Aequat. Math. 96 (2022), 85–114

0001-9054/22/010085-30
published online January 22, 2022
https://doi.org/10.1007/s00010-021-00861-3

Aequationes Mathematicae

Weighted Gaussian entropy and determinant inequalities

I. Stuhl, M. Kelbert, Y. Suhov, and S. Yasaei Sekeh

Abstract. We produce a series of results extending information-theoretical inequalities (discussed by Dembo–Cover–Thomas in 1988–1991) to a weighted version of entropy. Most of the resulting inequalities involve the Gaussian weighted entropy; they imply a number of new relations for determinants of positive-definite matrices. Unlike the Shannon entropy where the contribution of an outcome depends only upon its probability, the weighted (or context-dependent) entropy takes into account a 'value' of an outcome determined by a given weight function φ. An example of a new result is a weighted version of the strong Hadamard inequality (SHI) between the determinants of a positive-definite $d \times d$ matrix and its square blocks (sub-matrices) of different sizes. When $\varphi \equiv 1$, the weighted inequality becomes a 'standard' SHI; in general, the weighted version requires some assumptions upon φ. The SHI and its weighted version generalize a widely known 'usual' Hadamard inequality $\det \mathbf{C} \leq \prod_{j=1}^{d} C_{jj}$.

Mathematics Subject Classification. 60A10, 60B05, 60C05.

Keywords. Weight function, Weighted entropy, Weighted conditional entropy, Weighted mutual information, Weighted Gaussian entropy, Determinant inequalities.

1. Introduction

The aim of this paper is to continue an analysis of the weighted entropy by following [18] and in particular to give a number of new inequalities involving the determinants of positive-definite matrices. These inequalities can be considered as generalizations of inequalities discussed in [3,4,6,12].

The inequalities presented in the current paper are obtained by a unified method which is based on *weighted entropies* (WEs), in particular, on Gaussian WEs. The concept of a WE was introduced in the late 1960s–early 1970s; see, e.g., [2]. (Another term that can be used is a context-dependent or a preferential entropy.) The reader is referred to [18] where a number of notions and elementary inequalities were established for the WE, mirroring well-known facts about the standard (Shannon) entropy. We follow the system of notation from [3,6,18] with minor deviations. Methodologically, we follow the view that at the root of most WE inequalities is the weighted Gibbs inequality; see [[18], Theorem 1.3].

In recent years weighted entropy found applications in a wide variety of areas: medical statistics [9,17], investments [11], sensing for motion planning [13,14], data-compression [19], samples-selection in videos [1], data availability and file search efficiency in certain networks [21], slimming neural networks [8], rockburst predictions [23], multi-goal reinforcement learning [22], measurement of complexity in production [7], etc. A different, more theoretical, direction of study of the weight function and weighted entropy has been pursued in recent works [10,15,19,20]. In [19], for instance, the weight function (WF) is used to propose a further reduction in the storage space after data-compression (a refinement of Shannon's Noiseless Coding theorem). Results of [20] clarify various aspects of the weighted information (WI) and WE rates, establishing connections with important notions such as metric and topological entropy, pressure from the theory of dynamical systems.

The WE of a random element X taking values in a standard measure space (SMS) $(\mathcal{X},\mathfrak{M},\nu)$ with a WF: $x\in\mathcal{X}\mapsto\varphi(x)\geq 0$ is defined by

$$h^{\rm w}_{\varphi}(X)=h^{\rm w}_{\varphi}(f):=-\,\mathbb{E}\left[\varphi(X)\log\, f(X)\right]=-\int\limits_{\mathcal{X}}\varphi(x)f(x)\log\, f(x)\nu({\rm d}x), \quad (1.1)$$

assuming that φ is measurable and the integral in the right-hand side of (1.1) is absolutely convergent. Here $f=f(x)$ is the probability mass/density function (PM/DF) of X relative to measure ν. The symbol $\mathbb{E}$ stands for the expected value (relative to a probability distribution that is explicitly specified or emerges from the context in an unambiguous manner). This definition is immediately extended to a pair of random elements X_1, X_2 with values in $(\mathcal{X}_1,\mathfrak{M}_1,\nu_1)$ and $(\mathcal{X}_2,\mathfrak{M}_2,\nu_2)$ with a joint PM/DF $f=f_{X_1,X_2}$ and marginals $f_j=f_{X_j}$. Writing $f_{X_1,X_2}(x_1,x_2)=f_2(x_2)f_{1|2}(x_1|x_2)$, $x_j\in\mathcal{X}_j$, leads to the definition of the conditional WE

$$\begin{aligned}h^{\rm w}_{\varphi}(X_1|X_2)&=-\mathbb{E}\left[\varphi(X_1,X_2)\log\, f_{1|2}(X_1|X_2)\right]=h^{\rm w}_{\varphi}(X_1,X_2)-h^{\rm w}_{\psi_2}(X_2)\\&=-\int_{\mathcal{X}_1\times\mathcal{X}_2}\varphi(x_1,x_2)f(x_1,x_2)\log\, f_{1|2}(x_1|x_2)\nu_1({\rm d}x_1)\nu_2({\rm d}x_2),\end{aligned} \quad (1.2)$$

where ψ_2 is a reduced WF: $\psi_2(x_2)=\int_{\mathcal{X}_1}\varphi(x_1,x_2)f_{1|2}(x_1|x_2)\nu_1({\rm d}x_1)$. Further, the mutual weighted information (WI) between X_1 and X_2 is given by

$$\begin{aligned}i^{\rm w}_{\varphi}(X_1;X_2)&:=\mathbb{E}\left[\varphi(X_1,X_2)\log\,\frac{f(X_1,X_2)}{f_1(X_1)f_2(X_2)}\right]\\&=\int_{\mathcal{X}_1\times\mathcal{X}_2}\varphi(x_1,x_2)f(x_1,x_2)\log\frac{f(x_1,x_2)}{f_1(x_1)f_2(x_2)}\nu_1({\rm d}x_1)\nu_2({\rm d}x_2).\end{aligned} \quad (1.3)$$

For $\varphi(x_1, x_2) \equiv 1$, the above concepts are reduced to the corresponding standard ones.

In particular, the WE of a d-variate Gaussian random vector $\mathbf{X}$ with PDF $f^{\rm No}_{\mathbf{C}}$, with mean zero and covariance matrix $\mathbf{C}$, has the form

$$\begin{aligned} &h^{\rm w}_{\varphi}(\mathbf{X}) = h^{\rm w}_{\varphi}(f^{\rm No}_{\mathbf{C}}) = \frac{\alpha(\mathbf{C})}{2}\log\left((2\pi)^d\det\mathbf{C}\right) + \frac{\log \mathrm{e}}{2}\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{\Phi}\right) \\ &\text{where}\alpha(\mathbf{C}) = \mathbb{E}\varphi(\mathbf{X})\text{and}\mathbf{\Phi} = \mathbb{E}\left[\varphi(\mathbf{X})\mathbf{X}\mathbf{X}^{\rm T}\right]; \end{aligned} \tag{1.4}$$

for $\varphi \equiv 1$, we get the 'standard' Gaussian differential entropy

$$h(f^{\rm No}_{\mathbf{C}}) = \frac{1}{2}\log\left((2\pi\mathrm{e})^d\det\mathbf{C}\right).$$

Accordingly, we speak here of WE inequalities, in particular, weighted determinant inequalities (WDIs). The WDIs offered in the present paper are novel, at least to the best of our knowledge. In fact, the essence of this work is that we subsequently examined DIs from [3,6] for a possibility of a (direct) extension to non-constant weight functions; successful attempts are presented in the current paper. This reflects a particular feature of the present work: a host of new inequalities are obtained by an old method while [3,6] re-establish old inequalities by using a new method. An example is the so-called weighted Fano, Ky Fan, Hadamard, Cramér-Rao and Kullback inequalities (see [[16], Theorems 2.5, 3.5, 3.7, 5.1, 5.4, 5.3]); in this paper we continue this line of study by offering additional WDIs, although their list here is by no means complete.

To illustrate the situation, we mention here the weighted Ky Fan inequality (see [16, Theorem 3.5]). The original inequality asserts that $\sigma:\ \mathbf{C} \mapsto \log\det\mathbf{C}$ is a concave function on the set of $d \times d$ positive-definite matrices; it is proved in [3,6] by using an elegant and short argument based on properties of the standard entropy. Formally: $\forall\ \lambda_1, \lambda_2 \geq 0$ with $\lambda_1 + \lambda_2 = 1$,

$$\begin{aligned} &\sigma(\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2) - \lambda_1\sigma(\mathbf{C}_1) - \lambda_2\sigma(\mathbf{C}_2) \geq 0, \\ &\qquad \text{or, equivalently}, h(f^{\rm No}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) - \lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) - \lambda_2 h(f^{\rm No}_{\mathbf{C}_2}) \geq 0. \end{aligned} \tag{1.5}$$

The weighted Ky Fan inequality reads

$$h^{\rm w}_{\varphi}(f^{\rm No}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) - \lambda_1 h^{\rm w}_{\varphi}(f^{\rm No}_{\mathbf{C}_1}) - \lambda_2 h^{\rm w}_{\varphi}(f^{\rm No}_{\mathbf{C}_2}) \geq 0. \tag{1.6}$$

It involves an additional ingredient, a WF φ, and it holds provided that φ satisfies certain conditions involving λ_j and $\mathbf{C}_j$. For $\varphi \equiv 1$ these conditions are automatically satisfied. For a judicial choice of φ, the weighted inequality may lead to a positive right-hand side in (1.5), i.e., to an improvement of the original one. A similar picture emerges for other DIs. However, this direction needs further studies, including numerical simulations, which is beyond the scope of the current paper.

The present text is organized as follows. In Sect. 2 we work with a general setting, elaborating on properties of WEs which have been established earlier in [18]. Next, Sect. 3 summarizes some properties of Gaussian weighted entropies while Sect. 4 analyzes the behavior of weighted entropies under mappings; these sections also rely on Ref. [18]. The WDIs are presented in Sects. 5 and 6 as a sequel to the material from Sects. 2–4.

2. Random strings and reduced weight functions

A number of properties of the WE are related to a Cartesian product structure. Let random elements $X_1, \ldots, X_n$ be given, taking values in SMSs $(\mathcal{X}_i, \mathfrak{M}_i, \nu_i)$, $1 \le i \le n$. Set $\underline{\mathbf{X}}_1^n = \{X_1, \ldots, X_n\}$ and assume that $X_1, \ldots, X_n$ have a joint PM/DF $f_{\underline{\mathbf{X}}_1^n}(\underline{\mathbf{x}}_1^n)$, $\underline{\mathbf{x}}_1^n \in \mathcal{X}_1^n = \underset{1\le i\le n}{\times} \mathcal{X}_i$, relative to the measure $\nu_1^n = \underset{1\le i\le n}{\times} \nu_i$; for brevity we will sometimes set $f_{\underline{\mathbf{X}}_1^n} = f$. The joint WE of string $\underline{\mathbf{X}}_1^n$ is defined as

$$h^{\rm w}_\varphi(\underline{\mathbf{X}}_1^n) = -\mathbb{E}\left[\varphi(\underline{\mathbf{X}}_1^n)\log\, f(\underline{\mathbf{X}}_1^n)\right]$$
$$= -\int\limits_{\mathcal{X}_1^n} \varphi(\underline{\mathbf{x}}_1^n) f(\underline{\mathbf{x}}_1^n)\log\, f(\underline{\mathbf{x}}_1^n)\nu_1^n({\rm d}\underline{\mathbf{x}}_1^n). \tag{2.1}$$

Given a set $S \subseteq I = \{1, 2, \ldots, n\}$, define: $S^{\complement} = I \setminus S$, and

$$\underline{\mathbf{X}}(S) = \{X_i : i \in S\},\ \underline{\mathbf{X}}(S^{\complement}) = \{X_i : i \in S^{\complement}\},\ \underline{\mathbf{x}}(S)$$
$$= \{x_i : i \in S\},\ \underline{\mathbf{x}}(S^{\complement}) = \{x_i : i \in S^{\complement}\},$$
$$\mathcal{X}(S) = \times_{i\in S}\mathcal{X}_i,\ \ \mathcal{X}(S^{\complement}) = \times_{i\in S^{\complement}}\mathcal{X}_i.$$

Accordingly, the marginal PD/MF $f_{\underline{\mathbf{X}}(S)}(\underline{\mathbf{x}}(S))$ emerges, for which we will often write $f_S(\underline{\mathbf{x}}(S))$ or even $f(\underline{\mathbf{x}}(S))$ for short. Furthermore, given a WF: $\underline{\mathbf{x}}_1^n \mapsto \varphi(\underline{\mathbf{x}}_1^n) \ge 0$, we define the function $\psi(S) : \underline{\mathbf{x}}(S) \mapsto \psi(S; \underline{\mathbf{x}}(S)) \ge 0$ involving the conditional PM/DF $f_{\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S)}\Big(\underline{\mathbf{x}}(S^{\complement})\,|\,\underline{\mathbf{x}}(S)\Big)$:

$$\psi(S; \underline{\mathbf{x}}(S)) = \int\limits_{\mathcal{X}(S^{\complement})} \varphi(\underline{\mathbf{x}}_1^n) f_{\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S)}\Big(\underline{\mathbf{x}}(S^{\complement})\,|\,\underline{\mathbf{x}}(S)\Big)\nu_{\mathcal{X}(S^{\complement})}({\rm d}\underline{\mathbf{x}}(S^{\complement})) \tag{2.2}$$

where $\nu_{\mathcal{X}(S^{\complement})} = \underset{i\in S^{\complement}}{\times} \nu_i$. For brevity sometimes we again write $f_{S^{\complement}|S}$ instead of $f_{\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S)}$ or omit subscripts altogether. We also write ${\rm d}\underline{\mathbf{x}}(S)$ and ${\rm d}\underline{\mathbf{x}}(S^{\complement})$ instead of $\nu_{\mathcal{X}(S)}({\rm d}\underline{\mathbf{x}}(S))$ and $\nu_{\mathcal{X}(S)}({\rm d}\underline{\mathbf{x}}(S^{\complement}))$ and ${\rm d}\underline{\mathbf{x}}$ instead of $\nu_1^n({\rm d}\underline{\mathbf{x}}_1^n)$.

The function $\psi(S;\,\cdot\,)$ will play the role of a reduced (or induced) WF when we pass from $\underline{\mathbf{X}}_1^n$ to a sub-string $\underline{\mathbf{X}}(S)$. More precisely, set

$$h^{\rm w}_{\psi(S)}(\underline{\mathbf{X}}(S)) = -\mathbb{E}\left[\psi(S; \underline{\mathbf{X}}(S)\log f_S(\underline{\mathbf{X}}(S))\right]$$
$$= -\int\limits_{\mathcal{X}(S)} \psi(S; \underline{\mathbf{x}}(S)) f_S(\underline{\mathbf{x}}(S))\log\, f_S(\underline{\mathbf{x}}(S)){\rm d}\underline{\mathbf{x}}(S). \tag{2.3}$$

Next, for $k = 1, \dots, n$ define

$$h_k^{\mathrm{w},n} = \binom{n}{k}^{-1} \sum_{S \subseteq I:\ \#(S)=k} \frac{h^{\mathrm{w}}_{\psi(S)}(\underline{\mathbf{X}}(S))}{k}. \tag{2.4}$$

(Here and below, $\#(S)$ and $\#(S^{\complement})$ are the cardinalities of S and $S^{\complement}$.) Here $h_k^{\mathrm{w},n}$ renders the averaged WE (per string and per element) of a randomly drawn k-element sub-string in $\underline{\mathbf{X}}_1^n$.

