

Chapter 3

Fréchet Means in the Wasserstein Space \mathcal{W}_2



If H is a Hilbert space (or a closed convex subspace thereof) and $x_1, \dots, x_N \in H$, then the empirical mean $\bar{x}_N = N^{-1} \sum x_i$ is the unique element of H that minimises the sum of squared distances from the x_i 's.¹ That is, if we define

$$F(\theta) = \sum_{i=1}^N \|\theta - x_i\|^2, \quad \theta \in H,$$

then $\theta = \bar{x}_N$ is the unique minimiser of F . This is easily seen by “opening the squares” and writing

$$F(\theta) = F(\bar{x}_N) + N\|\theta - \bar{x}_N\|^2.$$

The concept of a Fréchet mean (Fréchet [55]) generalises the notion of mean to a more general metric space by replacing the usual “sum of squares” with a “sum of squared distances”, giving rise to the so-called *Fréchet functional*. A closely related notion is that of a *Karcher mean* (Karcher [78]), a term that describes stationary points of the sum of squares functional, when the latter is differentiable (see Sect. 3.1.6). Population versions of Fréchet means, assuming the space is endowed with a probability law, can also be defined, replacing summation by expectation with respect to that law.

Electronic Supplementary Material The online version of this chapter (https://doi.org/10.1007/978-3-030-38438-8_3) contains supplementary material.

¹ It should be remarked that this is a Hilbertian property (or at least a property linked to an inner product), not merely a linear property. In other words, it does not extend to Banach spaces. As an example, let $H = \mathbb{R}^2$ with the L_1 norm and consider the vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$ of the unit simplex. The mean of these is $(1/3, 1/3)$ but for (x, y) in the triangle,

$$F(x, y) = (x + y)^2 + (x + 1 - y)^2 + (1 - x + y)^2 = 2 + x^2 + y^2 + (x - y)^2$$

is minimised at $(0, 0)$.

Fréchet means are perhaps the most basic object of statistical interest, and this chapter studies such means when the underlying space is the Wasserstein space \mathcal{W}_2 . In general, existence and uniqueness of a Fréchet mean can be subtle, but we will see that the nature of optimal transport allows for rather clean statements in the case of Wasserstein space.

3.1 Empirical Fréchet Means in \mathcal{W}_2

3.1.1 The Fréchet Functional

As foretold in the preceding paragraph, the definition of a Fréchet mean requires the definition of an appropriate sum-of-squares functional, the *Fréchet functional*:

Definition 3.1.1 (Empirical Fréchet Functional and Mean) *The Fréchet functional associated with measures $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ is*

$$F : \mathcal{W}_2(\mathcal{X}) \rightarrow \mathbb{R} \quad F(\gamma) = \frac{1}{2N} \sum_{i=1}^N W_2^2(\gamma, \mu^i), \quad \gamma \in \mathcal{W}_2(\mathcal{X}). \quad (3.1)$$

A Fréchet mean of (μ^1, \dots, μ^N) is a minimiser of F in $\mathcal{W}_2(\mathcal{X})$ (if it exists).

In analysis, a Fréchet mean is often called a *barycentre*. We shall use the terminology of “Fréchet mean” that is arguably more popular in statistics.²

The factor $1/(2N)$ is irrelevant for the definition of Fréchet mean. It is introduced in order to have simpler expressions for the derivatives (Theorems 3.1.14 and 3.2.13) and to be compatible with a population version $\mathbb{E}W_2^2(\gamma, \Lambda)/2$ (see (3.3)).

The first reference that deals with empirical Fréchet means in $\mathcal{W}_2(\mathbb{R}^d)$ is the seminal paper of Agueh and Carlier [2]. They treat the more general weighted Fréchet functional

$$F(\gamma) = \frac{1}{2} \sum_{i=1}^N w_i W_2^2(\gamma, \mu^i), \quad 0 \leq w_i, \quad \sum_{i=1}^N w_i = 1,$$

but, for simplicity, we shall focus on the case of equal weights. (If all the w_i 's are rational, then the weighted functional can be encompassed in (3.1) by taking some of the μ^i 's to be the same. The case of irrational w_i 's is then treated with continuity arguments. Moreover, (3.3) encapsulates (3.1) as well as the weighted version when Λ can take finitely many values.)

² Interestingly, Fréchet himself [56] considered the Wasserstein metric between probability measures on \mathbb{R} , and some refer to this as the *Fréchet distance* (e.g., Dowson and Landau [44]), which is another reason to use this terminology.

3.1.2 Multimarginal Formulation, Existence, and Continuity

In [60], Gangbo and Świąch consider the following *multimarginal* Monge–Kantorovich problem. Let μ^1, \dots, μ^N be N measures in $\mathscr{W}_2(\mathscr{X})$ and let $\Pi(\mu^1, \dots, \mu^N)$ be the set of probability measures in \mathscr{X}^N having $\{\mu^i\}_{i=1}^N$ as marginals. The problem is to minimise

$$G(\pi) = \frac{1}{2N^2} \int_{\mathscr{X}^N} \sum_{i < j} \|x_i - x_j\|^2 d\pi(x_1, \dots, x_N), \quad \text{over } \pi \in \Pi(\mu^1, \dots, \mu^N).$$

The factor $1/(2N^2)$ is of course irrelevant for the minimisation and its purpose will be clarified shortly. If $N = 2$, we obtain the Kantorovich problem with quadratic cost. The probabilistic interpretation (as in Sect. 1.2) is that one is given random variables X_1, \dots, X_N with marginal probability laws μ^1, \dots, μ^N and seeks to construct a random vector $Y = (Y_1, \dots, Y_N)$ on \mathscr{X}^N such that $X_i \stackrel{d}{=} Y_i$ and

$$\frac{1}{2N^2} \mathbb{E} \sum_{i < j} \|Y_i - Y_j\|^2 \leq \frac{1}{2N^2} \mathbb{E} \sum_{i < j} \|Z_i - Z_j\|^2.$$

for any other random vector $Z = (Z_1, \dots, Z_N)$ such that $X_i \stackrel{d}{=} Z_i$. Intuitively, we seek a random vector with prescribed marginals but maximally correlated entries.

