Electron. Commun. Probab. **29** (2024), article no. 1, 1–13.
https://doi.org/10.1214/23-ECP572
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A note on the adapted weak topology in discrete time

Gudmund Pammer*

Abstract

The adapted weak topology is an extension of the weak topology for stochastic processes designed to adequately capture properties of underlying filtrations. With the recent work of Bart–Beiglböck–P. [7] as starting point, the purpose of this note is to recover with topological arguments the intriguing result by Backhoff–Bartl–Beiglböck–Eder [3] that all adapted topologies in discrete time coincide. We also derive new characterizations of this topology, including descriptions of its trace on the sets of Markov processes and processes equipped with their natural filtration. To emphasize the generality of the argument, we also describe the classical weak topology for measures on $\mathbb{R}^d$ by a weak Wasserstein metric based on the theory of weak optimal transport that was initiated by Gozlan–Roberto–Samson–Tetali [11].

Keywords: adapted weak topology; stochastic processes.
MSC2020 subject classifications: 60G05; 60B10; 49Q22.
Submitted to ECP on May 11, 2022, final version accepted on December 26, 2023.

1 Introduction

An essential difference in the study of random variables and stochastic processes is that the latter comes in conjuction with filtrations that are designed to model the flow of available information: Let us consider a path space $\mathcal{X} := \prod_{t=1}^N \mathcal{X}_t$ equipped with the product topology where $(\mathcal{X}_t, d_{\mathcal{X}_t})$ are Polish metric spaces and $N \in \mathbb{N}$ denotes the number of time steps. We write $\mathcal{P}(\mathcal{X})$ for the set of laws of stochastic processes, i.e., Borel probability measures on $\mathcal{X}$. Canonically, we identify $\mathbb{P} \in \mathcal{P}(\mathcal{X})$ with the process

$$\big(\mathcal{X}, (\sigma(X_{1:t}))_{t=1}^N, \sigma(X), \mathbb{P}, X\big), \tag{1.1}$$

where $X = X_{1:N}$ is the coordinate process on $\mathcal{X}$, $X_{1:t}$ denotes the projection from $\mathcal{X} \to \prod_{s=1}^t \mathcal{X}_s =: \mathcal{X}_{1:t}$, and $\sigma(X_{1:t})$ the σ-algebra generated by $X_{1:t}$. For $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_p(\mathcal{X})$, that are probabilities in $\mathcal{P}(\mathcal{X})$ with finite p-th moment, $p \in [1, \infty)$, the p-Wasserstein distance $\mathcal{W}_p$ is given by

$$\mathcal{W}_p^p(\mathbb{P}, \mathbb{Q}) := \inf_{\pi \in \mathrm{Cpl}(\mathbb{P}, \mathbb{Q})} \mathbb{E}_\pi\big[d_{\mathcal{X}}^p(X, Y)\big], \tag{1.2}$$

where $\mathrm{Cpl}(\mathbb{P}, \mathbb{Q})$ denotes the probabilities on $\mathcal{X} \times \mathcal{X}$ with marginals $\mathbb{P}$ and $\mathbb{Q}$, and $d_{\mathcal{X}}^p(x, y) := \sum_{t=1}^N d_{\mathcal{X}_t}^p(x_t, y_t)$. We endow $\mathcal{P}_p(\mathcal{X})$ with the topology induced by $\mathcal{W}_p$ and $\mathcal{P}(\mathcal{X})$ with the usual weak topology of measures, which is metrized by $\mathcal{W}_p$ when $d_{\mathcal{X}}$ is bounded.

*ETH Zurich, Switzerland. E-mail: gudmund.pammer@math.ethz.ch

The starting point for the study of *adapted* topologies is the fact that probabilistic operations and optimization problems, which crucially depend on filtrations, such as the Doob decomposition, the Snell envelope, optimal stopping, utility maximization, and stochastic programming, are typically not continuous w.r.t. the usual weak topology. These shortcomings are acknowledged by several authors from different communities, see, e.g., [1, 13, 18, 2, 7] for more details. The purpose of this note is to recover and strengthen the main result of Backhoff et al. [3], which states that all adapted topologies on $\mathcal{P}(\mathcal{X})$ coincide. In comparison to the original proof, our argument is more conceptional. At its core lies the elementary fact that comparable compact Hausdorff topologies agree, as shown in Lemma 1.9 below, which is a minor generalization of this observation.

1.1 Stochastic processes and the adapted weak topology

Subsequently, we want to consider topologies that incorporate the flow of information encoded in filtrations, for processes on general filtered probability spaces. Therefore, we follow the approach of [7] by introducing the notion of a *filtered process*.

Definition 1.1 (Filtered Process)**.** *A filtered process* $\mathbf{X}$ *with paths in* $\mathcal{X}$ *is a* 5*-tuplet*

$$\big(\Omega^{\mathbf{X}}, \mathcal{F}^{\mathbf{X}}, (\mathcal{F}^{\mathbf{X}}_t)_{t=1}^N, \mathbb{P}^{\mathbf{X}}, X\big), \tag{1.3}$$

consisting of a complete filtered probability space $(\Omega^{\mathbf{X}}, \mathcal{F}^{\mathbf{X}}, (\mathcal{F}^{\mathbf{X}}_t)_{t=1}^N, \mathbb{P}^{\mathbf{X}})$ *and an* $(\mathcal{F}^{\mathbf{X}}_t)_{t=1}^N$*-adapted stochastic process* X *with paths in* $\mathcal{X}$*. We write* FP *for the class of all filtered processes with paths in* $\mathcal{X}$*, and* FP_p *for the subclass of filtered processes that finitely integrate* $d^p_{\mathcal{X}}(\hat{x}, X)$ *for some* $\hat{x} \in \mathcal{X}$*.*

While a-priori FP is a proper class that contains a lot of redundancy, we will focus on equivalence classes $[\mathbf{X}]$ of filtered processes in the sense of Hoover–Keisler [13]. The corresponding factor space $\mathbf{FP}$ is a set, see, e.g., [6]. This factorization is analogous to classical L^p-theory, where equivalence classes modulo almost-sure equivalence are considered to obtain a Banach space. The equivalence relation between filtered processes can be characterized by an *adapted* version of the Wasserstein distance, known as the *adapted Wasserstein distance* $\mathcal{AW}_p$, as demonstrated in [7, Theorem 1.5]. We will provide a detailed introduction of this distance in Section 1.2. For $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}_p$, we have

$$\mathbf{Y} \in [\mathbf{X}] \iff \mathcal{AW}_p(\mathbf{X}, \mathbf{Y}) = 0,$$

and write $\mathbf{X} \equiv \mathbf{Y}$. Throughout this paper, we will use the same notation for elements in FP and their equivalence classes in $\mathbf{FP}$, denoting both by bold letters.

Henceforth, we consider the factor space $\mathbf{FP}$ and remark that equivalent processes share the same probabilistic properties, e.g. being predictable, being a martingale, having the same Doob decomposition, having the same Snell-envelope, ... Moreover, we write $\mathbf{FP}_p$ for those elements $\mathbf{X} \in \mathbf{FP}$ with $\mathbb{E}_{\mathbb{P}^{\mathbf{X}}}[d^p_{\mathcal{X}}(\hat{x}, X)] < \infty$ for some $\hat{x} \in \mathcal{X}$.

The topology induced by the adapted Wasserstein distance is denoted by $\tau_{\mathcal{AW}}$ and is called the *adapted weak topology*. Equipping $\mathbf{FP}$ with this topology results in a space with rich topological and geometric properties (see [7]). Importantly, we note that the value of $\mathcal{AW}_p(\mathbf{X}, \mathbf{Y})$ (as well as $\mathcal{CW}_p(\mathbf{X}, \mathbf{Y})$, which will be introduced below) is independent of the particular choice of representatives, thanks to the adapted block approximation introduced in [7]. Similarly, we can equip $\mathbf{FP}_p$ with p-th Wasserstein topology by defining

$$\mathcal{W}_p(\mathbf{X}, \mathbf{Y}) := \mathcal{W}_p(\mathscr{L}(X), \mathscr{L}(Y)).$$

However, it is important to remark that $\mathcal{W}_p$ is not point seperating on $\mathbf{FP}_p$, meaning that processes can have the same law but different information structure (as illustrated in [2,

Figure 1]). An essential feature of $\mathcal{AW}_p$ is the following Prokhorov-type result which will be applied at several occasions in the proofs:

Theorem 1.2 (Theorem 1.7 of [7]). *A set $M \subset \mathbf{FP}_p$ is $(\mathbf{FP}_p, \mathcal{AW}_p)$-precompact if and only if M is $(\mathbf{FP}_p, \mathcal{W}_p)$-precompact, that is $\{\mathscr{L}(X) \colon \mathbf{X} \in M\} \subset \mathcal{P}_p(\mathcal{X})$ is precompact.*

To emphasize the significance of Theorem 1.2 and to give the idea behind the main results, we formulate the following immediate corollary:

Corollary 1.3. *Let $d \colon \mathbf{FP}_p \times \mathbf{FP}_p \to \mathbb{R}^+$ be a metric on $\mathbf{FP}_p$ such that*

$$\mathcal{W}_p(\mathbf{X}, \mathbf{Y}) \le d(\mathbf{X}, \mathbf{Y}) \le \mathcal{AW}_p(\mathbf{X}, \mathbf{Y}). \tag{1.4}$$

Then d metrizes the adapted weak topology $\tau_{\mathcal{AW}}$.

