A Some elementary facts

We first list some useful inequalities for ψ_p and ψ_p^{-1}. Note that the estimates may not be the sharpest, but they suffice for our needs.

Proposition A.2. *For $p \geq 1$ and $x \geq 0$, let $\psi_p(x) = \exp(x^p) - 1$ and let $\psi_p^{-1}(x) = (\log(x+1))^{1/p}$ be its inverse. Then we have the following:*

(i) $\psi_p^2(x/2^{1/p}) \leq \psi_p(x)$.

(ii) $x\psi_p(x/4^{1/p}) \leq 2^{1/p}\psi_p(x/2^{1/p})$.

(iii) for $x \geq 0$ and $q \geq 1$, $\psi_p^{-1}(x^q) \leq q^{1/p}\psi_p^{-1}(x)$.

(iv) For $x \geq 1$, $\psi_p^{-1}(x) \leq (\log(x))^{1/p} + 1$.

Proof.

(i) For any $x \geq 0$,

$$\begin{aligned}\psi_p(x) &= \exp(x^p) - 1 = (\exp(x^p/2) - 1)(\exp(x^p/2) + 1) \geq (\exp(x^p/2) - 1)^2 \\ &= \psi_p^2(x/2^{1/p}).\end{aligned}$$

(ii) We only need to consider the case $x \geq 1$ since otherwise the inequality is obvious. Since $y \leq 2(\exp(y/4) + 1)$ for all $y \geq 1$, we have

$$x \leq 2^{1/p}(\exp(x^p/4) + 1)^{1/p} \leq 2^{1/p}(\exp(x^p/4p) + 1) \leq 2^{1/p}(\exp(x^p/4) + 1).$$

Then

$$\begin{aligned}x\psi_p(x/4^{1/p}) &= x(\exp(x^p/4) - 1) \\ &\leq 2^{1/p}(\exp(x^p/4) + 1)(\exp(x^p/4) - 1) \\ &= 2^{1/p}(\exp(x^p/2) - 1) \\ &= 2^{1/p}\psi_p(x/2^{1/p}).\end{aligned}$$

(iii) Since $x \geq 0$ and $q \geq 1$,

$$\psi_p^{-1}(x^q) = (\log(1 + x^q))^{1/p} \leq (\log(1 + x)^q)^{1/p} = q^{1/p}\psi_p^{-1}(x).$$

(iv) When $x \geq 1$,

$$\mathrm{e}x \geq x + 1 \implies \log x + 1 \geq \log(x + 1) \implies \log^{1/p}(x) + 1 \geq \psi_p^{-1}(x).$$

□

The following simple result is for converting between sums and integrals:

Proposition A.3. *For any $r \geq 2$, $K \in \mathbb{N}$, and a continuous nonincreasing $f : (0, +\infty) \to (0, +\infty)$, we have*

$$\sum_{k=1}^{K} r^{-k} f(r^{-k}) \leq r \int_0^1 f(\varepsilon)\, \mathrm{d}\varepsilon \leq r^2 \sum_{k=0}^{\infty} r^{-k} f(r^{-k}) \tag{A.1}$$

Proof. Using the monotonicity of f, we have

$$\begin{aligned}\sum_{k=1}^{K} r^{-k} f(r^{-k}) &\leq \sum_{k=1}^{K} r^{-k}(r-1) f(r^{-k}) \leq r \sum_{k=1}^{K} \int_{r^{-k-1}}^{r^{-k}} f(\varepsilon)\, \mathrm{d}\varepsilon \\ &\leq r \int_0^1 f(\varepsilon) \mathrm{d}\varepsilon \leq r \sum_{k=0}^{\infty} \int_{r^{-k}}^{r^{-k+1}} f(\epsilon)\, \mathrm{d}\varepsilon \leq r^2 \sum_{k=0}^{\infty} r^{-k} f(r^{-k}).\end{aligned}$$

□

B Omitted proofs

B.1 Proofs for Section 2

Proof of Proposition 1. It follows from the inequality $x\log(x+1) \le x\log x + 1$, $x \ge 0$, that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}+1\Big) \le \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\frac{\mathrm{d}\mu}{\mathrm{d}\nu}+1.$$