We refer to elements of $\Pi(\mu^1, \dots, \mu^N)$ (equivalently, joint laws of X_1, \dots, X_N) as *multicouplings* (of μ^1, \dots, μ^N). Just like in the Kantorovich problem, there always exists an optimal multicoupling π .

Let us now show how the multimarginal problem is equivalent to the problem of finding the Fréchet mean of μ^1, \dots, μ^N . The first thing to observe is that the objective function can be written as

$$G(\pi) = \int_{\mathscr{X}^N} \frac{1}{2N} \sum_{i=1}^N \|x_i - M(x)\|^2 d\pi(x), \quad M(x) = M(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i.$$

The next result shows that the Fréchet mean and the multicoupling problems are essentially the same.

Proposition 3.1.2 (Fréchet Means and Multicouplings) *Let $\mu^1, \dots, \mu^N \in \mathscr{W}(\mathscr{X})$. Then μ is a Fréchet mean of (μ^1, \dots, μ^N) if and only if there exists an optimal multicoupling $\pi \in \mathscr{W}(\mathscr{X}^N)$ of (μ^1, \dots, μ^N) such that $\mu = M\#\pi$, and furthermore $F(\mu) = G(\pi)$.*

Proof. Let π be an arbitrary multicoupling of (μ^1, \dots, μ^N) and set $\mu = M\#\pi$. Then $(x \mapsto x_i, M)\#\pi$ is a coupling of μ^i and μ , and therefore

$$\int_{\mathscr{X}^N} \|x_i - M(x)\|^2 d\pi(x) \geq W^2(\mu, \mu_i).$$

Summation over i gives $F(\mu) \leq G(\pi)$ and so $\inf F \leq \inf G$.

For the other inequality, let $\mu \in \mathscr{W}(\mathscr{X})$ be arbitrary. For each i , let π^i be an optimal coupling between μ and μ^i . Invoking the gluing lemma (Ambrosio and Gigli [10, Lemma 2.1]), we may glue all π^i 's using their common marginal μ . This procedure constructs a measure η on \mathscr{X}^{N+1} with marginals μ_1, \dots, μ_N, μ and its relevant projection π is then a multicoupling of μ_1, \dots, μ_N .

Since \mathscr{X} is a Hilbert space, the minimiser of $y \mapsto \sum \|x_i - y\|^2$ is $y = M(x)$. Thus

$$F(\mu) = \frac{1}{2N} \int_{\mathscr{X}^{N+1}} \sum_{i=1}^N \|x_i - y\|^2 d\eta(x, y) \geq \frac{1}{2N} \int_{\mathscr{X}^{N+1}} \sum_{i=1}^N \|x_i - M(x)\|^2 d\eta(x, y) = G(\pi).$$

In particular, $\inf F \geq \inf G$ and combining this with the established converse inequality we see that $\inf F = \inf G$. Observe also that the last displayed inequality holds as equality if and only if $y = M(x)$ η -almost surely, in which case $\mu = M\#\pi$. Therefore, if μ does not equal $M\#\pi$, then $F(\mu) > G(\pi) \geq F(M\#\pi)$, and μ cannot be optimal. Finally, if π is optimal, then

$$F(M\#\pi) \leq G(\pi) = \inf G = \inf F$$

establishing optimality of $\mu = M\#\pi$ and completing the proof.

Since optimal couplings exist, we deduce that so do Fréchet means.

Corollary 3.1.3 (Fréchet Means and Moments) *Any finite collection of measures $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathscr{X})$ admits a Fréchet mean μ , for all $p \geq 1$*

$$\int_{\mathscr{X}} \|x\|^p d\mu(x) \leq \frac{1}{N} \sum_{i=1}^N \int_{\mathscr{X}} \|x\|^p d\mu^i(x),$$

and when $p > 1$ equality holds if and only if $\mu^1 = \dots = \mu^N$.

Proof. Let π be a multicoupling of μ^1, \dots, μ^N such that $\mu = M_N\#\pi$ (Proposition 3.1.2). Then

$$\begin{aligned} \int_{\mathscr{X}} \|x\|^p d\mu(x) &= \int_{\mathscr{X}^N} \left\| \frac{1}{N} \sum_{i=1}^N x_i \right\|^p d\pi(x) \leq \frac{1}{N} \sum_{i=1}^N \int_{\mathscr{X}^N} \|x_i\|^p d\pi(x) \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathscr{X}} \|x\|^p d\mu^i(x). \end{aligned}$$

The statement about equality follows from strict convexity of $x \mapsto \|x\|^p$ if $p > 1$.

A further corollary of Proposition 3.1.2 is a bound on the support:

Corollary 3.1.4 *The support of any Fréchet mean is included in the set*

$$\frac{\text{supp}\mu^1 + \dots + \text{supp}\mu^N}{N} = \left\{ \frac{x_1 + \dots + x_N}{N} : x_i \in \text{supp}\mu^i \right\} \subseteq \text{conv} \left(\bigcup_{i=1}^N \text{supp}\mu^i \right).$$

In particular, if all the μ^i 's are supported on a common convex set K , then so is any of their Fréchet means.

The multimarginal formulation also yields a continuity property for the empirical Fréchet mean. Conditions for uniqueness will be given in the next subsection.

Theorem 3.1.5 (Continuity of Fréchet Means) *Suppose that $W_2(\mu_k^i, \mu^i) \rightarrow 0$ for $i = 1, \dots, N$ and let $\bar{\mu}_k$ denote any Fréchet mean of $(\mu_k^1, \dots, \mu_k^N)$. Then $(\bar{\mu}_k)$ stays in a compact set of $\mathcal{W}_2(\mathcal{X})$, and any limit point is a Fréchet mean of (μ^1, \dots, μ^N) .*

In particular, if μ^1, \dots, μ^N have a *unique* Fréchet mean $\bar{\mu}$, then $\bar{\mu}_k \rightarrow \bar{\mu}$ in $\mathcal{W}_2(\mathcal{X})$.

Proof. We sketch the steps of the proof here, with the full details given on page 63 of the supplement.

Step 1: tightness of $(\bar{\mu}_k)$. This is true because the collection of multicouplings is tight, and the mean function M is continuous.

Step 2: weak limits are limits in $\mathcal{W}_2(\mathcal{X})$. This holds because the mean function has linear growth.