Proof. Using (1.4) we need to show that any subsequence $(\mathbf{X}^{k'})_{k' \in \mathbb{N}}$ of a d-convergent sequence $(\mathbf{X}^k)_{k \in \mathbb{N}}$ with limit $\mathbf{X}$ admits an $\mathcal{AW}_p$-convergent subsequence with the same limit. To verify this, note that as $(\mathbf{X}^{k'})_{k' \in \mathbb{N}}$ is d-convergent, it is $\mathcal{W}_p$-precompact and thus, $\mathcal{AW}_p$-precompact by Theorem 1.2. Therefore, there exist $\mathbf{Y} \in \mathbf{FP}_p$ and a further $\mathcal{AW}_p$-convergent subsequence $(\mathbf{X}^{k''})_{k'' \in \mathbb{N}}$ with limit $\mathbf{Y}$. Again by (1.4), this sequence also converges w.r.t. d and the triangle inequality yields $d(\mathbf{X}, \mathbf{Y}) = 0$. Finally, as d is a metric, we get that $\mathbf{X} = \mathbf{Y}$, which concludes the proof. □

Remark 1.4. The conclusion fails for $d = \mathcal{W}_p$ since $\mathcal{W}_p$ does not separate points in $\mathbf{FP}_p$.

1.2 Adapted topologies

In order to capture the properties of filtrations, numerous authors have introduced extensions of the weak topology of measures on $\mathcal{P}(\mathcal{X})$, which we frame in our setting and briefly introduce below. For a thorough overview of the topic and introduction to those topologies we refer to [3] and the references therein.

(A) Aldous [1] introduces the *extended weak topology* τ_{A} by associating a process $\mathbf{X} \in \mathbf{FP}$ with a measure-valued martingale $\mathrm{pp}^1(\mathbf{X})$, the so-called *prediction process*, that is here

$$\mathrm{pp}^1(\mathbf{X}) := \big(\mathscr{L}(X | \mathcal{F}_t^{\mathbf{X}})\big)_{t=1}^N \in \mathcal{P}(\mathcal{X})^N, \tag{1.5}$$

where $\mathscr{L}(X|\mathcal{F}_t^{\mathbf{X}})$ is the conditional law of X given $\mathcal{F}_t^{\mathbf{X}}$. Then τ_{A} is defined as the initial topology induced by $\mathbf{X} \mapsto \mathscr{L}(\mathrm{pp}^1(\mathbf{X}))$ when $\mathcal{P}(\mathcal{P}(\mathcal{X})^N)$ is equipped with the weak topology.

(HK) Hoover–Keisler [13] introduce an increasing sequence of topologies τ_{HK}^r on $\mathbf{FP}$ where $r \in \mathbb{N} \cup \{0, \infty\}$ is called the rank. This is achieved by iterating Aldous' construction of the prediction process. Set $\mathrm{pp}^0(\mathbf{X}) := X$ and, recursively define, for $r \in \mathbb{N} \cup \{\infty\}$,

$$\mathrm{pp}^r(\mathbf{X}) := \big(\mathscr{L}(\mathrm{pp}^{r-1}(\mathbf{X}) | \mathcal{F}_t)\big)_{t=1}^N, \tag{1.6}$$

and $\mathrm{pp}(\mathbf{X}) := \mathrm{pp}^\infty(\mathbf{X})$. Analogously to (A), for $r \in \mathbb{N} \cup \{0, \infty\}$, τ_{HK}^r is given by the initial topology w.r.t. $\mathbf{X} \mapsto \mathscr{L}((\mathrm{pp}^k(\mathbf{X}))_{k=0}^r)$. We remark that τ_{HK}^0 is equivalent to weak convergence of the law, $\tau_{\mathrm{HK}}^1 = \tau_{\mathrm{A}}$, and $\tau_{\mathrm{HK}}^{N-1} = \tau_{\mathrm{HK}}^r$ for $r \ge N$ (see [7]) and simply write then $\tau_{\mathrm{HK}} := \tau_{\mathrm{HK}}^{N-1}$.

(OS) The optimal stopping topology τ_{OS} is defined in [3] as the initial topology w.r.t. the family of maps

$$\mathbf{X} \mapsto \inf \big\{ \mathbb{E}_{\mathbb{P}^{\mathbf{X}}}[c(\rho, X)] \colon \rho \text{ is } (\mathcal{F}_t^{\mathbf{X}})_{t=1}^N\text{-stopping time} \big\}, \tag{1.7}$$

where $c \colon \{1, \dots, N\} \times \mathcal{X} \to \mathbb{R}$ is continuous, bounded, and non-anticipative, that is, if $(t, x), (t, y) \in \{1, \dots, N\} \times \mathcal{X}$ with $x_{1:t} = y_{1:t}$ then $c(t, x) = c(t, y)$.

(H) The information topology τ_{H} of Hellwig [12] is based on a similar point of view as (A) and (HK). Properties of the filtration are encoded in the laws

$$\mathscr{L}\big(X_{1:t}, \mathscr{L}\big(X_{t+1:N}|\mathcal{F}_t^{\mathbf{X}}\big)\big), \quad 1 \le t \le N, \tag{1.8}$$

that are measures on $\mathcal{P}(\mathcal{X}_{1:t} \times \mathcal{P}(\mathcal{X}_{t+1:N}))$ and τ_H is given as the initial topology induced by the map that assigns $\mathbf{X} \in \mathrm{FP}$ the family of distributions in (1.8).

(BLO) Let the path space $\mathcal{X}$ be the N-fold product of a separable Banach space V, i.e., $\mathcal{X} = V^N$. In this setting, Bonnier–Liu–Oberhauser [9] embed $\mathbf{FP}$ into graded linear spaces V_r via higher rank expected signatures, where $r \in \mathbb{N} \cup \{0, \infty\}$ is again the rank, and define τ^r_{BLO} as the initial topology w.r.t. the corresponding embedding $\Phi_r \colon \mathbf{FP} \to V_r$.

Remark 1.5. In case that $d_{\mathcal{X}}$ is an unbounded metric on $\mathcal{X}$, we will fix for the rest of the paper $p \in [1, \infty)$ and consider the subset $\mathbf{FP}_p$ with the following topological adaptation. The topologies (A), (HK), (OS), (H) and (BLO) are then refined by additionally requiring continuity of

$$\mathbf{FP}_p \ni \mathbf{X} \mapsto \mathbb{E}_{\mathbb{P}^{\mathbf{X}}}\big[d^p_{\mathcal{X}}(\hat{x}, X)\big]. \tag{1.9}$$

To avoid notational excess, we state all results on $\mathbf{FP}_p$ for some $p \in [1, \infty)$. All results are also true when replacing $\mathbf{FP}_p$ with $\mathbf{FP}$ (and if necessary $d_{\mathcal{X}}$ with, for example, $d_{\mathcal{X}} \wedge 1$).

Besides using the powerful concept of initial topologies, various authors have constructed adapted topologies based on ideas from optimal transportation. The essence of this approach is to encode filtrations into constraints for the set of couplings and thereby construct modifications of the Wasserstein distance suitable for processes. To illustrate the idea, recall that optimal transport has so-called transport maps $T \colon \mathcal{X} \to \mathcal{X}$ at its core, satisfying the push-forward condition $T_\#\mathbb{P} = \mathbb{Q}$ for $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{X})$. We refer to [20] for a comprehensive overview on optimal transport. In our context, where $\mathbb{P}$ and $\mathbb{Q}$ are laws of processes, *causal optimal transport* suggests to use *adapted* maps in order to transport $\mathbb{P}$ to $\mathbb{Q}$, i.e., $T_\#\mathbb{P} = \mathbb{Q}$ and T is non-anticipative, which means

$$T(X) = \big(T_1(X_1), T_2(X_{1:2}), \ldots, T_N(X)\big).$$

When X resp. Y denote the first resp. second coordinate projection from $\mathcal{X} \times \mathcal{X} \to \mathcal{X}$, then this additional adaptedness constraint on couplings can be formulated as

$$\mathrm{Cpl}_{\mathrm{c}}(\mathbb{P}, \mathbb{Q}) := \big\{\pi \in \mathrm{Cpl}(\mathbb{P}, \mathbb{Q}) \colon X \perp_{X_{1:t}} Y_{1:t} \text{ under } \pi \text{ for } t = 1, \ldots, N-1\big\}, \tag{1.10}$$

where, for σ-algebras $\mathcal{A}, \mathcal{B}, \mathcal{C}$ on some probability space, $\mathcal{A} \perp_{\mathcal{B}} \mathcal{C}$ denotes conditional independence of $\mathcal{A}$ and $\mathcal{C}$ given $\mathcal{B}$. Elements of $\mathrm{Cpl}_{\mathrm{c}}(\mathbb{P}, \mathbb{Q})$ are called causal couplings. When one symmetrizes (1.10) one obtains the set of *bicausal couplings* $\mathrm{Cpl}_{\mathrm{bc}}(\mathbb{P}, \mathbb{Q})$, that are $\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mathbb{P}, \mathbb{Q})$ such that $(Y, X)_\#\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mathbb{Q}, \mathbb{P})$. These definitions can be easily extended to FP, which consists of filtered processes.

