A Note on the Evidence Approximation in Bayesian Experimental Design Models Based on an Orthonormal System

Yoshifumi Ukita
Yokohama College of Commerce
4-11-1 Higashiterao, Tsurumi-ku, Yokohama, Kanagawa 230-8577, Japan
E-mail: ukita@shodai.ac.jp

Abstract In this paper, we first present a theorem on the relationship between the traditional model and a model based on an orthonormal system in experimental design. Using the theorem, the former model can be converted to the latter, and vice versa. Next, we introduce prior distributions over the hyperparameters for experimental design models to consider fully Bayesian predictions. Combining the conversion and a previous result, we show that we can make an approximation in which we set the hyperparameters to specific values determined by maximizing the evidence function.

Keywords: evidence approximation, experimental design, linear model, Bayesian machine learning, orthonormal system, Fourier analysis

1. Introduction
In most areas of scientific research, experimentation is a major tool for acquiring new knowledge or a better understanding of the target phenomenon. Experiments usually aim to study how changes in various factors affect the results of a target, and what values the factors should be set to in order to improve the results. It is important to decide how to collect data, which has been studied in the research field of experimental design. Although experimental design originated from agricultural experiments, it is now also widely used in fields of natural sciences including engineering, biology, and medicine [1–7]. In recent decades, Bayesian prediction using prior knowledge as well as information from data has been widely used in various fields [8]. As it is likely that the researcher has prior knowledge of some aspect of the planned experiment, Bayesian experimental design has recently received much attention [9–11].

In machine learning, Bishop [12] introduced some analytical properties of Bayesian methods through models made from linear combinations of basis functions. In experimental design, the traditional model is often expressed through the effect of each factor [5–7]. On the other hand, it has also been shown that the model can be expressed in terms of orthonormal basis functions by using complex coefficients, making all parameters independent [13, 14]. In [15], using a model based on an orthonormal system and the result of [12], it was shown that a posterior distribution can be analytically derived within a Bayesian framework when the hyperparameters are known. However, the result for unknown hyperparameters has not been obtained.

In this paper, we first present a theorem on the relationship between the traditional model and a model based on an orthonormal system in experimental design. Using the theorem, the former model can be converted to the latter, and vice versa. Next, we introduce prior distributions over the hyperparameters for experimental design models to consider fully Bayesian predictions. Combining the conversion and the result of [12], we show that we can make an approximation in which we set the hyperparameters to specific values determined by maximizing the evidence function.

2. Preliminaries
2.1 Gaussian distribution
In the case of a single variable \( x \), the Gaussian distribution takes the form

\[
N(x | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)
\]

(1)

where \( \mu \) is the mean and \( \sigma^2 \) is the variance.

2.2 Multivariate Gaussian distribution
In the case of an \( N \)-dimensional vector \( x \), the multivariate Gaussian distribution takes the form

\[
N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)
\]

(2)

where \( \mu \) is an \( N \)-dimensional mean vector, \( \Sigma \) is an \( N \times N \) covariance matrix, \( |\Sigma| \) denotes the determinant of \( \Sigma \), and \( \Sigma^{-1} \) is the inverse of \( \Sigma \).
2.3 Galois field

Assume that $q$ is a prime power. Let $GF(q)$ be a Galois field of order $q$, where a Galois field is a field that contains a finite number of elements. We also use $GF(q)^n$ to denote the set of all $n$-tuples with entries from $GF(q)$. The elements of $GF(q)^n$ are referred to as vectors.

Example 1 Consider $GF(3) = \{0, 1, 2\}$ and $n = 2$. Then,

$$GF(3)^2 = \{00, 10, 20, 01, 21, 12, 22\} \quad (3)$$

3. Experimental Design Models

Experimental design is a research field on deciding how to collect data, and is now widely used in fields of natural sciences including engineering, biology, and medicine [1–7]. In this section, we introduce two formulations of experimental design models.

First, we explain the traditional model, which is expressed through the effect of each factor. This model clarifies how each factor affects the response variable [7]. Because we consider a Bayesian approach in this paper, we need to introduce a prior probability distribution over the model parameters. For the traditional model, it is easy to assume a prior probability distribution. However, this model is not suitable for deriving analytical properties of a Bayesian framework because it is not a linear combination of basis functions.

In contrast, models based on an orthonormal system are suitable for deriving analytical properties of a Bayesian framework. However, it is not possible to assume a prior probability distribution for models based on an orthonormal system because the models are expressed by using complex Fourier coefficients.