Consider the following condition: for all $S \subseteq I$,

$$\begin{aligned} &\forall i \in S, \text{with} S_i^- = \{j \in S :\ j < i\} and S_i^+ = \{j \in S :\ j > i\}, \\ &\int\limits_{\mathcal{X}(S)} \psi(S; \underline{\mathbf{x}}(S)) \Big\{ f(\underline{\mathbf{x}}(S)) - f(\underline{\mathbf{x}}(S_i^-)) f(x_i | \underline{\mathbf{x}}(S_i^-)) \\ &\qquad\qquad \times f(\underline{\mathbf{x}}(S_i^+) | \underline{\mathbf{x}}(S_i^-)) \Big\} \mathrm{d}\underline{\mathbf{x}}(S) \geq 0, \end{aligned} \tag{2.5}$$

with standard agreements when one of the sets $S_i^{\pm} = \emptyset$. Pictorially, Eq. (2.5) is an extension of inequality (1.27) from [18]; it means that for all $i \in S \subseteq I$, the induced WF $\psi(S; \cdot)$ is correlated more positively with the marginal PM/DF $f_S(\underline{\mathbf{x}}(S))$ than with the dependence-broken product $f_{S_i^-}(\underline{\mathbf{x}}(S_i^-)) f_{\{i\}|S_i^-}$ $(x_i | \underline{\mathbf{x}}(S_i^-)) f_{S_i^+|S_i^-}(\underline{\mathbf{x}}(S_i^+) | \underline{\mathbf{x}}(S_i^-))$. Another version of (essentially) the same property is Eq. (2.18) below.

Remark 2.1. The special choice of sets $S_i^{\pm}$ is not particularly important: it can be a general partition of $S \setminus \{i\}$ allowing us to use the chain rule for the conditional WE (see below).

Theorem 2.2. (Cf. [[3], Lemma 7] or [[6], Theorem 1]) *Let $h_k^{\mathrm{w},n}$ be defined as in (2.4) and assume* (2.5). *Then*

$$h_1^{\mathrm{w},n} \geq h_2^{\mathrm{w},n} \geq \cdots \geq h_{n-1}^{\mathrm{w},n} \geq h_n^{\mathrm{w},n}. \tag{2.6}$$

Remark 2.3. An overview of the proof of Theorem 2.3 below shows that its structure is the same as that of the quoted assertions from [3] and [6]. The noted differences are that (i) one uses the weighted Gibbs inequality in place of the standard one, and (ii) at every technical step, one has to re-calculate the involved WFs. Such a pattern persists in the proofs of the remaining statements in this section (Theorems 2.4–2.7 below). Therefore, we only highlight the notable differences in these proofs or entirely omit them from the paper.

Proof. Begin with the last inequality, $h_{n-1}^{\mathrm{w},n} \geq h_n^{\mathrm{w},n}$. Let $1 \leq i \leq n$ and choose $S = I$, $S_i^- = I_i^- = \{1, \dots, i-1\}$ and $S_i^+ = I_i^+ = \{i+1, \dots, n\}$, with $\{i\}^{\complement} = I_i^- \cup I_i^+$. Then the condition

$$\int\limits_{\mathcal{X}_1^n} \varphi(\underline{\mathbf{x}}) \left[f(\underline{\mathbf{x}}) - f_{\underline{\mathbf{X}}_1^{i-1}}(\underline{\mathbf{x}}_1^{i-1}) f(x_i | \underline{\mathbf{x}}_1^{i-1}) f(\underline{\mathbf{x}}_{i+1}^n | \underline{\mathbf{x}}_1^{i-1}) \right] \mathrm{d}\underline{\mathbf{x}} \geq 0 \ \ \text{(by virtue of (2.5))}, \tag{2.7}$$

yields:

$$h^{\rm w}_{\varphi}(\underline{\mathbf{X}}_1^n) = h^{\rm w}_{\varphi}(X_i|\underline{\mathbf{X}}(\{i\}^{\complement})) + h^{\rm w}_{\psi(\{i\}^{\complement})}(\underline{\mathbf{X}}(\{i\}^{\complement})) \qquad \text{by the chain rule} \tag{2.8}$$

$$\leq h^{\rm w}_{\psi(I_i^-)}(X_i|\underline{\mathbf{X}}_1^{i-1}) + h^{\rm w}_{\psi(\{i\}^{\complement})}(\underline{\mathbf{X}}(\{i\}^{\complement})) \qquad \text{by Theorem 1.3 from [18].} \tag{2.9}$$

Here reduced WFs $\psi(\{i\}^{\complement})$ and $\psi(I_i^-)$ are calculated according to the recipes in (2.2), (2.3).

Taking the sum, we obtain:

$$n\, h^{\rm w}_{\varphi}(\underline{\mathbf{X}}_1^n) \leq \sum_{i=1}^n h^{\rm w}_{\psi(\{i\}^{\complement})}(\underline{\mathbf{X}}(\{i\}^{\complement})) + \sum_{i=1}^n h^{\rm w}_{\psi(I_i^-)}(X_i|\underline{\mathbf{X}}_1^{i-1}). \tag{2.10}$$

By using the chain rule, $\sum\limits_{i=1}^n h^{\rm w}_{\psi(I_i^-)}(X_i|\underline{\mathbf{X}}_1^{i-1}) = h^{\rm w}_{\varphi}(\underline{\mathbf{X}}_1^n)$. The proof can now be completed as in the proof of [[6], Theorem 1] and then extended to k-element subsets $S = \{i_1, \ldots, i_k\} \subset I$. (We again need to invoke condition (2.5).) ☐

In Theorem 2.4 we extend the result of Theorem 2.2 to exponents of WEs for sub-strings in $\underline{\mathbf{X}}_1^n$.

Theorem 2.4. (Cf. [3, Corollary of Lemma 7] or [6, Corollary 1]) *Given $r > 0$, define:*

$$g_k^{{\rm w},n} = \binom{n}{k}^{-1} \sum_{S\subseteq I:\ \#(S)=k} \exp\left[r\, \frac{h^{\rm w}_{\psi(S)}(\underline{\mathbf{X}}(S))}{k}\right]. \tag{2.11}$$

Then, under assumption (2.5),

$$g_1^{{\rm w},n} \geq g_2^{{\rm w},n} \geq \cdots \geq g_{n-1}^{{\rm w},n} \geq g_n^{{\rm w},n}. \tag{2.12}$$

In Theorem 2.5 we analyse the averaged conditional WEs for sub-strings in $\underline{\mathbf{X}}_1^n$.

Theorem 2.5. (Cf. [6, Theorem 2]) *Let $p_k^{{\rm w},n}$ be defined as*

$$p_k^{{\rm w},n} = \binom{n}{k}^{-1} \sum_{S\subseteq I:\ \#(S)=k} \frac{h^{\rm w}_{\varphi}(\underline{\mathbf{X}}(S)|\underline{\mathbf{X}}(S^{\complement}))}{k}. \tag{2.13}$$

Then, under assumption (2.5), we have that

$$p_1^{{\rm w},n} \leq p_2^{{\rm w},n} \leq \cdots \leq p_{n-1}^{{\rm w},n} \leq p_n^{{\rm w},n}. \tag{2.14}$$

The next step is to pass to the (non-normalised) mutual WI.

Theorem 2.6. (Cf. [6, Corollary 2]) *Consider the averaged mutual WI between a subset (or a sub-string) and its complement:*

$$q_k^{\mathrm{w},n} = \binom{n}{k}^{-1} \sum_{S\subseteq I:\, \#(S)=k} \frac{i_\varphi^{\mathrm{w}}\Big(\underline{\mathbf{X}}(S);\underline{\mathbf{X}}(S^{\complement})\Big)}{k}, \tag{2.15}$$

and assume (2.5)*. Then*

$$q_1^{\mathrm{w},n} \geq q_2^{\mathrm{w},n} \geq \cdots \geq q_{n-1}^{\mathrm{w},n} \geq q_n^{\mathrm{w},n}. \tag{2.16}$$

Proof. The result is straightforward, from Theorems 2.2 and 2.5 and the following relation between the conditional WE and mutual WI:

$$i_\varphi^{\mathrm{w}}\Big(\underline{\mathbf{X}}(S);\underline{\mathbf{X}}(S^{\complement})\Big) = h_{\psi(S)}^{\mathrm{w}}(\underline{\mathbf{X}}(S)) - h_\varphi^{\mathrm{w}}\Big(\underline{\mathbf{X}}(S)\,\big|\,\underline{\mathbf{X}}(S^{\complement})\Big). \tag{2.17}$$

□

In Theorem 2.7 we consider the following condition: for all set S with $\# S \geq 2$ and $i, j \in S$ with $i \neq j$,

$$\int\limits_{\mathcal{X}_1^n} \varphi(\underline{\mathbf{x}}) f(\underline{\mathbf{x}}(S^{\complement})|\underline{\mathbf{x}}(S))\Big[f(\underline{\mathbf{x}}(S)) \\ - f(\underline{\mathbf{x}}(S\setminus\{i,j\}))\; f(x_i|\underline{\mathbf{x}}(S\setminus\{i,j\}))\; f(x_j|\underline{\mathbf{x}}(S\setminus\{i,j\}))\Big]\mathrm{d}\underline{\mathbf{x}} \geq 0. \tag{2.18}$$

The meaning of (2.18) is that for all S and i, j as above, the reduced WF $\psi_S(\underline{\mathbf{x}}(S))$ is correlated more positively with $f(\underline{\mathbf{x}}(S))$ than with the PM/DF $f(\underline{\mathbf{x}}(S\setminus\{i,j\}))f(x_i|\underline{\mathbf{x}}(S\setminus\{i,j\}))\; f(x_j|\underline{\mathbf{x}}(S\setminus\{i,j\}))$ where the conditional dependence between X_i and X_j is broken, given $\underline{\mathbf{X}}(S\setminus\{i,j\})$.

Theorem 2.7. (Cf. [6, Theorem 3]) *Define the average mutual WI as*

$$I_k^{\mathrm{w},n} = \binom{n}{k}^{-1} \sum_{S\subseteq I:\, \#(S)=k} i_\varphi^{\mathrm{w}}\Big(\underline{\mathbf{X}}(S);\underline{\mathbf{X}}(S^{\complement})\Big). \tag{2.19}$$

By the symmetry of the mutual WI, $I_k^{\mathrm{w},n} = I_{n-k}^{\mathrm{w},n}$*. Assume condition* (2.18)*. Then*

$$I_1^{\mathrm{w},n} \leq I_2^{\mathrm{w},n} \leq \cdots \leq I_{\lfloor n/2\rfloor}^{\mathrm{w},n}. \tag{2.20}$$

Proof. The proof follows that in [6]. Here a new concept is needed: mutual-conditional WIs

$$
\begin{aligned}
&i^{\mathrm{w}}_{\varphi}\Big[X_j;\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S\setminus\{j\})\Big]\\
&=\mathbb{E}\Big(\varphi(\underline{\mathbf{X}})\log\frac{f(X_j,\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S\setminus\{j\}))}{f(X_j|\underline{\mathbf{X}}(S\setminus\{j\}))f(\underline{\mathbf{X}}(S^{\complement})|\underline{\mathbf{X}}(S\setminus\{j\}))}\Big)\\
&\qquad=\int\limits_{\mathcal{X}_1^n}\varphi(\underline{\mathbf{x}})f(\underline{\mathbf{x}})\log\frac{f(x_j,\underline{\mathbf{x}}(S^{\complement})|\underline{\mathbf{x}}(S\setminus\{j\}))}{f(x_j|\underline{\mathbf{x}}(S\setminus\{j\}))f(\underline{\mathbf{x}}(S^{\complement})|\underline{\mathbf{x}}(S\setminus\{j\}))}\,\mathrm{d}\underline{\mathbf{x}},
\end{aligned}
\tag{2.21}
$$

$$
\begin{aligned}
i^{\mathrm{w}}_{\varphi}\Big[X_j;\underline{\mathbf{X}}(S\setminus\{j\})|\underline{\mathbf{X}}(S^{\complement})\Big]&=\mathbb{E}\Big(\varphi(\underline{\mathbf{X}})\log\,\frac{f(X_j,\underline{\mathbf{X}}(S\setminus\{j\})|\underline{\mathbf{X}}(S^{\complement}))}{f(X_j|\underline{\mathbf{X}}(S^{\complement}))f(\underline{\mathbf{X}}(S\setminus\{j\})|\underline{\mathbf{X}}(S^{\complement}))}\Big)\\
=\int\limits_{\mathcal{X}_1^n}\varphi(\underline{\mathbf{x}})f(\underline{\mathbf{x}})\log\,&\frac{f(x_j,\underline{\mathbf{x}}(S\setminus\{j\})|\underline{\mathbf{x}}(S^{\complement}))}{f(x_j|\underline{\mathbf{x}}(S^{\complement}))f(\underline{\mathbf{x}}(S\setminus\{j\})|\underline{\mathbf{x}}(S^{\complement}))}\,\mathrm{d}\underline{\mathbf{x}}.
\end{aligned}
\tag{2.22}
$$

With this at hand, the formal argument from [6] is repeated. □

3. Gaussian weighted entropies

As we said in the introduction, WDIs are connected with the Gaussian WE

$$
h^{\mathrm{w}}_{\varphi}(f^{\mathrm{No}}_{\mathbf{C}})=-\int\limits_{\mathbb{R}^d}\varphi(\mathbf{x})f^{\mathrm{No}}_{\mathbf{C}}(\mathbf{x})\log\,f^{\mathrm{No}}_{\mathbf{C}}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{3.1}
$$

Here $f^{\mathrm{No}}_{\mathbf{C}}$ stands for a normal probability density function (PDF) with mean $\mathbf{0}$ and covariance matrix $\mathbf{C}$:

$$
f^{\mathrm{No}}_{\mathbf{C}}(\mathbf{x})=\frac{1}{(2\pi)^{d/2}\big(\det\mathbf{C}\big)^{1/2}}\exp\left(-\frac{1}{2}\,\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}\right),\quad \mathbf{x}=(x_1,\dots,x_n)^{\mathrm{T}}\in\mathbb{R}^d. \tag{3.2}
$$

The Gaussian WE in (3.1) admits the following representation:

$$
h^{\mathrm{w}}_{\varphi}(f^{\mathrm{No}}_{\mathbf{C}})=\frac{\alpha_{\varphi}(\mathbf{C})}{2}\log\big[(2\pi)^d(\det\mathbf{C})\big]+\frac{\log\,\mathrm{e}}{2}\mathrm{tr}\,\left(\mathbf{C}^{-1}\mathbf{\Phi}^{\mathrm{No}}_{\mathbf{C}}\right). \tag{3.3}
$$

Here $\alpha_{\varphi}(\mathbf{C})>0$ and the positive-definite matrix $\mathbf{\Phi}^{\mathrm{No}}_{\mathbf{C}}$ are given by

$$
\alpha_{\varphi}(\mathbf{C})=\int\limits_{\mathbb{R}^d}\varphi(\mathbf{x}_1^d)f^{\mathrm{No}}_{\mathbf{C}}(\mathbf{x}_1^d)\mathrm{d}\mathbf{x}_1^d,\;\;\mathbf{\Phi}^{\mathrm{No}}_{\mathbf{C}}=\int\limits_{\mathbb{R}^d}\mathbf{x}_1^d\left(\mathbf{x}_1^d\right)^{\mathrm{T}}\varphi(\mathbf{x}_1^d)f^{\mathrm{No}}_{\mathbf{C}}(\mathbf{x}_1^d)\mathrm{d}\mathbf{x}_1^d. \tag{3.4}
$$

Throughout the paper we use a property of maximization of the WE $h^{\rm w}_\varphi(f)$ at $f = f^{\rm No}_{\mathbf{C}}$. Given a PDF f on $\mathbb{R}^d$, set:

$$\mathbf{\Phi} = \int\limits_{\mathbb{R}^d} \mathbf{x}_1^d \left(\mathbf{x}_1^d\right)^{\rm T} \varphi(\mathbf{x}_1^d) f(\mathbf{x}_1^d) {\rm d}\mathbf{x}_1^d. \tag{3.5}$$

Consider the following inequalities:

$$\begin{aligned} &\int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) \Big[f(\mathbf{x}) - f^{\rm No}_{\mathbf{C}}(\mathbf{x})\Big] {\rm d}\mathbf{x} \geq 0 \\ &\log\left[(2\pi)^d (\det \mathbf{C})\right] \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) \Big[f(\mathbf{x}) - f^{\rm No}_{\mathbf{C}}(\mathbf{x})\Big] {\rm d}\mathbf{x} + {\rm tr}\left[\mathbf{C}^{-1}\left(\mathbf{\Phi} - \mathbf{\Phi}^{\rm No}_{\mathbf{C}}\right)\right] \leq 0. \end{aligned} \tag{3.6}$$

Theorem 3.1. (Cf. [6, Example 3.2]) *Let* $\mathbf{X} = \mathbf{X}_1^d \sim f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, *be a random vector with PDF* f, *mean zero and covariance matrix*

$$\mathbf{C} = \mathbb{E}\left[\left(\mathbf{X}_1^d\right)\left(\mathbf{X}_1^d\right)^{\rm T}\right] = \int\limits_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^{\rm T} f(\mathbf{x}) {\rm d}\mathbf{x}. \tag{3.7}$$

Set:

$$\begin{aligned} \mathbf{\Phi}^{\rm No}_{\mathbf{C}} &= \mathbb{E}_{\mathbf{C}}\left[\left(\mathbf{X}_1^d\right)\left(\mathbf{X}_1^d\right)^{\rm T} \varphi(\mathbf{X}_1^d)\right] \\ &= \int\limits_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^{\rm T} \varphi(\mathbf{x}) f^{\rm No}_{\mathbf{C}}(\mathbf{x}) {\rm d}\mathbf{x} \end{aligned} \tag{3.8}$$

and suppose that (3.6) *is fulfilled. Then*

$$h^{\rm w}_\varphi(f) \leq h^{\rm w}_\varphi(f^{\rm No}_{\mathbf{C}}), \tag{3.9}$$

with equality iff $\varphi(f - f^{\rm No}_{\mathbf{C}}) = 0$ *a.s.*

The proof of Theorem 3.1 follows the argument in Example 3.2 from [18].