Using this and Jensen's inequality, we get

$$\begin{aligned}
\Big\langle \mu, \psi_p^{-1}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big)\Big\rangle &= \Big\langle \mu, \Big(\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}+1\Big)\Big)^{1/p}\Big\rangle \\
&\le \Big(\Big\langle \mu, \log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}+1\Big)\Big\rangle\Big)^{1/p} \\
&= \Big(\Big\langle \nu, \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}+1\Big)\Big\rangle\Big)^{1/p} \\
&\le \Big(\Big\langle \nu, \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big\rangle+1\Big)^{1/p} \\
&= \big(D(\mu\|\nu)+1\big)^{1/p}.
\end{aligned}$$

B.2 Proofs for Section 3

Proof of Lemma 1. To prove (4), we start with the Young-type inequality

$$xy \le \psi_p^*(x) + \psi_p(y), \qquad x, y \ge 0$$

where

$$\psi_p^*(x) := \sup_{y\ge 0}\big(xy - \psi_p(y)\big)$$

is the (one-sided) Legendre–Fenchel conjugate of ψ_p. While a closed-form expression for ψ_p^* is not available, we claim that we can bound it from above as $\psi_p^*(x) \le 2^{1/p}x\psi_p^{-1}(x)$, resulting in

$$xy \le 2^{1/p}x\psi_p^{-1}(x) + \psi_p(y). \tag{B.1}$$

To establish the claim, we write

$$\sup_{y\ge 0}\big(xy - \psi_p(y)\big) = \sup_{y\ge 0}\big(xy - (e^{y^p/2}-1)(e^{y^p/2}+1)\big)$$

and consider two cases:

- if $y \le 2^{1/p}\psi_p^{-1}(x)$, then
$$xy - (e^{y^p/2}-1)(e^{y^p/2}+1) \le 2^{1/p}x\psi_p^{-1}(x).$$
- if $y > 2^{1/p}\psi_p^{-1}(x)$, then
$$xy - (e^{y^p/2}-1)(e^{y^p/2}+1) \le (e^{y^p/2}-1)(y-(e^{y^p/2}+1)) \le 0.$$

Applying (B.1) with $x = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ and $y = g$ gives

$$g\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \le 2^{1/p}\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\psi_p^{-1}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big) + \psi_p(g),$$

so that

$$\begin{aligned}
\langle \mu, fg\rangle &= \Big\langle \nu, fg\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big\rangle \\
&\le \Big\langle \nu, \Big(2^{1/p}f\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\psi_p^{-1}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big) + f\psi_p(g)\Big)\Big\rangle \\
&= 2^{1/p}\Big\langle \mu, f\psi_p^{-1}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big)\Big\rangle + \langle \nu, f\psi_p(g)\rangle.
\end{aligned}$$

To prove (5), define the event

$$E := \left\{ \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \geq \frac{\exp(g^p/4) - 1}{\langle \nu, \exp(g^p)\rangle} \right\}.$$

Then, since $\langle \nu, \exp(g^p)\rangle \geq 1$,

$$\begin{aligned}
\int_E fg \,\mathrm{d}\mu &\leq 4^{1/p} \int f\Big(\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\langle \nu, \exp(g^p)\rangle + 1\Big)\Big)^{1/p} \mathrm{d}\mu \\
&\leq 4^{1/p} \int f\Big(\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu} + 1\Big) + \log\langle \nu, \exp(g^p)\rangle\Big)^{1/p} \mathrm{d}\mu \\
&\leq 4^{1/p} \int f\Big(\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu} + 1\Big)\Big)^{1/p} \mathrm{d}\mu + 4^{1/p}\int f\,\mathrm{d}\mu \cdot \Big(\log\langle \nu, \exp(g^p)\rangle\Big)^{1/p} \\
&= 4^{1/p}\Big\langle \mu, f\psi_p^{-1}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big)\Big\rangle + 4^{1/p}\|f\|_{L^1(\mu)}\big(\log\langle \nu, \exp(g^p)\rangle\big)^{1/p}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_{E^c} fg\,\mathrm{d}\mu &\leq \int fg \frac{\exp(g^p/4) - 1}{\langle \nu, \exp(g^p)\rangle}\,\mathrm{d}\nu \\
&\leq 2^{1/p}\int f \frac{\exp(g^p/2)}{\langle \nu, \exp(g^p)\rangle}\,\mathrm{d}\nu \\
&\leq 2^{1/p}\|f\|_{L^2(\nu)},
\end{aligned}$$

