



# Bayesian $A$ -optimal two-phase designs with a single blocking factor in each phase

Nha Vo-Thanh<sup>1</sup> · Hans-Peter Piepho<sup>1</sup>

Received: 22 October 2021 / Accepted: 1 July 2022 / Published online: 01 December 2022  
© The Author(s) 2022

## Abstract

Two-phase experiments are widely used in many areas of science (e.g., agriculture, industrial engineering, food processing, etc.). For example, consider a two-phase experiment in plant breeding. Often, the first phase of this experiment is run in a field involving several blocks. The samples obtained from the first phase are then analyzed in several machines (or days, etc.) in a laboratory in the second phase. There might be field-block-to-field-block and machine-to-machine (or day-to-day, etc.) variation. Thus, it is practical to consider these sources of variation as blocking factors. Clearly, there are two possible strategies to analyze this kind of two-phase experiment, i.e., blocks are treated as fixed or random. While there are a few studies regarding fixed block effects, there are still a limited number of studies with random block effects and when information of block effects is uncertain. Hence, it is beneficial to consider a Bayesian approach to design for such an experiment, which is the main goal of this work. In this paper, we construct a design for a two-phase experiment that has a single treatment factor, a single blocking factor in each phase, and a response that can only be observed in the second phase.

**Keywords**  $A$ -optimality · Bayesian optimal designs · Bayesian two-phase designs · Efficiency factors · Hill climbing · Quadrature rules · Linear mixed models

## 1 Introduction

Several kinds of experiments in agriculture, industry, food science, or bio-engineering are performed in two phases. For example, consider an experiment in plant breeding where the first phase is run in a field and the second phase is run in a laboratory. The samples obtained from the field are analyzed in the laboratory for further evaluation. Commonly, the first phase is done in multiple field blocks, while the second phase is performed with multiple machines or is performed in several days. In this example, it can be seen that the effects of interest (e.g., varieties or treatments) cannot be assessed in the first phase, but must be evaluated in the second phase. We call designs for such experiments two-phase designs (Brien 1983; Curnow 1959; McIntyre 1955; Wood et al. 1988). Furthermore, we expect that there is field-block-to-field-block

and machine-to-machine (or day-to-day, etc.) variation. In experimental jargon, these sources of variation are called blocking factors and the factors of interest are called treatment factors. In this example, each phase has its own block structure and these must be combined to form the proper analysis. Therefore, optimal two-phase designs must take the treatment factor(s) and the two blocking factors into account simultaneously.

It is well known that optimal designs depend on the statistical linear (or nonlinear) model envisioned for analysis (Atkinson et al. 2007). In the example above, two possible analytical strategies could be considered. With both strategies, the effects of interest are commonly considered as fixed, while different assumptions are made for the effects of the two blocking factors. In particular, the first one assumes the effects of blocks in both phases as fixed. This model is referred to as a linear fixed-effects model. The second one assumes the effects of blocks in both phases as random. This model is referred to as a linear mixed model. Under this model, the variances of random block effects and random residual errors are called variance components (Searle et al. 2009). In plant breeding, the second analytical strategy is commonly used because (i) the numbers of blocks in both

✉ Nha Vo-Thanh  
nha.vo-thanh@uni-hohenheim.de

Hans-Peter Piepho  
hans-peter.piepho@uni-hohenheim.de

<sup>1</sup> Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

phases might be large enough for the block variances to be estimated with sufficient accuracy, and (ii) block effects in both phases might be small. Hence, there may be substantial gains to be achieved by recovering the inter-block information (Yates 1939). The second strategy will be our focus in this paper.

There are several types of two-phase experiments in the current literature. For example, Jarrett and Ruggiero (2008) considered an application of two-phase experiments in gene expression microarrays, while Brien (2017) showed several applications of two-phase experiments, e.g., in plant pathology, greenhouse chamber experimentation, product storage, sensory evaluation, and laboratory analysis of crops. These applications involve single or multiple treatment factors, and single or multiple blocking factors in each phase. We refer to these two papers for those who are unfamiliar with applications of two-phase experiments in practice. In this paper, we limit our attention to a type of two-phase experiment where there is only a single treatment factor and a single blocking factor in each phase, which will be defined in more detail in Sect. 2.

In the current literature, there are only a few studies regarding constructing optimal two-phase designs under a linear fixed-effects model and a linear mixed model (Smith et al. 2006, 2015; Brien 2019; Vo-Thanh et al. 2022). When constructing optimal two-phase designs under a linear mixed model, we are often required to specify values of variance components, see Smith et al. (2006). Thus two-phase designs found under the linear mixed model are then optimal for these given values of variance components. In practice, however, these values are rarely known exactly before analyzing the data collected from an experiment, even though it may be possible to state a plausible range of values before an experiment. Furthermore, optimal designs found at given values of variance components might not be optimal at other values of variance components (Shah and Sinha 1989, Chapter 5). Hence, it is practical to construct optimal two-phase designs for a range of plausible values of variance components, referred to as *robust two-phase designs*. This task will be the focus of our work. To the best of our knowledge, there are no studies regarding finding such robust two-phase designs.

Two-phase designs often have complex blocking structures (e.g., non-orthogonal) when the block sizes in both phases are different. This will make the problem of searching for optimal two-phase designs more challenging. Therefore, it is advantageous to identify a suitable combined blocking pattern in both phases. Then the question is how to find a good allocation of treatments to this pattern. Vo-Thanh et al. (2022) studied extensively a nested row-column structure and showed that this pattern is indeed good for two-phase designs under a linear fixed-effects model. They have not, however,

considered this approach for constructing two-phase designs under a linear mixed model.

Studying robust two-phase designs requires a prior knowledge of variance components. The prior knowledge could be based on a well-educated guess or historical data of two-phase experiments. Here we assume that we do not have specific values for the variance components and we want to construct robust two-phase designs from scratch. Therefore, we consider various prior distributions for the variance components (Mylona et al. 2014; Goos and Mylona 2018), implementing these in a Bayesian framework (see a detailed review in Chaloner and Verdinelli (1995)). One of the major challenges with the Bayesian approach is that it is computationally expensive to get a good approximation for a Bayesian optimality criterion. Often, evaluating this requires integral techniques. Several works have utilized a quadrature approach to approximate the Bayesian criterion for studying several design problems such as designs with nonlinear models, and blocked and split-plot designs with linear mixed models (Gotwalt 2010; Gotwalt et al. 2009; Mylona et al. 2014). In this paper, we will use this approach to get a good approximation for the Bayesian criterion.

Several issues are studied in this paper. First, we propose a simple expression for computing the treatment information matrix for two-phase designs under a linear mixed model and a nested row-column pattern. Second, we derive an update formula to compute a generalized inverse of the treatment information matrix when finding optimal two-phase designs. Third, an overall efficiency factor and a Bayesian criterion are proposed to rank a two-phase design for given values of variance components and for a range of plausible values of variance components, respectively. From the literature, we will study three prior distributions, i.e., uniform, log-normal, and half-Cauchy distributions for variance components. We propose an approximation to compute the Bayesian criterion which is based on quadrature rules. We compare our proposed approximation with existing methods used for computing multidimensional integrals. We utilize the search framework proposed in Vo-Thanh and Piepho (2022) to find Bayesian optimal two-phase designs. We implement our work in a package for given values of variance components as well as for a range of plausible values of variance components with customized weights assigned for given values of the variance components. Finally, we derive two useful theorems when considering two-phase designs that can be constructed from balanced incomplete block designs.

Section 2 reviews a linear mixed model and proposes a simple expression for the treatment information matrix of two-phase designs under this model and the nested row-column pattern. Section 3.1 proposes an overall efficiency factor to rank a two-phase design for given values of variance components. This is then extended to a Bayesian criterion which ranks two-phase designs for a range of plausible values

of the variance components (Sect. 3.2). Section 3.3 considers several choices of prior distribution. Section 4 proposes several approximations for the Bayesian criterion computed under the three prior distributions mentioned above. Section 5 evaluates our proposed approximation and compares this with existing methods. In Sect. 6, we briefly describe a general search algorithm to find optimal Bayesian two-phase designs. In Sect. 7, we illustrate the search algorithm for various examples. Section 8 considers two-phase designs that can be found from balanced incomplete block designs and derives two useful theorems under this case. Section 9 concludes the paper along with suggestions for potential extensions of our work.

## 2 Linear model and information matrix

Suppose that we wish to arrange  $v$  treatments with  $r$  replications in  $b_1$  blocks of size  $k_1$  in the first phase (Phase 1). In the second phase (Phase 2), the  $n = vr$  samples are then arranged in  $b_2$  blocks of size  $k_2$ . The major challenge is to find a good allocation of treatments to blocks in both phases. In the following sections, we consider a linear model and an information matrix for this design problem.

### 2.1 Linear model

The linear model for our design problem is written as:

$$y = \mathbf{1}_n \mu + \mathbf{X} \boldsymbol{\tau} + \mathbf{Z}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{1}$$

where  $\mathbf{y}$  is an  $n \times 1$  response vector,  $\mathbf{1}_n$  is an  $n \times 1$  vector of ones,  $\mu$  is a general mean,  $\mathbf{X}$  is an  $n \times v$  design matrix for treatment effects,  $\boldsymbol{\tau}$  is a  $v \times 1$  vector of unknown treatment effects,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are  $n \times b_1$  and  $n \times b_2$  design matrices for block effects in both phases,  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are  $b_1 \times 1$  and  $b_2 \times 1$  vectors of unknown block effects in both phases, and  $\boldsymbol{\epsilon}$  is an  $n \times 1$  vector of unknown random residual errors. We assume that the expected value of  $\boldsymbol{\epsilon}$  is zero and the covariance matrix of  $\boldsymbol{\epsilon}$  is  $\text{var}(\boldsymbol{\epsilon}) = \sigma_0^2 \mathbf{I}_n$ , where  $\sigma_0^2$  is a constant variance of responses and  $\mathbf{I}_n$  is an identity matrix of size  $n$ . In the rest of this paper, we will assume that the effects of blocks in both phases are treated as random, while the treatment effects are treated as fixed. Under this assumption, model (1) is also called a mixed model (Searle et al. 2009, Chapter 1).

### 2.2 Information matrix

Let  $\text{var}(\boldsymbol{\beta}_1) = \sigma_1^2 \mathbf{I}_{b_1}$  and  $\text{var}(\boldsymbol{\beta}_2) = \sigma_2^2 \mathbf{I}_{b_2}$ . The treatment information matrix for model (1) can be written as:

$$\mathbf{C}_{\gamma_1 \gamma_2} = \mathbf{X}^\top \mathbf{P} \mathbf{X}, \tag{2}$$

where  $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{V}^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \mathbf{V}^{-1}$ ,  $\mathbf{V} = \gamma_1 \mathbf{Z}_1 \mathbf{Z}_1^\top + \gamma_2 \mathbf{Z}_2 \mathbf{Z}_2^\top + \mathbf{I}_n$ ,  $\gamma_1 = \sigma_1^2 / \sigma_0^2$ ,  $\gamma_2 = \sigma_2^2 / \sigma_0^2$ , and  $\gamma_1, \gamma_2 \in [0; +\infty)$ .

As mentioned earlier, Vo-Thanh et al. (2022) studied two-phase designs with several blocking patterns under the linear fixed-effects model. They showed that in a few special cases, two-phase designs obtained under a nested row-column pattern are indeed optimal. This still holds for designs obtained under the linear mixed model considered in this paper, where it is indeed sufficient to use a block design in one of the two phases to construct a two-phase design (e.g., see Sect. 7 and Brien et al. 2011, Big with Big). In complex cases, designs obtained under different blocking patterns (e.g., non-orthogonal blocking pattern) yield slightly better designs, compared to those obtained with a nested row-column pattern. However, in general, they showed that efficiencies of two-phase designs obtained under a nested row-column pattern do not differ much from those obtained under different blocking patterns. Thus, in this paper, we will limit our attention to two-phase designs under this pattern. This also implies  $\mathbf{C}_{\gamma_1 \gamma_2}$  and the inverse of  $\mathbf{V}$  can be expressed in tidy forms, as will be shown subsequently. Therefore,  $\mathbf{C}_{\gamma_1 \gamma_2}$  can be easily computed.