Hence, the ability to convert the former model to the latter, and vice versa, is desirable in a Bayesian experimental design framework [15]. We will show the relationship between the two formulations of experimental design models.

3.1 Notations for experimental design

Let $F_1, F_2, \ldots, F_n$ denote $n$ factors to be included in an experiment. Let each factor have $q$ levels, where $q$ is a prime power. The levels of each factor can be represented by $GF(q)$, and the combinations of levels can be represented by the $n$-tuples $x = (x_1, x_2, \ldots, x_n) \in GF(q)^n$. Let $t(x)$ denote the response of the experiment with level combination $x$.

3.2 Traditional experimental design model

In this paper, we are mostly concerned with second-order models, because in many cases the relationship between the response and the factors can be represented sufficiently with models of order no greater than 2 [1, 6].

We choose as a model the second-order model with $n$ factors*1.

$$t(x) = \mu + \sum_{i=1}^{n} \alpha_i(x_i) + \sum_{i=1}^{n-1} \sum_{m=i+1}^{n} \beta_{im}(x_i, x_m) + \epsilon \quad (4)$$

where $\mu$ is the general mean, $\alpha_i(x_i)$ is the effect of the $x_i$th level of factor $F_i$, $\beta_{im}(x_i, x_m)$ is the effect of the interaction of the $x_i$th level of factor $F_i$ and the $x_m$th level of factor $F_m$, and $\epsilon$ is a zero-mean Gaussian random variable with precision $\beta$, where the precision is the reciprocal of the variance.

Here, the constraints [7]

$$\sum_{\psi=0}^{q-1} \alpha_{\psi} = 0 \quad (5)$$

$$\sum_{\psi=0}^{q-1} \beta_{\psi, m} = 0 \quad (6)$$

$$\sum_{\psi=0}^{q-1} \beta_{\psi, m} = 0 \quad (7)$$

are generally assumed. These constraints have an average value of zero, and are set to find the relative difference, not an absolute value, of the effect of each level of the factors [6].

Example 2 Consider $n = 2$ and $q = 3$. Then, $\mu, \alpha_1(0), \alpha_1(1), \alpha_1(2), \alpha_2(0), \alpha_2(1), \alpha_2(2), \beta_{1,2}(0,0), \beta_{1,2}(1,0), \beta_{1,2}(2,0), \beta_{1,2}(0,1), \beta_{1,2}(0,2), \beta_{1,2}(1,1), \beta_{1,2}(1,2), \beta_{1,2}(2,1), \beta_{1,2}(2,2)$ are the parameters. The number of parameters is 16, but the number of independent parameters is 9 owing to the constraints in Eqs. (5)-(7).

Under the constraints in Eqs. (5)-(7), let $u \in \mathbb{R}^k$ denote the column vector of the independent parameters, where $k$ is the number of independent parameters.

Example 3 Consider $n = 2$ and $q = 3$. Then, $k = 9$ and

$$u = \begin{bmatrix} \mu \\ \alpha_1(0) \\ \alpha_1(1) \\ \alpha_2(0) \\ \alpha_2(1) \\ \beta_{1,2}(0,0) \\ \beta_{1,2}(1,0) \\ \beta_{1,2}(0,1) \\ \beta_{1,2}(1,1) \end{bmatrix}$$

3.3 Experimental design model based on an orthonormal system

In this section, we consider a model based on an orthonormal system [13]. First, the levels of each factor can be represented by $GF(q)$, which is a Galois field of order $q$, and the level combinations can be represented by the $n$-tuples $x \in GF(q)^n$. Let the Hamming weight $w(a)$ of vector $a = (a_1, a_2, \ldots, a_n)$ be defined as the number of nonzero components.

If we consider the expression based on an orthonormal system, Eq. (4) can be expressed by the following equation [13]:

*1 It is easy to extend the results of this paper to third-order models or higher-order models. In addition, instead of considering the interaction between all pairs of factors, it is also possible to consider the interaction between some pairs of factors.
\( t(x) = \sum_{a \in A} f_a X_a(x) + \epsilon \)  
(8)

where \( A = \{a | a(a) \leq 2, a \in GF(q)^n \} \), \( f_a \) is the \( a \)th Fourier coefficient parameter, \( X_a(x) \) is the \( a \)th basis function (character\(^2\)), and \( \epsilon \) is a zero-mean Gaussian random variable with precision \( \beta \).

