Optimal Experimental Designs in Multiple-Group Mixed Models

Habilitation

zur Erlangung des akademischen Grades

doctor rerum naturalium habilitatus

(Dr. rer. nat. habil.)

von Dr. Maryna Prus

geb. am 17.08.1982 in Gorlowka, Ukraine

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe
Prof. Dr. Stefanie Biedermann
Prof. Dr. France Mentré
Prof. Dr. Hans-Peter Piepho

eingereicht am: 10.03.2023
Verteidigung am: 8.11.2023

Summary

The present habilitation consists for five scientific papers: Prus (2022a), Prus (2023b), Prus (2023a), Prus and Filová (2023) and Prus and Piepho (2021). Also four earlier works (Prus (2020), Prus *et al.* (2020), Prus (2022b) and Harman and Prus (2018)) are related to the same topic.

In the habilitation optimal designs in multiple-group mixed models were investigated. In these models observational units (for simplicity called 'individuals') are allocated to several groups. Statistical properties of the individuals may differ from group to group. For each observational unit many observations are possible, repeated measurements are allowed. The response depends on unknown fixed and (unit-specific or 'individual') random effects.

In the present work the main focus is on optimal designs for the prediction of random effects or of linear combinations of fixed and random effects, based on the related best linear unbiased predictors (BLUP). Also for the estimation of fixed effects (BLUE) some new aspects, especially for the computation of designs, were considered. The resulting design criteria turned out to depend on several designs (group-designs) simultaneously. For such design criteria the general equivalence theorem proposed by Kiefer (1974) cannot be used.

In Prus (2022a) extended versions of the general equivalence theorem are provided. The (new) equivalence theorems for the multiple-design problems are based on the assumptions of convexity and differentiability of the design criteria. For the design problems with finite experimental regions optimality conditions were formulated with respect to the designs themselves (Theorem 1). For the case where the optimality criteria depend on the designs via information (or moment) matrices only, optimality conditions were formulated with respect to the information matrices (Theorem 2).

Prus (2023b) investigated optimal designs for the prediction of the individual random parameters and the group difference in two-groups random coefficient regression (RCR) models with multivariate response. A solution for optimal approximate designs is given in the form of optimality conditions for the linear and D-criteria. The optimality criteria depend on the designs via the (group-) information matrices, hence, the optimality conditions could be obtained using Theorem 2 from Prus (2022a).

Prus (2023a) considered optimal designs for the prediction in the RCR models in which only one observation per observational unit is possible. An analytical solution is given for optimal designs for the prediction of random effects for a group of selected individuals. The optimal designs have been obtained for individual effects via their arithmetic mean. The design criteria turned out to depend on two (group-) designs simultaneously. The solution is given by optimality conditions, which results from Theorem 2 in Prus (2022a).

Prus and Filová (2023) discussed the problem of the computation of optimal and efficient designs for fixed effects in the multiple-group mixed models considered in Prus (2023b). In this work equi- and invariance properties of designs have been analyzed, the proposed computational method is based on the algorithm by Harman *et al.* (2016) and allows for additional constrains on design.

Prus and Piepho (2021) considered allocation of trials to sub-regions in multi-environment crop variety testing. A linear mixed model with random genotype effects has been assumed. For this problem Bayesian optimal designs for an adjusted covariance matrix of genotype effects turn out to be optimal.

Some other related results are presented in the earlier works Prus (2020), Prus (2022b), Prus *et al.* (2020) and Harman and Prus (2018). Prus (2020) and Prus (2022b) investigated optimal designs in particular multiple-group RCR models with several treatment groups and a control group or two treatment groups, respectively. An application in medical research for simple

RCR models has been discussed in Prus *et al.* (2020). Harman and Prus (2018) proposed a computational approach for optimal designs with respect to the Compound Bayes Risk Criterion (CBRC). This approach has been used for computing Bayesian optimal designs in Prus (2023b) and Prus and Piepho (2021).

Zusammenfassung

Die vorliegende Habilitation besteht aus fünf wissenschaftlichen Arbeiten: Prus (2022a), Prus (2023b), Prus (2023a), Prus and Filová (2023) und Prus and Piepho (2021). Auch vier frühere Arbeiten (Prus (2020), Prus *et al.* (2020), Prus (2022b) und Harman and Prus (2018)) beziehen sich auf das Habilitationsthema.

In der Habilitation wurden optimale Versuchspläne (Designs) in gemischten Mehrgruppenmodellen untersucht. In diesen Modellen werden Beobachtungseinheiten ('Individuen') den Gruppen zugeordnet. Statistische Eigenschaften der Individuen können sich von Gruppe zu Gruppe unterscheiden. Für jede Beobachtungseinheit sind viele Beobachtungen möglich, wiederholte Messungen sind zugelassen. Die Antwortvariablen hängen von unbekannten festen und (individuellen) zufälligen Effekten ab.

Der Schwerpunkt der vorliegenden Arbeit liegt an optimalen Designs für die Vorhersage zufälliger Effekte bzw. linearer Kombinationen von festen und zufälligen Effekten, basierend auf der zugehörigen besten linearen erwartungstreuen Vorhersage (BLUP). Auch für die Schätzung von festen Effekten (BLUE) wurden einige neue Aspekte, insbesondere für die Berechnung von Designs, berücksichtigt. Die resultierenden Designkriterien hängen gleichzeitig von mehreren Designs (Gruppendesigns) ab. Für solche Designkriterien kann der von Kiefer (1974) vorgeschlagene allgemeine Äquivalenzsatz nicht verwendet werden.

In Prus (2022a) sind erweiterte Versionen des allgemeinen Äquivalenzsatzes dargestellt. Die (neuen) Äquivalenztheoreme für die Multiple-Design-Probleme basieren auf den Annahmen der Konvexität und Differenzierbarkeit der Designkriterien. Für die Designprobleme mit endlichen Versuchsbereichen wurden Optimalitätsbedingungen bezüglich der Designs selbst formuliert (Theorem 1). Für den Fall, wenn die Optimalitätskriterien nur über Informations- (oder Moment-) Matrizen von den Designs abhängen, wurden Optimalitätsbedingungen in Bezug auf die Informationsmatrizen formuliert (Theorem 2).

Prus (2023b) hat optimale Versuchspläne für die Vorhersage der individuellen zufälligen Parameter und der Gruppendifferenz in Zwei-Gruppen-Regressionsmodellen mit zufälligen Effekten (random coefficient regression, RCR) untersucht. Eine analytische Lösung für optimale approximative Versuchspäne ist in Form von Optimalitätsbedingungen für die L- und D-Kriterien gegeben. Die Optimalitätskriterien hängen von den Designs über die Gruppen-Informationsmatrizen ab. Die Optimalitätsbedingungen wurden mithilfe von Theorem 2 in Prus (2022a) erhalten.

Prus (2023a) hat optimale Versuchspläne für die Vorhersage in den RCR-Modellen, in denen nur eine Beobachtung pro Beobachtungseinheit möglich ist, untersucht. Eine analytische Lösung für optimale Designs zur Vorhersage zufälliger Effekte für eine Gruppe ausgewählter Individuen wurde vorgeschlagen. Die optimalen Designs für individuelle Effekte wurden über ihren arithmetischen Mittelwert erhalten. Die resultierenden Designkriterien hängen gleichzeitig von zwei Gruppen-Designs ab. Die Lösung ist gegeben in Form der Optimalitätsbedingungen, die sich aus Theorem 2 in Prus (2022a) ergeben.

In Prus and Filová (2023) wurde das Problem der Berechnung optimaler und effizienter Designs für feste Effekte in den in Prus (2023b) betrachteten gemischten Mehrgruppenmodellen berücksichtigt. In dieser Arbeit wurden Äqui- und Invarianzeigenschaften von Designs analysiert, die vorgeschlagene Berechnungsmethode basiert auf dem Algorithmus von Harman *et al.* (2016) und berücksichtigt zusätzliche Einschränkungen für Designs.

Prus and Piepho (2021) haben die Zuordnung der Feldversuche den Subregionen bei der Prüfung von Kulturpflanzensorten in mehreren Umgebungen betrachtet. Es wurde ein lineares gemischtes Modell mit zufälligen Genotypeffekten verwendet. Für dieses Versuchsplanungsproblem zeigen sich Bayes'sche optimale Designs für eine angepasste Kovarianzmatrix von Genotyp-

effekten optimal.

Einige weitere Ergebnisse, die für das Habilitationsthema relevant sind, wurden in den früheren Arbeiten Prus (2020), Prus (2022b), Prus *et al.* (2020) und Harman and Prus (2018) dargestellt. In Prus (2020) wurden optimale Designs in speziellen Mehrgruppen-RCR-Modellen mit mehreren Behandlungsgruppen und einer Kontrollgruppe untersucht. In Prus (2022b) ging es um Modelle mit zwei Behandlungsgruppen. Eine Anwendung in der medizinischen Forschung für einfache RCR-Modelle wurde in Prus *et al.* (2020) vorgeschlagen. Harman and Prus (2018) haben eine Berechnungsmethode für optimale Designs für das Compound Bayes Risk Criterion (CBRC) entwickelt. Diese Methode wurde zur Berechnung optimaler Bayes'scher Designs in Prus (2023b) und Prus and Piepho (2021) verwendet.

Extended Summary of Habilitation

Optimal Experimental Designs in Multiple-Group Mixed Models

of

Maryna Prus

at

Faculty of Mathematics

Otto von Guericke University of Magdeburg

Contents

1 Introduction

The subject of this work is multiple-group linear mixed models. These models are popular in many fields of statistical applications, for example in medical research, pharmacokinetic, agricultural sciences (see Fedorov and Jones (2005), Schmelter (2007), Kleinknecht *et al.* (2013)). In these models observational units (for simplicity called 'individuals') are allocated to several groups. Statistical properties of the individuals (for example variances and covariances of individual random effects) may differ from group to group. For each observational unit many observations are possible, repeated measurements are allowed. The response depends on unknown fixed and (usually unit-specific or 'individual') random effects.

The purpose of the work is to determine optimal designs (experimental settings) for the estimation / prediction of unknown parameters. The focus may be on fixed or random parameters as well as on their linear combinations. The designs are generally considered here for the best linear unbiased predictor (BLUP) (see e. g. Henderson (1975)). The design optimization problem for the best linear unbiased estimators (BLUE) for fixed parameters is well investigated in the literature (see e. g. Schmelter (2007). However some new aspects, especially for the computation of designs, are being considered in the present work.

For determining optimal designs, the mean squared error (MSE) matrix of the BLUP (or the covariance matrix of the BLUE) is being minimized via design criteria. The design criteria are functions of the MSE (or covariance) matrix as for example the determinant (D-criterion), the largest eigenvalue (E-criterion) or the trace (A-criterion) (see e. g. Fedorov and Hackl (1997), Pázman (1986), Pukelsheim (1993)). If the criterion of interest is convex and differentiable, the analytical solution may result in optimality conditions, that follow from the general equivalence theorem (see e.g. Kiefer (1974)). In general optimal designs do not follow directly from the optimality conditions. Therefore, a variety of computational methods have been developed (see e.g. Sagnol and Harman (2015), Harman *et al.* (2016)).

The analytical solutions for optimal designs available in the literature are proposed in general for design criteria depending on one design only (see e. g. Gladitz and Pilz (1982), Prus and Schwabe (2016)). In the multiple-group models, however, different designs for different groups are possible. Consequently, the resulting design criteria usually turn out to depend on several (group-) designs simultaneously (see e.g. Schmelter (2007)). For such design criteria, the general equivalence theorem cannot be used directly. Therefore, extended versions for the multiple-design problems have been developed in Prus (2022a). Results of this paper have been used for formulating optimality conditions for the multiple-group models discussed in Prus (2023b) and Prus (2023a). Prus (2023b) considered the two-groups random coefficient regression models with multivariate response. In Prus (2023a) optimal designs for the models with one observation per observational unit were investigated. Both papers provide analytical solutions for optimal designs in form of optimality conditions. An application in multi-environment trials has been considered in Prus and Piepho (2021).

The most methods for the computation of optimal or highly efficient designs proposed so far were developed for classical design criteria in the models without random effects (see e. g. Harman and Filová (2019)). Some methods are developed for non-linear mixed models (see e. g. Nyberg *et al.* (2012), Dumont *et al.* (2018)). These methods, however, do not allow for additional constraints (such as budget or other types of resources) that may arise in the experiment, and are focused on the approximate designs or consider only the D-criterion. The problem of the computation of optimal and efficient designs in the multiple-group mixed models has been discussed in Prus and Filová (2023). In this work equi- and invariance properties of designs are analyzed, the proposed computational method for exact designs is based on the algorithm by Harman *et al.* (2016) and allows for constrained design problems.

Some other related results are presented in earlier works Prus (2020), Prus (2022b), Prus *et al.* (2020) and Harman and Prus (2018). Prus (2020) and Prus (2022b) investigated optimal designs in particular multiple-group RCR models with several treatment groups and a control group or two treatment groups, respectively. An application in medical research for simple RCR models has been discussed in Prus *et al.* (2020). Harman and Prus (2018) proposed a computational approach for optimal designs with respect to the Compound Bayes Risk Criterion (CBRC), which includes the Bayesian linear criteria as a particular case. This approach has been used for computing optimal designs in Prus (2023b) and Prus and Piepho (2021).

The structure of the habilitation is presented explicitly in the next section. The main results are summarized in Section 3. The earlier works are briefly discussed in Section 4. The authors' contributions are given for all joint works directly after their descriptions. Section 5 includes a short discussion and an outlook. The scientific papers themselves are added after the discussion.

2 Structure of the Habilitation

The present commutative habilitation consists of the following scientific works:

- Prus, M. (2022a). Equivalence theorems for multiple-design problems with application in mixed models. Journal of Statistical Planning and Inference, 217, 153–164. (Accepted manuscript attached).
  https://doi.org/10.1016/j.jspi.2021.07.012
- Prus, M. (2023a). Optimal designs for prediction in random coefficient regression models with one observation per individual. Statistical Papers.
  https://doi.org/10.1007/s00362-023-01440-1
- Prus, M. (2023b). Optimal designs for prediction of random effects in two groups linear mixed models. Journal of Multivariate Analysis, 198. (Preprint attached).
  https://doi.org/10.1016/j.jmva.2023.105212
- Prus, M. and Filová, L. (2023). Computational aspects of experimental designs in multiple-group mixed models. Statistical Papers.
  https://link.springer.com/article/10.1007/s00362-023-01416-1
- Prus, M. and Piepho, H.-P. (2021). Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing. Journal of Agricultural, Biological and Environmental Statistics, 26, 267–288. (Accepted manuscript attached).
  https://doi.org/10.1007/s13253-020-00426-y

Other related publications:

- Harman, R., Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. Statistics & Probability Letters, 137, 135–141.
- Prus, M. (2020). Optimal designs in multiple-group random coefficient regression models. TEST, 29, 233-254.
- Prus, M. (2022a). Optimal designs for prediction in two treatment groups random coefficient regression models. In J. Pilz, T. A. Oliveira, K. Moder, and C. P. Kitsos, editors, Mindful Topics on Risk Analysis and Design of Experiments - Selected contributions from ICRA8, Vienna. Springer.
- Prus, M., Benda, N., and Schwabe, R. (2020). Optimal design in hierarchical random effect models for individual prediction with application in precision medicine. Journal of Statistical Theory and Practice, 14, 24.

3 Main Results of the Habilitation

In this section the main results of the habilitation, published in the five scientific papers listed in Section 2 (Prus (2022a), Prus (2023b), Prus (2023a), Prus and Filová (2023) and Prus and Piepho (2021)), will be described.

