \nu\|^*_{\mathrm{FM}}$ by $\|\mu - \nu_n\|^*_{\mathrm{FM}}$ using (31), (37) and convex optimization.

In computational practice, a key question is to determine what number of Dirac measures is needed such that the approximation of the arbitrary positive measure by the positive molecular measure is as small as desired. The smaller that number, the more efficient the norm $\|\mu_n - \nu_n\|_{\mathrm{FM}}$ or $\|\mu - \nu_n\|^*_{\mathrm{FM}}$ can be computed. The above mentioned empirical measures do approximate the original measure, but possibly with 'far too many' points in the support when compared to a molecular measure in which the weights may differ among the Dirac measures. While results (in [38]) exist for empirical measures, for approximation with arbitrary positive combinations of Dirac measures there are no results available to our knowledge that guarantee a minimal number of Dirac measures that is needed to obtain an approximation of the norm $\|\mu - \nu\|^*_{\mathrm{FM}}$ that is within tolerance. Also recall the discussion in the Introduction about the best approximation problem. These topics deserve further research.

6. Computing Dudley norms

So far we have been discussing expressions for and the computation of Fortet–Mourier norms only. The dual bounded Lipschitz norm on $\mathcal{M}(S)$, also known as Dudley norm, or flat metric – for the associated metric – is also considered often. It is given by

$$\|\mu\|^*_{\mathrm{BL}} := \sup_{f \in B^S_{\mathrm{BL}}} \langle \mu, f \rangle, \qquad B^S_{\mathrm{BL}} := \left\{ f \in \mathrm{BL}(S) : \|f\|_\infty + |f|_L \le 1 \right\}.$$

The results presented in Sections 2 and 3 do not readily generalize to the $\|\cdot\|^*_{\mathrm{BL}}$-norm. The main issue is, that the geometric shape of B^S_{BL} is more complicated than that of B^S_{FM}. For example, we could change the $\|\cdot\|_\infty$-norm of a function in B^S_{FM} 'independently' from its Lipschitz constant. For B^S_{BL} that is not possible, since the constraint $\|f\|_\infty + |f|_L \le 1$ should be maintained.

However, the results of Section 5 can be carried over to the $\|\cdot\|^*_{\mathrm{BL}}$-norm. The corresponding statement of (38) is still valid with 'FM' replaced by 'BL'. The description of $\iota(B^P_{\mathrm{BL}})$ becomes more awkward though. Without giving the lengthy proof here, we obtained

Lemma 6.1. *Let $n \ge 2$ and $P = \{x_1, \ldots, x_n\} \subset S$, consisting of distinct points. Put $d_{ij} := d(x_i, x_j)$. Then*

$$\begin{aligned}\iota(B^P_{\mathrm{BL}}) \;=\; & \bigcap_{1 \le i,j,k \le n,\ k \ne i \ne j \ne k} \left\{ f_k + (f_i - f_j) d_{ij}^{-1} \le 1 \right\} \cap \left\{ -f_k - (f_i - f_j) d_{ij}^{-1} \le 1 \right\} \\ & \cap \bigcap_{1 \le i,k \le n,\ i \ne k} \left\{ f_k + (f_k - f_i) d_{ik}^{-1} \le 1 \right\} \cap \left\{ -f_k - (f_k - f_i) d_{ik}^{-1} \le 1 \right\}.\end{aligned}$$

