

# Chapter 5

## Construction of Fréchet Means and Multicouplings



When given measures  $\mu^1, \dots, \mu^N$  are supported on the real line, computing their Fréchet mean  $\bar{\mu}$  is straightforward (Sect. 3.1.4). This is in contrast to the multivariate case, where, apart from the important yet special case of compatible measures, closed-form formulae are not available. This chapter presents an iterative procedure that provably approximates at least a Karcher mean with mild restrictions on the measures  $\mu^1, \dots, \mu^N$ . The algorithm is based on the differentiability properties of the Fréchet functional developed in Sect. 3.1.6 and can be interpreted as classical steepest descent in the Wasserstein space  $\mathscr{W}_2(\mathbb{R}^d)$ . It reduces the problem of finding the Fréchet mean to a succession of pairwise transport problems, involving only the Monge–Kantorovich problem between two measures. In the Gaussian case (or any location-scatter family), the latter can be done explicitly, rendering the algorithm particularly appealing (see Sect. 5.4.1).

This chapter can be seen as a complementary to Chap. 4. On the one hand, one can use the proposed algorithm to construct the regularised Fréchet–Wasserstein estimator  $\hat{\lambda}_n$  that approximates a population version (see Sect. 4.3). On the other hand, it could be that the object of interest is the sample  $\mu^1, \dots, \mu^N$  itself, but that the latter is observed with some amount of noise. If one only has access to proxies  $\hat{\mu}^1, \dots, \hat{\mu}^N$ , then it is natural to use their Fréchet mean  $\hat{\bar{\mu}}$  as an estimator of  $\bar{\mu}$ . The proposed algorithm can then be used, in principle, in order to construct  $\hat{\bar{\mu}}$ , and the consistency framework of Sect. 4.4 then allows to conclude that if each  $\hat{\mu}^i$  is consistent, then so is  $\hat{\bar{\mu}}$ .

After presenting the algorithm in Sect. 5.1, we make some connections to Procrustes analysis in Sect. 5.2. A convergence analysis of the algorithm is carried out in Sect. 5.3, after which examples are given in Sect. 5.4. An extension to infinitely many measures is sketched in Sect. 5.5.

---

**Electronic Supplementary Material** The online version of this chapter ([https://doi.org/10.1007/978-3-030-38438-8\\_5](https://doi.org/10.1007/978-3-030-38438-8_5)) contains supplementary material.

## 5.1 A Steepest Descent Algorithm for the Computation of Fréchet Means

Throughout this section, we assume that  $N$  is a fixed integer and consider a fixed collection

$$\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathbb{R}^d) \quad \text{with } \mu^1 \text{ absolutely continuous with bounded density,} \quad (5.1)$$

whose unique (Proposition 3.1.8) Fréchet mean  $\bar{\mu}$  is sought. It has been established that if  $\gamma$  is absolutely continuous then the associated Fréchet functional

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^n W_2^2(\mu^i, \gamma), \quad \gamma \in \mathcal{W}_2(\mathbb{R}^d),$$

has Fréchet derivative (Theorem 3.1.14)

$$F'(\gamma) = -\frac{1}{N} \sum_{i=1}^N \log_\gamma(\mu^i) = -\frac{1}{N} \sum_{i=1}^N (\mathbf{t}_\gamma^{\mu^i} - \mathbf{i}) \in \text{Tan}_\gamma \quad (5.2)$$

at  $\gamma$ . Let  $\gamma_j \in \mathcal{W}_2(\mathbb{R}^d)$  be an absolutely continuous measure, representing our current estimate of the Fréchet mean at step  $j$ . Then it makes sense to introduce a step size  $\tau_j > 0$ , and to follow the steepest descent of  $F$  given by the negative of the gradient:

$$\gamma_{j+1} = \exp_{\gamma_j}(-\tau_j F'(\gamma_j)) = \left[ \mathbf{i} + \tau_j \frac{1}{N} \sum_{i=1}^N \log_{\gamma_j}(\mu^i) \right] \# \gamma_j = \left[ \mathbf{i} + \tau_j \frac{1}{N} \sum_{i=1}^N (\mathbf{t}_{\gamma_j}^{\mu^i} - \mathbf{i}) \right] \# \gamma_j.$$

In order to employ further descent at  $\gamma_{j+1}$ , it needs to be verified that  $F$  is differentiable at  $\gamma_{j+1}$ , which amounts to showing that the latter stays absolutely continuous. This will happen for all but countably many values of the step size  $\tau_j$ , but necessarily if the latter is contained in  $[0, 1]$ :

**Lemma 5.1.1 (Regularity of the Iterates)** *If  $\gamma_0$  is absolutely continuous and  $\tau = \tau_0 \in [0, 1]$ , then  $\gamma_1 = \exp_{\gamma_0}(-\tau_0 F'(\gamma_0))$  is also absolutely continuous.*

The idea is that push-forwards of  $\gamma_0$  under monotone maps are absolutely continuous if and only if the monotonicity is strict, a property preserved by averaging. See page 118 in the supplement for the details.

Lemma 5.1.1 suggests that the step size should be restricted to  $[0, 1]$ . The next result suggests that the objective function essentially tells us that the optimal step size, achieving the maximal reduction of the objective function (thus corresponding to an approximate line search), is exactly equal to 1. It does not rely on finite-dimensional arguments and holds when replacing  $\mathbb{R}^d$  by a separable Hilbert space.

**Lemma 5.1.2 (Optimal Stepsize)** *If  $\gamma_0 \in \mathcal{W}_2(\mathbb{R}^d)$  is absolutely continuous, then*

$$F(\gamma_1) - F(\gamma_0) \leq -\|F'(\gamma_0)\|^2 \left[ \tau - \frac{\tau^2}{2} \right]$$

and the bound on the right-hand side of the last display is minimised when  $\tau = 1$ .

*Proof.* Let  $S_i = \mathbf{t}_{\gamma_0}^{\mu^i}$  be the optimal map from  $\gamma_0$  to  $\mu^i$ , and set  $W_i = S_i - \mathbf{i}$ . Then

$$2NF(\gamma_0) = \sum_{i=1}^N W_2^2(\gamma_0, \mu^i) = \sum_{i=1}^N \int_{\mathbb{R}^d} \|S_i - \mathbf{i}\|^2 d\gamma_0 = \sum_{i=1}^N \|W_i\|_{\mathcal{L}^2(\gamma_0)}^2, \quad (5.3)$$

Both  $\gamma_1$  and  $\mu^i$  can be written as push-forwards of  $\gamma_0$  and (2.3) gives the bound

$$W_2^2(\gamma_1, \mu^i) \leq \int_{\mathbb{R}^d} \left\| \left[ (1-\tau)\mathbf{i} + \frac{\tau}{N} \sum_{j=1}^N S_j \right] - S_i \right\|_{\mathbb{R}^d}^2 d\gamma_0 = \left\| -W_i + \frac{\tau}{N} \sum_{j=1}^N W_j \right\|_{\mathcal{L}^2(\gamma_0)}^2.$$

For brevity, we omit the subscript  $\mathcal{L}^2(\gamma_0)$  from the norms and inner products. Developing the squares, summing over  $i = 1, \dots, N$  and using (5.3) gives

$$\begin{aligned} 2NF(\gamma_1) &\leq \sum_{i=1}^N \|W_i\|^2 - 2\frac{\tau}{N} \sum_{i,j=1}^N \langle W_i, W_j \rangle + N\tau^2 \left\| \sum_{j=1}^N \frac{1}{N} W_j \right\|^2 \\ &= 2NF(\gamma_0) - 2N\tau \left\| \sum_{i=1}^N \frac{1}{N} W_i \right\|^2 + N\tau^2 \left\| \sum_{i=1}^N \frac{1}{N} W_i \right\|^2, \end{aligned}$$

and recalling that  $W_i = S_i - \mathbf{i}$  yields

$$F(\gamma_1) - F(\gamma_0) \leq \frac{\tau^2 - 2\tau}{2} \left\| \frac{1}{N} \sum_{i=1}^N W_i \right\|^2 = -\|F'(\gamma_0)\|^2 \left[ \tau - \frac{\tau^2}{2} \right].$$

To conclude, observe that  $\tau - \tau^2/2$  is maximised at  $\tau = 1$ .

In light of Lemmata 5.1.1 and 5.1.2, we will always take  $\tau_j = 1$ . The resulting iteration is summarised as Algorithm 1. A first step in the convergence analysis is that the sequence  $(F(\gamma_j))$  is nonincreasing and that for any integer  $k$ ,

$$\frac{1}{2} \sum_{j=0}^k \|F'(\gamma_j)\|^2 \leq \sum_{j=0}^k F(\gamma_j) - F(\gamma_{j+1}) = F(\gamma_0) - F(\gamma_{k+1}) \leq F(\gamma_0).$$

As  $k \rightarrow \infty$ , the infinite sum on the left-hand side converges, so  $\|F'(\gamma_j)\|^2$  must vanish as  $j \rightarrow \infty$ .