The important point here is there are no constraints on the parameters. This means that all parameters are independent.

**Example 4**  Consider \( n = 2 \) and \( q = 3 \). Then, \( f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8 \) are the parameters. The number of parameters is 9, and these parameters are independent. Here, from [16], \( X_a(x) = e^{2\pi i a x^3} \).

Let \( w \in \mathbb{C}^9 \) denote the column vector of the independent parameters, where \( k \) is the number of independent parameters.

\[
\begin{bmatrix}
    f_0 \\
    f_1 \\
    f_2 \\
    f_3 \\
    f_4 \\
    f_5 \\
    f_6 \\
    f_7 \\
    f_8
\end{bmatrix}
\]

\( \Box \)

### 3.4 Relationship between the two models

Now that we have reviewed the two models, we give the relationship between \( u \) in Sect. 3.2 and \( w \) in Sect. 3.3 in this subsection. First, the relationship is already partially provided in [14].

**Theorem 1**  [14]

Let \( \hat{\mu}, \hat{\alpha_i} (\varphi), \) and \( \hat{\beta}_{im} (\varphi, \psi) \) be the unbiased estimator of the general mean \( \mu \), that of \( \alpha_i (\varphi) \) and that of \( \beta_{im} (\varphi, \psi) \) in the model of Sect. 3.2, respectively.

Let \( \hat{f}_{0,0}, \hat{f}_{0,0,0,0,0} \), and \( \hat{f}_{0,0,0,0,0,0,0,0,0} \) be the unbiased estimators of the Fourier coefficients \( f_{0,0}, f_{0,0,0,0,0}, \) and \( f_{0,0,0,0,0,0,0,0,0,0} \) in the model of Sect. 3.3, respectively.

Then, the following equations hold:

\[
\hat{\mu} = \hat{f}_{0,0} 
\]

\[ \hat{\alpha_i} (\varphi) = \sum_{a \in GF(q)} X_{a,0}(\varphi) \hat{f}_{0,0,0,0,0} \]

\[ \hat{\beta}_{im} (\varphi, \psi) = \sum_{a \in GF(q)} \sum_{a \in GF(q)} X_{a,0}(\varphi) X_{a,0}(\psi) \hat{f}_{0,0,0,0,0,0,0,0,0} \]

Using Eqs. (9)-(11), we can construct a \( k \times k \) matrix \( M \) that satisfies

\[
u = Mw \]

where

\[ \Box \]

\[^2\] For the characters, refer to \([13,16]\).
$$R_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \omega_3 & \omega_3^2 & \omega_3 & \omega_3^2 \\ \omega_3 & \omega_3 & \omega_3^2 & \omega_3^2 \\ \omega_3 & 1 & 1 & \omega_3 \end{bmatrix}$$ (16)

and $\omega_3 = e^{-i \pi / 3}$.

Using Eq. (12), we can convert $w$ in the model of Sect. 3.3 to $u$ in the model of Sect. 3.2. On the other hand, however, to convert $u$ to $w$, another theorem for the relationship is required. The theorem is given below.

**Theorem 2**  Regarding the relationship between the Fourier coefficients and the effect of the main factor, the following equation holds:

$$f_{\omega,0\omega,0\ldots 0} = \frac{1}{q} \sum_{q=0}^{\infty} X_q(\varphi) \hat{X}_q(\varphi)$$ (17)

where $X_q(\varphi)$ is the complex conjugate of $X_q(\varphi)$.

Next, for the relationship between the Fourier coefficients and the effect of the interaction, the following equation holds:

$$f_{\omega,0\omega,0\ldots 0\ldots 0} = \frac{1}{q} \sum_{q=0}^{\infty} \sum_{q=0}^{\infty} X_q(\varphi) \bar{X}_q(\varphi) \beta_{\omega,0}(\varphi, \psi)$$ (18)

**Proof of Theorem 2:** See Appendix.

Using Eqs. (9), (17), and (18) and the constraints in Eqs. (5)-(7), we can construct a $k \times k$ matrix $M^{-1}$ that satisfies the following equation:

$$w = M^{-1} u$$ (19)

where

$$M^{-1} = \begin{bmatrix} 1 & S_1 & \cdots & 0 \\ S_1 & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & S_2 \end{bmatrix}$$