Definition 1.6 (Causal and Bicausal Couplings)**.** *Let $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}$. For $s, t \in \{0, \ldots, N\}$ we denote by $\mathcal{F}^{\mathbf{X},\mathbf{Y}}_{s,t}$ the σ-algebra on $\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Y}}$ given by $\mathcal{F}^{\mathbf{X}}_s \otimes \mathcal{F}^{\mathbf{Y}}_t$ under the convention that $\mathcal{F}^{\mathbf{X}}_0$ and $\mathcal{F}^{\mathbf{Y}}_0$ are the corresponding trivial σ-algebras. A probability π on the measurable space $(\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Y}}, \mathcal{F}^{\mathbf{X}} \otimes \mathcal{F}^{\mathbf{Y}})$ is called causal if, under π,*

$$\mathcal{F}^{\mathbf{X},\mathbf{Y}}_{N,0} \perp_{\mathcal{F}^{\mathbf{X},\mathbf{Y}}_{t,0}} \mathcal{F}^{\mathbf{X},\mathbf{Y}}_{0,t}. \tag{1.11}$$

We call π bicausal if it additionally satisfies

$$\mathcal{F}^{\mathbf{X},\mathbf{Y}}_{0,N} \perp_{\mathcal{F}^{\mathbf{X},\mathbf{Y}}_{0,t}} \mathcal{F}^{\mathbf{X},\mathbf{Y}}_{t,0}. \tag{1.12}$$

Finally, we write $\mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})$ resp. $\mathrm{Cpl}_{\mathrm{bc}}(\mathbf{X}, \mathbf{Y})$ for the set of causal resp. bicausal probabilities with first marginal $\mathbb{P}^{\mathbf{X}}$ and second marginal $\mathbb{P}^{\mathbf{Y}}$.

Remark 1.7. Even though the sets defined in Definition 1.6 depend on the specific filtered processes $\mathbf{X}, \mathbf{Y}$, we can show the following: If $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}_p$, $\mathbf{X}' \in [\mathbf{X}]$, $\mathbf{Y}' \in [\mathbf{Y}]$, and $\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})$ resp. $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mathbf{X}, \mathbf{Y})$, then there is, for every $\epsilon > 0$, a coupling $\pi' \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}', \mathbf{Y}')$ resp. $\pi' \in \mathrm{Cpl}_{\mathrm{bc}}(\mathbf{X}', \mathbf{Y}')$ such that $\mathcal{W}_p((X,Y)_\# \pi, (X', Y')_\# \pi') < \epsilon$. This follows from the adapted block approximation (c.f. [7, Theorem A.4]) and the fact that gluing preserves causality (see Corollary 2.12).

(SCW) Lassalle [17] and Backhoff et al. [4] coin the notion of causality, see Definition 1.6, and introduce the causal Wasserstein "distance" $\mathcal{CW}_p$ on $\mathcal{P}_p(\mathcal{X})$. For $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}_p$ we define

$$\mathcal{CW}_p^p(\mathbf{X}, \mathbf{Y}) := \inf_{\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})} \mathbb{E}_\pi \big[d_{\mathcal{X}}^p(X, Y) \big], \tag{1.13}$$

and note that by Remark 1.7, (1.13) is independent of the choice of representative. By abuse of notation, we view $\mathcal{CW}_p$ from now on as a function on $\mathbf{FP}_p \times \mathbf{FP}_p$. Clearly, $\mathcal{CW}_p$ is not a metric as it lacks symmetry, which motivates to consider the so-called *symmetrized causal Wasserstein distance*, see [3],

$$\mathcal{SCW}_p(\mathbf{X}, \mathbf{Y}) := \max \big\{ \mathcal{CW}_p(\mathbf{X}, \mathbf{Y}), \mathcal{CW}_p(\mathbf{Y}, \mathbf{X}) \big\}, \tag{1.14}$$

which constitutes a metric on $\mathbf{FP}_p$. We write $\tau_{\mathcal{SCW}}$ for the induced topology.

(AW) Instead of symmetrizing as in (1.14), one can directly symmetrize the definition on the level of couplings via the notion of bicausal couplings. Approaches in this spirit but to different extents go back to Rüschendorf [19], Pflug–Pichler [18], Bion-Nadal–Talay [8], and Bartl et al. [7]. We define the adapted Wasserstein distance of $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}_p$

$$\mathcal{AW}_p^p(\mathbf{X}, \mathbf{Y}) := \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mathbf{X}, \mathbf{Y})} \mathbb{E}_\pi \big[d_{\mathcal{X}}^p(X, Y) \big], \tag{1.15}$$

and, as above, view $\mathcal{AW}_p$ as a function on $\mathbf{FP}_p \times \mathbf{FP}_p$. The adapted Wasserstein distance is a metric on $\mathbf{FP}_p$, and we denote its induced topology by $\tau_{\mathcal{AW}}$.

(CW) Finally, we introduce here a new mode of convergence, called the topology of causal convergence, denoted by $\tau_{\mathcal{CW}}$. A neighbourhood basis for $\tau_{\mathcal{CW}}$ at a fixed $\mathbf{X} \in \mathbf{FP}_p$ is given by sets of the form $\big\{ \mathbf{Y} \in \mathbf{FP}_p \colon \mathcal{CW}_p(\mathbf{X}, \mathbf{Y}) < \epsilon \big\}$, where $\epsilon > 0$. In other words, $\tau_{\mathcal{CW}}$ can be equivalently described by

$$\mathbf{X}^k \to \mathbf{X} \text{ in } \tau_{\mathcal{CW}} \iff \mathcal{CW}_p(\mathbf{X}, \mathbf{X}^k) \to 0. \tag{1.16}$$

Remark 1.8. It is apparent from the definitions in (1.2), (1.13), (1.14) and (1.15) that

$$\mathcal{W}_p(\mathbf{X}, \mathbf{Y}) \le \mathcal{CW}_p(\mathbf{X}, \mathbf{Y}) \le \mathcal{SCW}_p(\mathbf{X}, \mathbf{Y}) \le \mathcal{AW}_p(\mathbf{X}, \mathbf{Y}), \tag{1.17}$$

for $\mathbf{X}, \mathbf{Y} \in \mathbf{FP}_p$. Hence, we have $\tau_{\mathcal{W}} \subseteq \tau_{\mathcal{CW}} \subseteq \tau_{SCW} \subseteq \tau_{\mathcal{AW}}$.

1.3 Characterizations of the adapted weak topology

In this subsection we formulate the main results of this paper. We recall that a topological space is said to be sequential if every sequentially closed subset is closed, where a subset A is said to be sequentially closed if every sequence $(x_n)_{n \in \mathbb{N}}$ in A that converges, has its limit in A. The core ingredient in order to prove the main results, Theorems 1.10 and 1.12, and also Proposition 1.13, is the following simple observation:

Lemma 1.9. *Let $(\mathcal{A}, \tau'), (\mathcal{A}, \tau)$ be topological spaces that satisfy the following:*

(1) $(\mathcal{A}, \tau')$ and $(\mathcal{A}, \tau)$ are sequential topological spaces.

(2) The topology τ is a least as fine as τ', that is $\tau \supseteq \tau'$.

(3) If $M \subset \mathcal{A}$ is $(\mathcal{A}, \tau')$-precompact then M is $(\mathcal{A}, \tau)$-precompact.

(4) $(\mathcal{A}, \tau')$ is Hausdorff.

Then we have $(\mathcal{A}, \tau) = (\mathcal{A}, \tau')$.

Note that Lemma 1.9 in combination with Theorem 1.2 have Corollary 1.3 as a consequence. Next, we provide characterizations of the adapted weak topology on $\mathbf{FP}_p$.

