The Spectral-Domain $\mathcal{W}_2$ Wasserstein Distance for Elliptical Processes and the Spectral-Domain Gelbrich Bound

Song Fang$^1$ and Quanyan Zhu$^1$

Abstract—In this short note, we introduce the spectral-domain $\mathcal{W}_2$ Wasserstein distance for elliptical stochastic processes in terms of their power spectra. We also introduce the spectral-domain Gelbrich bound for processes that are not necessarily elliptical.

I. INTRODUCTION

The Wasserstein distance (see, e.g., [1], [2] and the references therein) is an important metric from optimal transport theory (see, e.g., [3]–[5] and the references therein). In this note, we consider the average $\mathcal{W}_2$ Wasserstein distance between stationary stochastic processes that are elliptical with the same density generator, which can be naturally characterized by a spectral-domain expression in terms of the power spectra of the processes (Theorem I). On the other hand, when the stochastic processes are not necessarily elliptical, the average $\mathcal{W}_2$ Wasserstein distance is lower bounded by a spectral-domain generalization of the Gelbrich bound (Corollary I).

II. PRELIMINARIES

Throughout the note, we consider zero-mean real-valued continuous random variables and random vectors, as well as discrete-time stochastic processes. We represent random variables and random vectors using boldface letters, e.g., $\mathbf{x}$, while the probability density function of $\mathbf{x}$ is denoted as $p_{\mathbf{x}}$. We denote the sequence $\mathbf{x}_0, \ldots, \mathbf{x}_k$ by $\mathbf{x}_{0:k}$ for simplicity, which, by a slight abuse of notation, is also identified with the random vector $[\mathbf{x}_T^\top, \ldots, \mathbf{x}_{T+k}^\top]^\top$.

We denote the covariance matrix of an $m$-dimensional random vector $\mathbf{x}$ by $\Sigma_{\mathbf{x}} = \mathbb{E} [\mathbf{x} \mathbf{x}^\top]$. In the scalar case, the variance of $\mathbf{x}$ is denoted by $\sigma^2_{\mathbf{x}}$. An $m$-dimensional stochastic process $\{\mathbf{x}_k\}$ is said to be stationary if its auto-correlation $R_{\mathbf{x}}(i,k) = \mathbb{E} [\mathbf{x}_i \mathbf{x}_{i+k}^\top]$ depends only on $k$, and can thus be denoted as $R_{\mathbf{x}}(k)$ for simplicity. The power spectrum of a stationary process $\{\mathbf{x}_k\}$ is then defined as

$$\Phi_{\mathbf{x}}(\omega) = \sum_{k=-\infty}^{\infty} R_{\mathbf{x}}(k) e^{-j\omega k}.$$ 

In the scalar case, the power spectrum of $\{\mathbf{x}_k\}$ is denoted as $S_{\mathbf{x}}(\omega)$. Note that throughout the note, all covariance matrices and power spectra are assumed to be positive definite.

The $\mathcal{W}_p$ Wasserstein distance (see, e.g., [1], [2]) is defined as follows.

**Definition 1:** The $\mathcal{W}_p$ (for $p \geq 1$) Wasserstein distance between distribution $p_{\mathbf{x}}$ and distribution $p_{\mathbf{y}}$ is defined as

$$\mathcal{W}_p (p_{\mathbf{x}}; p_{\mathbf{y}}) = \left( \inf_{x,y} \mathbb{E} [||x - y||^p] \right)^{\frac{1}{p}},$$

where $x$ and $y$ denote $m$-dimensional random vectors with distributions $p_{\mathbf{x}}$ and $p_{\mathbf{y}}$, respectively.

Particularly when $p = 2$, the $\mathcal{W}_2$ distance is given by

$$\mathcal{W}_2 (p_{\mathbf{x}}; p_{\mathbf{y}}) = \sqrt{\mathbb{E} [||x - y||^2]}.$$

The following lemma (see, e.g., [1], [2], [5]) provides an explicit expression for the $\mathcal{W}_2$ distance between elliptical distributions with the same density generator. Note that Gaussian distributions are a special class of elliptical distributions (see, e.g., [1]). Note also that hereinafter the random vectors are assumed to be zero-mean for simplicity.