Step 3: the limit is a Fréchet mean of (μ^1, \dots, μ^N) . From Corollary 3.1.3, it follows that $\bar{\mu}_k$ must be sought on some fixed bounded set in $\mathcal{W}_2(\mathcal{X})$. On such sets, the Fréchet functionals are uniformly Lipschitz, so their minimisers converge as well.

3.1.3 Uniqueness and Regularity

A general situation in which Fréchet means are unique is when the Fréchet functional is strictly convex. In the Wasserstein space, this requires some regularity, but weak convexity holds in general. Absolutely continuous measures on infinite-dimensional \mathcal{X} are defined in Definition 1.6.4.

Proposition 3.1.6 (Convexity of the Fréchet Functional) *Let $\Lambda, \gamma_i \in \mathcal{W}_2(\mathcal{X})$ and $t \in [0, 1]$. Then*

$$W_2^2(t\gamma_1 + (1-t)\gamma_2, \Lambda) \leq tW_2^2(\gamma_1, \Lambda) + (1-t)W_2^2(\gamma_2, \Lambda). \quad (3.2)$$

When Λ is absolutely continuous, the inequality is strict unless $t \in \{0, 1\}$ or $\gamma_1 = \gamma_2$.

Remark 3.1.7 *The Wasserstein distance is not convex along geodesics. That is, if we replace the linear interpolant $t\gamma_1 + (1-t)\gamma_2$ by McCann's interpolant, then $t \mapsto W_2^2(\gamma_t, \Lambda)$ is not necessarily convex (Ambrosio et al. [12, Example 9.1.5]).*

Proof. Let $\pi_i \in \Pi(\gamma_i, \Lambda)$ be optimal and notice that the linear interpolant $t\pi_1 + (1-t)\pi_2 \in \Pi(t\gamma_1 + (1-t)\gamma_2, \Lambda)$, so that

$$W_2^2(t\gamma_1 + (1-t)\gamma_2, \Lambda) \leq \int_{\mathcal{X}^2} \|x-y\|^2 d[t\pi_1 + (1-t)\pi_2](x, y),$$

which is (3.2). When Λ is absolutely continuous and $t \in (0, 1)$, equality in (3.2) holds if and only if $\pi_t = t\pi_1 + (1-t)\pi_2 = (\mathbf{t}_\Lambda^{t\gamma_1 + (1-t)\gamma_2} \times \mathbf{i})\#\Lambda$. But π_t is supported on the graphs of two functions: $\mathbf{t}_\Lambda^{\gamma_1}$ and $\mathbf{t}_\Lambda^{\gamma_2}$. Consequently, equality can hold only if these two maps equal Λ -almost surely, or, equivalently, if $\gamma_1 = \gamma_2$.

As a corollary, we deduce that the Fréchet mean is unique if one of the measures μ^i is absolutely continuous, and this extends to the population version (see Proposition 3.2.7).

We conclude this subsection by stating an important regularity property in the Euclidean case. See Agueh and Carlier [2, Proposition 5.1] for a proof.

Proposition 3.1.8 (L_∞ -Regularity of Fréchet Means) *Let $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathbb{R}^d)$ and suppose that μ^1 is absolutely continuous with density bounded by M . Then the Fréchet mean of $\{\mu^i\}$ is absolutely continuous with density bounded by $N^d M$ and is consequently a Karcher mean.*

In Theorem 5.5.2, we extend Proposition 3.1.8 to the population level.

3.1.4 The One-Dimensional and the Compatible Case

When $\mathscr{X} = \mathbb{R}$, there is a simple expression for the Fréchet mean because $\mathscr{W}_2(\mathbb{R})$ can be imbedded in a Hilbert space. Indeed, recall that

$$W_2(\mu, \nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L_2(0,1)}$$

(see Sect. 2.3.2 or 1.5). In view of that, $\mathscr{W}_2(\mathbb{R})$ can be seen as the convex closed subset of $L_2(0, 1)$ formed by equivalence classes of left-continuous nondecreasing functions on $(0, 1)$: any quantile function is left-continuous and nondecreasing, and any such function G can be seen to be the inverse function of the distribution function, the *right-continuous inverse* of G

$$F(x) = \inf\{t \in (0, 1) : G(t) > x\} = \sup\{t \in (0, 1) : G(t) \leq x\}.$$

(See, for example, Bobkov and Ledoux [25, Appendix A].) Therefore, the Fréchet mean of $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathbb{R})$ is the measure μ having quantile function

$$F_\mu^{-1} = \frac{1}{N} \sum_{i=1}^N F_{\mu^i}.$$

The Fréchet mean is thus unique. This is no longer true in higher dimension, unless some regularity is imposed on the measures (Proposition 3.2.7).

Boissard et al. [28] noticed that compatibility of μ^1, \dots, μ^N according to Definition 2.3.1 allows for a simple solution to the Fréchet mean problem, as in the one-dimensional case. Recall from Proposition 3.1.2 that this is equivalent to the multimarginal problem. Returning to the original form of G , we obtain an easy lower bound for any $\pi \in \Pi(\mu^1, \dots, \mu^N)$:

$$G(\pi) = \frac{1}{2N^2} \int_{\mathcal{X}^N} \sum_{i < j} \|x_i - x_j\|^2 d\pi(x_1, \dots, x_N) \geq \frac{1}{2N^2} \sum_{i < j} W_2^2(\mu^i, \mu^j),$$

because the (i, j) -th marginal of π is a coupling of μ^i and μ^j . Thus, if equality above holds for π , then π is optimal and $M\#\pi$ is the Fréchet mean by Proposition 3.1.2. This is indeed the case for $\pi = (\mathbf{i}, \mathbf{t}_{\mu^1}^{\mu^2}, \dots, \mathbf{t}_{\mu^1}^{\mu^N})\#\mu^1$ because the compatibility gives:

$$\begin{aligned} \int_{\mathcal{X}^N} \|x_i - x_j\|^2 d\pi(x_1, \dots, x_N) &= \int_{\mathcal{X}} \left\| \mathbf{t}_{\mu^1}^{\mu^i} - \mathbf{t}_{\mu^1}^{\mu^j} \right\|^2 d\mu^1 \\ &= \int_{\mathcal{X}} \left\| \mathbf{t}_{\mu^1}^{\mu^i} \circ \mathbf{t}_{\mu^j}^{\mu^i} - \mathbf{i} \right\|^2 d\mu^j = W_2^2(\mu^i, \mu^j). \end{aligned}$$

We may thus conclude, in a slightly more general form (γ was μ^1 above):

Theorem 3.1.9 (Fréchet Mean of Compatible Measures) *Suppose that $\{\gamma, \mu^1, \dots, \mu^N\}$ are compatible measures. Then*

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

is the Fréchet mean of (μ^1, \dots, μ^N) .