where the first inequality is by the definition of E, the second inequality follows from Proposition A.2(ii), and the third inequality is by Cauchy–Schwarz. Putting everything together, we get (5).

B.3 Proofs for Section 4

Proof of Theorem 1. It follows from the independence of $Z_1, \ldots, Z_n$ that $\mathrm{gen}(w, S)$ is $(\sigma/\sqrt{n})$-subgaussian, so

$$\mathbf{E}\left[\psi_2\left(\frac{|\mathrm{gen}(w,S)|}{\sigma\sqrt{6/n}}\right)\right] \leq 1, \qquad \forall w \in \mathcal{W}. \tag{B.2}$$

Using Lemma 1 with $\mu = P_{W|S}$, $\nu = Q_W$, $f(w) = \sigma\sqrt{6/n}$, and $g(w) = \frac{|\mathrm{gen}(w,S)|}{\sigma\sqrt{6/n}}$, we have

$$\langle P_{W|S}, |\mathrm{gen}(\cdot, S)|\rangle \leq \sqrt{\frac{12\sigma^2}{n}}\left(\left\langle P_{W|S}, \psi_2^{-1}\left(\frac{\mathrm{d}P_{W|S}}{\mathrm{d}Q_W}\right)\right\rangle + \left\langle Q_W, \psi_2\left(\frac{|\mathrm{gen}(\cdot, S|}{\sigma\sqrt{6/n}}\right)\right\rangle\right).$$

Taking expectations of both sides w.r.t. P_S and using Fubini's theorem and (B.2), we get (6).

Proof of Corollary 1. Applying Proposition 1 conditionally on S gives

$$\Big\langle P_{W|S}, \psi_2^{-1}\Big(\frac{\mathrm{d}P_{W|S}}{\mathrm{d}Q_W}\Big)\Big\rangle \leq \sqrt{D(P_{W|S}\|Q_W) + 1},$$

where the divergence $D(P_{W|S}\|Q_W)$, being a function of S, is a random variable. Substituting this into (6) and using Jensen's inequality, the definition of conditional divergence, and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \leq \sqrt{2(a+b)}$, we get

$$\mathbf{E}[|\mathrm{gen}(W,S)|] \leq \sqrt{\frac{24\sigma^2}{n}\Big(D(P_{W|S}\|Q_W|P_S) + 4\Big)}.$$

Taking the infimum of both sides w.r.t. Q_W and using (2), we get (7).

Proof of Theorem 2. For each fixed $(w, \tilde{s})$, the random variable $\delta(w, \tilde{s}, \varepsilon) := |\sum_{i=1}^n \varepsilon_i(\ell(w, z_i') - \ell(w, z_i))|$ is $\sigma(w, \tilde{s})$-subgaussian, where $\sigma(w, \tilde{s}) := \big(\sum_{i=1}^n (\ell(w, z_i') - \ell(w, z_i))^2\big)^{1/2}$. Thus,

$$\mathbf{E}_\varepsilon[\zeta(w, \tilde{s}, \varepsilon)] := \mathbf{E}_\varepsilon\left[\psi_2\left(\frac{\delta(w, \tilde{s}, \varepsilon)}{\sqrt{6}\sigma(w, \tilde{s})}\right)\right] \le 1, \qquad \forall (w, \tilde{s}). \tag{B.3}$$