$S_1$ is the $(q-1) \times (q-1)$ matrix given by

$$S_1 = \frac{1}{q} \begin{bmatrix} X_1'(0)X_1'(0) & X_1'(q-1)X_1'(q-1) \\ X_2'(0)X_2'(0) & X_2'(q-1)X_2'(q-1) \\ \cdots & \cdots \\ X_{q-1}'(0)X_{q-1}'(0) & X_{q-1}'(q-1)X_{q-1}'(q-1) \\ \cdots & \cdots \\ X_1'(q-2)X_1'(q-2) & X_1'(q-1)X_1'(q-1) \\ \cdots & \cdots \\ \cdots & \cdots \\ X_{q-1}'(q-2)X_{q-1}'(q-2) & X_{q-1}'(q-1)X_{q-1}'(q-1) \end{bmatrix}$$

and $S_2$ is the $(q-1)^2 \times (q-1)^2$ matrix given by

$$S_2 = \frac{1}{q^2} \begin{bmatrix} X_1'(0)X_1'(0) - X_1'(q-1)X_1'(q-1) \\ X_2'(0)X_2'(0) - X_2'(q-1)X_2'(q-1) \\ \cdots \\ \cdots \\ X_{q-1}'(0)X_{q-1}'(0) - X_{q-1}'(q-1)X_{q-1}'(q-1) \\ \cdots \\ X_1'(q-2)X_1'(q-2) - X_1'(q-1)X_1'(q-1) \\ \cdots \\ \cdots \\ X_{q-1}'(q-2)X_{q-1}'(q-2) - X_{q-1}'(q-1)X_{q-1}'(q-1) \end{bmatrix}$$

Using Eqs. (19) and (20),

$$[f_{00} \ f_{10} \ f_{20} \ f_{01} \ f_{02} \ f_{11} \ f_{12} \ f_{21} \ f_{22}] = \begin{bmatrix} \mu \\ \alpha_1(0) \\ \alpha_2(0) \\ \alpha_2(1) \\ \beta_{1,2}(0,0) \\ \beta_{1,2}(1,0) \\ \beta_{1,2}(0,1) \\ \beta_{1,2}(1,1) \end{bmatrix}$$ (21)

and

$$M^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & S_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & S_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & S_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & S_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_2 \end{bmatrix}$$

where the $2 \times 2$ matrix $S_1$ and the $4 \times 4$ matrix $S_2$ are given by

$$S_1 = \frac{1}{3} \begin{bmatrix} 1 - \omega_1 & \omega_3^2 & \omega_3 & \omega_3^2 \\ 1 - \omega_3 & 1 & \omega_3 & \omega_3^2 \\ \omega_3 & \omega_3 & 1 & \omega_3 \\ \omega_3^2 & \omega_3 & \omega_3^2 & 1 \end{bmatrix}$$ (22)

$$S_2 = \frac{1}{3} \begin{bmatrix} -\omega_1 & \omega^2_3 & \omega_3 & \omega_3 \\ 1 & -\omega_1 & \omega_3 & \omega_3 \\ -\omega_3 & \omega_3 & 1 & \omega_3 \\ -\omega_3^2 & \omega_3 & \omega_3 & 1 \end{bmatrix}$$ (23)

Using Eqs. (12) and (19), the traditional model can be converted to the model based on an orthonormal system, and vice versa. This conversion is desirable for deriving analytical properties in a Bayesian experimental design framework.

4. **Bayesian Experimental Design**

In Sect. 3, it is shown that models expressed through the effect of each factor can be converted to those based on an orthonormal system. As the model based on an orthonormal system is
a linear combination of basis functions, we can apply some analytical properties derived by Bishop [12] to this model. This section uses the results of [12, Sect. 3.3] to show that the posterior distribution can also be analytically derived in the Bayesian experimental framework.

We use the following notation.
- The data set of inputs is \(X = \{x_1, \ldots, x_N\}\) with corresponding target values \(t(x_1), \ldots, t(x_N)\).
- The variables \(t(x_1), \ldots, t(x_N)\) are represented by a column vector denoted by \(t\).
- \(K\) is the number of parameters.