**Lemma 1:** Consider $m$-dimensional elliptical random vectors $\mathbf{x}$ and $\mathbf{y}$ with the same density generator, while with covariance matrices $\Sigma_{\mathbf{x}}$ and $\Sigma_{\mathbf{y}}$, respectively. The $\mathcal{W}_2$ distance between distribution $p_{\mathbf{x}}$ and distribution $p_{\mathbf{y}}$ is given by

$$\mathcal{W}_2 (p_{\mathbf{x}}; p_{\mathbf{y}}) = \sqrt{\text{tr} \left( \Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}} - 2 (\Sigma_{\mathbf{x}}^{\frac{1}{2}} \Sigma_{\mathbf{y}} \Sigma_{\mathbf{x}}^{\frac{1}{2}})^\frac{1}{2} \right)}.$$

Meanwhile, the Gelbrich bound (see, e.g., [1]) is given as follows, which provides a generic lower bound for the $\mathcal{W}_2$ distance between distributions that are not necessarily elliptical.

**Lemma 2:** Consider $m$-dimensional random vectors $\mathbf{x}$ and $\mathbf{y}$ with covariance matrices $\Sigma_{\mathbf{x}}$ and $\Sigma_{\mathbf{y}}$, respectively. The $\mathcal{W}_2$ distance between distribution $p_{\mathbf{x}}$ and distribution $p_{\mathbf{y}}$ is lower bounded by

$$\mathcal{W}_2 (p_{\mathbf{x}}; p_{\mathbf{y}}) \geq \sqrt{\text{tr} \left( \Sigma_{\mathbf{x}} + \Sigma_{\mathbf{y}} - 2 (\Sigma_{\mathbf{x}}^{\frac{1}{2}} \Sigma_{\mathbf{y}} \Sigma_{\mathbf{x}}^{\frac{1}{2}})^\frac{1}{2} \right)}.$$

III. SPECTRAL-DOMAIN $\mathcal{W}_2$ WASSERSTEIN DISTANCE AND GELBRICH BOUND

We first present the definition of average Wasserstein distance for stochastic processes.

**Definition 2:** Consider $m$-dimensional stochastic processes $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$. The average $\mathcal{W}_p$ distance between $p_{\{\mathbf{x}_k\}}$ and $p_{\{\mathbf{y}_k\}}$ is defined as

$$\mathcal{W}_p \left( p_{\{\mathbf{x}_k\}}; p_{\{\mathbf{y}_k\}} \right) = \left( \inf_{\{\mathbf{x}_k\}, \{\mathbf{y}_k\}} \limsup_{i \to \infty} \mathbb{E} [||\mathbf{x}_{0:i} - \mathbf{y}_{0:i}||^p] \right)^{\frac{1}{p}}.$$  (1)

---

1 Song Fang and Quanyan Zhu are with the Department of Electrical and Computer Engineering, New York University, New York, USA song.fang@nyu.edu; quanyan.zhu@nyu.edu
In the case of \( p = 2 \), the average \( \mathcal{W}_2 \) distance is defined as
\[
\mathcal{W}_2 \left( p(x_k) ; p(y_k) \right) = \left( \inf_{\{x_k\} : \{y_k\}} \lim_{i \to \infty} \sup_{i \to \infty} \frac{E \left[ \|x_{0,\ldots,i} - y_{0,\ldots,i}\|_2^2 \right]}{i + 1} \right)^{\frac{1}{2}}.
\] (2)

We now proceed to present the main results of this note. The following Theorem provides a spectral-domain expression for the average \( \mathcal{W}_2 \) distance between elliptical processes.

