THE ENTROPY DISTANCE BETWEEN THE WIENER
AND STATIONARY GAUSSIAN MEASURES

U. KEICH

THE ENTROPY DISTANCE BETWEEN THE WIENER AND STATIONARY GAUSSIAN MEASURES

U. Keich

Investigating the entropy distance between the Wiener measure, $W_{t_0,\tau}$, and stationary Gaussian measures, $Q_{t_0,\tau}$ on the space of continuous functions $C[t_0 - \tau, t_0 + \tau]$, we show that in some cases this distance can essentially be computed. This is done by explicitly computing a related quantity which in effect is a valid approximation of the entropy distance, provided it is sufficiently small; this will be the case if τ/t_0 is small. We prove that $H(W_{t_0,\tau}, Q_{t_0,\tau}) > \tau/2t_0$, and then show that $\tau/2t_0$ is essentially the typical case of such entropy distance, provided the mean and the variance of the stationary measures are set "appropriately".

Utilizing a similar technique, we estimate the entropy distance between the Ornstein-Uhlenbeck measure and other stationary Gaussian measures on $C[1-\tau, 1+\tau]$. Using this result combined with a variant of the triangle inequality for the entropy distance, which we devise, yields an upper bound on the entropy distance between stationary Gaussian measures which are absolutely continuous with respect to the Wiener measure.

1. Introduction.

Motivated by the study of stationary approximations to non-stationary stochastic processes, we ask how well can Brownian motion be approximated by stationary Gaussian processes. We look at a finite time interval and measure the quality of the approximation using the entropy distance.

Let P and Q be Gaussian measures on a common probability space. The entropy distance between P and Q, $H(P,Q)$, is finite if and only if the two measures are absolutely continuous with respect to one another; in that case:

$$H(P,Q) = E^P \left(\log \frac{dP}{dQ} \right) + E^Q \left(\log \frac{dQ}{dP} \right),$$

where E^P is the expectation under the measure P and $\frac{dP}{dQ}$ is the Radon-Nikodym derivative of P with respect to Q (see e.g. [**2**]).

Let $t_0 > 0$, $0 < \tau < t_0$, and let $W_{t_0,\tau}$ be the restriction of the Wiener measure to the space of continuous functions on $[t_0-\tau, t_0+\tau]$, $C[t_0-\tau, t_0+\tau]$.

In other words, $W_{t_0,\tau}$ is the measure induced by the standard Brownian motion observed between times $t_0 - \tau$ and $t_0 + \tau$. As a Gaussian measure it is characterized by its correlation $R(t,s) = t \wedge s$ with $t, s \in [t_0 - \tau, t_0 + \tau]$, and by its vanishing mean.

A Gaussian measure on $C[t_0 - \tau, t_0 + \tau]$, $Q_{t_0,\tau}$, is *stationary* if its mean is constant and its correlation, S, is a Töeplitz function. Here S denotes both the correlation $S(t,s)$ and the auto-correlation $S(r)$ with $r = t - s \in [-2\tau, 2\tau]$. Let $\mathfrak{S}$ denote the class of stationary Gaussian measures on $C[t_0 - \tau, t_0 + \tau]$ that are absolutely continuous with respect to $W_{t_0,\tau}$.

Consider the 1:1 and onto map $\psi : C[t_0 - \tau, t_0 + \tau] \mapsto C[1 - \tau/t_0, 1 + \tau/t_0]$ defined by $\psi(f)(s) \overset{d}{=} f(st_0)/\sqrt{t_0}$. This map and $W_{t_0,\tau}$ induce on $C[1 - \tau/t_0, 1 + \tau/t_0]$ the measure $W_{1,\tau/t_0}$ and, similarly, $Q_{t_0,\tau}$ induces the stationary Gaussian measure $\tilde{Q}$, determined by its correlation $\tilde{S}(t,s) = S((t-s)t_0)/t_0$. Note that $H(\tilde{Q}, W_{1,\tau/t_0}) = H(Q_{t_0,\tau}, W_{t_0,\tau})$; therefore $\inf_{Q_{t_0,\tau} \in \mathfrak{S}} H(Q_{t_0,\tau}, W_{t_0,\tau})$ depends only on τ/t_0.

We start our investigation with a couple of examples, the details of which can be found in appendix A. Consider the entropy distance between $W^\tau \overset{d}{=} W_{1,\tau}$ and the Ornstein-Uhlenbeck measure, $Q^\tau \overset{d}{=} Q_{1,\tau}$, defined by its vanishing mean and $S = \exp(-|t - s|/2)$. In this case, $H(Q^\tau, W^\tau) = \tau/(2(1-\tau))$. In the second example, Q^τ is defined by $S(t,s) \overset{d}{=} 1 - |t - s|/2$ with $|t - s| \leq 2\tau < 4$ and, again, $H(Q^\tau, W^\tau) = \tau/(2(1-\tau))$.

Note that in these examples $H \to \infty$, as $\tau \to 1$, and that for small τ, $H \sim \tau/2$. We would like to know how much can we possibly improve on that. It can be shown that $\inf_{Q_{t_0,\tau} \in \mathfrak{S}} H(Q_{t_0,\tau}, W_{t_0,\tau})$ is always attained (see appendix B). However, as we next show, the minimizer does not significantly improve on our examples.

Claim 1.1. For any $Q_{t_0,\tau} \in \mathfrak{S}$,

$$H(Q_{t_0,\tau}, W_{t_0,\tau}) \geq \sqrt{2}\sqrt{\frac{2 - \tau/t_0}{1 - \tau/t_0}} - 2.$$

As a corollary we find that as $\tau \to t_0$, $H \to \infty$ and $H \geq 1/2(\tau/t_0) + 7/16(\tau/t_0)^2$. Thus, for small τ/t_0, the minimal entropy distance is, up to leading order, the same as in our examples. As we show next, these examples are rather typical.

Let P and Q be two Gaussian measures on $C[t_0 - \tau, t_0 + \tau]$ with vanishing means. Let P_n and Q_n be their restrictions to $2^n + 1$ equally spaced points in $[t_0 - \tau, t_0 + \tau]$, with correlations R_n and S_n. Then $H(P, Q) = \lim H(P_n, Q_n)$ [2]. Let K_n be a root of R_n, i.e., $R_n = K_n K_n^*$, and let $T_n = K_n^{-1} S_n K_n^{-*}$,

with K^{-*} being a short for $(K^{-1})^*$. Then, as we show in [3],

$$H(P_n, Q_n) = \frac{1}{2} \operatorname{Tr}(T_n + T_n^{-1} - 2I_n) = \frac{1}{2} \sum_{i=0}^{2^n} \frac{(\lambda_i^n - 1)^2}{\lambda_i^n},$$

where λ_i^n are the positive eigenvalues of T_n. In our case $Q = Q_{t_0,\tau} \in \mathfrak{S}$ with an autocorrelation S, and $P = W_{t_0,\tau}$ with $R = t \wedge s$. Since R_n^{-1} is essentially a second order difference operator, if we choose K_n to be the Cholesky factorization of R_n, with $\delta \overset{d}{=} \delta_n \overset{d}{=} 2\tau/2^n$, and $S_k \overset{d}{=} S(k\delta)$, we have:

(1)
$T =$

$$\begin{pmatrix} \frac{S_0}{t_0-\tau} & \frac{1}{\sqrt{t_0-\tau}}\frac{S_1-S_0}{\sqrt{\delta}} & \frac{1}{\sqrt{t_0-\tau}}\frac{S_2-S_1}{\sqrt{\delta}} & \frac{1}{\sqrt{t_0-\tau}}\frac{S_3-S_2}{\sqrt{\delta}} & \cdots & \frac{1}{\sqrt{t_0-\tau}}\frac{S_n-S_{n-1}}{\sqrt{\delta}} \\ * & 2\frac{S_0-S_1}{\delta} & \frac{2S_1-S_0-S_2}{\delta} & \frac{2S_2-S_1-S_3}{\delta} & \cdots & \frac{2S_{n-1}-S_{n-2}-S_n}{\delta} \\ * & * & 2\frac{S_0-S_1}{\delta} & \frac{2S_1-S_0-S_2}{\delta} & \cdots & \frac{2S_{n-2}-S_{n-3}-S_{n-1}}{\delta} \\ * & * & * & 2\frac{S_0-S_1}{\delta} & \ddots & \vdots \\ & & & & \ddots & \frac{2S_1-S_0-S_2}{\delta} \\ * & * & * & * & & 2\frac{S_0-S_1}{\delta} \end{pmatrix}$$

where the $*$'s are filled in according to the symmetry of T.