3.1 Equivalence Theorems for Multiple-Design Problems

Prus (2022a) considered design problems with optimality criteria depending on several designs simultaneously. This work can be seen as the basic paper of the habilitation. It was motivated by the fact that in many multiple-group mixed models, especially in the random coefficient regression (RCR) models with several groups of observational units, design criteria depend on several (group-) designs (see e. g. Schmelter (2007)). For such design criteria the general equivalence theorem proposed by Kiefer (1974) cannot be used. Therefore, for solving design optimization problems in the multiple-group models an extended version of this theorem had to be developed.

Prus (2022a) formulated equivalence theorems for the multiple-design problems based on the assumptions of convexity and differentiability of the design criteria. For the design problems with finite experimental regions optimality conditions were formulated with respect to the designs themselves (Theorem 1). For the case where the optimality criteria depend on the designs via information (or moment) matrices only, optimality conditions were formulated with respect to the information matrices (Theorem 2). The proposed optimality conditions were applied to the multiple-group RCR models: Optimal designs for the estimation of fixed effects have been discussed. As a by-product the following result has been obtained: If all observational units have the same statistical properties and there are no group-specific design-restrictions, optimal designs in the single-group models are also optimal as group-designs in the multiple-group models. In this case the group sizes have no influence on the designs. However, if the numbers of observations differ from group to group, optimal group-designs may depend on the numbers of observations and the group sizes. This behavior has been illustrated by the example of straight line regression models.

The results from this work have been used directly for formulating optimality conditions for the design problems considered in Prus (2023b) and Prus (2023a) (see Sections 3.2 and 3.3).

3.2 Optimal Designs in Multiple-Group RCR Models

Prus (2023b) investigated optimal designs for the prediction of individual random parameters and the group difference in the two-groups RCR models with multivariate response. Note that in this and other works observational units are often called for simplicity 'individuals'. However, in reality observational units may be (besides people or animals) studies, centers, plots, etc. The solution for optimal approximate designs in the two-groups RCR models is given in the form of optimality conditions for the linear and D-criteria. As the optimality criteria depend on the designs via the (group-) information matrices, the optimality conditions could be obtained using Theorem 2 from Prus (2022a) (see Section 3.1). Optimal designs for the prediction of the group differences coincide, in the particular case of the same regression function and the same design region for both groups, with the Bayesian optimal designs for an adjusted covariance matrix of random effects. Optimal approximate and exact designs for Bayesian linear criteria were computed using "OptimalDesign" package in R (see Harman and Filová (2019)) and the approach proposed by Harman and Prus (2018) (see Section 4) for Bayesian non-singular linear criteria.

3.3 Optimal Designs in Multiple-Group Models with one Observation per Individual

Prus (2023a) considered optimal designs for the prediction of random effects in the RCR models in which only one observation per observational unit (individual) is possible. For only two observational units, these models may be in general considered as a particular case of the RCR models discussed in Prus (2023b) (Prus (2022a)) with one individual per group. However, the theory developed in that paper for optimal designs for the prediction, is based on the condition of full column rank design matrices for both groups, i. e. the number of observations for each observational unit should not be smaller than the number of unknown fixed parameters. This condition is obviously not satisfied in the case of only one observation per individual. Therefore, the results obtained in Prus (2023b) could not be used.

In Prus (2023a) analytical solution is given for optimal designs for the prediction of random effects for a group of selected individuals. The optimal designs have been obtained for individual effects (individual deviations from the population parameter) via their arithmetic mean. The design criteria turned out to depend on two (group-) designs simultaneously. The solution is given by optimality conditions, which result from Theorem 2 in Prus (2022a) (see Section 3.1). The analytical results are illustrated by simple examples of straight line regression models.

3.4 Computational Aspects for Optimal Designs

An analytical solution for optimal designs for the estimation of fixed effects in the multiple-group RCR models has been obtained in Prus (2022a). (Optimal designs for prediction of random effects in these models were considered in Prus (2023b).) This solution was given in form of optimality conditions, that allow to check if a candidate design is optimal or not. However, the question arises how the candidate designs should be chosen. The answer to this question is in general far from obvious. Hence, the problem of the computation of optimal or highly efficient designs remained open.

Prus and Filová (2023) focused on computational aspects for optimal designs in the multiple-group mixed models. In the first step of this work equi- and invariance properties of approximate optimal designs were considered. These properties have been used to fix the support points and, consequently, to reduce the number of unknown variables in first- and second-order models on a symmetric square. These results were used to determine optimal designs analytically in a few particular cases, as shown in the examples. In the second step, a modified version of the algorithm of Harman *et al.* (2016) has been developed for the computation of exact designs. As it has been illustrated by several examples, this modified algorithm is a useful tool for such computations, even in the cases where there are several nontrivial constraints on the design.

Authors' contributions: Prus, M. 60%, **Filová, L.** 40%**.**

3.5 Application in Multi-Environment Trials

Prus and Piepho (2021) considered allocation of trials to sub-regions in multi-environment crop variety testing. A linear mixed model with random genotype effects has been assumed. Due to its complicated covariance structure, this model is not a particular case of the RCR models considered in Prus (2023b). The allocation of locations for different sub-regions with respect to the prediction of genotype effects and their pairwise linear contrasts was optimized for A- and particular linear (weighted A-) criteria. The proposed approach is based on the method of best linear unbiased prediction (BLUP). For this problem, Bayesian optimal designs for an adjusted covariance matrix of genotype effects turned out to be optimal. In the example, two kinds of

models with respect to the covariance structure were considered: first-order factor-analytic and compound symmetry. The resulting designs in both cases depend on the covariance structure, observational errors variance and the total number of locations in all sub-regions. The only exception is the standard A-criterion for the compound symmetry covariance structure: in this case, balanced designs are optimal. For the weighted A-criterion the weighted design turned out to be highly efficient for compound symmetry. However, one should be more careful with the choice of design in case of the factor-analytic covariance structure, especially if there are large differences between variances of genotype effects for different sub-regions. In this work Bayesian optimal designs have been computed using the approach proposed by Harman and Prus (2018). Linear constraints on designs have been used to avoid the situation with zero locations in some sub-regions.

Authors' contributions: Prus, M. $70-80\%$**, Piepho, H.-P.** $20-30\%$**.**

4 Other Related Publications

This section briefly describes the results of the four earlier works of the author: Prus (2020), Prus *et al.* (2020), Prus (2022b) and Harman and Prus (2018). These works are not directly included in the habilitation. They are, however, related to the same topic. Results from these works were used in the habilitation.

4.1 Some Particular Cases

In Prus (2020) multiple-group RCR models with several treatment groups and a control group have been considered. The A-, D- and E-optimality criteria for the estimation of population parameters and for the prediction of individual treatment effects have been obtained using the covariance matrix of the BLUE (best linear unbiased estimator) and the mean squared error matrix of the BLUP (best linear unbiased predictor), respectively. The optimal designs (optimal group sizes) turned out to be different for the estimation and the prediction and do not coincide with those in the corresponding fixed-effects model (one-way layout).

Prus (2022b) considered RCR models with two treatment groups. The A- and D-optimality criteria for the estimation of the population parameters and the prediction of the individual random effects have been derived. The efficiency of the balanced design, which assigns equal group sizes, is relatively high only for small values of the variances of random effects and decreases with increasing variance.

4.2 Application in Medical Research

Prus *et al.* (2020) discussed the problem of optimal designs in particular multi-center models. In these models individuals are divided into several centers (laboratories, studies, etc.). Then in each center individuals are assigned either to the treatment or to the control group. Only one observation is available per individual. Optimal designs are obtained for the A-criterion of the average of the mean squared error for the prediction of the center specific treatment effects. The performance of the resulting optimal designs is compared with that of a conventional balanced design with respect to the treatment and control allocations. The investigations show that balanced designs are far from being optimal, for instance, if the treatment effects vary strongly as compared to the residual errors, and more subjects should be recruited to the active treatment group.

Authors' contributions: Prus, M. more than 50%**.**

4.3 Computation of Optimal Designs for CBRC

Harman and Prus (2018) developed an approach for the computation of optimal approximate and exact designs with respect to a general Compound Bayes Risk Criterion (CBRC). This more general criterion includes various specific criteria, for example the Bayesian non-singular linear criterion. The proposed approach even permits to solve problems under additional linear constraints on the designs. This approach has been used for the computation of optimal designs in Prus (2023b) and Prus and Piepho (2021).

Authors' contributions: Prus, M. 40%**, Harman, R.** 60%**.**

5 Discussion and Outlook

The present habilitation consists for five scientific papers: Prus (2022a), Prus (2023b), Prus (2023a), Prus and Filová (2023) and Prus and Piepho (2021). Also four earlier works of the author (Prus (2020), Prus *et al.* (2020), Prus (2022b) and Harman and Prus (2018)) are related to the topic. In the habilitation optimal designs in multiple-group mixed models were investigated. In the basic paper Prus (2022a) extended versions of the general equivalence theorem are provided. Prus (2023b) and Prus (2023a) considered optimal designs for the prediction of random effects in two kinds of multiple-group RCR models. Analytical solutions are proposed in form of optimality conditions for candidate designs. Prus and Filová (2023) considered computational aspects of optimal designs for fixed effects in the models considered in Prus (2022a) and Prus (2023b). Prus and Piepho (2021) considered an application in agricultural sciences: allocation of trials to sub-regions in multi-environment crop variety testing. The earlier works Prus (2020) and Prus *et al.* (2020) considered optimal designs in particular cases of the multiple-group RCR models. Prus *et al.* (2020) proposed an application in medical research. Harman and Prus (2018) developed an approach for computing optimal designs for CBRC.

In the present works optimal designs depend on the covariance matrices of random effects in the model, and are therefore locally optimal. In the future research it may be reasonable to consider robust design criteria (as minimax-criterion or maximin-efficient), that are independent of the variance parameters. Also assuming particular covariance structure may be a solution (as for example the compound-symmetry structure in Prus and Piepho (2021)). The problem of computation of designs in solved only for the estimation of fixed effects in the multiple-group RCR models (by Prus and Filová (2023)). For the prediction of random effects (Prus (2023b), Prus (2023a)) this problem remains open (except of some particular cases, where Bayesian designs turned out to be optimal). Computational aspects for optimal designs may be considered in more detail in the future. Moreover, the results obtained in Prus (2023b) and Prus (2023a) cover the situations with only one observation per individual or if the number of observations m is not smaller than the number p of unknown fixed effects. The problem of sparse designs, $1 < m < p$, also remains open. Another reasonable direction of future research is extension of results obtained in Prus and Piepho (2021) for optimal designs in multi-environment trials. In practice, the linear mixed models used in this research should e. g. incorporate the influence of environmental factors and allow for multi-annual experiments.

Acknowledgment

This research was partially supported by Grant SCHW 531/16 of the German Research Foundation (DFG). A part of the research has been performed as the author was visiting Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis in Nice and Comenius University of Bratislava in framework of the postdoc program (PKZ 91672803) of German Academic Exchange Service (DAAD). The last two papers were submitted as the author was a visiting researcher and lecturer in Linköping University, STIMA. The final steps were performed while the author was working as a researcher in Biostatistics Division of Hohenheim University.

References

Dumont, C., Lestini, G., Nagard, H. L., Mentré, F., Comets, E., and Nguyen, T. T. (2018). PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Computer methods and programs in biomedicine*, **156**, 217–229.

Fedorov, V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*. Springer, New York.

Fedorov, V. and Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248.

Gladitz, J. and Pilz, J. (1982). Construction of optimal designs in random coefficient regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, **13**, 371–385.

Harman, R. and Filová, L. (2019). Package 'OptimalDesign'. https://cran.r-project.org/web/packages/OptimalDesign/index.html.

Harman, R. and Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. *Statistics and Probability Letters*, **137**, 135–141.

Harman, R., Bachratá, A., and Filová, L. (2016). Construction of efficient experimental designs under multiple resource constraints. *Appl. Stochastic Models Bus. Ind.*, **32**, 3–17.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–477.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, **2**, 849–879.

Kleinknecht, K., Möhring, J., Singh, K., Zaidi, P., Atlin, G., and Piepho, H.-P. (2013). Comparison of the performance of blue and blup for zoned indian maize data. *Crop Science*, **53**, 1384–1391.

Nyberg, J., Ueckert, S., Strömberg, E. A., Hennig, S., Karlsson, M. O., and Hooker, A. C. (2012). Poped: an extended, parallelized, nonlinear mixed effects models optimal design tool. *Computer methods and programs in biomedicine*, **108**(2), 789–805.

Pázman, A. (1986). *Foundations of optimum Experimental Design*. D. Reidel Publishing Company, New York.

Prus, M. (2020). Optimal designs in multiple group random coefficient regression models. *TEST*, **29**, 233–254.

Prus, M. (2022a). Equivalence theorems for multiple-design problems with application in mixed models. *Journal of Statistical Planning and Inference*, **217**, 153–164.

Prus, M. (2022b). Optimal designs for prediction in two treatment groups random coefficient regression models. In J. Pilz, T. A. Oliveira, K. Moder, and C. P. Kitsos, editors, *Mindful Topics on Risk Analysis and Design of Experiments - Selected contributions from ICRA8, Vienna*. Springer.

Prus, M. (2023a). Optimal designs for prediction in random coefficient regression models with one observation per individual. *Statistical Papers*. https://doi.org/10.1007/s00362-023-01440-1.

Prus, M. (2023b). Optimal designs for prediction of random effects in two groups linear mixed models. *Journal of Multivariate Analysis*, **198**. https://doi.org/10.1016/j.jmva.2023.105212.

Prus, M. and Filová, L. (2023). Computational aspects of experimental designs in multiple-group mixed models. *Statistical Papers*. https://link.springer.com/article/10.1007/s00362-023-01416-1.

Prus, M. and Piepho, H.-P. (2021). Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing. *Journal of Agricultural, Biological and Environmental Statistics*, **26**, 267–288.

Prus, M. and Schwabe, R. (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B*, **78**, 175–191.

Prus, M., Benda, N., and Schwabe, R. (2020). Optimal design in hierarchical random effect models for individual prediction with application in precision medicine. *Journal of Statistical Theory and Practice*, **14**, 24.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Sagnol, G. and Harman, R. (2015). Computing exact D-optimal designs by mixed integer second-order cone programming. *The Annals of Statistics*, (43), 2198–2224.

Schmelter, T. (2007). *Experimental Design For Mixed Models With Application to Population Pharmacokinetic Studies*. Ph.D. thesis, Otto-von-Guericke University Magdeburg.

Equivalence theorems for multiple-design problems with application in mixed models

Maryna Prus

Abstract

In the present paper we consider design criteria which depend on several designs simultaneously. We formulate equivalence theorems based on information matrices (if criteria depend on designs via information matrices) or with respect to the designs themselves (for finite design regions). We apply the obtained optimality conditions to multiple-group random coefficient regression models and illustrate the results by simple examples.

Keywords: Optimal design, optimality condition, multiple-group, mixed model, random coefficient regression, multivariate response

1 Introduction

The subject of this work is multiple-design problems - optimization problems with optimality criteria depending on several designs simultaneously. Such optimality criteria can be, for example, commonly used design criteria for estimation of unknown model parameters in cases when the covariance matrix depends on several designs (see e. g. Fedorov and Jones (2005), Schmelter (2007a). For such criteria the general equivalence theorem proposed in Kiefer (1974) cannot be used directly. In Fedorov and Jones (2005) optimal designs were obtained for specific regression functions. In Schmelter (2007a) particular group-wise identical designs have been discussed.

In this paper we formulate equivalence theorems for two kinds of multiple-design problems: 1) problems on finite experimental regions and 2) problems with optimality criteria depending on designs via information matrices. For both cases we assume the optimality criteria to be convex and differentiable in the designs themselves or in the information matrices, respectively. In case 1) we formulate optimality conditions with respect to the designs directly (as proposed in Whittle (1973) for one-design problems). These results can be useful in situations when design criteria cannot be presented as functions of information matrices (see e. g. Bose and Mukerjee (2015). In case 2) optimality conditions are formulated with respect to the information matrices. Therefore, no additional restrictions of the experimental regions are needed.