The equivalence of τ_{HK} and the adapted Wasserstein-topology, $\tau_{\mathcal{AW}}$, is due to [7] whereas the characterization in terms of the symmetric causal Wasserstein-topology, $\tau_{\mathcal{SCW}}$, is novel. Moreover, we remark that the equivalence of the higher rank expected signature-topology, τ_{BLO} and τ_{HK} was already known when, for $t \in \{1, \ldots, N\}$, $\mathcal{X}_t = V$ and V is a compact subset of a separable Banach space, see [9, Theorem 2].

Theorem 1.10. *On $\mathbf{FP}_p$, we have*

$$\tau_{\mathrm{HK}} = \tau_{\mathcal{SCW}} = \tau_{\mathcal{AW}}. \tag{1.18}$$

If $\mathcal{X}_t = \mathbb{R}^d$, $1 \le t \le N$, then these topologies also coincide with $\tau^{N-1}_{\mathrm{BLO}}$, and $\tau^r_{\mathrm{BLO}} = \tau^r_{\mathrm{HK}}$.

When restricting to sets of processes that have simpler information structure, such as Markov processes or processes equipped with their natural filtration, there are simpler ways to characterize the adapted weak topology. This motivates the definition of higher-order Markov processes where the transition probabilities are allowed to depend the current state as well as on past states.

Definition 1.11. *Let $n \in \mathbb{N} \cup \{\infty\}$. We call a process $\mathbf{X} \in \mathbf{FP}_p$ n-th order Markovian (or n-th order Markov process) if, for all $1 \le t \le N$,*

$$\mathscr{L}(X_{t+1}|\mathcal{F}^{\mathbf{X}}_t) = \mathscr{L}(X_{t+1}|X_{1\vee(t-n):t}) \quad \text{almost surely.} \tag{1.19}$$

The set of all n-th order Markov processes is denoted by $\mathbf{FP}^{\mathrm{Markov}}_{p,n}$. Moreover, we may call ∞-th order Markov processes plain and write $\mathbf{FP}^{\mathrm{plain}}_p := \mathbf{FP}^{\mathrm{Markov}}_{p,\infty}$.

We equip $\mathbf{FP}^{\mathrm{Markov}}_{p,n}$ with the initial topology τ^n_{Markov} induced by the family of maps $\mathbf{X} \mapsto \mathscr{L}(T^n_t(X)) \in \mathcal{P}_p(\mathcal{X}_{1\vee(t-n+1):t} \times \mathcal{P}_p(\mathcal{X}_{t+1}))$ for $1 \le t \le N-1$, where

$$T^n_t(X) := \big(X_{1\vee(t-n+1):t}, \mathscr{L}(X_{t+1}|\mathcal{F}^{\mathbf{X}}_t)\big). \tag{1.20}$$

The next result recovers and generalizes the main result of [3]. The novelty of the next result is two-fold: On the one hand, the case $n = \infty$ recovers the results of [3] and additionally gives a new description in terms of $\tau^\infty_{\mathrm{Markov}}$. On the other hand, the case $n \in \mathbb{N}$ extends this result to the subset of n-th order Markov processes.

Theorem 1.12 (All Adapted Topologies are Equal)**.** *Let $n, r \in \mathbb{N} \cup \{\infty\}$. Then the trace on $\mathbf{FP}^{\mathrm{Markov}}_{p,n}$ of the topologies τ_{A}, τ^r_{HK}, τ_{OS}, τ_{H}, $\tau_{\mathcal{CW}}$, $\tau_{\mathcal{SCW}}$ and $\tau_{\mathcal{AW}}$ are the same. In particular, they all coincide with τ^n_{Markov}.*

1.4 Characterizations of weak topologies

Proposition 1.13 shows that Lemma 1.9 applies beyond the framework of the adapted weak topology. Specifically, the p-Wasserstein topology on $\mathcal{P}_p(\mathbb{R}^d)$ can be metrized by

$$\mathcal{V}^p_p(\mathbb{P}, \mathbb{Q}) := \max\big\{V_p(\mathbb{P}, \mathbb{Q}), V_p(\mathbb{Q}, \mathbb{P})\big\}, \tag{1.21}$$

where $\mathbb{R}^d$ is equipped with the Euclidean norm $|\cdot|$, and

$$V^p_p(\mathbb{P}, \mathbb{Q}) := \inf_{\pi \in \mathrm{Cpl}(\mathbb{P}, \mathbb{Q})} \mathbb{E}_\pi\Big[\big|\mathbb{E}_\pi\big[X - Y\big|X\big]\big|^p\Big]. \tag{1.22}$$

Proposition 1.13. *The p-Wasserstein topology on $\mathcal{P}_p(\mathbb{R}^d)$ is metrized by $\mathcal{V}_p$.*

The minimization problem in (1.22) falls within the framework of *weak optimal transport*, which is a generalization of optimal transport. In particular, (1.22) vanishes if and only if there exists a martingale coupling between $\mathbb{P}$ and $\mathbb{Q}$. For more information on this topic, we refer to [11, 5] and the references therein.

Likewise, this point of view can be applied in order to recover [14, Theorem 2.8]: Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. We denote by τ_W and τ_S, the weak resp. strong topology on $\mathcal{P}(\mathcal{X})$, that is the topology induced by the mappings $\mu \mapsto \int f\,d\mu$ where $f\colon \mathcal{X} \to [0,1]$ is continuous resp. measurable. Similarly, we write τ_{SW} for the initial topology on $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ induced by the family of mappings $\mu \mapsto \int f\,d\mu$, where $f\colon \mathcal{X}\times\mathcal{Y} \to [0,1]$ is measurable and $y \mapsto f(x,y)$ is continuous for every $x \in \mathcal{X}$.

Lemma 1.14. *Let $\Pi \subseteq \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ and $\{\mathrm{pj}_{\mathcal{X}}\pi \colon \pi \in \Pi\}$ be τ_S-sequentially precompact, then τ_{SW} and τ_W coincide on Π, that is $(\Pi,\tau_{SW}) = (\Pi,\tau_W)$.*

2 Proofs

To prove the main results, we first verify the assumptions of Lemma 1.9. In the process, we utilize a variant of a well-known fact on martingales: if $X = (X_1, X_2, X_3)$ is a martingale, then

$$X_1 \sim X_3 \implies X_1 = X_2 = X_3 \quad \text{almost surely.} \tag{2.1}$$

Here, $X_1 \sim X_3$ means that X_1 and X_3 are identically distributed. We recall that a process $X = (X_t)_{t=1}^N$ taking values in $\mathcal{P}(\mathcal{X})$ is called a measure-valued martingale with values in $\mathcal{P}(\mathcal{X})$ if, for $f \in C_b(\mathcal{X})$, the real-valued, bounded process $(X_t(f))_{t=1}^N$ is an $(\sigma(X_{1:t}))_{t=1}^N$-martingale. For $\rho \in \mathcal{P}(\mathcal{X})$, we write $\rho(f)$ to denote the integral $\int f d\rho$.

Lemma 2.1. *Let $X = (X_1, X_2, X_3)$ be a measure-valued martingale taking values in $\mathcal{P}(\mathcal{X})$, where $\mathcal{X}$ is a Polish space. If $X_1 \sim X_3$, then $X_1 = X_2 = X_3$ almost surely.*

Proof. Since there exists a countable family in $C_b(\mathcal{X})$ that separates points in $\mathcal{P}(\mathcal{X})$, it suffices to show, for $f \in C_b(\mathcal{X})$, $X_1(f) = X_2(f) = X_3(f)$ a.s., which follows from (2.1). □

2.1 Properties of $\mathbf{FP}_{p,n}^{\mathrm{Markov}}$

First, we justify that the n-Markov property is preserved under equivalence.