Let

(2) $\varphi(W_{t_0,\tau}, Q_{t_0,\tau}) \overset{d}{=} \frac{1}{2} \lim_n \sum_i (\lambda_i^n - 1)^2 = \frac{1}{2} \lim_n \operatorname{Tr} (T_n - I)^2.$

Claim 1.2.

$$\varphi(W_{t_0,\tau}, Q_{t_0,\tau}) = \frac{1}{2} \left(\frac{S_0}{t_0 - \tau} - 1 \right)^2 + \frac{1}{2} \int_0^{2\tau} S''(t)^2 (2\tau - t)\, dt$$

$$+ \frac{1}{t_0 - \tau} \int_0^{2\tau} S'(t)^2\, dt.$$

The importance of the last claim is that $\varphi(W_{t_0,\tau}, Q_{t_0,\tau})$ can be used to approximate $H(W_{t_0,\tau}, Q_{t_0,\tau})$:

Claim 1.3. For $\varphi < 1$,

$$\left| \frac{\varphi - H}{\varphi} \right| \leq \frac{\sqrt{2\varphi}}{1 - \sqrt{2\varphi}}.$$

Thus, $\varphi(W_{t_0,\tau}, Q_{t_0,\tau})$ is a good approximation of $H(W_{t_0,\tau}, Q_{t_0,\tau})$ if it is small. Suppose $Q_{t_0,\tau} \in \mathfrak{S}$, then Shepp shows that on $(0, 2\tau)$, S' is absolutely continuous and S'' satisfies $\int_0^{2\tau} S''(t)^2 (2\tau - t)\, dt < \infty$, and that $S^+(0) = -1/2$ where S^+ is the derivative from the right [7]. Thus, with S fixed,

τ sufficiently small and $S(0) = t_0$, $\varphi(W_{t_0,\tau}, Q_{t_0,\tau}) \sim \tau/2t_0$ and therefore $H(W_{t_0,\tau}, Q_{t_0,\tau}) \sim \tau/2t_0$, as is the case in our examples.

What if $V \overset{d}{=} S(0) \neq t_0$?

Claim 1.4.   (i) If $V > t_0$, then

$$H(W_{t_0,\tau}, Q_{t_0,\tau}) > \frac{1}{2} \frac{(V/t_0 - 1)^2}{V/t_0}.$$

(ii) If $V < t_0$, then

$$H(W_{t_0,\tau}, Q_{t_0,\tau}) > \frac{1}{2} \frac{((1+\delta)V/t_0 - 1)^2}{(1+\delta)V/t_0},$$

where $1 + \delta \overset{d}{=} 1/(1 - \tau/t_0)$.

(iii) In either case, for a fixed S (corresponding to $Q \in \mathfrak{S}$),

$$H(W_{t_0,\tau}, Q_{t_0,\tau}) = \frac{1}{2} \frac{(V/t_0 - 1)^2}{V/t_0} + o(1) \qquad \text{as } \tau \to 0.$$

Similarly, when the constant mean, μ, of $\tilde{Q}_{t_0,\tau} \in \mathfrak{S}$ does not vanish, $H(W_{t_0,\tau}, Q_{t_0,\tau})$ is rather large: let $Q_{t_0,\tau} \in \mathfrak{S}$ be obtained from $\tilde{Q}_{t_0,\tau}$ by removing the constant drift μ, then:

Claim 1.5.

$$H(W_{t_0,\tau}, \tilde{Q}_{t_0,\tau}) > H(W_{t_0,\tau}, Q_{t_0,\tau}) + \frac{1}{2} \frac{\mu^2}{t_0 - \tau},$$

and if $V = t_0$, then $H(W_{t_0,\tau}, \tilde{Q}_{t_0,\tau}) = \mu^2/t_0 + o(1)$ as $\tau \to 0$.

So far we dealt with the entropy distance between $W^\tau = W_{1,\tau}$ and $Q^\tau \in \mathfrak{S}$. We next consider the entropy distance between two stationary Gaussian measures in $\mathfrak{S}$. Again, we start with an example. Let $\tilde{S} = e^{-|t-s|/2}$ be the Ornstein-Uhlenbeck correlation, and let $S = 1 - |t - s|/2$. Using the chain rule, we get

$$\frac{d\tilde{Q}}{dQ} = \frac{d\tilde{Q}}{dW} \frac{dW}{dQ},$$

both Radon-Nikodym derivatives on the right-hand side being known (see appendix A). The entropy is easily obtained:

$$H^\tau(\tilde{Q}, Q) = \frac{1}{2} \frac{e^{-\tau} - (1 - \tau)}{2 - \tau} = \frac{1}{8}\tau^2 + O(\tau^3).$$

This is an order of magnitude smaller than $\tau/2$. As before, the example is rather typical and to prove that, we first estimate the entropy distance between the Ornstein-Uhlenbeck measure, $\tilde{Q}^\tau$ and any other $Q^\tau \in \mathfrak{S}$, and then use a variant of the triangle inequality.

Let $\tilde{K}_n$ denote the Cholesky factorization of $\tilde{S}_n$, and let $T_n = \tilde{K}_n^{-1} S_n \tilde{K}_n^{-*}$. Again, by Claim 1.3, $H(\tilde{Q}^\tau, Q^\tau) = \lim_n \text{Tr}(T_n + T_n^{-1} - 2I_n)/2$ can be approximated by $\varphi(\tilde{Q}^\tau, Q^\tau) \overset{d}{=} \lim_n \sum_i \left(\lambda_i^n(\tau) - 1 \right)^2 / 2$, where:

Claim 1.6.

(3)

$$\varphi(\tilde{Q}^\tau, Q^\tau) = (S(0) - 1)^2 + 2 \int_0^{2\tau} S'(t)^2 \, dt + \left(S^2(2\tau) - 1 \right) + \frac{1}{2} \int_0^{2\tau} S^2(t) \, dt$$

$$+ \int_0^{2\tau} S''(t)^2 (2\tau - t) \, dt + \frac{1}{16} \int_0^{2\tau} S(t)^2 (2\tau - t) \, dt$$

$$- \frac{1}{2} \int_0^{2\tau} S''(t) S(t) (2\tau - t) \, dt.$$

It is not hard to see that if S is fixed with $S(0) = 1$ and if S'' is bounded on $(0, 2\tau_0)$ for some $\tau_0 > 0$, then the right hand side of (3) is bounded by $c\tau^2$ for some constant $c > 0$ and any $\tau < \tau_0$. Thus, there exists another constant c such that $H(\tilde{Q}^\tau, Q^\tau) < c\tau^2$ in this case. More generally:

Claim 1.7. Suppose $\grave{Q}^\tau, Q^\tau \in \mathfrak{S}$ are two measures on $C[1 - \tau, 1 + \tau]$ such that $\grave{S}(0) = S(0) = 1$ and that $\grave{S}''$ and S'' are bounded on $(0, 2\tau_0)$ for some $\tau_0 > 0$. Then, $H(\grave{Q}^\tau, Q^\tau) < c\tau^2$ for some constant $c > 0$ and any $\tau < \tau_0$.

The proof is an immediate corollary of the previous discussion and of Theorem 1 which follows.

The entropy distance is not a metric; it fails to satisfy the triangle inequality even for one-dimensional correlations: if P, Q_1 and Q_2 are three (mean zero) Gaussian measures on the line with variances $4, 2$ respectively 6, then

$$H(Q_1, Q_2) = \frac{2}{3} > \frac{1}{4} + \frac{1}{12} = H(Q_1, P) + H(P, Q_2).$$

However, one can prove the following variant of the triangle inequality. Here P, Q_1 and Q_2 are Gaussian measures on $L^2[0, T]$ with $T < \infty$.