We apply the equivalence theorems to multiple-group random coefficient regression (RCR) models. In these models observational units (individuals) are assigned to groups. Within one group a single design (group-design) for all individuals is assumed. Group-designs for individuals from different groups are in general not the same. The assumption of the same design for all individuals within one group may be not feasible for some applications. This simplification is needed for analytical purposes. Most of the commonly used design criteria in multiple-group RCR models are functions of several group-designs. The particular case of these models with one observation per individual has been considered in Graßhoff *et al.* (2012). Mentré *et al.* (1997) and Dumont *et al.* (2018) proposed computational methods for determining D-optimal designs based on the normality assumption. In Prus (2015), ch. 6, models with group-specific mean parameters were briefly discussed. Bludowsky *et al.* (2015), Kunert *et al.* (2010), Lemme *et al.* (2015) and Prus (2020) considered models with particular regression functions and specific covariance structures of random effects. In Entholzner *et al.* (2005) and Prus and Schwabe (2016)

a single design was assumed for all observational units. Other particular cases of multiple-design problems have been considered in Wierich (1989) and Schwabe (1996).

In this work, we propose an analytical approach for determining optimal designs with respect to all convex and differentiable criteria in multiple-design problems. In particular, linear and D-criteria in the multiple-group models are considered in detail. The proposed approach is based on moment assumptions only (no distributional assumptions are needed).

The paper has the following structure: Section 2 provides equivalence theorems for the multiple-design problems. In Section 3 we apply the obtained optimality conditions to the multiple-group RCR models. In Section 4 we illustrate the results by a simple example. The paper is concluded by a short discussion in Section 5.

2 Optimality Conditions for Multiple-Design Problems

We consider a multiple-design problem in which $\xi_1, \ldots, \xi_s$ are probability measures (designs) on experimental regions $\mathcal{X}_1, \ldots, \mathcal{X}_s$, respectively, and ϕ is a design criterion which depends on $\xi_1, \ldots, \xi_s$ simultaneously and has to be minimized. Ξ_i denotes the set of all designs on $\mathcal{X}_i$, $i = 1, \ldots, s$. For any $x_i \in \mathcal{X}_i$, δ_{x_i} denotes the particular design ξ_i with all observations at point x_i. For convenience we use the notation $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_s)$ for a vector of designs $\xi_i \in \Xi_i$, $i = 1, \ldots, s$. Then $\boldsymbol{\xi} \in \Xi$ for $\Xi = \times_{i=1}^{s} \Xi_i$, where "$\times$" denotes the Cartesian product.

As an example, consider the multi-center trials model (see e. g. Fedorov and Jones (2005), where t_{ij} denote the number of patients on treatment j in center i for $i = 1, \ldots, s$ and $j = 1, \ldots, \ell$. Then the design ξ_i may be determined by the numbers of patients for each treatment in the i-th center: $t_{i1}, \ldots, t_{i\ell}$.

In Section 2.1 we consider the multiple-design problems, where all design regions are assumed to be finite. We formulate an equivalence theorem (Theorem 1) with respect to the designs directly.

In Section 2.2 we consider the design criteria depending on designs via information matrices and we propose an equivalence theorem (Theorem 2) based on this structure. In this case no additional restrictions for the experimental regions are needed.

2.1 Optimality conditions in case of finite design regions

In this section we restrict ourselves on the optimization problems on finite design regions: $|\mathcal{X}_i| = k_i < \infty$ for all $i = 1, \ldots, s$. $\phi : \Xi \to (-\infty; \infty]$ denotes a design criterion. We use the notation $\Phi(\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})$ for the directional derivative of ϕ at $\boldsymbol{\xi}$ in direction of $\tilde{\boldsymbol{\xi}}$:

$$\Phi(\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) = \lim_{\alpha \searrow 0} \frac{1}{\alpha} \Big(\phi((1-\alpha)\boldsymbol{\xi} + \alpha\tilde{\boldsymbol{\xi}}) - \phi(\boldsymbol{\xi}) \Big). \tag{1}$$

Further we define the partial directional derivative of ϕ at ξ_i in direction of $\tilde{\xi}_i$ as follows:

$$\Phi_{\xi_{i'}, i' \neq i}(\xi_i, \tilde{\xi}_i) = \Phi(\boldsymbol{\xi}, \breve{\boldsymbol{\xi}}), \tag{2}$$

where $\breve{\boldsymbol{\xi}} = (\breve{\xi}_1, \ldots, \breve{\xi}_s)$ with $\breve{\xi}_{i'} = \xi_{i'}$, $i' = 1, \ldots, s$, $i' \neq i$, and $\breve{\xi}_i = \tilde{\xi}_i$.

Theorem 1. *Let $\phi : \Xi \to (-\infty; \infty]$ be convex and differentiable.*

a) The following statements are equivalent:

(i) $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ minimizes $\phi(\boldsymbol{\xi})$

(ii) $\sum_{i=1}^{s} \Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \xi_i) \geq 0, \ \forall\, \xi_i \in \Xi_i, \ i = 1, \ldots, s$

(iii) $\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi_i)\geq 0,\ \forall\,\xi_i\in\Xi_i,\ i=1,\ldots,s$

(iv) $\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\delta_{x_i})\geq 0,\ \forall\,x_i\in\mathcal{X}_i,\ i=1,\ldots,s.$

b) Let $\boldsymbol{\xi}^*=(\xi^*_1,\ldots,\xi^*_s)$ *minimize* $\phi(\boldsymbol{\xi})$*. Let* x_i *be a support point of* ξ^*_i*,* $i\in\{1,\ldots,s\}$*. Then* $\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\delta_{x_i})=0$*.*

c) Let $\boldsymbol{\xi}^*=(\xi^*_1,\ldots,\xi^*_s)$ *minimize* $\phi(\boldsymbol{\xi})$*. Then the point* $(\boldsymbol{\xi}^*,\boldsymbol{\xi}^*)$ *is a saddle point of* Φ *in that*

$$\Phi(\boldsymbol{\xi}^*,\boldsymbol{\xi})\geq 0=\Phi(\boldsymbol{\xi}^*,\boldsymbol{\xi}^*)\geq\Phi(\tilde{\boldsymbol{\xi}},\boldsymbol{\xi}^*),\quad\forall\,\boldsymbol{\xi},\tilde{\boldsymbol{\xi}}\in\Xi \tag{3}$$

and the point (ξ^*_i,ξ^*_i) *is a saddle point of* $\Phi_{\xi^*_{i'},i'\neq i}$ *in that*

$$\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi_i)\geq 0=\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi^*_i)\geq\Phi_{\xi^*_{i'},i'\neq i}(\tilde{\xi}_i,\xi^*_i),\quad\forall\,\xi_i,\tilde{\xi}_i\in\Xi_i, \tag{4}$$

for all $i=1,\ldots,s$*.*

Proof. a) **(i)⇔(ii)**:

For this proof we present designs in form of row vectors $\xi_i=(w_{i1},\ldots,w_{ik_i})$, where $w_{it}\geq 0$ is the weight of observations at x_{it}, the t-th point of the experimental region $\mathcal{X}_i$, $t=1,\ldots,k_i$, $\sum_{t=1}^{k_i}w_{it}=1$ (see e. g. Boyd and Vandenberghe (2004), ch. 7). Then $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_s)$ is the full (row) vector of all weights of observations at all points of all experimental regions.

We use the notations $\nabla_{\xi_i}\phi$ for the gradient of ϕ with respect to ξ_i: $\nabla_{\xi_i}\phi=\left(\frac{\partial\phi}{\partial w_{it}}\right)_{t=1,\ldots,k_i}$, and $\nabla_{\xi}\phi$ for the gradient of ϕ with respect to $\boldsymbol{\xi}$, which means $\nabla_{\xi}\phi=(\nabla_{\xi_1}\phi,\ldots,\nabla_{\xi_s}\phi)$. (Gradients $\nabla_{\xi_i}\phi$ and $\nabla_{\xi}\phi$ are row vectors).

According to convex optimization theory (see e. g. Boyd and Vandenberghe (2004), ch. 4) $\boldsymbol{\xi}^*$ minimizes ϕ iff $\Phi(\boldsymbol{\xi}^*,\boldsymbol{\xi})\geq 0$ for all $\boldsymbol{\xi}\in\Xi$. Then the equivalence of (i) and (ii) follows from

$$\begin{aligned}\Phi(\boldsymbol{\xi}^*,\boldsymbol{\xi}) &= \nabla_{\xi}\phi(\boldsymbol{\xi}^*)(\boldsymbol{\xi}-\boldsymbol{\xi}^*)^\top\\ &= \sum_{i=1}^{s}\nabla_{\xi_i}\phi(\boldsymbol{\xi}^*)(\xi_i-\xi^*_i)^\top\\ &= \sum_{i=1}^{s}\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi_i).\end{aligned}$$

(ii)⇔(iii): (iii)⇒(ii) Straightforward

(ii)⇒(iii): Let $\exists\ \tilde{\xi}_{i_1}\in\mathcal{X}_i$ with $\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_{i_1},\tilde{\xi}_{i_1})<0$. Let $\xi_{i_1}=\tilde{\xi}_{i_1}$ and $\xi_i=\xi^*_i$, $\forall\,i\neq i_1$. Then for all $i\neq i_1$ we have $\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi_i)=\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi^*_i)=0$, which results in

$$\begin{aligned}\sum_{i=1}^{s}\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi_i) &= \Phi_{\xi^*_{i'},i'\neq i}(\xi^*_{i_1},\tilde{\xi}_{i_1})+\sum_{i\in\{1,\ldots,s\}\setminus i_1}\Phi_{\xi^*_{i'},i'\neq i}(\xi^*_i,\xi^*_i)\\ &= \Phi_{\xi^*_{i'},i'\neq i}(\xi^*_{i_1},\tilde{\xi}_{i_1})<0.\end{aligned}$$

(iii)⇔(iv): (iii)⇒(iv) Straightforward

(iv)⇒(iii) Let $x_i=x_{it}$ be the t-th point in $\mathcal{X}_i$, $t=1,\ldots,k_i$. Then the one-point design with all observations at x_i is given by $\delta_{x_{it}}=\mathbf{e}_t$, where $\mathbf{e}_t$ is the t-th unit (row) vector of

length k_i. A design ξ_i can be written as $\xi_i = \sum_{t=1}^{k_i} w_{it}$. Then the directional derivative of ϕ at ξ_i^* in direction of ξ_i can be presented in form

$$\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \xi_i) = \sum_{t=1}^{k_i} w_{it} \nabla_{\xi_i} \phi(\xi^*)(\mathbf{e}_t - \xi_i^*)^\top,$$

which results in

$$\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \xi_i) = \sum_{t=1}^{k_i} w_{it} \Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \delta_{x_{it}}) \geq 0.$$

b) Let the support point $x_i = x_{it'}$ be the t'-th point in $\mathcal{X}_i$, $t' \in 1, \ldots, k_i$. Then for $\xi_i^* = (w_{i1}^*, \ldots, w_{ik_i}^*)$ we have $w_{it'}^* > 0$ and

$$\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \xi_i^*) = \sum_{t=1}^{k_i} w_{it}^* \Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \delta_{x_{it}}).$$

Let $\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \delta_{x_{it'}}) > 0$. Then since $\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \delta_{x_{it}}) \geq 0$, $t = 1, \ldots, k_i$, we obtain $\Phi_{\xi_{i'}^*, i' \neq i}(\xi_i^*, \xi_i^*) > 0$.

c) The left-hand sides of both (3) and (4) are straightforward. From formula (1) and convexity of ϕ we obtain

$$\Phi(\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}^*) \leq \phi(\boldsymbol{\xi}^*) - \phi(\tilde{\boldsymbol{\xi}}),$$

which is non-positive for optimal $\boldsymbol{\xi}^*$ and all $\tilde{\boldsymbol{\xi}} \in \Xi$. Similarly using formula (2) we obtain the right-hand side of (4).

□

2.2 Optimality conditions based on information matrices

We use the notation $\mathbf{M}_i(\xi_i)$ for a matrix which characterizes a design ξ_i. We assume $\mathbf{M}_i(\xi_i)$ to be symmetric and positive semi-definite and to satisfy the condition

$$\mathbf{M}_i(\xi_i) = \int_{\mathcal{X}_i} \mathbf{M}_i(\delta_{x_i}) \xi_i(\mathrm{d}x_i) \tag{5}$$

for all $i = 1, \ldots, s$. We call this matrix information matrix of a design ξ_i. However, we do not require any specific form of $\mathbf{M}_i(\xi_i)$. In particular cases it may be, for example, the Fisher information or the moment matrix in sense of Pukelsheim (1993). $\mathcal{M}_i$ denotes the set of all information matrices $\mathbf{M}_i(\xi_i)$, $\xi_i \in \Xi_i$. For $\mathbb{M}(\boldsymbol{\xi}) = (\mathbf{M}_1(\xi_1), \ldots, \mathbf{M}_s(\xi_s))$, $\mathcal{M}$ denotes the set of all $\mathbb{M}(\boldsymbol{\xi})$, $\boldsymbol{\xi} \in \Xi$. Then $\mathcal{M} = \times_{i=1}^s \mathcal{M}_i$ and $\mathcal{M}$ is convex. $\phi : \mathcal{M} \to (-\infty; \infty]$ is a design criterion. $\Phi(\mathbb{M}, \tilde{\mathbb{M}})$ denotes the directional derivative of ϕ at $\mathbb{M}$ in direction of $\tilde{\mathbb{M}}$:

$$\Phi(\mathbb{M}, \tilde{\mathbb{M}}) = \lim_{\alpha \searrow 0} \frac{1}{\alpha} \left(\phi((1-\alpha)\mathbb{M} + \alpha\tilde{\mathbb{M}}) - \phi(\mathbb{M}) \right). \tag{6}$$

We define the partial directional derivative of ϕ at $\mathbf{M}_i$ in direction of $\tilde{\mathbf{M}}_i$ as follows:

$$\Phi_{\mathbf{M}_{i'}, i' \neq i}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) = \Phi(\mathbb{M}, \breve{\mathbb{M}}), \tag{7}$$

where $\breve{\mathbf{M}}_{i'} = \mathbf{M}_{i'}$, $i' = 1, \ldots, s$, $i' \neq i$, and $\breve{\mathbf{M}}_i = \tilde{\mathbf{M}}_i$.

Theorem 2. *Let $\phi : \mathcal{M} \to (-\infty; \infty]$ be convex and differentiable.*

a) The following statements are equivalent:

(i) $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ *minimizes* $\phi(\mathbb{M}(\boldsymbol{\xi}))$

(ii) $\sum_{i=1}^{s} \Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i)) \geq 0, \ \forall \xi_i \in \Xi_i, \ i = 1, \ldots, s$

(iii) $\Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i)) \geq 0, \ \forall \xi_i \in \Xi_i, \ i = 1, \ldots, s$

(iv) $\Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\delta_{x_i})) \geq 0, \ \forall x_i \in \mathcal{X}_i, \ i = 1, \ldots, s.$

b) Let $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ *minimize* $\phi(\mathbb{M}(\boldsymbol{\xi}))$. *Let* x_i *be a support point of* ξ_i^*, $i \in \{1, \ldots, s\}$. *Then* $\Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\delta_{x_i})) = 0$.

c) Let $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ *minimize* $\phi(\mathbb{M}(\boldsymbol{\xi}))$. *Then the point* $(\mathbb{M}(\boldsymbol{\xi}^*), \mathbb{M}(\boldsymbol{\xi}^*))$ *is a saddle point of* Φ *in that*

$$\Phi(\mathbb{M}(\boldsymbol{\xi}^*), \mathbb{M}(\boldsymbol{\xi})) \geq 0 = \Phi(\mathbb{M}(\boldsymbol{\xi}^*), \mathbb{M}(\boldsymbol{\xi}^*)) \geq \Phi(\mathbb{M}(\tilde{\boldsymbol{\xi}}), \mathbb{M}(\boldsymbol{\xi}^*)), \quad \forall \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} \in \Xi \tag{8}$$

and the point $(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i^*))$ *is a saddle point of* $\Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}$ *in that*

$$\begin{aligned} \Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i)) \geq 0 \quad &= \quad \Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i^*)) \\ &\geq \quad \Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\tilde{\xi}_i), \mathbf{M}_i(\xi_i^*)), \quad \forall \xi_i, \tilde{\xi}_i \in \Xi_i, \end{aligned} \tag{9}$$

for all $i = 1, \ldots, s$.