Lemma 2.2. *Let $n \in \mathbb{N}\cup\{\infty\}$, and $\mathbf{X}, \mathbf{Y} \in \mathrm{FP}$ with $\mathbf{X} \equiv \mathbf{Y}$. Then $\mathbf{X}$ is n-Markovian if and only if $\mathbf{Y}$ is n-Markovian.*

Proof. By Definition 1.11 the property of being n-Markovian can be deduced from observing the law of the corresponding first-order prediction process. Hence, we conclude by the fact that $\mathbf{X} \equiv \mathbf{Y}$ readily implies $\mathscr{L}(\mathrm{pp}^1(\mathbf{X})) = \mathscr{L}(\mathrm{pp}^1(\mathbf{Y}))$. □

Lemma 2.3. *If $\mathbf{X} \in \mathbf{FP}_p^{plain}$, $\mathbf{Y} \in \mathbf{FP}_p$ and $\mathscr{L}(X) = \mathscr{L}(Y)$, then $\mathcal{CW}_p(\mathbf{Y},\mathbf{X}) = 0$. In particular, if additionally $\mathbf{Y} \in \mathbf{FP}_p^{plain}$, then $\mathbf{X} = \mathbf{Y}$.*

Proof. We work with a representative of $\mathbf{Y}$. As Y_t is $\mathcal{F}_t^{\mathbf{Y}}$-measurable, the coupling $\pi := (\mathrm{id}_{\Omega^{\mathbf{Y}}}, Y)_\#\mathbb{P}^{\mathbf{Y}}$ is causal from $\mathbf{Y}$ to $\mathbf{X} := (\mathcal{X}, \sigma(X_{1:t})_t, \sigma(X), \mathscr{L}(Y), X)$, where X denotes the canonical process on $\mathcal{X}$. If $\mathbf{Y}$ is plain, we have by Definition 1.11 that $\mathscr{L}(Y|\mathcal{F}_t^{\mathbf{Y}}) = \mathscr{L}(Y|Y_{1:t})$ $\mathbb{P}^{\mathbf{Y}}$-a.s., which translate to $X \perp_{X_{1:t}} \mathcal{F}_{t,0}^{\mathbf{Y},\mathbf{X}}$ under π as $X = Y$ π-a.s. Thus, π is bicausal and $\mathcal{AW}_p(\mathbf{X},\mathbf{Y}) = 0$. □

Corollary 2.4. *Given $n, m \in \mathbb{N} \cup \{\infty\}$ with $n \le m$, we have $\mathbf{FP}_{p,n}^{\mathrm{Markov}} \subseteq \mathbf{FP}_{p,m}^{\mathrm{Markov}}$. Moreover, processes in $\mathbf{FP}_{p,n}^{\mathrm{Markov}}$ are uniquely defined by their law, that is, for $\mathbf{X}, \mathbf{Y} \in \mathbf{FP}_{p,n}^{\mathrm{Markov}}$*

$$\mathscr{L}(X) = \mathscr{L}(Y) \implies \mathbf{X} = \mathbf{Y}.$$

Proof. The first claim is a direct consequence of the definition of n-th resp. m-th order Markov processes. The second claim then readily follows from Lemma 2.3. □

Lemma 2.5. *$(\mathbf{FP}_{p,n}^{\mathit{Markov}}, \tau^n_{\mathit{Markov}})$ is a sequential Hausdorff space.*

Proof. First, we remark that, for $1 \le t \le N-1$, the map $\mathbf{X} \mapsto \mathscr{L}(T_t^n(X))$ takes values in the Polish (and therefore first countable) space $\mathcal{P}_p(\mathcal{X}_{1\vee(t-n+1):t} \times \mathcal{P}_p(\mathcal{X}_{t+1}))$, thus $\tau_{n,\mathrm{Markov}}$ is sequential.

Next, let $\mathbf{X}, \mathbf{Y} \in \mathbf{FP}_{p,n}^{\mathrm{Markov}}$ with $\mathscr{L}(T_t^n(X)) = \mathscr{L}(T_t^n(Y))$ for $1 \le t \le N-1$. Consequently, we find $\mathscr{L}(X_{1\vee(t-n+1):t+1}) = \mathscr{L}(Y_{1\vee(t-n+1):t+1})$ and the existence of a measurable map $f_t \colon \mathcal{X}_{1\vee(t-n+1):t} \to \mathcal{P}(\mathcal{X}_{t+1})$ such that almost surely

$$f_t(X_{1\vee(t-n+1):t}) = \mathscr{L}(X_{t+1}|X_{1\vee(t-n+1):t}), \quad f_t(Y_{1\vee(t-n+1):t}) = \mathscr{L}(Y_{t+1}|Y_{1\vee(t-n+1):t}).$$

In particular, we have for $t = n$ that $\mathscr{L}(X_{1:n+1}) = \mathscr{L}(Y_{1:n+1})$.

We proceed to show $\mathscr{L}(X) = \mathscr{L}(Y)$. Assume that we have already shown $\mathscr{L}(X_{1:t}) = \mathscr{L}(Y_{1:t})$ for some $n+1 \le t \le N-1$. By the disintegration theorem and the definition of n-th order Markovian, we may write

$$\mathscr{L}(X_{1:t+1}) = \mathscr{L}(X_{1:t}) \otimes f_t(X_{t-n+1:t}) = \mathscr{L}(Y_{1:t}) \otimes f_t(Y_{t-n+1:t}) = \mathscr{L}(Y_{1:t+1}),$$

where we use the notation $\mu \otimes k$ for $\mu \in \mathcal{P}(\mathcal{X}_{1:t})$ and a measurable kernel $k \colon \mathcal{X}_{1:t} \to \mathcal{P}(\mathcal{X}_{t+1})$ to denote the gluing of μ with k, that is the probability defined by

$$\mu \otimes k(A \times B) = \textstyle\int_A k(x, B)\, \mu(dx) \quad A \in \mathcal{B}(X_{1:t}), B \in \mathcal{B}(X_{t+1}).$$

This concludes the inductive step.

Finally, we can apply Lemma 2.3 and conclude $\mathbf{X} = \mathbf{Y}$. □

Lemma 2.6. *Let $\mathbf{X}, \mathbf{Y} \in \mathbf{FP}_p$ with $\mathcal{CW}_p(\mathbf{X}, \mathbf{Y}) = \mathcal{CW}_p(\mathbf{Y}, \mathbf{X}) = 0$, then $\mathbf{X} = \mathbf{Y}$. In particular, $\mathcal{SCW}_p(\mathbf{X}, \mathbf{Y}) = 0$ if and only if $\mathcal{AW}_p(\mathbf{X}, \mathbf{Y}) = 0$.*

Proof. As the values of $\mathcal{CW}_p$, $\mathcal{SCW}_p$ and $\mathcal{AW}_p$ are independent of the representatives of $\mathbf{X}$ and $\mathbf{Y}$, we choose representatives such that $\mathcal{F}_N^{\mathbf{X}} = \mathcal{F}^{\mathbf{X}}$, $\mathcal{F}_N^{\mathbf{Y}} = \mathcal{F}^{\mathbf{Y}}$, and $(\Omega^{\mathbf{X}}, \mathcal{F}^{\mathbf{X}})$ and $(\Omega^{\mathbf{Y}}, \mathcal{F}^{\mathbf{Y}})$ are standard Borel spaces. This is possible by [7, Subsection 1.3] and, in this case, one has by standard arguments that the values of $\mathcal{CW}_p$, $\mathcal{SCW}_p$ and $\mathcal{AW}_p$ are attained. Let $\pi \in \mathrm{Cpl_c}(\mathbf{X}, \mathbf{Y})$ and $\pi' \in \mathrm{Cpl_c}(\mathbf{Y}, \mathbf{X})$ with $X = Y$ π- and π'-almost surely. This allows us to consider the conditionally independent product of π and π' denoted by $\hat\pi := \pi \dot\otimes \pi' \in \mathrm{Cpl}(\mathbf{X}, \mathbf{Y}, \mathbf{X})$, see Definition 2.9. Here, $\mathrm{Cpl}(\mathbf{X}, \mathbf{Y}, \mathbf{X})$ denotes the set of couplings with marginals $\mathbb{P}^{\mathbf{X}}$, $\mathbb{P}^{\mathbf{Y}}$ and $\mathbb{P}^{\mathbf{X}}$. In order to distinguish the coordinates, we will write $\tilde{\mathbf{X}}$ resp. $\tilde X$ for the second $\mathbf{X}$-coordinate. By induction we show that

$$\mathrm{pp}^k(\mathbf{X}) = \mathrm{pp}^k(\mathbf{Y}) = \mathrm{pp}^k(\tilde{\mathbf{X}}) \quad \hat\pi\text{-a.s.}, \tag{2.2}$$

for all $k \in \mathbb{N} \cup \{0\}$. Since we know that $X = Y = \tilde X$ $\hat\pi$-almost surely, we have verified (2.2) when $k = 0$. Assume that (2.2) holds for some k. By causality of π' and Lemma 2.11 we find, for $1 \le t \le N$,

$$\mathcal{F}_{0,N,0}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}} \perp_{\mathcal{F}_{0,t,0}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}} \mathcal{F}_{0,0,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}, \quad \mathcal{F}_{N,0,0}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}} \perp_{\mathcal{F}_{t,0,0}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}} \mathcal{F}_{0,t,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}, \tag{2.3}$$

where we naturally extend the notation introduced in Definition 1.6 in order to write products of multiple σ-algebras. Since $\mathrm{pp}^k(\mathbf{X})$ and $\mathrm{pp}^k(\mathbf{Y})$ are measurable w.r.t. $\mathcal{F}_N^{\mathbf{X}}$ and $\mathcal{F}_N^{\mathbf{Y}}$ resp., we obtain by combining (2.2), (2.3), and the tower property