Theorem 1.

$$H(Q_1, Q_2) \le 4H(Q_1, P) + 4H(P, Q_2) + 4H(Q_1, P)H(P, Q_2).$$

If we define $\bar{H}(P, Q) \overset{d}{=} \sqrt{H(P, Q)}$, then the last inequality leads to an "almost" triangle inequality:

Corollary 1.

$$\bar{H}(Q_1, Q_2) \le 2 \big[\bar{H}(Q_1, P) + \bar{H}(P, Q_2) \big].$$

Theorem 1 is set up in the context of Gaussian measures on a Hilbert space. Such a measure P is completely determined by its covariance operator R and mean μ. Prohorov proved that P exists if and only if R is a trace class operator (e.g. [4, Thm I.2.3]). In the case of $L^2[0,T]$, the covariance operator can be identified with a covariance function $R \in L^2([0,T] \times [0,T])$, the latter being the kernel of the trace class integral operator. The basic result here is due to Rao and Varadarajan [5]:

RV-Theorem. $H(P,Q) < \infty$ if and only if there exists a Hilbert-Schmidt operator, G with a spectrum $\sigma(G) > -1$, such that

$$S = R + R^{\frac{1}{2}}GR^{\frac{1}{2}},$$

and $\mu - \nu \in \mathfrak{D}(R^{-\frac{1}{2}}) = \mathfrak{D}(S^{-\frac{1}{2}})$,

where $\mathfrak{D}(R^{-\frac{1}{2}})$ is the domain of the self-adjoint operator $R^{-\frac{1}{2}}$. In the finite case, the operator

$$F \overset{d}{=} R^{-\frac{1}{2}}SR^{-\frac{1}{2}} + R^{\frac{1}{2}}S^{-1}R^{\frac{1}{2}} - 2I,$$

is a well defined, symmetric, positive-definite trace class operator, and a slight variation on a result by Sekine [6] yields:

Lemma 1.8.

$$H(P,Q) = \frac{1}{2}\operatorname{Tr} F + \frac{1}{2}\left|R^{-\frac{1}{2}}(\mu - \nu)\right|^2 + \frac{1}{2}\left|S^{-\frac{1}{2}}(\mu - \nu)\right|^2.$$

We prove Theorem 1 by dealing separately with the entropy "due to the correlations", $H_c \overset{d}{=} \frac{1}{2}\operatorname{Tr} F$, and the part that comes from the means $H_m \overset{d}{=} \frac{1}{2}\left|R^{-\frac{1}{2}}(\mu - \nu)\right|^2 + \frac{1}{2}\left|S^{-\frac{1}{2}}(\mu - \nu)\right|^2$. It follows from Lemma 1.8 and some algebraic manipulations, that H_c itself obeys an analogue of Theorem 1. H_m, however, cannot be bounded independently of H_c.

Finally, in appendix C, we contrast the entropy distance to W^τ with the L^2 distance between the correlations. Analogously to the entropy distance we find that all the correlations representing stationary Gaussian measures that are absolutely continuous with respect to the Wiener measure are, to leading order, at the same L^2 distance from the Wiener correlation.

2. The proofs.

Proof of Claim 1.1. By the scaling argument that was mentioned in the introduction, it suffices to prove the claim for $t_0 = 1$ and $\tau < 1$. Recall that W_n^τ is the restriction of W^τ to the σ-field generated by $2^n + 1$ equally spaced points in $[t_0 - \tau, t_0 + \tau]$, and that R_n is the corresponding correlation matrix. Thus, with $S(0) = V$ and $S(2\tau) = \gamma V$ for some γ with $|\gamma| < 1$,

$$R_0 \overset{d}{=} \begin{bmatrix} 1-\tau & 1-\tau \\ 1-\tau & 1+\tau \end{bmatrix} \qquad\qquad S_0 \overset{d}{=} \begin{bmatrix} V & \gamma V \\ \gamma V & V \end{bmatrix}.$$

$$H_0 \stackrel{d}{=} H(W_0^\tau, Q_0^\tau)$$

$$= \frac{1}{2} \operatorname{Tr} \left(R_0^{-1} S_0 + S_0^{-1} R_0 - 2I_2 \right)$$

$$= \frac{1}{2} V \frac{1 - \gamma(1-\tau)}{\tau(1-\tau)} + \frac{1}{V} \frac{1 - \gamma(1-\tau)}{1 - \gamma^2} - 2.$$

Minimizing H_0 with respect to V we find that $V_{\min} = \sqrt{2\tau(1-\tau)/(1-\gamma^2)}$ and therefore

$$H_0 \geq \sqrt{2} \frac{1 - \gamma(1-\tau)}{\sqrt{\tau(1-\tau)(1-\gamma^2)}} - 2.$$

Minimizing the right hand side with respect to γ, we learn that $\gamma_{min} = 1 - \tau$, whence

$$H_0 \geq \sqrt{2} \sqrt{\frac{2-\tau}{1-\tau}} - 2.$$

The proof is completed by the obvious inequality $H(W^\tau, Q^\tau) \geq H_0$. Note that

$$H_0 \geq \sqrt{2} \sqrt{\frac{2-\tau}{1-\tau}} - 2 \geq \tau/2 + 7/16\tau^2,$$

as can be verified directly. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Claim 1.2. With $T = T_n(\tau) = (t_{ij})_{0 \leq i, j \leq 2^n}$,

$$(4) \qquad \sum_{i=1}^{2^n} t_{i0}^2 + t_{0i}^2 = \frac{2}{t_0 - \tau} \sum_{i=1}^{2^n} (S_i - S_{i-1})^2 \frac{1}{\delta} \longrightarrow \frac{2}{t_0 - \tau} \int_0^{2\tau} S'(t)^2 \, dt.$$

Let

$$\bar{S}(r) \stackrel{d}{=} S(r) - \left(1 - \frac{|r|}{2} \right).$$

Then $\bar{S}$ is an even function, with absolutely continuous derivative $\bar{S}'$ on $(-2\tau, 2\tau)$, and $\bar{S}'' \equiv S''$ on $(0, 2\tau)$. Let

$$\bar{S}''(t, s) \stackrel{d}{=} \bar{S}''(t - s).$$

Then $\bar{S}''$ is a Töeplitz function of the square $\Omega \stackrel{d}{=} [0, 2\tau] \times [0, 2\tau]$, and

$$\int_0^{2\tau} \int_0^{2\tau} \bar{S}''(t-s)^2 \, dt \, ds = \int_0^{2\tau} S''(t)^2 (2\tau - t) \, dt < \infty,$$

whence $\bar{S}'' \in L^2(\Omega)$.

Consider Ω equipped with the σ-field $\mathcal{F}_n$, generated by the squares $\{[(i-1)\delta, i\delta) \times [(j-1)\delta, j\delta) : i, j = 1 \ldots 2^n\}$ and with Lebesgue's measure. Let $\psi_n \stackrel{d}{=} E[\bar{S}''|\mathcal{F}_n]$, then

$$\psi_n(x, y) = \frac{1}{\delta^2} \int_{(i-1)\delta}^{i\delta} \int_{(j-1)\delta}^{j\delta} \bar{S}''(t-s)\, dt\, ds \qquad \begin{matrix} (i-1)\delta \leq x < i\delta \\ (j-1)\delta \leq y < j\delta \end{matrix},$$

and since $\bar{S}'' \in L^2(\Omega)$, by the L^2 martingale convergence theorem, $\psi_n \to \bar{S}''$ a.e. and in L^2, in particular, $\|\psi_n\| \longrightarrow \|\bar{S}''\|$. Integrating, one finds that

$$t_{ij} = \frac{1}{\delta} \int_{(i-1)\delta}^{i\delta} \int_{(j-1)\delta}^{j\delta} \bar{S}''(t-s)\, dt\, ds \qquad 1 \leq i \neq j \leq n$$

$$t_{ii} - 1 = \frac{1}{\delta} \int_{(i-1)\delta}^{i\delta} \int_{(i-1)\delta}^{i\delta} \bar{S}''(t-s)\, dt\, ds \qquad 1 \leq i \leq n.$$