Proof. a) **(i)⇔(ii)**:

We use for the gradients of ϕ with respect to $\mathbf{M}_i$ and $\mathbb{M}$ the notations

$$\nabla_{M_i} \phi = \left(\frac{\partial \phi}{\partial m_{kl}} \right)_{k,l}, \quad \mathbf{M}_i = (m_{kl})_{k,l}$$

and

$$\nabla_{M} \phi = \left(\frac{\partial \phi}{\partial m_{kl}} \right)_{k,l}, \quad \mathbb{M} = (m_{kl})_{k,l},$$

respectively.

$\boldsymbol{\xi}^*$ minimizes ϕ iff $\Phi(\mathbb{M}(\boldsymbol{\xi}^*), \mathbb{M}(\boldsymbol{\xi})) \geq 0$ for all $\boldsymbol{\xi} \in \Xi$. The directional derivative can be computed by formula

$$\Phi(\mathbb{M}, \tilde{\mathbb{M}}) = \frac{\partial \phi}{\partial \alpha} \left((1 - \alpha)\mathbb{M} + \alpha \tilde{\mathbb{M}} \right) |_{\alpha = 0}. \tag{10}$$

Then using some matrix differentiation rules (see e. g. Seber (2007), ch. 17) we receive

$$\begin{aligned} \Phi(\mathbb{M}(\boldsymbol{\xi}^*), \mathbb{M}(\boldsymbol{\xi})) \quad &= \quad \operatorname{tr} \left(\nabla_M \phi(\mathbb{M}(\boldsymbol{\xi}^*)) (\mathbb{M}(\boldsymbol{\xi}) - \mathbb{M}(\boldsymbol{\xi}^*))^\top \right) \\ &= \quad \sum_{i=1}^{s} \operatorname{tr} \left(\nabla_{M_i} \phi(\mathbb{M}(\boldsymbol{\xi}^*)) (\mathbf{M}_i(\xi_i) - \mathbf{M}_i(\xi_i^*)) \right) \\ &= \quad \sum_{i=1}^{s} \Phi_{\mathbf{M}_{i'}(\xi_{i'}^*), i' \neq i}(\mathbf{M}_i(\xi_i^*), \mathbf{M}_i(\xi_i)), \end{aligned}$$

which implies the equivalence of (i) and (ii).

(ii)⇔(iii): (iii)⇒(ii) Straightforward

(ii)⇒(iii): Let $\exists\ \tilde{\xi}_{i_1}$ with $\Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_{i_1}(\xi^*_{i_1}),\mathbf{M}_{i_1}(\tilde{\xi}_{i_1})) < 0$. Then for $\xi_{i_1} = \tilde{\xi}_{i_1}$ and $\xi_i = \xi^*_i$, $\forall\, i \neq i_1$, we obtain

$$\begin{aligned}\sum_{i=1}^{s} \Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\xi_i)) &= \Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_{i_1}(\xi^*_{i_1}),\mathbf{M}_{i_1}(\tilde{\xi}_{i_1})) \\ &\quad + \sum_{i\in\{1,\ldots,s\}\setminus i_1} \Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\xi_i)) \\ &= \Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_{i_1}(\xi^*_{i_1}),\mathbf{M}_{i_1}(\tilde{\xi}_{i_1})) < 0.\end{aligned}$$

(iii)⇔(iv): (iii)⇒(iv) Straightforward

(iv)⇒(iii) The directional derivative of ϕ at $\mathbf{M}_i$ in direction of $\tilde{\mathbf{M}}_i$ is linear in the second argument:

$$\Phi_{\mathbf{M}_{i'},i'\neq i}(\mathbf{M}_i,\tilde{\mathbf{M}}_i) = \mathrm{tr}\left(\nabla_{M_i}\phi(\mathbb{M})(\tilde{\mathbf{M}}_i - \mathbf{M}_i)\right).$$

Then using formula (5) we obtain

$$\Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\xi_i)) = \int_{\mathcal{X}_i} \Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\delta_{x_i}))\xi_i(\mathrm{d}x_i) \qquad (11)$$

for each $\xi_i \in \Xi_i$.

b) The result follows from formula (11), $\Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\delta_{x_i})) \geq 0$, for all $x_i \in \mathcal{X}_i$, and $\Phi_{\mathbf{M}_{i'}(\xi^*_{i'}),i'\neq i}(\mathbf{M}_i(\xi^*_i),\mathbf{M}_i(\xi^*_i)) = 0$.

c) The left-hand sides of both (8) and (9) are straightforward. From convexity of ϕ and formula (6) we obtain

$$\Phi(\mathbb{M}(\tilde{\boldsymbol{\xi}}),\mathbb{M}(\boldsymbol{\xi}^*)) \leq \phi(\mathbb{M}(\boldsymbol{\xi}^*)) - \phi(\mathbb{M}(\tilde{\boldsymbol{\xi}})), \quad \forall\, \tilde{\boldsymbol{\xi}} \in \Xi,$$

which implies the right-hand side of (8). Similarly using formula (7) we obtain the right-hand side of (9).

□

3 Optimal Designs in Multiple-Group RCR Models

We consider the multiple-group RCR models in which N observational units are assigned to s groups: n_i observational units in the i-th group, $\sum_{i=1}^{s} n_i = N$. The group sizes n_i are fixed. Experimental designs are assumed to be the same for all observational units within one group (group-design): m_i observations per unit in design points x_{ih}, $h = 1,\ldots,m_i$, in group i. However, for units from different groups experimental designs are in general not the same: $m_{i'} \neq m_{i''}$ and (or) $x_{i'h} \neq x_{i''h}$, $i' \neq i''$.

Note that the experimental settings $x_{i1},\ldots,x_{im_i}$ in group i are not necessarily all distinct (repeated measurements are not excluded).

Note also that observational units (often called individuals in the literature) are usually expected to be people, animals or plants. However, they may also be studies, centers, clinics, plots, etc.

3.1 Model specification and estimation of unknown parameters

In multiple-group random coefficient regression models the h-th observation of the j-th observational unit in the i-th group is given by the following l-dimensional random column vector

$$\mathbf{Y}_{ijh} = \mathbf{G}_i(x_{ih})\boldsymbol{\beta}_{ij} + \boldsymbol{\varepsilon}_{ijh}, \quad x_{ih} \in \mathcal{X}_i, \quad h = 1, \ldots, m_i, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, s, \tag{12}$$

where $\mathbf{G}_i$ denotes a group-specific $(l \times p)$ matrix of known regression functions in group i (in particular case $l = 1$: $\mathbf{G}_i = \mathbf{f}_i^\top$, where $\mathbf{f}_i$ is a p-dimensional column vector of regression functions), experimental settings x_{ih} come from some experimental region $\mathcal{X}_i$, $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, \ldots, \beta_{ijp})^\top$ are unit-specific random parameters with unknown mean $\boldsymbol{\beta}_0$ and given $(p \times p)$ covariance matrix $\mathbf{D}_i$, $\boldsymbol{\varepsilon}_{ijh}$ denote column vectors of observational errors with zero mean and non-singular $(l \times l)$ covariance matrix $\boldsymbol{\Sigma}_i$. All observational errors and all random parameters are assumed to be uncorrelated.

For the vector $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij1}^\top, ..., \mathbf{Y}_{ijm_i}^\top)^\top$ of observations at the j-th observational unit in the i-th group we obtain

$$\mathbf{Y}_{ij} = \mathbf{F}_i\boldsymbol{\beta}_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, s, \tag{13}$$

where $\mathbf{F}_i = (\mathbf{G}_i^\top(x_{i1}), ..., \mathbf{G}_i^\top(x_{im_i}))^\top$ is the design matrix in group i and $\boldsymbol{\varepsilon}_{ij} = (\boldsymbol{\varepsilon}_{ij1}^\top, ..., \boldsymbol{\varepsilon}_{ijm_i}^\top)^\top$. Then the vector $\mathbf{Y}_i = (\mathbf{Y}_{i1}^\top, \ldots, \mathbf{Y}_{in_i}^\top)^\top$ of all observations in group i is given by

$$\mathbf{Y}_i = (\mathbb{I}_{n_i} \otimes \mathbf{F}_i)\,\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, s, \tag{14}$$

where $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}^\top, \ldots, \boldsymbol{\beta}_{in_i}^\top)^\top$, $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}^\top, \ldots, \boldsymbol{\varepsilon}_{in_i}^\top)^\top$, $\mathbb{I}_{n_i}$ is the $n_i \times n_i$ identity matrix and "$\otimes$" denotes the Kronecker product, and the total vector $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_s^\top)^\top$ of all observations in all groups results in

$$\mathbf{Y} = \begin{pmatrix} \mathbb{1}_{n_1} \otimes \mathbf{F}_1 \\ \ldots \\ \mathbb{1}_{n_s} \otimes \mathbf{F}_s \end{pmatrix} \boldsymbol{\beta}_0 + \tilde{\boldsymbol{\varepsilon}} \tag{15}$$

with

$$\tilde{\boldsymbol{\varepsilon}} = \text{block-diag}\,(\mathbb{I}_{n_1} \otimes \mathbf{F}_1, \ldots, \mathbb{I}_{n_s} \otimes \mathbf{F}_s)\,\mathbf{b} + \boldsymbol{\varepsilon},$$

where block-diag$(\mathbf{A}_1, \ldots, \mathbf{A}_s)$ is the block-diagonal matrix with blocks $\mathbf{A}_1, \ldots, \mathbf{A}_s$, $\mathbf{b} = \boldsymbol{\beta} - \mathbb{1}_N \otimes \boldsymbol{\beta}_0$ for $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_s^\top)^\top$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_s^\top)^\top$ and $\mathbb{1}_{n_i}$ is the column vector of length n_i with all entries equal to 1.

Using Gauss-Markov theory we obtain the following best linear unbiased estimator for the mean parameters $\boldsymbol{\beta}_0$:

$$\hat{\boldsymbol{\beta}}_0 = \left[\sum_{i=1}^{s} n_i((\tilde{\mathbf{F}}_i^\top\tilde{\mathbf{F}}_i)^{-1} + \mathbf{D}_i)^{-1}\right]^{-1} \sum_{i=1}^{s} n_i((\tilde{\mathbf{F}}_i^\top\tilde{\mathbf{F}}_i)^{-1} + \mathbf{D}_i)^{-1}\hat{\boldsymbol{\beta}}_{0,i}, \tag{16}$$

where $\hat{\boldsymbol{\beta}}_{0,i} = (\tilde{\mathbf{F}}_i^\top\tilde{\mathbf{F}}_i)^{-1}\tilde{\mathbf{F}}_i^\top\tilde{\bar{\mathbf{Y}}}_i$ is the estimator based only on observations in the i-th group, $\tilde{\mathbf{F}}_i = (\mathbb{I}_{n_i} \otimes \boldsymbol{\Sigma}_i^{-1/2})\mathbf{F}_i$ and $\tilde{\bar{\mathbf{Y}}}_i = (\mathbb{I}_{n_i} \otimes \boldsymbol{\Sigma}_i^{-1/2})\bar{\mathbf{Y}}_i$ (for the mean observational vector $\bar{\mathbf{Y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{Y}_{ij}$ in the i-th group and the symmetric positive definite matrix $\boldsymbol{\Sigma}_i^{1/2}$ with the property $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\Sigma}_i^{1/2}$) are the transformed design matrix and the transformed mean observational vector with respect to the covariance structure of observational errors.

Note that BLUE (16) exists only if all matrices $\tilde{\mathbf{F}}_i^\top\tilde{\mathbf{F}}_i$ are non-singular. Therefore, we restrict ourselves on the case where design matrices $\mathbf{F}_i$ are of full column rank for all i.

The covariance matrix of the best linear unbiased estimator $\hat{\boldsymbol{\beta}}_0$ is given by

$$\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}_0\right) = \left[\sum_{i=1}^{s} n_i((\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i)^{-1} + \mathbf{D}_i)^{-1}\right]^{-1}. \tag{17}$$

In Fedorov and Jones (2005) similar results were obtained for the multi-center trials models.

3.2 Design criteria

We define an exact design in group i as

$$\xi_i = \begin{pmatrix} x_{i1}, \ldots, x_{ik_i} \\ m_{i1}, \ldots, m_{ik_i} \end{pmatrix},$$

where $x_{i1}, \ldots, x_{ik_i}$ are the (distinct) experimental settings in $\mathcal{X}_i$ with the related numbers of observations $m_{i1}, \ldots, m_{ik_i}$, $\sum_{h=1}^{k_i} m_{ih} = m_i$. For analytical purposes we also introduce approximate designs:

$$\xi_i = \begin{pmatrix} x_{i1}, ..., x_{ik_i} \\ w_{i1}, ..., w_{ik_i} \end{pmatrix},$$

where $w_{ih} \geq 0$ denotes the weight of observations at x_{ih}, $h = 1, \ldots, k_i$, and $\sum_{h=1}^{k_i} w_{ih} = 1$.

We will use the following notation for the information matrix in group i:

$$\mathbf{M}_i(\xi_i) = \sum_{h=1}^{k_i} w_{ih}\, \tilde{\mathbf{G}}_i(x_{ih})^\top \tilde{\mathbf{G}}_i(x_{ih}), \tag{18}$$

where $\tilde{\mathbf{G}}_i = \boldsymbol{\Sigma}_i^{-1/2}\mathbf{G}_i$. For exact designs we have $w_{ih} = m_{ih}/m_i$ and

$$\mathbf{M}_i(\xi_i) = \frac{1}{m_i}\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i.$$

We will also use the notation $\boldsymbol{\Delta}_i = m_i\mathbf{D}_i$ for the adjusted dispersion matrix of random effects in group i.

Then we extend the definition of the variance-covariance matrix (17) with respect to approximate designs $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_s)$:

$$\mathrm{Cov}_\xi = \left[\sum_{i=1}^{s} n_i m_i \left(\mathbf{M}_i(\xi_i)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right]^{-1}. \tag{19}$$

Further we focus on the linear (L-) and determinant (D-) criteria for estimation of the mean parameters $\boldsymbol{\beta}_0$. The linear criterion is defined as

$$\phi_L = \mathrm{tr}\left[\mathrm{Cov}\left(\mathbf{L}\hat{\boldsymbol{\beta}}_0\right)\right], \tag{20}$$

where $\mathbf{L}$ is some linear transformation of $\boldsymbol{\beta}_0$. For approximate designs we obtain

$$\phi_L(\boldsymbol{\xi}) = \mathrm{tr}\left(\left[\sum_{i=1}^{s} n_i m_i \left(\mathbf{M}_i(\xi_i)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right]^{-1}\mathbf{V}\right), \tag{21}$$

where $\mathbf{V} = \mathbf{L}^\top\mathbf{L}$.

Remark 1. *The A-, and c-criteria for estimation of* $\boldsymbol{\beta}_0$ *are the particular linear criteria with* $\mathbf{V} = \mathbb{I}_p$ *and* $\mathbf{V} = \mathbf{c}\mathbf{c}^\top$*, for a* p*-dimensional real column vector* $\mathbf{c}$*, respectively.*

The D-criterion is defined as the logarithm of the determinant of the covariance matrix of the estimation, which results in

$$\phi_D(\boldsymbol{\xi}) = -\ln\det\left(\sum_{i=1}^{s} n_i m_i \left(\mathbf{M}_i(\xi_i)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right) \tag{22}$$

for approximate designs.