### 4.1 Likelihood function

As explained in Sect. 3.3, it is assumed that the target variable \(t(x)\) is given by a deterministic function with additive noise, which is a zero-mean Gaussian random variable with precision \(\beta\). Hence, using Eqs. (8) and (19), the likelihood function is

\[
p(t|X, u, \beta) = \prod_{i=1}^{N} \mathcal{N}(t(x_i)|\phi(x_i)^T M^{-1} u, \beta^{-1})
\]

(24)

\[
= \mathcal{N}(t|\Phi M^{-1} u, \beta^{-1} I)
\]

(25)

where

\[
\phi(x)^T = \begin{bmatrix} X_{a_1}(x) & X_{a_2}(x) & \ldots & X_{a_K}(x) \end{bmatrix}
\]

(26)

and

\[
\Phi = \begin{bmatrix}
X_{a_1}(x_1) & X_{a_2}(x_1) & \ldots & X_{a_K}(x_1) \\
X_{a_1}(x_2) & X_{a_2}(x_2) & \ldots & X_{a_K}(x_2) \\
\vdots & \vdots & \ddots & \vdots \\
X_{a_1}(x_N) & X_{a_2}(x_N) & \ldots & X_{a_K}(x_N)
\end{bmatrix}
\]

(27)

### 4.2 Prior distribution

The same as in [12], the corresponding conjugate prior is given by a Gaussian distribution of the form

\[
p(u) = \mathcal{N}(u|m_0, S_0)
\]

(28)

### 4.3 Posterior distribution

Then the posterior probability is given by

\[
p(u|X, t, \beta) = \mathcal{N}(u|m_N, S_N)
\]

(29)

where

\[
m_N = S_N \left(\beta(M^{-1})^T \Phi^T t + S_0^{-1} m_0\right)
\]

(30)

\[
S_N^{-1} = \beta(M^{-1})^T \Phi^T \Phi M^{-1} + S_0^{-1}
\]

(31)

and \(T\) denotes the conjugate transpose (Hermitian transpose).

Except for the inclusion of \(M\), the proof is the same as that in [12].

### 4.4 Bayesian optimality

There are many criteria for optimal design, such as A-optimality, and D-optimality [3]. For example, for A-optimality in this linear model, we have to find the design that minimizes the trace of the posterior variance. However, it is generally not easy to calculate the posterior variance for all designs.

Here, we focus on a subclass of designs for which \(\Phi^T \Phi = N I\) holds. The designs are called orthogonal designs [13] (equivalently, orthogonal arrays [4]). If we use orthogonal designs, the posterior variance is directly given by the following equation.

\[
S_N = \left(\beta N(M^{-1})^T M^{-1} + S_0^{-1}\right)^{-1}
\]

(32)

### 5. Evidence Approximation

We would like to make fully Bayesian predictions by marginalizing with respect to \(u, \alpha, \text{ and } \beta\) by introducing prior distributions over the hyperparameters \(\alpha\) and \(\beta\). However, complete marginalization over all of these variables is analytically intractable [8,12].

This section uses the results of [12, Sects. 3.5 and 7.2] to show that we can make an approximation in which we set the hyperparameters to specific values determined by maximizing the evidence function obtained by first integrating over \(u\) in the Bayesian experimental framework.

In this section, the same as in [12, Sect. 7.2.1], we assume a prior distribution over \(u\) of the form

\[
p(u|\alpha) = \prod_{i=1}^{K} \mathcal{N}(u_i|0, \alpha_i^{-1})
\]

(33)

where \(\alpha_i\) represents the precision of the corresponding parameter \(u_i\) and \(\alpha\) denotes \((\alpha_1, \ldots, \alpha_K)^T\). Let \(A\) denote the precision matrix, that is, \(A = \text{diag}(\alpha_i)\).

#### 5.1 Integrating over the weight parameters \(u\)

The evidence function \(p(t|X, \alpha, \beta)\) is obtained by integrating over the weight parameters \(u\). Using Eqs. (24) and (33), the evidence function is given by

\[
p(t|X, \alpha, \beta) = \int p(t|X, u, \beta) p(u|\alpha) du
\]

\[
= \int \prod_{i=1}^{N} \frac{1}{(2\pi\beta^{-1})^{1/2}} \cdot \exp\left\{-\beta(t(x_i) - \phi(x_i)^T M^{-1} u)^2\right\}
\]

\[
\cdot \prod_{i=1}^{K} \frac{1}{(2\pi\alpha_i^{-1})^{1/2}} \exp\left\{-\frac{\alpha_i}{2} u_i^2\right\} du
\]

\[
= \frac{1}{(2\pi\beta^{-1})^{N/2}} \prod_{i=1}^{K} (2\pi\alpha_i^{-1})^{1/2} \int \exp\left\{-\frac{\beta}{2} \|t - \Phi M^{-1} u\|^2 - \frac{1}{2} u^T Au\right\} du
\]