Thus,

$$\lim_n \left[\sum_{1 \leq i \neq j \leq n} t_{ij}^2 + \sum_{i=1}^{2^n} (t_{ii} - 1)^2 \right] = \lim_n \|\psi_n\|^2 = \int_0^{2\tau} S''(t)^2 (2\tau - t)\, dt,$$

which together with (4) and (2) completes the proof. $\qquad \square$

Proof of Claim 1.3.

$$|H - \varphi| = \left| \frac{1}{2} \lim_n \sum_i \left[\frac{(\lambda_i^n - 1)^2}{\lambda_i^n} - (\lambda_i^n - 1)^2 \right] \right|$$

$$\leq \frac{1}{2} \lim_n \sum_i \frac{(\lambda_i^n - 1)^2 |1 - \lambda_i^n|}{\lambda_i^n}$$

$$\leq \varphi \, \frac{1}{\underline{\lim}_n \min_i \lambda_i^n} \, \overline{\lim}_n \max_i |1 - \lambda_i^n|.$$

The proof is completed by noting that

$$\overline{\lim}_n \max_i |1 - \lambda_i^n| \leq \overline{\lim}_n \sqrt{\sum_i (\lambda_i^n - 1)^2} = \sqrt{2\varphi},$$

and that

$$\underline{\lim}_n \min_i \lambda_i^n \geq 1 - \overline{\lim}_n \sqrt{\sum_i (\lambda_i^n - 1)^2} = 1 - \sqrt{2\varphi}.$$

$$\square$$

Proof of Claim 1.4.   (i) Suppose $t_0 = 1$, and assume that $\lambda_0^n \geq \lambda_1^n \geq \cdots \geq \lambda_{2^n}^n$. With $\langle, \rangle$ being the standard inner-product,

$$\lambda_0^n \geq \langle T_n e_1, e_1 \rangle = \frac{V}{1 - \tau} > V > 1.$$

It follows that

$$H \geq \frac{1}{2} \frac{(\lambda_0^n - 1)^2}{\lambda_0^n} > \frac{1}{2} \frac{(V - 1)^2}{V},$$

which proves (i) when $t_0 = 1$. The aforementioned scaling argument completes the proof.

(ii) Without loss of generality assume that $t_0 = 1$, and let $1 + \delta \stackrel{d}{=} (1 - \tau)^{-1}$. If τ (equivalently δ), is sufficiently small, then

$$\lambda_{2^n}^n \leq \langle T_n e_1, e_1 \rangle = V(1 + \delta) < 1.$$

Hence,

$$H \geq \frac{[V(1 + \delta) - 1]^2}{V(1 + \delta)}.$$

(iii) Again, without loss of generality $t_0 = 1$ and assume $V > 1$. Then, as we saw in the proof of (i), $\lambda_0^n = \lambda_0^n(\tau) > V/(1 - \tau)$, therefore

(5)
$$(\lambda_0^n - 1)^2 > \left(\frac{V}{1 - \tau} - 1 \right)^2.$$

Since by Claim 1.2

$$\sum_i (\lambda_i^n(\tau) - 1)^2 = \left(\frac{S_0}{1 - \tau} - 1 \right)^2$$
$$+ \int_0^{2\tau} S''(t)^2 (2\tau - t)\, dt + \frac{2}{1 - \tau} \int_0^{2\tau} S'(t)^2\, dt$$
$$= \left(\frac{V}{1 - \tau} - 1 \right)^2 + \tau + o(\tau),$$

it follows that:

(a)
$$\varlimsup_n \sum_1^{2^n} (\lambda_i^n(\tau) - 1)^2 \leq \tau + o(\tau).$$

Note that λ_0^n is not included in the summation above.

(b)
$$\varlimsup_n (\lambda_0^n - 1)^2 \leq \left(\frac{V}{1 - \tau} - 1 \right)^2 + O(\tau).$$

As in Claim 1.3, we can deduce from (a) that $\varlimsup_n \sum_1^{2^n} (\lambda_i^n(\tau) - 1)^2 / \lambda_i^n = O(\tau)$. From (b) we learn that for some constant c,

$$\frac{V}{1 - \tau} \leq \varliminf_n \lambda_0^n(\tau) \leq \varlimsup_n \lambda_0^n(\tau) \leq V + c\tau.$$

Since $H = \lim_n \sum_0^{2^n} (\lambda_i^n(\tau) - 1)^2 / 2\lambda_i^n$, it follows that $H = (V-1)^2/2V + o(1)$ as $\tau \to 0$. The case $V < 1$ is proved similarly.
$\square$

Proof of Claim 1.5. Let $\boldsymbol{\mu} \in \mathbb{R}^{2^n+1}$ be the vector with constant entries μ. Then, with E^P denoting expectation with respect to the P measure and $\boldsymbol{x} \in \mathbb{R}^{2^n+1}$:

$$H(W_n^\tau, \tilde{Q}_n^\tau) = -\frac{1}{2} E^{W_n^\tau} \left[\langle R_n^{-1} \boldsymbol{x}, \boldsymbol{x} \rangle - \langle S_n^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{x} - \boldsymbol{\mu} \rangle \right]$$
$$+ \frac{1}{2} E^{\tilde{Q}_n^\tau} \left[\langle R_n^{-1} \boldsymbol{x}, \boldsymbol{x} \rangle - \langle S_n^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{x} - \boldsymbol{\mu} \rangle \right]$$
$$= H(W_n^\tau, Q_n^\tau) + \frac{1}{2} \langle S_n^{-1} \boldsymbol{\mu}, \boldsymbol{\mu} \rangle + \frac{1}{2} \langle R_n^{-1} \boldsymbol{\mu}, \boldsymbol{\mu} \rangle.$$

It happens that with $\boldsymbol{e}_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{2^n+1}$, $\mu K \boldsymbol{e}_1 = \sqrt{t_0 - \tau}\boldsymbol{\mu}$ and therefore

$$\langle R_n^{-1} \boldsymbol{\mu}, \boldsymbol{\mu} \rangle = \frac{\mu^2}{t_0 - \tau} \langle R_n^{-1} K \boldsymbol{e}_1, K \boldsymbol{e}_1 \rangle = \frac{\mu^2}{t_0 - \tau}.$$

Hence, $H(W_{t_0,\tau}, \tilde{Q}_{t_0,\tau}) > H(W_{t_0,\tau}, Q_{t_0,\tau}) + \frac{1}{2}\mu^2/(t_0 - \tau)$. Note that $\langle R_n^{-1} \boldsymbol{\mu}, \boldsymbol{\mu} \rangle = \mu^2/t_0 + o(1)$ as $\tau \to 0$, and since $S(0) = t_0$ implies that $\overline{\lim}_n \max_i |\lambda_i^n(\tau) - 1| = o(1)$, it follows that $\overline{\lim}_n \max_i |\lambda_i^n - 1|^{-1} = o(1)$ and therefore,

$$\overline{\lim_n} \langle S_n^{-1} \boldsymbol{\mu}, \boldsymbol{\mu} \rangle = \overline{\lim_n} \frac{\mu^2}{t_0 - \tau} \langle T_n^{-1} \boldsymbol{e}_1, \boldsymbol{e}_1 \rangle = \mu^2/t_0 + o(1).$$

$\square$

Proof of Claim 1.6. Let $\tilde{K} = \tilde{K}_n$ denote the Cholesky factorization of $\tilde{S}_n$, i.e., $\tilde{K}$ is a lower triangular matrix with $\tilde{S} = \tilde{K}\tilde{K}^*$, and let $T_n = \tilde{K}^{-1} S_n \tilde{K}^{-*}$. As in (1) we can explicitly compute T_n:

(6) $T_n =$

$$\begin{pmatrix} S_0 & yS_1 - xS_0 & yS_2 - xS_1 & yS_3 - xS_2 & \ldots & yS_n - xS_{n-1} \\ * & bS_0 - 2aS_1 & bS_1 - a(S_0 + S_2) & bS_2 - a(S_1 + S_3) & \ldots & bS_{n-1} - a(S_{n-2} + S_n) \\ * & * & bS_0 - 2aS_1 & bS_1 - a(S_0 + S_2) & \ldots & bS_{n-2} - a(S_{n-3} + S_{n-1}) \\ * & * & * & bS_0 - 2aS_1 & \ddots & \vdots \\ & & & & \ddots & \\ * & * & * & & & bS_1 - a(S_0 + S_2) \\ * & * & * & & & bS_0 - 2aS_1 \end{pmatrix},$$

where $S_k = S(k\delta)$ and

$$a = \frac{e^{-\delta/2}}{1 - e^{-\delta}}, \qquad b = \frac{1 + e^{-\delta}}{1 - e^{-\delta}}, \qquad c = \frac{1}{1 - e^{-\delta}},$$
$$x = \frac{e^{-\delta/2}}{\sqrt{1 - e^{-\delta}}}, \qquad y = \frac{1}{\sqrt{1 - e^{-\delta}}}.$$