The nested row-column pattern is described as follows. Let  $p = \text{gcd}(b_1, b_2)$ , and  $q = \text{gcd}(k_1, k_2)$ , where  $\text{gcd}$  is the greatest common divisor. After some algebra, we have  $n = pqb_1 p b_2 p = b_1 k_1 = b_2 k_2$ , where  $b_{1p} = b_1 / p$ , and  $b_{2p} = b_2 / p$ . Moreover, it is elementary to show that  $b_1 = p b_{1p}$ ,  $b_2 = p b_{2p}$ ,  $k_1 = q b_{2p}$ , and  $k_2 = q b_{1p}$ . From these facts, we can use a nested row-column pattern involving  $p$  superblocs (rectangles), each of which has  $b_{1p}$  rows and  $b_{2p}$  columns. Rows in superblocs are blocks in Phase 1, while columns in superblocs are blocks in Phase 2. The intersection between a row and a column contains  $q$  experimental units. This pattern will be illustrated in Sect. 5 (Example I). Under the nested row-column pattern, the design matrices for the block effects in both phases can be represented as:

$$\begin{aligned} \mathbf{Z}_1 &= \mathbf{I}_p \otimes \mathbf{I}_{b_{1p}} \otimes \mathbf{1}_{q b_{2p}} \text{ and} \\ \mathbf{Z}_2 &= \mathbf{I}_p \otimes \mathbf{1}_{b_{1p}} \otimes \mathbf{I}_{b_{2p}} \otimes \mathbf{1}_q, \end{aligned} \tag{3}$$

where  $\otimes$  is a Kronecker product. Thus,  $\mathbf{V}$  can be written as:

$$\begin{aligned} \mathbf{V} &= \mathbf{I}_p \\ &\quad \otimes (\gamma_1 \mathbf{I}_{b_{1p}} \otimes \mathbf{J}_{q b_{2p}} + \gamma_2 \mathbf{J}_{b_{1p}} \otimes \mathbf{I}_{b_{2p}} \otimes \mathbf{J}_q + \mathbf{I}_{q b_{1p} b_{2p}}), \end{aligned} \tag{4}$$

where  $\mathbf{J}_{q b_{2p}}$ ,  $\mathbf{J}_{b_{1p}}$ , and  $\mathbf{J}_q$  are square matrices of ones of sizes  $q b_{2p}$ ,  $b_{1p}$ , and  $q$ , respectively. As shown in Eq. (4),  $\mathbf{V}$  is a block diagonal matrix, where any main diagonal block is a matrix  $\mathbf{V}_0 = \gamma_1 \mathbf{I}_{b_{1p}} \otimes \mathbf{J}_{q b_{2p}} + \gamma_2 \mathbf{J}_{b_{1p}} \otimes \mathbf{I}_{b_{2p}} \otimes \mathbf{J}_q + \mathbf{I}_{q b_{1p} b_{2p}}$ . Thus, to find the inverse of  $\mathbf{V}$ , it is sufficient to find the inverse of  $\mathbf{V}_0$ , which can be expressed according to the following proposition.

**Proposition 1** Given a matrix  $\mathbf{V}_0 = \gamma_1 \mathbf{I}_{b_{1p}} \otimes \mathbf{J}_{qb_{2p}} + \gamma_2 \mathbf{J}_{b_{1p}} \otimes \mathbf{I}_{b_{2p}} \otimes \mathbf{J}_q + \mathbf{I}_{qb_{1p}b_{2p}}$ , there are two useful results as shown below:

(i) The inverse of this matrix is

$$\mathbf{V}_0^{-1} = -\varphi_1 \mathbf{I}_{b_{1p}} \otimes \mathbf{J}_{qb_{2p}} - \varphi_2 \mathbf{J}_{b_{1p}} \otimes \mathbf{I}_{b_{2p}} \otimes \mathbf{J}_q + \varphi_3 \mathbf{J}_{qb_{1p}b_{2p}} + \varphi_4 \mathbf{I}_{qb_{1p}b_{2p}}, \tag{5}$$

where  $\varphi_1 = \frac{\gamma_1}{1+q\gamma_1b_{2p}}$ ,  $\varphi_2 = \frac{\gamma_2}{1+q\gamma_2b_{1p}}$ ,  $\varphi_3 = \frac{q\gamma_1\gamma_2(2+q\gamma_2b_{1p})}{(1+q\gamma_2b_{1p})(1+q\gamma_1b_{2p})(1+qb_{1p}\gamma_2+qb_{2p}\gamma_1)}$ , and  $\varphi_4 = 1$ .

(ii) Furthermore,

$$\mathbf{V}^{-1} \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{V}^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \mathbf{V}^{-1} = \varphi_5 \mathbf{J}_v, \tag{6}$$

where  $\varphi_5 = \frac{-\varphi_1qb_{2p}-\varphi_2qb_{1p}+\varphi_3qb_{1p}b_{2p}+\varphi_4}{n}$ .

The proof of this proposition can be found in Supplementary Section A. From this proposition,  $\mathbf{C}_{\gamma_1\gamma_2}$  can be written as:

$$\mathbf{C}_{\gamma_1\gamma_2} = -\varphi_1 \mathbf{N}_1 \mathbf{N}_1^\top - \varphi_2 \mathbf{N}_2 \mathbf{N}_2^\top + \varphi_3 \mathbf{X}^\top \mathbf{I}_p \otimes \mathbf{J}_{qb_{1p}b_{2p}} \mathbf{X} + \varphi_4 \mathbf{X}^\top \mathbf{X} - \varphi_5 r^2 \mathbf{J}_v, \tag{7}$$

where  $\mathbf{N}_1 = \mathbf{X}^\top \mathbf{Z}_1$  and  $\mathbf{N}_2 = \mathbf{X}^\top \mathbf{Z}_2$  are the treatment-block incidence matrices in Phases 1 and 2, respectively, and  $\varphi_i$  are identified in Proposition 1 ( $i = 1, 2, \dots, 5$ ).

### 3 Optimality criteria and choice of prior distribution

#### 3.1 Overall efficiency factor

In the theory of optimal designs, it is well known that designs can be ranked by the *A*-, *D*-, *G*-, *I*-, *C*- or *E*-optimality criterion (Atkinson et al. 2007). The choice of a criterion depends on the goal of experiments. For instance, if we are interested in precise estimations of the model parameters, both *D*- and *A*-optimality criteria are good choices. If we are interested in precise predictions of the response, both *I*- and *G*-optimality criteria are good choices. These aforementioned criteria are prevalent and are implemented in many software packages (e.g., JMP, SAS PROC OPTEX). When we are interested in treatment comparisons and a fixed-effects model is used, there is another criterion, called an *overall efficiency factor* (Shah and Sinha 1989, Chapter 3). It is computed based on the average variance of the estimated difference between pairs of treatments. It is highly preferable to rank designs for this purpose and has a connection with *A*-optimality criterion (Shah and Sinha 1989, Section 3.3 of Chapter 3). Shah

and Sinha (1989, Chapter 5) extended this criterion to rank block designs and row-column designs under a linear mixed model. There is, however, little theoretical work on this criterion under a linear mixed model. In this paper, we first utilize this criterion to rank two-phase designs for given values of variance components  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ . Second, we will extend this criterion to cover the case in which we have prior information on  $\boldsymbol{\gamma}$ . We limit our attention to two-phase designs for which  $\text{rank}(\mathbf{C}_\boldsymbol{\gamma}) = v - 1$  so that all treatment comparisons are estimable.

Given a value of  $\boldsymbol{\gamma}$ , the average variance of pairwise treatment comparisons  $\bar{V}_\boldsymbol{\gamma}$  is computed as:

$$\begin{aligned} \bar{V}_\boldsymbol{\gamma} &= \frac{2\sigma_0^2}{v(v-1)} (v \text{tr}(\boldsymbol{\Omega}_\boldsymbol{\gamma}) - \mathbf{1}_n^\top \boldsymbol{\Omega}_\boldsymbol{\gamma} \mathbf{1}_n) \\ &= \frac{2\sigma_0^2}{v-1} \text{tr}(\boldsymbol{\Omega}_\boldsymbol{\gamma}), \end{aligned} \tag{8}$$

where  $\boldsymbol{\Omega}_\boldsymbol{\gamma}$  is the Moore–Penrose inverse of  $\mathbf{C}_\boldsymbol{\gamma}$ , and  $\mathbf{1}_n^\top \boldsymbol{\Omega}_\boldsymbol{\gamma} \mathbf{1}_n = 0$  since  $\mathbf{1}_n^\top \mathbf{C}_\boldsymbol{\gamma} \mathbf{1}_n = 0$ . We rank designs according to

$$A_\boldsymbol{\gamma} = \frac{2\sigma_0^2}{r\bar{V}_\boldsymbol{\gamma}}. \tag{9}$$

Generally, we seek designs that maximize  $A_\boldsymbol{\gamma}$ , which is the same as minimizing  $\bar{V}_\boldsymbol{\gamma}$ . From Eqs. (8) and (9), it can be seen that  $A_\boldsymbol{\gamma}$  is also the harmonic mean of the non-zero eigenvalues of  $\frac{1}{r}\mathbf{C}_\boldsymbol{\gamma}$ . Given this connection, both criteria will lead to *A*-optimal two-phase designs.

#### 3.2 Bayesian criterion

As introduced earlier, optimal designs found at given values of variance components might not be optimal at another set of values of variance components. Therefore, it is advantageous to consider a criterion that can rank designs for a wide range of values of variance components. To achieve this, we develop a Bayesian criterion to rank two-phase designs for all  $\boldsymbol{\gamma} \in \mathcal{D} \subset [0; +\infty) \times [0; +\infty) = \mathbb{R}_+^2$ , where  $\mathbb{R}_+ = [0; +\infty)$ . The general Bayesian criterion can be expressed as

$$A_B = \iint_{\boldsymbol{\gamma} \in \mathcal{D}} u(\boldsymbol{\gamma}) p(\boldsymbol{\gamma}) d\boldsymbol{\gamma}, \tag{10}$$

where  $u(\boldsymbol{\gamma})$  is a utility function and  $p(\boldsymbol{\gamma})$  is a prior distribution summarizing our knowledge for  $\boldsymbol{\gamma}$ . In other words,  $A_B$  is the expected utility of a design obtained by integrating the product of the utility function and the prior for  $\boldsymbol{\gamma}$  over  $\mathcal{D}$  (Chaloner and Verdinelli 1995, see Eq. (1)). Our main goal

is to maximize or minimize the expected utility, depending on the choice of the utility function. There are several ways to choose a utility function (entropy). Here, we consider  $u(\boldsymbol{\gamma}) = A_{\boldsymbol{\gamma}}$  and  $p(\boldsymbol{\gamma}) = p_1(\gamma_1)p_2(\gamma_2)$ , where  $p_1(\gamma_1)$  and  $p_2(\gamma_2)$  are two independent priors for  $\gamma_1$  and  $\gamma_2$ , respectively. Thus, Eq. (10) can be written as:

$$A_B = \iint_{\boldsymbol{\gamma} \in \mathcal{D}} A_{\boldsymbol{\gamma}} p_1(\gamma_1) p_2(\gamma_2) d\gamma_1 d\gamma_2. \tag{11}$$

Because  $0 \leq \varphi_1 \leq \frac{1}{qb_{2p}}$  and  $0 \leq \varphi_2 \leq \frac{1}{qb_{1p}}$ , it is convenient to use  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$  instead of  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$  when deriving an approximation of  $A_B$  (see Proposition 1). Thus, Eq. (11) can be written as

$$A_B = \iint_{\boldsymbol{\varphi} \in \tilde{\mathcal{D}}} A_{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2} p_1(\varphi_1) p_2(\varphi_2) d\varphi_1 d\varphi_2, \tag{12}$$

where  $\tilde{\mathcal{D}}$  is obtained by transforming  $\mathcal{D}$  using  $\boldsymbol{\gamma} \rightarrow \boldsymbol{\varphi}$ . As before, designs with high  $A_B$  are preferable.

### 3.3 Choice of prior distribution

In Bayesian inference, selecting prior information is crucial. Generally, there are several ways to choose a prior, such as a noninformative, a highly informative, or a weakly informative prior. As mentioned in the introduction, in this paper, we assume that we do not have a prior knowledge on the variance components. Therefore, we use a noninformative prior (Liu and Wasserman 2022, Chapter 13). Noninformative priors are often selected from known probability distributions (Liu and Wasserman 2022, Chapter 13). The history of objective priors in Bayesian inference can be found in Kass and Wasserman (1996) and Ghosh (2011). Consonni et al. (2018) provided a review on the choice of prior distributions. The selection of probability distributions is often based on improper priors (i.e., a prior probability density that is not integrable), flat priors (or invariant priors), Jeffreys’s prior, and conjugate priors (Liu and Wasserman 2022, Chapter 13). Gelman (2006) studied a few prior distributions for variance components for a simple hierarchical model (i.e., one treatment and one blocking factor), i.e., uniform, inverse-gamma, and half-Cauchy distributions. The author showed that the uniform distribution belongs to the class of flat priors. The inverse-gamma and half-Cauchy distributions belong to the class of conjugate priors, meaning that the posterior distributions of these are also inverse-gamma or half-Cauchy distributions. The author recommended the uniform and half-Cauchy distributions since the inverse-gamma distribution is sensitive when the variance component approaches zero. Furthermore, because the variance ratios ( $\boldsymbol{\gamma}$ ) can only take

positive values, another possible prior distribution for  $\boldsymbol{\gamma}$  is the log-normal distribution. For these reasons, in the next section, we will study three prior distributions for variance components, i.e., the uniform, log-normal, and half-Cauchy distributions.

## 4 Approximations of the Bayesian criterion

It might be intractable to compute  $A_B$  because of the complexity of the joint function  $A_{\boldsymbol{\varphi}} p_1(\varphi_1) p_2(\varphi_2)$ . Therefore, it is useful to propose an approximation for computing  $A_B$ . There are several ways to approximate  $A_B$  using numerical integration techniques, e.g., Monte Carlo methods (based on sequences of pseudorandom numbers), quasi-Monte Carlo methods (based on low-discrepancy sequences), Gaussian (or Gauss) quadrature rules (Davis and Rabinowitz 2007; Kythe and Schäferkötter 2004; Stoer and Bulirsch 2013). These techniques compute the numerical value of a definite integral. It is, however, time-consuming to compute  $A_B$  using Monte Carlo and quasi-Monte Carlo methods since they require a large number of sample points. In contrast, Gauss quadrature rules propose well-chosen sample points and weights corresponding to the sample points so that  $A_B$  can be computed within a reasonable computing time. Therefore, we will study this technique to get a numerical solution for  $A_B$ . In the next section, we discuss how to choose quadrature rules for one-dimensional integrals. We then extend this to two-dimensional integrals in which we use a full grid to obtain quadrature rules.