A conditional form of Theorem 3.1 is Theorem 3.2 below. The corresponding assertion for the standard entropy was noted in earlier literature. See, e.g., [3, p. 1516]: the proof of Theorem 29, item (c), the reference to a conditional version of [3, Lemma 5].

Given a $d \times d$ positive-definite matrix $\mathbf{C}$ and $p = 1, \ldots, d-1$, write $\mathbf{C}$ in the block form:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_1^p & \mathbf{C}_{d-p}^p \\ \mathbf{C}_p^{d-p} & \mathbf{C}_{p+1}^d \end{pmatrix} \tag{3.10}$$

where $\mathbf{C}_{d-p}^p$ and $\mathbf{C}_p^{d-p}$ are mutually transposed $p \times (d-p)$ and $(d-p) \times p$ matrices: $\left(\mathbf{C}_{d-p}^p\right)^{\rm T} = \mathbf{C}_p^{d-p}$. Set: $\mathbf{D} = \mathbf{C}_{d-p}^p \left(\mathbf{C}_{p+1}^d\right)^{-1}$ and $\mathbf{K}_1^p = \mathbf{C}_1^p - \mathbf{C}_{d-p}^p \left(\mathbf{C}_{p+1}^d\right)^{-1} \mathbf{C}_p^{d-p}$. If $\mathbf{X} = \mathbf{X}_1^d$ is a random vector (RV) with PDF $f_{\mathbf{X}}$ and covariance matrix $\mathbf{C}$ then $\mathbf{C}_1^p$ represents the covariance matrix for vector $\mathbf{X}_1^p$, with PDF $f_{\mathbf{X}_1^p}(\mathbf{x}_1^p)$. Let $\mathbf{X}_{p+1}^d$ stand for the residual/remaining random vector and set $f_{\mathbf{X}_1^p|\mathbf{X}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d) = \dfrac{f_{\mathbf{X}}(\mathbf{x}_1^d)}{f_{\mathbf{X}_1^p}(\mathbf{x}_{p+1}^d)}$. Also denote by $\mathbf{N}$, $\mathbf{N}_1^p$ and $\mathbf{N}_{p+1}^d$ the corresponding Gaussian vectors, with PDFs $f_{\mathbf{N}}(\mathbf{x}) = f_{\mathbf{C}}^{\mathrm{No}}(\mathbf{x})$, $f_{\mathbf{N}_1^p}(\mathbf{x}_1^p) = f_{\mathbf{C}_1^p}^{\mathrm{No}}(\mathbf{x}_1^p)$ and $f_{\mathbf{N}_1^p|\mathbf{N}_{p+1}^d}^{\mathrm{No}}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d)$. Finally, for a given WF: $\mathbf{x} \in \mathbb{R}^d \mapsto \varphi(\mathbf{x})$ set:

$$
\begin{aligned}
\psi(\mathbf{x}_{p+1}^d) &= \int\limits_{\mathbb{R}^p} \varphi(\mathbf{x}) f_{\mathbf{N}_1^d|\mathbf{N}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d)\mathrm{d}\mathbf{x}_1^p,\\
\alpha(\mathbf{C}_{p+1}^d) &= \int\limits_{\mathbb{R}^{d-p}} \psi(\mathbf{x}_{p+1}^d) f_{\mathbf{N}_{p+1}^d}(\mathbf{x}_{p+1}^d)\mathrm{d}\mathbf{x}_{p+1}^d,\ \alpha(\mathbf{C}) = \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{N}}(\mathbf{x})\mathrm{d}\mathbf{x},\\
\mathbf{\Psi}_{\mathbf{N}_{p+1}^d} &= \int\limits_{\mathbb{R}^{d-p}} \left[\mathbf{x}_{p+1}^d \left(\mathbf{x}_{p+1}^d\right)^{\mathrm{T}}\right] \psi(\mathbf{x}_{p+1}^d) f_{\mathbf{N}_{p+1}^d}(\mathbf{x}_{p+1}^d)\mathrm{d}\mathbf{x}_{p+1}^d,\\
\mathbf{\Phi}_{\mathbf{N}} &= \int\limits_{\mathbb{R}^d} \left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) \varphi(\mathbf{x}) f_{\mathbf{N}_1^d}(\mathbf{x})\mathrm{d}\mathbf{x}.
\end{aligned}
\tag{3.11}
$$

Also, consider inequalities

$$
\begin{aligned}
I_1 &= \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}_{p+1}^d}(\mathbf{x}_{p+1}^d)\Big[f_{\mathbf{X}_1^p|\mathbf{X}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d) - f_{\mathbf{N}_1^p|\mathbf{N}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d)\Big]\mathrm{d}\mathbf{x} \geq 0,\\
I_2 &= \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x})\Big[f_{\mathbf{X}}(\mathbf{x}) - f_{\mathbf{N}}(\mathbf{x})\Big]\Big\{\log\ [(2\pi)^p \det\ (\mathbf{K}_1^p)]\\
&\quad + (\log \mathrm{e})\left[\left(\mathbf{x}_1^p - \mathbf{D}\mathbf{x}_{p+1}^d\right)^{\mathrm{T}} (\mathbf{K}_1^p)^{-1} \left(\mathbf{x}_1^p - \mathbf{D}\mathbf{x}_{p+1}^d\right)\right]\Big\}\mathrm{d}\mathbf{x} \leq 0.
\end{aligned}
\tag{3.12}
$$

Theorem 3.2. *Assume that inequalities (3.12) are satisfied. Then the following inequality holds true:*

$$
\begin{aligned}
h_\varphi^{\mathrm{w}}(\mathbf{X}_1^d|\mathbf{X}_{p+1}^d) &= -\int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \log\ f_{\mathbf{X}_1^p|\mathbf{X}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d)\mathrm{d}\mathbf{x}\\
&\leq h_\varphi^{\mathrm{w}}(\mathbf{N}_1^p|\mathbf{N}_{p+1}^d) = h_\varphi^{\mathrm{w}}(\mathbf{N}) - h_\psi^{\mathrm{w}}(\mathbf{N}_{p+1}^d)\\
&= \frac{\alpha(\mathbf{C})}{2} \log\left[(2\pi)^d \det \mathbf{C}\right] + \frac{\log \mathrm{e}}{2} \mathrm{tr}\ \left[\mathbf{C}^{-1}\mathbf{\Phi}_{\mathbf{N}}\right]\\
&\quad - \frac{\alpha(\mathbf{C}_{p+1}^d)}{2} \log\left[(2\pi)^{d-p} \det \mathbf{C}_{p+1}^d\right]\\
&\quad - \frac{\log \mathrm{e}}{2} \mathrm{tr}\ \left[\left(\mathbf{C}_{p+1}^d\right)^{-1} \mathbf{\Psi}_{\mathbf{N}_{p+1}^d}\right]\ .
\end{aligned}
\tag{3.13}
$$

Proof. Note that the conditional distribution $\mathbf{N}_1^p|\mathbf{N}_{p+1}^d$ coincides with $\mathbf{N}_{\mathbf{K}_1^p}$, a Gaussian variable with mean $\mathbf{D}\mathbf{x}_{p+1}^d$ and covariance matrix $\mathbf{K}_1^p$. Denote the random variable with conditional distribution $\mathbf{X}_1^p|\mathbf{X}_{p+1}^d$ by $\mathbf{X}_{1,p}^{p+1,d}$. Set:

$$\hat{\psi}(\mathbf{X}_{p+1}^d) = \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}_1^p|\mathbf{X}_{p+1}^d}(\mathbf{x}_1^p|\mathbf{x}_{p+1}^d)\mathrm{d}\mathbf{x}_1^p.$$

By the weighted Gibbs inequality, under assumption (3.12),

$$\int\limits_{\mathbb{R}^{d-p}} f_{\mathbf{X}_{p+1}^d}(\mathbf{x}_{p+1}^d) i^{\mathrm{w}}_{\hat{\psi}(\mathbf{X}_{p+1}^d)}(\mathbf{X}_{1,p}^{p+1,d}|\mathbf{N}_{\mathbf{K}_1^p})\mathrm{d}\mathbf{x}_{p+1}^d \geq I_1 \geq 0. \tag{3.14}$$

Hence,

$$\begin{aligned} h^{\mathrm{w}}_{\varphi}(\mathbf{X}_{1,p}^{p+1,d}) \leq \frac{1}{2} \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \Big[&\log[(2\pi)^p \det(\mathbf{K}_1^p)] \\ &+ (\log e)[(\mathbf{x}_1^p - \mathbf{D}\mathbf{x}_{p+1}^d)^{\mathrm{T}} (\mathbf{K}_1^p)^{-1} (\mathbf{x}_1^p - \mathbf{D}\mathbf{x}_{p+1}^d)]\Big]\mathrm{d}\mathbf{x} \end{aligned} \tag{3.15}$$

implying (3.13) in view of the inequality $I_2 \leq 0$ in (3.12). □

4. Weighted entropies under mappings

In this section we give a series of general results (Theorems 4.1–4.3 and Theorem 4.4) reflecting properties of WEs under mappings of random variables (an example is a sum $X+Y$). Of a special importance for us is Theorem 4.3 used in Sect. 5. In essence, Theorems 4.1–4.3 can be interpreted as versions of data-processing inequalities for the WE, and directly generalize their counterparts from [18]. Hence, we omit their proofs.

Let $(\mathcal{X}, \mathfrak{X}, \nu_{\mathcal{X}})$, $(\mathcal{Y}, \mathfrak{Y}, \nu_Y)$ be a pair of SMSs and suppose X, Y are random elements in $(\mathcal{X}, \mathfrak{X})$, $(\mathcal{Y}, \mathfrak{Y})$ and f_X, f_Y are PM/DFs, relative to measures $\nu_{\mathcal{X}}$, $\nu_{\mathcal{Y}}$, respectively. Suppose $\eta : (\mathcal{X}, \mathfrak{X}) \to (\mathcal{Y}, \mathfrak{Y})$ is a measurable map onto, and that $\nu_{\mathcal{Y}}(B) = \nu_{\mathcal{X}}(\eta^{-1}B)$, $B \in \mathfrak{Y}$. Consider the partition of $\mathcal{X}$ with elements $\mathcal{B}(y) = \{x \in \mathcal{X} :\ \eta x = y\}$ and let $\nu_{\mathcal{X}}(\,\cdot\,|y)$ be the family of induced measures on $\mathcal{B}(y)$, $y \in \mathcal{Y}$. Suppose that $f_Y(y) = \displaystyle\int_{\mathcal{B}(y)} f_X(x)\nu(\mathrm{d}x|y)$ and for $x \in \mathcal{B}(y)$ let $f_{X|Y}(x|y) = \dfrac{f_X(x)}{f_Y(y)}$ denote the PM/DF of X conditional on $Y = y$. (Recall, $f_{X|Y}(\,\cdot\,|y)$ is a family of PM/DFs defined for f_Y-a.a $y \in \mathcal{Y}$ such that $\displaystyle\int_{\mathcal{X}} G(x) f_X(x)\nu_{\mathcal{X}}(\mathrm{d}x) = \int_{\mathcal{Y}}\int_{\mathcal{B}(y)} G(x) f_{X|Y}(x|y)\nu_{\mathcal{X}}(\mathrm{d}x|y) f_Y(y)\nu_{\mathcal{Y}}(\mathrm{d}y)$ for any non-negative measurable function G.)

Theorem 4.1. (Cf. [16, Lemma 1.6]) *Suppose that a WF: $x \in \mathcal{X} \mapsto \varphi(x) \geq 0$ obeys*

$$\int\limits_{\mathcal{X}} \varphi(x) f_X(x) \Big[f_{X|Y}(x|\eta x) - 1 \Big] \nu_{\mathcal{X}}(\mathrm{d}x) \leq 0 \tag{4.1}$$

and set

$$\psi(y) = \int\limits_{\mathcal{B}(y)} \varphi(x) f_{X|Y}(x|y) \nu(\mathrm{d}x|y), \;\; y \in \mathcal{Y}. \tag{4.2}$$

Then

$$\begin{aligned} h^{\mathrm{w}}_{\varphi}(X) \geq h^{\mathrm{w}}_{\psi}(Y) &= -\int\limits_{\mathcal{Y}} \psi(y) f_Y(y) \log\, f_Y(y) \nu_Y(\mathrm{d}y), \quad \textit{or} \\ h^{\mathrm{w}}_{\varphi}(X|Y) &= -\int\limits_{\mathcal{X}} \varphi(x) f_X(x) \log\, f_{X|Y}(x|y(x)) \nu_{\mathcal{X}}(\mathrm{d}x) \geq 0, \end{aligned} \tag{4.3}$$

with equality iff $\varphi(x)\big[f_{X|Y}(x|\eta x) - 1\big] = 0$ for f-a.a. $x \in \mathcal{X}$.

In particular, suppose that for f_Y-a.a. $y \in \mathcal{Y}$ a set $\mathcal{B}(y)$ contains at most countably many values and $\nu(\,\cdot\,|y)$ is a counting measure with $\nu_1(x) = 1$, $x \in \mathcal{B}(y)$. Then the value $f_{X|Y}(x|\eta x)$ yields the conditional probability $\mathbb{P}(X = x|Y = \eta x)$, which is ≤ 1 for f_Y-a.a. $y \in \mathcal{Y}$. Then $h^{\mathrm{w}}_{\varphi}(X|Y) \geq 0$ and the inequality is strict unless, modulo φ, map η is one-to-one.