The following estimates are based on

$$a = \frac{1}{\delta} - \frac{1}{24}\delta + O(\delta^3), \qquad b = \frac{2}{\delta} + \frac{1}{6}\delta + O(\delta^3),$$

$$x = \frac{1}{\sqrt{\delta}} - \frac{1}{4}\sqrt{\delta} + O(\delta^{3/2}), \qquad y = \frac{1}{\sqrt{\delta}} + \frac{1}{4}\sqrt{\delta} + O(\delta^{3/2}).$$

$$\sum_{k=1}^{2^n} t_{0k}^2 = \sum_{k=1}^{2^n}(yS_k - xS_{k-1})^2$$

$$= \sum \left[\frac{1}{\sqrt{\delta}}(S_k - S_{k-1}) + \frac{\sqrt{\delta}}{4}(S_k + S_{k-1}) + O(\delta^{3/2}) \right]^2$$

$$= \sum \left[\frac{1}{\sqrt{\delta}}(S_k - S_{k-1}) \right]^2 + \frac{1}{2}\sum(S_k^2 - S_{k-1}^2) + \frac{1}{16}\sum \delta(S_k + S_{k-1})^2$$

$$+ \sum O(\delta)(S_k - S_{k-1}) + \sum O(\delta^2)(S_k + S_{k-1}) + \sum O(\delta^3).$$

We have six terms on the right hand side. The last three are all $o(1)$ as $n \to \infty$, while the first three converge to the corresponding three terms in (3) (to be precise, one half of each term is obtained this way). What remains is

$$\sum_{1 \le i \ne j \le 2^n} t_{ij}^2 + \sum_{i=0}^{2^n}(t_{ii} - 1)^2.$$

Note that, with $\bar{S} = S - (1 - |r|/2)$, and $i \ge 1$,

$$t_{ii} - 1 = bS_0 - 2aS_1 - 1$$

$$= \left[\frac{2}{\delta}S_0 - \frac{2}{\delta}S_1 - 1 \right] + \left[\frac{\delta}{6}S_0 + \frac{\delta}{12}S_1 \right] + O(\delta^3)$$

$$= -\frac{1}{\delta}\int_{(i-1)\delta}^{i\delta}\int_{(i-1)\delta}^{i\delta} \bar{S}''(t - s)\,dt\,ds$$

$$+ \left[\frac{1}{6}S_0 + \frac{1}{24}(S_{-1} + S_1) \right]\delta + O(\delta^3).$$

Similarly, for $1 \le i \ne j \le 2^n$,

$$t_{ij} = bS_{i-j} - a(S_{i-j-1} + S_{i-j+1})$$

$$= -\frac{1}{\delta}\int_{(i-1)\delta}^{i\delta}\int_{(j-1)\delta}^{j\delta} \bar{S}''(t - s)\,dt\,ds$$

$$+ \left[\frac{1}{6}S_{i-j} + \frac{1}{24}(S_{i-j-1} + S_{i-j+1}) \right]\delta + O(\delta^3).$$

Hence, with $\psi_n = E[\bar{S}''|\mathcal{F}_n]$ as in the proof of Claim 1.2,

$$\sum_{1 \leq i \neq j \leq n} t_{ij}^2 + \sum_{i=0}^{2^n} (t_{ii} - 1)^2$$

$$= \|\psi_n\|^2 + \sum_{i,j} \left[\frac{1}{6} S_{i-j} + \frac{1}{24}(S_{i-j-1} + S_{i-j+1}) \right]^2 \delta^2$$

$$- 2 \sum_{i,j} \left[\frac{1}{6} S_{i-j} + \frac{1}{24}(S_{i-j-1} + S_{i-j+1}) \right] \int_{(i-1)\delta}^{i\delta} \int_{(j-1)\delta}^{j\delta} \bar{S}''(t-s) \, dt \, ds$$

$$+ o(1).$$

Recall that $\|\psi_n\| \longrightarrow \|\bar{S}''\|$ and let $n \to \infty$ to obtain the last three terms in (3). $\qquad \square$

Proof of Lemma 1.8. Let $P \sim (R, \mu)$, i.e. P is the Gaussian measure on $H = L^2[0, T]$ which is determined by the correlation $R(t, s)$ and mean μ_t ($t, s \in [0, T]$), and let $Q \sim (S, \nu)$ be absolutely continuous with respect to P. Note that neither is assumed to be stationary. The proof is essentially a translation of a result by Sekine [**6**] which is set in the context of an abstract Wiener space [**4**, Sec I.4]. Let $\mathfrak{D} \overset{d}{=} \mathfrak{D}(R^{-\frac{1}{2}})$. If we equip $\mathfrak{D}$ with the inner-product

$$(x, y)_R \overset{d}{=} (R^{-\frac{1}{2}}x, R^{-\frac{1}{2}}y),$$

then the resulting space, $\mathfrak{D}_R$, is a Hilbert space and $R^{\frac{1}{2}}$ is a unitary map from H onto $\mathfrak{D}_R$. With i being the natural inclusion of $\mathfrak{D}_R$ in H, the triplet $(i, \mathfrak{D}_R, H)$ is an abstract Wiener space.

Since $P \sim Q$, by the RV-theorem there exists a Hilbert-Schmidt operator G on H with $\sigma(G) > -1$ such that

$$A \overset{d}{=} RS^{-1} = R^{\frac{1}{2}}(I + G)^{-1} R^{-\frac{1}{2}},$$

is a well defined linear operator on $\mathfrak{D}$. Moreover, one can show that A is a bounded, symmetric, positive-definite operator on $\mathfrak{D}_R$, and $A - I$ is a Hilbert-Schmidt operator on $\mathfrak{D}_R$: The Hilbert-Schmidt norm of $A - I$ is the same as that of $(I + G)^{-1} - I$ (a Hilbert-Schmidt operator on H). Note that for $x, y \in \mathfrak{D}$,

$$(x, y)_S = (Ax, y)_R.$$

Indeed, one can define A in this manner in the more general setting of an abstract Wiener space where the RV-theorem does not apply [**8**, Thm 10.1].

Sekine shows that, with α_i being the eigenvalues of A,

$$H(P|Q) = \frac{1}{2} \left| A^{-\frac{1}{2}}(\mu - \nu) \right|_R^2 - \frac{1}{2} \log \prod \alpha_i^{-1} e^{(\alpha_i^{-1} - 1)}.$$

Since SR^{-1} and RS^{-1}, as operators on $\mathfrak{D}_S$, respectively $\mathfrak{D}_R$, have reciprocal eigenvalues, it follows that

$$H(P,Q) = \frac{1}{2}\sum\left(\alpha_i + \alpha_i^{-1} - 2\right) + \frac{1}{2}\left|S^{-\frac{1}{2}}(\mu-\nu)\right|^2 + \frac{1}{2}\left|R^{-\frac{1}{2}}(\mu-\nu)\right|^2.$$

Finally, since $RS^{-1}|_{\mathfrak{D}_R}$ and $R^{\frac{1}{2}}S^{-1}R^{\frac{1}{2}}|_H$ have the same eigenvalues, our lemma is proved. $\square$

Proof of Theorem 1. For Gaussian measures $P \sim (R,\mu)$ and $Q \sim (S,\nu)$, we define

$$H_c(P,Q) \stackrel{d}{=} \frac{1}{2}\operatorname{Tr}\left(R^{-\frac{1}{2}}SR^{-\frac{1}{2}} + R^{\frac{1}{2}}S^{-1}R^{\frac{1}{2}} - 2I\right).$$

Note that H_c is the distance "due to the correlations", i.e., $H_c(P,Q) = H(P',Q')$ where $P' \sim (R,0)$ and $Q' \sim (S,0)$. We also define

$$H_m(P,Q) \stackrel{d}{=} \frac{1}{2}\left|S^{-\frac{1}{2}}(\mu-\nu)\right|^2 + \frac{1}{2}\left|R^{-\frac{1}{2}}(\mu-\nu)\right|^2,$$

so that

$$H(P,Q) = H_c(P,Q) + H_m(P,Q).$$

Let $P \sim (R,\mu)$, $Q_1 \sim (S_1,\nu_1)$ and $Q_2 \sim (S_2,\nu_2)$ be three Gaussian measures on $L^2[0,T]$.