Note that optimal designs depend on the group sizes n_i only via the proportions n_i/N, $i = 1, \ldots, s$, and are, hence, independent of the total number of observational units N itself. This statement is easy to verify by formulas (21) and (22).

3.3 Optimality conditions

To make use of the equivalence theorems proposed in Section 2 we verify the convexity of the design criteria.

Lemma 1. *The L- and D-criteria for estimation of the mean parameters* $\boldsymbol{\beta}_0$ *are convex with respects to* $\mathbb{M}(\boldsymbol{\xi}) = (\mathbf{M}_1(\xi_1), \ldots, \mathbf{M}_s(\xi_s))$.

Proof. The function $\phi(\mathbf{N}) = \mathbf{N}^{-1}$ is matrix-convex for any positive definite matrix $\mathbf{N}$, i. e.

$$(\alpha\mathbf{N}_1 + (1-\alpha)\mathbf{N}_2)^{-1} \leq \alpha\mathbf{N}_1^{-1} + (1-\alpha)\mathbf{N}_2^{-1} \tag{23}$$

in Loewner ordering for any $\alpha \in [0,1]$ and all positive definite $\mathbf{N}_1$ and $\mathbf{N}_2$ (see e. g. Seber (2007), ch. 10). Since ϕ is non-increasing in Loewner ordering, it is easy to verify that $\psi_i(\mathbf{M}_i) = \left(\mathbf{M}_i^{-1} + \boldsymbol{\Delta}_i\right)^{-1}$ is matrix-concave for any positive definite $\mathbf{M}_i$ (see e. g. Bernstein (2018), ch. 10). Then $\psi(\mathbf{M}_1, \ldots, \mathbf{M}_s) = \sum_{i=1}^{s} n_i m_i \psi_i(\mathbf{M}_i)$ is matrix-concave with respects to $\mathbb{M} = (\mathbf{M}_1, \ldots, \mathbf{M}_s)$. The functions $\phi_1(\mathbf{N}) = -\ln\det(\mathbf{N})$ and $\phi_2(\mathbf{N}) = \operatorname{tr}\left(\mathbf{N}^{-1}\mathbf{V}\right)$ are non-increasing in Loewner ordering and convex for any positive definite matrix $\mathbf{N}$ and any positive semi-definite matrix $\mathbf{V}$ as the standard D- and L-criteria (see e. g. Pázman (1986), ch. 4, or Fedorov and Leonov (2013), ch. 2). Then the functions $\phi_1 \circ \psi$ and $\phi_2 \circ \psi$ are convex. □

We formulate optimality conditions for criteria (21) and (22) based on the results of Theorem 2.

Theorem 3. *Approximate designs* $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ *are L-optimal for estimation of the mean parameters* $\boldsymbol{\beta}_0$ *iff*

$$\begin{aligned}
&\operatorname{tr}\left\{\tilde{\mathbf{G}}_i(x_i)\left[\mathbf{M}_i(\xi_i^*)^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\mathbf{V}\right.\right.\\
&\quad\left.\left.\cdot\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\mathbf{M}_i(\xi_i^*)^{-1}\right]\tilde{\mathbf{G}}_i(x_i)^\top\right\}\\
&\leq \operatorname{tr}\left\{\mathbf{M}_i(\xi_i^*)^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\mathbf{V}\right.\\
&\quad\left.\cdot\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right\}
\end{aligned} \tag{24}$$

for $x_i \in \mathcal{X}_i$, $i = 1, \ldots, s$.

For support points of ξ_i^* *equality holds in* (24).

Proof. We obtain the results using Lemma 1 and parts a) (equivalence of (i) and (iv)) and b) of Theorem 2 for the partial directional derivatives

$$\begin{aligned}\boldsymbol{\Phi}_{L,\mathbf{M}_{i'},i'\neq i}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) &= -n_i m_i \operatorname{tr}\left\{\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1} \left(\mathbf{M}_i^{-1} + \boldsymbol{\Delta}_i\right)^{-1} \mathbf{M}_i^{-1}(\tilde{\mathbf{M}}_i - \mathbf{M}_i)\right. \\ &\quad \left. \cdot\, \mathbf{M}_i^{-1}\left(\mathbf{M}_i^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1} \mathbf{V}\right\} \qquad (25)\end{aligned}$$

for $i = 1, \ldots, s$. □

Theorem 4. *Approximate designs* $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ *are D-optimal for estimation of the mean parameters* $\boldsymbol{\beta}_0$ *iff*

$$\begin{aligned}&\operatorname{tr}\left\{\tilde{\mathbf{G}}_i(x_i)\left[\mathbf{M}_i(\xi_i^*)^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\right.\right. \\ &\quad \left.\left. \cdot \left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\mathbf{M}_i(\xi_i^*)^{-1}\right]\tilde{\mathbf{G}}_i(x_i)^\top\right\} \\ &\leq \operatorname{tr}\left\{\mathbf{M}_i(\xi_i^*)^{-1}\left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1}\right. \\ &\quad \left. \cdot \left(\mathbf{M}_i(\xi_i^*)^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right\} \qquad (26)\end{aligned}$$

for $x_i \in \mathcal{X}_i$, $i = 1, \ldots, s$.

For support points of ξ_i^* *equality holds in* (26).

Proof. The optimality condition follows from Lemma 1 and Theorem 2 with the partial directional derivatives

$$\begin{aligned}\boldsymbol{\Phi}_{D,\mathbf{M}_{i'},i'\neq i}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) &= -n_i m_i \operatorname{tr}\left\{\left[\sum_{r=1}^{s} n_r m_r \left(\mathbf{M}_r^{-1} + \boldsymbol{\Delta}_r\right)^{-1}\right]^{-1} \left(\mathbf{M}_i^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right. \\ &\quad \left. \cdot\, \mathbf{M}_i^{-1}(\tilde{\mathbf{M}}_i - \mathbf{M}_i)\mathbf{M}_i^{-1}\left(\mathbf{M}_i^{-1} + \boldsymbol{\Delta}_i\right)^{-1}\right\} \qquad (27)\end{aligned}$$

for $i = 1, \ldots, s$. □

Note that the results of Theorems 3 and 4 coincide for $l = 1$ and $n_1 = n_2 = 1$ with the optimality conditions for group-wise identical designs in Schmelter (2007b), ch. 8.

3.4 Particular case

We consider the particular multiple-group models, where the regression matrices $\mathbf{G}_i$, the numbers m_i of observations per observational unit, the covariance matrices $\mathbf{D}_i$ and $\boldsymbol{\Sigma}_i$ of random effects and observational errors and the experimental regions $\mathcal{X}_i$ are the same among all groups. For these models we have $m_i = m$, $\boldsymbol{\Delta}_i = \boldsymbol{\Delta}$ and $\mathbf{M}_i(\xi_i) = \mathbf{M}(\xi_i)$, $i = 1, \ldots, s$. Then *L*- and *D*-criteria (21) and (22) simplify to

$$\phi_L(\boldsymbol{\xi}) = \frac{1}{m}\operatorname{tr}\left(\left[\sum_{i=1}^{s} n_i \left(\mathbf{M}(\xi_i)^{-1} + \boldsymbol{\Delta}\right)^{-1}\right]^{-1} \mathbf{V}\right) \qquad (28)$$

and

$$\phi_D(\boldsymbol{\xi}) = -\ln\det\left(\sum_{i=1}^{s} n_i \left(\mathbf{M}(\xi_i)^{-1} + \boldsymbol{\Delta}\right)^{-1}\right) \tag{29}$$

(neglecting the constant term $-p\ln m$ in (29)).

We denote by ξ_L^* an optimal design for the classical linear criterion

$$\phi_L(\xi) = \operatorname{tr}\left(\mathbf{M}(\xi)^{-1}\mathbf{V}\right) \tag{30}$$

and ξ_D^* is an optimal design for the D-criterion in the single-group model

$$\phi_D(\xi) = \ln\det\left(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta}\right). \tag{31}$$

Then it can be easily verified that the designs $\xi_i^* = \xi_L^*$ and $\xi_i^* = \xi_D^*$, $i = 1, \ldots, s$, satisfy the optimality conditions in Theorems 3 and 4, respectively (see Fedorov and Hackl (1997), ch. 5, for the optimality condition with respect to D-criterion (31)).

Corollary 1. *L-optimal designs in the fixed effects model are L-optimal as group-designs in the multiple-group RCR model in which the numbers of observations m_i, the regression matrices $\mathbf{G}_i$, the covariance matrices of random effects and observational errors $\mathbf{D}_i$ and $\boldsymbol{\Sigma}_i$ and the experimental regions $\mathcal{X}_i$ are the same for all groups.*

Corollary 2. *D-optimal designs in the single-group RCR model are D-optimal as group-designs in the multiple-group RCR model in which the numbers of observations m_i, the regression matrices $\mathbf{G}_i$, the covariance matrices of random effects and observational errors $\mathbf{D}_i$ and $\boldsymbol{\Sigma}_i$ and the experimental regions $\mathcal{X}_i$ are the same for all groups.*

The latter statements are expected results in that all observational units in all groups have the same statistical properties and there are no group-specific restrictions on designs. Note that the group sizes have no influence on the designs in this case. However, as we will see in Section 4, even for statistically identical observational units, optimal designs may depend on the group sizes n_i if the numbers m_i of observations per unit differ from group to group.

4 Example: Straight Line Regression

We consider the two-groups model of general form (12) with the regression functions $\mathbf{G}_i(x) = (1, x)^\top$ on the design region $\mathcal{X}_i = [0, 1]$, $i = 1, 2$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih} + \boldsymbol{\varepsilon}_{ijh}, \quad h = 1, \ldots, m_i, \quad j = 1, \ldots, n_i. \tag{32}$$

The covariance structures of random effects and observational errors are given by $\mathbf{D}_i = \operatorname{diag}(d_1, d_2)$ and $\Sigma_i = 1$ for both groups. Group sizes n_i and numbers observations per unit m_i are in general not the same for the first and the second group.

For model (32) the left-hand sides of optimality conditions (24) and (26) are parabolas in x_i with positive leading terms. Then L- and D-optimal approximate designs have the form

$$\xi_i = \begin{pmatrix} 0 & 1 \\ 1 - w_i & w_i \end{pmatrix}, \tag{33}$$

where w_i denotes the weight of observations in point 1 for the i-th group, and the information matrices are given by

$$\mathbf{M}(\xi_i) = \begin{pmatrix} 1 & w_i \\ w_i & w_i \end{pmatrix}. \tag{34}$$

Table 1: A-optimal designs in random intercept model in dependence on group sizes n_i and numbers of observations m_i for $d_1 = 1$

Case no.	n_1	n_2	m_1	m_2	m_1/m_2	w_1^*	$1-w_1^*$	w_2^*	$1-w_2^*$
1	1	1	2	8	1/4	0.298	0.702	0.450	0.550
2	1	1	5	5	1	0.414	0.586	0.414	0.586
3	1	1	8	2	4	0.450	0.550	0.298	0.702
4	1	1	4	16	1/4	0.300	0.700	0.450	0.550
5	1	1	10	10	1	0.414	0.586	0.414	0.586
6	1	1	16	4	4	0.450	0.550	0.300	0.700
7	1	2	2	8	1/4	0.256	0.744	0.439	0.561
8	1	2	5	5	1	0.414	0.586	0.414	0.586
9	1	2	8	2	4	0.460	0.540	0.338	0.662
10	1	2	4	16	1/4	0.258	0.742	0.439	0.561
11	1	2	10	10	1	0.414	0.586	0.414	0.586
12	1	2	16	4	4	0.460	0.540	0.339	0.661

4.1 Random intercept

We consider first the particular case of model (32) in which only the intercept $\boldsymbol{\beta}_{ij1}$ is random, i. e. $d_2 = 0$. We focus on the A- and D-criteria, which are given by (21) for $\mathbf{V} = \mathbb{I}_2$ and (22), respectively. The D-criterion for the random intercept model is given by

$$\phi_D(w_1, w_2) = -\ln\left(\sum_{i=1}^{2} \frac{n_i m_i}{d_1 m_i + 1} \sum_{i=1}^{2} \frac{n_i m_i w_i (d_1 m_i (1-w_i) + 1)}{d_1 m_i + 1} - \left(\sum_{i=1}^{2} \frac{n_i m_i w_i}{d_1 m_i + 1}\right)^2\right). \quad (35)$$

For all values of n_i and m_i, this function achieves its minimum at point $w_1^* = w_2^* = 0.5$, which coincides with the optimal design in the fixed effects model and in the single-group random intercept model (see Schwabe and Schmelter (2008). The A-criterion for the random intercept model is given by

$$\phi_A(w_1, w_2) = \frac{\sum_{i=1}^{2} \frac{n_i m_i (d_1 m_i w_i (1-w_i) + 1 + w_i)}{d_1 m_i + 1}}{\sum_{i=1}^{2} \frac{n_i m_i}{d_1 m_i + 1} \sum_{i=1}^{2} \frac{n_i m_i w_i (d_1 m_i (1-w_i) + 1)}{d_1 m_i + 1} - \left(\sum_{i=1}^{2} \frac{n_i m_i w_i}{d_1 m_i + 1}\right)^2}. \quad (36)$$

In contrast to the D-criterion, the A-optimal weights in general depend on the group sizes n_i and the numbers of observations m_i. Some numerical results for $d_1 = 1$, are presented in Table 1. As we can see in the table, if the numbers of observations m_i are the same for both groups: cases 2, 5, 8 and 11, the optimal weight $w_A^* = 0.414$ in the fixed effects model is also optimal for the multiple-group model, which is in accordance with Corollary 1. In these cases optimal designs are independent of the group sizes n_i and the numbers of observations m_i. For all other cases the optimal weight w_i^* is smaller (larger) than w_A^* if the number of observations m_i is smaller (larger) than the mean number of observations $(m_1 + m_2)/2$. For the same group sizes ($n_1 = n_2$) optimal designs "swap places" if the numbers of observations "swap places": the optimal weight w_1^* in case 1 is the same as the optimal weight w_2^* in case 3 (same for cases 4 and 6). This property, however, does not hold for different group sizes (cases 7 and 9 or 10 and 12).

Table 2: D-optimal designs in random slope model in dependence on group sizes n_i and numbers of observations m_i for $d_2 = 1$

Case no.	n_1	n_2	m_1	m_2	m_1/m_2	w_1^*	$1-w_1^*$	w_2^*	$1-w_2^*$
1	1	1	2	8	1/4	0.725	0.275	0.181	0.819
2	1	1	5	5	1	0.290	0.710	0.290	0.710
3	1	1	8	2	4	0.181	0.819	0.725	0.275
4	1	1	4	16	1/4	0.579	0.421	0.145	0.855
5	1	1	10	10	1	0.232	0.768	0.232	0.768
6	1	1	16	4	4	0.145	0.855	0.579	0.421
7	1	2	2	8	1/4	0.823	0.177	0.206	0.794
8	1	2	5	5	1	0.290	0.710	0.290	0.710
9	1	2	8	2	4	0.155	0.845	0.618	0.382
10	1	2	4	16	1/4	0.651	0.349	0.163	0.837
11	1	2	10	10	1	0.232	0.768	0.232	0.768
12	1	2	16	4	4	0.125	0.875	0.500	0.500

4.2 Random slope

Now we consider the particular case of straight line regression model (32) in which only the slope $\boldsymbol{\beta}_{ij2}$ is random: $d_1 = 0$. For this model the D-criterion is given by

$$\phi_D(w_1, w_2) = -\ln\left(\sum_{i=1}^{2} \frac{n_i m_i w_i}{d_2 m_i w_i + 1} \sum_{i=1}^{2} n_i m_i (1 - w_i)\right). \tag{37}$$

In contrast to the random intercept model, D-optimal designs for the random slope depend on the group sizes and the numbers of observations. Numerical results for $d_2 = 1$, are presented in Table 2. As we can see in the table, if $m_1 = m_2$ optimal designs are the same for both groups (cases 2, 5, 8 and 11). However, they depend on the numbers of observations m_i themselves (optimal weights in cases 2 and 8 differ from those in cases 5 and 11). This phenomenon is in accordance with Corollary 2 since the D-optimal weight w_D^* in the single-group model (which minimizes criterion (31)) also depends on the number of observations (via matrix $\boldsymbol{\Delta}$). In contrast to the random intercept model, for the random slope the optimal weight w_i^* is larger (smaller) than w_D^* if m_i is smaller (larger) than $(m_1 + m_2)/2$.