### 4.1 Quadrature rules for one-dimensional integrals

We study how to choose quadrature rules for a one-dimensional integral, which is also called a univariate integral. Suppose that we want to compute the integral  $\int_{l_1}^{l_2} \phi(x) f(x) dx$ , where  $l_1$  and  $l_2$  are scalars ( $l_1 < l_2$ ),  $f(x)$  is a smooth function that can be well approximated by polynomials of degree  $2d - 1$  on the interval  $[l_1, l_2]$ , and  $\phi(x)$  is a nonnegative weight function. In numerical integration, it is well known that

$$I_{\phi, f} = \int_{l_1}^{l_2} \phi(x) f(x) dx \approx \sum_{i=1}^d \omega_i f(\xi_i), \tag{13}$$

where  $d$  is the number of sample points used,  $\omega_i$  are quadrature weights and  $\xi_i$  are the roots of a  $d$ th orthogonal polynomial used for the weight function ( $\xi_i$  is also called *abscissa* or *node*) (Stoer and Bulirsch 2013). Depending on a weight function, we can identify quadrature nodes and corresponding quadrature weights. Here, we consider three weight functions, i.e.,  $\phi(x) = 1$ ,  $e^{-x^2}$ , or  $\frac{1}{1+x^2}$ . These weight func-

**Table 1** Gauss quadrature rules for  $\phi(x)$  (Gautschi 2004, Table 1.1)

Interval	$\phi(x)$	Orthogonal polynomial	$\omega_i$
$[-1, 1]$	1	Legendre: $L_d(x) = \frac{2d-1}{d}L_{d-1}(x) - \frac{d-1}{d}L_{d-2}(x)$ , where $L_0(x) = 1$ and $L_1(x) = x$	$\frac{2}{(1-\xi_i^2)[L'_d(\xi_i)]^2}$
$(-\infty, +\infty)$	$e^{-x^2}$	Hermite: $H_{d+1}(x) = 2xH_d(x) - 2dH_{d-1}(x)$ , where $H_0(x) = 1$ $H_1(x) = 2x$	$\frac{2^{d-1}d!\sqrt{\pi}}{d^2[H_{d-1}(\xi_i)]^2}$
$[0, +\infty)$	$\frac{1}{1+x^2}$	Chebyshev: $T_{d+1}(x) = 2xT_d(x) - T_{d-1}(x)$ , where $T_0(x) = 1$ , $T_1(x) = x$	$2(e_i)_1^2$

Notice that  $\xi_i$  are the roots of a  $d$ th orthogonal polynomial used for the weight function.  $(e_i)_1$  are the first components of the eigenvectors of Jacobi matrices (Trefethen and Bau 1997, Chapter 37)

tions correspond to the three prior distributions studied in this paper.

For  $\phi(x) = 1$ , it is well known that Legendre polynomials can be used to obtain nodes and weights to compute  $\int_{-1}^{+1} \phi(x)f(x)dx$ . This is called Gauss–Legendre quadrature (Gautschi 2004). For  $\phi(x) = e^{-x^2}$ , Hermite polynomials can be used to obtain nodes and weights to compute  $\int_{-\infty}^{+\infty} \phi(x)f(x)dx$ . This is called Gauss–Hermite quadrature. For  $\phi(x) = \frac{1}{1+x^2}$ , we can apply a combination between Lanczos and Gram–Schmidt procedures to Chebyshev polynomials. Details will be discussed in the next paragraphs. The Legendre, Hermite, and Chebyshev polynomials and their weights are shown in Table 1.

Goos and Mylona (2018) also provided an excellent tutorial for using Gauss quadrature rules for one-dimensional integrals for several forms of  $\phi(x)$ , including Gauss–Hermite, Gauss–Stieltjes–Wigert, generalized Gauss–Laguerre, and Gauss–Jacobi quadrature. These techniques were illustrated with split-plot experiments. They did not, however, consider the weight  $\phi(x) = \frac{1}{1+x^2}$  that we will also study in this paper.

Let  $\chi$  be a random variable that has the probability density function  $p(x)$ . Here we want to compute the posterior function  $\int u(x)p(x)dx$ , where  $u(x)$  is a utility function. In the following, we explain how to apply Gauss–Legendre quadrature, Gauss–Hermite quadrature, and the combination between Lanczos and Gram–Schmidt procedures to the Chebyshev polynomials to compute the posterior functions for the three prior distributions.

*Uniform distributions:* Consider  $\chi \sim \mathcal{U}(l_1, l_2)$  where  $\mathcal{U}(l_1, l_2)$  is a continuous uniform distribution with the two boundaries  $l_1$  and  $l_2$  ( $l_1 < l_2$ ). Parameters  $l_1$  and  $l_2$  are called hyperparameters. The probability density function  $p(x)$  is equal to  $\frac{1}{l_2-l_1}$  when  $l_1 \leq x \leq l_2$  and zero otherwise. Therefore, our integral is of the form  $\frac{1}{l_2-l_1} \int_{l_1}^{l_2} u(x)dx$ . As mentioned above, the Gauss–Legendre quadrature is applied for the interval of  $[-1, 1]$ . So, we need to perform a change of interval to obtain  $\int_{l_1}^{l_2} u(x)dx$  as follows:

$$\int_{l_1}^{l_2} u(x)dx = \frac{l_2 - l_1}{2} \int_{-1}^1 u\left(\frac{l_2 - l_1}{2}m + \frac{l_1 + l_2}{2}\right) dm. \tag{14}$$

By applying the Gauss–Legendre quadrature rule with  $d$  points  $(\xi_i, \omega_i)$ , the above equation can be approximated as:

$$\frac{1}{l_2 - l_1} \int_{l_1}^{l_2} u(x)dx \approx \frac{1}{2} \sum_{i=1}^d \omega_i u\left(\frac{l_2 - l_1}{2}\xi_i + \frac{l_1 + l_2}{2}\right). \tag{15}$$

*Log-normal distributions:* Consider  $\chi \sim \mathcal{LN}(\mu, \psi^2)$ , where  $\mu$  and  $\psi$  are the location parameter and the scale parameter of the log-normal distribution. Under this case, we need to compute  $\int_0^{+\infty} \frac{1}{x\psi\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\psi}\right)^2} u(x)dx$ . By using a change of variable twice and applying the Gauss–Hermite quadrature rule, our integral can be approximated as follows:

$$\begin{aligned} &\int_0^{+\infty} u(x) \frac{1}{x\psi\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\psi}\right)^2} dx \\ &= \int_{-\infty}^{+\infty} e^{-m^2} \frac{1}{\sqrt{\pi}} u(e^{\mu+\sqrt{2}\psi m}) dm \\ &\approx \sum_{i=1}^d \omega_i \frac{1}{\sqrt{\pi}} u(e^{\mu+\sqrt{2}\psi \xi_i}), \end{aligned} \tag{16}$$

where  $x = e^z$ ,  $m = \frac{z-\mu}{\sqrt{2}\psi}$  or  $z = \mu + \sqrt{2}\psi m$ ,  $dx = e^z dz$ ,  $dz = \sqrt{2}\psi dm$ , and  $\xi_i$  and  $\omega_i$  are nodes and weights obtained from the Gauss–Hermite quadrature rule.

*Half-Cauchy distributions:* Assume that  $\chi \sim \mathcal{HC}(\mu, \psi)$ , where  $\mu$  is the location parameter (specifying the location of the peak of the distribution), and  $\psi$  is the scale parameter which specifies the half-width at half-maximum. Our integral is of the form:  $\int_{\mu}^{+\infty} \frac{2}{\pi\psi(1+(\frac{x-\mu}{\psi})^2)} u(x)dx$ .

One of the possible approaches to approximate this is to use the well-known technique *Lanczos* with a different inner product, which is similar to the *Stieltjes* procedure. This approach can be applied to obtain the orthogonal polynomial for an arbitrary weight function, see Olver (2022). In what follows, we provide a brief workflow to obtain  $d$  nodes and its weights for an arbitrary weight function  $\phi(x)$ . It is a modification of Gauss quadrature rules. The general workflow of Gauss quadrature rules can be found in textbooks or works on numerical integration (e.g., Golub and Welsch 1969; Kythe and Schäferkötter 2004). Traditionally, this workflow

is based on a monomial basis  $\{1, x, x^2, \dots, x^d\}$  with an inner product  $\langle \cdot, \cdot \rangle$  defined as  $\langle p_1, p_2 \rangle = \int_{l_1}^{l_2} \phi(x) p_1(x) p_2(x) dx$ , where  $p_1(x)$  and  $p_2(x)$  are functions chosen from the monomial basis. Next, these polynomials are orthonormalized to orthogonal polynomials, which can be done using the well-known *Gram–Schmidt orthogonalization*. Let  $\{q_0(x), q_1(x), q_2(x), \dots, q_d(x)\}$  be a set of orthogonal polynomials obtained from the Gram–Schmidt orthogonalization. The Gauss quadrature points are then the roots of  $q_d(x)$ . Furthermore, as is reviewed in, e.g., Chapter 37 of Trefethen and Bau (1997), these roots (and also the quadrature weights) can be obtained by solving a tridiagonal  $d \times d$  eigenvalue problem whose entries are obtained by a well-known algorithm *Lanczos* (Lanczos 1950). As mentioned in Johnson (2019), for numerical stability when constructing Gauss quadrature rules computationally for arbitrary weight functions, the author suggested applying the Lanczos process to polynomials in the basis  $\{T_0, T_1, T_2, \dots, T_d\}$  of *Chebyshev polynomials*, after rescaling the interval  $[l_1, l_2]$  to  $[-1, 1]$ , rather than using the monomial basic. We call this *Gauss–Chebyshev–Lanczos* approach. It is implemented in the Julia QuadGK.jl package (Johnson 2022b). It is used to obtain nodes and weights for half-Cauchy distributions considered in this paper.

By applying the aforementioned technique, our integral can be approximated as follows:

$$\int_{\mu}^{+\infty} u(x) \frac{2}{\pi \sigma (1 + (\frac{x-\mu}{\psi})^2)} dx \approx \sum_{i=1}^d \omega_i u(\xi_i), \tag{17}$$

where  $\xi_i$  and  $\omega_i$  are nodes and weights obtained from the Gauss–Chebyshev–Lanczos approach.

### 4.2 Quadrature rules for two-dimensional integrals

We extend the quadrature rules for one-dimensional to two-dimensional integrals, which are used to approximate  $A_B$ . Without loss of generality, we assume that  $f(x, y)$  is a smooth function on the domain  $\mathcal{D} = [l_1, l_2] \times [l_1, l_2]$ . This can be accomplished for bounded rectangular domains by a proper change of interval. Suppose that we want to compute  $\iint_{\mathcal{D}} \phi_1(x) \phi_2(y) f(x, y) dx dy$ , where  $f(x, y)$  can be well approximated by polynomials of degree  $2d - 1$  within  $\mathcal{D}$ , and  $\phi_1(x)$  and  $\phi_2(y)$  are nonnegative weight functions. Using the tensor product rule (full-grid), our integral can be written as:

$$\begin{aligned} I_{\phi_1 \phi_2, f} &= \iint_{\mathcal{D}} \phi_1(x) \phi_2(y) f(x, y) dx dy \\ &\approx \sum_{i=1}^d \sum_{j=1}^d \omega_{1i} \omega_{2j} f(\xi_i, \xi_j), \end{aligned} \tag{18}$$

where  $d$  is the number of sample points used for each dimension,  $\omega_{1i}$  and  $\omega_{2i}$  are quadrature weights obtained from the first and second dimensions, and  $\xi_i$  and  $\xi_j$  are the roots of  $d$ th orthogonal polynomials used in the first and second dimensions. Under the tensor product rule, this approximation requires  $d^2$  sample points. Notice that it is possible to use the sparse grid approach, originally proposed by Smolyak (1963) to obtain sample points. However, this approach is recommended only when the number of dimensions is greater than two. Also, under this approach, we recommend using nested quadrature rules instead of Gauss quadrature rules, for example, Fejér and Clenshaw–Curtis. In the rest of this article, we will use the right-hand side of Eq. (18) to approximate  $A_B$ . The approximated value is denoted as  $\tilde{A}_{QB}$ .

### 4.3 Multidimensional integration–benchmark

Another technique to compute a multidimensional integral is based on a multidimensional *h-adaptive* integration (Berntsen et al. 1991; Genz and Malik 1980), which uses a globally adaptive subdivision strategy. This algorithm recursively partitions the integration domain into smaller subdomains, and then applies the same integration rule to each, until convergence is achieved. It is best suited for a moderate number of dimensions (e.g.,  $< 7$ ), and is superseded for high-dimensional integrals by other methods (e.g., Monte Carlo variants or sparse grids). A general structure of adaptive integration algorithms consists of the following four main steps: (1) choose some subregion(s) from a set of subregions, (2) subdivide the chosen subregion(s), (3) apply the integration rule to the new subregions; update the subregion set, and (4) update global integral and error estimates; check for convergence. Genz and Malik (1980) improved the adaptive subdivision strategy to facilitate parallelization at the subregion level and allow for more than one integrand, proposed several different integration rules that can be selected for use with the algorithm, and proposed a new error estimation method. This approach is implemented in the Julia package HCubature by Johnson (2022a). In the next section, we will use this package to evaluate the performance of our approximated Bayesian criterion proposed in the previous section.