Lemma 2.1.

$$H_c(Q_1,Q_2) \le 2H_c(Q_1,P) + 2H_c(P,Q_2) + 2H_c(Q_1,P)H_c(P,Q_2).$$

Proof. Let A_1 and A_2 be boundedly invertible operators on L^2. Then

$$\left(A_1A_2 + (A_2A_1)^{-1} - 2I\right) + \left(A_1A_2^{-1} + A_1^{-1}A_2 - 2I\right)$$
$$= (A_1 + A_1^{-1} - 2I)(A_2 + A_2^{-1} - 2I)$$
$$+ 2(A_1 + A_1^{-1} - 2I) + 2(A_2 + A_2^{-1} - 2I).$$

Assume now that for $i = 1, 2$, $A_i = I + G_i$ where G_i are Hilbert-Schmidt operators with spectra $\sigma(G_i) > -1$. It is not hard to see that there exist Hilbert-Schmidt operators $\tilde{G}_i$ such that $A_i^{-1} = I + \tilde{G}_i$ (and $\sigma(\tilde{G}_i) > -1$). In this case,

$$A_1A_2 + (A_2A_1)^{-1} - 2I = G_1 + G_2 + G_1G_2 + \tilde{G}_1 + \tilde{G}_2 + \tilde{G}_1\tilde{G}_2.$$

Since $G_i + \tilde{G}_i = A_i + A_i^{-1} - 2I$ are trace class, and the product of two Hilbert-Schmidt operators is trace class too, we find that $A_1A_2 + (A_2A_1)^{-1} - 2I$ is a trace class operator and that

$$\operatorname{Tr}\left(A_1A_2 + (A_2A_1)^{-1} - 2I\right) = \operatorname{Tr}\left(A_1A_2 + (A_1A_2)^{-1} - 2I\right).$$

Similarly, $A_1A_2^{-1} + A_1^{-1}A_2 - 2I$ is trace class and hence

$$\text{Tr}\left(A_1A_2 + (A_1A_2)^{-1} - 2I\right) + \text{Tr}\left(A_1A_2^{-1} + A_2A_1^{-1} - 2I\right)$$
$$= 2\,\text{Tr}\left(A_1 + A_1^{-1} - 2I\right) + 2\,\text{Tr}\left(A_2 + A_2^{-1} - 2I\right)$$
$$+ \text{Tr}\left(A_1 + A_1^{-1} - 2I\right)\left(A_2 + A_2^{-1} - 2I\right).$$

Note that

$$\text{Tr}\left(A_1A_2 + (A_1A_2)^{-1} - 2I\right) = \text{Tr}\left(A_2^{\frac{1}{2}}A_1A_2^{\frac{1}{2}} + A_2^{-\frac{1}{2}}A_1^{-1}A_2^{-\frac{1}{2}} - 2I\right) \geq 0,$$

and that for the positive-definite trace class operators $F_i = A_i + A_i^{-1} - 2$, (with $\{x_j\}$ the orthonormal basis of eigenvectors of F_2),

$$\text{Tr}\,F_1F_2 = \sum_j \langle F_1F_2x_j, x_j \rangle \leq \sum_j \|F_2\|\langle F_1x_j, x_j\rangle \leq (\text{Tr}\,F_1)(\text{Tr}\,F_2),$$

so that

$$(7) \quad \text{Tr}\left(A_1A_2^{-1} + A_2A_1^{-1} - 2I\right)$$
$$\leq 2\,\text{Tr}\left(A_1 + A_1^{-1} - 2I\right) + 2\,\text{Tr}\left(A_2 + A_2^{-1} - 2I\right)$$
$$+ \text{Tr}\left(A_1 + A_1^{-1} - 2I\right)\text{Tr}\left(A_2 + A_2^{-1} - 2I\right).$$

We can assume that $H_c(Q_1, P) < \infty$, and $H_c(P, Q_2) < \infty$. In this case $A_i \stackrel{d}{=} R^{-\frac{1}{2}}S_iR^{-\frac{1}{2}}$ satisfy the conditions imposed above and it can be verified that $R^{-\frac{1}{2}}S_1^{\frac{1}{2}}$ and $S_1^{-\frac{1}{2}}R^{\frac{1}{2}}$ are well defined and bounded, so

$$H_c(Q_1, Q_2)$$
$$= \frac{1}{2}\,\text{Tr}\left[S_1^{-\frac{1}{2}}S_2S_1^{-\frac{1}{2}} + S_1^{\frac{1}{2}}S_2^{-1}S_1^{\frac{1}{2}} - 2I\right]$$
$$= \frac{1}{2}\,\text{Tr}\left[R^{-\frac{1}{2}}S_1^{\frac{1}{2}}(S_1^{-\frac{1}{2}}S_2S_1^{-\frac{1}{2}} + S_1^{\frac{1}{2}}S_2^{-1}S_1^{\frac{1}{2}} - 2I)S_1^{-\frac{1}{2}}R^{\frac{1}{2}}\right]$$
$$= \frac{1}{2}\,\text{Tr}\left[(R^{-\frac{1}{2}}S_2R^{-\frac{1}{2}})(R^{\frac{1}{2}}S_1^{-1}R^{\frac{1}{2}}) + (R^{-\frac{1}{2}}S_1R^{-\frac{1}{2}})(R^{\frac{1}{2}}S_2^{-1}R^{\frac{1}{2}}) - 2I\right]$$
$$= \frac{1}{2}\,\text{Tr}\left[A_1A_2^{-1} + A_2A_1^{-1} - 2I\right]$$
$$\leq 2H_c(Q_1, P) + 2H_c(P, Q_2) + 2H_c(Q_1, P)H_c(P, Q_2).$$

The last inequality is due to (7) and Lemma 1.8. □

Lemma 2.2. *If $H(Q_1, P) < \infty$ and $H(P, Q_2) < \infty$, then*

$$H_m(Q_1, Q_2) \leq 4H_m(Q_1, P) + 4H_m(P, Q_2)$$
$$+ 4H_m(Q_1, P)H_c(P, Q_2) + 4H_c(Q_1, P)H_m(P, Q_2).$$

Proof. We can assume without loss of generality that $\mu \equiv 0$, thus $\nu_1, \nu_2 \in \mathfrak{D}(R^{-\frac{1}{2}}) = \mathfrak{D}(S_1^{-\frac{1}{2}}) = \mathfrak{D}(S_2^{-\frac{1}{2}})$. The next claim is the heart of the proof.