Let $(\mathcal{X}, \mathfrak{X}, \nu_{\mathcal{X}})$, $(\mathcal{Y}, \mathfrak{Y}, \nu_{\mathcal{Y}})$, $(\mathcal{Z}, \mathfrak{Z}, \nu_{\mathcal{Z}})$ be a triple of SMSs and suppose X, Y, Z are random elements in $(\mathcal{X}, \mathfrak{X})$, $(\mathcal{Y}, \mathfrak{Y})$, $(\mathcal{Z}, \mathfrak{Z})$. Let f_X be the PM/DF for X relative to measure $\nu_{\mathcal{X}}$ and $f_{Y,Z}$ the joint PM/DF for Y, Z relative to the measures $\nu_{\mathcal{Y}} \times \nu_{\mathcal{Z}}$. Further, set $f_Z(z) = \displaystyle\int_{\mathcal{Y}} f(y,z) \nu_{\mathcal{Y}}(\mathrm{d}y)$ and $f_{Y|Z}(y|z) = \dfrac{f_{Y,Z}(y,z)}{f_Z(z)}$.

Consider the partition of $\mathcal{X}$ with elements $\mathcal{B}(y,z) = \{x \in \mathcal{X} : \;\; \eta x = y, \zeta x = z\}$, and let $\nu_{\mathcal{X}}(\,\cdot\,|y,z)$ be the family of induced measures on $\mathcal{B}(y,z)$, $(y,z) \in \mathcal{Y} \times \mathcal{Z}$. Set:

$$f_{Y,Z}(y,z) = \int\limits_{\mathcal{B}(y,z)} f_X(x) \nu_{\mathcal{X}}(\mathrm{d}x|y,z) \tag{4.4}$$

and for $x \in \mathcal{B}(y,z)$ let $f_{X|Y,Z}(x|y,z) = \dfrac{f_X(x)}{f_{Y,Z}(y,z)}$ denote the PM/DF of X conditional on $Y = y$, $Z = z$. (Recall, $f_{X|Y,Z}(\,\cdot\,|y,z)$ is a family of PM/DFs defined for $f_{Y,Z}$-a.a $(y,z)\in\mathcal{Y}\times\mathcal{Z}$ such that

$$\begin{aligned}&\int\limits_{\mathcal{X}} G(x)f_X(x)\nu_{\mathcal{X}}(\mathrm{d}x)\\ &\quad=\int\limits_{\mathcal{Y}\times\mathcal{Z}}\int\limits_{\mathcal{B}(y,z)} G(x)f_{X|Y,Z}(x|y,z)\nu_{\mathcal{X}}(\mathrm{d}x|y,z)f_{Y,Z}(y,z))\nu_{\mathcal{Y}}(\mathrm{d}y)\nu_{\mathcal{Z}}(\mathrm{d}z) \quad (4.5)\end{aligned}$$

for any non-negative measurable function G.)

Theorem 4.2. (Cf. [16, Lemma 1.9]) *Suppose that*

$$\eta:(\mathcal{X},\mathfrak{X})\to(\mathcal{Y},\mathfrak{Y}),\ \ \zeta:(\mathcal{X},\mathfrak{X})\to(\mathcal{Z},\mathfrak{Z}) \tag{4.6}$$

is a pair of measurable maps onto, and that

$$\nu_{\mathcal{Y}}(A)=\nu_{\mathcal{X}}(\eta^{-1}A),\ A\in\mathfrak{Y},\ \ \nu_{\mathcal{Z}}(B)=\nu_{\mathcal{X}}(\zeta^{-1}B),\ B\in\mathfrak{Z}. \tag{4.7}$$

Assume that a WF: $x\mapsto\varphi(x)\geq 0$ obeys

$$\int_{\mathcal{X}}\varphi(x)f(x)\Big[f_{X|Y,Z}(x|\eta x,\zeta x)-1\Big]\nu_{\mathcal{X}}(\mathrm{d}x)\leq 0 \tag{4.8}$$

and set

$$\psi(y,z)=\int\limits_{\mathcal{B}(y,z)}\varphi(x)f_{X|Y,Z}(x|y,z)\nu(\mathrm{d}x|y,z). \tag{4.9}$$

Then

$$\begin{aligned}&-\int\limits_{\mathcal{Y}\times\mathcal{Z}}\psi(y,z)f_{Y,Z}(y,z)\log\,f_{Y|Z}(y|z)\nu_{\mathcal{Y}}(\mathrm{d}y)\nu_{\mathcal{Z}}(\mathrm{d}z)\\ &\qquad=h^{\mathrm{w}}_{\psi}(Y|Z)\leq h^{\mathrm{w}}_{\varphi}(X|Z)=-\int\limits_{\mathcal{X}}\varphi(x)f_X(x)\log\,f_{X|Z}(x|\zeta x)\nu(\mathrm{d}x);\end{aligned} \tag{4.10}$$

equality iff $\varphi(x)\big[f_{X|Y,Z}(x|\eta x,\zeta x)-1\big]=0$ for f_X-a.a. $x\in\mathcal{X}$.

As in Theorem 4.1, assume $\mathcal{B}(y,z)$ consists of at most countably many values and $\nu(x|y,z)=1$, $x\in\mathcal{B}(y,z)$ for $f_{Y,Z}$-a.a. $(y,z)\in\mathcal{Y}\times\mathcal{Z}$. Then the value $f_{X|Y,Z}(x|y,z)$ yields the conditional probability $\mathbb{P}(X=x|Y=y,Z=z)$, for $f_{Y,Z}$-a.a. $y,z\in\mathcal{Y}\times\mathcal{Z}$. Then $h^{\mathrm{w}}_{\varphi}(X|Z)\geq h^{\mathrm{w}}_{\psi}(Y|Z)$, with equality iff, modulo φ, the map $x\mapsto(\eta x,\zeta x)$ is one-to-one.

Let $f_{X,Y}$ be the joint PM/DF for X,Y relative to measure $\nu_{\mathcal{X}}\times\nu_{\mathcal{Y}}$ and set

$$f_Y(y)=\int\limits_{\mathcal{X}}f_{X,Y}(x,y)\nu_{\mathcal{X}}(\mathrm{d}x),\ \ f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{4.11}$$

Suppose that

$$\xi:(\mathcal{Y},\mathfrak{Y})\to(\mathcal{Z},\mathfrak{Z}) \tag{4.12}$$

is a measurable map onto, and that

$$\nu_{\mathcal{Z}}(C)=\nu_{\mathcal{Y}}(\xi^{-1}C),\ C\in\mathfrak{Z}. \tag{4.13}$$

Consider a partition of $\mathcal{Y}$ with elements $\mathcal{C}(z)=\{y\in\mathcal{Y}:\xi y=z\}$ and let $\nu_{\mathcal{Y}}(\,\cdot\,|z)$ be the family of induced measures on $\mathcal{C}(z)$, $z\in\mathcal{Z}$. Given $(x,z)\in\mathcal{X}\times\mathcal{Z}$ and $y\in\mathcal{C}(z)$, let

$$f_{X,Z}(x,z)=\int\limits_{\mathcal{C}(z)} f_{X,Y}(x,y)\nu_{\mathcal{Y}}(\mathrm{d}y|z),\ f_Z(z)=\int\limits_{\mathcal{X}} f_{X,Z}(x,z)\nu_{\mathcal{X}}(\mathrm{d}x), \tag{4.14}$$

and

$$f_{X|Z}(x|z)=\frac{f_{X,Z}(x,z)}{f_Z(z)},\ \ f_{Y|Z}(y|z)=\frac{f_Y(y)}{f_Z(z)}. \tag{4.15}$$

Theorem 4.3. (Cf. [16, Lemma 1.11]) *Assume that a WF: $(x,y)\mapsto\varphi(x,y)\geq 0$ obeys*

$$\int\limits_{\mathcal{X}\times\mathcal{Y}}\varphi(x,y)\left[f_{X,Y}(x,y)-f_Z(\xi y)f_{X|Z}(x|\xi y)f_{Y|Z}(y|\xi y)\right]\nu_{\mathcal{X}}(\mathrm{d}x)\nu_{\mathcal{Y}}(\mathrm{d}y)\geq 0 \tag{4.16}$$

and set

$$\psi(x,z)=\int\limits_{\mathcal{C}(z)}\varphi(x,y)f_{Y|Z}(y|z)\nu_{\mathcal{Y}}(\mathrm{d}y|z). \tag{4.17}$$

Then

$$\begin{aligned}&-\int\limits_{\mathcal{X}\times\mathcal{Z}}\psi(x,z)f_{X,Z}(x,z)\log\, f_{X|Z}(y|z)\nu_{\mathcal{X}}(\mathrm{d}x)\nu_{\mathcal{Z}}(\mathrm{d}z)\\ &\qquad=h^{\rm w}_{\psi}(X|Z)\geq h^{\rm w}_{\varphi}(X|Y)=-\int\limits_{\mathcal{X}\times\mathcal{Y}}\varphi(x,y)f_X(x)\log\, f_{X|Y}(x|y)\nu_{\mathcal{X}}(\mathrm{d}x)\nu_{\mathcal{Y}}(\mathrm{d}y).\end{aligned} \tag{4.18}$$

Furthermore, equality in (4.18) holds iff X and Y are conditionally independent given Z modulo φ, i.e. $\varphi(x,y)\left[f_{X,Y}(x,y)-f_Z(\xi y)f_{X|Z}(x|\xi y)f_{Y|Z}(y|\xi y)\right]=0$.

We will use an alternative notation $h^{\rm w}_{\varphi}(\mathbf{X})=h^{\rm w}_{\varphi}(f_{\mathbf{X}})$ where $\mathbf{X}=\mathbf{X}_1^d=(X_1,\ldots,X_d)^T$ is a d-dimensional random vector with PDF $f_{\mathbf{X}}(\mathbf{x})$. (A change in the notation is motivated by the emphasis on linearity properties in $\mathbb{R}^d$ absent in $\mathcal{X}^n$.) In this context, we employ the notation $\mathbf{X}\sim f_{\mathbf{X}}$, $\mathbf{Y}\sim f_{\mathbf{Y}}$, $(\mathbf{X},\mathbf{Y})\sim f_{\mathbf{X},\mathbf{Y}}$ and $(\mathbf{X}|\mathbf{Y})\sim f_{\mathbf{X}|\mathbf{Y}}$ where $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})=\dfrac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}$.

Theorem 4.4 below mimics a result in [3, Lemma 5] extending from the case of a standard entropy to that of the WE. A number of facts are related to the conditional WE

$$h^{\rm w}_{\varphi}(\mathbf{X}|\mathbf{Y}) = -\int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(\mathbf{x},\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \log\, f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}){\rm d}\mathbf{x}{\rm d}\mathbf{y} \tag{4.19}$$

or, more generally,

$$h^{\rm w}_{\widetilde{\varphi}}(\mathbf{U}|\mathbf{V}) = -\int_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{\varphi}(\mathbf{u},\mathbf{v}) f_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v}) \log\, f_{\mathbf{U}|\mathbf{V}}(\mathbf{u}|\mathbf{v}){\rm d}\mathbf{u}{\rm d}\mathbf{v}. \tag{4.20}$$

Here a pair $(\mathbf{U},\mathbf{V})$ is a function of $(\mathbf{X},\mathbf{Y})$ with a joint PM/DF $f_{\mathbf{U},\mathbf{V}}$, marginal PM/DFs $f_{\mathbf{U}}$, $f_{\mathbf{V}}$ and conditional PM/DF $f_{\mathbf{U}|\mathbf{V}}(\mathbf{u}|\mathbf{v}) = \dfrac{f_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v})}{f_{\mathbf{V}}(\mathbf{v})}$. (Viz., $\mathbf{U}=\mathbf{Y}$, $\mathbf{V}=\mathbf{X}+\mathbf{Y}$.) WF $\widetilde{\varphi}$ may or may not be involved with the map $(\mathbf{X},\mathbf{Y})\mapsto(\mathbf{U},\mathbf{V})$.

Theorem 4.4. (Cf. [3, Lemma 5]) *Suppose* $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent random vectors of dimension d, with PDFs* $f_{\mathbf{X}}$ *and* $f_{\mathbf{Y}}$*:*

$$(\mathbf{X},\mathbf{Y})\sim f_{\mathbf{X},\mathbf{Y}}\quad \textit{where}\quad f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y}),\quad \mathbf{x},\mathbf{y}\in\mathbb{R}^d. \tag{4.21}$$

Assume that WF: $(\mathbf{x},\mathbf{y})\in\mathbb{R}^d\times\mathbb{R}^{d'}\to\varphi^*(\mathbf{x},\mathbf{y})=\varphi(\mathbf{x}+\mathbf{y},\mathbf{y})\geq 0$ *obeys*

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} \varphi^*(\mathbf{x},\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})\Big[f_{\mathbf{X}+\mathbf{Y}}(\mathbf{x}+\mathbf{y})-f_{\mathbf{X}}(\mathbf{x})\Big]{\rm d}\mathbf{x}{\rm d}\mathbf{y}\geq 0 \tag{4.22}$$

and set

$$\begin{aligned}\theta(\mathbf{x}) &= \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}+\mathbf{Y}}(\mathbf{y}|\mathbf{x}+\mathbf{y}){\rm d}\mathbf{y}\\ \theta^*(\mathbf{x}) &= \int\limits_{\mathbb{R}^d}\varphi(\mathbf{x}+\mathbf{y},\mathbf{y})f_{\mathbf{Y}}(\mathbf{y}){\rm d}\mathbf{y},\ \ \mathbf{x}\in\mathbb{R}^d.\end{aligned} \tag{4.23}$$

Then

$$h^{\rm w}_{\theta}(\mathbf{X}+\mathbf{Y})\geq h^{\rm w}_{\theta^*}(\mathbf{X}), \tag{4.24}$$

with equality iff $\varphi(\mathbf{x},\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})\Big[f_{\mathbf{X}}(\mathbf{x})-f_{\mathbf{X}+\mathbf{Y}}(\mathbf{x}+\mathbf{y})\Big]=0$ *for Lebesgue-a.a.* $(\mathbf{x},\mathbf{y})\in\mathbb{R}^d\times\mathbb{R}^d$.

Proof. The following relations (a)–(c) hold true:

$$\begin{aligned}&\text{(a)}\quad h^{\rm w}_{\theta}(\mathbf{X}+\mathbf{Y})\geq h^{\rm w}_{\varphi}(\mathbf{X}+\mathbf{Y}|\mathbf{Y}),\\ &\text{(b)}\quad h^{\rm w}_{\varphi}(\mathbf{X}+\mathbf{Y}|\mathbf{Y}) = h^{\rm w}_{\varphi^*}(\mathbf{X}|\mathbf{Y}),\\ &\text{(c)}\quad h^{\rm w}_{\varphi^*}(\mathbf{X}|\mathbf{Y}) = h^{\rm w}_{\theta^*}(\mathbf{X}).\end{aligned} \tag{4.25}$$

Here inequality (a) comes from the sub-additivity of the WE, see [6, Theorem 1.8] or Eq. (1.32) from [18]. Next, (b) and (c) are derived by applying the following equations:

$$\begin{aligned}
h^{\rm w}_\varphi(\mathbf{X}+\mathbf{Y}|\mathbf{Y}) &= \int_{\mathbb{R}^d} f_{\mathbf{Y}}(\mathbf{y}) h^{\rm w}_\varphi(\mathbf{X}+\mathbf{Y}|\mathbf{Y}=\mathbf{y})\mathrm{d}\mathbf{y}\\
&= -\int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{X}+\mathbf{Y}|\mathbf{Y}}(\mathbf{x}+\mathbf{y}|\mathbf{y})\log f_{\mathbf{X}+\mathbf{Y}|\mathbf{Y}}(\mathbf{x}+\mathbf{y}|\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\\
&= -\int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\log f_{\mathbf{X}+\mathbf{Y}|\mathbf{Y}}(\mathbf{x}+\mathbf{y}|\mathbf{y})\mathrm{d}\mathbf{x})\mathrm{d}\mathbf{y}\\
&= -\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y}\right] f_{\mathbf{X}}(\mathbf{x})\log f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x} = h^{\rm w}_{\varphi^*}(\mathbf{X}|\mathbf{Y}) = h^{\rm w}_{\theta^*}(\mathbf{X}).
\end{aligned} \tag{4.26}$$

These equations also allow us to complete the proof of Theorem 4.4. □

5. Miscellaneous weighted determinant inequalities

In this section we present a host of WDIs derived from the properties of WEs. As we said before, the proposed inequalities hold when WF $\varphi \equiv 1$ (in this case the stated conditions are trivially fulfilled). To stress parallels with 'standard' DIs, we provide references to [3] or [6] in each case under consideration.