Applying Lemma 1 conditionally on $(\tilde{S}, \varepsilon)$ with $\mu = \bar{P}_{W|\tilde{S}\varepsilon}$, $\nu = Q_{W|\tilde{S}}$, $f(w) = \sigma(w, \tilde{S})$, $g(w) = \zeta(w, \tilde{S}, \varepsilon)$, we obtain

$$\begin{aligned}
&\big\langle \bar{P}_{W|\tilde{S}\varepsilon}, \sigma(\cdot, \tilde{S})\zeta(\cdot, \tilde{S}, \varepsilon)\big\rangle \\
&\le \sqrt{2}\left\langle \bar{P}_{W|\tilde{S}\varepsilon}, \sigma(\cdot, \tilde{S})\psi_2^{-1}\left(\frac{\mathrm{d}\bar{P}_{W|\tilde{S},\varepsilon}}{\mathrm{d}Q_{W|\tilde{S}}}\right)\right\rangle + \Big\langle Q_{W|\tilde{S}}, \sigma(\cdot, \tilde{S})\psi_2(\zeta(\cdot, \tilde{S}, \varepsilon))\Big\rangle \\
&\le \sqrt{2}\|\Delta(\tilde{S})\|_{\ell^2}\left(\left\langle \bar{P}_{W|\tilde{S}\varepsilon}, \psi_2^{-1}\left(\frac{\mathrm{d}\bar{P}_{W|\tilde{S},\varepsilon}}{\mathrm{d}Q_{W|\tilde{S}}}\right)\right\rangle + \Big\langle Q_{W|\tilde{S}}, \psi_2(\zeta(\cdot, \tilde{S}, \varepsilon))\Big\rangle\right).
\end{aligned}$$

Taking expectations of both sides w.r.t. $\tilde{S}$ and ε, then using Fubini's theorem, (B.3), and the inequality $\mathbf{E}_P[|\mathrm{gen}(W, S)|] \le \frac{1}{n}\mathbf{E}_{\bar{P}}[\delta(W, \tilde{S}, \varepsilon)]$, we obtain (8).

Proof of Corollary 2. For any $Q_{W|\tilde{S}}$, using Proposition 1, Cauchy–Schwarz, and the independence of (Z_i', Z_i), we have

$$\begin{aligned}
&\mathbf{E}_{\bar{P}}\left[\|\Delta(\tilde{S})\|_{\ell^2}\psi_2^{-1}\left(\frac{\mathrm{d}\bar{P}_{W|\tilde{S},\varepsilon}}{\mathrm{d}Q_{W|\tilde{S}}}\right)\right] \\
&\le \sqrt{\mathbf{E}_{\bar{P}}[\|\Delta(\tilde{S})\|^2_{\ell^2}]\big(D(\bar{P}_{W|\tilde{S}\varepsilon}\|Q_{W|\tilde{S}}|\bar{P}_{\tilde{S}\varepsilon}) + 1\big)} \\
&= \sqrt{n\mathbf{E}[\Delta(Z, Z')^2]\big(D(\bar{P}_{W|\tilde{S}\varepsilon}\|Q_{W|\tilde{S}}|\bar{P}_{\tilde{S}\varepsilon}) + 1\big)}.
\end{aligned}$$

Substituting this estimate into (8), taking the infimum of both sides w.r.t. $Q_{W|\tilde{S}}$, and using (3), we get (9).

B.4 Proofs for Section 5

Proof of Theorem 3. Let

$$\begin{aligned}
\delta(u, v, z, z') &:= \big(\ell(u, z') - \ell(v, z')\big) - \big(\ell(u, z) - \ell(v, z)\big), \\
\delta(u, v, \tilde{s}) &:= \sum_{i=1}^n \delta(u, v, z_i, z_i'), \\
\zeta(u, v, \tilde{s}) &:= \frac{|\delta(u, v, \tilde{s})|}{\sqrt{6}\sigma(u, v, \tilde{s})}.
\end{aligned}$$