**Theorem 1:** Consider \( m \)-dimensional stationary stochastic processes \( \{x_k\} \) and \( \{y_k\} \) that are elliptical with the same density generator. Suppose that their distributions are given respectively by \( p(x_k) \) and \( p(y_k) \), while the power spectra are given respectively as \( \Phi_X(\omega) \) and \( \Phi_Y(\omega) \). The average \( \mathcal{W}_2 \) distance between \( p(x_k) \) and \( p(y_k) \) is given by
\[
\mathcal{W}_2 \left( p(x_k) ; p(y_k) \right) = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} \text{tr} \left\{ W \left[ \Phi_X(\omega), \Phi_Y(\omega) \right] \right\} d\omega},
\] (3)

where
\[
W \left[ \Phi_X(\omega), \Phi_Y(\omega) \right] = \Phi_X(\omega) + \Phi_Y(\omega) - 2 \left[ \Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega) \right]^\frac{1}{2}.
\] (4)

**Proof:** Note first that since \( \{x_k\} \) and \( \{y_k\} \) are stationary, we have
\[
\mathcal{W}_2 \left( p(x_k) ; p(y_k) \right) = \left( \inf_{\{x_k\} : \{y_k\}} \lim_{i \to \infty} \sup_{i \to \infty} \frac{E \left[ \|x_{0,\ldots,i} - y_{0,\ldots,i}\|_2^2 \right]}{i + 1} \right)^{\frac{1}{2}}
\]
\[
= \left( \inf_{\{x_k\} : \{y_k\}} \lim_{i \to \infty} \sup_{i \to \infty} \frac{E \left[ \|x_{0,\ldots,i} - y_{0,\ldots,i}\|_2^2 \right]}{i + 1} \right)^{\frac{1}{2}}
\]
\[
= \lim_{i \to \infty} \inf_{x_{0,\ldots,i}, y_{0,\ldots,i}} E \left[ \|x_{0,\ldots,i} - y_{0,\ldots,i}\|_2^2 \right]^{\frac{1}{2}}
\]
It then follows from Definition and Lemma that
\[
\inf_{x_{0,\ldots,i}, y_{0,\ldots,i}} E \left[ \|x_{0,\ldots,i} - y_{0,\ldots,i}\|_2^2 \right] = \text{tr} \left[ \Sigma_{x_{0,\ldots,i}} + \Sigma_{y_{0,\ldots,i}} - 2 \left( \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \right)^{\frac{1}{2}} \right]
\]
\[
= \text{tr} \left( \Sigma_{x_{0,\ldots,i}} \right) + \text{tr} \left( \Sigma_{y_{0,\ldots,i}} \right) - 2 \text{tr} \left( \left( \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \right)^{\frac{1}{2}} \right),
\]
since \( x_{0,\ldots,i} \) and \( y_{0,\ldots,i} \) are elliptical with the same density generator. Meanwhile, it is known from [6], [7] that
\[
\lim_{i \to \infty} \frac{\text{tr} \left( \Sigma_{x_{0,\ldots,i}} \right)}{i + 1} = \frac{1}{2\pi} \int_0^{2\pi} \text{tr} \left\{ \Phi_X(\omega) \right\} d\omega,
\]
and
\[
\lim_{i \to \infty} \frac{\text{tr} \left( \Sigma_{y_{0,\ldots,i}} \right)}{i + 1} = \frac{1}{2\pi} \int_0^{2\pi} \text{tr} \left\{ \Phi_Y(\omega) \right\} d\omega.
\]
It remains to prove that
\[
\lim_{i \to \infty} \frac{\text{tr} \left( \left( \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \right)^{\frac{1}{2}} \right)}{i + 1}
\]
\[
= \frac{1}{2\pi} \int_0^{2\pi} \text{tr} \left\{ \left( \Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega) \right)^{\frac{1}{2}} \right\} d\omega.
\]
Note that \( \Phi_X(\omega) \) and \( \Phi_Y(\omega) \) are positive definite. As such, \( \Phi_X^\frac{1}{2}(\omega) \) is also positive definite (thus invertible), and
\[
\Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega) = \Phi_X^\frac{1}{2}(\omega) \left[ \Phi_X(\omega) \Phi_Y(\omega) \right] \Phi_X^\frac{1}{2}(\omega),
\]
meaning that
\[
\Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega)
\]
are similar. Consequently, they share the same eigenvalues (which are all positive since \( \Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega) \) is positive definite), while the square roots of the eigenvalues are also the same. Hence,
\[
\text{tr} \left\{ \Phi_X^\frac{1}{2}(\omega) \Phi_Y(\omega) \Phi_X^\frac{1}{2}(\omega) \right\} = \text{tr} \left\{ \Phi_X(\omega) \Phi_Y(\omega) \right\}^{\frac{1}{2}}.
\]
Similarly, it can be proved that
\[
\lim_{i \to \infty} \frac{\text{tr} \left( \left( \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \right)^{\frac{1}{2}} \right)}{i + 1}
\]
\[
= \lim_{i \to \infty} \frac{\sum_{j=0}^{m-1} \lambda_j \left( \Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}} \right)^{\frac{1}{2}}}{i + 1}
\]
\[
= \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=1}^{m} \lambda_j \left( \Phi_X(\omega) \Phi_Y(\omega) \right)^{\frac{1}{2}} d\omega,
\]
\[
= \frac{1}{2\pi} \int_0^{2\pi} \sum_{j=1}^{m} \lambda_j \left( \Phi_X(\omega) \Phi_Y(\omega) \right)^{\frac{1}{2}} d\omega.
\]
Herein, we have used the fact that the eigenvalues of \( (\Sigma_{x_{0,\ldots,i}} \Sigma_{y_{0,\ldots,i}})^{\frac{1}{2}} \) are given by the square roots of
the eigenvalues of \( \Sigma_{x_0,\ldots,x_{y-1}} \), with the former being denoted as \( \lambda_j \left( \Sigma_{x_0,\ldots,x_{y-1}} \right)^\dagger \) while the latter as \( \lambda_j \left( \Sigma_{x_0,\ldots,x_{y-1}} \right) \). Similarly, the eigenvalues of \( \{ \Phi_x (\omega) \Phi_y (\omega) \} \) are given by the square roots of the eigenvalues of \( \Phi_x (\omega) \Phi_y (\omega) \), with the former being denoted as \( \lambda_j \left( \Phi_x (\omega) \Phi_y (\omega) \right)^\dagger \) while the latter as \( \lambda_j \left( \Phi_x (\omega) \Phi_y (\omega) \right) \). Consequently,