Theorem 5.1. (Cf. [3, Theorem 2]) *Let* $\mathbf{X}$, $\mathbf{Y}$ *be independent d-variate normal vectors with zero means and covariance matrices* $\mathbf{C}_1$, $\mathbf{C}_2$, *respectively:* $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y})$, $\mathbf{x},\mathbf{y}\in\mathbb{R}^d$, *where* $f_{\mathbf{X}} = f^{\rm No}_{\mathbf{C}_1}$, $f_{\mathbf{Y}} = f^{\rm No}_{\mathbf{C}_2}$. *Given a WF:* $(\mathbf{x},\mathbf{y})\in\mathbb{R}^d\times\mathbb{R}^d \mapsto \varphi(\mathbf{x},\mathbf{y})\geq 0$, *positive on an open domain in* $\mathbb{R}^d\times\mathbb{R}^d$, *consider a quantity* β *and* $d\times d$ *matrices* $\mathbf{\Theta}$, $\mathbf{\Theta}^*$*:*

$$\begin{aligned}
&\beta = \int_{\mathbb{R}^d} \theta(\mathbf{x}) f^{\rm No}_{\mathbf{C}_1+\mathbf{C}_2}(\mathbf{x})\mathrm{d}\mathbf{x},\\
&\mathbf{\Theta} = \int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^{\rm T}\theta(\mathbf{x}) f^{\rm No}_{\mathbf{C}_1+\mathbf{C}_2}(\mathbf{x})\mathrm{d}\mathbf{x},\quad \mathbf{\Theta}^* = \int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^{\rm T}\theta^*(\mathbf{x}) f^{\rm No}_{\mathbf{C}_1}(\mathbf{x})\mathrm{d}\mathbf{x}
\end{aligned} \tag{5.1}$$

where θ *and* θ^* *are as in* (4.23)*:*

$$\theta(\mathbf{x}) = \int_{\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}+\mathbf{Y}}(\mathbf{y}|\mathbf{x}+\mathbf{y})\mathrm{d}\mathbf{y},\quad \theta^*(\mathbf{x}) = \int_{\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y}. \tag{5.2}$$

Assume the condition emulating (4.22)*:*

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(\mathbf{x}+\mathbf{y},\mathbf{y}) f^{\rm No}_{\mathbf{C}_2}(\mathbf{y})\Big[f^{\rm No}_{\mathbf{C}_1+\mathbf{C}_2}(\mathbf{x}+\mathbf{y}) - f^{\rm No}_{\mathbf{C}_1}(\mathbf{x})\Big]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \geq 0. \tag{5.3}$$

Then

$$\beta\log\left[\frac{\det(\mathbf{C}_1+\mathbf{C}_2)}{\det\mathbf{C}_1}\right] + (\log e)\Big\{\mathrm{tr}\Big[(\mathbf{C}_1+\mathbf{C}_2)^{-1}\boldsymbol{\Theta}\Big] - \mathrm{tr}\left(\mathbf{C}_1^{-1}\boldsymbol{\Theta}^*\right)\Big\} \geq 0. \tag{5.4}$$

Proof. Using Theorem 4.4 and Eq. (3.3), we can write:

$$\frac{1}{2}\log\Big[(2\pi)^d(\det(\mathbf{C}_1+\mathbf{C}_2))\Big]\int\limits_{\mathbb{R}^d}\theta(\mathbf{x})f^{\rm No}_{\mathbf{C}_1+\mathbf{C}_2}(\mathbf{x})\mathrm{d}\mathbf{x} + \frac{\log \mathrm{e}}{2}\mathrm{tr}\left[(\mathbf{C}_1+\mathbf{C}_2)^{-1}\boldsymbol{\Theta}\right]$$
$$\geq \frac{1}{2}\log\Big[(2\pi)^d(\det\mathbf{C}_1)\Big]\int\limits_{\mathbb{R}^d}\theta^*(\mathbf{x})f^{\rm No}_{\mathbf{C}_1}(\mathbf{x})\mathrm{d}\mathbf{x} + \frac{\log \mathrm{e}}{2}\mathrm{tr}\left(\mathbf{C}_1^{-1}\boldsymbol{\Theta}^*\right). \tag{5.5}$$

Next, $\int\limits_{\mathbb{R}^d}\theta^*(\mathbf{x})f^{\rm No}_{\mathbf{C}_1}(\mathbf{x})\mathrm{d}\mathbf{x} = \beta$. The inequality in (5.4) then follows. □

Remark 5.2. Note that (5.4) is equivalent to:

$$\beta\log\left[\det(\mathbf{I}+\mathbf{C}_1^{-1}\mathbf{C}_2)\right] + (\log e)\mathrm{tr}\Big[(\mathbf{C}_1+\mathbf{C}_2)^{-1}\boldsymbol{\Theta}^* - \mathbf{C}_1^{-1}\boldsymbol{\Theta}^* + (\mathbf{C}_1+\mathbf{C}_2)^{-1}\widetilde{\boldsymbol{\Theta}}\Big] \geq 0 \tag{5.6}$$

where

$$\widetilde{\boldsymbol{\Theta}} = \int\limits_{\mathbb{R}^d\times\mathbb{R}^d}\left(\mathbf{x}\mathbf{y}^{\rm T}+\mathbf{y}\mathbf{x}^{\rm T}+\mathbf{y}\mathbf{y}^{\rm T}\right)\varphi(\mathbf{x}+\mathbf{y},\mathbf{y})f^{\rm No}_{\mathbf{C}_2}(\mathbf{y})f^{\rm No}_{\mathbf{C}_1}(\mathbf{x})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}. \tag{5.7}$$

This claim is verified by observing that $\boldsymbol{\Theta} = \boldsymbol{\Theta}^* + \widetilde{\boldsymbol{\Theta}}$.

Remark 5.3. As above, we can assume that $\mathbf{C}_2$ is a matrix of size $d'\times d'$, agreeing that in the sum $\mathbf{C}_1+\mathbf{C}_2$, matrix $\mathbf{C}_2$ is identified as a top left block (say). This is possible because in Eqs. (5.4) and (5.6) we do not use the inverse $\mathbf{C}_2^{-1}$ or the determinant $\det\mathbf{C}_2$.

To this end, recall the following theorem from [16]:

Theorem 5.4. *Let* $\mathbf{G}$ *and* $\mathbf{G}+\mathbf{E}$ *be non-singular matrices where* $\mathbf{E}$ *is a matrix of rank one. Let* $g = \mathrm{tr}\left(\mathbf{E}\mathbf{G}^{-1}\right)$. *Then* $g \neq -1$ *and*

$$(\mathbf{G}+\mathbf{E})^{-1} = \mathbf{G}^{-1} - \frac{1}{1+g}\mathbf{G}^{-1}\,\mathbf{E}\,\mathbf{G}^{-1}. \tag{5.8}$$

The above equation is essentially the Sherman-Morrison formula (see [5], p. 161).

Assuming that $\mathbf{C}_2 = \mathbf{E}$ has rank 1 and letting $g = \mathrm{tr}\,(\mathbf{E}\mathbf{C}_1^{-1})$, inequality (5.4) turns into the following inequality:

$$\beta \log\left[\frac{\det(\mathbf{C}_1+\mathbf{E})}{\det \mathbf{C}_1}\right] + (\log e)\left[-\mathrm{tr}\left(\frac{\mathbf{C}_1^{-1}\mathbf{E}\mathbf{C}_1}{1+g}\boldsymbol{\Theta}^*\right) + \mathrm{tr}\left\{(\mathbf{C}_1+\mathbf{E})^{-1}\widetilde{\boldsymbol{\Theta}}\right\}\right] \geq 0. \quad (5.9)$$

The techniques developed so far allow us to prove Theorem 5.5 below rendering a weighted form of the Szasz theorem. Suppose $\mathbf{C}$ is a positive definite $d\times d$ matrix. Given $1 \leq k \leq d$ and a set $S \subseteq I^{(d)} = \{1,\dots,d\}$ with $\#(S) = k$, let $\mathbf{C}(S)$ be the $k\times k$ sub-matrix of $\mathbf{C}$ formed by the rows and columns with indices $i \in S$. With every S we associate a Gaussian random vector $\mathbf{X}(S) \sim f^{\mathrm{No}}_{\mathbf{C}(S)}$ considered as a sub-collection of $\mathbf{X} \sim f^{\mathrm{No}}_{\mathbf{C}}$. Accordingly, conditional PDFs emerge, $f^{\mathrm{No}}_{S|S'}(\mathbf{x}(S)|\mathbf{x}(S'))$, for pairs of sets S, S' with $S\cap S' = \emptyset$, where $\mathbf{x}(S) \in \mathbb{R}^{\#(S)}$, $\mathbf{x}(S') \in \mathbb{R}^{\#(S')}$. [The PDF $f^{\mathrm{No}}_{S|S'}$ is expressed in terms of block sub-matrices forming the inverse matrix $\mathbf{C}(S\cup S')^{-1}$.]

Further, let a function $\varphi(\mathbf{x}) \geq 0$, $\mathbf{x}\in\mathbb{R}^d$, be given, which is positive on an open domain in $\mathbb{R}^d$ and set, as in (2.2),

$$\psi(S;\mathbf{x}(S)) = \int\limits_{\mathbb{R}^{\#(S^{\complement})}} \varphi(\mathbf{x}) f^{\mathrm{No}}_{S^{\complement}|S}(\mathbf{x}(S^{\complement})\,|\,\mathbf{x}(S))\mathrm{d}\mathbf{x}(S^{\complement}). \quad (5.10)$$

Furthermore, define:

$$\tau(S) = \mathrm{tr}\,\left[\mathbf{C}(S)^{-1}\boldsymbol{\Phi}(S)\right],\ \ \mathrm{T}(k) = \sum_{S\subseteq I^{(d)}:\#(S)=k} \tau(S) \quad (5.11)$$

where matrix $\boldsymbol{\Phi}(S)$ is given by

$$\boldsymbol{\Phi}(S) = \boldsymbol{\Phi}(\mathbf{C}(S)) = \int\limits_{\mathbb{R}^{\#(S)}} \mathbf{x}(S)\mathbf{x}(S)^{\mathrm{T}}\psi(S;\mathbf{x}(S)) f^{\mathrm{No}}_{\mathbf{C}(S)}(\mathbf{x}(S))\mathrm{d}\mathbf{x}(S). \quad (5.12)$$

(For $S = I^{(d)}$, we write simply $\boldsymbol{\Phi}$; cf. (3.4).) Finally, set:

$$\alpha(S) = \alpha(\mathbf{C}(S)) = \int\limits_{\mathbb{R}^{\#(S)}} \psi(S;\mathbf{x}(S)) f^{\mathrm{No}}_{\mathbf{C}(S)}(\mathbf{x}(S))\mathrm{d}\mathbf{x}(S),$$

$$\mathrm{A}(k) = \sum_{S\subseteq I^{(d)}:\#(S)=k} \alpha(S) \quad (5.13)$$

and

$$\lambda(S) = \alpha(S)\log\det \mathbf{C}(S),\ \ \Lambda(k) = \sum_{S\subseteq I^{(d)}:\#(S)=k} \lambda(S). \quad (5.14)$$

Consider the following condition repeating (2.5) for the Gaussian case:

$$\forall i \in S \subseteq I, \text{with} S_i^- = \{j \in S:\ j < i\} \text{and} S_i^+ = \{j \in S:\ j > i\},$$
$$\int\limits_{(\mathbb{R}^{\#(S)}} \psi(S;\mathbf{x}(S))\Big\{f^{\rm No}_{\mathbf{C}(S)}(\mathbf{x}(S))$$
$$- f^{\rm No}_{\mathbf{C}(S_i^-)}(\mathbf{x}(S_i^-))f^{\rm No}_{\{i\}|S_i^-}(x_i|\mathbf{x}(S_i^-))f^{\rm No}_{S_i^+|S_i^-}(\mathbf{x}(S_i^+)|\mathbf{x}(S_i^-))\Big\}{\rm d}\mathbf{x}(S) \geq 0. \tag{5.15}$$

Theorem 5.5 below follows from Theorem 2.2:

Theorem 5.5. (Cf. [3, Theorem 4] or [6, Theorem 31]) *Assume condition* (5.15). *Then the quantity* $m(k) = m(k, \mathbf{C}, \varphi)$ *defined by*

$$m(k) = \binom{d}{k}^{-1}\left[\frac{\log\ \Lambda(k)}{2k} + \frac{\log(2\pi)}{2}A(k) + \frac{\log\ {\rm e}}{2k}T(k)\right] \tag{5.16}$$

is decreasing in $k = 1, \ldots, d$*:*

$$m(1) \geq \cdots \geq m(d). \tag{5.17}$$

Proof. For $\mathbf{X}(S) \sim f^{\rm No}_{\mathbf{C}(S)}$ we have, by using (3.3):

$$\frac{h^{\rm w}_{\psi(S)}(\mathbf{X}(S))}{k} = \frac{\alpha(S)}{2k}\log\ \left[(2\pi)^k \det\mathbf{C}(S)\right] + \frac{\log e}{2k}{\rm tr}\left[\mathbf{C}(S)^{-1}\mathbf{\Phi}(S)\right]. \tag{5.18}$$

Therefore,

$$m(k) = \binom{d}{k}^{-1}\sum_{S:|S|=k}\frac{h^{\rm w}_{\psi(S)}(\mathbf{X}(S))}{k}. \tag{5.19}$$

Invoking Theorem 2.2 completes the proof. □

Theorem 5.6. (Cf. [3, Theorem 5] or [6, Theorem 32]) *Assuming* (5.15)*, for all* $r > 0$ *the values*

$$s(k) = \binom{d}{k}^{-1}\sum_{S\subseteq I^{(d)}:\ \#(S)=k}\Lambda(k)^{1/2k}\exp\left\{r\left[\frac{\log(2\pi)}{2}A(k) + \frac{\log\ {\rm e}}{2k}{\rm T}(k)\right]\right\} \tag{5.20}$$

obey

$$s(1) \geq \cdots \geq s(d). \tag{5.21}$$

Proof. The assertion follows readily from Theorem 2.4. □

Our next goal is to establish inequalities for Toeplitz determinants extending [3, Theorem 6] or [6, Theorem 27]. Recall, $\mathbf{C} = (C_{ij})$ is a $d \times d$ Toeplitz matrix if $C_{ij} = C_{kl}$ whenever $|i-j| = |k-l|$. A more restrictive property is cyclic Toeplitz where $C_{ij} = C_{kl}$ whenever ${\rm dist}_d(i,j) = {\rm dist}_d(k,l)$. Here, for

$1 \leq i < j \leq d$ the cyclic distance $\mathrm{dist}_d(i,j) = \min[j-i, d-j+i]$; it is then extended to a metric with $\mathrm{dist}_d(i,j) = \mathrm{dist}_d(j,i)$ and $\mathrm{dist}_d(i,i) = 0$. As before, we consider sub-matrices $\mathbf{C}(S)$ where $S \subseteq I^{(d)} = \{1, \dots, d\}$ and the Gaussian random vectors $\mathbf{X}(S) \sim f^{\mathrm{No}}_{\mathbf{C}(S)}$ as sub-collections in $\mathbf{X}_1^d = (X_1, \dots, X_d)^T \sim f^{\mathrm{No}}_{\mathbf{C}}$. A special role is played by $S = I_{i,j}$ where $I_{i,j}$ stands for a segment of positive integers $\{i, i+1, \dots, j\}$ of cardinality $j - i + 1$ where $1 \leq i < j \leq d$. In particular, for $S = I_{1,k}$, we set: $\mathbf{C}(S) = \mathbf{C}_k$ and deal with vectors $\mathbf{X}_1^k \sim f^{\mathrm{No}}_{\mathbf{C}_k}$, $1 \leq k \leq d$, with $\mathbf{C}_d = \mathbf{C}$.