**Remark 5.1.3** *The proof of Proposition 3.1.2 suggests a generalisation of Algorithm 1 to arbitrary measures in  $\mathcal{W}_2(\mathbb{R}^d)$  even if none are absolutely continuous. One can verify that Lemmata 5.1.2 and 5.3.5 (below) also hold in this setup, so it may be that convergence results also apply in this setup. The iteration no longer has the interpretation as steepest descent, however.*

---

**Algorithm 1** Steepest descent via Procrustes analysis

---

- (A) Set a tolerance threshold  $\varepsilon > 0$ .
  - (B) For  $j = 0$ , let  $\gamma_j$  be an arbitrary absolutely continuous measure.
  - (C) For  $i = 1, \dots, N$  solve the (pairwise) Monge problem and find the optimal transport map  $\mathbf{t}_{\gamma_j}^{\mu^i}$  from  $\gamma_j$  to  $\mu^i$ .
  - (D) Define the map  $T_j = N^{-1} \sum_{i=1}^N \mathbf{t}_{\gamma_j}^{\mu^i}$ .
  - (E) Set  $\gamma_{j+1} = T_j \# \gamma_j$ , i.e. push-forward  $\gamma_j$  via  $T_j$  to obtain  $\gamma_{j+1}$ .
  - (F) If  $\|F'(\gamma_{j+1})\| < \varepsilon$ , stop, and output  $\gamma_{j+1}$  as the approximation of  $\bar{\mu}$  and  $\mathbf{t}_{\gamma_{j+1}}^{\mu^i}$  as the approximation of  $\mathbf{t}_{\bar{\mu}}^{\mu^i}$ ,  $i = 1, \dots, N$ . Otherwise, return to step (C).
- 

## 5.2 Analogy with Procrustes Analysis

Algorithm 1 is similar in spirit to another procedure, *generalised Procrustes analysis*, that is used in shape theory. Given a subset  $B \subseteq \mathbb{R}^d$ , most commonly a finite collection of labelled points called *landmarks*, an interesting question is how to mathematically define the *shape* of  $B$ . One way to reach such a definition is to disregard those properties of  $B$  that are deemed irrelevant for what one considers this shape should be; typically, these would include its location, its orientation, and/or its scale. Accordingly, the shape of  $B$  can be defined as the equivalence class consisting of all sets obtained as  $gB$ , where  $g$  belongs to a collection  $\mathcal{G}$  of transformations of  $\mathbb{R}^d$  containing all combinations of rotations, translations, dilations, and/or reflections (Dryden and Mardia [45, Chapter 4]).

If  $B_1$  and  $B_2$  are two collections of  $k$  landmarks, one may define the distance between their shapes as the infimum of  $\|B_1 - gB_2\|^2$  over the group  $\mathcal{G}$ . In other words, one seeks to *register*  $B_2$  as close as possible to  $B_1$  by using elements of the group  $\mathcal{G}$ , with distance being measured as the sum of squared Euclidean distances between the transformed points of  $B_2$  and those of  $B_1$ . In a sense, one can think about the shape problem and the Monge problem as dual to each other. In the former, one is given constraints on how to optimally carry out the registration of the points with the cost being judged by how successful the registration procedure is. In the latter, one imposes that the registration be done *exactly*, and evaluates the cost by how much the space must be deformed in order to achieve this.

The optimal  $g$  and the resulting distance can be found in closed-form by means of *ordinary Procrustes analysis* [45, Section 5.2]. Suppose now that we are given

$N > 2$  collections of points,  $B_1, \dots, B_N$ , with the goal of minimising the sum of squares  $\|g_i B_i - g_j B_j\|^2$  over  $g_i \in \mathcal{G}$ .<sup>1</sup> As in the case of Fréchet means in  $\mathcal{W}_2(\mathbb{R}^d)$  (Sect. 3.1.2), there is a formulation in terms of sum of squares from the average  $N^{-1} \sum g_j B_j$ . Unfortunately, there is no explicit solution for this problem when  $d \geq 3$ . Like Algorithm 1, generalised Procrustes analysis (Gower [66]; Dryden and Mardia [45, p. 90]) tackles this “multimatching” setting by iteratively solving the pairwise problem as follows. Choose one of the configurations as an initial estimate/template, then register every other configuration to the template, employing ordinary Procrustes analysis. The new template is then given by the linear average of the registered configurations, and the process is iterated subsequently.

Paralleling this framework, Algorithm 1 iterates the two steps of registration and linear averaging given the current template  $\gamma_j$ , but in a different manner:

- (1) **Registration:** by finding the optimal transportation maps  $\mathbf{t}_{\gamma_j}^{\mu^i}$ , we identify each  $\mu^i$  with the element  $\mathbf{t}_{\gamma_j}^{\mu^i} - \mathbf{i} = \log_{\gamma_j}(\mu^i)$ . In this sense, the collection  $(\mu^1, \dots, \mu^N)$  is viewed in the common coordinate system given by the tangent space at the template  $\gamma_j$  and is registered to it.
- (2) **Averaging:** the registered measures are averaged linearly, using the common coordinate system of the registration step (1), as elements in the linear space  $\text{Tan}_{\gamma_j}$ . The linear average is then retracted back onto the Wasserstein space via the exponential map to yield the estimate at the  $(j+1)$ -th step,  $\gamma_{j+1}$ .

Notice that in the Procrustes sense, the maps that register each  $\mu^i$  to the template  $\gamma_j$  are  $\mathbf{t}_{\mu^i}^{\gamma_j}$ , the inverses of  $\mathbf{t}_{\gamma_j}^{\mu^i}$ . We will not use the term “registration maps” in the sequel, to avoid possible confusion.

## 5.3 Convergence of Algorithm 1

In order to tackle the issue of convergence, we will use an approach that is specific to the nature of optimal transport. This is because the Hessian-type arguments that are used to prove similar convergence results for steepest descent on Riemannian manifolds (Afsari et al. [1]) or Procrustes algorithms (Le [86]; Groisser [67]) do not apply here, since the Fréchet functional may very well fail to be twice differentiable.

In fact, even in Euclidean spaces, convergence of steepest descent usually requires a Lipschitz bound on the derivative of  $F$  (Bertsekas [19, Subsection 1.2.2]). Unfortunately,  $F$  is not known to be differentiable at discrete measures, and these constitute a dense set in  $\mathcal{W}_2$ ; consequently, this Lipschitz condition is very unlikely to hold. Still, this specific geometry of the Wasserstein space affords some advantages; for instance, we will place no restriction on the starting point for the iteration, except that it be absolutely continuous; and no assumption on how “spread out” the collection  $\mu^1, \dots, \mu^N$  is will be necessary as in, for example, [1, 67, 86].

---

<sup>1</sup> One needs to add an additional constraint to prevent registering all the collections to the origin.

**Theorem 5.3.1 (Limit Points are Karcher Means)** *Let  $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathbb{R}^d)$  be probability measures and suppose that one of them is absolutely continuous with a bounded density. Then, the sequence generated by Algorithm 1 stays in a compact set of the Wasserstein space  $\mathcal{W}_2(\mathbb{R}^d)$ , and any limit point of the sequence is a Karcher mean of  $(\mu^1, \dots, \mu^N)$ .*

Since the Fréchet mean  $\bar{\mu}$  is a Karcher mean (Proposition 3.1.8), we obtain immediately:

**Corollary 5.3.2 (Wasserstein Convergence of Steepest Descent)** *Under the conditions of Theorem 5.3.1, if  $F$  has a unique stationary point, then the sequence  $\{\gamma_j\}$  generated by Algorithm 1 converges to the Fréchet mean of  $\{\mu^1, \dots, \mu^N\}$  in the Wasserstein metric,*

$$W_2(\gamma_j, \bar{\mu}) \longrightarrow 0, \quad j \rightarrow \infty.$$

Alternatively, combining Theorem 5.3.1 with the optimality criterion Theorem 3.1.15 shows that the algorithm converges to  $\bar{\mu}$  when the appropriate assumptions on  $\{\mu^i\}$  and the Karcher mean  $\mu = \lim \gamma_j$  are satisfied. This allows to conclude that Algorithm 1 converges to the unique Fréchet mean when  $\mu^i$  are Gaussian measures (see Theorem 5.4.1).

The proof of Theorem 5.3.1 is rather elaborate, since we need to use specific methods that are tailored to the Wasserstein space. Before giving the proof, we state two important consequences. The first is the uniform convergence of the optimal maps  $\mathbf{t}_{\gamma_j}^{\mu^i}$  to  $\mathbf{t}_{\bar{\mu}}^{\mu^i}$  on compacta. This convergence does not immediately follow from the Wasserstein convergence of  $\gamma_j$  to  $\bar{\mu}$ , and is also established for the inverses. Both the formulation and the proof of this result are similar to those of Theorem 4.4.3.