For each fixed $(u, v) \in \mathcal{W}^2$, $\delta(u, v, Z_i, Z_i')$, $1 \le i \le n$, are i.i.d. symmetric random variables. Therefore, introducing a tuple $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ of i.i.d. Rademacher random variables independent of everything else and using the fact that the joint distributions of $\big(\delta(u, v, Z_i, Z_i')\big)_{i=1}^n$ and $\big(\varepsilon_i\delta(u, v, Z_i, Z_i')\big)_{i=1}^n$ are the same, we see that

$$\mathbf{E}[\psi_2(\zeta(u, v, \tilde{S}))] = \mathbf{E}_{\tilde{S}}\mathbf{E}_\varepsilon\left[\psi_2\left(\frac{|\sum_{i=1}^n \varepsilon_i\delta(u, v, Z_i, Z_i')|}{\sqrt{6}\sigma(u, v, \tilde{S})}\right)\right] \le 1,$$

where the inequality follows from the fact that, conditionally on S and S', the random variables $\sum_{i=1}^n \varepsilon_i\delta(u, v, Z_i, Z_i')$ are $\sigma(u, v, \tilde{S})$-subgaussian.

Now, given $Q_W \in \mathcal{P}(\mathcal{W})$ and a family of couplings $P_{UV|S=s} \in \Pi(P_{W|S=s}, Q_W)$, it follows from the above definitions and from (10) that

$$\mathbf{E}[\mathrm{gen}(W, S)] \le \frac{1}{n}\mathbf{E}[|\delta(U, V, \tilde{S})|] = \frac{\sqrt{6}}{n}\mathbf{E}[\sigma(U, V, \tilde{S})\zeta(U, V, \tilde{S})]. \tag{B.4}$$

Picking any $\rho_{UV} \in \mathcal{P}(\mathcal{W} \times \mathcal{W})$ such that $P_{UV|S=s} \ll \rho_{UV}$ for all $s \in \mathcal{Z}^n$ and applying Lemma 1, we get

$$\begin{aligned}&\langle P_{UV|S}, \sigma(\cdot, \tilde{S})\zeta(\cdot, \tilde{S})\rangle \\ &\qquad \le 2\Big\langle P_{UV|S}, \sigma(\cdot, \tilde{S})\psi_2^{-1}\Big(\frac{\mathrm{d}P_{UV|S}}{\mathrm{d}\rho_{UV}}\Big)\Big\rangle + \sqrt{2}\Big\langle \rho_{UV}, \sigma(\cdot, \tilde{S})\psi_2\Big(\frac{\zeta(\cdot, \tilde{S})}{\sqrt{2}}\Big)\Big\rangle.\end{aligned}$$

Using the inequality $\psi_2^2(x/\sqrt{2}) \le \psi_2(x)$ (see Proposition A.2(i)), Cauchy–Schwarz, and (B.4), we have

$$\mathbf{E}\Big[\sigma(u, v, \tilde{S})\psi_2\Big(\frac{\zeta(u, v, \tilde{S})}{\sqrt{2}}\Big)\Big] \le \sqrt{\mathbf{E}[\sigma^2(u, v, \tilde{S})]}, \qquad \forall (u, v) \in \mathcal{W} \times \mathcal{W}.$$

Putting everything together and taking expectations w.r.t. S and S', we obtain (12).

Proof of Corollary 3. For σ defined in Theorem 3, we have

$$\sigma^2(u, v, \tilde{S}) \le 2\sum_{i=1}^n \Big(\big(\ell(u, Z_i') - \ell(v, Z_i')\big)^2 + \big(\ell(u, Z_i) - \ell(v, Z_i)\big)^2\Big).$$

Taking conditional expectations given U, V, S and using Jensen's inequality gives

$$\begin{aligned}\mathbf{E}[\sigma(U, V, \tilde{S})|U, V, S] &\le \sqrt{\mathbf{E}[\sigma^2(U, V, \tilde{S})|U, V, S]} \\ &\le \sqrt{2n}\big(d_\ell(U, V) + d_{S,\ell}(U, V)\big).\end{aligned}$$

An analogous argument gives

$$\sqrt{\mathbf{E}[\sigma^2(\bar{U}, \bar{V}, \tilde{S})|\bar{U}, \bar{V}]} \le 2\sqrt{n}d_\ell(\bar{U}, \bar{V}).$$

Substituting these estimates into (12) gives the desired result.