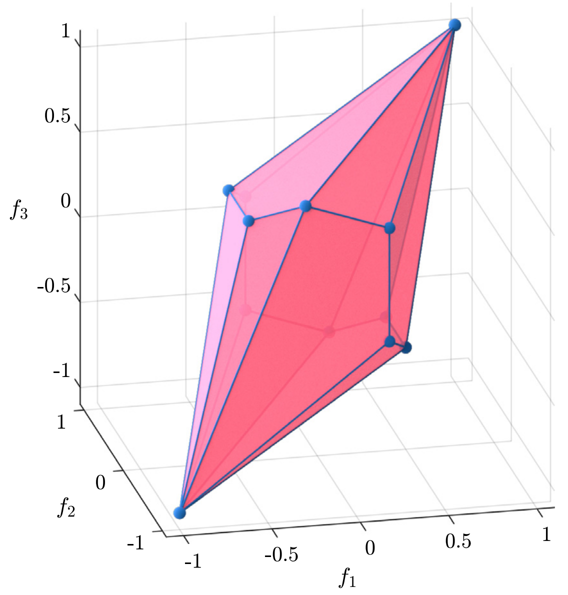

Fig. 2. The unit ball B^P_{BL} for the norm $\|\cdot\|_{\mathrm{BL}}$ on the space $\mathrm{BL}(P, d)$, where $P = \{x_1, x_2, x_3\}$ and $f \in \mathrm{BL}(P, d)$ is represented by $(f_i) \in \mathbb{R}^3$ with $f_i := f(x_i)$. The defining conditions are given by those in Lemma 6.1, while the metric d is defined by $d_{ij} := d(x_i, x_j)$ with $d_{12} = 1$, $d_{13} = 2$, $d_{23} = 3$.

In Fig. 2 an example is shown of a unit ball B^P_{BL} for P consisting of three points. Compare the more complex geometric structure with those of B^P_{FM} presented in Fig. 1.

Thus, one has — with P as in Lemma 6.1 and $0 \neq \alpha_i \in \mathbb{R}$:

$$\Big\|\sum_{i=1}^n \alpha_i \delta_{x_i}\Big\|^*_{\mathrm{BL}} = \max_{f\in\iota(B^P_{\mathrm{BL}})} \sum_{i=1}^n \alpha_i f_i = -\min_{f\in\iota(B^P_{\mathrm{BL}})} \Big(-\sum_{i=1}^n \alpha_i f_i\Big), \tag{40}$$

which optimum can again be found by a linear programming algorithm, now by using Lemma 6.1 for the standard description of the domain of optimization as intersection of half-spaces.

Data availability

No data was used for the research described in the article.

Acknowledgements

We thank M.A. Müller for preparing the graphics showing the unit balls in the spaces of bounded Lipschitz functions over a finite set of points in Figs. 1 and 2. Moreover, we thank one of the anonymous reviewers for pointing to applications of metrics on measures (generalized Wasserstein metrics in particular) to the field of image processing and analysis.

Appendix A. MATLAB implementations

A.1. FM-distance between a positive linear combination of Dirac measures and a positive measure

```
% Computes ||\nu-\mu||_FM^* for \nu in Mol^+(S), \mu in M^+(S),
% S=(Smin, Smax)\subset\R (possibly Smin=-Inf, Smax=+Inf)
% \mu=h d\lambda, \mu abs ct wrt Lebesgue measure, \nu=sum a_i delta_{s_i}
Smin = 0;
Smax = Inf;
a = [1/2 2 1/5 1/3];  n = length(a);
P = [0 1/3 1/2 3];
h = @(s) exp(-s^2);
integrand = @(s,f)h(s)*max(-1,max(f-abs(P-s)));
psi = @(f) -dot(a,f)+integral(@(s) integrand(s,f),Smin,Smax,'ArrayValued',true);
A = [];
for i = 1:n
   for j = i+1:n
      B = zeros(1,n);
      B(i) = abs(P(i)-P(j))^(-1);
      B(j) = -B(i);
      A = [A;B];
   end
end
A = [A;eye(n)];
A = [A;-A];
b = ones(n^2+n,1);
f0 = zeros(1,n);
[f,val] = fmincon(psi,f0,A,b);
f
norm = -val
```