## 5 Evaluating approximations of the Bayesian criterion

We evaluate the performance of the approximations of Bayesian criterion based on quadrature rules and HCubature. The approximated value of  $A_B$  found using HCubature is denoted as  $\tilde{A}_{HB}$ . For our proposed approximated Bayesian criterion using quadrature rules, up to 20 nodes are used for each dimension, meaning that we approximate  $A_{\gamma}$  up

to  $2 \times 20 - 1 = 39$  degrees, which requires the evaluation of up to 400 points to get an approximated value of  $A_B$ . The approximated value of  $A_B$  under this approach is denoted as  $\tilde{A}_{QB}$ . Let us consider the two-phase experiment that was briefly mentioned in the introduction as follows.

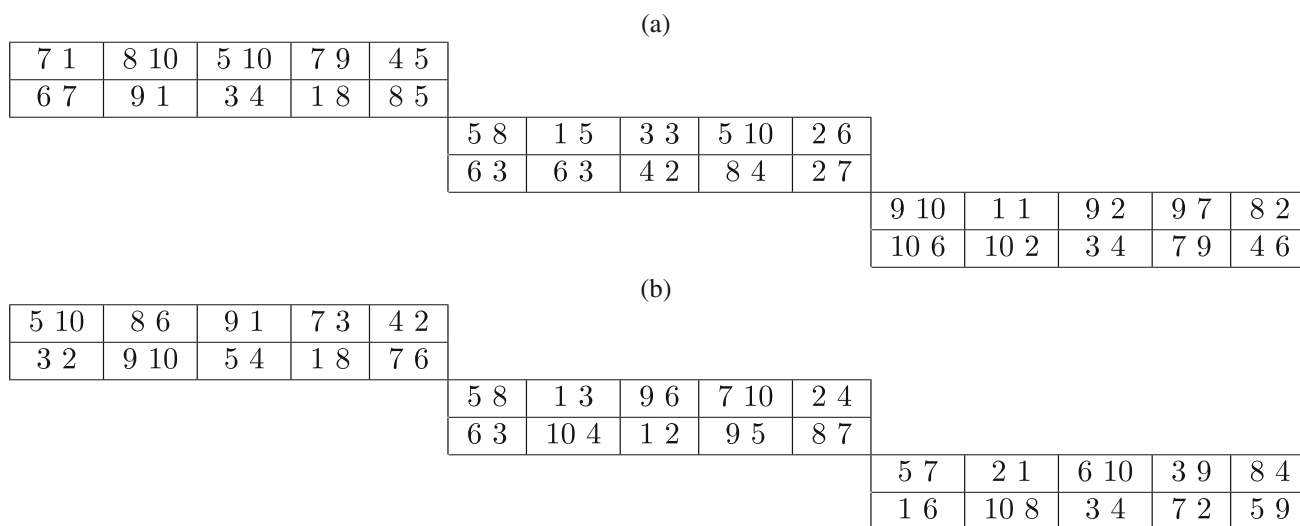
*Example I:* Suppose that an experimenter wants to test 10 treatments in a field involving 6 blocks in Phase 1, each of which can test 10 treatments. The 60 samples obtained from this phase are then analyzed in a laboratory in Phase 2. Since a machine in the laboratory can only process four samples at a time, we need 15 machines to process all the samples. Because there might be field-block-to-field-block and machine-to-machine variation, it is essential to consider these sources of variation as blocking factors. Following the notation used in Sect. 2, parameters for this experiment are  $v = 10, r = 6, b_1 = 6, k_1 = 10, b_2 = 15,$  and  $k_2 = 4$ . Under these parameters, a nested row-column pattern has  $p = \text{gcd}(b_1, b_2) = 3$  superblocks, and each of which has  $b_{1p} = b_1/p = 2$  rows and  $b_{2p} = b_2/p = 5$  columns. There are  $q = \text{gcd}(k_1, k_2) = 2$  experimental units in a cell formed by the intersection of a row and a column (Fig. 1a).

First, we generate a random two-phase design for Example I to evaluate approximations of the Bayesian criterion. This design is created by randomly assigning treatments to the nested row-column pattern, see Fig. 1a. Next, we identify hyperparameters for each distribution (i.e.,  $l_1$  and  $l_2$  in uniform distributions,  $\mu$  and  $\psi$  in normal and half-Cauchy distributions). These hyperparameters can be based on an educated guess or empirical datasets. Here we suppose that we do not have empirical datasets. Thus, it is reasonable to consider two scenarios for  $\gamma_i$ : (1)  $0 \leq \gamma_i \leq 1$  and (2)

$\gamma_i > 1 (i = 1, 2)$ . In the following analysis, we choose  $[0, 1]$  and  $[1, 10]$  as two boundaries for  $\gamma$  under uniform distributions. We use  $\tilde{A}_{HB}$  as a benchmark.

Figure 2 displays various approximations for  $A_B$  corresponding to four combinations of variance ratios under the uniform distributions. Figures 2a–d correspond to the four combinations of variance ratios. For each figure, the  $x$  axis shows the number of nodes used to approximate  $A_B$  under the quadrature rules, while the  $y$  axis shows the approximated values of  $A_B$  found either from the quadrature rules or HCubature. A blue line corresponds to the value of  $\tilde{A}_{HB}$ , while orange markers correspond to the values of  $\tilde{A}_{QB}$ .

Figure 2a shows approximations for  $A_B$  when the variance ratios  $\gamma_1$  and  $\gamma_2$  are small. In this case, the value of  $\tilde{A}_{HB}$  is approximated to 0.79090. The approximated values of  $A_B$  using Gauss–Legendre converge to  $\tilde{A}_{HB}$  after 5 nodes. When one of the variance ratios is small and the other variance ratio is large, the values of  $\tilde{A}_{QB}$  and  $\tilde{A}_{HB}$  are slightly smaller than the ones obtained from Fig. 2a. This is clearly shown in Figs. 2b and 2c. When both variance ratios are large, it is expected that the block effects in both phases can be considered as fixed. In this case, the values of  $\tilde{A}_{QB}$  are smallest compared to other values of  $\tilde{A}_{QB}$  shown in Figs. 2a–c. Under this analysis, it is sufficient to use 5 nodes to get a good approximated value of  $A_B$ . This indicates that using Gauss–Legendre reduces not only computing times but also gives a good approximation for  $A_B$ . This is mainly because it requires a smaller number of sample points and well-chosen sample points, compared to prevalent approaches such as Monte Carlo and quasi-Monte Carlo.



**Fig. 1** (a) A random allocation of treatments to a nested row-column pattern for Example I. (b) A Bayesian optimal two-phase design for Example I is found with  $\gamma_1 \sim \mathcal{U}(0, 1)$  and  $\gamma_2 \sim \mathcal{U}(0, 1)$ . In both pan-

els, rectangles are superblocks. Rows in superblocks are blocks in Phase 1, while columns in superblocks are blocks in Phase 2

Based on these findings, we propose a search framework in the next section to find a Bayesian optimal two-phase design where we use 10 nodes to approximate the Bayesian criterion. As an illustration, a Bayesian optimal two-phase design for Example I found with  $\gamma_1 \sim \mathcal{U}(0, 1)$  and  $\gamma_2 \sim \mathcal{U}(0, 1)$  is shown in Fig. 1b. As shown in this figure, one of the good properties of this design is that treatments have equal replications in each superblock. This property resembles one of the properties of balanced incomplete block designs in which all pairs of treatments occur together within a block an equal number of times (Shah and Sinha 1989). For instance, treatment 1 occurs twice in each superblock. This is opposed to the design shown in Fig. 1a, where treatment 1 occurs unequally in each superblock.

Likewise, we use two log-normal distributions, i.e.,  $\mathcal{LN}(0, \frac{1}{2})$  and  $\mathcal{LN}(1, \frac{1}{2})$ , as priors for both variance ratios to evaluate values of  $\tilde{A}_{HB}$  and  $\tilde{A}_{QB}$ . When  $\gamma_i \sim \mathcal{LN}(0, \frac{1}{2})$ , it is expected that the variance ratios are small, while when  $\gamma_i \sim \mathcal{LN}(1, \frac{1}{2})$ , it is expected that the variance ratios are large ( $i = 1, 2$ ). Figure 3 shows various approximated values of  $A_B$  with the four combinations of the two variance ratios. The structures of this figure and Fig. 2 are identical. Here we use a quantile function (i.e., inverse cumulative distribution function) to get an upper bound for the two distributions with the probability value of 0.999. This avoids the problem when both variance ratios approach infinity ( $+\infty$ ) when using the HCubature approach.

Similar to the uniform distributions, the values of  $\tilde{A}_{QB}$  obtained from the Gauss–Hermite rule converged to the value of  $\tilde{A}_{HB}$  with 5 nodes for the four cases considered in Fig. 3. Finally, we considered the Gauss–Legendre rule to get an approximated value of  $A_B$  where the weight function and  $A_\gamma$  are combined as a joint function. This approach also worked quite well, compared to the Gauss–Hermite approach. It converges to  $\tilde{A}_{HB}$  after 10 nodes.

We consider two half-Cauchy distributions  $\mathcal{HC}(0, 1)$  and  $\mathcal{HC}(1, 10)$  as priors for  $\gamma$ . Similar to the uniform and log-normal distributions, these distributions represent the small and large values of variance ratios as discussed before. Figure 4 plots various approximated values of  $A_B$  for the four scenarios. The structure of this figure is analogous to the structure of Figs. 2 and 3. Here we also use a quantile function to get an upper bound for the two distributions with the probability value of 0.99. This avoids the problem when both variance ratios approach infinity ( $+\infty$ ). Likewise, the quadrature approaches converge to  $\tilde{A}_{HB}$  after 10 nodes. However, the Gauss–Legendre approach converges more slowly to  $\tilde{A}_{HB}$  than the Gauss–Chebyshev–Lanczos approach.

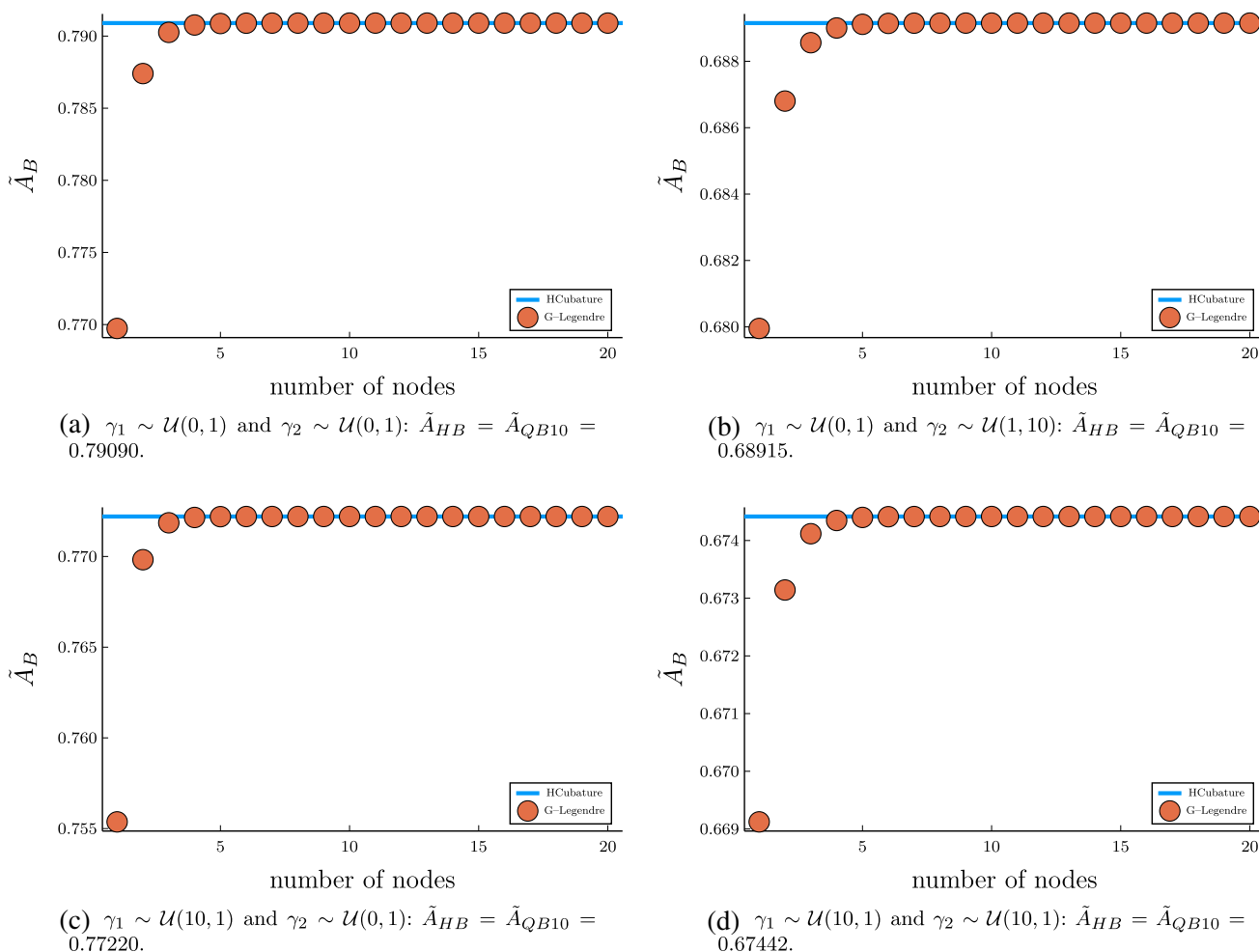
Note that it is possible to use the HCubature approach to find Bayesian optimal two-phase designs. However, it is costly to use this approach when searching for a large

two-phase design (i.e., a design with a large number of treatments). This is because update formulae cannot be applied under the HCubature approach. For example, we compared computing times of finding Bayesian two-phase designs for Example I with one iteration. Using the HCubature approach, it took 159.21 (s), 1050.56 (s), and 3467.97 (s) to obtain two-phase designs for the uniform, log-normal, and half-Cauchy distributions, respectively. However, using the quadrature rules, the average computing time for uniform, log-normal, and half-Cauchy distributions was less than 1 (s).