Accordingly, we say that WF: $\mathbf{x} \in \mathbb{R}^d \mapsto \varphi(\mathbf{x}) \geq 0$ has a Toeplitz property if the value of the reduced WF $\psi(I_{i,j}; \mathbf{x}_i^j)$ coincides with $\psi(I_{i+k,j+k}; \mathbf{x}_{i+k}^{j+k})$, provided that arguments $\mathbf{x}_i^j = \mathbf{x}(I_{i,j})$ and $\mathbf{x}_{i+k}^{j+k} = \mathbf{x}(I_{i+k,j+k})$ are shifts of each other (with $x_{i+s}(I_{i,j}) = x_{i+k+s}(I_{i+k,j+k})$, for $0 \leq s \leq j - i$), where $1 \leq i < j \leq d$ and $1 \leq i + k < j + k \leq d$. An example is where $\mathbf{C}$ is cyclic Toeplitz and φ has a product-form: $\varphi(\mathbf{x}) = \prod\limits_{1 \leq i \leq d} \widehat{\varphi}(x_i)$. Recall, the reduced WF in question involves the conditional PDF $f^{\mathrm{No}}_{I^{\complement}_{i,j}|I_{i,j}}(\mathbf{x}(I^{\complement}_{i,j})|\mathbf{x}_i^j)$:

$$\psi(I_{i,j}; \mathbf{x}_i^j) = \int\limits_{\mathbb{R}^{d-j+i-1}} \varphi(\mathbf{x}) f^{\mathrm{No}}_{I^{\complement}_{i,j}|I_{i,j}}(\mathbf{x}(I^{\complement}_{i,j})|\mathbf{x}_i^j) \mathrm{d}\mathbf{x}(I^{\complement}_{i,j}) \quad \text{where} \quad I^{\complement}_{i,j} = I_{1,d} \setminus I_{i,j}. \tag{5.22}$$

For $S = I_{1,k}$, $1 \leq k \leq d$, in accordance with (3.3),

$$h_{\psi(k)}(\mathbf{X}_1^k) = h_{\psi(I_{1,k})}(\mathbf{X}_1^k) = \frac{\alpha(\mathbf{C}_k)}{2} \log \left[(2\pi)^k \det \mathbf{C}_k\right] + \frac{\log e}{2} \mathrm{tr}\left[\mathbf{C}_k^{-1} \mathbf{\Psi}_k\right]. \tag{5.23}$$

Here the value $\alpha(\mathbf{C}_k) = \alpha(k, \mathbf{C}, \varphi)$ and the $k \times k$ matrix $\mathbf{\Psi}_k = \mathbf{\Psi}_k(k, \mathbf{C}, \psi)$ are given by

$$\alpha(\mathbf{C}_k) = \int\limits_{\mathbb{R}^k} \psi(k; \mathbf{x}_1^k) f^{\mathrm{No}}_{\mathbf{C}_k}(\mathbf{x}_1^k) \mathrm{d}\mathbf{x}_1^k, \quad \mathbf{\Psi}_k = \int\limits_{\mathbb{R}^k} \mathbf{x}_1^k \left(\mathbf{x}_1^k\right)^{\mathrm{T}} \psi(k; \mathbf{x}_1^k) f^{\mathrm{No}}_{\mathbf{C}_k}(\mathbf{x}_1^k) \mathrm{d}\mathbf{x}_1^k \tag{5.24}$$

and $\psi(k) = \psi(I_{1,k})$. (For $k = d$, the subscript k will be omitted.)

Theorem 5.7. (Cf. [3, Theorem 6] or [6, Theorem 27]) *Suppose $\mathbf{C}_d$ is a positive definite $d \times d$ Toeplitz matrix and φ has the Toeplitz property. Consider the map $k \in \{1, \dots, d\} \mapsto a(k) = a(k, \mathbf{C}, \varphi)$ where*

$$a(k) = \alpha(\mathbf{C}_k) \left\{\log(2\pi) + \log \left[(\det \mathbf{C}_k)^{1/k}\right]\right\} + \frac{\log e}{k} \mathrm{tr}\left[\mathbf{C}_k^{-1} \mathbf{\Psi}_k\right]. \tag{5.25}$$

Assuming condition (5.15), the value $a(k)$ is decreasing in k: $a(1) \geq \cdots \geq a(d)$.

Proof. By using the Toeplitz property of $\mathbf{C}$ and φ, we can write

$$h^{\mathrm{w}}_{\psi(I_{1,k})}(X_k|\mathbf{X}_1^{k-1}) = h^{\mathrm{w}}_{\psi(I_{2,k+1})}(X_{k+1}|\mathbf{X}_2^k). \tag{5.26}$$

Next, Theorem 4.3 yields:

$$h^{\rm w}_{\psi(I_{2,k+1})}(X_{k+1}|\mathbf{X}_2^k) \geq h^{\rm w}_{\psi(I_{1,k+1})}(X_{k+1}|\mathbf{X}_1^k). \tag{5.27}$$

From (5.26) and (5.27) we conclude that $h^{\rm w}_{\psi(I_{1,k})}(X_k|\mathbf{X}_1^{k-1})$ is decreasing in k. Thus the running average also decreases. On the other hand, by the chain rule

$$\frac{1}{k}\, h_{\psi(I_{1,k})}(\mathbf{X}_1^k) = \frac{1}{k}\sum_{i=1}^{k} h^{\rm w}_{\psi(I_{1,i})}(X_i|\mathbf{X}_1^{i-1}). \tag{5.28}$$

Consequently $\dfrac{1}{k}\, h_{\psi(I_{1,k})}(\mathbf{X}_1^k)$ decreases in k too. Referring to Eqs. (5.24) and (5.23) leads directly to the result. $\square$

Theorem 5.8. (Cf. [6, Theorem 33]) *Given a WF: $\mathbf{x} \in \mathbb{R}^d \mapsto \varphi(\mathbf{x})$, assume condition*

$$\int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) \left[f^{\rm No}_C(\mathbf{x}) - \prod_{i=1}^{d} f^{\rm No}_{C_{ii}}(x_i) \right] {\rm d}\mathbf{x} \geq 0. \tag{5.29}$$

Then the quantity

$$\begin{aligned} w(k) = w(k,\mathbf{C},\varphi) &= \binom{d}{k}^{-1} \frac{\alpha(\mathbf{C})}{2k} \log \left[\prod_{S\subseteq I_n:\ \#(S)=k} \frac{(2\pi)^d(\det\,\mathbf{C})}{(2\pi)^{d-k}(\det\,\mathbf{C}(S^{\complement}))} \right] \\ &\quad + \binom{d}{k}^{-1} \frac{\log\ e}{2k} \sum_{S\subseteq I_n:\ \#(S)=k} \left\{ {\rm tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] - {\rm tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right] \right\} \end{aligned} \tag{5.30}$$

is increasing in k, with

$$w(1) \leq \cdots \leq w(d). \tag{5.31}$$

Proof. Using the chain rule for the conditional WE, we can write

$$\begin{aligned} h^{\rm w}_\varphi(\mathbf{X}(S)|\mathbf{X}(S^{\complement})) &= h^{\rm w}_\varphi(\mathbf{X}(S),\mathbf{X}(S^{\complement})) - h^{\rm w}_{\psi(S^{\complement})}(\mathbf{X}(S^{\complement})) \\ &= \frac{\alpha(\mathbf{C})}{2}\log\ \left[(2\pi)^d({\rm det}\mathbf{C})\right] + \frac{\log\ e}{2}{\rm tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] \\ &- \frac{\alpha(\mathbf{C})}{2}\log\ \left[(2\pi)^{d-k}({\rm det}\mathbf{C}(S^{\complement}))\right] - \frac{\log\ e}{2}{\rm tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right]. \end{aligned} \tag{5.32}$$

Here $\alpha(\mathbf{C}) = \int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f^{\rm No}_{\mathbf{C}}(\mathbf{x}){\rm d}\mathbf{x} = \int\limits_{\mathbb{R}^{\#(S^{\complement})}} \psi(\mathbf{x}(S^{\complement})) f^{\rm No}_{\mathbf{C}(S^{\complement})}(\mathbf{x}(S^{\complement})){\rm d}\mathbf{x}(S^{\complement})$. Therefore,

$$h^{\mathrm{w}}_{\varphi}(\mathbf{X}(S)|\mathbf{X}(S^{\complement}))$$
$$= \frac{\alpha(\mathbf{C})}{2} \log \left[\frac{(2\pi)^d(\det \mathbf{C})}{(2\pi)^{d-k}(\det \mathbf{C}(S^{\complement}))}\right] + \frac{\log\, e}{2}\left\{\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] - \mathrm{tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right]\right\}. \tag{5.33}$$

After that we apply Theorem 2.5 which completes the proof. □

Remark 5.9. Note that the outermost inequality, $w(1) \leq w(d)$, can be rewritten as

$$\alpha(\mathbf{C}) \log \left[(2\pi)^d(\det \mathbf{C})\right] + \log\, e\; \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] \geq \alpha(\mathbf{C}) \log \left[\prod_{i=1}^{d} \frac{2\pi(\det \mathbf{C})}{\det \mathbf{C}(I_1^{i-1} \cup I_{i+1}^d)}\right]$$
$$+ \log\, e \sum_{i=1}^{d}\left\{\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] - \mathrm{tr}\left[\mathbf{C}(I_1^{i-1} \cup I_{i+1}^d)^{-1}\mathbf{\Phi}(I_1^{i-1} \cup I_{i+1}^d)\right]\right\}. \tag{5.34}$$

One can note that for ordinary entropies, the outermost inequality can be interpreted as an inequality involving estimation errors; see [6, p. 1517].

Our next goal is to establish additional WDIs by using Theorem 2.7. For this purpose, we first analyse the mutual Gaussian WI, $i^{\mathrm{w}}_{\varphi}(\mathbf{X}(S);\mathbf{X}(S^{\complement}))$. According to the definition of the mutual WI in [18], we can write

$$i^{\mathrm{w}}_{\varphi}(\mathbf{X}(S);\mathbf{X}(S^{\complement})) = h^{\mathrm{w}}_{\psi(S)}(\mathbf{X}(S)) - h^{\mathrm{w}}_{\varphi}(\mathbf{X}(S)|\mathbf{X}(S^{\complement})). \tag{5.35}$$

Then, in accordance with (5.33), we have

$$i^{\mathrm{w}}_{\varphi}(\mathbf{X}(S);\mathbf{X}(S^{\complement})) = \frac{\alpha(\mathbf{C})}{2} \log \left[\frac{(\det\mathbf{C}(S))\;(\det\mathbf{C}(S^{\complement}))}{(\det\mathbf{C})}\right]$$
$$+ \frac{\log\, e}{2}\left\{\mathrm{tr}\left[\mathbf{C}(S)^{-1}\mathbf{\Phi}(S)\right] + \mathrm{tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right] - \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right]\right\}. \tag{5.36}$$

In Theorems 5.10 and 5.11 we consider the following condition (5.37) stemming from (2.18): $\forall\ S \subseteq \{1,\ldots,n\}$ with $\#\,S \geq 2$ and $i,j \in S$ with $i \neq j$,

$$\int_{\mathbb{R}^d} \varphi(\mathbf{x}) f^{\mathrm{No}}_{S^{\complement}|S}(\mathbf{x}(S^{\complement})|\mathbf{x}(S))\Big[f^{\mathrm{No}}_{\mathbf{C}(S)}(\mathbf{x}(S))$$
$$-f^{\mathrm{No}}_{\mathbf{C}(S\setminus\{i,j\})}(\mathbf{x}(S\setminus\{i,j\}))\; f^{\mathrm{No}}_{i|S\setminus\{i,j\}}(\mathbf{x}(S\setminus\{i,j\}))\; f^{\mathrm{No}}_{j|S\setminus\{i,j\}}(x_j|\mathbf{x}(S\setminus\{i,j\}))\Big]\mathrm{d}\mathbf{x} \geq 0. \tag{5.37}$$

The proof of Theorems 5.10 and 5.11 is done with the help of Theorem 2.6, assuming that $X_1, X_2, \ldots, X_d$ are normally distributed with covariance matrix $\mathbf{C}$.

Theorem 5.10. (Cf. [6, Theorem 34]) *Assume condition* (5.37). *Let*

$$u(k) = \binom{d}{k}^{-1} \frac{\alpha(\mathbf{C})}{2k} \log \left[\prod_{S \subseteq I^{(d)}:\ \#(S)=k} \frac{(\det \mathbf{C}(S))\ (\det \mathbf{C}(S^{\complement}))}{(\det \mathbf{C})} \right]$$
$$+ \binom{d}{k}^{-1} \frac{\log\ e}{2k} \sum_{S \subseteq I_n:\ \#(S)=k} \left\{ \mathrm{tr}\left[\mathbf{C}(S)^{-1}\mathbf{\Phi}(S)\right] + \mathrm{tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right] - \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] \right\}. \tag{5.38}$$

Then

$$u(1) \geq u(2) \geq \cdots \geq u(d-1) \geq u(d). \tag{5.39}$$

Theorem 5.11. (Cf. [6, Theorem 35]) *Under condition* (5.37), *let*

$$\begin{aligned} z(k) = \binom{d}{k}^{-1} \frac{\alpha(\mathbf{C})}{2} \log \left[\prod_{S \subseteq I^{(d)}:\ \#(S)=k} \frac{(\det \mathbf{C}(S))\ (\det \mathbf{C}(S^{\complement}))}{(\det \mathbf{C})} \right] \\ + \binom{d}{k}^{-1} \frac{\log\ e}{2} \sum_{S \subseteq I^{(d)}:\ \#(S)=k} \\ \left\{ \mathrm{tr}\left[\mathbf{C}(S)^{-1}\mathbf{\Phi}(S)\right] + \mathrm{tr}\left[\mathbf{C}(S^{\complement})^{-1}\mathbf{\Phi}(S^{\complement})\right] - \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] \right\}. \end{aligned} \tag{5.40}$$

Then

$$z(1) \geq z(2) \geq \cdots \geq z(\lfloor d/2 \rfloor). \tag{5.41}$$

6. Weighted Hadamard-type inequalities

In this section we group several results related to the weighted Hadamard inequality (WHI); cf. [6, Theorem 3.7]. The WHI inequality asserts that for a $d \times d$ positive definite matrix $\mathbf{C}$, under condition (5.29) we have:

$$\begin{aligned} \alpha(\mathbf{C}) \log((2\pi)^d \prod_i \mathbf{C}_{ii}) + (\log\ \mathrm{e}) \sum_i \mathbf{C}_{ii}^{-1} \Phi_{ii} \\ -\alpha(\mathbf{C}) \log((2\pi)^d \det \mathbf{C}) - (\log\ \mathrm{e}) \mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{\Phi}\right) \geq 0, \end{aligned} \tag{6.1}$$

with equality iff $\mathbf{C}$ is diagonal. Recall, $\alpha(\mathbf{C}) = \alpha_\varphi(\mathbf{C})$ and $\mathbf{\Phi} = \mathbf{\Phi}_{\mathbf{C}} = \Phi_{\mathbf{C},\varphi}$ are as in (3.4). For $\varphi \equiv 1$, it becomes $\det \mathbf{C} \leq \prod\limits_{1 \leq i \leq d} \mathbf{C}_{ii}$, the famous inequality due to Hadamard.