A population version is given in Theorem 5.5.3.

3.1.5 The Agueh–Carlier Characterisation

Agueh and Carlier [2] provide a useful sufficient condition for γ to be the Fréchet mean. When $\mathcal{X} = \mathbb{R}^d$, this condition is also necessary [2, Proposition 3.8], hence characterising Fréchet means in \mathbb{R}^d . It will allow us to easily deduce some equivariance results for Fréchet means with respect to independence (Lemma 3.1.11) and rotations (3.1.12). More importantly, it provides a sufficient condition under which a local minimum of F is a global minimum (Theorem 3.1.15) and the same idea can be used to relate the population Fréchet mean to the expected value of the optimal maps (Theorem 4.2.4). Recall that ϕ^* denotes the Legendre transform of ϕ , as defined on page 14.

Proposition 3.1.10 (Fréchet Means and Potentials) *Let $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ be absolutely continuous, let $\gamma \in \mathcal{W}_2(\mathcal{X})$ and denote by ϕ_i^* the convex potentials of $\mathbf{t}_{\mu^i}^{\gamma}$. If $\phi_i = \phi_i^{**}$ are such that*

$$\frac{1}{N} \sum_{i=1}^N \phi_i(x) \leq \frac{1}{2} \|x\|^2, \quad \forall x \in \mathcal{X}, \quad \text{with equality } \gamma\text{-almost surely,}$$

then γ is the unique Fréchet mean of μ^1, \dots, μ^N .

Proof. Uniqueness follows from Proposition 3.2.7. If $\theta \in \mathcal{W}_2(\mathcal{X})$ is any measure, then the Kantorovich duality yields

$$\begin{aligned} W_2^2(\gamma, \mu^i) &= \int_{\mathcal{X}} \left(\frac{1}{2} \|x\|^2 - \phi_i(x) \right) d\gamma(x) + \int_{\mathcal{X}} \left(\frac{1}{2} \|y\|^2 - \phi_i^*(y) \right) d\mu^i(y); \\ W_2^2(\theta, \mu^i) &\geq \int_{\mathcal{X}} \left(\frac{1}{2} \|x\|^2 - \phi_i(x) \right) d\theta(x) + \int_{\mathcal{X}} \left(\frac{1}{2} \|y\|^2 - \phi_i^*(y) \right) d\mu^i(y). \end{aligned}$$

Summation over i gives the result.

A population version of this result, based on similar calculations, is given in Theorem 4.2.4.

The next two results are formulated in \mathbb{R}^d because then the converse of Proposition 3.1.10 is proven to be true. If one could extend [2, Proposition 3.8] to any separable Hilbert \mathcal{X} , then the two lemmata below will hold with \mathbb{R}^d replaced by \mathcal{X} . The simple proofs are given on page 66 of the supplement.

Lemma 3.1.11 (Independent Fréchet Means) *Let μ^1, \dots, μ^N and ν^1, \dots, ν^N be absolutely continuous measures in $\mathcal{W}_2(\mathbb{R}^{d_1})$ and $\mathcal{W}_2(\mathbb{R}^{d_2})$ with Fréchet means μ and ν , respectively. Then the independent coupling $\mu \otimes \nu$ is the Fréchet mean of $\mu^1 \otimes \nu^1, \dots, \mu^N \otimes \nu^N$.*

By induction (or a straightforward modification of the proof), one can show that the Fréchet mean of $(\mu^i \otimes \nu^i \otimes \rho^i)$ is $\mu \otimes \nu \otimes \rho$, and so on.

Lemma 3.1.12 (Rotated Fréchet Means) *If μ is the Fréchet mean of the absolutely continuous measures μ^1, \dots, μ^N and U is orthogonal, then $U\#\mu$ is the Fréchet mean of $U\#\mu^1, \dots, U\#\mu^N$.*

3.1.6 Differentiability of the Fréchet Functional and Karcher Means

Since we seek to minimise the Fréchet functional F , it would be helpful if F were differentiable, because we could then find at least local minima by solving the equation $F' = 0$. This observation of Karcher [78] leads to the notion of *Karcher mean*.

Definition 3.1.13 (Karcher Mean) *Let F be a Fréchet functional associated with some random measure Λ in $\mathcal{W}_2(\mathcal{X})$. Then γ is a Karcher mean for Λ if F is differentiable at γ and $F'(\gamma) = 0$.*

Of course, if γ is a Fréchet mean for the random measure Λ and F is differentiable at γ , then $F'(\gamma)$ must vanish. In this subsection, we build upon the work of Ambrosio et al. [12] and determine the derivative of the Fréchet functional. This will not only allow for a simple characterisation of Karcher means in terms of the optimal maps $\mathbf{t}_\gamma^\Lambda$ (Proposition 3.2.14), but will also be the cornerstone of the construction of a steepest

descent algorithm for empirical calculation of Fréchet means. The differentiability holds at the population level too (Theorem 3.2.13).