(34)
\[ E(u) = \frac{\beta}{2} \| t - \Phi M^{-1} u \|^2 + \frac{1}{2} u^T A u \]  

Here, \( E(u) \) can be written as

\[ E(u) = \frac{\beta}{2} (t - \Phi M^{-1} u)^T (t - \Phi M^{-1} u) + \frac{1}{2} u^T A u \]

\[ = \frac{1}{2} (\beta t^T - 2\beta u^T \Phi M^{-1} u + \beta u^T u) + \frac{1}{2} u^T A u \]

\[ = \frac{1}{2} (\beta t^T - 2\beta u^T \Phi M^{-1} u + u^T (\beta (M^{-1})^T \Phi^T \Phi M^{-1} + A) u) \]

\[ = \frac{1}{2} (\beta t^T - 2\beta u^T \Phi M^{-1} u + u^T Bu) \]  

(36)

where

\[ B = \beta (M^{-1})^T \Phi^T \Phi M^{-1} + A \]  

(37)

and \( B \) represents the precision matrix of the posterior distribution.

In Eq. (36), by completing the square with respect to \( u \), we have

\[ = \frac{1}{2} (u - m_n)^T B (u - m_n) + E(m_n) \]  

(38)

where

\[ m_n = \beta (M^{-1})^T \Phi^T t \]

\[ = \beta B^{-1} (M^{-1})^T \Phi^T t \]  

(39)

and

\[ E(m_n) = \frac{1}{2} (\beta t^T - m_n^T B m_n) \]  

(40)

Then, from Eq. (38),

\[ \int \exp \{-E(u)\} \, du \]

\[ = \int \exp \left\{ -\frac{1}{2} (u - m_n)^T B (u - m_n) \right\} \exp \{-E(m_n)\} \, du \]

\[ = (2\pi|B|^{1/2})^{-1/2} \exp \{-E(m_n)\} \]  

(41)

Using Eqs. (34) and (41), we can write the log of the evidence function in the form

\[ \ln p(U, X, \alpha, \beta) \]

\[ = \frac{N}{2} \ln \beta - \frac{N}{2} \ln 2\pi + \frac{1}{2} \sum_{i=1}^k \ln \alpha_i \]

\[ + \frac{1}{2} \ln |B^{-1}| - E(m_n) \]  

(42)

5.2 Maximizing the evidence function with respect to \( \alpha_i \)

First we consider the derivative of \( E(m_n) \) with respect to \( \alpha_i \). From Eq. (40), \( E(m_n) \) can be written as follows:

\[ E(m_n) = \frac{1}{2} (\beta t^T - m_n^T B m_n) \]

\[ = \frac{1}{2} (\beta t^T - 2m_n^T B m_n + m_n^T B m_n) \]

\[ = \frac{1}{2} (\beta t^T - 2m_n^T B B^{-1}(M^{-1})^T \Phi \beta + m_n^T (\beta (M^{-1})^T \Phi^T \Phi M^{-1} + A)m_n) \]

\[ = \frac{1}{2} (\beta t^T - 2m_n^T (M^{-1})^T \Phi \beta + m_n^T (\beta (M^{-1})^T \Phi^T \Phi M^{-1} + A)m_n) \]

\[ = \frac{1}{2} (\beta t^T - M^{-1} m_n)^T (t - \Phi M^{-1} m_n) + m_n^T A m_n \]  

(43)

From Eq. (43), we have

\[ \frac{\partial}{\partial \alpha_i} E(m_n) = \frac{1}{2} m_i^2 \]  

(44)

Next, consider the derivative of \( \ln |B^{-1}| \) with respect to \( \alpha_i \),

\[ \frac{\partial}{\partial \alpha_i} \ln |B^{-1}| \]

\[ = Tr \left( B \frac{\partial B^{-1}}{\partial \alpha_i} \right) \]

\[ = Tr \left( B \left( -B^{-1} \frac{\partial B}{\partial \alpha_i} B^{-1} \right) \right) \]

\[ = -Tr \left( \frac{\partial B}{\partial \alpha_i} B^{-1} \right) \]

\[ = -B_{ii}^{-1} \]  

(45)

in which \( B_{ii}^{-1} \) is the \( i \)th diagonal component of the matrix \( B^{-1} \).