B.5 Proofs for Section 6

Proof of Theorem 5. Using the definition of $\bar{\ell}$, we have

$$\mathbf{E}[\mathrm{gen}(W, S)] = \frac{1}{n}\sum_{k=1}^K \mathbf{E}\Bigg[\sum_{i=1}^n \big(\bar{\ell}(W_k, Z_i) - \bar{\ell}(W_{k-1}, Z_i)\big)\Bigg].$$

Applyng Lemma 1 conditionally on S with $f(u, v) = d(u, v)$, $g(u, v) = \frac{|\sum_{i=1}^n (\bar{\ell}(u,, Z_i) - \bar{\ell}(v, Z_i))|}{\sqrt{n}d(u,v)}$, $\mu = P_{W_k W_{k-1}|S}$ and $\nu = \rho_{W_k W_{k-1}}$, taking expectations w.r.t. P_S, and using (13) gives the desired result.

Proof of Corollary 4. For each $k \ge 1$, let $\rho_{W_k W_{k-1}} = P_{W_k W_{k-1}}$. Then

$$\frac{\mathrm{d}P_{W_k W_{k-1}|S}}{\mathrm{d}P_{W_k W_{k-1}}} = \frac{\mathrm{d}P_{W_k W_{k-1} S}}{\mathrm{d}(P_{W_k W_{k-1}} \otimes P_S)} = \frac{\mathrm{d}P_{S|W_k W_{k-1}}}{\mathrm{d}P_S} = \frac{\mathrm{d}P_{S|W_k}}{\mathrm{d}P_S},$$

where we have made use of Bayes' rule and the fact that $S \perp\!\!\!\perp W_{k-1}|W_k$. Using this in (14) together with Proposition 1 gives (15). An application of Cauchy–Schwarz and Jensen gives (16).

Proof of Corollary 5. For each $k \ge 1$, let $\rho_{W_k W_{k-1}} = P_{W_k W_{k-1}}$. Notice that, by disintegration and the choice of couplings,

$$\mathbf{E}[d(\bar{W}_k, \bar{W}_{k-1})] = \mathbf{E}\Big[\int d(u, v)P_{W_k W_{k-1}|S}(\mathrm{d}u, \mathrm{d}v)\Big] \le \mathbf{E}[\mathsf{W}_2(P_{W_k|S}, P_{W_{k-1}|S})],$$

where we have used the fact that $\mathsf{W}_2(\cdot, \cdot)$ dominates $\mathsf{W}_1(\cdot, \cdot)$ [23]. Since $P_{W_k|S}$ are points on the geodesic connecting $P_{W|S}$ and P_W, we have

$$\sum_{k=1}^K \mathsf{W}_2(P_{W_k|S}, P_{W_{k-1}|S}) = \sum_{k=1}^K (t_k - t_{k-1})\mathsf{W}_2(P_{W|S}, P_W) = \mathsf{W}_2(P_{W|S}, P_W).$$

Using this together with Cauchy-Schwarz and Proposition 1, we obtain (17).

B.6 Proofs for Section 7

Proof of Theorem 6. Applying (5) conditionally on S with $f(w) = \sigma\sqrt{6/n}$, $g(w) = \frac{|\mathrm{gen}(w,S)|}{\sigma\sqrt{6/n}}$, $\mu = P_{W|S}$, and $\nu = Q_W$, we have

$$\begin{aligned}&\langle P_{W|S}, |\mathrm{gen}(W,S)|\rangle \\ &\le \sqrt{\frac{24\sigma^2}{n}}\left(1 + \left\langle P_{W|S}, \psi_2^{-1}\left(\frac{\mathrm{d}P_{W|S}}{\mathrm{d}Q_W}\right)\right\rangle + \left(\log\left\langle Q_W, \exp\left(\frac{\mathrm{gen}^2(\cdot,S)}{6\sigma^2/n}\right)\right\rangle\right)^{1/2}\right).\end{aligned}$$

Since $\ell(w,S)$ is $(\sigma/\sqrt{n})$-subgaussian for all w, Markov's inequality gives, for any $0 < \delta < 1$,

$$\mathbf{P}\left[\left\langle Q_W, \exp\left(\frac{\mathrm{gen}^2(\cdot,S)}{6\sigma^2/n}\right)\right\rangle > \frac{2}{\delta}\right] \le \frac{\delta}{2}\left\langle P_S \otimes Q_W, \exp\left(\frac{\mathrm{gen}^2(\cdot,\cdot)}{6\sigma^2/n}\right)\right\rangle \le \delta,$$

which concludes the proof.