It turns out that the tangent bundle structure described in Sect. 2.3 gives rise to a differentiable structure in the Wasserstein space. Fix $\mu^0 \in \mathscr{W}_2(\mathscr{X})$ and consider the function

$$F_0 : \mathscr{W}_2(\mathscr{X}) \rightarrow \mathbb{R}, \quad F_0(\gamma) = \frac{1}{2} W_2^2(\gamma, \mu^0).$$

Ambrosio et al. [12, Corollary 10.2.7] show that when γ is absolutely continuous,

$$\lim_{W_2(v, \gamma) \rightarrow 0} \frac{F_0(v) - F_0(\gamma) + \int_{\mathscr{X}} \langle \mathbf{t}_\gamma^{\mu^0}(x) - x, \mathbf{t}_\gamma^v(x) - x \rangle d\gamma(x)}{W_2(v, \gamma)} = 0.$$

Parts of the proof of this result (the limit superior above is ≤ 0 ; the limit inferior is bounded below) are reproduced in Proposition 3.2.12. The integral above can be seen as the inner product

$$\langle \mathbf{t}_\gamma^{\mu^0} - \mathbf{i}, \mathbf{t}_\gamma^v - \mathbf{i} \rangle$$

in the space $\mathscr{L}_2(\gamma)$ that includes as a (closed) subspace the tangent space Tan_γ . In terms of this inner product and the log map, we can write

$$F_0(v) - F_0(\gamma) = -\langle \log_\gamma(\mu^0), \log_\gamma(v) \rangle + o(W_2(v, \gamma)), \quad v \rightarrow \gamma \text{ in } \mathscr{W}_2,$$

so that F_0 is Fréchet-differentiable³ at γ with derivative

$$F'_0(\gamma) = -\log_\gamma(\mu^0) = -(\mathbf{t}_\gamma^{\mu^0} - \mathbf{i}) \in \text{Tan}_\gamma.$$

By linearity, one immediately obtains:

Theorem 3.1.14 (Gradient of the Fréchet Functional) *Fix a collection of measures $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathscr{X})$. When $\gamma \in \mathscr{W}_2(\mathscr{X})$ is absolutely continuous, the Fréchet functional*

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^N W_2^2(\gamma, \mu^i), \quad \gamma \in \mathscr{W}_2(\mathscr{X})$$

is Fréchet-differentiable and

$$F'(\gamma) = -\frac{1}{N} \sum_{i=1}^N \log_\gamma(\mu^i) = -\frac{1}{N} \sum_{i=1}^N (\mathbf{t}_\gamma^{\mu^i} - \mathbf{i}).$$

It follows from this that an absolutely continuous $\gamma \in \mathscr{W}_2(\mathscr{X})$ is a Karcher mean if and only if the average of the optimal maps is the identity. If in addition one μ^i is absolutely continuous with bounded density, then the Fréchet mean $\bar{\mu}$ is absolutely

³ The notion of Fréchet derivative is also named after Maurice Fréchet, but is not directly related to Fréchet means.

continuous by Proposition 3.1.8, so it is a Karcher mean. The result extends to the population version; see Proposition 3.2.14.

It may happen that a collection μ^1, \dots, μ^N of absolutely continuous measures have a Karcher mean that is not a Fréchet mean; see Álvarez-Esteban et al. [9, Example 3.1] for an example in \mathbb{R}^2 . But a Karcher mean γ is “almost” a Fréchet mean in the following sense. By Proposition 3.2.14, $N^{-1} \sum \mathbf{t}_\gamma^{\mu^i}(x) = x$ for γ -almost all x . If, on the other hand, the equality holds for all $x \in \mathscr{X}$, then γ is the Fréchet mean by taking integrals and applying Proposition 3.1.10. One can hope that under regularity conditions, the γ -almost sure equality can be upgraded to equality everywhere. Indeed, this is the case:

Theorem 3.1.15 (Optimality Criterion for Karcher Means) *Let $U \subseteq \mathbb{R}^d$ be an open convex set and let $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathbb{R}^d)$ be probability measures on U with bounded strictly positive densities g^1, \dots, g^N . Suppose that an absolutely continuous Karcher mean γ is supported on U with bounded strictly positive density f there. Then γ is the Fréchet mean of μ^1, \dots, μ^N if one of the following holds:*

1. $U = \mathbb{R}^d$ and the densities f, g^1, \dots, g^N are of class $C^{0,\alpha}$ for some $\alpha > 0$;
2. U is bounded and the densities f, g^1, \dots, g^N are bounded below on U .

Proof. The result exploits Caffarelli’s regularity theory for Monge–Ampère equations in the form of Theorem 1.6.7. In the first case, there exist C^1 (in fact, $C^{2,\alpha}$) convex potentials φ_i on \mathbb{R}^d with $\mathbf{t}_\gamma^{\mu^i} = \nabla \varphi_i$, so that $\mathbf{t}_\gamma^{\mu^i}(x)$ is a singleton for all $x \in \mathbb{R}^d$. The set $\{x \in \mathbb{R}^d : \sum \mathbf{t}_\gamma^{\mu^i}(x)/N \neq x\}$ is γ -negligible (and hence Lebesgue negligible) and open by continuity. It is therefore empty, so $F'(\gamma) = 0$ everywhere, and γ is the Fréchet mean (see the discussion before the theorem).

In the second case, by the same argument we have $\sum \mathbf{t}_\gamma^{\mu^i}(x)/N = x$ for all $x \in U$. Since U is convex, there must exist a constant C such that $\sum \varphi_i(x) = C + N\|x\|^2/2$ for all $x \in U$, and we may assume without loss of generality that $C = 0$. If one repeats the proof of Proposition 3.1.10, then $F(\gamma) \leq F(\theta)$ for all $\theta \in P(U)$. By continuity considerations, the inequality holds for all $\theta \in P(\overline{U})$ (Theorem 2.2.7) and since \overline{U} is closed and convex, γ is the Fréchet mean by Corollary 3.1.3.

3.2 Population Fréchet Means

In this section, we extend the notion of empirical Fréchet mean to the population level, where Λ is a random element in $\mathscr{W}_2(\mathscr{X})$ (a measurable mapping from a probability space to $\mathscr{W}_2(\mathscr{X})$). This requires a different strategy, since it is not clear how to define the analogue of the multicouplings at that level of abstraction. However, it is important to point out that when there is more structure in Λ , multicouplings can be defined as laws of stochastic processes; see Pass [102] for a detailed account of the problem in this case.

In analogy with (3.1), we define:

Definition 3.2.1 (Population Fréchet Mean) Let Λ be a random measure in $\mathcal{W}_2(\mathcal{X})$. The Fréchet mean of Λ is the minimiser (if it exists and is unique) of the Fréchet functional

$$F(\gamma) = \frac{1}{2} \mathbb{E} W_2^2(\gamma, \Lambda), \quad \gamma \in \mathcal{W}_2(\mathcal{X}). \quad (3.3)$$

Since W_2 is continuous and nonnegative, the expectation is well-defined.