$$\mathrm{pp}_t^{k+1}(\tilde{\mathbf{X}}) = \mathbb{E}\Big[\mathscr{L}\big(\mathrm{pp}^k(\mathbf{Y})|\mathcal{F}_{0,t,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}\big)|\mathcal{F}_{0,0,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}\Big] = \mathbb{E}\Big[\mathrm{pp}_t^{k+1}(\mathbf{Y})|\mathcal{F}_{0,0,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}\Big],$$

and similarly, $\mathrm{pp}_t^{k+1}(\mathbf{Y}) = \mathbb{E}[\mathrm{pp}_t^{k+1}(\mathbf{X})|\mathcal{F}_{0,t,t}^{\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}}}]$. Hence, we conclude the inductive step by observing that the triplet $(\mathrm{pp}_t^{k+1}(\tilde{\mathbf{X}}), \mathrm{pp}_t^{k+1}(\mathbf{Y}), \mathrm{pp}_t^{k+1}(\mathbf{X}))$ satisfies the assumptions of Lemma 2.1. In particular, we have shown that $\mathscr{L}(\mathrm{pp}(\mathbf{X})) = \mathscr{L}(\mathrm{pp}(\mathbf{Y}))$, whence $\mathbf{X} = \mathbf{Y}$ by [7, Theorem 4.11]. □

Proposition 2.7. *Let $n \in \mathbb{N} \cup \{\infty\}$ and $M \subseteq \mathbf{FP}_{p,n}^{\mathit{Markov}}$ be τ_{Markov}^n-precompact. Then M is precompact in $(M, \tau_{\mathcal{AW}})$.*

Proof. Let $(\mathbf{X}^k)_{k\in\mathbb{N}}$ be a τ_{Markov}^n-converging sequence in $\mathbf{FP}_{p,n}^{\mathrm{Markov}}$ with limit $\mathbf{X} \in \mathbf{FP}_{p,n}^{\mathrm{Markov}}$. First, we convince ourselves that $(\mathscr{L}(X^k))_{k\in\mathbb{N}}$ converges to $\mathscr{L}(X)$. Assume that we have already shown that $\mathscr{L}(X_{1:t}^k) \to \mathscr{L}(X_{1:t})$ for some $1 \le t \le N-1$. The conditionally independent product $\dot{\otimes}$, see [10, Definition 2.8], allows us to rewrite

$$\mathscr{L}(X_{1:t}, \mathscr{L}(X_{t+1}|X_{1:t})) = \mathscr{L}(X_{1:t})\dot{\otimes}\mathscr{L}(T_t^n(X)).$$

By [10, Theorem 4.1], that reads in our context as continuity of $\dot{\otimes}$ at $(\mathscr{L}(X_{1:t}), \mathscr{L}(T_t^n(X)))$, we obtain that $\mathscr{L}(X_{1:t+1}^k) \to \mathscr{L}(X_{1:t+1})$.

Hence, $(\mathscr{L}(X^k))_{k\in\mathbb{N}}$ is convergent and therefore tight. Thus, there exists by Theorem 1.2 a subsequence of $(\mathbf{X}^k)_{k\in\mathbb{N}}$ converging in $\tau_{\mathcal{AW}}$ to some $\mathbf{Y} \in \mathbf{FP}_p$. Due to $\tau_{\mathcal{AW}}$-continuity, we get $\mathscr{L}(T_t^n(X)) = \lim_{j\to\infty} \mathscr{L}(T_t^n(X^{k_j})) = \mathscr{L}(T_t^n(Y))$. Hence, there exist measurable maps $f_t \colon \mathcal{X}_{1\vee(t-n+1):t} \to \mathcal{P}(\mathcal{X}_{t+1})$ with the property

$$f_t(Y_{1\vee(t-n+1):t}) = \mathscr{L}(Y_{t+1}|\mathcal{F}_t^{\mathbf{Y}}) \quad \text{almost surely.}$$

In other words, $\mathbf{Y} \in \mathbf{FP}_{p,n}^{\mathrm{Markov}}$. Therefore the sequence $(\mathbf{X}^k)_{k\in\mathbb{N}}$ is also precompact in $(\mathbf{FP}_{p,n}^{\mathrm{Markov}}, \tau_{\mathcal{AW}})$, which concludes the proof. □

Proposition 2.8. *Let $M \subseteq \mathbf{FP}_p^{\mathrm{plain}}$ be precompact in $(\mathbf{FP}_p^{\mathrm{plain}}, \tau_{\mathcal{CW}})$. Then M is precompact in $(\mathbf{FP}_p^{\mathrm{plain}}, \tau_{\mathcal{AW}})$.*

Proof. Let M be precompact in $(\mathbf{FP}_p^{\mathrm{plain}}, \tau_{\mathcal{CW}})$. Since $\tau_{\mathcal{W}} \subseteq \tau_{\mathcal{CW}}$, there exists by Theorem 1.2 a $\tau_{\mathcal{AW}}$-convergent subsequence with limit $\mathbf{Y}$ for some $\mathbf{Y} \in \mathbf{FP}_p$. Since $\mathcal{CW}_p$ is by Lemma 2.13 (1-Lipschitz) continuous w.r.t. $\mathcal{AW}_p$, we find $\mathcal{CW}_p(\mathbf{X}, \mathbf{Y}) = \lim_j \mathcal{CW}_p(\mathbf{X}, \mathbf{X}^{k_j}) = 0$, and conclude by Lemma 2.3 that $\mathbf{Y}(= \mathbf{X}) \in \mathbf{FP}_p^{\mathrm{plain}}$. □

2.2 Causal gluing

This section is devoted to develop auxiliary results concerning the composition of causal couplings with matching intermediary marginal. We recall that due to [7] we can always assume w.l.o.g. that all spaces under consideration are standard Borel. Therefore, we assume for the rest of the section that we have chosen representatives of $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbf{FP}$ such that $(\Omega^{\mathbf{X}}, \mathcal{F}_N^{\mathbf{X}}), (\Omega^{\mathbf{Y}}, \mathcal{F}_N^{\mathbf{Y}})$, and $(\Omega^{\mathbf{Z}}, \mathcal{F}_N^{\mathbf{Z}})$ are standard Borel and $\mathcal{F}^{\mathbf{X}} = \mathcal{F}_N^{\mathbf{X}}, \mathcal{F}^{\mathbf{Y}} = \mathcal{F}_N^{\mathbf{Y}}$, and $\mathcal{F}^{\mathbf{Z}} = \mathcal{F}_N^{\mathbf{Z}}$.

Definition 2.9. *Let $\gamma \in \mathrm{Cpl}(\mathbf{X}, \mathbf{Y})$ and $\eta \in \mathrm{Cpl}(\mathbf{Y}, \mathbf{Z})$. We define the* conditionally independent product *of γ and η as the probability on $\big(\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Y}} \times \Omega^{\mathbf{Z}}, \mathcal{F}_{N,N,N}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}\big)$ satisfying for any U, bounded and $\mathcal{F}_{N,N,N}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}$-measurable, that*

$$\int U \, d\gamma\dot{\otimes}\eta = \int\int U(\omega^{\mathbf{X}}, \omega^{\mathbf{Y}}, \omega)\, \eta_{\omega^{\mathbf{Y}}}(d\omega^{\mathbf{Z}})\, \gamma(d\omega^{\mathbf{X}}, d\omega^{\mathbf{Y}}), \tag{2.4}$$

where $\eta_{\omega^{\mathbf{Y}}}$ is a $\mathcal{F}^{\mathbf{Y}}_N$-measurable kernel satisfying η-almost surely $\eta_{\omega^{\mathbf{Y}}} = \mathscr{L}_\eta(\omega^{\mathbf{Z}}|\mathcal{F}^{\mathbf{Y},\mathbf{Z}}_{N,0})$. Due to symmetry reasons, we have

$$\int U \, d\gamma\dot{\otimes}\eta = \iint U(\omega^{\mathbf{X}}, \omega^{\mathbf{Y}}, \omega^{\mathbf{Z}}) \, (\eta_{\omega^{\mathbf{Y}}} \otimes \gamma_{\omega^{\mathbf{Y}}})(d\omega^{\mathbf{X}}, d\omega^{\mathbf{Z}}) \, \mathbb{P}^{\mathbf{Y}}(d\omega^{\mathbf{Y}}). \tag{2.5}$$

The term (2.5) clarifies the naming of $\gamma\dot{\otimes}\eta$ as the conditional independent product: conditionally on $\omega^{\mathbf{Y}}$ the knowledge of $\omega^{\mathbf{X}}$ does not affect $\omega^{\mathbf{Z}}$ and vice versa. This suggests the following probabilistic formulation.