Proof of Theorem 7. The argument is almost identical to the proof of Theorem 3, with the difference that (5) is used for decorrelation.

To lighten the notation, let $\pi_k^S := P_{W_k W_{k-1}|S}$ and $\rho_k := \rho_{W_k W_{k-1}}$. Use the same definitions of δ, σ, ζ as in the proof of Theorem 3. Then, applying (5) with $f(\cdot) = \sigma(\cdot, \tilde{S})$, $g(\cdot) = \zeta(\cdot, \tilde{S})$, $\mu = \pi_k^S$, $\nu = \rho_k$, we have

$$\begin{aligned}\langle \pi_k^S, \sigma(\cdot,\tilde{S})\zeta(\cdot,\tilde{S})\rangle \le \sqrt{2}\|\sigma(\cdot,\tilde{S})\|_{L^2(\rho_k)} &+ 2\left\langle \pi_k^S, \sigma(\cdot,\tilde{S})\psi_2^{-1}\left(\frac{\mathrm{d}\pi_k^S}{\mathrm{d}\rho_k}\right)\right\rangle \\ &+ 2\|\sigma(\cdot,\tilde{S})\|_{L^1(\pi_k^S)}\sqrt{\log\langle \rho_k, \exp\left(\zeta^2(\cdot,\tilde{S})\right)\rangle}.\end{aligned}$$

By Markov's inequality and the union bound, for any $0 < \delta < 1$,

$$\mathbf{P}\left[\exists k \text{ s.t. } \left\langle \rho_k, \exp\left(\zeta^2(\cdot,\tilde{S})\right)\right\rangle > \frac{2}{p_k\delta}\right] \le \sum_k \frac{p_k\delta}{2}\left\langle P_{\tilde{S}} \otimes \rho_k, \exp\left(\zeta^2(\cdot,\cdot)\right)\right\rangle \le \delta.$$

Using this together with the estimate $\sigma(\cdot,\tilde{S}) \le 2\sqrt{n}d_{\tilde{S},\ell}(\cdot)$ yields the result in the statement.

B.7 Proofs for Section 8

Proof of Theorem 8. Without loss of generality, we assume $\mathrm{diam}(T) = 1$. Let Q be the Markov kernel from Ω to T defined by $Q(\cdot|\omega) = \delta_{\tau(\omega)}(\cdot)$; in particular, $\nu(\cdot) = \int_\Omega \mathbb{P}(\mathrm{d}\omega)Q(\cdot|\omega)$.

Fix some $r \ge 2$. For each $k \ge 0$ and each $t \in T$, let $B_k(t) := B(t, r^{-k})$. Since T is finite, there exists some $K \in \mathbb{N}$, such that $B_K(t) = \{t\}$ for all $t \in T$. Let $\mu \in \mathcal{P}(T)$ be given. Define the following sequence of Markov kernels from Ω to T:

$$Q_k(\cdot|\omega) := \frac{\mu(\cdot \cap B_k(\tau(\omega)))}{\mu(B_k(\tau(\omega)))}, \qquad k = 0, \dots, K.$$