## 6 A search algorithm

In the literature, most search algorithms for comparative experiments focus on variants of simulated annealing (Kirkpatrick et al. 1983) and Tabu search (Glover 1989), see a review in Vo-Thanh and Piepho (2022). Recently, Vo-Thanh and Piepho (2022) considered a variant of hill climbing (Appleby et al. 1961) to obtain optimal row-column designs with a complex blocking structure, and equal or unequal treatment replications. They showed that their proposed algorithm is competitive with the variants of simulated annealing and Tabu search used to search for optimal designs for comparative experiments. In this paper, we utilize their work to obtain Bayesian two-phase designs. The workflow is briefly described as follows.

Generally, a computer search starts with an initial design, which is also a candidate design. We then iteratively improve this design in a number of steps ( $\rho_0$ ). At each step, given a candidate design, a better design can be found via neighbor designs of the current candidate design. A neighbor design is obtained from a given design by applying a swapping mechanism to this design. In this paper, we consider a swapping mechanism where we exchange two treatments in the design matrix  $\mathbf{X}$ , while fixing the nested row-column pattern ( $\mathbf{Z}$ ). The next candidate design is selected among the neighbor designs generated by performing all possible swaps in  $\mathbf{X}$ . This selection is based on the accepting and rejecting strategies considered in Vo-Thanh and Piepho (2022), using  $\tilde{A}_{QB}$  as a main search criterion. The selected candidate design will be used in the next iteration. A pseudocode of the search framework is shown in Algorithm 1, while details of this algorithm can be found in Vo-Thanh and Piepho (2022). As pointed out in Vo-Thanh and Piepho (2022), it is essential to use an efficient update formula when computing a generalized inverse of the treatment information matrix before and after interchanging treatments. Here we adapted their strategies to derive an update formula allowing to swap two treatments in the design matrix  $\mathbf{X}$ , while fixing the design matrix  $\mathbf{Z}$  (see Supplementary Section B). In the next section, we apply this search framework to find Bayesian optimal two-phase designs.



**Fig. 2** Approximations of the Bayesian criterion for a two-phase design are computed under uniform distributions. For each subfigure, the  $x$  axis shows the number of nodes used in Gauss–Legendre (G-Legendre) rang-

ing from 1 up to 20 nodes. The  $y$  axis shows various approximated values of  $A_B$ .  $\tilde{A}_{QB10}$  is the value of  $\tilde{A}_{QB}$  found using 10 nodes

**Algorithm 1** A search framework for finding Bayesian optimal two-phase designs.

**Require:** a random starting design ( $\mathbf{D}_s$ ).  
 $\mathbf{D}_c \leftarrow \mathbf{D}_s$   
 $\mathbf{D}_o \leftarrow \mathbf{D}_s$   
**for**  $i = 1$  to  $\rho_0$  **do**  
 step 1: generate all possible neighbor designs for the candidate design ( $\mathbf{D}_c$ ).  
 step 2: compute  $A_{QB}$  of all neighbor designs.  
 step 3: select the best design in the neighbor designs.  
 step 4: choose a next candidate design and update  $\mathbf{D}_c$ .  
 step 5: update the optimal design ( $\mathbf{D}_o$ ) if possible.  
**end for**  
**return** an optimal two-phase design ( $\mathbf{D}_o$ ) found during the search.

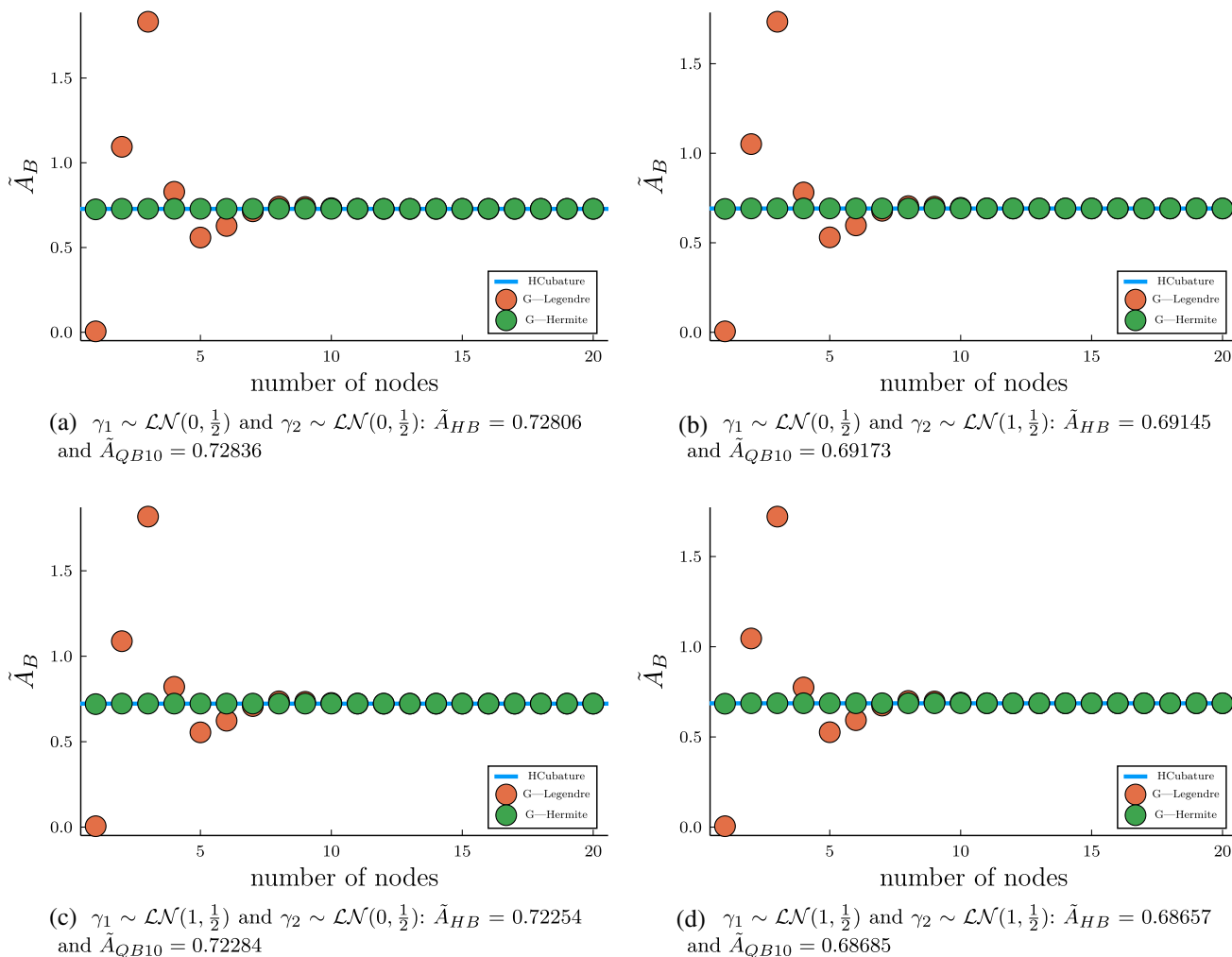
**7 Bayesian two-phase designs**

We illustrate the search framework to find Bayesian optimal two-phase designs for 16 parameter settings shown in Table 2. The first column of this table represents an identifi-

cation (ID) for the examples. The remaining columns show the design information as mentioned in Sect. 2. We study the three prior distributions for these 16 design settings. Based on the preliminary results in Sect. 5, we applied Gauss quadrature rules with 10 nodes for each dimension to approximate the Bayesian criterion.

As mentioned in Sect. 5, it is practical to choose a prior distribution for a variance ratio that represents small and large variances. In the following, we use the notation  $\boldsymbol{\gamma}_{ab}$  to represent the choice of the two variance ratios where  $a$  and  $b$  represent the choices of  $\gamma_1$  and  $\gamma_2$ . There are four possible choices of variance ratios considered in this section, i.e.,  $\boldsymbol{\gamma}_{--}$ ,  $\boldsymbol{\gamma}_{-+}$ ,  $\boldsymbol{\gamma}_{+-}$  and  $\boldsymbol{\gamma}_{++}$ , where  $-$  and  $+$  represent small and large variance ratios. For instance,  $\boldsymbol{\gamma}_{-+}$  represents the case where  $\gamma_1$  is chosen to be small, while  $\gamma_2$  is chosen to be large.

For each of the three distributions studied in this paper, we applied our proposed algorithm to search for optimal designs with four combinations of two variance ratios as



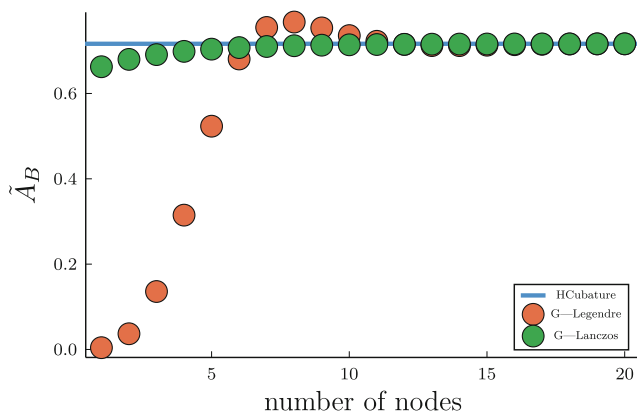
**Fig. 3** Approximations of the Bayesian criterion for a two-phase design are computed under the log-normal distributions. For each subfigure, the x axis shows the number of nodes used in Gauss quadrature rules

(i.e., Gauss–Legendre and Gauss–Hermite) ranging from 1 up to 20 nodes. The y axis shows various approximated values of  $A_B$ .  $\tilde{A}_{QB10}$  is the value of  $\tilde{A}_{QB}$  found using Gauss–Hermite with 10 nodes

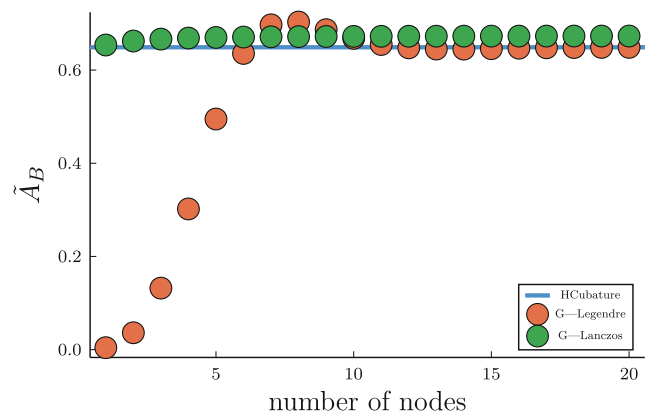
discussed above. In total, for each of the three distributions, we obtained 64 Bayesian optimal two-phase designs. Parameters used in the search algorithm are  $(\rho_0, \rho_1, \rho_2, \rho_3) = (400, 100, 400, 1)$  where  $\rho_0$  is the number of iterations used to search for a design,  $\rho_1$  is the length of a fitness array,  $\rho_2$  is the length of a visited candidate array,  $\rho_3$  is a percentage of neighbor designs. Details of these parameters can be found in Vo-Thanh and Piepho (2022). The implementation of our algorithm can be found in the supplementary material of this paper (see JCode.7z). The algorithm was run under Julia 1.5.3 on Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz and 16GB of memory. The average computing time required to find 64 two-phase designs for each distribution is about 13,690.68 s ( $\approx 3.8$  h). In the following section, we detail our findings.

### 7.1 Uniform distributions

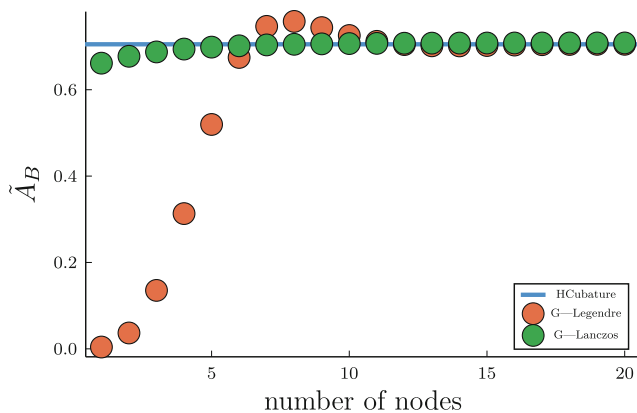
Figure 5 shows three panels that represent various information of the 64 Bayesian two-phase designs obtained from the four combinations  $(\gamma_{--}, \gamma_{-+}, \gamma_{+-}$  and  $\gamma_{++})$ . Figure 5a shows the values of  $\tilde{A}_{HB}$  of the 64 optimal designs found with  $\gamma_{--}, \gamma_{-+}, \gamma_{+-}$  and  $\gamma_{++}$ , where “–” and “+” represent the two uniform distributions  $\mathcal{U}(0, 1)$  and  $\mathcal{U}(1, 10)$ , respectively. For each combination, the values of  $\tilde{A}_{HB}$  of 16 designs range from 0.64188 to 0.96998. This indicates that these optimal designs are able to estimate pairwise treatment comparisons with a high design efficiency. As shown in this panel, it can be seen that, if  $k_1 \leq k_2$ ,  $\tilde{A}_{HB\gamma_{++}} \leq \tilde{A}_{HB\gamma_{+-}} \leq \tilde{A}_{HB\gamma_{-+}} \leq \tilde{A}_{HB\gamma_{--}}$ , where  $\tilde{A}_{HB\gamma_{ab}}$  is the value of  $\tilde{A}_{HB}$  of a design obtained from  $\gamma_{ab}$ . If  $k_1 \geq k_2$ , we have  $\tilde{A}_{HB\gamma_{++}} \leq \tilde{A}_{HB\gamma_{-+}} \leq \tilde{A}_{HB\gamma_{+-}} \leq \tilde{A}_{HB\gamma_{--}}$ . For Examples 1 and 2, the numbers of blocks in both phases are identical.