Lemma 2.10. *Let $\gamma \in \mathrm{Cpl}(\mathbf{X}, \mathbf{Y})$ and $\eta \in \mathrm{Cpl}(\mathbf{Y}, \mathbf{Z})$, and let $\mathcal{G}$ be a σ-algebra with $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,N,0} \subseteq \mathcal{G} \subseteq \mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,N,0}$. Then, we have $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,N,0} \perp_{\mathcal{G}} \mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,N,N}$ under $\gamma\dot{\otimes}\eta$.*

Proof. The assertion follows from (2.5) coupled with [15, Proposition 5.8]. □

Lemma 2.11. *Let $\gamma \in \mathrm{Cpl}(\mathbf{X}, \mathbf{Y})$ and $\eta \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{Y}, \mathbf{Z})$. We have, for $1 \le t \le N$,*

(1) under $\gamma\dot{\otimes}\eta$: $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,N,0} \perp_{F^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,t,0}} \mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,t}$;

if furthermore $\gamma \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})$, then we have

(2) under $\gamma\dot{\otimes}\eta$: $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,0,0} \perp_{F^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,0,0}} \mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,t}$.

Proof. To show item (1), let W be bounded and $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,t}$ measurable. From Lemma 2.10, we derive the first equality in

$$\mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,N,0}\right] = \mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,N,0}\right] = \mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,0}\right], \tag{2.6}$$

whereas the second stems from causality of η. Here, this causality yields that under $\gamma\dot{\otimes}\eta$, conditionally on $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,0}$, $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,N,0}$ is independent of $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{0,t,t}$. Since the last term in (2.6) is $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,t,0}$-measurable, the tower property yields item (1).

To establish item (2), let W be as above. Note that causality of γ provides that, conditionally on $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,0,0}$, $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,0,0}$ is independent of $\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,t,0}$ under $\gamma\dot{\otimes}\eta$. Using that in addition to item (1) and the tower property, we conclude

$$\begin{aligned}\mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,0,0}\right] &= \mathbb{E}_{\gamma\dot{\otimes}\eta}\left[\mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,t,0}\right]|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{N,0,0}\right] \\ &= \mathbb{E}_{\gamma\dot{\otimes}\eta}\left[\mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,t,0}\right]|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,0,0}\right] = \mathbb{E}_{\gamma\dot{\otimes}\eta}\left[W|\mathcal{F}^{\mathbf{X},\mathbf{Y},\mathbf{Z}}_{t,0,0}\right].\end{aligned}$$ □

Corollary 2.12. *Let $\gamma \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})$ and $\eta \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{Y}, \mathbf{Z})$. Writing $\mathrm{pj}_{\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Z}}}$ for the projection onto $\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Z}}$, we have $(\mathrm{pj}_{\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Z}}})_\# \gamma\dot{\otimes}\eta \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Z})$.*

Proof. This result is a direct consequence of item (2) of Lemma 2.11. □

Lemma 2.13. *Let $\mathbf{X} \in \mathbf{FP}_p$. The map $\mathbf{FP}_p \ni \mathbf{Y} \mapsto \mathcal{CW}_p(\mathbf{X}, \mathbf{Y})$ is 1-Lipschitz w.r.t. $\mathcal{SCW}_p$.*

Proof. Let $\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Y})$ and $\pi' \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{Y}, \mathbf{Z})$, then $(\mathrm{pj}_{\Omega^{\mathbf{X}} \times \Omega^{\mathbf{Z}}})_\# \pi\dot{\otimes}\pi' \in \mathrm{Cpl}_{\mathrm{c}}(\mathbf{X}, \mathbf{Z})$ by Corollary 2.12. Hence, we compute by Minkowski's inequality

$$\mathcal{CW}_p(\mathbf{X}, \mathbf{Z}) \le \left(\mathbb{E}_{\pi\dot{\otimes}\pi'}\left[d^p_{\mathcal{X}}(X, Z)\right]\right)^{\frac{1}{p}} \le \left(\mathbb{E}_{\pi}\left[d^p_{\mathcal{X}}(X, Y)\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}_{\pi'}\left[d^p_{\mathcal{X}}(Y, Z)\right]\right)^{\frac{1}{p}},$$

and conclude $\left|\mathcal{CW}_p(\mathbf{X}, \mathbf{Z}) - \mathcal{CW}_p(\mathbf{X}, \mathbf{Y})\right| \le \mathcal{SCW}_p(\mathbf{Y}, \mathbf{Z})$. □

2.3 Postponed proofs of Section 1

Proof of Lemma 1.9. Due to (2) it remains to show that convergence in $(\mathcal{A},\tau')$ implies convergence in $(\mathcal{A},\tau)$. To this end, let $(y^k)_{k\in\mathbb{N}}$ be a sequence in $(\mathcal{A},\tau')$ converging to y. By (3) we find a subsequence $(y^{k_j})_{j\in\mathbb{N}}$ that converges in $(\mathcal{A},\tau)$ to some element z. Again by (2), we have that $(y^{k_j})_{j\in\mathbb{N}}$ also converges in $(\mathcal{A},\tau')$ to z, which yields by (4) that $y=z$. Therefore, y is the only $(\mathcal{A},\tau)$-accumulation point of $(y^k)_{k\in\mathbb{N}}$, from where we conclude that $(y^k)_{k\in\mathbb{N}}$ has to converge to y in $(\mathcal{A},\tau)$. □

Proof of Theorem 1.10. It is evident from [7, Theorem 3.10], [7, Lemma 4.7] and [7, Lemma 4.10] that $\tau_{\mathcal{AW}}$ and τ_{HK} coincide.

Using the notation of Lemma 1.9, we let $(\mathcal{B},\tau) := (\mathbf{FP}_p,\tau_{\mathcal{AW}})$ and $\mathcal{A}=\mathcal{B}$. By Remark 1.8, resp. [9, Proposition 6] we have for $\tau'\in\{\tau_{\mathcal{SCW}},\tau_{\mathrm{BLO}}^{N-1}\}$ that $\tau'\subseteq\tau$. By Lemma 2.6, resp. [9, Theorem 4] we find that τ' is Hausdorff. Moreover, we obtain $\tau_{\mathcal{W}}\subseteq\tau'$ from Remark 1.8 resp. [9, Proposition 8], where $\tau_{\mathcal{W}}$ is the topology of p-Wasserstein convergence of the laws. Since $\tau_{\mathcal{W}}$ and τ have the same precompact sets by Theorem 1.2, we conclude the same for τ'. Hence, all assumptions of Lemma 1.9 are met, which yields the first two assertions of the theorem.

The last assertion of the theorem follows mutatis mutandis. □

Proof of Theorem 1.12. Let $\mathcal{A}:=\mathbf{FP}_{p,n}^{\mathrm{Markov}}$, $\mathcal{B}:=\mathbf{FP}_p$, and $\tau=\tau_{\mathcal{AW}}$.

It is evident (either by construction, from Theorem 1.10, or from [3, Lemma 7.5]) that τ_{Markov}^n is coarser than τ_{H}, τ_{A}, τ_{HK}^r, τ_{OS}, τ_{AW} and τ_{SCW}. Similarly, we have that all of these topologies are coarser than $\tau_{\mathcal{AW}}$. We remark that $\tau_{\mathcal{AW}}\supseteq\tau_{\mathrm{OS}}$ can be seen due to the fact that the map which maps $\mathbf{X}\in\mathbf{FP}_p$ to its Snell envelope is $\tau_{\mathcal{AW}}$-continuous.

Thus, it suffices to show that $(\mathcal{A},\tau')=(\mathcal{A},\tau)$ for $\tau'\in\{\tau_{\mathrm{Markov}}^n,\tau_{\mathcal{CW}}\}$. We proceed by verifying the assumptions in Lemma 1.9: Item (1) follows from Lemma 2.5 resp. is evident by construction. Item (2) is satisfied, since it is easy to see that $\tau_{\mathrm{H}}\subseteq\tau_{\mathrm{HK}}^1\subseteq\tau_{\mathrm{HK}}=\tau_{\mathrm{AW}}$ (where the last equality is due to Theorem 1.10) resp. by Remark 1.8. Item (3) is proven in Proposition 2.7 resp. Proposition 2.8. Finally, item (4) is due to Corollary 2.4 resp. Lemma 2.3. □

Proof of Proposition 1.13. Let $\mathcal{A}=\mathcal{B}=\mathcal{P}_p(\mathbb{R}^d)$ and $\tau=\tau_{\mathcal{W}}$. It is straightforward to check that $\mathcal{V}_p$ is a pseudometric and $\mathcal{V}_p\le\mathcal{W}_p$. Moreover, as a simple consequence of (2.1) we find that $\mathcal{V}_p$ separates points: If $\mathcal{V}_p(\mathbb{P},\mathbb{Q})=0$, then there exist couplings $\pi\in\mathrm{Cpl}(\mathbb{P},\mathbb{Q})$ and $\tilde\pi\in\mathrm{Cpl}(\mathbb{Q},\mathbb{P})$ with $x=\int y\,\pi_x(dy)$ π-a.s. and $x=\int y\,\tilde\pi_x(dy)$ $\tilde\pi$-a.s. Let $X=(X_t)_{t=1}^3$ be a Markov process with $(X_1,X_2)\sim\pi$ and $(X_2,X_3)\sim\pi'$. Thus, X is a martingale in its generated filtration where initial and terminal distribution coincide. By (2.1), we find $X_1=X_2=X_3$, hence, $\mathbb{P}=\mathbb{Q}$ and $\mathcal{V}_p$ is a metric on $\mathcal{P}_p(\mathbb{R}^d)$. We write $\tau_{\mathcal{V}}$ for the topology induced by $\mathcal{V}_p$ and get $\tau_{\mathcal{V}}\subseteq\tau_{\mathcal{W}}$. It remains to verify Item (3) of Lemma 1.9.