**Theorem 5.3.3 (Uniform Convergence of Optimal Maps)** *Under the conditions of Corollary 5.3.2, there exist sets  $A, B^1, \dots, B^N \subseteq \mathbb{R}^d$  such that  $\bar{\mu}(A) = 1 = \mu^1(B^1) = \dots = \mu^N(B^N)$  and*

$$\sup_{\Omega_1} \left\| \mathbf{t}_{\gamma_j}^{\mu^i} - \mathbf{t}_{\bar{\mu}}^{\mu^i} \right\| \xrightarrow{j \rightarrow \infty} 0, \quad \sup_{\Omega_2^i} \left\| \mathbf{t}_{\mu^i}^{\gamma_j} - \mathbf{t}_{\mu^i}^{\bar{\mu}} \right\| \xrightarrow{j \rightarrow \infty} 0, \quad i = 1, \dots, N,$$

for any pair of compacta  $\Omega_1 \subseteq A$ ,  $\Omega_2^i \subseteq B^i$ . If in addition all the measures  $\mu^1, \dots, \mu^N$  have the same support, then one can choose all the sets  $B^i$  to be the same.

The other consequence is convergence of the optimal multicouplings.

**Corollary 5.3.4 (Convergence of Multicouplings)** *Under the conditions of Corollary 5.3.2, the sequence of multicouplings*

$$\left( \mathbf{t}_{\gamma_j}^{\mu^1}, \dots, \mathbf{t}_{\gamma_j}^{\mu^N} \right) \# \gamma_j$$

of  $\{\mu^1, \dots, \mu^N\}$  converges (in Wasserstein distance on  $(\mathbb{R}^d)^N$ ) to the optimal multicoupling  $(\mathbf{t}_{\bar{\mu}}^{\mu^1}, \dots, \mathbf{t}_{\bar{\mu}}^{\mu^N}) \# \bar{\mu}$ .

The proofs of Theorem 5.3.3 and Corollary 5.3.4 are given at the end of the present section.

The proof of Theorem 5.3.1 is achieved by establishing the following facts:

1. The sequence  $(\gamma_j)$  stays in a compact subset of  $\mathcal{W}_2(\mathbb{R}^d)$  (Lemma 5.3.5).
2. Any limit of  $(\gamma_j)$  is absolutely continuous (Proposition 5.3.6 and the paragraph preceding it).
3. Algorithm 1 acts continuously on its argument (Corollary 5.3.8).

Since it has already been established that  $\|F'(\gamma_j)\| \rightarrow 0$ , these three facts indeed suffice.

**Lemma 5.3.5** *The sequence generated by Algorithm 1 stays in a compact subset of the Wasserstein space  $\mathcal{W}_2(\mathbb{R}^d)$ .*

*Proof.* For all  $j \geq 1$ ,  $\gamma_j$  takes the form  $M_n \# \pi$ , where  $M_N(x_1, \dots, x_N) = \bar{x}$  and  $\pi$  is a multicoupling of  $\mu^1, \dots, \mu^N$ . The compactness of this set has been established in Step 2 of the proof of Theorem 3.1.5; see page 63 in the supplement, where this is done in a more complicated setup.

A closer look at the proof reveals that a more general result holds true. Let  $\mathcal{A}$  denote the steepest descent iteration, that is,  $\mathcal{A}(\gamma_j) = \gamma_{j+1}$ . Then the image of  $\mathcal{A}$ ,  $\{\mathcal{A}\mu : \mu \in \mathcal{W}_2(\mathbb{R}^d) \text{ absolutely continuous}\}$  has a compact closure in  $\mathcal{W}_2(\mathbb{R}^d)$ . This is also true if  $\mathbb{R}^d$  is replaced by a separable Hilbert space.

In order to show that a weakly convergent sequence  $(\gamma_j)$  of absolutely continuous measures has an absolutely continuous limit  $\gamma$ , it suffices to show that the densities of  $\gamma_j$  are uniformly bounded. Indeed, if  $C$  is such a bound, then for any open  $O \subseteq \mathbb{R}^d$ ,  $\liminf \gamma_k(O) \leq C \text{Leb}(O)$ , so  $\gamma(O) \leq C \text{Leb}(O)$  by the portmanteau Lemma 1.7.1. It follows that  $\gamma$  is absolutely continuous with density bounded by  $C$ . We now show that such  $C$  can be found that applies to all measures in the image of  $\mathcal{A}$ , hence to all sequences resulting from iterations of Algorithm 1.

**Proposition 5.3.6 (Uniform Density Bound)** *For each  $i = 1, \dots, N$  denote by  $g^i$  the density of  $\mu^i$  (if it exists) and  $\|g^i\|_\infty$  its supremum, taken to be infinite if  $g^i$  does not exist (or if  $g^i$  is unbounded). Let  $\gamma_0$  be any absolutely continuous probability measure. Then the density of  $\gamma_1 = \mathcal{A}(\gamma_0)$  is bounded by the  $1/d$ -th harmonic mean of  $\|g^i\|_\infty$ .*

$$C_\mu = \left[ \frac{1}{N} \sum_{i=1}^N \frac{1}{\|g^i\|_\infty^{1/d}} \right]^{-d}.$$

The constant  $C_\mu$  depends only on the measures  $(\mu^1, \dots, \mu^N)$ , and is finite as long as one  $\mu^i$  has a bounded density, since  $C_\mu \leq N^d \|g^i\|_\infty$  for any  $i$ .

*Proof.* Let  $h_i$  be the density of  $\gamma_i$ . By the change of variables formula, for  $\gamma_0$ -almost any  $x$

$$h_1(\mathbf{t}_{\gamma_0}^{\gamma_1}(x)) = \frac{h_0(x)}{\det \nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)}; \quad g^i(\mathbf{t}_{\gamma_0}^{\mu^i}(x)) = \frac{h_0(x)}{\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)}, \quad \text{when } g^i \text{ exists.}$$

(Convex functions are twice differentiable almost surely (Villani [125, Theorem 14.25]), hence these gradients are well-defined  $\gamma_0$ -almost surely.) We seek a lower bound on the determinant of  $\nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)$ , which by definition equals

$$N^{-d} \det \sum_{i=1}^N \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x).$$

Such a bound is provided by the Brunn–Minkowski inequality (Stein and Shakarchi [121, Section 1.5]) for symmetric positive semidefinite matrices

$$[\det(A+B)]^{1/d} \geq [\det A]^{1/d} + [\det B]^{1/d},$$

which, applied inductively, yields

$$[\det \nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)]^{1/d} \geq \frac{1}{N} \sum_{i=1}^N [\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)]^{1/d}.$$

From this, we obtain an upper bound for  $h_1$ :

$$\frac{1}{h_1^{1/d}(\mathbf{t}_{\gamma_0}^{\gamma_1}(x))} = \frac{\det^{1/d} \sum_{i=1}^N \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)}{N h_0^{1/d}(x)} \geq \frac{1}{N} \sum_{i=1}^N \frac{1}{[g^i(\mathbf{t}_{\gamma_0}^{\mu^i}(x))]^{1/d}} \geq \frac{1}{N} \sum_{i=1}^N \frac{1}{\|g^i\|_\infty^{1/d}} = C_\mu^{-1/d}.$$

Let  $\Sigma$  be the set of points where this inequality holds, then  $\gamma_0(\Sigma) = 1$ . Hence

$$\gamma_1(\mathbf{t}_{\gamma_0}^{\gamma_1}(\Sigma)) = \gamma_0[(\mathbf{t}_{\gamma_0}^{\gamma_1})^{-1}(\mathbf{t}_{\gamma_0}^{\gamma_1}(\Sigma))] \geq \gamma_0(\Sigma) = 1.$$

Thus,  $\gamma_1$ -almost surely and for all  $i$ ,

$$h_1(y) \leq C_\mu.$$

The third statement (continuity of  $\mathcal{A}$ ) is much more subtle to establish, and its rather lengthy proof is given next. In view of Proposition 5.3.6, the uniform bound on the densities is not a hindrance for the proof of convergence of Algorithm 1.

**Proposition 5.3.7** *Let  $(\gamma_n)$  be a sequence of absolutely continuous measures with uniformly bounded densities, suppose that  $W_2(\gamma_n, \gamma) \rightarrow 0$ , and let*

$$\eta_j = \left( \mathbf{t}_{\gamma_j}^{\mu^1}, \dots, \mathbf{t}_{\gamma_j}^{\mu^n}, \mathbf{i} \right) \# \gamma_j, \quad \eta = \left( \mathbf{t}_{\gamma}^{\mu^1}, \dots, \mathbf{t}_{\gamma}^{\mu^n}, \mathbf{i} \right) \# \gamma.$$

*Then  $\eta_j \rightarrow \eta$  in  $\mathcal{W}_2([\mathbb{R}^d]^{N+1})$ .*

*Proof.* As has been established in the discussion before Proposition 5.3.6, the limit  $\gamma$  must be absolutely continuous, so  $\eta$  is well-defined.

In view of Theorem 2.2.1, it suffices to show that if  $h : (\mathbb{R}^d)^{N+1} \rightarrow \mathbb{R}$  is any continuous nonnegative function such that



$$|h(t_1, \dots, t_N, y)| \leq \frac{2}{N} \sum_{i=1}^N \|t_i\|^2 + 2\|y\|^2,$$

then

$$\int_{\mathbb{R}^d} g_n \, d\gamma_n = \int_{(\mathbb{R}^d)^{N+1}} h \, d\eta_n \rightarrow \int_{(\mathbb{R}^d)^{N+1}} h \, d\eta = \int_{\mathbb{R}^d} g \, d\gamma, \quad g_n(x) = h(\mathbf{t}_{\gamma_j}^{\mu^1}(x), \dots, \mathbf{t}_{\gamma_j}^{\mu^n}(x), x),$$

and  $g$  defined analogously. The proof, given in full detail on page 124 of the supplement, is sketched here.