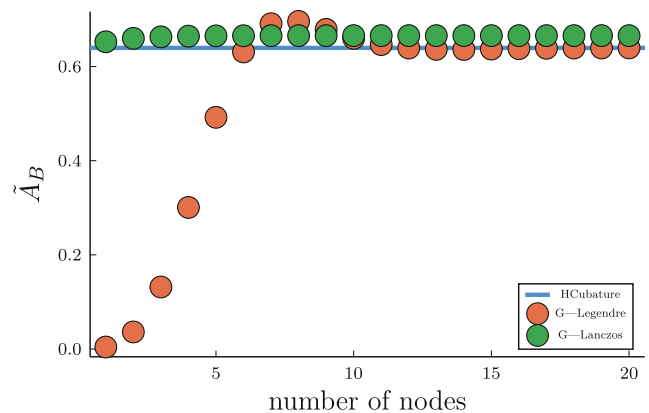
(a)  $\gamma_1 \sim \mathcal{HC}(0, 1)$  and  $\gamma_2 \sim \mathcal{HC}(0, 1)$ :  $\tilde{A}_{HB} = 0.716498$  and  $\tilde{A}_{QB10} = 0.714122$ .



(b)  $\gamma_1 \sim \mathcal{HC}(0, 1)$  and  $\gamma_2 \sim \mathcal{HC}(1, 10)$ :  $\tilde{A}_{HB} = 0.648934$  and  $\tilde{A}_{QB10} = 0.67230$ .



(c)  $\gamma_1 \sim \mathcal{HC}(1, 10)$  and  $\gamma_2 \sim \mathcal{HC}(0, 1)$ :  $\tilde{A}_{HB} = 0.705224$  and  $\tilde{A}_{QB10} = 0.706876$ .



(d)  $\gamma_1 \sim \mathcal{HC}(1, 10)$  and  $\gamma_2 \sim \mathcal{HC}(1, 10)$ :  $\tilde{A}_{HB} = 0.639667$  and  $\tilde{A}_{QB10} = 0.665738$ .

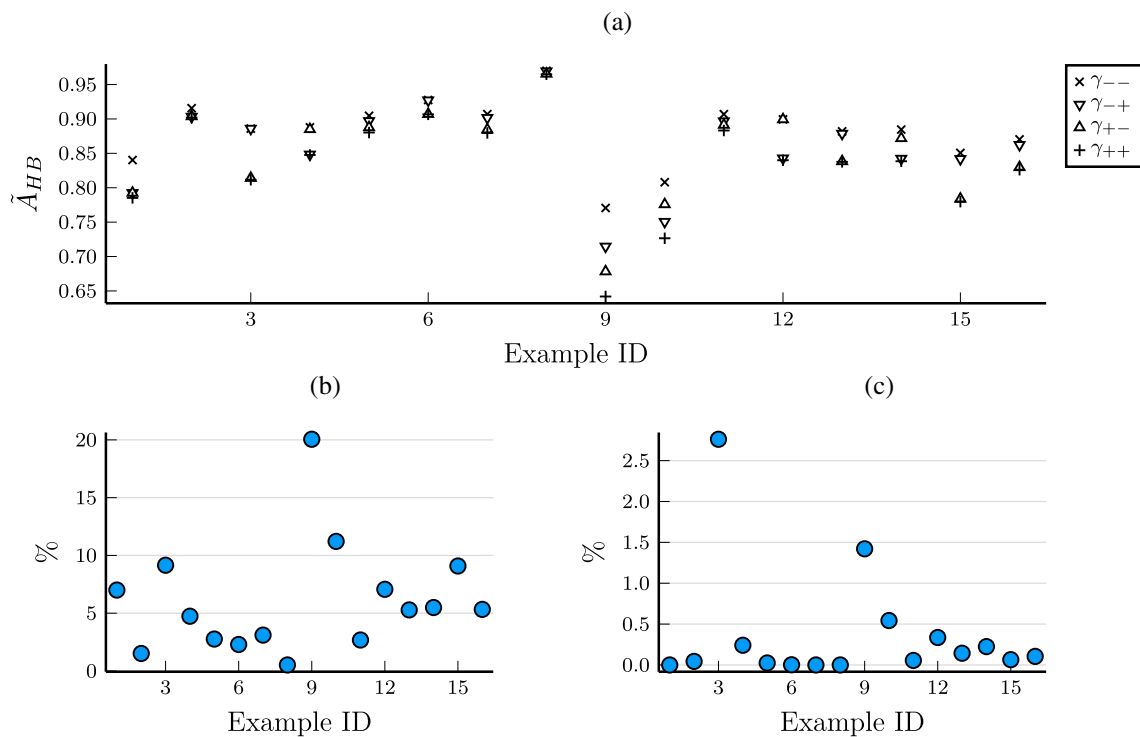
**Fig. 4** Approximations of the Bayesian criterion for a two-phase design are computed based on half-Cauchy distributions. For each subfigure, the x axis shows the number of nodes used in Gauss quadrature

rules ranging from 1 up to 20 nodes. The y axis shows various approximated values of  $A_B$ .  $\tilde{A}_{QB10}$  is the value of  $\tilde{A}_{QB}$  found using Gauss-Chebyshev-Lanczos with 10 nodes

For this case, it is best to confound the blocks in both phases as advocated in Brien et al. (2011, Big with Big). This is achieved under the nested row-column structure. For Example 1, the designs obtained with the four combinations of variance ratios are balanced incomplete block designs. For Examples 3 and 4, the Bayesian optimal two-phase designs are nested designs. Vo-Thanh et al. (2022) recommended constructing a design in a phase with smaller block size first. Then blocks in this resulting design can be grouped to form blocks in the other phase. This grouping is done via the nested row-column pattern. For Example 5, the Bayesian optimal two-phase design is a row-column design with  $q = 1$  (one experimental unit per cell). For Examples 6, 7, and 8, the Bayesian optimal two-phase designs are row-column designs with  $q > 1$  (more than one experimental unit per cell). For Examples 9 and 10, the Bayesian optimal designs are nested row-column designs with  $q = 1$  (one experimental unit per cell). For Examples 11 to 16, the Bayesian optimal designs

are nested row-column designs with  $q > 1$  (more than one experimental unit per cell).

Figure 5b shows the percentage change of  $\tilde{A}_{HB}$  of the Bayesian optimal designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ . The percentage change is computed as follows. We take the value of  $\tilde{A}_{HB}$  of a design obtained from  $\gamma_{++}$  and subtract the value of  $\tilde{A}_{HB}$  of a design obtained from  $\gamma_{--}$ . The result is a gain or loss. We take the gain or loss and divide it by the value of  $\tilde{A}_{HB}$  of the design obtained from  $\gamma_{++}$ . Finally, we multiply the result by 100 to arrive at the percentage change in the design efficiency. As shown in this panel, the percentage change can vary between 0.52% to 20%. As expected, the values of  $\tilde{A}_{HB}$  of designs obtained under  $\gamma_{++}$  are slightly lower than the values of  $\tilde{A}_{HB}$  of designs obtained under  $\gamma_{--}$ . This indicates there is a substantial benefit when considering an analysis with the linear mixed model, which is often the case in plant breeding. There are seven examples where the gaps are less than 5% (Examples 2, 4, 5, 6, 7, 8, and 11). There are seven examples where the gaps are between 5% to 10% (Examples



**Fig. 5** (a) Values of  $\tilde{A}_{HB}$  of the Bayesian optimal two-phase designs obtained under the uniform distributions. The x axis shows Example IDs, while the y axis shows approximated values of  $A_B$  ( $\tilde{A}_{HB}$ ). (b) Percentage change of  $\tilde{A}_{HB}$  of Bayesian two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $\tilde{A}_{HB}$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The y axis shows the percentage change of  $\tilde{A}_{HB}$ . (c) Percentage change of  $A_\infty$  of the

Bayesian two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$  under the linear mixed model, and  $A_\infty$  is the efficiency of a design evaluated under the fixed-effects model, where  $A_\infty$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The y axis shows the percentage change of  $A_\infty$ . Notice that:  $\gamma_{--} = \{\mathcal{U}(0, 1), \mathcal{U}(0, 1)\}$ ,  $\gamma_{+-} = \{\mathcal{U}(0, 1), \mathcal{U}(1, 10)\}$ ,  $\gamma_{+0} = \{\mathcal{U}(1, 10), \mathcal{U}(0, 1)\}$ , and  $\gamma_{++} = \{\mathcal{U}(1, 10), \mathcal{U}(1, 10)\}$

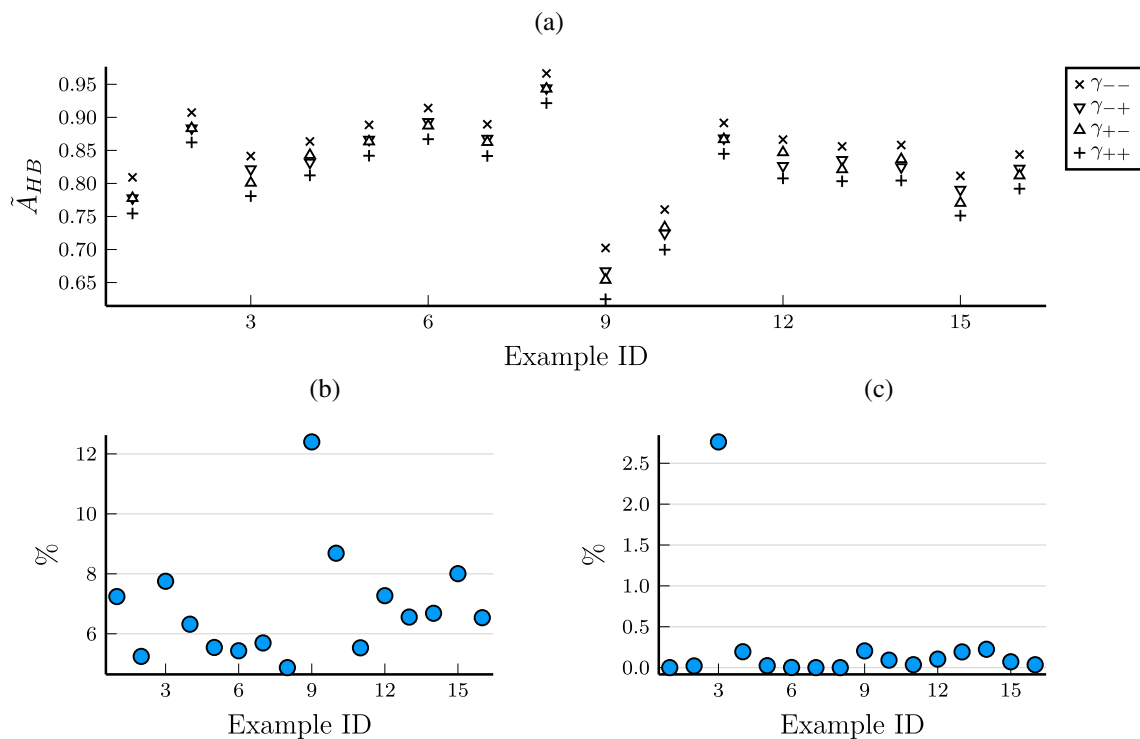
1, 3, 12, 13, 14, 15, and 16). There are two examples where the gaps are around 11% and 19.5% (Examples 9 and 10). As shown in Table 2, the block sizes in Examples 9 and 10 in both phases are relatively small, while the numbers of blocks in both phases are large. We conjecture that this might be a reason why we have seen a large gap. Therefore, in this case, we would suggest considering an analysis with random block effects instead of fixed block effects in both phases for the recovery of inter-block information.

Suppose that we are also interested in the performance of our designs under the fixed-effects model discussed in the introduction. This is relevant because optimal designs obtained under a mixed model are often disconnected under a fixed-effects model. However, this is not the case with our Bayesian optimal two-phase designs since these designs are optimized for a range of plausible values of variance components. In fact, Fig. 5c shows the percentage change of  $A_\infty$  of Bayesian two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $A_\infty$  is the design efficiency of a design computed under the fixed-effects model, and  $A_\infty$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. As shown in this figure, these designs are indeed connected

under the fixed-effects model. Furthermore, the losses of design efficiencies for all but Example 3 are relatively small. This suggests that our Bayesian two-phase designs could be used for analysis with a linear fixed-effects model.