We begin with the weighted version of the strong Hadamard inequality (WSHI). This inequality (and other inequalities in this section) will involve determinants $\det \mathbf{C}(S)$ of sub-matrices $\mathbf{C}(S)$ in $\mathbf{C}$ where, as before, S is a subset of $I^{(d)} = \{1, \ldots, d\}$ of a special type. Namely, we fix $p \in \{1, \ldots, d-1\}$ and consider the segment $I_{p+1,d} = \{p+1, \ldots, d\}$, segment $I_{1,p} = \{1, \ldots, p\}$

and unions $\{i\}\cup I_{p+1,d}$ and $I_{1,i}\cup I_{p+1,d}=I^{\complement}_{i+1,p}$ where $i\in I_{1,p}$. We deal with the related entry C_{ii} in $\mathbf{C}$ and sub-matrices

$$\mathbf{C}^d_{p+1}=\mathbf{C}(I_{p+1,d}),\ \mathbf{C}^{i-1}_1=\mathbf{C}(I_{1,i-1}),\ \mathbf{C}(\{i\}\cup I_{p+1,d})\ \text{ and }\ \mathbf{C}(I_{1,i}\cup I_{p+1,d})$$

and Gaussian random variables X_i and vectors $\mathbf{X}^d_{p+1}=\mathbf{X}(I_{p+1,d})$, $\mathbf{X}^{i-1}_1=\mathbf{X}(I_{1,i-1})$, $X_i\vee\mathbf{X}^d_{p+1}=\mathbf{X}(\{i\}\cup I_{p+1,d})$ and $\mathbf{X}^i_1\vee\mathbf{X}^d_{p+1}=\mathbf{X}(I_{1,i}\cup I_{p+1,d})$ using symbols x_i, $\mathbf{x}^d_{p+1}$, $\mathbf{x}^{i-1}_1$, and $\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1}$ for their respective values. For simplicity, let us omit henceforth the subscript No indicating normality. Then the PDFs

$$f_{\mathbf{X}^d_{p+1}}(\mathbf{x}^d_{p+1})=f_{\mathbf{C}^d_{p+1}}(\mathbf{x}^d_{p+1})\text{ and }f_{\mathbf{X}^i_1\vee\mathbf{X}^d_{p+1}}(\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1})=f_{\mathbf{C}(I_{1,i}\cup I_{p+1,d})}(\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1})$$

emerge, as well as conditional PDFs $f_{X_i|\mathbf{X}^d_{p+1}}(x_i|\mathbf{x}^d_{p+1})$ and $f_{\mathbf{X}^{i-1}_1|\mathbf{X}^d_{p+1}}(\mathbf{x}^{i-1}_1|\mathbf{x}^d_{p+1})$. Viz., $\mathbf{X}^i_1\vee\mathbf{X}^d_{p+1}$ and $\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1}$ stand for the concatenated vectors $(X_1,\dots,X_i,X_{p+1},\dots,X_d)^T$ and $(x_1,\dots,x_i,x_{p+1},\dots,x_d)$, each with $i+d-p$ entries. As above (see (3.4)), for a given WF: $\mathbf{x}\in\mathbb{R}^d\mapsto\varphi(\mathbf{x})$ we consider numbers $\alpha(\mathbf{C}^p_1)=\alpha_\varphi(p,\mathbf{C})$ and matrices $\mathbf{\Phi}_{\mathbf{C}^p_1}=\mathbf{\Phi}_{p,\mathbf{C},\varphi}$:

$$\alpha(\mathbf{C}^p_1)=\int\limits_{\mathbb{R}^d}\varphi(\mathbf{x}^p_1)f_{\mathbf{C}^p_1}(\mathbf{x}^p_1)\mathrm{d}\mathbf{x}^p_1,\quad \mathbf{\Phi}_{\mathbf{C}^p_1}=\int\limits_{\mathbb{R}^d}\mathbf{x}^p_1\left(\mathbf{x}^p_1\right)^{\mathrm{T}}\varphi(\mathbf{x}^p_1)f_{\mathbf{C}}(\mathbf{x}^p_1)\mathrm{d}\mathbf{x}^p_1. \tag{6.2}$$

We also set

$$\begin{aligned}\mathbf{\Phi}^d_{p+1}&=\int\limits_{\mathbb{R}^{p-d}}\mathbf{x}^d_{p+1}\left(\mathbf{x}^d_{p+1}\right)^{\mathrm{T}}\psi(I_{p+1,d};\mathbf{x}^d_{p+1})\,f_{\mathbf{X}^d_{p+1}}(\mathbf{x}^d_{p+1})\,\mathrm{d}\mathbf{x}^d_{p+1},\\ \mathbf{\Phi}(\{i\}\cup I_{p+1,d})&=\int\limits_{\mathbb{R}^{p-d+1}}(x_i\vee\mathbf{x}^d_{p+1})\left(x_i\vee\mathbf{x}^d_{p+1}\right)^{\mathrm{T}}\\ &\quad\times\psi(\{i\}\cup I_{p+1,d};x_i\vee\mathbf{x}^d_{p+1})f_{X_i\vee\mathbf{X}^d_{p+1}}(x_i\vee\mathbf{x}^d_{p+1})\mathrm{d}(x_i\vee\mathbf{x}^d_{p+1}),\end{aligned} \tag{6.3}$$

with reduced WFs $\psi(I_{p+1,d})$ and $\psi(\{i\}\cup I_{p+1,d})$ calculated as in (2.2), for $S=I_{p+1,d}$ and $S=\{i\}\cup I_{p+1,d}$.

Furthermore, we will assume in Theorem 6.1 that, $\forall\ i=1,\dots,p$, the reduced WF $\psi(S)$ with $S=\{1,\dots i,p+1,\dots d\}=I^{\complement}_{i+1,p}$ obeys

$$\begin{aligned}&\int\limits_{\mathbb{R}^{i+d-p}}\psi(I^{\complement}_{i+1,p};\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1})\Big\{f_{\mathbf{X}^i_1\vee\mathbf{X}^d_{p+1}}(\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1})\\ &\quad-f_{\mathbf{X}^d_{p+1}}(\mathbf{x}^d_{p+1})f_{X_i|\mathbf{X}^d_{p+1}}(x_i|\mathbf{x}^d_{p+1})f_{\mathbf{X}^{i-1}_1|\mathbf{X}^d_{p+1}}(\mathbf{x}^{i-1}_1|\mathbf{x}^d_{p+1})\Big\}\mathrm{d}(\mathbf{x}^i_1\vee\mathbf{x}^d_{p+1})\geq 0.\end{aligned} \tag{6.4}$$

The 'standard' SHI is

$$\frac{\det \mathbf{C}}{\det \mathbf{C}^d_{p+1}} \leq \prod_{1\leq i\leq p} \frac{\det \mathbf{C}(\{i\}\cup I_{p+1,d})}{\det \mathbf{C}^d_{p+1}}$$
$$\text{or} \quad \log\, \det \mathbf{C} + (p-1)\log\, \det \mathbf{C}^d_{p+1} \leq \sum_{1\leq i\leq p} \log\, \det \mathbf{C}(\{i\}\cup I_{p+1,d}). \tag{6.5}$$

The WE approach offers the following WSHI:

Theorem 6.1. (Cf. [3, Theorem 8] or [6, Theorem 28]) *Under condition* (6.4), *for* $1 \leq p < d$,

$$\begin{aligned} &\alpha(\mathbf{C}) \log\, \left[(2\pi)^d \det \mathbf{C}\right] + (\log\, e)\mathrm{tr}\,(\mathbf{C}^{-1}\boldsymbol{\Phi}) \\ &\quad + (p-1)\Big\{\alpha(\mathbf{C}^d_{p+1}) \log\, \left[(2\pi)^{d-p} \det \mathbf{C}^d_{p+1}\right] + (\log\, e)\mathrm{tr}\,[(\mathbf{C}^d_{p+1})^{-1}\boldsymbol{\Phi}^d_{p+1}]\Big\} \\ &\quad \leq \sum_{1\leq i\leq p} \Bigg\{\alpha(\mathbf{C}(\{i\}\cup I_{p+1,d})) \log\, \left[(2\pi)^{d-p+1}\det \mathbf{C}(\{i\}\cup I_{p+1,d})\right] \\ &\quad + (\log\, e)\mathrm{tr}\,[\mathbf{C}(\{i\}\cup I_{p+1,d})^{-1}\boldsymbol{\Phi}(\{i\}\cup I_{p+1,d})]\Bigg\}. \end{aligned} \tag{6.6}$$

Proof. We use the same idea as in [6, Theorem 3.7]. Recalling (6.13) we can write

$$\begin{aligned} h^{\mathrm{w}}_{\varphi}(\mathbf{X}^p_1|\mathbf{X}^d_{p+1}) &= \frac{1}{2}\, \log\, \left[(2\pi)^d \det \mathbf{C}\right]\, \alpha(\mathbf{C}) + \frac{\log\, e}{2}\mathrm{tr}\,(\mathbf{C}^{-1}\boldsymbol{\Phi}) \\ &\quad - \frac{1}{2} \log\, \left[(2\pi)^{d-p} \det \mathbf{C}^d_{p+1}\right]\, \alpha(\mathbf{C}^d_{p+1}) - \frac{\log\, e}{2}\mathrm{tr}\,[(\mathbf{C}^d_{p+1})^{-1}\boldsymbol{\Phi}^d_{p+1}], \end{aligned} \tag{6.7}$$

Cf. Eqs. (5.12), (5.13), (5.24). Furthermore, by the subadditivity of the conditional WE (see [6, Theorem 1.8]), under assumption (6.4) we can write

$$h^{\mathrm{w}}_{\varphi}(\mathbf{X}^p_1|\mathbf{X}^d_{p+1}) \leq \sum_{i=1}^{p} h^{\mathrm{w}}_{\psi(\{i\}\cup I_{p+1,d})}(X_i|\mathbf{X}^d_{p+1}). \tag{6.8}$$

Here for $i = 1, \dots, p$, again in agreement with (6.13),

$$\begin{aligned} h^{\mathrm{w}}_{\psi(\{i\}\cup I_{p+1,d})}(X_i|\mathbf{X}^d_{p+1}) &= \frac{1}{2} \log\left[(2\pi)^{d-p+1}\det \mathbf{C}(\{i\}\cup I_{p+1,d})\right] \alpha(\mathbf{C}(\{i\}\cup I_{p+1,d})) \\ &\quad + \frac{\log\, e}{2}\mathrm{tr}\, \mathbf{C}(\{i\}\cup I_{p+1,d})^{-1}\boldsymbol{\Phi}(\{i\}\cup I_{p+1,d}) \\ &\quad - \frac{1}{2} \log\, \left[(2\pi)^{d-p} \det \mathbf{C}^d_{p+1}\right]\, \alpha(\mathbf{C}^d_{p+1}) \\ &\quad - \frac{\log\, e}{2}\mathrm{tr}\,[(\mathbf{C}^d_{p+1})^{-1}\boldsymbol{\Phi}^d_{p+1}]. \end{aligned} \tag{6.9}$$

Substituting into (6.8) yields the assertion of the theorem. $\square$

Our next result, Theorem 6.2, gives an extension of [3, Lemma 9] (or [6, Lemma 8]). Let $\widehat{\mathbf{C}}_{dd}=\mathbf{C}_{dd}-\mathbf{C}^1_{d-1}\left(\mathbf{C}_1^{d-1}\right)^{-1}\left(\mathbf{C}^1_{d-1}\right)^{\mathrm{T}}$ be the mean square error of estimate of X_d by observations $\mathbf{X}_1^{d-1}$. Then

$$\widehat{\mathbf{C}}_{dd}=\frac{\det\mathbf{C}}{\det\mathbf{C}_1^{d-1}},\quad\text{or}\quad\log\,\widehat{\mathbf{C}}_{dd}+\log\,\det\mathbf{C}_1^{d-1}-\log\,\det\mathbf{C}=0. \tag{6.10}$$

Remarkably, Theorem 6.2 does not require assumption (6.4), in fact, this is a purely algebraic identity.

Theorem 6.2. (Cf. [3, Lemma 9] or [6, Lemma 8]) *The following equality holds true:*

$$\begin{aligned}\alpha(\widehat{\mathbf{C}}_{dd})\log\left[(2\pi)\widehat{\mathbf{C}}_{dd}\right]+\alpha(\mathbf{C}_1^{d-1})\log\left[(2\pi)^{d-1}\det\,\mathbf{C}_1^{d-1}\right]-\alpha(\mathbf{C})\log\left[(2\pi)^{d}\det\,\mathbf{C}\right]\\ =(\log\,\mathrm{e})\,\mathrm{tr}\,\left[\mathbf{C}^{-1}\mathbf{\Phi}\right]-(\log\,\mathrm{e})\,\mathrm{tr}\,\left[\left(\mathbf{C}_1^{d-1}\right)^{-1}\mathbf{\Phi}_1^{d-1}\right]-(\log\,\mathrm{e})\,\widehat{\mathbf{C}}_{dd}^{-1}\Phi_{dd}.\end{aligned} \tag{6.11}$$

Proof. Using the conditional normality of X_d given $\mathbf{X}_1^{d-1}$, we can write

$$h^{\mathrm{w}}_{\varphi}(X_d|\mathbf{X}_1^{d-1})=\frac{\alpha(\widehat{\mathbf{C}}_{dd})}{2}\log\,\left[(2\pi)\widehat{\mathbf{C}}_{dd}\right]+\frac{\log\,\mathrm{e}}{2}\,\widehat{\mathbf{C}}_{dd}^{-1}\Phi_{dd}.$$

On the other hand,

$$h^{\mathrm{w}}_{\varphi}(X_d|\mathbf{X}_1^{d-1})=h^{\mathrm{w}}_{\varphi}(\mathbf{X}_1^{d})-h^{\mathrm{w}}_{\psi(I_{1,d-1})}(\mathbf{X}_1^{d-1}), \tag{6.12}$$

and therefore

$$\begin{aligned}&\frac{\alpha(\widehat{\mathbf{C}}_{dd})}{2}\log\,\left[(2\pi)\widehat{\mathbf{C}}_{dd}\right]+\frac{\log\,\mathrm{e}}{2}\,\widehat{\mathbf{C}}_{dd}^{-1}\Phi_{dd}\\ &\quad=\frac{\alpha(\mathbf{C})}{2}\log\,\left[(2\pi)^d\det\,\mathbf{C}\right]+\frac{\log\,\mathrm{e}}{2}\,\mathrm{tr}\,\mathbf{C}^{-1}\mathbf{\Phi}\\ &\qquad-\frac{\alpha(\mathbf{C}_1^{d-1})}{2}\log\,\left[(2\pi)^{d-1}\det\,\mathbf{C}_1^{d-1}\right]-\frac{\log\,\mathrm{e}}{2}\,\mathrm{tr}\,\left[\left(\mathbf{C}_1^{d-1}\right)^{-1}\mathbf{\Phi}_1^{(d-1)}\right].\end{aligned} \tag{6.13}$$