The A-criterion for the random slope model is given by

$$\phi_A(w_1, w_2) = \frac{\sum_{i=1}^{2} \frac{n_i m_i (d_2 m_i w_i (1-w_i) + 1 + w_i)}{d_2 m_i w_i + 1}}{\sum_{i=1}^{2} \frac{n_i m_i w_i}{d_2 m_i w_i + 1} \sum_{i=1}^{2} n_i m_i (1 - w_i)}. \tag{38}$$

If the numbers of observations are the same for both groups: $m_1 = m_2$, the optimal weights are also the same: $w_1^* = w_2^* = 0.414$, which is in accordance with Corollary 1. In case of different numbers of observations the solution depends on m_1 and m_2 and on the group sizes n_1 and n_2. If the group sizes are the same: $n_1 = n_2 = 1$, the optimal values $(w_1^*; w_2^*)$ are given by $(1; 0.263)$ and $(0.263; 1)$ for $m_1 = 2$, $m_2 = 8$ and $m_1 = 8$, $m_2 = 2$, respectively. For $m_1 = 4$, $m_2 = 16$ and $m_1 = 16$, $m_2 = 4$ we obtain $(1; 0.262)$ and $(0.262; 1)$. These solutions lead to singular information matrices. However, the behavior of the A-optimal designs is in general similar to that of the D-optimal designs presented in Table 2: The optimal weight w_i^* in the group with smaller numbers of observations ($m_i < m_{i'}$) is larger than the optimal weight $w_{i'}^*$ in the other one. The difference

between the two weights w_1^* and w_2^* for the A-criterion is significantly larger than for the D-criterion. Also in case $n_1 = 1$, $n_2 = 2$ optimal designs behave similarly. For $m_1 = 2$, $m_2 = 8$ and $m_1 = 8$, $m_2 = 2$ we obtain the values $(1; 0.326)$ and $(0.178, 1)$, respectively. For $m_1 = 4$, $m_2 = 16$ and $m_1 = 16$, $m_2 = 4$ the solutions are given by $(1; 0.324)$ and $(0.184; 1)$. Although these designs are singular (and hence not applicable), they can be used for determining optimal exact designs. For example, it can be easily verified that in case $m_1 = 2$, $m_2 = 8$, $n_1 = n_2 = 1$, the optimal exact design is given by $m_{11} = 1$, $m_{21} = 3$, i.e. 1 observation in both $x = 1$ and $x = 0$ for all units in the first group, 3 observations in $x = 1$ and 5 observations in $x = 0$ for the second group.

5 Discussion

In this work we considered design problems depending on several designs simultaneously. We proposed equivalence theorems based on the assumptions of convexity and differentiability of the optimality criteria. For design problems with finite experimental regions we formulated optimality conditions with respect to the designs themselves (Theorem 1). If the optimality criteria depend on the designs via information matrices only, optimality conditions are formulated with respect to the information matrices (Theorem 2).

We applied the proposed optimality conditions to the multiple-group RCR models. If all observational units have the same statistical properties and there are no group-specific design-restrictions, optimal designs in the single-group models are also optimal as group-designs in the multiple-group models. In this case the group sizes have no influence on the designs. However, if the numbers of observations differ from group to group, optimal group-designs may depend on the numbers of observations and the group sizes. This behavior has been illustrated by the example of straight line regression models. For more complicated models, it may be difficult to fix support points. The design problem can be solved numerically by using general-purpose tools for constrained optimization. However, these tools are imprecise due to the design problem being multi-dimensional. Standard tools for computing optimal designs, for example OptimalDesign package (see Harman and Filová (2019), cannot be used directly because of the complexity of the design criteria. Also the approach proposed in Harman and Prus (2018) for the compound Bayes risk criterion, which covers a lot of linear criteria in mixed models, is useless for multiple-design problems. Therefore, for the computation of optimal approximate and exact designs some new approach has to be developed. This work is planned for a future research.

The proposed results solve the problem of optimal designs for the estimation of fixed effects in the multiple-group RCR models. The problem of prediction of random parameters remains, however, open. Moreover, we assumed fixed numbers of observational units n_i and observations per unit m_i. Design optimization with respect to these numbers may be an interesting direction for a future research.

Acknowledgment

This research was supported by grant SCHW 531/16 of the German Research Foundation (DFG). The author is grateful to Norbert Gaffke for fruitful discussions on optimality conditions in Theorem 2 and convexity in Lemma 1. Also discussions with Rainer Schwabe on application in mixed models have been very helpful. The author thanks three anonymous referees and the associate editor for helpful comments which significantly improved the presentation of the results.

References

Bernstein, D. S. (2018). *Scalar, Vector, and Matrix Mathematics.* Princeton University Press.

Bludowsky, A., Kunert, J., and Stufken, J. (2015). Optimal designs for the carryover model with random interactions between subjects and treatments. *Australian and New Zealand Journal of Statistics*, **57**, 517–533.

Bose, M. and Mukerjee, R. (2015). Optimal design measures under asymmetric errors, with application to binary design points. *Journal of Statistical Planning and Inference*, **159**, 28–36.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press.

Dumont, C., Lestini, G., Nagard, H. L., Mentré, F., Comets, E., and Nguyen, T. T. (2018). PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Computer methods and programs in biomedicine*, **156**, 217–229.

Entholzner, M., Benda, N., Schmelter, T., and Schwabe, R. (2005). A note on designs for estimating population parameters. *Biometrical Letters - Listy Biometryczne*, **42**, 25–41.

Fedorov, V. and Hackl, P. (1997). *Model-Oriented Design of Experiments.* Springer, New York.

Fedorov, V. and Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248.

Fedorov, V. and Leonov, S. (2013). *Optimal Design for Nonlinear Response Models.* CRC Press, Boca Raton.

Graßhoff, U., Doebler, A., Holling, H., and Schwabe, R. (2012). Optimal design for linear regression models in the presence of heteroscedasticity caused by random coefficients. *Journal of Statistical Planning and Inference*, **142**, 1108–1113.

Harman, R. and Filová, L. (2019). Package 'OptimalDesign'. https://cran.r-project.org/web/packages/OptimalDesign/index.html.

Harman, R. and Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. *Statistics and Probability Letters*, **137**, 135–141.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, **2**, 849–879.

Kunert, J., Martin, R. J., and Eccleston, J. (2010). Optimal block designs comparing treatments with a control when the errors are correlated. *Journal of Statistical Planning and Inference*, **140**, 2719–2738.

Lemme, F., van Breukelen, G. J. P., and Berger, M. P. F. (2015). Efficient treatment allocation in two-way nested designs. *Statistical Methods in Medical Research*, **24**, 494–512.

Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, **84**, 429–442.

Pázman, A. (1986). *Foundations of optimum Experimental Design.* D. Reidel Publishing Company, New York.

Prus, M. (2015). *Optimal Designs for the Prediction in Hierarchical Random Coefficient Regression Models*. Ph.D. thesis, Otto-von-Guericke University, Magdeburg.

Prus, M. (2020). Optimal designs in multiple group random coefficient regression models. *TEST*, **29**, 233–254.

Prus, M. and Schwabe, R. (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B*, **78**, 175–191.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Schmelter, T. (2007a). Considerations on group-wise identical designs for linear mixed models. *Journal of Statistical Planning and Inference*, **137**, 4003–4010.

Schmelter, T. (2007b). *Experimental Design For Mixed Models With Application to Population Pharmacokinetic Studies*. Ph.D. thesis, Otto-von-Guericke University Magdeburg.

Schwabe, R. (1996). Optimal design of experiments for multivariate response in two-factor linear models. In A. K. Gupta and V. L. Girko, editors, *Multidimensional Statistical Analysis and Theory of Random Matrices. Proceedings of the Sixth Eugene Lukacs Symposium*, pages 235–242. Bowling Green, OH. VSP, Utrecht.

Schwabe, R. and Schmelter, T. (2008). On optimal designs in random intercept models. *Tatra Mountains Mathematical Publications*, **39**, 145–153.

Seber, G. A. F. (2007). *A Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society. Series B*, **35**, 123–130.

Wierich, W. (1989). An application of the Kiefer-Wolfowitz equivalence theorem to ANOVA models with additive regression. *Journal of Statistical Planning and Inference*, **21**, 277–283.

Statistical Papers (2023) 64:1057–1068
https://doi.org/10.1007/s00362-023-01440-1

REGULAR ARTICLE

Optimal designs for prediction in random coefficient regression with one observation per individual

Maryna Prus[1,2]

Received: 31 December 2022 / Revised: 28 February 2023 / Published online: 23 March 2023

Abstract
The subject of this work is random coefficient regression models with only one observation per observational unit (individual). An analytical solution in form of optimality conditions is proposed for optimal designs for the prediction of individual random effect for a group of selected individuals. The behavior of optimal designs is illustrated by the example of linear regression models.

Keywords Experimental design · Mixed model · Prediction · Random effects

1 Introduction

The subject of this work is random coefficient regression (RCR) models in which only one observation per observational unit (individual) is possible. These models are popular e.g. in psychology (see Freund and Holling (2008)) and pharmacokinetics (see Patan and Bogacka (2007)). The main purpose of the present paper is to obtain an analytical solution for the designs that are optimal for the prediction of random effects. Optimal designs for prediction in RCR models have been discussed e.g. in Gladitz and Pilz (1982), Fedorov and Hackl (1997), Prus and Schwabe (2016) and Prus (2023). In Gladitz and Pilz (1982) and Prus and Schwabe (2016) the number of observations per individual is required to be not smaller than the number of unknown model parameters, which excludes the case of only one observation. Also the same design for all individuals has been assumed in the both papers. Fedorov and Hackl (1997) investigated models with specific regression functions. The models considered in the present work may be seen as a particular case (with one individual per group) of the multiple-group models discussed in Prus (2023). However, the solution developed in that paper is based on the assumption of sufficient number of observations per

✉ Maryna Prus
maryna.prus@uni-hohenheim.de

1 Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

2 Division of Statistics and Machine Learning, Linköping University, Linköping, Sweden

individual, i.e. not smaller than the number of parameters. Models with only one observation per individual were discussed e.g. in Patan and Bogacka (2007) and Graßhoff et al. (2012). Patan and Bogacka (2007) investigated non-linear mixed-effects models. Graßhoff et al. (2012) proposed a solution for optimal designs in form of an optimality condition for RCR models. In these papers optimal design were determined for estimation of *fixed* effects. However, not much has been done for prediction of *random* effects.

In the present work we determine optimal designs for prediction of *random* effects for a group of selected individuals in RCR models with one observation per individual. The obtained design criterion (linear criterion) results in a multiple-design problem, for which the standard approach for design optimization proposed in Kiefer (1974) cannot be applied. We make use of the optimality conditions for multiple-design problems proposed in Prus (2022). The obtained analytical results are illustrated by examples of linear regression models with particular covariance matrices.

The paper has the following structure: In the second section the model is specified and the mean squared error (MSE) matrix for the best linear unbiased prediction (BLUP) of the random effects is determined. In Sect. 3 the linear design criterion for the prediction and the resulting optimality conditions are formulated. The analytical results are illustrated by examples in Sect. 4. The paper is concluded by a short discussion in Sect. 5.

2 Model specification

In this work we the consider RCR models with only one observation per observational unit (individual) of the following form:

$$Y_i = \mathbf{f}(x_i)^\top \boldsymbol{\beta}_i + \varepsilon_i, \quad i = 1, \ldots, n, \quad x_i \in \mathcal{X}, \tag{1}$$

where Y_i is an observation at the i-th individual, n is the number of individuals, $\mathbf{f} = (f_1, \ldots, f_p)^\top$ is a vector of known regression functions, experimental settings x_i come from an experimental region $\mathcal{X}$. The observational errors ε_i are assumed to have zero mean and common variance $\sigma^2 > 0$. The individual parameters $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip})^\top$ have unknown expected value (population mean) $\mathrm{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ and known covariance matrix $\mathrm{Cov}(\boldsymbol{\beta}_i) = \sigma^2 \mathbf{D}$. All individual parameters $\boldsymbol{\beta}_i$ and all observational errors ε_i are assumed to be uncorrelated.

According to Graßhoff et al. (2012) the covariance matrix of the best linear unbiased estimator (BLUE) $\hat{\boldsymbol{\beta}}$ for the population parameter (fixed effects) $\boldsymbol{\beta}$ is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{M}^{-1}, \tag{2}$$

where

$$\mathbf{M} = \sum_{i=1}^{n} \mathbf{M}(x_i), \tag{3}$$

$$\mathbf{M}(x_i) = \mathbf{g}(x_i)\mathbf{g}(x_i)^\top, \tag{4}$$

$$\mathbf{g}(x_i) = \frac{1}{\sqrt{\mathbf{f}(x_i)^\top \mathbf{D}\mathbf{f}(x_i) + 1}}\mathbf{f}(x_i). \tag{5}$$

In this work we focus on random effects, in particular on the individual deviations from the mean: $\mathbf{b}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}$ (see e.g. Prus and Schwabe (2013)). Namely, we consider the situation in which our main interest is in some selected individuals. We determine optimal experimental settings for prediction of the individual effects for the k selected individuals: $\boldsymbol{\Psi} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{b}_i$, for $k \in [p, n-p]$. We assume $n \geq 2p$, otherwise $[p, n-p] = \emptyset$. Note that the order of the individuals does not matter in this case. Therefore, we can consider k first individuals without loss of generality.

Further we search for experimental settings that minimize the MSE matrix of the BLUP $\hat{\boldsymbol{\Psi}}$ for the individual deviations $\boldsymbol{\Psi}$.

Lemma 1 *The MSE matrix of the BLUP $\hat{\boldsymbol{\Psi}}$ is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \sigma^2 \left(\frac{1}{k}\mathbf{D} - \frac{1}{k^2}\mathbf{D}\mathbf{M}_k\mathbf{D} + \frac{1}{k^2}\mathbf{D}\mathbf{M}_k\mathbf{M}^{-1}\mathbf{M}_k\mathbf{D} \right), \tag{6}$$

where $\mathbf{M}_k = \sum_{i=1}^{k} \mathbf{M}(x_i)$.

The proof of Lemma 1 is deferred to Appendix A.

3 Experimental design

We consider the following two groups of individuals: the k selected individuals build the first group (Group 1), and the second group (Group 2) consists of the $n-k$ remaining individuals. We also allow for group-specific restrictions with respect to experimental settings, i.e. the experimental regions for the two groups of individuals may differ from each other. In practice it can be useful in case of some particular restrictions in two centers/clinics/etc. Further $\mathcal{X}_\ell$ denotes the experimental region for group ℓ, $\ell = 1, 2$, and $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$. The particular case $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$ will be later considered more detailed. We add the group index to experimental settings for clear notation and we define an exact design in group ℓ as

$$\xi_{\ell,e} = \begin{pmatrix} x_{\ell 1}, \ldots, x_{\ell N_\ell} \\ m_{\ell 1}, \ldots, m_{\ell N_\ell} \end{pmatrix},$$

where $x_{\ell j} \in \mathcal{X}_\ell$ are the support point of $\xi_{\ell,e}$, $m_{\ell j} > 0$ is the number of observations at $x_{\ell j}$ with $\sum_{j=1}^{N_1} m_{1j} = k$ and $\sum_{j=1}^{N_2} m_{2j} = n-k$ and $N_\ell = |\{x_{\ell 1}, \ldots, x_{\ell N_\ell}\}|$. Note that $m_{\ell j}$ can be larger than one in case if observations for more than one individual are taken at point $x_{\ell j}$. Note also that the set $\{x_{11}, \ldots, x_{1N_1}\}$ coincides with $\{x_1, \ldots x_k\}$ in model (1) (and $\{x_{21}, \ldots, x_{2N_2}\}$ coincides with $\{x_{k+1}, \ldots x_n\}$).