A.2. FM-norm of a linear combination of Dirac measures

```
function [norm,f] = FMdualnorm(a,dist)
%a=[a_1 a_2 ... a_n], mu=sum_i=1^n a_i delta_s_i, dist=[d_ij]_i,j=1^n matrix
%norm =||mu||_FM^*, f=ext pt for which <mu,f>=norm
n=length(a);
A=[];
for i=1:n
   for j=i+1:n
      B=zeros(1,n);
      B(i)=dist(i,j)^(-1);
      B(j)=-B(i);
      A=[A;B];
```

```
   end
end
A=[A;eye(n)];
A=[A;-A];
b=ones(n^2+n,1); %m=#rows of A=2n+2(n choose 2)=2n+n(n-1)=n^2+n
minus_a=-a;
f=linprog(minus_a,A,b);
norm=a*f;
```

References

[1] A.S. Ackleh, B.G. Fitzpatrick, H.R. Thieme, Rate distributions and survival of the fittest: a formulation on the space of measures, Discrete Contin. Dyn. Syst. Ser. B 5 (4) (2005) 917–928.

[2] A.S. Ackleh, D.F. Marshall, H.E. Heatherly, B.G. Fitzpatrick, Survival of the fittest in a generalized logistic model, Math. Models Methods Appl. Sci. 9 (9) (1999) 1379–1391.

[3] A.S. Ackleh, N. Saintier, J. Skrzeczkowski, Sensitivity equations for measure-valued solutions of transport equations, Math. Biosci. Eng. 17 (1) (2019) 514–537.

[4] T. Alkurdi, S.C. Hille, O. van Gaans, Ergodicity and stability of a dynamical system perturbed by impulsive random interventions, J. Math. Anal. Appl. 63 (2013) 480–494.

[5] L. Ambrosio, M. Fornasier, M. Morandotti, G. Savaré, Spatially inhomogeneous evolutionary games, Comm. Pure Appl. Math. 74 (2021) 1353–1402.

[6] M. Bačák, Convex Analysis and Optimization in Hadamard Spaces, Vol. 22, De Gruyter, 2014.

[7] D.P. Bertsekas, Convex Optimization Algorithms, Athena Scientific, Belmont, 2015.

[8] V.I. Bogachev, Measure Theory, Volume II, Springer-Verlag, Berlin-Heidelberg, 2007.

[9] S. Boyd, L. VandenBerghe, Convex Optimization, Cambridge University Press, 2004.

[10] J.A. Carillo, R.M. Colombo, P. Gwiazda, A. Ulikowska, Structured populations, cell growth and balance laws, J. Differential Equations 252 (2012) 3245–3277.

[11] R.L. Church, Z. Drezner, A. Tamir, Extensions to the Weber problem, Comput. Oper. Res. 143 (2022) 105786.

[12] G.H. Cottet, S. Mas-Gallic, A particle method to solve the Navier–Stokes system, Numer. Math. 57 (1990) 805–827.

[13] D. Czapla, S.C. Hille, K. Horbacz, H. Wodjewódka-Ściążko, Continuous dependence of an invariant measure on the jump rate of a piecewise-deterministic Markov process, Math. Biosci. Eng. 17 (2) (2019) 1059–1073.

[14] Bertsimas. D., Introduction to Linear Optimization, Athena Scientific, Belmont, 1997.

[15] F. Deutsch, Existence of best approximations, J. Approx. Theory 28 (2) (1980) 132–154.

[16] J. Dolbeault, C. Schmeiser, The two-dimensional Keller–Segel model after blow-up, Discrete Contin. Dyn. Syst. 25 (1) (2009) 109–121.

[17] Z. Drezner, K. Klamroth, A. Schöbel, G.O. Wesolowsky, The Weber problem, in: Facility Location, Springer-Verlag, 2002, pp. 1–36.

[18] R. Dudley, Convergence of Baire measures, Studia Math. 27 (3) (1966) 251–268.

[19] J.H.M. Evers, Evolution Equations for Systems Governed By Social Interactions (Ph.D. thesis), Technische Universiteit Eindhoven, 2015.