**Step 1: Truncation.** Since  $\gamma_n$  converge in the Wasserstein space, they satisfy the uniform integrability (2.4) and absolute continuity (2.7) by Theorem 2.2.1. Consequently,  $g_{n,R} = \min(g_n, 4R)$  is uniformly close to  $g_n$ :

$$\sup_n \int [g_n(x) - g_{n,R}(x)] \, d\gamma_n(x) \rightarrow 0, \quad R \rightarrow \infty.$$

We may thus replace  $g_n$  by a bounded version  $g_{n,R}$ .

**Step 2: Convergence of  $g_n$  to  $g$ .** By Proposition 1.7.11, the optimal maps  $\mathbf{t}_{\gamma_n}^{\mu^i}$  converge to  $\mathbf{t}_{\gamma}^{\mu^i}$  and (since  $h$  is continuous),  $g_n \rightarrow g$  uniformly on “nice” sets  $\Omega \subseteq E = \text{supp } \gamma$ . Write

$$\int g_{n,R} \, d\gamma_n - \int g_R \, d\gamma = \int g_R \, d(\gamma_n - \gamma) + \int_{\Omega} (g_{n,R} - g_R) \, d\gamma_n + \int_{\mathbb{R}^d \setminus \Omega} (g_{n,R} - g_R) \, d\gamma_n.$$

**Step 3: Bounding the first two integrals.** The first integral vanishes as  $n \rightarrow \infty$ , by the portmanteau Lemma 1.7.1, and the second by uniform convergence.

**Step 4: Bounding the third integral.** The integrand is bounded by  $8R$ , so it suffices to bound the measures of  $\mathbb{R}^d \setminus \Omega$ . This is a bit technical, and uses the uniform density bound on  $(\gamma_n)$  and the portmanteau lemma.

**Corollary 5.3.8 (Continuity of  $\mathcal{A}$ )** *If  $W_2(\gamma_n, \gamma) \rightarrow 0$  and  $\gamma_n$  have uniformly bounded densities, then  $\mathcal{A}(\gamma_n) \rightarrow \mathcal{A}(\gamma)$ .*

*Proof.* Choose  $h$  in the proof of Proposition 5.3.7 to depend only on  $y$ .

*Proof (Proof of Corollary 5.3.4).* Choose  $h$  in the proof of Proposition 5.3.7 to depend only on  $t_1, \dots, t_n$ .

*Proof (Proof of Theorem 5.3.3).* Let  $E = \text{supp } \bar{\mu}$  and set  $A^i = E^{\text{den}} \cap \{x : \mathbf{t}_{\bar{\mu}}^{\mu^i}(x) \text{ is univalued}\}$ . As  $\bar{\mu}$  is absolutely continuous,  $\bar{\mu}(A^i) = 1$ , and the same is true for  $A = \bigcap_{i=1}^N A^i$ . The first assertion then follows from Proposition 1.7.11.

The second statement is proven similarly. Let  $E^i = \text{supp } \mu^i$  and notice that by absolute continuity the  $B^i = (E^i)^{\text{den}} \cap \{x : \mathbf{t}_{\mu^i}^{\mu^i}(x) \text{ is univalued}\}$  has measure 1 with respect to  $\mu^i$ . Apply Proposition 1.7.11. If in addition  $E^1 = \dots = E^N$ , then  $\mu^i(B) = 1$  for  $B = \bigcap B^i$ .

## 5.4 Illustrative Examples

As an illustration, we implement Algorithm 1 in several scenarios for which pairwise optimal maps can be calculated explicitly at every iteration, allowing for fast computation without error propagation. In each case, we give some theory first, describing how the optimal maps are calculated, and then implement Algorithm 1 on simulated examples.

### 5.4.1 Gaussian Measures

No example illustrates the use of Algorithm 1 better than the Gaussian case. This is so because optimal maps between centred nondegenerate Gaussian measures with covariances  $A$  and  $B$  have the explicit form (see Sect. 1.6.3)

$$\mathbf{t}_A^B(x) = A^{-1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{-1/2}x, \quad x \in \mathbb{R}^d,$$

with the obvious slight abuse of notation. In contrast, the Fréchet mean of a collection of Gaussian measures (one of which nonsingular) does not admit a closed-form formula and is only known to be a Gaussian measure whose covariance matrix  $\Gamma$  is the unique invertible root of the matrix equation

$$\Gamma = \frac{1}{N} \sum_{i=1}^N \left[ \Gamma^{1/2} S_i \Gamma^{1/2} \right]^{1/2}, \quad (5.4)$$

where  $S_i$  is the covariance matrix of  $\mu^i$ .

Given the formula for  $\mathbf{t}_A^B$ , application of Algorithm 1 to Gaussian measures is straightforward. The next result shows that, in the Gaussian case, the iterates must converge to the unique Fréchet mean, and that (5.4) can be derived from the characterisation of Karcher means.

**Theorem 5.4.1 (Convergence in Gaussian Case)** *Let  $\mu^1, \dots, \mu^N$  be Gaussian measures with zero means and covariance matrices  $S_i$  with  $S_1$  nonsingular, and let the initial point  $\gamma_0$  be  $\mathcal{N}(0, \Gamma_0)$  with  $\Gamma_0$  nonsingular. Then the sequence of iterates generated by Algorithm 1 converges to the unique Fréchet mean of  $(\mu^1, \dots, \mu^N)$ .*

*Proof.* Since the optimal maps are linear, so is their mean and therefore  $\gamma_k$  is a Gaussian measure for all  $k$ , say  $\mathcal{N}(0, \Gamma_k)$  with  $\Gamma_k$  nonsingular. Any limit point of  $\gamma$  is a Karcher mean by Theorem 5.3.1. If we knew that  $\gamma$  itself were Gaussian, then it actually must be the Fréchet mean because  $N^{-1} \sum \mathbf{t}_\gamma^{\mu^i}$  equals the identity everywhere on  $\mathbb{R}^d$  (see the discussion before Theorem 3.1.15).

Let us show that every limit point  $\gamma$  is indeed Gaussian. It suffices to prove that  $(\Gamma_k)$  is a bounded sequence, because if  $\Gamma_k \rightarrow \Gamma$ , then  $\mathcal{N}(0, \Gamma_k) \rightarrow \mathcal{N}(0, \Gamma)$  weakly, as can be seen from either Lehmann–Scheffé’s theorem (the densities converge) or Lévy’s continuity theorem (the characteristic functions converge).

To see that  $(\Gamma_k)$  is bounded, observe first that for any centred (Gaussian or not) measure  $\mu$  with covariance matrix  $S$ ,

$$W_2^2(\mu, \delta_0) = \text{tr}[S],$$

where  $\delta_0$  is a Dirac mass at the origin. (This follows from the spectral decomposition of  $S$ .) Therefore

$$0 \leq \text{tr}[\Gamma_k] = W_2^2(\gamma_k, \delta_0)$$

is bounded uniformly, because  $\{\gamma_k\}$  stays in a Wasserstein-compact set by Lemma 5.3.5. If we define  $C = \sup_k \text{tr}[\Gamma_k] < \infty$ , then all the diagonal elements of  $\Gamma_k$  are bounded uniformly. When  $A$  is symmetric and positive semidefinite,  $2|A_{ij}| \leq A_{ii} + A_{jj}$ . Consequently, all the entries of  $\Gamma_k$  are bounded uniformly by  $C$ , which means that  $(\Gamma_k)$  is a bounded sequence.

From the formula for the optimal maps, we see that if  $\Gamma$  is the covariance of the Fréchet mean, then

$$I = \frac{1}{N} \sum_{i=1}^N \Gamma^{-1/2} \left[ \Gamma^{1/2} S_i \Gamma^{1/2} \right]^{1/2} \Gamma^{-1/2}$$

and we recover the fixed point equation (5.4).

If the means are nonzero, then the optimal maps are affine and the same result applies; the Fréchet mean is still a Gaussian measure with covariance matrix  $\Gamma$  and mean that equals the average of the means of  $\mu^i$ ,  $i = 1, \dots, N$ .

Figure 5.1 shows density plots of  $N = 4$  centred Gaussian measures on  $\mathbb{R}^2$  with covariances  $S_i \sim \text{Wishart}(I_2, 2)$ , and Fig. 5.2 shows the density of the resulting Fréchet mean. In this particular example, the algorithm needed 11 iterations starting from the identity matrix. The corresponding optimal maps are displayed in Fig. 5.3. It is apparent from the figure that these maps are linear, and after a more careful reflection one can be convinced that their average is the identity. The four plots in the figure are remarkably different, in accordance with the measures themselves having widely varying condition numbers and orientations;  $\mu^3$  and more so  $\mu^4$  are very concentrated, so the optimal maps “sweep” the mass towards zero. In contrast, the optimal maps to  $\mu^1$  and  $\mu^2$  spread the mass out away from the origin.