### 7.2 Log-normal distributions

In this section, we consider two log-normal distributions  $\mathcal{LN}(0, 0.5)$  and  $\mathcal{LN}(1, 0.5)$  covering the two bounds  $[0, 1]$  and  $[1, 10]$  for the two variance ratios as indicated before. Likewise, “-” and “+” represent the two log-normal distributions  $\mathcal{LN}(0, 0.5)$  and  $\mathcal{LN}(1, 0.5)$ , respectively. Figure 6 shows three panels that represent information on the 64 Bayesian two-phase designs obtained from the four scenarios ( $\gamma_{--}$ ,  $\gamma_{+-}$ ,  $\gamma_{+0}$  and  $\gamma_{++}$ ). Its structure is analogous to that of Fig. 5. As discussed before, we want to avoid setting the variance ratios to infinity when computing  $A_\gamma$ . Therefore, we use a quantile function to get upper bounds when finding the values of  $\tilde{A}_{HB}$ , where we choose the value of probability of 0.99. Hence,  $\tilde{A}_{HB}$  can be evaluated. As shown in Fig. 6a, the values of  $\tilde{A}_{HB}$  of the Bayesian optimal two-phase designs obtained from the four scenarios range from



**Fig. 6** (a) Values of  $\tilde{A}_{HB}$  of the optimal two-phase designs obtained under the log-normal distributions. The  $x$  axis shows Example IDs, while the  $y$  axis shows the values of  $\tilde{A}_{HB}$ . (b) Percentage change of  $\tilde{A}_{HB}$  of the Bayesian optimal two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $\tilde{A}_{HB}$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The  $y$  axis shows the percentage change of  $\tilde{A}_{HB}$ .

(c) Percentage change of  $A_\infty$  of Bayesian two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $A_\infty$  is the efficiency of a design computed under the fixed-effects model, and  $A_\infty$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The  $y$  axis shows the percentage change of  $A_\infty$ . Notice that  $\gamma_{--} = \{\mathcal{LN}(0, 0.5), \mathcal{LN}(0, 0.5)\}$ ,  $\gamma_{-+} = \{\mathcal{LN}(0, 0.5), \mathcal{LN}(1, 0.5)\}$ ,  $\gamma_{+-} = \{\mathcal{LN}(1, 0.5), \mathcal{LN}(0, 0.5)\}$ , and  $\gamma_{++} = \{\mathcal{LN}(1, 0.5), \mathcal{LN}(1, 0.5)\}$

0.62485 to 0.96641. For Example 1, we also obtained four balanced incomplete block designs as were obtained with the uniform distributions. As shown in Fig. 6b, its pattern is similar to the pattern of Fig. 5b, where the percentage changes are less than 10% for all but Example 9. When considering the fixed-effects model, all designs are indeed connected as shown in Fig. 6c. For Examples 3, the percentage change (loss) is the largest among other examples, which is around 2.6%. If experimenters accept a minor loss ( $< 3\%$ ), these designs could be used to plan a two-phase experiment when the linear fixed-effects model is of interest.

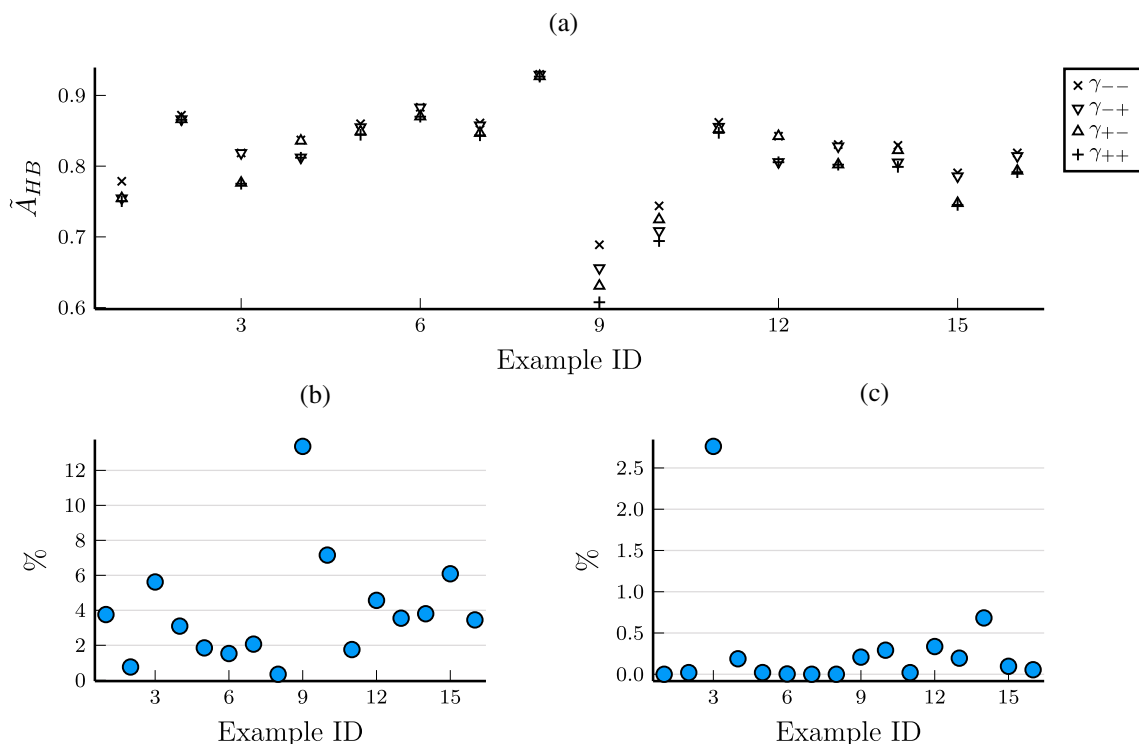
### 7.3 Half-Cauchy distributions

Two half-Cauchy distributions are studied in this section, i.e.,  $\mathcal{HC}(0, 1)$  and  $\mathcal{HC}(1, 10)$  as priors for  $\gamma$ . Under this case, “-” and “+” represent  $\mathcal{HC}(0, 1)$  and  $\mathcal{HC}(1, 10)$ , respectively. Likewise, Fig. 7 shows information on the 64 Bayesian designs found under the two half-Cauchy distributions. The structure of this figure is identical to the structure of Fig. 5. The values of  $\tilde{A}_{HB}$  of the Bayesian optimal designs range from 0.60759 to 0.92974. For the percentage changes

(Fig. 7b), there is also a similar pattern as we have seen with those obtained under the uniform and log-normal distributions. The losses for all but Example 9 are less than 10%, while the loss for Example 9 is around 12%. In terms of the linear fixed-effects model, all designs are also connected and the losses are at most 2.76%. Overall, the general picture here is that we obtain consistent results as with the uniform and log-normal distributions.

## 8 Bayesian balanced incomplete block designs and their dual designs

As discussed in the previous section, when the block sizes in both phases are identical, it is best to confound blocks in both phases to construct a two-phase design. This is indeed an optimal strategy as shown in Brien et al. (2011). Therefore, in such a case, we can use a block design to construct a two-phase design, which will be our focus in this section. In particular, we focus on a balanced incomplete block design and its dual design, where the dual design is obtained by interchanging the roles of treatments and blocks (Shah and



**Fig. 7** (a) Values of  $\tilde{A}_{HB}$  of the Bayesian optimal two-phase designs obtained under the half-Cauchy distributions. The  $x$  axis shows Example IDs, while the  $y$  axis shows values of  $\tilde{A}_{HB}$ . (b) Percentage change of  $\tilde{A}_{HB}$  of the Bayesian optimal two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $\tilde{A}_{HB}$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The  $y$  axis shows the percentage change of  $\tilde{A}_{HB}$ . (c) Percentage change of  $A_\infty$  of the

Bayesian optimal two-phase designs obtained from  $\gamma_{--}$  and  $\gamma_{++}$ , where  $A_\infty$  is the efficiency of a design computed under the fixed-effects model, and  $A_\infty$  of the Bayesian two-phase design obtained from  $\gamma_{++}$  is used as a reference. The  $y$  axis shows the percentage change of  $A_\infty$ . Notice that  $\gamma_{--} = \{\mathcal{HC}(0, 1), \mathcal{HC}(0, 1)\}$ ,  $\gamma_{-+} = \{\mathcal{HC}(0, 1), \mathcal{HC}(1, 10)\}$ ,  $\gamma_{+-} = \{\mathcal{HC}(1, 10), \mathcal{HC}(0, 1)\}$ , and  $\gamma_{++} = \{\mathcal{HC}(1, 10), \mathcal{HC}(1, 10)\}$

**Table 2** Design parameters for constructing two-phase designs

Example ID	$v$	$r$	$b_1$	$k_1$	$b_2$	$k_2$	$p$	$q$	Design type
1	7	3	7	3	3	7	7	3	BD
2	20	7	20	7	20	7	20	7	BD
3	6	5	10	3	5	6	5	3	ND
4	49	3	7	21	21	7	7	7	ND
5	44	4	16	11	11	16	1	1	RCD
6	68	3	17	12	4	51	1	3	RCD
7	42	2	7	12	4	21	1	3	RCD
8	40	4	8	20	5	32	1	4	RCD
9	12	4	16	3	12	4	4	1	NRCD
10	35	4	20	7	28	5	4	1	NRCD
11	30	4	12	10	10	12	2	2	NRCD
12	10	6	6	10	15	4	3	2	NRCD
13	42	4	28	6	8	21	4	3	NRCD
14	10	6	10	6	15	4	5	2	NRCD
15	20	5	25	4	10	10	5	2	NRCD
16	42	5	35	6	15	14	5	2	NRCD

Notice that BD stands for block design, ND stands for nested design, RCD stands for row-column design, and NRCD stands for nested row-column design

Sinha 1989, Chapters 2 and 3). Theorem 1 provides a useful result to compute a Bayesian balanced block design, while Theorem 2 provides a useful result to compute a Bayesian block design where its dual design is balanced.

**Theorem 1** Consider a balanced incomplete block design with its parameters  $v, r, b$ , and  $k$  in a single phase, where  $b$  and  $k$  are the number of blocks and the block size, respectively, and  $b \geq v$  (Fisher’s inequality). Let  $\gamma$  be the ratio between block and residual error variances. The values of  $A_B$  computed from the three prior distributions can be characterized as follows:

- a uniform distribution  $\gamma \sim \mathcal{U}(l_1, l_2)$ :

$$A_B = \frac{1}{l_2 - l_1} \left( \left( 1 + \frac{\lambda - r}{kr} \right) \gamma - \frac{\lambda - r}{rk^2} \ln|1 + k\gamma| \right) \Big|_{l_1}^{l_2}. \tag{19}$$

- a log-normal distribution  $\gamma \sim \mathcal{LN}(\mu, \psi^2)$ :

$$A_B = \int_0^{+\infty} \frac{1}{\gamma \psi \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\ln(\gamma) - \mu}{\psi} \right)^2} \frac{1}{r} \left( r + \frac{(\lambda - r)\gamma}{1 + k\gamma} \right) d\gamma. \tag{20}$$

- a half-Cauchy distribution  $\gamma \sim \mathcal{HC}(\mu, \psi)$ :

$$\begin{aligned}
 A_B = & \frac{2}{\pi} \arctan\left(\frac{\gamma - \mu}{\psi}\right) \Big|_{\mu}^{+\infty} \\
 & + \frac{(\lambda - r)}{\pi r} \frac{-2\psi \ln(|k\gamma + 1|)}{k(k\psi^2 + k\mu^2 + 2\mu) + 1} \Big|_{\mu}^{+\infty} \\
 & + \frac{(\lambda - r)}{\pi r} \frac{\psi \ln((\gamma - \mu)^2 + \psi^2)}{k(k\psi^2 + k\mu^2 + 2\mu) + 1} \Big|_{\mu}^{+\infty} \\
 & + \frac{(\lambda - r)}{\pi r} \frac{2\left(k\psi^2 + k\mu^2 + \mu\right) \arctan\left(\frac{\gamma - \mu}{\psi}\right)}{k(k\psi^2 + k\mu^2 + 2\mu) + 1} \Big|_{\mu}^{+\infty},
 \end{aligned} \tag{21}$$

where  $\lambda = \frac{r(k-1)}{v-1}$ ,  $\arctan(x)$  is the arctangent of  $x$ , and  $\ln(x)$  is the natural logarithm of  $x$ .