3.2.1 Existence, Uniqueness, and Continuity

Existence and uniqueness of Fréchet means on a general metric space M are rather delicate questions. Usually, existence proofs are easier: for example, since the Fréchet functional F is continuous on M (as we show below), one often invokes local compactness of M in order to establish existence of a minimiser. Unfortunately, a different strategy is needed when $M = \mathcal{W}_2(\mathcal{X})$, because the Wasserstein space is not locally compact (Proposition 2.2.9).

The first thing to notice is that F is indeed continuous (this is clear for the empirical version). This is a consequence of the triangle inequality and holds when $\mathcal{W}_2(\mathcal{X})$ is replaced by any metric space.

Lemma 3.2.2 (Finiteness of the Fréchet Functional) If F is not identically infinite, then it is finite and locally Lipschitz everywhere on $\mathcal{W}_2(\mathcal{X})$.

Proof. Assume that F is finite at γ . If θ is any other measure in $\mathcal{W}_2(\mathcal{X})$, write

$$2F(\gamma) - 2F(\theta) = \mathbb{E}[W_2(\gamma, \Lambda) - W_2(\theta, \Lambda)][W_2(\gamma, \Lambda) + W_2(\theta, \Lambda)].$$

Since $x \leq 1 + x^2$ for all x , the triangle inequality in $\mathcal{W}_2(\mathcal{X})$ yields

$$\begin{aligned} 2|F(\gamma) - F(\theta)| &\leq W_2(\gamma, \theta)[2\mathbb{E}W_2(\gamma, \Lambda) + W_2(\theta, \gamma)] \\ &\leq W_2(\gamma, \theta)[2\mathbb{E}W_2^2(\gamma, \Lambda) + 2 + W_2(\theta, \gamma)]. \end{aligned}$$

Since $F(\gamma) < \infty$, this shows that F is finite everywhere and the right-hand side vanishes as $\theta \rightarrow \gamma$ in $\mathcal{W}_2(\mathcal{X})$. Now that we know that F is continuous, the same upper bound shows that it is in fact locally Lipschitz.

Example: let (a_n) be a sequence of positive numbers that sum up to one. Let $x_n = 1/a_n$ and suppose that Λ equals $\delta_{\{x_n\}} \in \mathcal{W}_2(\mathbb{R})$ with probability a_n . Then

$$\mathbb{E}W_2^2(\Lambda, \delta_0) = \sum_{n=1}^{\infty} a_n x_n^2 = \sum_{n=1}^{\infty} 1/a_n = \infty,$$

and by Lemma 3.2.2 F is identically infinite. Henceforth, we say that F is finite when the condition in Lemma 3.2.2 holds.

Using the lower semicontinuity (2.5), one can prove existence on \mathbb{R}^d rather easily. (The empirical means exist even in infinite dimensions by Corollary 3.1.3.)

Proposition 3.2.3 (Existence of Fréchet Means) *The Fréchet functional associated with any random measure Λ in $\mathscr{W}_2(\mathbb{R}^d)$ admits a minimiser.*

Proof. The assertion is clear if F is identically infinite. Otherwise, let (γ_n) be a minimising sequence. We wish to show that the sequence is tight. Define $L = \sup_n F(\gamma_n) < \infty$ and observe that since $x \leq 1 + x^2$ for all $x \in \mathbb{R}$,

$$\mathbb{E}W_2(\gamma_n, \Lambda) \leq 1 + \mathbb{E}W_2^2(\gamma_n, \Lambda) \leq 2L + 1, \quad n = 1, 2, \dots$$

By the triangle inequality

$$L' = \mathbb{E}W_2(\delta_0, \Lambda) \leq W_2(\delta_0, \gamma_1) + \mathbb{E}W_2(\gamma_1, \Lambda) \leq W_2(\delta_0, \gamma_1) + 2L + 1$$

so that for all n

$$\left(\int_{\mathbb{R}^d} \|x\|^2 d\gamma_n(x) \right)^{1/2} = W_2(\gamma_n, \delta_0) \leq \mathbb{E}W_2(\gamma_n, \Lambda) + \mathbb{E}W_2(\Lambda, \delta_0) \leq 2L + 1 + L' < \infty.$$

Since closed and bounded sets in \mathbb{R}^d are compact, it follows that (γ_n) is a tight sequence. We may assume that $\gamma_n \rightarrow \gamma$ weakly, then use (2.5) and Fatou's lemma to obtain

$$2F(\gamma) = \mathbb{E}W_2^2(\gamma, \Lambda) \leq \mathbb{E} \liminf_{n \rightarrow \infty} W_2^2(\gamma_n, \Lambda) \leq \liminf_{n \rightarrow \infty} \mathbb{E}W_2^2(\gamma_n, \Lambda) = 2 \inf F.$$

Thus, γ is a minimiser of F , and existence is established.

When \mathscr{X} is an infinite-dimensional Hilbert space, existence still holds under a compactness assumption. We first prove a result about the support of the Fréchet mean. At the empirical level, one can say more about the support (see Corollary 3.1.4).

Proposition 3.2.4 (Support of Fréchet Mean) *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ and let $K \subseteq \mathscr{X}$ be a convex closed set such that $\mathbb{P}[\Lambda(K) = 1] = 1$. If γ minimises F , then $\gamma(K) = 1$.*

Remark 3.2.5 *For any closed $K \subseteq \mathscr{X}$ and any $\alpha \in [0, 1]$, the set $\{\Lambda \in \mathscr{W}_p(\mathscr{X}) : \Lambda(K) \geq \alpha\}$ is closed in $\mathscr{W}_p(\mathscr{X})$ because $\{\Lambda \in P(\mathscr{X}) : \Lambda(K) \geq \alpha\}$ is weakly closed by the portmanteau lemma (Lemma 1.7.1).*

The proof amounts to a simple projection argument; see page 70 in the supplement.

Corollary 3.2.6 *If there exists a compact convex K satisfying the hypothesis of Proposition 3.2.4, then the Fréchet functional admits a minimiser supported on K .*

Proof. Proposition 3.2.4 allows us to restrict the domain of F to $\mathscr{W}_2(K)$, the collection of probability measures supported on K . Since this set is compact in $\mathscr{W}_2(\mathscr{X})$ (Corollary 2.2.5), the result follows from continuity of F .