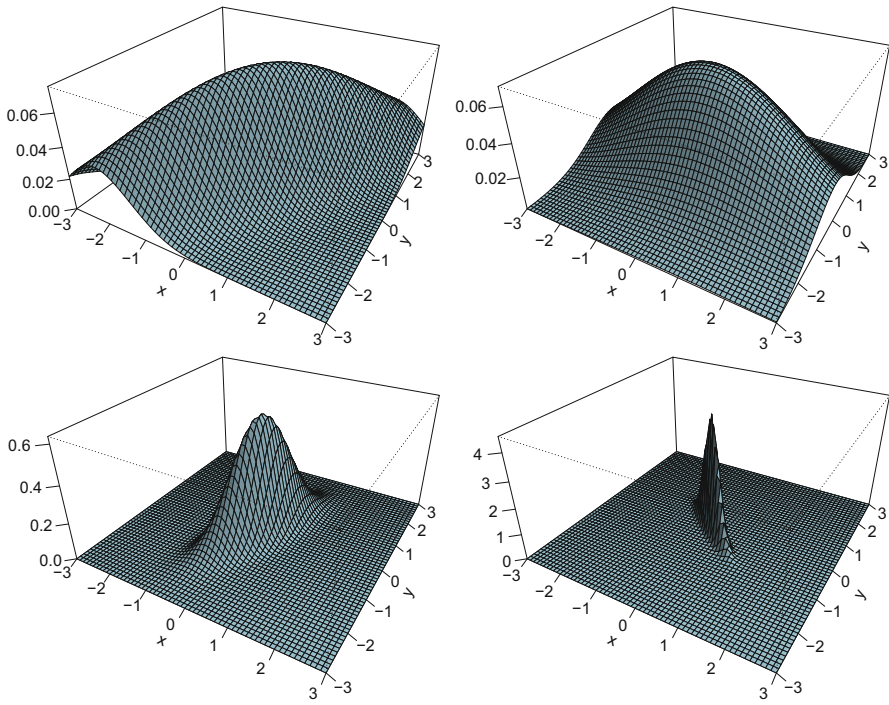


Fig. 5.1: Density plot of four Gaussian measures in  $\mathbb{R}^2$ .

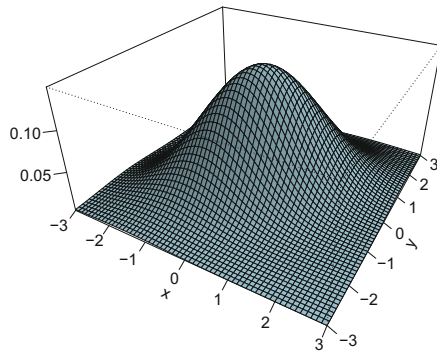


Fig. 5.2: Density plot of the Fréchet mean of the measures in Fig. 5.1

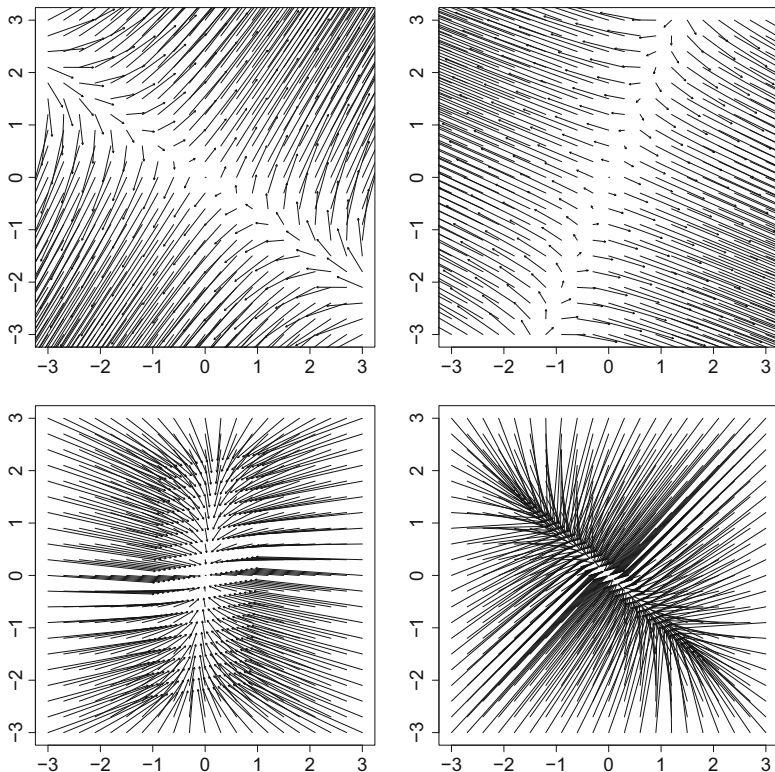


Fig. 5.3: Gaussian example: vector fields depicting the optimal maps  $x \mapsto \mathbf{t}_{\mu}^{\mu^i}(x)$  from the Fréchet mean  $\bar{\mu}$  of Fig. 5.2 to the four measures  $\{\mu^i\}$  of Fig. 5.1. The order corresponds to that of Fig. 5.1

### 5.4.2 Compatible Measures

We next discuss the behaviour of the algorithm when the measures are compatible. Recall that a collection  $\mathcal{C} \subseteq \mathcal{W}_2(\mathcal{X})$  is *compatible* if for all  $\gamma, \rho, \mu \in \mathcal{C}$ ,  $\mathbf{t}_{\mu}^{\nu} \circ \mathbf{t}_{\gamma}^{\mu} = \mathbf{t}_{\gamma}^{\nu}$  in  $L_2(\gamma)$  (Definition 2.3.1). Boissard et al. [28] showed that when this condition holds, the Fréchet mean of  $(\mu^1, \dots, \mu^N)$  can be found by simple computations involving the *iterated barycentre*. We again denote by  $\gamma_0$  the initial point of Algorithm 1, which can be any absolutely continuous measure.

**Lemma 5.4.2 (Compatibility and Convergence)** *If  $\gamma_0 \cup \{\mu^i\}$  is compatible, then Algorithm 1 converges to the Fréchet mean of  $(\mu^i)$  after a single step.*

*Proof.* By definition, the next iterate is

$$\gamma_1 = \left[ \frac{1}{N} \sum_{i=1}^N \mathbf{t}_{\gamma_0}^{\mu^i} \right] \# \gamma_0,$$

which is the Fréchet mean by Theorem 3.1.9.

In this case, Algorithm 1 requires the calculation of  $N$  pairwise optimal maps, and this can be reduced to  $N - 1$  if the initial point is chosen to be  $\mu^1$ . This is the same computational complexity as the calculation of the iterated barycentre proposed in [28].

When the measures have a common copula, finding the optimal maps reduces to finding the optimal maps between the one-dimensional marginals (see Lemma 2.3.3) and this can be done using quantile functions as described in Sect. 1.5. The marginal Fréchet means are then plugged into the common copula to yield the joint Fréchet mean. We next illustrate Algorithm 1 in three such scenarios.

#### 5.4.2.1 The One-Dimensional Case

When the measures are supported on the real line, there is no need to use the algorithm since the Fréchet mean admits a closed-form expression in terms of quantile functions (see Sect. 3.1.4). We nevertheless discuss this case briefly because we build upon this construction in subsequent examples. Given that  $\mathbf{t}_{\mu}^{\nu} = F_{\nu}^{-1} \circ F_{\mu}$ , we may apply Algorithm 1 starting from one of these measures (or any other measure). Figure 5.4 plots  $N = 4$  univariate densities and the Fréchet mean yielded by the algorithm in two different scenarios. At the left, the densities were generated as

$$f^i(x) = \frac{1}{2} \phi \left( \frac{x - m_1^i}{\sigma_1^i} \right) + \frac{1}{2} \phi \left( \frac{x - m_2^i}{\sigma_2^i} \right), \quad (5.5)$$

with  $\phi$  the standard normal density, and the parameters generated independently as

$$m_1^i \sim U[-13, -3], \quad m_2^i \sim U[3, 13], \quad \sigma_1^i, \sigma_2^i \sim \text{Gamma}(4, 4).$$

At the right of Fig. 5.4, we used a mixture of a shifted gamma and a Gaussian:

$$f^i(x) = \frac{3}{5} \frac{\beta_i^3}{\Gamma(3)} (x - m_3^i)^2 e^{-\beta_i(x - m_3^i)} + \frac{2}{5} \phi(x - m_4^i), \quad (5.6)$$

with

$$\beta^i \sim \text{Gamma}(4, 1), \quad m_3^i \sim U[1, 4], \quad m_4^i \sim U[-4, -1].$$

The resulting Fréchet mean density for both settings is shown in thick light blue, and can be seen to capture the bimodal nature of the data. Even though the Fréchet mean of Gaussian mixtures is not a Gaussian mixture itself, it is approximately so, provided that the peaks are separated enough. Figure 5.5 shows the optimal maps pushing the Fréchet mean  $\bar{\mu}$  to the measures  $\mu^1, \dots, \mu^N$  in each case. If one ignores the “middle part” of the  $x$  axis, the maps appear (approximately) affine for small

values of  $x$  and for large values of  $x$ , indicating how the peaks are shifted. In the middle region, the maps need to “bridge the gap” between the different slopes and intercepts of these affine maps.