**Theorem 2** Consider a block design with its parameters  $v, r, b$ , and  $k$  in a single phase, where its dual design with parameters  $\tilde{v} = b, \tilde{r} = k, \tilde{b} = v, \tilde{k} = r$  is a balanced incomplete block design, and  $\tilde{b} \leq \tilde{v}$  (Fisher’s inequality). Let  $\gamma$  be the ratio between block and residual error variances. The values of  $A_B$  computed from the three prior distributions can be characterized as follows:

- a uniform distribution  $\gamma \sim \mathcal{U}(l_1, l_2)$ :

$$A_B = \frac{(\alpha_1\alpha_3 - \alpha_1\alpha_2)(\ln|\alpha_3\gamma + \alpha_1|) + \alpha_2\alpha_3\gamma}{\alpha_3^2(l_2 - l_1)} \Big|_{l_1}^{l_2}. \tag{22}$$

- a log-normal distribution  $\gamma \sim \mathcal{LN}(\mu, \psi^2)$ :

$$A_B = \int_0^{+\infty} \frac{1}{\gamma\psi\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(\gamma)-\mu}{\psi}\right)^2} \frac{\alpha_1 + \alpha_2\gamma}{\alpha_1 + \alpha_3\gamma} d\gamma. \tag{23}$$

- a half-Cauchy distribution  $\gamma \sim \mathcal{HC}(\mu, \psi)$ :

$$\begin{aligned}
 A_B = & \frac{1}{\pi} \frac{2\alpha_1(\alpha_3 - \alpha_2)\psi \ln(|\alpha_3\gamma + \alpha_1|)}{\alpha_3^2\psi^2 + \alpha_3^2\mu^2 + 2\alpha_1\alpha_3\mu + \alpha_1^2} \Big|_{\mu}^{+\infty} \\
 & - \frac{1}{\pi} \frac{\alpha_1(\alpha_3 - \alpha_2)\psi \ln((\gamma - \mu)^2 + \psi^2)}{\alpha_3^2\psi^2 + \alpha_3^2\mu^2 + 2\alpha_1\alpha_3\mu + \alpha_1^2} \Big|_{\mu}^{+\infty} \\
 & + \frac{2}{\pi} \frac{(\alpha_2\alpha_3\psi^2 + \alpha_2\alpha_3\mu^2) \arctan\left(\frac{\gamma - \mu}{\psi}\right)}{\alpha_3^2\psi^2 + \alpha_3^2\mu^2 + 2\alpha_1\alpha_3\mu + \alpha_1^2} \Big|_{\mu}^{+\infty} \\
 & + \frac{2}{\pi} \frac{(\alpha_1\alpha_3\mu + \alpha_1\alpha_2\mu + \alpha_1^2) \arctan\left(\frac{\gamma - \mu}{\psi}\right)}{\alpha_3^2\psi^2 + \alpha_3^2\mu^2 + 2\alpha_1\alpha_3\mu + \alpha_1^2} \Big|_{\mu}^{+\infty},
 \end{aligned} \tag{24}$$

where  $\alpha_1 = r(v - 1)(b - 1)$ ,  $\alpha_2 = (v - 1)kb(r - 1)$ , and  $\alpha_3 = k(b(v - b)(r - 1) + r(b - 1)^2)$ .

The proofs of Theorems 1 and 2 can be found in Supplementary Sections C and D.

### 8.1 Application

In this section, we apply Theorem 1 to compute  $A_B$  for Example 1. This is because the design parameters of this example are  $v = 7, r = 3, b_1 = 7$  and  $k_1 = 3$ , which satisfy the necessary condition of a balanced incomplete design since  $\lambda = r(k_1 - 1)/(v - 1) = 1$ . From Theorem 1, it is easy to verify that  $A_\gamma = \frac{1}{r}\left(r + \frac{(\lambda-r)\gamma}{1+k_1\gamma}\right) = \frac{3+7\gamma}{3+9\gamma}$ . The values of  $A_B$  can be computed according to Eqs. (19)–(21).

### 9 Conclusion

In this paper, we presented a general framework to search for Bayesian two-phase designs based on the variant of hill climbing studied in Vo-Thanh and Piepho (2022). We considered three families of distributions, i.e., uniform, log-normal, and half-Cauchy distributions. We considered Gauss quadrature rules to approximate the Bayesian criterion. This reduces the computational expense of searching for Bayesian optimal two-phase designs. Furthermore, we proposed two theorems in the case when two-phase designs can be constructed from balanced incomplete block designs. These theorems can be used as upper bounds for  $A_\gamma$  of two-phase designs under the linear mixed model. Such upper bounds have not yet received much attention in the literature.

An important novel feature of our approach is that it can handle Bayesian two-phase designs for various parameter settings and can find an efficient Bayesian two-phase design within a reasonable time. This is because we used the Gauss quadrature rules to approximate the Bayesian criterion where it is adequate to use 10 nodes to get a good approximated value of  $A_B$  for the three prior distributions considered in this paper. Furthermore, we applied Gauss quadrature rules for an arbitrary weight (based on a half-Cauchy distribution) that has not yet been considered in the current literature on Bayesian two-phase designs.

It is worthy to mention that if we have informative prior information for the variance components, our search framework can also be applied, replacing  $p(\gamma)$  by the informative prior information and  $\mathcal{D}$  by the specific domain. This should be quite common in practice, because historical data on previous trials of two-phase experiments as the one that is planned is often available. From this, we should be able to set up an informative prior for variance components. Here, the main bottleneck is the assembly of the historical data in an accessible database, which might require substantial effort in practice.

Even though we have studied several prior distributions, in general, it is not straightforward to propose a recommendation for selecting a right prior distribution in the case where researchers have no specific prior information about the value of variance components. In this case, our recommendation is

to use a flat prior distribution (e.g., a uniform distribution) in searching for two-phase optimal designs. Alternatively, we could perform a simulation study where we obtain many designs for various prior distributions. This could be done using our proposed framework. Based on the result, we can select the best design that is optimal for all various prior distributions. It is also acknowledged that an experiment may target multiple responses, in which case a multivariate model could be considered, but such an extension is beyond the scope of this paper.

We suggest several subjects that merit further research. First, it is possible to extend our proposed approach to handle two-phase designs with more than two blocking factors and/or one treatment factor. The first challenge is to identify a blocking structure involving several blocking factors and a treatment structure involving several treatment factors. The second challenge is to derive a simple form for the treatment information matrix so that update formulae can be proposed. This is very crucial to avoid computational burden when searching for a large two-phase design. The last challenge is to approximate a Bayesian criterion. This can be done using quadrature approaches. However, it might take a great amount of time to search for a large two-phase design. For example, the current nested row-column pattern can be extended to two-phase designs where there are two blocking factors in Phase 1 and one blocking factor in Phase 2. The remaining challenge here is to derive a simple form so that the treatment information matrix can be computed easily. This could be done by extending Proposition 1. Another example is that when there are two treatment factors and two blocking factors, it might be possible that two treatment factors can be combined using a factorial structure. Then, the challenge is to derive a simple form for the treatment information matrix. Second, our proposed approximation of the Bayesian criterion can be adapted to search for block designs and row-column designs under linear mixed models. Third, even though we limited our attention to three families of prior distributions, the quadrature rules can be applied to different prior distributions. Particularly, the Gauss–Chebyshev–Lanczos approach could be employed here to get nodes and weights for an arbitrary weight function if a quadrature rule for this weight function is unavailable. Fourth, for some design parameters, it might be possible to obtain better two-phase designs with a complex blocking pattern, compared to those obtained under a nested row-column pattern. It will take a great amount of computing time, however, to search for a Bayesian two-phase design in this case. This is because efficient update formulae will not be available. However, this option might be worthy to look at in further research, although we conjecture that the efficiencies of these designs will not differ too much, compared to those obtained under a nested row-column pattern.

In this paper, we have not considered Jeffreys's prior. One of the key features of this prior is that it is invariant under a change of coordinates for the parameter vector  $\boldsymbol{\gamma}$ . This makes it of special interest for use with scale parameters. It is, however, rather complex to directly compute Jeffreys's prior since it requires the square root of the determinant of the Fisher information matrix. Therefore, it would be practical to propose an approximation for this prior. This could be done by using the work of Natarajan and Kass (2000) who considered an approximation for Jeffreys's prior.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11222-022-10126-x>.

**Acknowledgements** We would like to acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) via grant PI 377/19-2. We are further grateful to an associate editor and two anonymous referees for their invaluable comments.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Appleby, J.S., Blake, D.V., Newman, E.A.: Techniques for producing school timetables on a computer and their application to other scheduling problems. *Comput. J.* **3**, 237–245 (1961)
- Atkinson, A.C., Donev, A.N., Tobias, R.D.: *Optimum Experimental Designs*, with SAS, vol. 34. Oxford University Press, Oxford (2007)
- Berntsen, J., Espelid, T.O., Genz, A.: An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Trans. Math. Softw.* **17**, 437–451 (1991)
- Brien, C.J.: Analysis of variance tables based on experimental structure. *Biometrics* **39**, 53–59 (1983)
- Brien, C.J.: Multiphase experiments in practice: A look back. *Aust. N. Z. J. Stat.* **59**, 327–352 (2017)
- Brien, C.J.: Multiphase experiments with at least one later laboratory phase. II. Nonorthogonal designs. *Aust. N. Z. J. Stat.* **61**, 234–268 (2019)
- Brien, C.J., Harch, B.D., Correll, R.L., Bailey, R.A.: Multiphase experiments with at least one later laboratory phase. I. Orthogonal designs. *J. Agric. Biol. Environ. Stat.* **16**, 422–450 (2011)
- Chaloner, K., Verdinelli, I.: Bayesian experimental design: a review. *Stat. Sci.* **10**, 273–304 (1995)

- Consonni, G., Fouskakis, D., Liseo, B., Ntzoufras, I.: Prior distributions for objective Bayesian analysis. *Bayesian Anal.* **13**, 627–679 (2018)
- Curnow, R.N.: The analysis of a two phase experiment. *Biometrics* **15**, 60–73 (1959)
- Davis, P.J., Rabinowitz, P.: *Methods of Numerical Integration*. Dover Publications, Mineola (2007)
- Gautschi, W.: *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press, Oxford (2004)
- Gelman, A.: Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1**, 515–534 (2006)
- Genz, A.C., Malik, A.A.: Remarks on Algorithm 006: An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *J. Comput. Appl. Math.* **6**, 295–302 (1980)
- Ghosh, M.: Objective priors: An introduction for frequentists. *Stat. Sci.* **26**, 187–202 (2011)
- Glover, F.: Tabu search—part I. *ORSA J. Comput.* **1**, 190–206 (1989)
- Golub, G.H., Welsch, J.H.: Calculation of Gauss quadrature rules. *Math. Comput.* **23**, 221–230 (1969)
- Goos, P., Mylona, K.: Quadrature methods for Bayesian optimal design of experiments with nonnormal prior distributions. *J. Comput. Graph. Stat.* **27**, 179–194 (2018)
- Gotwalt, C.M.: Addendum to “Fast computation of designs robust to parameter uncertainty for nonlinear settings”. *Technometrics* **52**, 137–137 (2010)
- Gotwalt, C.M., Jones, B.A., Steinberg, D.M.: Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* **51**, 88–95 (2009)
- Jarrett, R.G., Ruggiero, K.: Design and analysis of two-phase experiments for gene expression microarrays-part I. *Biometrics* **64**, 208–216 (2008)
- Johnson, S.G.: Accurate solar-power integration: solar-weighted Gaussian quadrature. arXiv preprint. [arXiv:1912.06870](https://arxiv.org/abs/1912.06870) (2019)
- Johnson, S.G.: HCubature.jl: a pure-Julia implementation of multidimensional “h-adaptive” integration. <https://github.com/JuliaMath/HCubature.jl> (2022a)
- Johnson, S.G.: QuadGK.jl: one-dimensional numerical integration in Julia using adaptive Gauss-Kronrod quadrature. <https://github.com/JuliaMath/QuadGK.jl> (2022b)
- Kass, R.E., Wasserman, L.: The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* **91**, 1343–1370 (1996)
- Kirkpatrick, S., Gelatt Jr, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**, 671–680 (1983)
- Kythe, P.K., Schäferkötter, M.R.: *Handbook of Computational Methods for Integration*. CRC Press, New York (2004)
- Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand.* **45**, 255–282 (1950)
- Liu, H., Wasserman, L.: *Statistical Machine Learning by Han Liu and Larry Wasserman* (Unpublished Book). <http://www.stat.cmu.edu/~larry/=sml/Bayes.pdf> (2022)
- McIntyre, G.A.: Design and analysis of two phase experiments. *Biometrics* **11**, 324–334 (1955)
- Mylona, K., Goos, P., Jones, B.: Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation. *Technometrics* **56**, 132–144 (2014)
- Natarajan, R., Kass, R.E.: Reference Bayesian methods for generalized linear mixed models. *J. Am. Stat. Assoc.* **95**, 227–237 (2000)
- Olver, S.: Quasi matrices, orthogonal polynomials, and Lanczos. <http://approximatelyfunctioning.blogspot.com/2020/09/quasi-matrices-orthogonal-polynomials.html#footnote-2> (2022)
- Searle, S.R., Casella, G., McCulloch, C.E.: *Variance Components*. Wiley, Hoboken (2009)
- Shah, K.R., Sinha, B.K.: *Theory of Optimal Designs*, vol. 54. Springer, Berlin (1989)
- Smith, A.B., Lim, P., Cullis, B.R.: The design and analysis of multi-phase plant breeding experiments. *J. Agric. Sci.* **144**, 393–409 (2006)
- Smith, A.B., Butler, D.G., Cavanagh, C.R., Cullis, B.R.: Multi-phase variety trials using both composite and individual replicate samples: a model-based design approach. *J. Agric. Sci.* **153**, 1017–1029 (2015)
- Smolyak, S.A.: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR* **148**, 1042–1045 (1963)
- Stoer, J., Bulirsch, R.: *Introduction to Numerical Analysis*, vol. 12. Springer, Berlin (2013)
- Trefethen, L.N., Bau, D.: *Numerical Linear Algebra*, vol. 50. SIAM, Philadelphia (1997)
- Vo-Thanh, N., Bailey, R.A., Brien, C.J., Piepho, H.-P.: Construction of two-phase designs for experiments with a single blocking factor in each phase. Working paper (2022)
- Vo-Thanh, N., Piepho, H.-P.: Generating designs for comparative experiments with two blocking factors. Working paper (2022)
- Wood, J.T., Williams, E.R., Speed, T.P.: Non-orthogonal block structure in two-phase designs. *Aust. J. Stat.* **30**, 225–237 (1988)
- Yates, F.: The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen.* **9**, 136–156 (1939)

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