[20] J.H.M. Evers, S.C. Hille, A. Muntean, Measure-valued mass evolution problems with flux boundary conditions and solution-dependent velocities, SIAM J. Math. Anal. 48 (3) (2016) 1929–1953.

[21] P. Gwiazda, S.C. Hille, K. Lyczek, A. Swierczewska-Gwiazda, Differentiability in perturbation parameter of measure solutions to perturbed transport equation, Kinet. Relat. Models 17 (2) (2019) 1093–1108.

[22] A. Herrero, J.J.L. Velázquez, A blow-up mechanism for a chemotaxis model, Ann. Sc. Norm. Super. Pisa Cl. Sci. 24 (4) (1997) 633–683.

[23] S.C. Hille, K. Horbacz, T. Szarek, Existence of a unique invariant measure for a class of equicontinuous Markov operators with application to a stochastic model for an autoregulated gene, Ann. Math. Blaise Pascal 23 (2) (2016) 171–217.

[24] S.C. Hille, E.S. Theewis, Norming and dense sets of extreme points of the unit ball in spaces of bounded Lipschitz functions, 2023, Preprint available on arXiv:2305.08500 submitted for publication.

[25] S.C. Hille, D.T.H. Worm, Embedding of semigroups of Lipschitz maps into positive linear semigroups on ordered Banach spaces generated by measures, Integral Equations Operator Theory 63 (3) (2009) 351–371.
[26] J. Jabłoński, A. Marciniak-Czochra, Efficient algorithms computing distances between Radon measures on $\mathbb{R}$, 2013, arXiv preprint.
[27] H.W. Kuhn, R.E. Kuenne, An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics, J. Reg. Sci. 4 (2) (1962) 21–33.
[28] A. Lasota, J. Myjak, Fractals, semifractals and Markov operators, Int. J. Bifurcation Chaos 9 (2) (1999) 307–325.
[29] A. Lasota, J. Myjak, T. Szarek, Markov operators with a unique invariant measure, J. Math. Anal. Appl. 276 (2002) 343–356.
[30] J. Lellmann, D.A. Lorenz, C. Schönlieb, T. Valkonen, Imaging with Kantorovich-Rubinstein discrepancy, SIAM J. Imaging Sci. 7 (2014) 2833–2859.
[31] E.J. McShane, Extension of range of functions, Bull. Amer. Math. Soc. 40 (1934) 837.
[32] J. Pachl, Uniform spaces and measures, in: Fields Institute Monographs, Volume 30, the Fields Institute for Research in the Mathematical Sciences, Springer, New York, 2013.
[33] B. Piccoli, F. Rossi, Generalized Wasserstein distance and its application to transport equations with source, Arch. Ration. Mech. Anal. 211 (2014) 335–358.
[34] Piccoli, B.F. Rossi, M. Tournus, A Wasserstein norm for signed measures, with application to a non local transport equation with source term, 2021, HAL preprint. https://arxiv.org/abs/hal-01665244v5.
[35] B. Piccoli, A. Tosin, Time-evolving measures and macroscopic modeling of pedestrian flow, Arch. Ration. Mech. Anal. 199 (2011) 707–738.
[36] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1972.
[37] B. Schmitzer, B. Wirth, A framework for Wasserstein-1-type metrics, J. Convex Anal. 26 (2019) 353–396.
[38] B.K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, G.R.G Lanckriet, On the empirical estimation of integral probability metrics, Electr. J. Stat. 6 (2012) 1550–1599.
[39] C. Villani, Topics in Optimal Transportation, American Mathematical Society, Providence, 2003.
[40] C. Villani, Optimal transport, in: Old and New, Grundlehren der mathematischen Wissenschaften, Vol. 338, Spinger-Verlag, Berlin, 2009.
[41] N. Weaver, Lipschitz Algebras, World Scientific Publishing, Singapore, 1999.