Observe that $Q_0 = \mu$ and $Q_K = Q$. Then, since

$$\langle \mathbb{P} \otimes \mu, X\rangle = \int_{\Omega\times T} \mathbb{P}(\mathrm{d}\omega)\mu(\mathrm{d}t)X_t(\omega) = \int_T \mu(\mathrm{d}t)\mathbf{E}[X_t] = 0,$$

we can write

$$\begin{aligned}\mathbf{E}[X_\tau] &= \langle \mathbb{P} \otimes Q, X\rangle - \langle \mathbb{P} \otimes \mu, X\rangle \\ &= \langle \mathbb{P} \otimes Q_K, X\rangle - \langle \mathbb{P} \otimes Q_0, X\rangle \\ &= \sum_{k=1}^K \langle \mathbb{P} \otimes Q_k - \mathbb{P} \otimes Q_{k-1}, X\rangle \\ &= \sum_{k=1}^K \int_\Omega \left(\int_T X_t(\omega)Q_k(\mathrm{d}t|\omega) - \int_T X_t(\omega)Q_{k-1}(\mathrm{d}t|\omega)\right)\mathbb{P}(\mathrm{d}\omega) \\ &\le \sum_{k=1}^K \int_\Omega \int_{T\times T} |X_u(\omega) - X_v(\omega)|Q_k(\mathrm{d}u|\omega)Q_{k-1}(\mathrm{d}v|\omega)\mathbb{P}(\mathrm{d}\omega).\end{aligned}$$

Applying (4) conditionally on ω with

$$\begin{aligned}
f(u,v) &= \mathbf{1}_{B_k(\tau(\omega))}(u)\mathbf{1}_{B_{k-1}(\tau(\omega))}(v)d(u,v), \\
g(u,v) &= \frac{|X_u(\omega)-X_v(\omega)|}{d(u,v)}, \\
\mu(\mathrm{d}u,\mathrm{d}v) &= Q_k(\mathrm{d}u|\omega)\otimes Q_{k-1}(\mathrm{d}v|\omega), \\
\nu(\mathrm{d}u,\mathrm{d}v) &= \mu(\mathrm{d}u)\otimes\mu(\mathrm{d}v),
\end{aligned}$$

and using the fact that $Q_k(\cdot|\omega)$ is supported on $B_k(\tau(\omega))$ and $B_k(\tau(\omega)) \subseteq B_{k-1}(\tau(\omega))$, we have

$$\begin{aligned}
&\int_{T\times T} |X_u(\omega)-X_v(\omega)| Q_k(\mathrm{d}u|\omega)Q_{k-1}(\mathrm{d}v|\omega) \\
&\le 2^{1/p} r^{-k+1}\psi_p^{-1}\left(\frac{1}{\mu(B_k(\tau(\omega)))^2}\right) + r^{-k+1}\int_{T\times T}\psi_p\left(\frac{|X_u(\omega)-X_v(\omega)|}{d(u,v)}\right)\mu(\mathrm{d}u)\mu(\mathrm{d}v) \\
&\le 2^{2/p} r^{-k+1}\psi_p^{-1}\left(\frac{1}{\mu(B_k(\tau(\omega)))}\right) + r^{-k+1}\int_{T\times T}\psi_p\left(\frac{|X_u(\omega)-X_v(\omega)|}{d(u,v)}\right)\mu(\mathrm{d}u)\mu(\mathrm{d}v),
\end{aligned}$$

where the first term in the last step is due to Proposition A.2(iii). Then, using the increment condition and Proposition A.3, we have

$$\begin{aligned}
\mathbf{E}[X_\tau] &\le 2^{2/p}\sum_{k=1}^{K} r^{-k+1}\int_T \psi_p^{-1}\left(\frac{1}{\mu(B(t,r^{-k}))}\right)\nu(\mathrm{d}t) + \sum_{k=1}^{K} r^{-k+1} \\
&\le 1 + 2^{2/p} r^2 \int_T\int_0^1 \psi_p^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)\mathrm{d}\varepsilon\,\nu(\mathrm{d}t).
\end{aligned}$$

Since $1/\mu(B(t,\varepsilon)) \ge 1$, we can apply Proposition A.2(iv) to obtain the inequality

$$\mathbf{E}[X_\tau] \le 2^{2/p} r^2\left(2 + \int_T\int_0^1\left(\log\frac{1}{\mu(B(t,\varepsilon))}\right)^{1/p}\mathrm{d}\varepsilon\,\nu(\mathrm{d}t)\right).$$

We can now take $r=2$ to get the desired result when $\mathrm{diam}(T)=1$; the general finite-diameter case follows by straightforward rescaling.