\[
\lim_{i \to \infty} \left[ \frac{i}{i+1} \right] = \frac{1}{2\pi} \int_0^{2\pi} \left\{ \Phi_x (\omega) \Phi_y (\omega) \right\} d\omega = \frac{1}{2\pi} \int_0^{2\pi} \left\{ \Phi_x (\omega) \Phi_y (\omega) \right\} d\omega.
\]

This concludes the proof.\[ \blacksquare \]

Note that it is known from the proof of Theorem I that

\[
\text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right] = \text{tr \[ \Phi_x (\omega) \Phi_y (\omega) \].}
\]

Accordingly, \( W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] \) can equivalently be rewritten as

\[
W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] = \Phi_x (\omega) + \Phi_y (\omega) - 2 \left[ \Phi_x (\omega) \Phi_y (\omega) \right]^{\frac{1}{2}}\]

(6)

Moreover, if it is further assumed that \( \Phi_x (\omega) \Phi_y (\omega) = \Phi_y (\omega) \Phi_x (\omega) \), then \( W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] \) reduces to

\[
W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] = \Phi_x (\omega) + \Phi_y (\omega) - 2 \Phi_x (\omega) \Phi_y (\omega) \]

(8)

and accordingly, \( W_2 \left( p_{(x)} ; p_{(y)} \right) \) reduces to

\[
W_2 \left( p_{(x)} ; p_{(y)} \right) = \sqrt{ \frac{1}{2\pi} \int_0^{2\pi} \left\{ \Phi_x (\omega) + \Phi_y (\omega) - 2 \Phi_x (\omega) \Phi_y (\omega) \right\} d\omega \}
\]

This coincides with the Hellinger distance proposed in [8]. However, in general when \( \Phi_x (\omega) \) and \( \Phi_y (\omega) \) do not necessarily commute, it can be verified that

\[
W_2 \left( p_{(x)} ; p_{(y)} \right) \geq \sqrt{ \frac{1}{2\pi} \int_0^{2\pi} \left\| \Phi_x (\omega) - \Phi_y (\omega) \right\|_F^2 d\omega,}
\]

(10)

since

\[
\text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right] \geq \text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right] = \text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right],
\]

(11)

where the inequality is due to the Araki–Lieb–Thirring inequality [9]; from this it also follows that

\[
\text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right] \geq \text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right] \geq \text{tr} \left[ \Phi_x (\omega) \Phi_y (\omega) \right],
\]

(12)

which is a property that may be useful in other settings; note that herein the equality has been proved in the proof of Theorem I.