From the convexity (3.2), one obtains a simple criterion for uniqueness. See Definition 1.6.4 for absolute continuity in infinite dimensions.

Proposition 3.2.7 (Uniqueness of Fréchet Means) *Let Λ be a random measure in $\mathcal{W}_2(\mathcal{X})$ with finite Fréchet functional. If Λ is absolutely continuous with positive (inner) probability, then the Fréchet mean of Λ is unique (if it exists).*

Remark 3.2.8 *It is not obvious that the set of absolutely continuous measures is measurable in $\mathcal{W}_2(\mathcal{X})$. We assume that there exists a Borel set $A \subset \mathcal{W}_2(\mathcal{X})$ such that $\mathbb{P}(\Lambda \in A) > 0$ and all measures in A are absolutely continuous.*

Proof. By taking expectations in (3.2), one sees that F is convex on $\mathcal{W}_2(\mathcal{X})$ with respect to linear interpolants. From Proposition 3.1.6, we conclude that

$$\Lambda \text{ absolutely continuous} \implies \gamma \mapsto \frac{1}{2}W_2^2(\gamma, \Lambda) \text{ strictly convex.}$$

As F was already shown to be weakly convex in any case, it follows that

$$\mathbb{P}(\Lambda \text{ absolutely continuous}) > 0 \implies F \text{ strictly convex.}$$

Since strictly convex functionals have at most one minimiser, this completes the proof.

We state without proof an important consistency result (Le Gouic and Loubes [87, Theorem 3]). Since $\mathcal{W}_2(\mathcal{X})$ is a complete and separable metric space, we can define the “second degree” Wasserstein space $\mathcal{W}_2(\mathcal{W}_2(\mathcal{X}))$. The law of a random measure Λ is in $\mathcal{W}_2(\mathcal{W}_2(\mathcal{X}))$ if and only if the corresponding Fréchet functional is finite.

Theorem 3.2.9 (Consistency of Fréchet Means) *Let Λ_n, Λ be random measures in $\mathcal{W}_2(\mathbb{R}^d)$ with finite Fréchet functionals and laws $\mathbb{P}_n, \mathbb{P} \in \mathcal{W}_2(\mathcal{W}_2(\mathbb{R}^d))$. If $\mathbb{P}_n \rightarrow \mathbb{P}$ in $\mathcal{W}_2(\mathcal{W}_2(\mathbb{R}^d))$, then any sequence λ_n of Fréchet means of Λ_n has a W_2 -limit point λ , which is a Fréchet mean of Λ .*

See the Bibliographical Notes for a more general formulation.

Corollary 3.2.10 (Wasserstein Law of Large Numbers) *Let Λ be a random measure in $\mathcal{W}_2(\mathbb{R}^d)$ with finite Fréchet functional and let Λ_1, \dots be a sample from Λ . Assume λ is the unique Fréchet mean of Λ (see Proposition 3.2.7). Then almost surely, the sequence of empirical Fréchet means of $\Lambda_1, \dots, \Lambda_n$ converges to λ .*

Proof. Let \mathbb{P} be the law of Λ and let \mathbb{P}_n be its empirical counterpart (a random element in $\mathcal{W}_2(\mathcal{W}_2(\mathbb{R}^d))$). Like in the proof of Proposition 2.2.6 (with \mathcal{X} replaced by the complete separable metric space $\mathcal{W}_2(\mathbb{R}^d)$), almost surely $\mathbb{P}_n \rightarrow \mathbb{P}$ in $\mathcal{W}_2(\mathcal{W}_2(\mathbb{R}^d))$ and Theorem 3.2.9 applies.

Under a compactness assumption, one can give a direct proof for the law of large numbers as in Theorem 3.1.5. This is done on page 71 in the supplement.

3.2.2 The One-Dimensional Case

As a generalisation of the empirical version, we have:

Theorem 3.2.11 (Fréchet Means in $\mathscr{W}_2(\mathbb{R})$) *Let Λ be a random measure in $\mathscr{W}_2(\mathbb{R})$ with finite Fréchet functional. Then the Fréchet mean of Λ is the unique measure λ with quantile function $F_\lambda^{-1}(t) = \mathbb{E}F_\Lambda^{-1}(t)$, $t \in (0, 1)$.*

Proof. Since $L_2(0, 1)$ is a Hilbert space, the random element $F_\Lambda^{-1} \in L_2(0, 1)$ has a unique Fréchet mean $g \in L_2(0, 1)$, defined by the relations $\langle g, f \rangle = \mathbb{E} \langle F_\Lambda^{-1}, f \rangle$ for all $f \in L_2(0, 1)$. On page 72 of the supplement, we show that g can be identified with F_λ^{-1} .

Interestingly, no regularity is needed in order for the Fréchet mean to be unique. This is not the case for higher dimensions, see Proposition 3.2.7. If there is some regularity, then one can state Theorem 3.2.11 in terms of optimal maps, because F_Λ^{-1} is the optimal map from $\text{Leb}|_{[0,1]}$ to Λ . If $\gamma \in \mathscr{W}_2(\mathbb{R})$ is any absolutely continuous (or even just continuous) measure, then Theorem 3.2.11 can be stated as follows: the Fréchet mean of Λ is the measure $[\mathbb{E}t_\gamma^\Lambda] \# \gamma$. A generalisation of this result to compatible measures (Definition 2.3.1) can be carried out in the same way, since compatible measures are imbedded in a Hilbert space, using the Bochner integrals for the definition of the expected optimal maps (see Sect. 2.4).