For analytical purposes we also introduce approximate designs:

$$\xi_\ell = \begin{pmatrix} x_{\ell 1}, \ldots, x_{\ell \tilde{N}_\ell} \\ w_{\ell 1}, \ldots, w_{\ell \tilde{N}_\ell} \end{pmatrix},$$

where $w_{\ell j} > 0$ denotes the weight of observations at $x_{\ell j}$, $\sum_{j=1}^{\tilde{N}_\ell} w_{\ell j} = 1$, $\ell = 1, 2$, and $\tilde{N}_\ell = |\{x_{\ell 1}, \ldots, x_{\ell \tilde{N}_\ell}\}|$. Note that $\tilde{N}_\ell = N_\ell$ for exact designs. Further we use the notation $\xi = (\xi_1, \xi_2)$ for the pair of group designs.

We also define the information (or moment) matrices for the first and second design as

$$\mathbf{M}_{1,\xi} = \sum_{j=1}^{\tilde{N}_1} w_{1j} \mathbf{M}(x_{1j}),$$

and

$$\mathbf{M}_{2,\xi} = \sum_{j=1}^{\tilde{N}_2} w_{2j} \mathbf{M}(x_{2j}),$$

respectively.

For exact designs we obtain $w_{1j} = m_{1j}/k$, $w_{2j} = m_{2j}/(n-k)$, $\mathbf{M}_k = k\mathbf{M}_{1,\xi}$ and

$$\mathbf{M} = k\mathbf{M}_{1,\xi} + (n-k)\mathbf{M}_{2,\xi}.$$

In this work we focus on the linear criterion for the prediction of individual deviations for the selected individuals, which is defined for exact designs as

$$\Phi_L = \operatorname{tr}\left(\operatorname{Cov}(\mathcal{L}\hat{\boldsymbol{\Psi}} - \mathcal{L}\boldsymbol{\Psi})\right), \tag{7}$$

where $\mathcal{L}$ denotes the transformation matrix.

Neglecting the constants that have no influence on designs, we obtain for approximate designs the following results.

Theorem 1 *The linear criterion for the prediction of individual effects* $\boldsymbol{\Psi}$ *is given by*

$$\Phi_L(\xi) = -tr\left[\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi}^{-1} + \frac{1}{n-k}\mathbf{M}_{2,\xi}^{-1}\right)^{-1}\right], \tag{8}$$

where $\tilde{\mathbf{L}} = \mathbf{DLD}$ *and* $\mathbf{L} = \mathcal{L}^\top\mathcal{L}$.

The proof of this result is deferred to Appendix B.

Note that we assume matrices $\mathbf{M}_{1,\xi}$ and $\mathbf{M}_{2,\xi}$ and consequently $\mathbf{M}_k$ non-singular, which requires $k \in [p, n-p]$. Otherwise linear criterion (7) cannot be written in form (8).

As we can see by formula (8), the linear criterion depends on two designs simultaneously. In this case the general equivalence theorem (see Kiefer (1974)) cannot be directly applied. Instead we use the extended version for multiple-design problems proposed in Prus (2022). To make use of the equivalence theorems presented in that work, we have to verify convexity of the criterion.

Lemma 2 *The linear criterion for the prediction of individual effects* $\boldsymbol{\Psi}$ *is convex with respect to* $(\mathbf{M}_{1,\xi}, \mathbf{M}_{2,\xi})$.

For the proof of Lemma 2 see Appendix C.

As the linear criterion for the prediction of the individual deviations is differentiable and convex with respect to both moment matrices, optimality conditions can be formulated.

Theorem 2 *Approximate designs* $\xi^* = (\xi_1^*, \xi_2^*)$ *are L-optimal for the prediction of the individual effects* $\boldsymbol{\Psi}$ *iff*

$$\begin{aligned}&\mathbf{g}(x)^\top \mathbf{M}_{1,\xi^*}^{-1}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\mathbf{M}_{1,\xi^*}^{-1}\mathbf{g}(x)\\ &\leq tr\left[\mathbf{M}_{1,\xi^*}^{-1}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\right],\end{aligned} \tag{9}$$

for all $x \in \mathcal{X}_1$, *and*

$$\begin{aligned}&\mathbf{g}(x)^\top \mathbf{M}_{2,\xi^*}^{-1}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\mathbf{M}_{2,\xi^*}^{-1}\mathbf{g}(x)\\ &\leq tr\left[\mathbf{M}_{2,\xi^*}^{-1}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi^*}^{-1}+\frac{1}{n-k}\mathbf{M}_{2,\xi^*}^{-1}\right)^{-1}\right],\end{aligned} \tag{10}$$

for all $x \in \mathcal{X}_2$.

For support points of ξ_1^* *and* ξ_2^* *equalities hold in* (9) *and* (10), *respectively.*

The proof is deferred to Appendix D.

Note that optimal designs depend on the dispersion matrix of random effects $\mathbf{D}$, the total number of individuals n and the number of selected individuals k.

Note also that in case where the design region is the same for all individuals $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$, optimal designs may be also the same for both groups, i.e. $\mathbf{M}_{1,\xi^*} = \mathbf{M}_{2,\xi^*}$. In this situation design criterion (8) simplifies to

$$\Phi_L(\xi) = -\mathrm{tr}\left[\tilde{\mathbf{L}}\mathbf{M}_{1,\xi}\right], \tag{11}$$

which is linear in design, i.e. singular designs (that result in a singular moment matrix) or all permissible designs may be optimal. The optimality conditions simplify to the following inequality:

$$\mathbf{g}(x)^\top \tilde{\mathbf{L}}\mathbf{g}(x) \leq \mathrm{tr}\left(\tilde{\mathbf{L}}\mathbf{M}_{1,\xi^*}\right), \quad \forall x \in \mathcal{X}. \tag{12}$$

Behavior of such designs will be illustrated by Example 4.1.

4 Examples

We consider the linear regression model

$$Y_i = \boldsymbol{\beta}_{i1} + \boldsymbol{\beta}_{i2} x_i + \varepsilon_i, \tag{13}$$

which is the particular case of model (1) with $\mathbf{f}(x) = (1, x)^\top$, and we assume the diagonal structure of the covariance matrix of random effects: $\mathbf{D} = \mathrm{diag}(d_1, d_2)$.

Further we distinguish between the two different cases: the same design region for all individuals (Example 4.1) and different design regions for different groups (Example 4.2). We focus on the A-optimality criterion, i.e. $\mathbf{L} = \mathbb{I}_p$, where $\mathbb{I}_p$ denotes the $p \times p$ identity matrix.

4.1 Example 1

In this example we consider the situation where there are no particular restrictions for the designs for the selected individuals (that are of our main interest) or for all other individuals, i.e. the design region is the same for both groups: $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$. In particular, we consider symmetric design regions: $\mathcal{X} = [-a, a]$, for $a > 0$. Further we focus on the designs that are the same for both groups: $\xi_1 = \xi_2$ with all observations at the endpoints:

$$\xi_1 = \begin{pmatrix} -a & a \\ 1-w & w \end{pmatrix}, \tag{14}$$

where w is the weight of observations at point a. The total number of individuals n and the number of selected individuals k have no influence on designs in the present case and do not need to be specified. If optimal designs of form (14) exist, they assign $w = 0$ or $w = 1$ (and make the moment matrix singular) or all values of w are equally good.

For the present model we obtain

$$\mathbf{g}(x) = \frac{1}{\sqrt{d_2 x^2 + d_1 + 1}} (1, x)^\top,$$

and

$$\mathbf{M}_{1,\xi} = \mathbf{M}_{2,\xi} = \frac{1}{d_2 a^2 + d_1 + 1} \begin{pmatrix} 1 & a(1-2w) \\ a(1-2w) & a^2 \end{pmatrix}.$$

The simplified linear criterion (11) is given by

$$\Phi_L(\xi) = -\frac{d_2a^2 + d_1^2}{d_2a^2 + d_1 + 1},$$

which is independent of the designs, i.e. all designs (all $w \in (0, 1)$) are optimal or there is no optimal designs of form (14) at all. The optimality condition (12) is given by

$$\frac{d_2(d_1^2 - d_2(d_1 + 1))(a^2 - x^2)}{(d_2a^2 + d_1 + 1)(d_2x^2 + d_1 + 1)} \leq 0, \quad \forall x \in [-a, a],$$

which is satisfied only in case

$$\{d_2 \geq \frac{d_1^2}{d_1 + 1}\} \cup \{d_2 = 0\}. \tag{15}$$

Hence, under condition (15) all design are equally good. Otherwise, there is no solution with $\xi_1 = \xi_2$, where ξ_1 is of form (14).

4.2 Example 2

In this example we assume different design regions $\mathcal{X}_1$ and $\mathcal{X}_2$. In particular, we concentrate on symmetric design regions of different lengths: $\mathcal{X} = [-1, 1]$ and $\mathcal{X} = [-a, a]$, for $a > 0$. We consider the endpoints-designs

$$\xi_1 = \begin{pmatrix} -1 & 1 \\ 1 - w_1 & w_1 \end{pmatrix}, \tag{16}$$

and

$$\xi_2 = \begin{pmatrix} -a & a \\ 1 - w_2 & w_2 \end{pmatrix}, \tag{17}$$

where w_1 and w_2 are the weights of observations at points 1 and a for the first and the second group, respectively. We simplified the notations w_{12} and w_{22} to w_1 and w_2, as there are only two support points for each design. Further we consider two different cases for the covariance matrix of random effects: random intercept and random slope.

Case 1 $d_1 = d$ **and** $d_2 = 0$

In the case of random intercept we have

$$\mathbf{g}(x) = \frac{1}{\sqrt{d+1}}(1, x)^\top,$$
$$\mathbf{M}_{1,\xi} = \frac{1}{d+1}\begin{pmatrix} 1 & 1 - 2w \\ 1 - 2w & 1 \end{pmatrix},$$

and

$$\mathbf{M}_{2,\xi} = \frac{1}{d+1}\begin{pmatrix} 1 & a(1-2w) \\ a(1-2w) & a^2 \end{pmatrix}.$$

Linear criterion (8) for this model will not be presented here because of its complexity. We use software Maple2020 for further computations. For the following points the first derivative of the criterion function is zero:

$$w_1 = aw_2 - \frac{a}{2} + \frac{1}{2}, \quad (w_1, w_2) \in [0,1]^2. \tag{18}$$

All designs with property (18) turn out to be optimal, i.e. they satisfy optimality conditions (9) and (10) for all possible values of a. Note that the designs are independent of the variance parameter d.

Case 2 $d_1 = 0$ **and** $d_2 = d$

For the random slope we obtain

$$\mathbf{g}(x) = \frac{1}{\sqrt{da^2+1}}(1,x)^\top,$$
$$\mathbf{M}_{1,\xi} = \frac{1}{da^2+1}\begin{pmatrix} 1 & 1-2w \\ 1-2w & 1 \end{pmatrix},$$

and

$$\mathbf{M}_{2,\xi} = \frac{1}{da^2+1}\begin{pmatrix} 1 & a(1-2w) \\ a(1-2w) & a^2 \end{pmatrix}.$$

Using the same approach as for the random intercept case, it can be verified that all designs with

$$w_1 = \frac{a + 2w_2 - 1}{2a}, \quad (w_1, w_2) \in [0,1]^2, \tag{19}$$

are optimal for all $a > 0$.

Note that in both cases: random intercept and random slope, the obtained designs are optimal for all values of the total number of individuals n and the number of selected individuals $k \in [p, n-p]$. Note also that $a = 1$ would lead to $w_1 = w_2$, which is in accordance with Example 4.1.

5 Discussion

We considered RCR models in which only one observation per individual is possible. We focused on individual random effects for a group of selected individuals, in particular on the mean random effect in the group. The solution for optimal designs is proposed in form of optimality conditions. The number of selected individuals k is

assumed to be between the number of parameters p and $n - p$, where n is the total number of individuals, which is possible only for models with $n \geq 2p$. If the main interest is in a very small, i.e. $k < p$, or a very large number of selected individuals, another approach is, however, needed. Optimal designs are determined for the BLUP of the random effects, which depends on the variance parameters (covariance matrices). The variance parameters are assumed to be known, which is in general not the case in practice. In a practical situation, where the covariance matrices have to be estimated, we deal with estimated BLUP (EBLUP). The obtained optimal designs in general depend on the variance parameters and are, therefore, locally optimal. The problem of local optimality may be solved for particular models by considering robust design criteria, for example minimax-criterion. For specific covariance structure, optimal designs may be independent of variance parameters, which has been illustrated by the example of the linear regression models with random intercept and random slope. All observational errors were assumed to have the same variance. Heteroscedastic errors, especially depending on the experimental settings (as considered by Graßhoff et al. (2012)), are analytically more challenging. However, an extension may be considered in a future research. Moreover optimal designs have been obtained for individual effects (individual deviations from the population parameter) via their arithmetic mean. Design optimization for the prediction of the individual deviations or individual parameters themselves turned out to be more challenging. It may also be a subject of future investigations, especially for particular models.

Acknowledgements This research was partially supported by Grant SCHW 531/16 of the German Research Foundation (DFG). A part of the research has been performed as the author was visiting Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis in Nice in framework of the postdoc program (PKZ 91672803) of German Academic Exchange Service (DAAD). The author thanks two anonymous referees for helpful comments which improved the presentation of the results.

Funding Open Access funding enabled and organized by Projekt DEAL.

A Proof of Lemma 1

Model (1) can be rewritten in vector form as follows

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \varepsilon_i, \quad i = 1, \ldots, n, \quad x_i \in \mathcal{X}, \tag{20}$$

with $\mathbf{X} = (\mathbf{f}(x_1), \ldots \mathbf{f}(x_n))^\top$, $\mathbf{Z} =$ block-diag $(\mathbf{f}(x_1)^\top, \ldots \mathbf{f}(x_n)^\top)$ and block-diag $(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is the block-diagonal matrix with

blocks $\mathbf{A}_1, \ldots, \mathbf{A}_n$. This model is the classical linear mixed model considered e.g. in Henderson (1975).