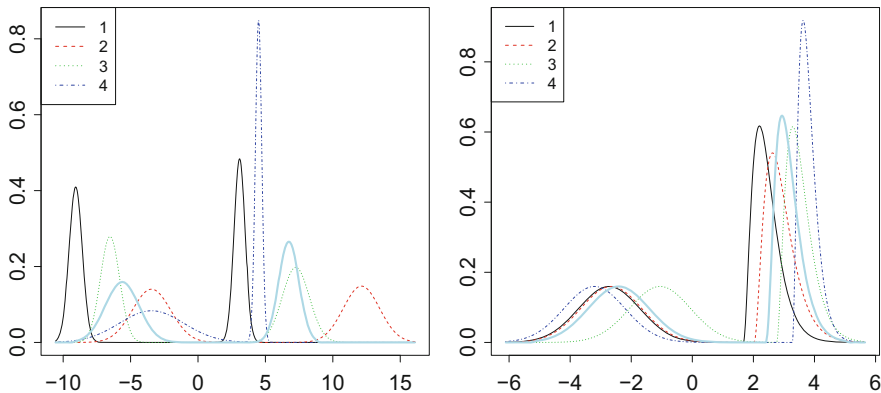


Fig. 5.4: Densities of a bimodal Gaussian mixture (left) and a mixture of a Gaussian with a gamma (right), with the Fréchet mean density in light blue

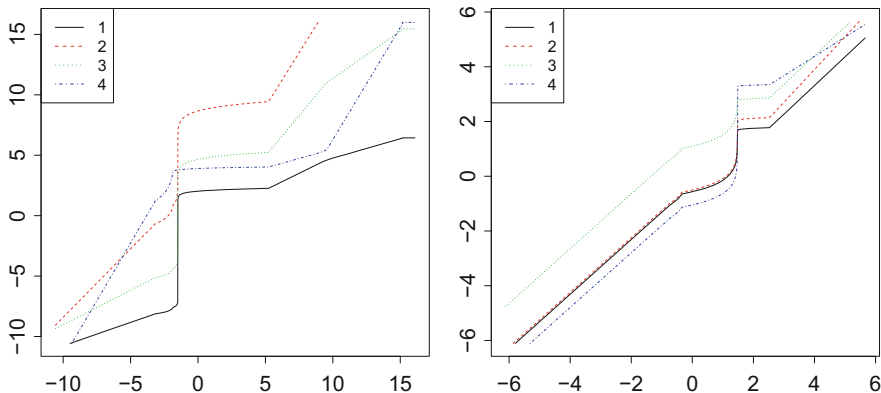


Fig. 5.5: Optimal maps  $t_{\bar{\mu}}^{\mu^i}$  from the Fréchet mean  $\bar{\mu}$  to the four measures  $\{\mu^i\}$  in Fig. 5.4. The left plot corresponds to the bimodal Gaussian mixture, and the right plot to the Gaussian/gamma mixture

**5.4.2.2 Independence**

We next take measures  $\mu^i$  on  $\mathbb{R}^2$ , having independent marginal densities  $f_X^i$  as in (5.5), and  $f_Y^i$  as in (5.6). Figure 5.6 shows the density plot of  $N = 4$  such measures, constructed as the product of the measures from Fig. 5.4. One can distinguish the

independence by the “parallel” structure of the figures: for every pair  $(y_1, y_2)$ , the ratio  $g(x, y_1)/g(x, y_2)$  does not depend on  $x$  (and vice versa, interchanging  $x$  and  $y$ ). Figure 5.7 plots the density of the resulting Fréchet mean. We observe that the Fréchet mean captures the four peaks and their location. Furthermore, the parallel nature of the figure is preserved in the Fréchet mean. Indeed, by Lemma 3.1.11 the Fréchet mean is a product measure. The optimal maps, in Fig. 5.10, are the same as in the next example, and will be discussed there.

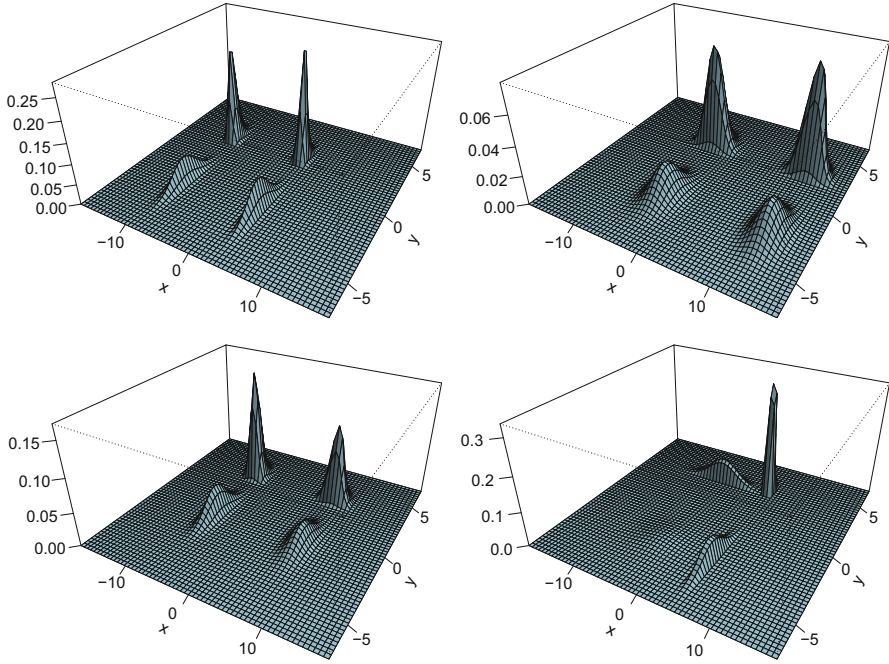


Fig. 5.6: Density plots of the four product measures of the measures in Fig. 5.4

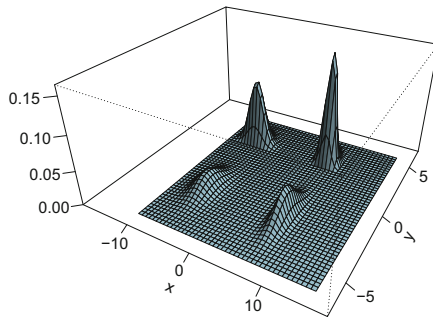


Fig. 5.7: Density plot of the Fréchet mean of the measures in Fig. 5.6



### 5.4.2.3 Common Copula

Let  $\mu^i$  be a measure on  $\mathbb{R}^2$  with density

$$g^i(x, y) = c(F_X^i(x), F_Y^i(y))f_X^i(x)f_Y^i(y),$$

where  $f_X^i$  and  $f_Y^i$  are random densities on the real line with distribution functions  $F_X^i$  and  $F_Y^i$ , and  $c$  is a copula density. Figure 5.8 shows the density plot of  $N = 4$  such measures, with  $f_X^i$  generated as in (5.5),  $f_Y^i$  as in (5.6), and  $c$  is the Frank(-8) copula density, while Fig. 5.9 plots the density of the Fréchet mean obtained. (For ease of comparison we use the same realisations of the densities that appear in Fig. 5.4.) The Fréchet mean can be seen to preserve the shape of the density, having four clearly distinguished peaks. Figure 5.10, depicting the resulting optimal maps, allows for a clearer interpretation: for instance, the leftmost plot (in black) shows more clearly that the map splits the mass around  $x = -2$  to a much wider interval; and conversely a very large amount of mass is sent to  $x \approx 2$ . This rather extreme behaviour matches the peak of the density of  $\mu^1$  located at  $x = 2$ .