To this end, let $(\mathbb{P}^k)_{k\in\mathbb{N}}$ converge to $\mathbb{P}$ in $\tau_{\mathcal{V}}$ and we want to show $\mathcal{W}_p$-relative compactness of the sequence. By [5, Lemma 6.1], we have

$$V_p(\mathbb{P}^k,\mathbb{P})=\inf_{\mathbb{Q}\le_{\mathrm{cx}}\mathbb{P}}\mathcal{W}_p(\mathbb{P}^k,\mathbb{Q}), \tag{2.7}$$

where $\le_{\mathrm{cx}}$ denotes the convex order on $\mathcal{P}_1(\mathbb{R}^d)$. Recall that, for $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^d)$, $\mu\le_{\mathrm{cx}}\nu$ if and only if $\int f\,d\mu\le\int f\,d\nu$ for all convex $f\colon\mathbb{R}^d\to\mathbb{R}$. Due to compactness of closed balls of finite radius in $\mathbb{R}^d$, it is easy to see, for example, by the De la Vallée-Poussin theorem for uniform integrability and [20, Definition 6.8], that the set $\{\mathbb{Q}\le_{\mathrm{cx}}\mathbb{P}\}$ is $\mathcal{W}_p$-compact in $\mathcal{P}_p(\mathbb{R}^d)$. Hence, we find by standard arguments the existence of $\mathbb{Q}^k\le_{\mathrm{cx}}\mathbb{P}$ attaining (2.7). Consequently,

$$\lim_{k\to\infty}V_p(\mathbb{P},\mathbb{P}^k)=\lim_{k\to\infty}\mathcal{W}_p(\mathbb{P}^k,\mathbb{Q}^k)=0,$$

which in particular yields $\mathcal{W}_p$-relative compactness of $\{\mathbb{P}^k\colon k\in\mathbb{N}\}$. □

Proof of Lemma 1.14. We need to show $(\Pi, \tau_W) = (\Pi, \tau_f)$, where τ_f is the initial topology w.r.t. $\mu \mapsto \int g \, d\mu$ for $g \in C_b(\mathcal{X} \times \mathcal{Y}) \cup \{f\}$ and $f \colon \mathcal{X} \times \mathcal{Y} \to [0,1]$ is measurable with $f(x, \cdot) \in C_b(\mathcal{Y})$ for every $x \in \mathcal{X}$.

Assume that $\mathcal{Y}$ is compact, then $(C_b(\mathcal{Y}), |\cdot|_\infty)$ is Polish. As $F \colon \mathcal{X} \to C_b(\mathcal{Y}) \colon x \mapsto f(x, \cdot)$ is measurable, there is by [16, Theorem 13.11] a Polish topology $\hat{\tau}^{\mathcal{X}}$, leaving the Borel sets unchanged, so that F is $\hat{\tau}^{\mathcal{X}}$-continuous. Therefore, $f \in C_b(\hat{\mathcal{X}} \times \mathcal{Y})$, where $\hat{\mathcal{X}}$ denotes $\mathcal{X}$ equipped with $\hat{\tau}^{\mathcal{X}}$. As $\{\mathrm{pj}_{\mathcal{X}} \pi \colon \pi \in \Pi\}$ is τ_S-sequentially precompact, it is precompact in $\mathcal{P}(\hat{\mathcal{X}})$, and by Prokhorov's theorem, if $A \subseteq \Pi$ is τ_W-precompact then A is τ_f-precompact.

For general $\mathcal{Y}$, there is for any τ_W-precompact set $A \subseteq \Pi$ an increasing sequence of compacts $(\mathcal{Y}_n)_{n \in \mathbb{N}}$ with $\inf_{\mu \in A} \mu(\mathcal{Y}_n) \to 1$. We obtain that A_n is τ_f-precompact, where $A_n := \{\mu|_{\mathcal{X} \times \mathcal{Y}_n} / \mu(\mathcal{X} \times \mathcal{Y}_n) : \mu \in A\}$. Thus, a sequence $(\mu_k)_{k \in \mathbb{N}}$ in A admits $(\mu_{k_j})_{j \in \mathbb{N}}$ so that, for every $n \in \mathbb{N}$, $(\mu_{k_j}|_{\mathcal{X} \times \mathcal{Y}_n} / \mu_{k_j}(\mathcal{X} \times \mathcal{Y}_n))_{j \in \mathbb{N}}$ is τ_f-convergent. Hence, $(\mu_{k_j})_{j \in \mathbb{N}}$ is τ_f-convergent, A is τ_f-precompact, and we conclude by Lemma 1.9, $(\Pi, \tau_W) = (\Pi, \tau_f)$. □

References

[1] David J. Aldous, *Weak convergence and general theory of processes*, Unpublished monograph: Department of Statistics, University of California, Berkeley, July 1981.

[2] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder, *Adapted Wasserstein distances and stability in mathematical finance*, Finance Stoch. **24** (2020), no. 3, 601–632. MR4118990

[3] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder, *All adapted topologies are equal*, Probab. Theory Relat. Fields **178** (2020), no. 3-4, 1125–1172. MR4168395

[4] Julio Backhoff-Veraguas, Mathias Beiglböck, Yiqing Lin, and Anastasiia Zalashko, *Causal transport in discrete time and applications*, SIAM Journal on Optimization **27** (2017), no. 4, 2528–2562. MR3738324

[5] Julio Backhoff-Veraguas, Mathias Beiglböck, and Gudmund Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calculus of Variations and Partial Differential Equations **58** (2019), no. 6, 1–28. MR4029731

[6] Julio Backhoff-Veraguas, Mathias Beiglböck, and Gudmund Pammer, *Weak monotone rearrangement on the line*, Electronic Communications in Probability **25** (2020). MR4069738

[7] Daniel Bartl, Mathias Beiglböck, and Gudmund Pammer, *The Wasserstein space of stochastic processes*, arXiv:2104.14245 (2023).

[8] Jocelyne Bion-Nadal and Denis Talay, *On a Wasserstein-type distance between solutions to stochastic differential equations*, Ann. Appl. Probab. **29** (2019), no. 3, 1609–1639. MR3914552

[9] Patric Bonnier, Chong Liu, and Harald Oberhauser, *Adapted topologies and higher rank signatures*, Ann. Appl. Probab. **33** (2023), no. 3, 2136–2175. MR4583667

[10] Manu Eder, *Compactness in adapted weak topologies*, arXiv:1905.00856 (2019).

[11] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. MR3706606

[12] M. F. Hellwig, *Sequential decisions under uncertainty and the maximum theorem*, J. Math. Econom. **25** (1996), no. 4, 443–464. MR1394333

[13] Douglas N. Hoover and H. Jerome Keisler, *Adapted probability distributions*, Transactions of the American Mathematical Society **286** (1984), no. 1, 159–201. MR0756035

[14] Jean Jacod and Jean Mémin, *Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité*, Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French), Lecture Notes in Math., vol. 850, Springer, Berlin-New York, 1981, pp. 529–546. MR622586

[15] O. Kallenberg, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, 1997. MR1464694

[16] A. S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, New York, 1995. MR1321597

[17] Rémi Lassalle, *Causal transport plans and their Monge–Kantorovich problems*, Stoch. Anal. Appl. **36** (2013), 452–484. MR3784142

[18] G. Ch. Pflug and A. Pichler, *A distance for multistage stochastic optimization models*, SIAM J. Optim. **22** (2012), no. 1, 1–23. MR2902682

[19] L. Rüschendorf, *Fréchet-bounds and their applications*, Advances in probability distributions with given marginals (Rome, 1990), Math. Appl., vol. 67, Kluwer Acad. Publ., Dordrecht, 1991, pp. 151–187. MR1215951

[20] C. Villani, *Optimal transport. old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, 2009. MR2459454