Claim 2.3.

$$\left|S_1^{-\frac{1}{2}}\nu_2\right|^2 \le 2\left|R^{-\frac{1}{2}}\nu_2\right|^2 + 2H_c(Q_1,P)\left|R^{-\frac{1}{2}}\nu_2\right|^2.$$

Proof. Since $H_c(Q_1,P) < \infty$, $T = R^{-\frac{1}{2}}S_1 R^{-\frac{1}{2}}$ is a symmetric, positive-definite and bounded operator with a bounded inverse $T^{-1} = R^{\frac{1}{2}}S_1^{-1}R^{\frac{1}{2}}$ (e.g. [**4**, thm I.3.2]). Note that for $x \ge 0$,

$$\frac{1}{x} \le \left(\frac{1}{x} + x - 2\right) + 2,$$

so for the symmetric, positive-definite operator T we have

$$T^{-1} \le \left(T^{-1} + T - 2I\right) + 2I = F + 2I.$$

It follows that

$$\left|S_1^{-\frac{1}{2}}\nu_2\right|^2 = \langle T^{-1}R^{-\frac{1}{2}}\nu_2, R^{-\frac{1}{2}}\nu_2\rangle \quad \le \langle FR^{-\frac{1}{2}}\nu_2, R^{-\frac{1}{2}}\nu_2\rangle + 2\left|R^{-\frac{1}{2}}\nu_2\right|^2.$$

We finish by noting that for any symmetric, positive-definite F and a vector x,

$$\langle Fx, x\rangle \le \operatorname{Tr} F|x|^2.$$

$\square$

Using the last claim we prove the lemma:

$$\begin{aligned}
H_m(Q_1,Q_2) &= \frac{1}{2}\left|S_1^{-\frac{1}{2}}(\nu_1-\nu_2)\right|^2 + \frac{1}{2}\left|S_2^{-\frac{1}{2}}(\nu_1-\nu_2)\right|^2 \\
&\le \left|S_1^{-\frac{1}{2}}\nu_1\right|^2 + \left|S_1^{-\frac{1}{2}}\nu_2\right|^2 + \left|S_2^{-\frac{1}{2}}\nu_1\right|^2 + \left|S_2^{-\frac{1}{2}}\nu_2\right|^2 \\
&\le \left|S_1^{-\frac{1}{2}}\nu_1\right|^2 + 2\left|R^{-\frac{1}{2}}\nu_2\right|^2 + 2H_c(Q_1,P)\left|R^{-\frac{1}{2}}\nu_2\right|^2 \\
&\quad + \left|S_2^{-\frac{1}{2}}\nu_2\right|^2 + 2\left|R^{-\frac{1}{2}}\nu_1\right|^2 + 2H_c(P,Q_2)\left|R^{-\frac{1}{2}}\nu_1\right|^2 \\
&\le 4H_m(Q_1,P) + 4H_m(P,Q_2) \\
&\quad + 4H_m(Q_1,P)H_c(P,Q_2) + 4H_c(Q_1,P)H_m(P,Q_2).
\end{aligned}$$

$\square$

Theorem 1 is now a trivial consequence of the last couple of lemmas. $\square$

Appendix A. Concrete Examples.

A.1. Triangle correlation.

$$S(t,s) = 1 - \frac{|t-s|}{2} \quad \text{with} \quad t,s \in [1-\tau, 1+\tau] \quad \text{and} \quad |t-s| < 4.$$

In this case, the Radon-Nikodym derivatives can be found explicitly. One way is by computing

(8)

$$R^{-1} = \begin{bmatrix} \frac{1}{\delta} + \frac{1}{1-\tau} & -\frac{1}{\delta} & 0 & 0 & \cdots & 0 \\ -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} & 0 & \cdots & 0 \\ 0 & -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} & \cdots & 0 \\ \vdots & & & & \ddots & \\ 0 & 0 & \cdots & -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} \\ 0 & 0 & \cdots & 0 & -\frac{1}{\delta} & \frac{1}{\delta} \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} \frac{1}{\delta} + \frac{1}{4-2\tau} & -\frac{1}{\delta} & 0 & \cdots & 0 & \frac{1}{4-2\tau} \\ -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} & 0 & \cdots & 0 \\ 0 & -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} & \cdots & 0 \\ \vdots & & & & \ddots & \\ 0 & \cdots & 0 & -\frac{1}{\delta} & \frac{2}{\delta} & -\frac{1}{\delta} \\ \frac{1}{4-2\tau} & 0 & \cdots & 0 & -\frac{1}{\delta} & \frac{1}{\delta} + \frac{1}{4-2\tau} \end{bmatrix}$$

$$S^{-1} - R^{-1} = \begin{bmatrix} \frac{\tau-3}{(4-2\tau)(1-\tau)} & 0 & \cdots & 0 & \frac{1}{4-2\tau} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{4-2\tau} & 0 & \cdots & 0 & \frac{1}{4-2\tau} \end{bmatrix}$$

$$S^{-1}R = \begin{bmatrix} \frac{1-\tau}{2-\tau} & * & * & \cdots & * & -\frac{1-\tau}{2-\tau} \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ \frac{1-\tau}{2-\tau} & * & * & \cdots & * & \frac{3-\tau}{2-\tau} \end{bmatrix},$$

from which you find

$$\det S^{-1}R = \frac{1-\tau}{2-\tau}\frac{3-\tau}{2-\tau} + \frac{1-\tau}{2-\tau}\frac{1-\tau}{2-\tau} = \frac{2(1-\tau)}{2-\tau},$$

and

$$
\frac{dW_n}{dQ_n}
$$

$$
= \frac{1}{\sqrt{\det S^{-1}R}} \exp\left\{ \frac{1}{2} \left\langle (S^{-1} - R^{-1})\boldsymbol{x}, \boldsymbol{x} \right\rangle \right\}
$$

$$
= \sqrt{\frac{2-\tau}{2(1-\tau)}} \exp\left\{ \frac{1}{2} \left(\frac{\tau-3}{(4-2\tau)(1-\tau)} \boldsymbol{x}_0^2 + \frac{2}{4-2\tau} \boldsymbol{x}_0 \boldsymbol{x}_n + \frac{1}{4-2\tau} \boldsymbol{x}_n^2 \right) \right\}.
$$

This is just the Radon-Nikodym derivative of the two distributions sampled only at the end points, i.e., it is the Radon-Nikodym derivative of the 2 dimensional (mean 0) Gaussian vectors with correlations $R_0 = \left(\begin{smallmatrix} 1-\tau & 1-\tau \\ 1-\tau & 1+\tau \end{smallmatrix} \right)$ and $S_0 = \left(\begin{smallmatrix} 1 & 1-\tau \\ 1-\tau & 1 \end{smallmatrix} \right)$.

Claim A.1. $\frac{dW_n}{dQ_n}$ *is independent of* n, *therefore it is* $\frac{dW}{dQ}$.

We just showed that $\frac{dW_n}{dQ_n}$ is independent of n. Let $\mathcal{F}_n$ be the σ-field generated by sampling the paths at $2^n + 1$ points. Then $\frac{dW_n}{dQ_n}$ is $\mathcal{F}_n$ measurable, and $\left(\frac{dW_n}{dQ_n}, \mathcal{F}_n \right)$ is a martingale which obviously converges in $L^1(dQ)$ to itself, i.e. it *is* $\frac{dW}{dQ}$.

There is an alternative way to find $\frac{dW}{dQ}$. Let $L(t; \alpha, \beta)$ be the linear interpolation between the points (τ, α) and (σ, β), i.e.,

$$
L(t; \alpha, \beta) \overset{d}{=} \frac{t-\tau}{\sigma-\tau}(\beta - \alpha) + \alpha.
$$

Let $X \in C[\tau, \sigma]$ be a generic path, and define $Y \in C[\tau, \sigma]$ as

$$
Y_t \overset{d}{=} X_t - L(t; X_\tau, X_\sigma).
$$

It is easy to check that $E^Q(Y_t X_\tau) = E^Q(Y_t X_\sigma) = 0$ and also $E^W(Y_t X_\tau) = E^W(Y_t X_\sigma) = 0$. As the means are 0, we find that Y is independent of X_τ and X_σ under both measures. Furthermore,

$$
E^W Y_t Y_s = \frac{(s-\tau)(\sigma-t)}{\sigma-\tau} = E^Q Y_t Y_s \qquad \tau \le s \le t \le \sigma,
$$

that is, Y has exactly the same distribution under W as under Q. In fact, under both measures, Y is just a tied Brownian motion. Here is the gist of how to use the above for computing $\frac{dW}{dQ}$. Let $f_{\tau,s,t,\sigma}^{W/Q}$ be the joint density of $(X_\tau, X_s, X_t, X_\sigma)$ under W, respectively Q. Let $f_{(s,t)|(\tau,\sigma)}^{W/Q}$ be the associated conditional density. Obviously,

$$
f_{\tau,s,t,\sigma}(x_\tau, x_s, x_t, x_\sigma) = f_{(s,t)|(\tau,\sigma)}(x_s, x_t | x_\tau, x_\sigma) f_{(\tau,\sigma)}(x_\tau, x_\sigma)
$$