Note also that in the scalar case (when \( m = 1 \)), supposing that the power spectra of \( \{ x_k \} \) and \( \{ y_k \} \) are given respectively by \( S_x (\omega) \) and \( S_y (\omega) \), the average \( W_2 \) distance between \( p_{(x)} \) and \( p_{(y)} \) is given by

\[
W_2 \left( p_{(x)} ; p_{(y)} \right) = \sqrt{ \frac{1}{2\pi} \int_0^{2\pi} \left\{ S_x (\omega) + S_y (\omega) - 2 \left[ S_x (\omega) S_y (\omega) \right]^{\frac{1}{2}} \right\} d\omega \}
\]

(13)

The subsequent Corollary I presents the spectral-domain Gelbrich bound for processes that are not necessarily elliptical, which follows directly from Theorem I.

Corollary I: Consider \( m \)-dimensional stationary stochastic processes \( \{ x_k \} \) and \( \{ y_k \} \) that are not necessarily elliptical. Suppose that their distributions are given respectively as \( \Phi_x (\omega) \) and \( \Phi_y (\omega) \). The average \( W_2 \) distance between \( p_{(x)} \) and \( p_{(y)} \) is lower bounded by

\[
W_2 \left( p_{(x)} ; p_{(y)} \right) \geq \sqrt{ \frac{1}{2\pi} \int_0^{2\pi} \left\{ W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] \right\} d\omega \}
\]

(14)

where

\[
W \left[ \Phi_x (\omega) , \Phi_y (\omega) \right] = \Phi_x (\omega) + \Phi_y (\omega) - 2 \left[ \Phi_x (\omega) \Phi_y (\omega) \right]^{\frac{1}{2}}\]

(15)

In the scalar case, the spectral-domain Gelbrich bound is given by

\[
W_2 \left( p_{(x)} ; p_{(y)} \right) \geq \sqrt{ \frac{1}{2\pi} \int_0^{2\pi} \left\{ S_x (\omega) + S_y (\omega) - 2 \left[ S_x (\omega) S_y (\omega) \right]^{\frac{1}{2}} \right\} d\omega \}
\]

(16)

IV. CONCLUSION

In this note, we have introduced the spectral-domain \( W_2 \) Wasserstein distance for elliptical stochastic processes. We have also introduced the spectral-domain Gelbrich bound for processes that are not necessarily elliptical. It might be interesting to examine the implications of the results in future (cf. [10] for instance).
REFERENCES

[1] G. Peyré and M. Cuturi, “Computational optimal transport: With applications to data science,” Foundations and Trends® in Machine Learning, vol. 11, no. 5-6, pp. 355–607, 2019.

[2] V. M. Panaretos and Y. Zemel, An Invitation to Statistics in Wasserstein Space. Springer, 2020.

[3] C. Villani, Topics in Optimal Transportation. American Mathematical Society, 2003.

[4] ——, Optimal Transport: Old and New. Springer, 2008.

[5] F. Santambrogio, Optimal Transport for Applied Mathematicians. Birkhäuser, Springer, 2015.

[6] J. Gutiérrez-Gutiérrez and P. M. Crespo, ”Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems,” IEEE Transactions on Information Theory, vol. 54, no. 12, pp. 5671–5680, 2008.

[7] A. Lindquist and G. Picci, Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification. Springer, 2015.

[8] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger versus Kullback-Leibler multivariable spectrum approximation,” IEEE Transactions on Automatic Control, vol. 53, no. 4, pp. 954–967, 2008.

[9] H. Araki, “On an inequality of Lieb and Thirring,” Letters in Mathematical Physics, vol. 19, no. 2, pp. 167–170, 1990.

[10] S. Fang and Q. Zhu, “Independent elliptical distributions minimize their Wasserstein distance from independent elliptical distributions with the same density generator,” arXiv preprint, 2020.