The result then follows. □

The next assertion, Theorem 6.3, extends the result of [3, Theorem 9] (or [3, Theorem 29]) that, $\forall\,p=1,\ldots,d$, $\mathbf{C}\mapsto\log\,\dfrac{\det\mathbf{C}}{\det\mathbf{C}_1^p}$ is a concave function of a positive definite $d\times d$ matrix $\mathbf{C}$. We will write matrix $\mathbf{C}$ in the block form similar to (3.10):

$$\mathbf{C}=\begin{pmatrix}\mathbf{C}_1^p & \mathbf{C}_{d-p}^p\\ \mathbf{C}_p^{d-p} & \mathbf{C}_{p+1}^d\end{pmatrix}, \tag{6.14}$$

with $\left(\mathbf{C}^p_{d-p}\right)^{\mathrm{T}} = \mathbf{C}^{d-p}_p$. Set $\mathbf{D} = \mathbf{C}^p_{d-p}\ (\mathbf{C}^d_{p+1})^{-1}$ and $\mathbf{K}^p_1 = \mathbf{C}^p_1 - \mathbf{C}^p_{d-p}\ (\mathbf{C}^d_{p+1})^{-1}\ \mathbf{C}^{d-p}_p$. Consider the following inequalities

$$\int\limits_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}^d_{p+1}}(\mathbf{x}^d_{p+1})\Big[f_{\mathbf{X}^d_1|\mathbf{X}^d_{p+1}}(\mathbf{x}^p_1|\mathbf{x}^d_{p+1}) - f_{\mathbf{Y}^p_1|\mathbf{Y}^d_{p+1}}(\mathbf{x}^p_1|\mathbf{x}^d_{p+1})\Big]\mathrm{d}\mathbf{x} \geq 0 \quad (6.15)$$

and

$$\int\limits_{\mathbb{R}^d} \varphi(\mathbf{x})\Big[f_{\mathbf{X}}(\mathbf{x}) - f_{\mathbf{C}}(\mathbf{x})\Big]\Big\{\log\ [(2\pi)^p \det\ (\mathbf{K}^p_1)] + (\log\ \mathrm{e})\ \Big[\left(\mathbf{x}^p_1 - \mathbf{D}\mathbf{x}^d_{p+1}\right)^{\mathrm{T}} (\mathbf{K}^p_1)^{-1} \left(\mathbf{x}^p_1 - \mathbf{D}\mathbf{x}^d_{p+1}\right)\Big]\Big\}\mathrm{d}\mathbf{x} \leq 0. \quad (6.16)$$

Theorem 6.3. (Cf. [3, Theorem 9] or [6, Theorem 29]) *Assume that* $\mathbf{C} = \lambda\mathbf{C}' + (1-\lambda)\mathbf{C}''$ *where* $\mathbf{C}$, $\mathbf{C}'$ *and* $\mathbf{C}''$ *are positive definite* $d\times d$ *matrices and* $\lambda \in [0,1]$. *Given a WF:* $\mathbf{x} \mapsto \varphi(\mathbf{x}) \geq 0$ *and* $1 \leq p \leq d$, *define:*

$$\begin{aligned}\mu(\mathbf{C}) &= h^{\mathrm{w}}_{\varphi}(\mathbf{X}^p_1|\mathbf{X}^d_{p+1}) \\ &= \frac{1}{2}\Big\{\alpha(\mathbf{C})\log\ \left[(2\pi)^d \det \mathbf{C}\right] + (\log\ \mathrm{e})\ \mathrm{tr}\ \left[\mathbf{C}^{-1}\mathbf{\Phi}_{\mathbf{C}}\right] \\ &\quad - \alpha(\mathbf{C}^d_{p+1})\log\ \left[(2\pi)^p \det \mathbf{C}^d_{p+1}\right] - (\log\ \mathrm{e})\ \mathrm{tr}\ \left[\left(\mathbf{C}^d_{p+1}\right)^{-1}\mathbf{\Phi}_{\mathbf{C}^d_{p+1}}\right]\Big\},\end{aligned} \quad (6.17)$$

and similarly with $\mu(\mathbf{C}')$ *and* $\mu(\mathbf{C}'')$. *Then, under conditions* (6.15) *and* (6.16),

$$\mu(\mathbf{C}) \geq \lambda\mu(\mathbf{C}') + (1-\lambda)\mu(\mathbf{C}''). \quad (6.18)$$

Proof. Again we essentially follow the method in [3] with modifications developed in [18]. Fix two $d \times d$ positive definite matrices $\mathbf{C}'$ and $\mathbf{C}''$ and set $\mathbf{X}' \sim f_{\mathbf{C}'}$, $\mathbf{X}'' \sim f_{\mathbf{C}''}$. Given $\lambda \in [0,1]$, consider a random variable Θ taking values $\vartheta = 1, 2$ with probabilities λ and $1-\lambda$ independently of $(\mathbf{X}', \mathbf{X}'')$. Next, set

$$\mathbf{X} = \begin{cases}\mathbf{X}', & \text{when} \Theta = 1,\\ \mathbf{X}'', & \text{when} \Theta = 2.\end{cases}$$

Then $\mathbf{X} \sim (\lambda f_{\mathbf{C}'} + (1-\lambda) f_{\mathbf{C}''})$ and the covariance matrix $\mathrm{Cov}\,\mathbf{X} = \lambda\mathbf{C}' + (1-\lambda)\mathbf{C}'' = \mathbf{C}$.

With the WF $\widetilde{\varphi}(\mathbf{x}^d_1, \vartheta) = \varphi(\mathbf{x}^d_1)$, use [6, Theorem 2.1] and Theorem 3.2 from Sect. 3 (which is possible under (6.15) and (6.16)) and write:

$$h^{\mathrm{w}}_{\widetilde{\varphi}}(\mathbf{X}^d_{p+1}|\mathbf{X}^p_1, \Theta) \leq h^{\mathrm{w}}_{\varphi}(\mathbf{X}^d_{p+1}|\mathbf{X}^p_1) \leq h^{\mathrm{w}}_{\varphi}(\mathbf{Y}^d_{p+1}|\mathbf{Y}^p_1). \quad (6.19)$$

Here $\mathbf{Y}$ stands for the Gaussian random vector with the PDF $f_{\mathbf{C}}(\mathbf{x}^d_1)$. The LHS in (6.19) coincides with $\lambda\mu(\mathbf{C}') + (1-\lambda)\mu(\mathbf{C}'')$ and the RHS with $\mu(\mathbf{C})$. This completes the proof. □

In a particular case $p = d - 1$, the function $\mathbf{C} \mapsto \dfrac{\det \mathbf{C}}{\det \mathbf{C}_1^{d-1}}$ is also concave. (See [3, Theorem 10] or [6, Theorem 30].) It is challenging to establish a weighted version of this assertion. In this paper we make a step towards such a result: see Theorem 6.5 below. A crucial part is played by Theorem 4.3, with X represented by the random variable $Z_d \sim f_{A_{dd}+B_{dd}}$ and Y associated with the independent Gaussian pair of vectors $(\mathbf{X}_1^{d-1}, \mathbf{Y}_1^{d-1})$ having the joint PDF

$$f_{\mathbf{X}_1^{d-1},\mathbf{Y}_1^{d-1}}(\mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) = f_{\mathbf{A}_1^{d-1}}(\mathbf{x}_1^{d-1}) f_{\mathbf{B}_1^{d-1}}(\mathbf{y}_1^{d-1}).$$

The random element Z from Theorem 4.3 is represented by $\mathbf{Z}_1^{d-1} = \mathbf{X}_1^{d-1} + \mathbf{Y}_1^{d-1}$, and the map ξ takes $(\mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) \mapsto \mathbf{x}_1^{d-1} + \mathbf{y}_1^{d-1}$.

Next, introduce a WF

$$(z, \mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) \in \mathbb{R} \times \mathbb{R}^{d-1} \times \mathbb{R}^{d-1} \mapsto \varphi(z, \mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) \tag{6.20}$$

and consider the following inequality involving conditional normal PDFs $f_{Z_d|\mathbf{X}_1^{d-1},\mathbf{Y}_1^{d-1}}$ and $f_{Z_d|\mathbf{Z}_1^{d-1}}$:

$$\begin{aligned}\int\limits_{\mathbb{R}\times\mathbb{R}^{d-1}\times\mathbb{R}^{d-1}} & \varphi(z, \mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) f_{\mathbf{A}_1^{d-1}}(\mathbf{x}_1^{d-1}) f_{\mathbf{B}_1^{d-1}}(\mathbf{y}_1^{d-1}) \\ & \times \Big[f_{Z_d|\mathbf{X}_1^{d-1},\mathbf{Y}_1^{d-1}}(z|\mathbf{x}_1^{d-1}, \mathbf{y}_1^{d-1}) - f_{Z_d|\mathbf{Z}_1^{d-1}}(z|\mathbf{x}_1^{d-1} + \mathbf{y}_1^{d-1}) \Big] \mathrm{d}z \mathrm{d}\mathbf{x}_1^{d-1} \mathbf{y}_1^{d-1} \geq 0.\end{aligned} \tag{6.21}$$

Theorem 6.4. *Let* $\mathbf{A}$, $\mathbf{B}$ *be two positive definite* $d \times d$ *matrices and* $\mathbf{X} \sim f_{\mathbf{A}}$, $\mathbf{Y} \sim f_{\mathbf{B}}$ *be the corresponding independent Gaussian vectors, with* $\mathbf{Z}{:=}\mathbf{X}+\mathbf{Y} \sim f_{\mathbf{A}+\mathbf{B}}$. *Then, under condition* (6.21),

$$h^{\mathrm{w}}_{\psi}(Z_d|\mathbf{Z}_1^{d-1}) \geq h^{\mathrm{w}}_{\varphi}(X_d + Y_d|\mathbf{X}_1^{d-1}, \mathbf{Y}_1^{d-1}). \tag{6.22}$$

Proof. The assertion follows by virtue of (3.13) and Theorem 4.3. □

Finally, combining (5.34) and (6.1), we offer

Theorem 6.5. (Cf. [6, Corollary 4]) *Given a* $d \times d$ *positive definite matrix* $\mathbf{C}$, *assume condition* (5.29). *Then*

$$\begin{aligned}\alpha(\mathbf{C}) \log & \left[\prod_{i=1}^{d} \frac{2\pi(\det \mathbf{C})}{\det \mathbf{C}(I_1^{i-1} \cup I_{i+1}^d)} \right] \\ & + \log\, e \sum_{i=1}^{d} \left\{ \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] - \mathrm{tr}\left[\mathbf{C}(I_1^{i-1} \cup I_{i+1}^d)^{-1} \mathbf{\Phi}(I_1^{i-1} \cup I_{i+1}^d)\right] \right\} \\ \leq\ & \alpha(\mathbf{C}) \log \big((2\pi)^d \det \mathbf{C} \big) + (\log \mathrm{e}) \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{\Phi}\right] \\ \leq\ & \alpha(\mathbf{C}) \log \Big((2\pi)^d \prod_i \mathbf{C}_{ii} \Big) + (\log \mathrm{e}) \sum_i \mathbf{C}_{ii}^{-1} \Phi_{ii}.\end{aligned} \tag{6.23}$$

Acknowledgements

I. S. and Y. S. thank the Math Department, Penn State University for hospitality and support. M. K. was supported by the Russian Academic Excellence Project “5–100”. S. Y. S. acknowledges support from NSF under grant NSF2053480.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Bappy, J.H., Paul, S., Tuncel, E., Roy-Chowdhury, A.K.: Exploiting typicality for selecting informative and anomalous samples in videos. IEEE Trans. Image Process. **28**(10), 5214–5226 (2019)
[2] Belis, M., Guiasu, S.: A quantitative and qualitative measure of information in cybernetic systems. IEEE Trans. Inf. Theory **14**, 593–594 (1968)
[3] Cover, T., Thomas, J.: Determinant inequalities via information theory. SIAM J. Matrix Anal. Appl. **9**(3), 384–392 (1988)
[4] Cover, T., Thomas, J.: Elements of Information Theory. Wiley, New York (2006)
[5] Dahlquist, G., Björck, A. (eds.): Numerical Methods, p. 161. Prentice-Hall, Englewood Cliffs (1974)
[6] Dembo, A., Cover, T., Thomas, J.: Information-theoretic inequalities. IEEE Trans. Inform. Theory **37**(6), 1501–1518 (1991)
[7] Frizelle, G., Suhov, Y.M.: The measurement of complexity in production and other commercial systems. Proc. R. Soc. A **464**, 2649–2668 (2008)
[8] Ganesh, M.R., Blanchard, D., Corso, J.J., Yasaei Sekeh, S.: Slimming neural networks using adaptive connectivity scores. arXiv:2006.12463
[9] Kasianova, K., Kelbert, M., Mozgunov, P.: Response adaptive designs for Phase II trials with binary endpoint based on context-dependent information measures. Comput. Stat. Data Anal. **158**, 107187 (2021)
[10] Kelbert, M., Stuhl, I., Suhov, Y.: Weighted entropy and its use in computer science and beyond. In: Rykov, V., Singpurwalla, N., Zubkov, A. (eds.) Analytical and Computational Methods in Probability Theory. ACMPT 2017. Lecture Notes in Computer Science, vol. 10684, pp. 293–308. Springer, Cham (2017)
[11] Kelbert, M., Stuhl, I., Suhov, Y.: Weighted entropy and optimal portfolios for risk-averse Kelly investments. Aequationes Math. **92**(1), 165–200 (2018)
[12] Kelbert, M., Suhov, Y.: Information Theory and Coding by Example. CUP, Cambridge (2013)
[13] MacDonald, R.A., Smith, S.L.: Active sensing for motion planning in uncertain environments via mutual information policies. Int. J. Robot. Res. **38**(2–3), 146–161 (2019)
[14] MacDonald, R.A., Smith, S.L.: Reactive motion planning in uncertain environments via mutual information policies. In: Goldberg, K., Abbeel, P., Bekris, K., Miller, L. (eds.) Algorithmic Foundations of Robotics XII. Springer Proceedings in Advanced Robotics, vol. 13, pp. 256–271. Springer, Cham (2020)
[15] Man’ko, V.I., Seilov, Z.: Weighted information and weighted entropic inequalities for qutrit states. J. Russ. Laser Res. **37**(6), 591–597 (2016)
[16] Miller, K.S.: On the inverse of the sum of matrices. Math. Mag. **54**(2), 67–72 (1981)
[17] Mozgunov, P., Jaki, T.: An information theoretic phase I–II design for molecularly targeted agents that does not require an assumption of monotonicity. J. R. Stat. Soc. Ser. C Appl. Stat. **68**(2), 347–367 (2019)

[18] Suhov, Y., Stuhl, I., Yasaei Sekeh, S., Kelbert, M.: Basic inequalities for weighted entropies. Aequationes Math. **90**(4), 817–848 (2016)
[19] Suhov, Y., Stuhl, I.: Selected data compression: a refinement of Shannon's principle. In: Rykov, V., Singpurwalla, N., Zubkov, A. (eds) Analytical and Computational Methods in Probability Theory. ACMPT 2017. Lecture Notes in Computer Science, vol. 10684, pp. 309–321. Springer, Cham (2017)
[20] Suhov, Y., Stuhl, I.: Weighted information and entropy rates. arXiv:1612.09169
[21] Yan, L., Shen, H., Chen, K., Liu, G.: MobileCopy: improving data availability and file search efficiency in delay tolerant networks against correlated node failure. IEEE Trans. Mob. Comput. (2019). https://doi.org/10.1109/TMC.2019.2939792
[22] Zhao, R., Sun, X., Tresp, V.: Maximum entropy-regularized multi-goal reinforcement learning. arXiv:1905.08786
[23] Zheng, Y., Zhong, H., Fang, Y., Zhang, W., Liu, K., Fang, J.: Rockburst prediction model based on entropy weight integrated with grey relational BP neural network. Fang. Adv. Civil Eng. (2019). https://doi.org/10.1155/2019/3453614

I. Stuhl and Y. Suhov
Mathematics Department
Penn State University
University Park, State College PA 16802
USA
e-mail: ius68@psu.edu

M. Kelbert
National Research University the Higher School of Economics
Moscow
Russian Federation

Y. Suhov
DPMMS
University of Cambridge
Cambridge
UK

and

IPIT RAS
Moscow
Russian Federation

S. Yasaei Sekeh
School of Computing and Information Science
University of Maine
Orono ME
USA

Received: August 18, 2020
Revised: November 25, 2021
Accepted: November 26, 2021