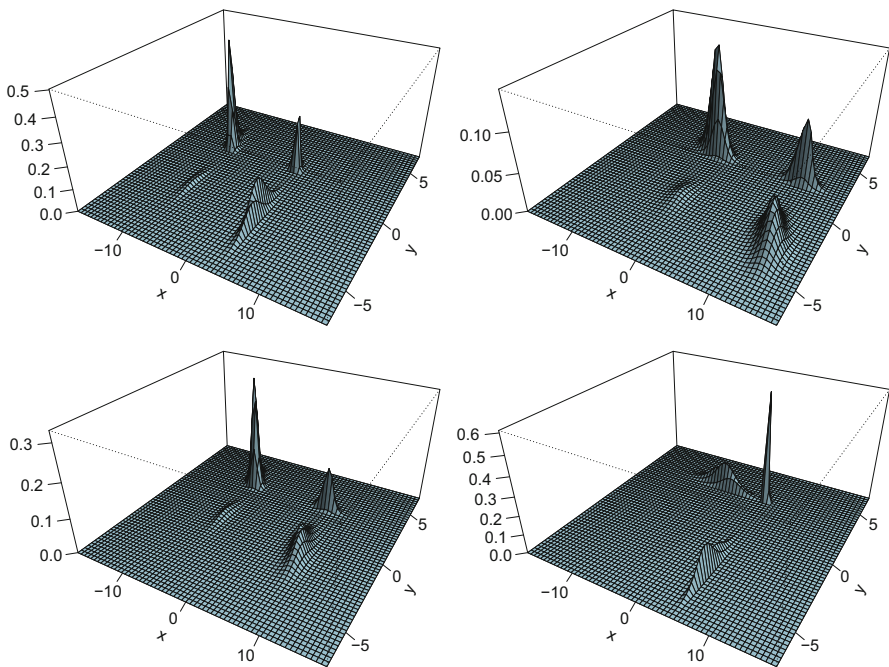


Fig. 5.8: Density plots of four measures in  $\mathbb{R}^2$  with Frank copula of parameter  $-8$

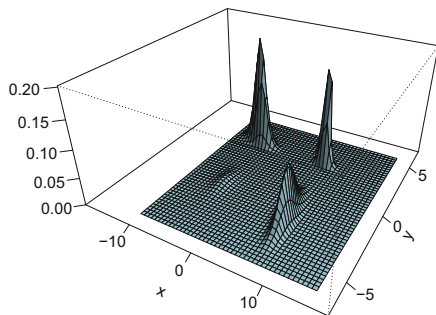


Fig. 5.9: Density plot of the Fréchet mean of the measures in Fig. 5.8

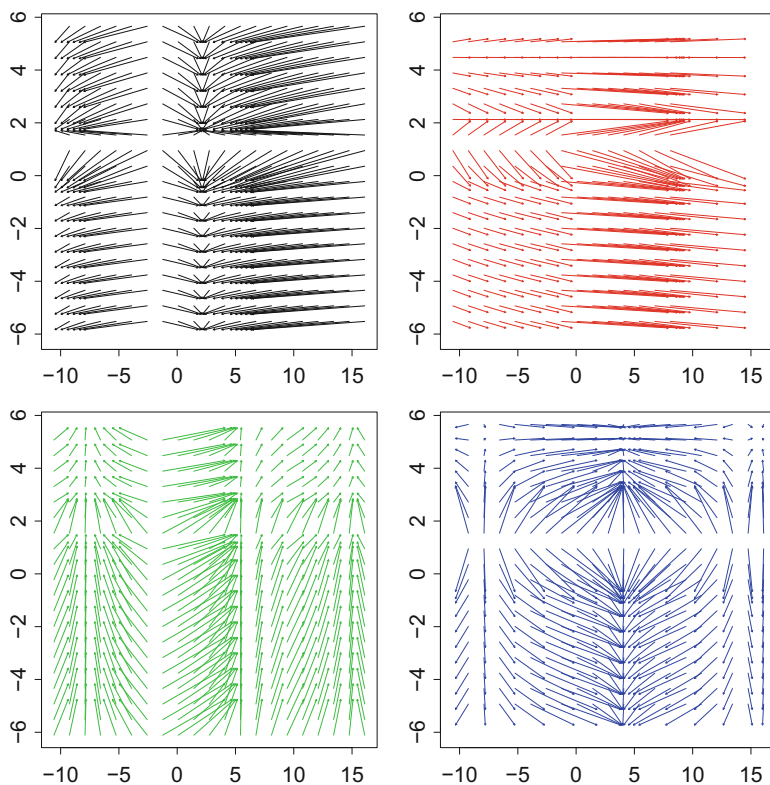


Fig. 5.10: Frank copula example: vector fields of the optimal maps  $\mathbf{t}_{\bar{\mu}}^{\mu^i}$  from the Fréchet mean  $\bar{\mu}$  of Fig. 5.9 to the four measures  $\{\mu^i\}$  of Fig. 5.8. The colours match those of Fig. 5.4

### 5.4.3 Partially Gaussian Trivariate Measures

We now apply Algorithm 1 in a situation that entangles two of the previous settings. Let  $U$  be a fixed  $3 \times 3$  real orthogonal matrix with columns  $U_1, U_2, U_3$  and let  $\mu^i$  have density

$$g^i(y_1, y_2, y_3) = g^i(y) = f^i(U_3^t y) \frac{1}{2\pi\sqrt{\det S^i}} \exp \left[ -\frac{(U_1^t y, U_2^t y)(S^i)^{-1} \begin{pmatrix} U_1^t y \\ U_2^t y \end{pmatrix}}{2} \right],$$

with  $f^i$  bounded density on the real line and  $S^i \in \mathbb{R}^{2 \times 2}$  positive definite. We simulated  $N = 4$  such densities with  $f^i$  as in (5.5) and  $S^i \sim \text{Wishart}(I_2, 2)$ . We apply Algorithm 1 to this collection of measures and find their Fréchet mean (see the end of this subsection for precise details on how the optimal maps were calculated). Figure 5.11 shows level set of the resulting densities for some specific values. The bimodal nature of  $f^i$  implies that for most values of  $a$ ,  $\{x : f^i(x) = a\}$  has four elements. Hence, the level sets in the figures are unions of four separate parts, with each peak of  $f^i$  contributing two parts that form together the boundary of an ellipsoid in  $\mathbb{R}^3$  (see Fig. 5.12). The principal axes of these ellipsoids and their position in  $\mathbb{R}^3$  differ between the measures, but the Fréchet mean can be viewed as an average of those in some sense.

In terms of orientation (principal axes) of the ellipsoids, the Fréchet mean is most similar to  $\mu^1$  and  $\mu^2$ , whose orientations are similar to one another.

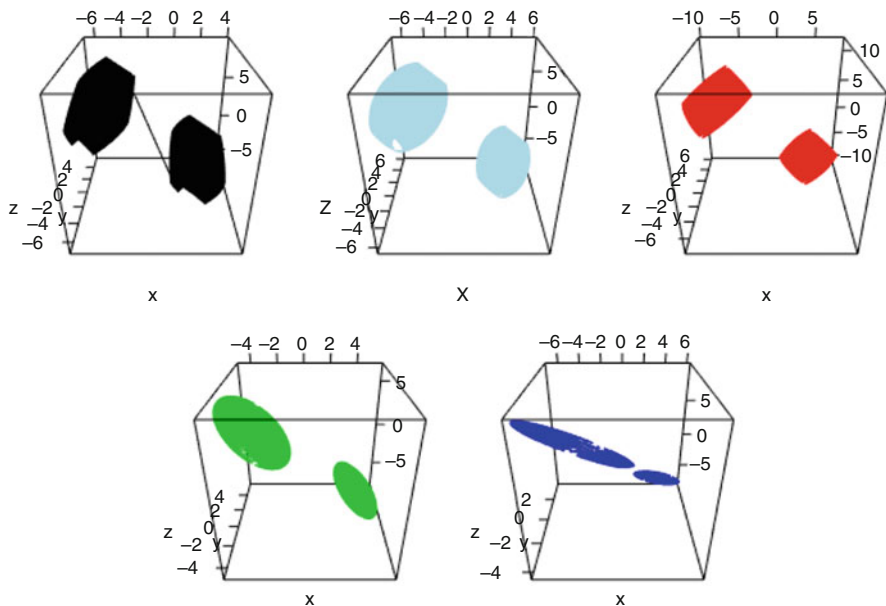


Fig. 5.11: The set  $\{v \in \mathbb{R}^3 : g^i(v) = 0.0003\}$  for  $i = 1$  (black), the Fréchet mean (light blue),  $i = 2, 3, 4$  in red, green, and dark blue, respectively

Let us now see how the optimal maps are calculated. If  $Y = (y_1, y_2, y_3) \sim \mu^i$ , then the random vector  $(x_1, x_2, x_3) = X = U^{-1}Y$  has joint density

$$f^i(x_3) \exp \left[ -\frac{(x_1, x_2)(\Sigma^i)^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}{2} \right] \frac{1}{2\pi\sqrt{\det \Sigma^i}},$$

so the probability law of  $X$  is  $\rho^i \otimes \nu^i$  with  $\rho^i$  centred Gaussian with covariance matrix  $\Sigma^i$  and  $\nu^i$  having density  $f^i$  on  $\mathbb{R}$ . By Lemma 3.1.11, the Fréchet mean of  $(U^{-1}\#\mu^i)$  is the product measure of that of  $(\rho^i)$  and that of  $(\nu^i)$ ; by Lemma 3.1.12, the Fréchet mean of  $(\mu^i)$  is therefore

$$U\#(\mathcal{N}(0, \Sigma) \otimes f), \quad f = F', \quad F^{-1}(q) = \frac{1}{N} \sum_{i=1}^N F_i^{-1}(q), \quad F_i(x) = \int_{-\infty}^x f^i(s) ds,$$

where  $\Sigma$  is the Fréchet–Wasserstein mean of  $\Sigma_1, \dots, \Sigma_N$ .