and also

$$
f_{(s,t)|(\tau,\sigma)}(x_s, x_t | x_\tau, x_\sigma)
$$

$$= f_{(Y_s,Y_t)|(X_\tau,X_\sigma)} \left(x_t - L(t; x_\tau, x_\sigma), x_s - L(s; x_\tau, x_\sigma) \big| (x_\tau, x_\sigma) \right)$$
$$= f_{(Y_s,Y_t)} \left(x_t - L(t; x_\tau, x_\sigma), x_s - L(s; x_\tau, x_\sigma) \right)$$

as Y is independent of the end points of the X process under both measures. Hence

$$f^Q_{\tau,s,t,\sigma}(x_\tau, x_s, x_t, x_\sigma)$$
$$= f^Q_{(Y_s,Y_t)} \left(x_t - L(t; x_\tau, x_\sigma), x_s - L(s; x_\tau, x_\sigma) \right) f^Q_{(\tau,\sigma)}(x_\tau, x_\sigma)$$
$$= f^W_{\tau,s,t,\sigma}(x_\tau, x_s, x_t, x_\sigma) \frac{f^Q_{(\tau,\sigma)}(x_\tau, x_\sigma)}{f^W_{(\tau,\sigma)}(x_\tau, x_\sigma)}.$$

A standard argument now shows that $\frac{dW}{dQ} = \frac{f^Q_{(\tau,\sigma)}(x_\tau,x_\sigma)}{f^W_{(\tau,\sigma)}(x_\tau,x_\sigma)}$.

Once you have $\frac{dW}{dQ}$, computing the entropy is trivial:

$$H = \frac{\tau}{2(1-\tau)}.$$

It should be noted that the entropy itself can be recovered by computing T (cf (1)) and T^{-1}; both are rather easy to compute due to their special structure.

A.2. The Ornstein-Uhlenbeck Process.

$$S(t,s) = e^{-|t-s|/2}.$$

Here also you can compute the Radon-Nikodym derivative. Let W_T be the Wiener measure on $C[T, T+2\tau]$. By the chain rule,

$$\frac{dQ}{dW_{1-\tau}} = \frac{dQ}{dW_1} \frac{dW_1}{dW_{1-\tau}}.$$

The Ornstein-Uhlenbeck process satisfies: $dX = db - \frac{1}{2}X\,dt$, with a standard Brownian motion b, so Girsanov's formula implies:

$$\frac{dQ}{dW_1} = \exp \left\{ -\int_0^{2\tau} \frac{1}{2} X_s \, dX_s - \frac{1}{2} \int_0^{2\tau} \frac{1}{4} X_s^2 \, ds \right\}$$
$$= \exp \left\{ -\frac{1}{4} \left[X_t^2 - t \right]_0^{2\tau} - \frac{1}{8} \int_0^{2\tau} X_s^2 \, ds \right\}$$

and, with considerably less effort, we get

$$\frac{dW_1}{dW_{1-\tau}} = \sqrt{1-\tau} \, \exp \left\{ -\frac{1}{2} X_0^2 \left(1 - \frac{1}{1-\tau} \right) \right\}.$$

Now you have the Radon-Nikodym derivative in front of you and from it you compute:

$$E^Q \log \frac{dQ}{dW_{1-\tau}} = \frac{1}{2} \log(1-\tau) - \frac{1}{2} + \frac{\tau}{4} + \frac{1}{2(1-\tau)}$$

$$E^{W_{1-\tau}} \log \frac{dW_{1-\tau}}{dQ} = -\frac{1}{2} \log(1-\tau) - \frac{\tau}{4}.$$

Finally,

$$H = \frac{1}{2} \frac{\tau}{1-\tau},$$

exactly as for the triangle correlation. In this case, you can also compute the Radon-Nikodym derivative as a limit of finite-dimensional derivatives since the matrix S is readily invertible.

Appendix B. The existence of a minimizer.

The following is adopted from [1, pp. 32-40]. Let X be a Polish space, i.e., a complete separable metric space, and let $\mathcal{P}(X)$ denote the set of probability measures on X. Using weak-convergence one can introduce a (metrizable) topology on $\mathcal{P}(X)$. Let $P, Q \in \mathcal{P}(X)$, then the relative entropy of P given Q is defined by

$$H(P|Q) \overset{d}{=} E^P \log \frac{dP}{dQ}.$$

It can be shown that $H(P|Q)$ is a convex lower semicontinuous function of $(P, Q) \in \mathcal{P}(X) \times \mathcal{P}(X)$, and that for every finite M the set $\{Q : H(Q|P) \leq M\}$ is compact. It follows that $H(P, Q) = H(P|Q) + H(Q|P)$ is a lower semicontinuous function of (P, Q). Thus, if for a given P the set $\{Q : H(Q|P) < \infty\}$ is non empty, there exists a $Q_0 \in \mathcal{P}(X)$ which minimizes $H(P, Q)$. Finally, since the set of stationary Gaussian measures is closed, the above argument will yield a stationary minimizer in our case as well.

Appendix C. On the L^2 distance.

The L^2 distance between correlations R and S is

$$\|R - S\|^2 = \int_{1-\tau}^{1+\tau} \int_{1-\tau}^{1+\tau} |R(t, s) - S(t, s)|^2 \, dt \, ds.$$

If R is any correlation, then the best L^2 approximation to R by a Töeplitz operator is obtained by averaging R along its diagonals:

$$S(r) = \frac{1}{2\tau - r} \int_{1-\tau}^{1+\tau-r} R(s+r, s) \, ds \qquad 0 \leq r \leq 2\tau.$$

The problem is that, in general, S would not be positive-definite. Besides, this L^2 distance seems unnatural from a probabilistic standpoint. In the case of $R = t \wedge s$, S turns out to be $1 - |r|/2$ which happens to be positive-definite (see appendix A.1) and we get $\|R - S\|^2 = \frac{2}{3}\tau^4$. More generally, if S is stationary with $S(r) = 1 - \frac{|r|}{2} + o(r)$, then $\|R - S\|^2 = \frac{2}{3}\tau^4 + o(\tau^4)$ so, as in the case of the entropy, we find (to leading order) that the distance between

Brownian motion and any stationary process that is absolutely continuous with respect to it is always the same. This may be contrasted with the entropy distance between Brownian motion and $S(r) = 1 - \frac{|r|}{2} + |r|^{3/2}$ which is infinite.

It is also easily verified that the L^2 distance between any two stationary correlations of the above type $(S(r) = 1 - \frac{|r|}{2} + o(r))$ is $o(\tau^2)$ which is reminiscent of the case of the entropy distance in the sense that it is smaller than the distance to the Brownian motion.

The L^2 (as opposed to the entropy) distance to Brownian motion remains finite for stationary correlations such as $S(r) = 1 - \alpha|r| + o(|r|)$ with $\alpha \neq \frac{1}{2}$:

$$\|S - R\|^2 = \frac{4}{3}(2\alpha^2 - 2\alpha + 1)\tau^4 + o(\tau^4).$$

This is minimized for $\alpha = \frac{1}{2}$, as it should be.

References

[1] P. Dupuis and R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, Inc, 1997.

[2] T. Hida and M. Hitsuda, *Gaussian Processes*, AMS, 1993.

[3] U. Keich, *Absolute continuity between the Wiener and Stationary Gaussian Measures*, preprint, 1998.

[4] H.H. Kuo, *Gaussian Measures in Banach Spaces*, Lecture Notes in Mathematics, **463**, Springer-Verlag, 1975.

[5] R.C. Rao and V.S. Varadarajan, *Discrimination of Gaussian processes*, Sankhyā Ser. A, **25** (1963), 303-330.

[6] J. Sekine, *The Hilbert Riemannian structure of equivalent Gaussian measures associated with the Fisher information*, **32** (1995), 71-95.

[7] L.A. Shepp, *Radon-Nikodym derivatives of Gaussian measures*, Ann. Math. Stat., **37** (1966), 321-354.

[8] Y. Yamasaki, *Measures on Infinite Dimensional Spaces*, World Scientific Publishing Co. Pte. Ltd., 1985.

CALTECH
PASADENA, CA 91125
E-mail address: keich@ama.caltech.edu