3.2.3 Differentiability of the Population Fréchet Functional

We now use the Fubini result (Proposition 2.4.9) in order to extend the differentiability of the Fréchet functional to the population version. This will follow immediately if we can interchange the expectation and the derivative in the form

$$F'(\gamma) = \frac{1}{2}(\mathbb{E}W_2^2)'(\gamma, \Lambda) = \mathbb{E} \left(\frac{1}{2}W_2^2 \right)'(\gamma, \Lambda) = -\mathbb{E}(t_\gamma^\Lambda - \mathbf{i}).$$

In order to do this, we will use dominated convergence in conjunction with uniform bounds on the slopes

$$u(\theta, \Lambda) = \frac{0.5W_2^2(\theta, \Lambda) - 0.5W_2^2(\theta_0, \Lambda) + \int_{\mathscr{X}} \langle t_{\theta_0}^\Lambda - i, t_{\theta_0}^\theta - i \rangle d\theta_0}{W_2(\theta, \theta_0)}, \quad u(\theta_0, \Lambda) = 0. \quad (3.4)$$

Proposition 3.2.12 (Slope Bounds) *Let θ_0 , Λ , and θ be probability measures with θ_0 absolutely continuous, and set $\delta = W_2(\theta, \theta_0)$. Then*

$$\frac{1}{2}\delta - W_2(\theta_0, \Lambda) - \sqrt{2W_2^2(\theta_0, \delta_0) + 2W_2^2(\Lambda, \delta_0)} \leq u(\theta, \Lambda) \leq \frac{1}{2}\delta,$$

where u is defined by (3.4). If the measures are compatible in the sense of Definition 2.3.1, then $u(\theta, \Lambda) = \delta/2$.

The proof is a slight variation of Ambrosio et al. [12, Theorem 10.2.2 and Proposition 10.2.6], and the details are given on page 72 of the supplement.

Theorem 3.2.13 (Population Fréchet Gradient) *Let Λ be a random measure with finite Fréchet functional F . Then F is Fréchet-differentiable at any absolutely continuous θ_0 in the Wasserstein space, and $F'(\theta_0) = \mathbb{E} \mathbf{t}_{\theta_0}^\Lambda - \mathbf{i} \in \mathcal{L}_2(\theta_0)$. More precisely,*

$$\frac{F(\theta) - F(\theta_0) + \int_{\mathcal{X}} \langle \mathbb{E} \mathbf{t}_{\theta_0}^\Lambda - \mathbf{i}, \mathbf{t}_{\theta_0}^\theta - \mathbf{i} \rangle d\theta_0}{W_2(\theta, \theta_0)} \rightarrow 0, \quad \theta \rightarrow \theta_0 \text{ in } \mathcal{W}_2.$$

Thus, the Fréchet derivative of F can be identified with the map $-(\mathbb{E} \mathbf{t}_{\theta_0}^\Lambda - \mathbf{i})$ in the tangent space at θ_0 , a subspace of $\mathcal{L}_2(\theta_0)$.

Proof. Introduce the slopes $u(\theta, \Lambda)$ defined by (3.4). Then for all $\Lambda, u(\theta, \Lambda) \rightarrow 0$ as $W_2(\theta, \theta_0) \rightarrow 0$, by the differentiability properties established above. Let us show that $\mathbb{E}u(\theta, \Lambda) \rightarrow 0$ as well. By Proposition 3.2.12, the expectation of u is bounded above by a constant that does not depend on Λ , and below by the negative of

$$\begin{aligned} \mathbb{E}W_2(\theta_0, \Lambda) + \mathbb{E}\sqrt{2W_2^2(\theta_0, \delta_0) + 2W_2^2(\Lambda, \delta_0)} \\ \leq \sqrt{2}W_2(\theta_0, \delta_0) + \mathbb{E}W_2(\theta_0, \Lambda) + \sqrt{2}\mathbb{E}W_2(\Lambda, \delta_0). \end{aligned}$$

Both expectations are finite by the hypothesis on Λ because the Fréchet functional is finite. The dominated convergence theorem yields

$$\mathbb{E}u(\theta, \Lambda) = \frac{F(\theta) - F(\theta_0) + \mathbb{E} \int_{\mathcal{X}} \langle \mathbf{t}_{\theta_0}^\Lambda - \mathbf{i}, \mathbf{t}_{\theta_0}^\theta - \mathbf{i} \rangle d\theta_0}{W_2(\theta, \theta_0)} \rightarrow 0, \quad W_2(\theta_0, \theta) \rightarrow 0.$$

The measurability of the integral and the result then follow from Fubini's theorem (see Proposition 2.4.9).

Proposition 3.2.14 *Let Λ be a random measure in $\mathcal{W}_2(\mathcal{X})$ with finite Fréchet functional F , and let γ be absolutely continuous in $\mathcal{W}_2(\mathcal{X})$. Then γ is a Karcher mean of Λ if and only if $\mathbb{E} \mathbf{t}_\gamma^\Lambda - \mathbf{i} = \mathbf{0}$ in $\mathcal{L}_2(\gamma)$. Furthermore, if γ is a Fréchet mean of Λ , then it is also a Karcher mean.*

The characterisation of Karcher means follows immediately from Theorem 3.2.13. The other statement is that the derivative vanishes at the minimum, which is fairly obvious intuitively; see page 73 in the supplement.

3.3 Bibliographical Notes

Proposition 3.1.2 is essentially due to Agueh and Carlier [2, Proposition 4.2], who show it on \mathbb{R}^d (see also Zemel and Panaretos [134, Theorem 2]). An earlier result in a compact setting can be found in Carlier and Ekeland [33]. The formulation given here is from Masarotto et al. [91]. A more general version is provided by Le Gouic and Loubes [87, Theorem 8].

Lemmata 3.1.11 and 3.1.12 are from [135], but were known earlier (e.g., Bonneel et al. [30]).

Proposition 3.1.6 is a simplified version of Álvarez-Esteban et al. [8, Theorem 2.8] (see [8, Corollary 2.9]).

Propositions 3.2.3 and 3.2.7 are from Bigot and Klein [22], who also show the law of large numbers (Corollary 3.2.10) and deal with the one-dimensional setup (Theorem 3.2.11) in a compact setting. Section 2.4 appears to be new, but see the discussion in its beginning for other measurability results.

Barycentres can be defined for any $p \geq 1$ as the measures minimising $\mu \mapsto \mathbb{E}W_p^p(\Lambda, \mu)$. (Strictly speaking, these are not Fréchet means unless $p = 2$.) Le Gouic and Loubes [87] show Proposition 3.2.3 and Theorem 3.2.9 in this more general setup, where \mathbb{R}^d can be replaced by any separable locally compact geodesic space.

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.