From Eqs. (42), (44), and (45), we can write the derivative of the log of the evidence function in the form

\[ \frac{\partial}{\partial \alpha_i} \ln p(U, X, \alpha, \beta) = \frac{1}{2} \left( \frac{\gamma_i}{m_i^2} - 1 - \frac{1}{2} m_i^2 \right) \]

(46)

We set Eq. (46) to zero and obtain the following re-estimation equation:

\[ \alpha_i = \frac{\gamma_i}{m_i^2} \]  

(47)

where \( \gamma_i \) is defined as \( \gamma_i = 1 - \alpha_i B_{ii}^{-1} \).

5.3 Maximizing the evidence function with respect to \( \beta \)

First, from Eq. (43), we have

\[ \frac{\partial}{\partial \beta} E(m_n) = \frac{1}{2} (t - \Phi M^{-1} m_n)^T \]

(48)

Next, consider the derivative of \( \ln |B^{-1}| \) with respect to \( \beta \).
\[
\frac{\partial}{\partial \beta} \ln |B^{-1}| = Tr \left( B \frac{\partial B^{-1}}{\partial \beta} \right) = Tr \left( B \left( -B^{-1} \frac{\partial B}{\partial \beta} B^{-1} \right) \right) = -Tr \left( \frac{\partial B}{\partial \beta} B^{-1} \right) = -Tr \left( (M^{-1})^T \Phi^T \Phi M^{-1} B^{-1} \right) \tag{49}
\]

Here, the following equation holds.

\[
(M^{-1})^T \Phi^T \Phi M^{-1} B^{-1} = (M^{-1})^T \Phi^T \Phi M^{-1} B^{-1} + \beta^{-1} AB^{-1} - \beta^{-1} AB^{-1} = \beta^{-1} (\beta(M^{-1})^T \Phi^T \Phi M^{-1} + A)B^{-1} - \beta^{-1} AB^{-1} = \beta^{-1} B(B^{-1} - AB^{-1}) = \beta^{-1} (I - AB^{-1}) \tag{50}
\]

Using Eqs. (49) and (50), we have

\[
\frac{\partial}{\partial \beta} \ln |B^{-1}| = -\beta^{-1} Tr(I - AB^{-1}) = -\beta^{-1} \sum_i \left( 1 - \alpha_i B^{-1}_{ii} \right) = -\beta^{-1} \sum_i \gamma_i \tag{51}
\]

From Eqs. (42), (48), and (51), we can write the derivative of the log of the evidence function in the form

\[
\frac{\partial}{\partial \beta} \ln p(t|X, \alpha, \beta) = \frac{1}{2} \left( \beta^{-1} N - \beta^{-1} \sum_i \gamma_i - ||t - \Phi M^{-1} m_0||^2 \right) \tag{52}
\]

We set Eq. (52) to zero and obtain the following re-estimation equation:

\[
\beta^{-1} = \frac{||t - \Phi M^{-1} m_0||^2}{N - \sum_i \gamma_i} \tag{53}
\]

Using Eqs. (47) and (53), we can set the hyperparameters to specific values determined by maximizing the evidence function obtained by first integrating over \( u \). For the precise procedure, refer to [12, Sects. 3.5 and 7.2].

From the above, it is clear that we can also make the evidence approximation in the Bayesian experimental framework by using the results of [12]. The difference between the equations in [12] and those in the experimental framework is that the matrix \( M \) is included only in the experimental framework. Note that the matrices \( M, M^{-1}, (M^{-1})^T \) are fixed and can be calculated before the start of the procedure. Therefore, it is expected that the precise procedure of the evidence approximation in [12] can be directly applied to the experimental framework.

### 6. Conclusions

In this paper, we first presented a theorem on the relationship between the traditional model and a model based on an orthonormal system in experimental design. Using the theorem, the former model can be converted to the latter, and vice versa. Next, we introduced prior distributions over the hyperparameters for experimental design models to consider fully Bayesian predictions. Combining the conversion and the result of [12], we showed that we can make an approximation in which we set the hyperparameters to specific values determined by maximizing the evidence function.

### Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17K00316.

### References

[1] A. C. Atkinson and A. N. Donev: Optimum Experimental Designs, Oxford University Press, 1992.

[2] R. A. Bailey: Design of Comparative Experiments, Cambridge University Press, 2008.

[3] K. Chaloner and I. Verdinelli: Bayesian experimental design: a review, Stat. Sci., Vol.10, No.3, pp.273-304, 1995.

[4] A. S. Hedayat, N. J. A. Sloane and J. Stufken: Orthogonal Arrays: Theory and Applications, Springer, 1999.

[5] T. Okuno and T. Haga: Experimental Designs, Baifukan, 1969.