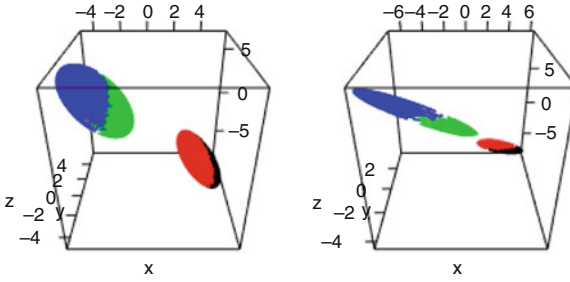


Fig. 5.12: The set  $\{v \in \mathbb{R}^3 : g^i(v) = 0.0003\}$  for  $i = 3$  (left) and  $i = 4$  (right), with each of the four different inverses of the bimodal density  $f^i$  corresponding to a colour

Starting at an initial point  $\gamma_0 = U\#(\mathcal{N}(0, \Sigma_0) \otimes \nu_0)$ , with  $\nu_0$  having continuous distribution  $F_{\nu_0}$ , the optimal maps are  $U \circ \mathbf{t}_0^i \circ U^{-1} = \nabla(\varphi_0^i \circ U^{-1})$  with

$$\mathbf{t}_0^i(x_1, x_2, x_3) = \begin{pmatrix} \mathbf{t}_{\Sigma_0}^{\Sigma^i}(x_1, x_2) \\ F_j^{-1} \circ F_{\nu_0}(x_3) \end{pmatrix}$$

the gradients of the convex function

$$\varphi_0^i(x_1, x_2, x_3) = (x_1, x_2) \mathbf{t}_{\gamma_0}^{\Sigma^i} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \int_0^{x_3} F_j^{-1}(F_{\nu_0}(s)) ds,$$

where we identify  $\mathbf{t}_{\gamma_0}^{\Sigma^i}$  with the positive definite matrix  $(\Sigma^i)^{1/2}[(\Sigma^i)^{1/2}\Sigma_0(\Sigma^i)^{1/2}]^{-1/2}(\Sigma^i)^{1/2}$  that pushes forward  $\mathcal{N}(0, \Sigma_0)$  to  $\mathcal{N}(0, \Sigma^i)$ . Due to the one-dimensionality, the algorithm finds the third component of the rotated measures after one step, but the convergence of the Gaussian component requires further iterations.

### 5.5 Population Version of Algorithm 1

Let  $\Lambda \in \mathscr{W}_2(\mathbb{R}^d)$  be a random measure with finite Fréchet functional. The population version of (5.1) is

$$q = \mathbb{P}(\Lambda \text{ absolutely continuous with density bounded by } M) > 0 \quad \text{for some } M < \infty, \tag{5.7}$$

which we assume henceforth. This condition is satisfied if and only if

$$\mathbb{P}(\Lambda \text{ absolutely continuous with bounded density}) > 0.$$

These probabilities are well-defined because the set

$$\mathscr{W}_2(\mathbb{R}^d; M) = \{\mu \in \mathscr{W}_2(\mathbb{R}^d) : \mu \text{ absolutely continuous with density bounded by } M\}$$

is weakly closed (see the paragraph before Proposition 5.3.6), hence a Borel set of  $\mathscr{W}_2(\mathbb{R}^d)$ .

In light of Theorem 3.2.13, we can define a population version of Algorithm 1 with the iteration function

$$\mathscr{A}(\gamma) = \mathbb{E} \mathbf{t}_\gamma^\Lambda, \quad \gamma \in \mathscr{W}_2(\mathbb{R}^d) \text{ absolutely continuous.}$$

The (Bochner) expectation is well-defined in  $\mathscr{L}_2(\gamma)$  because the random map  $\mathbf{t}_\gamma^\Lambda$  is measurable (Lemma 2.4.6). Since  $\mathscr{L}_2(\gamma)$  is a Hilbert space, the law of large numbers applies there, and results for the empirical version carry over to the population version by means of approximations. In particular:

**Lemma 5.5.1** *Any descent iterate  $\gamma$  has density bounded by  $q^{-d}M$ , where  $q$  and  $M$  are as in (5.7).*

*Proof.* The result is true in the empirical case, by Proposition 5.3.6. The key point (observed by Pass [102, Subsection 3.3]) is that the number of measures does not appear in the bound  $q^{-d}M$ .

Let  $\Lambda_1, \dots$  be a sample from  $\Lambda$  and let  $q_n$  be the proportion of measures in  $(\Lambda_1, \dots, \Lambda_n)$  that have density bounded by  $M$ . Then both  $n^{-1} \sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i} \rightarrow \mathbb{E} \mathbf{t}_\gamma^\Lambda$  and  $q_n \rightarrow q$  almost surely by the law of large numbers. Pick any  $\omega$  in the probability space for which this happens and notice that (invoking Lemma 2.4.5)

$$\mathscr{A}(\gamma) = \left[ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i} \right] \# \gamma = \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i} \right] \# \gamma.$$

Let  $\lambda_n$  denote the measure in the last limit. By Proposition 5.3.6, its density is bounded by  $q_n^{-d}M \rightarrow q^{-d}M$  almost surely, so for any  $C > q^{-d}M$  and  $n$  large,  $\lambda_n$  has density bounded by  $C$ . By the portmanteau Lemma 1.7.1, so does  $\lim \lambda_n = [\mathbb{E} \mathbf{t}_\gamma^\Lambda] \# \gamma$ . Now let  $C \searrow q^{-d}M$ .

Though it follows that every Karcher mean of  $\Lambda$  has a bounded density, we cannot yet conclude that the same bound holds for the Fréchet mean, because we need an a-priori knowledge that the latter is absolutely continuous. This again can be achieved by approximations:

**Theorem 5.5.2 (Bounded Density for Population Fréchet Mean)** *Let  $\Lambda \in \mathcal{W}_2(\mathbb{R}^d)$  be a random measure with finite Fréchet functional. If  $\Lambda$  has a bounded density with positive probability, then the Fréchet mean of  $\Lambda$  is absolutely continuous with a bounded density.*

*Proof.* Let  $q$  and  $M$  be as in (5.7),  $\Lambda_1, \dots$  be a sample from  $\Lambda$ , and  $q_n$  the proportion of  $(\Lambda_i)_{i \leq n}$  with density bounded by  $M$ . The empirical Fréchet mean  $\lambda_n$  of the sample  $(\Lambda_1, \dots, \Lambda_n)$  has a density bounded by  $q_n^{-d}M$ . The Fréchet mean  $\lambda$  of  $\Lambda$  is unique by Proposition 3.2.7, and consequently  $\lambda_n \rightarrow \lambda$  in  $\mathcal{W}_2(\mathbb{R}^d)$  by the law of large numbers (Corollary 3.2.10). For any  $C > \limsup q_n^{-d}M$ , the density of  $\lambda$  is bounded by  $C$  by the portmanteau Lemma 1.7.1, and the limsup is  $q^{-d}M$  almost surely. Thus, the density is bounded by  $q^{-d}M$ .

In the same way, one shows the population version of Theorem 3.1.9:

**Theorem 5.5.3 (Fréchet Mean of Compatible Measures)** *Let  $\Lambda : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathcal{W}_2(\mathbb{R}^d)$  be a random measure with finite Fréchet functional, and suppose that with positive probability  $\Lambda$  is absolutely continuous and has bounded density. If the collection  $\{\gamma\} \cup \Lambda(\Omega)$  is compatible and  $\gamma$  is absolutely continuous, then  $[\mathbf{Et}_\gamma^\Lambda] \# \gamma$  is the Fréchet mean of  $\Lambda$ .*

It is of course sufficient that  $\{\gamma\} \cup \Lambda(\Omega \setminus \mathcal{N})$  be compatible for some null set  $\mathcal{N} \subset \Omega$ .

## 5.6 Bibliographical Notes

The algorithm outlined in this chapter was suggested independently in this steepest descent form by Zemel and Panaretos [134] and in the form a fixed point equation iteration by Álvarez-Esteban et al. [9]. These two papers provide different alternative proofs of Theorem 5.3.1. The exposition here is based on [134]. Although longer and more technical than the one in [9], the formalism in [134] is amenable to directly treating the optimal maps (Theorem 5.3.3) and the multicouplings (Corollary 5.3.4). On the flip side, it is noteworthy that the proof of the Gaussian case (Theorem 5.4.1) given in [9] is more explicit and quantitative; for instance, it shows the additional property that the traces of the matrix iterates are monotonically increasing.

Developing numerical schemes for computing Fréchet means in  $\mathcal{W}_2(\mathbb{R}^d)$  is a very active area of research, and readers are referred to the recent monograph of Peyré and Cuturi [103, Section 9.2] for a survey.

In recent work, Backhoff-Varaguas et al. [15] propose a stochastic steepest descent for finding Karcher means of a population Fréchet functional associated with a random measure  $\Lambda$ . At iterate  $j$ , one replaces  $\gamma_j$  by

$$[t_j \mathbf{t}_{\gamma_j}^{\mu_j} + (1 - t_j) \mathbf{i}] \# \gamma_j, \quad \mu_j \sim \Lambda.$$

The analogue of Theorem 5.3.1 holds under further conditions.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