[6] I. Takahashi: Combinatorial Theory and Its Application, Iwanami Syoten, 1979.

[7] H. Toutenburg and Shalabh: Statistical Analysis of Designed Experiments (Third Edition), Springer, 2009.

[8] J. O. Berger: Statistical Decision Theory and Bayesian Analysis, Springer, 1985.

[9] A. Alexanderian, P. J. Gloop and O. Ghattas: On Bayesian A- and D-optimal experimental designs in infinite dimensions, Bayesian Anal., Vol.11, No.3, pp.671-695, 2016.

[10] S. Ji, Y. Xue and L. Carin: Bayesian compressive sensing, IEEE Trans. Signal Process., Vol.56, No.6, pp.2346-2356, 2008.

[11] M. Jones, M. Goldstein, P. Jonathan and D. Randell: Bayes linear analysis for Bayesian optimal experimental design, J. Stat. Plann. Inference, Vol.171, pp.115-129, 2016.

[12] C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006.

[13] Y. Ukita, T. Saito, T. Matsushima and S. Hirasawa: A note on a sampling theorem for functions over \( GF(q^n) \) domain, IEICE Trans. Fundam., Vol.E93-A, No.6, pp.1024-1031, 2010.

[14] Y. Ukita and T. Matsushima: A Note on relation between the Fourier coefficients and the effects in the experimental design, Proc. 9th Int. Conf. Inf., Comm. and Signal Process., (ICICS 2011), pp.1-5, 2011.

[15] Y. Ukita: A note on Bayesian experimental design model based on an orthonormal system, Yokohama Shodai Ronshu, Vol.51, No.1, pp.1-10, 2017.

[16] E. M. Stein and R. Shakarchi: Fourier Analysis: An Introduction, Princeton University Press, 2003.

### Appendix

#### Proof of Theorem 2

The right side of Eq. (17) can be written as follows:

\[ \text{Encryption} \]

\[ \text{Decryption} \]

\[ \text{Key Generation} \]
we obtain Eq. (17). was used for the transformation from Eq. (54) to Eq. (55). Hence, we obtain Eq. (18).

Next, the right side of Eq. (18) can be written as follows:

\[
\frac{1}{q^2} \sum_{q=0}^{q-1} \sum_{\psi=0}^{q-1} X_n^*(\varphi) X_m(\psi) \hat{f}_{0, 0}(\varphi, \psi)
\]

\[
= \frac{1}{q^2} \sum_{q=0}^{q-1} X_n^*(\varphi) X_m(\psi) \hat{f}_{0, 0}(\varphi, \psi)
\]

\[
\times \left\{ \left[ \sum_{\frac{q'}{q} \in \mathbb{GF}(q)} \sum_{\frac{\psi'}{\psi} \in \mathbb{GF}(\psi)} X_n^*(\varphi) X_m(\psi) \hat{f}_{0, 0}(\varphi, \psi) \right] \right\}
\]

\[
= \sum_{\frac{q'}{q} \in \mathbb{GF}(q)} \sum_{\frac{\psi'}{\psi} \in \mathbb{GF}(\psi)} \left[ \frac{1}{q^2} \sum_{q=0}^{q-1} \sum_{\psi=0}^{q-1} X_n^*(\varphi) X_m(\psi) \hat{f}_{0, 0}(\varphi, \psi) \right]
\]

\[
\times \hat{f}_{0, 0}(\varphi', \psi') \quad (57)
\]

\[
= \hat{f}_{0, 0}(\varphi', \psi') \quad (58)
\]

was used for the transformation from Eq. (57) to Eq. (58). Hence, we obtain Eq. (18).

Yoshifumi Ukita received his B.E. degree and M.E. degree in industrial and management systems engineering from Waseda University, Tokyo, Japan, in 1994 and 1996, respectively, and his Ph.D. degree in applied mathematics from Waseda University in 2007. From 1998 to 2000, he was a Research Associate in industrial and management systems engineering at Waseda University. He is currently a Professor at the Department of Management and Information, Yokohama College of Commerce, Kanagawa, Japan. From January to April 2012, he was an Exchange Professor of the University of Pittsburgh at Bradford. From August 2013 to August 2014, he was a Visiting Fellow at the University of Cambridge. His research interests include experimental designs, signal processing, and artificial intelligence. He is a member of the IEICE, IPSJ, JSAI, and IEEE.

(Received May 14, 2018; revised August 25, 2018)