Then the MSE matrix for the BLUP $\hat{\mathbf{b}}$ for random effects $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ can be computed as follows:

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{b}} - \mathbf{b}) &= \left(\mathbf{W} - \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\right)^{-1} \\ &= \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}^{-1}\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}^{-1},\end{aligned}$$

where $\mathbf{G} = \mathrm{Cov}\,(\mathbf{b})$, $\mathbf{R} = \mathrm{Cov}\,(\boldsymbol{\varepsilon})$ and $\mathbf{W} = \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$. For $\mathbf{R} = \sigma^2\mathbb{I}_n$ and $\mathbf{G} = \sigma^2\mathbb{I}_n \otimes \mathbf{D}$ we obtain

$$\begin{aligned}\mathbf{W}^{-1} &= \sigma^2\left(\mathbb{I}_n \otimes \mathbf{D} - \text{block-diag}\left(\mathbf{D}\mathbf{M}_{x_1}\mathbf{D}, \ldots, \mathbf{D}\mathbf{M}_{x_n}\mathbf{D}\right)\right), \\ \left(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}^{-1}\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} - \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X}\right)^{-1} &= -\mathbf{M}^{-1},\end{aligned}$$

and

$$\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}^{-1} = \left(\mathbf{M}_{x_1}\mathbf{D}, \ldots, \mathbf{M}_{x_n}\mathbf{D}\right),$$

which result in

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{b}} - \mathbf{b}) = \sigma^2&\left(\mathbb{I}_n \otimes \mathbf{D} - \text{block-diag}\left(\mathbf{D}\mathbf{M}_{x_1}\mathbf{D}, \ldots, \mathbf{D}\mathbf{M}_{x_n}\mathbf{D}\right)\right) \\ &+\sigma^2\begin{pmatrix}\mathbf{D}\mathbf{M}_{x_1}\mathbf{M}^{-1}\mathbf{M}_{x_1}\mathbf{D} & \ldots & \mathbf{D}\mathbf{M}_{x_1}\mathbf{M}^{-1}\mathbf{M}_{x_n}\mathbf{D} \\ \ldots & \ldots & \ldots \\ \mathbf{D}\mathbf{M}_{x_n}\mathbf{M}^{-1}\mathbf{M}_{x_1}\mathbf{D} & \ldots & \mathbf{D}\mathbf{M}_{x_n}\mathbf{M}^{-1}\mathbf{M}_{x_n}\mathbf{D}\end{pmatrix}.\end{aligned}$$

Then from $\boldsymbol{\Psi} = \frac{1}{k}\left(\left(\mathbf{1}_k^\top, \mathbf{0}_{n-k}^\top\right) \otimes \mathbb{I}_p\right) b$ and consequently

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \frac{1}{k^2}\left(\mathbf{1}_k^\top, \mathbf{0}_{n-k}^\top\right) \otimes \mathbb{I}_p\mathrm{Cov}(\hat{\mathbf{b}} - \mathbf{b})\left(\mathbf{1}_k^\top, \mathbf{0}_{n-k}^\top\right)^\top \otimes \mathbb{I}_p,$$

follows Eq. (6).

B Proof of Theorem 1

To determine the linear criterion for approximate designs, we use the basic formula (7) and we replace matrices $\mathbf{M}_k$ and $\mathbf{M}$ by $k\mathbf{M}_{1,\xi}$ and $k\mathbf{M}_{1,\xi} + (n-k)\mathbf{M}_{2,\xi}$, respectively. Then we suppress the first term ($\frac{1}{k}\mathrm{tr}(\mathbf{D})$), which is independent of designs, and the multiplicator σ^2/k^2, and we obtain

$$\Phi_L(\xi) = \mathrm{tr}\left\{\mathbf{L}\mathbf{D}\left[k\mathbf{M}_{1,\xi}\left(k\mathbf{M}_{1,\xi} + (n-k)\mathbf{M}_{2,\xi}\right)^{-1}k\mathbf{M}_{1,\xi} - k\mathbf{M}_{1,\xi}\right]\mathbf{D}\right\},$$

which can be easily simplified to (8) using the properties of the trace and the standard formula for the inverse of sum of two non-singular matrices.

C Proof of Lemma 2

The function $h(\mathbf{N}) = \mathbf{N}^{-1}$ is non-increasing in Loewner ordering and matrix-convex for any positive definite matrix $\mathbf{N}$. Then

$$\frac{1}{k}\mathbf{M}_{1,\xi}^{-1} + \frac{1}{n-k}\mathbf{M}_{2,\xi}^{-1},$$

is matrix-convex and

$$\left(\frac{1}{k}\mathbf{M}_{1,\xi}^{-1} + \frac{1}{n-k}\mathbf{M}_{2,\xi}^{-1}\right)^{-1},$$

is matrix-concave in $(\mathbf{M}_{1,\xi}, \mathbf{M}_{2,\xi})$, respectively (see e. g. Bernstein (2018), ch. 10). Consequently,

$$-\mathrm{tr}\left[\tilde{\mathbf{L}}\left(\frac{1}{k}\mathbf{M}_{1,\xi}^{-1} + \frac{1}{n-k}\mathbf{M}_{2,\xi}^{-1}\right)^{-1}\right],$$

is convex with respect to $(\mathbf{M}_{1,\xi}, \mathbf{M}_{2,\xi})$ for any positive semi-definite matrix $\tilde{\boldsymbol{L}}$.

D Proof of Theorem 2

According to Theorem 2 in Prus (2022), designs $\xi^* = (\xi_1^*, \xi_2^*)$ minimize a convex criterion Φ if and only if the directional derivative of Φ at $(\mathbf{M}_{1,\xi^*}, \mathbf{M}_{2,\xi^*})$ in the direction of $(\mathbf{g}(x)\mathbf{g}(x)^\top, \mathbf{M}_{2,\xi^*})$ is non-negative for all $x \in \mathcal{X}_1$, and the directional derivative of Φ at $(\mathbf{M}_{1,\xi^*}, \mathbf{M}_{2,\xi^*})$ in the direction of $(\mathbf{M}_{2,\xi^*}, \mathbf{g}(x)\mathbf{g}(x)^\top)$ is non-negative for all $x \in \mathcal{X}_2$. For support points of ξ_1^* and ξ_2^* the related directional derivatives are equal to zero. By computing the directional derivatives for linear criterion (8), i. e. $\Phi = \Phi_L(\xi)$, and setting them non-negative, we obtain inequalities (9) and (10).

References

Bernstein DS (2018) Scalar, vector, and matrix mathematics. Princeton University Press, Princeton

Fedorov V, Hackl P (1997) Model-oriented design of experiments. Springer, New York

Freund PA, Holling H (2008) Creativity in the classroom: a multilevel analysis investigating the impact of creativity and reasoning ability on GPA. Creat Res J 20:309–318

Gladitz J, Pilz J (1982) Construction of optimal designs in random coefficient regression models. Math Oper Stat Ser Stat 13:371–385

Graßhoff U, Doebler A, Holling H, Schwabe R (2012) Optimal design for linear regression models in the presence of heteroscedasticity caused by random coefficients. J Stat Plann Infer 142:1108–1113

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–477

Kiefer J (1974) General equivalence theory for optimum designs (approximate theory). Ann Stat 2:849–879

Patan M, Bogacka B (2007) Efficient sampling windows for parameter estimation in mixed effects models. In: López-Fidalgo J, Rodríguez-Díaz JM, Torsney B (eds) mODa 8—advances in model-oriented design and analysis. Physica, Heidelberg, pp 147–155

Prus M (2022) Equivalence theorems for multiple-design problems with application in mixed models. J Stat Plann Infer 217:153–164

Prus M (2023) Optimal designs for prediction of random effects in two groups linear mixed models. https://www.researchgate.net/publication/353660007_Optimal_designs_for_prediction_of_random_effects_in_multiple-group_linear_mixed_models

Prus M, Schwabe R (2013) Optimal designs for the prediction of individual effects in random coefficient regression. In: Uciński D, Atkinson AC, Patan M (eds) mODa 10—advances in model-oriented design and analysis. Springer, Cham, pp 211–218

Prus M, Schwabe R (2016) Optimal designs for the prediction of individual parameters in hierarchical models. J R Stat Soc B 78:175–191

Optimal designs for prediction of random effects in two-groups models with multivariate response

Maryna Prus

Abstract

We propose an analytical solution for optimal designs for the prediction of individual random effects and the group difference in two-groups models with multivariate response. The solution is given by optimality conditions for approximate designs. In particular two-groups models with the same regression function for both groups, Bayesian optimal designs are optimal for the prediction of the group difference. The results are illustrated by examples of linear and bi-linear regression.

Keywords: Mixed model, multiple-group model, multivariate response, best linear unbiased prediction (BLUP), optimality condition, experimental design

1 Introduction

The subject of this paper is two-groups linear mixed models, where observational units (further called "individuals") are assigned to groups. Individuals within one group are assumed to have the same statistical properties. Group sizes, numbers of observations per individual, experimental regions, regression functions and variance parameters may, however, differ from group to group. Optimal designs for the estimation of fixed (population) parameters in these models have been considered in Schmelter (2007) and Prus (2021). Computational methods for determining efficient exact designs are presented in Dumont *et al.* (2018) and Filová and Prus (2021). However, besides population parameters the main focus may be on random effects as, for example, genotype effects in multi-environmental trials (see e. g. Prus and Piepho (2021)), treatment and control parameters (see e. g. Prus *et al.* (2020)) or center-specific effects in multi-center trials (see e. g. Fedorov and Jones (2005) or Lemme *et al.* (2015)). Prus and Schwabe (2016) considered optimal designs for the prediction of random effects in models with the same design for all individuals (one-group models). Optimal designs in fixed-effects models with multivariate response have been discussed e. g. in Schwabe (1996). Liu *et al.* (2019) has proposed an extension of the results presented in Prus and Schwabe (2016) for multivariate response models.

In this work we consider multiple-group models, where individual designs may differ from group to group. For simplicity, we restrict ourselves on the experiments with two groups of individuals. Such experiments are popular in many statistical applications, for example, in medical research: comparing of two treatments or treatment and placebo. We consider optimal designs for the predictions of the following random effects: the individual random parameters and the group difference, which is given by the difference between the two group means. The analytical solution for the designs is given by optimality conditions for the linear and D-criteria. For the prediction of the group difference, in the particular case of the same regression function and the same experimental region for both groups, Bayesian designs (that minimize Bayesian information matrix) turn out to be optimal.

The paper has the following structure: In Section 2 the two-groups mixed models are specified and the best linear unbiased predictions for the individual random parameters and the group difference are introduced. Section 3 provides an analytical solution for the designs, which are optimal for the prediction of the individual parameters or the group difference. In Section 4 analytical results are illustrated by examples of linear and bi-linear regression models. The paper is concluded by a short discussion.

2 Model Specification and Predictions

In this work we consider mixed effects models with two groups of observational units. The observational units may be people, clinics, studies, centers, plots, etc. Further we call them "individuals" for simplicity. The individuals are randomly allocated to the first or the second group with the group sizes n_1 or n_2, respectively. The total number of individuals is $n = n_1 + n_2$. The vector of observations at individual j in group i is given by

$$\mathbf{Y}_{ij} = \mathbf{F}_i \boldsymbol{\beta}_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, 2, \quad j = 1, \ldots, n_i, \tag{1}$$

where $\mathbf{F}_i$ is the design matrix in group i, ε_{ij} denote observational errors, which are assumed to have zero mean and a given non-singular covariance matrix $\mathrm{Cov}(\varepsilon_{ij}) = \mathbf{V}_i$. (Structure and dimension for design matrices $\mathbf{F}_i$ and covariance matrices $\mathbf{V}_i$ will be specified in Section 3.) $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, \ldots, \beta_{ijp})^\top$ denote individual random parameters with an unknown expected value (population mean parameter) $\mathrm{E}(\boldsymbol{\beta}_{ij}) = \boldsymbol{\beta}_0$ and a given non-singular $(p \times p)$ covariance matrix $\mathrm{Cov}(\boldsymbol{\beta}_{ij}) = \mathbf{D}_i$. All individual parameters $\boldsymbol{\beta}_{ij}$ and all observational errors ε_{ij} are assumed to be uncorrelated.

Note that the results presented in this section are valid for general design matrices $\mathbf{F}_i$ and covariance matrices $\mathbf{V}_i$. Further in Section 3 these two matrices will be specified. However, it is worthwhile mentioning already now that multiple observations (repeated measurements) are possible for each individual. Note also that the number of observations per individual may differ from group to group. (See Section 3 for more details.)

The best linear unbiased estimator (BLUE) for the population mean parameters $\boldsymbol{\beta}_0$ is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_0 \;=\; & \left[n_1((\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1} + \mathbf{D}_1)^{-1} + n_2((\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1} + \mathbf{D}_2)^{-1} \right]^{-1} \\ & \cdot \left[n_1((\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1} + \mathbf{D}_1)^{-1} \hat{\boldsymbol{\beta}}_{0,1} + n_2((\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1} + \mathbf{D}_2)^{-1} \hat{\boldsymbol{\beta}}_{0,2} \right] \end{aligned} \tag{2}$$

where $\hat{\boldsymbol{\beta}}_{0,i} = (\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i)^{-1} \tilde{\mathbf{F}}_i^\top \tilde{\mathbf{Y}}_i$, $\tilde{\mathbf{F}}_i = \mathbf{V}_i^{-1/2} \mathbf{F}_i$, $\tilde{\mathbf{Y}}_i = \mathbf{V}_i^{-1/2} \bar{\mathbf{Y}}_i$, $\bar{\mathbf{Y}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{Y}_{ij}$ and $\mathbf{V}_i^{-1/2} = (\mathbf{V}_i^{1/2})^{-1}$ for the symmetric positive definite matrix $\mathbf{V}_i^{1/2}$ with $\mathbf{V}_i = \mathbf{V}_i^{1/2} \mathbf{V}_i^{1/2}$ (see Prus (2021)). The covariance matrix of the BLUE $\hat{\boldsymbol{\beta}}_0$ is given by

$$\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}_0\right) = \left[n_1((\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1} + \mathbf{D}_1)^{-1} + n_2((\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1} + \mathbf{D}_2)^{-1} \right]^{-1}. \tag{3}$$

The next theorem provides the best linear unbiased predictor (BLUP) for the individual random effects.

Theorem 1. *The BLUP for the individual parameters $\boldsymbol{\beta}_{ij}$ for individual j in group i is given by*

$$\hat{\boldsymbol{\beta}}_{ij} = \left(\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i + \mathbf{D}_i^{-1} \right)^{-1} \left(\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{Y}}_{ij} + \mathbf{D}_i^{-1} \hat{\boldsymbol{\beta}}_0 \right),$$

where $\tilde{\mathbf{Y}}_{ij} = \mathbf{V}_i^{-1/2} \mathbf{Y}_{ij}$.

A similar result has been obtained in Fedorov and Jones (2005) for the multi-center trials models.

Besides the individual parameters themselves, the group difference $\boldsymbol{\mu} = \bar{\boldsymbol{\beta}}_1 - \bar{\boldsymbol{\beta}}_2$ for $\bar{\boldsymbol{\beta}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{\beta}_{ij}$, may be of prior interest. The BLUP for the group difference $\boldsymbol{\mu}$ can be easily determined using Theorem 1 and the relation

$$\boldsymbol{\mu} = \left(\left(\frac{1}{n_1} \mathbb{1}_{n_1}^\top, -\frac{1}{n_2} \mathbb{1}_{n_2}^\top \right) \otimes \mathbb{I}_p \right) \boldsymbol{\beta},$$

where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top$ for $\boldsymbol{\beta}_i = \left(\boldsymbol{\beta}_{i1}^\top, \ldots \boldsymbol{\beta}_{in_i}^\top\right)^\top$, $\mathbb{1}_s$ is the vector of length s with all entries equal to 1, $\mathbb{I}_s$ is the $(s \times s)$ identity matrix and $\otimes$ denotes the Kronecker product.

Corollary 1. *The BLUP for the group difference $\boldsymbol{\mu}$ is given by*

$$\hat{\boldsymbol{\mu}} = \left(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1 + \mathbf{D}_1^{-1}\right)^{-1} \left(\tilde{\mathbf{F}}_1^\top \tilde{\tilde{\mathbf{Y}}}_1 + \mathbf{D}_1^{-1} \hat{\boldsymbol{\beta}}_0\right) - \left(\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2 + \mathbf{D}_2^{-1}\right)^{-1} \left(\tilde{\mathbf{F}}_2^\top \tilde{\tilde{\mathbf{Y}}}_2 + \mathbf{D}_2^{-1} \hat{\boldsymbol{\beta}}_0\right).$$

We measure the performance of a predictor in terms of its mean squared error (MSE) matrix, which is presented for $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top)^\top$, $\hat{\boldsymbol{\beta}}_i = (\hat{\boldsymbol{\beta}}_{i1}^\top, \ldots \hat{\boldsymbol{\beta}}_{in_i}^\top)^\top$, by the next theorem.

Lemma 1. *The MSE matrix of the BLUP $\hat{\boldsymbol{\beta}}$ is given by*

$$\operatorname{Cov}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^\top & \mathbf{B}_{22} \end{pmatrix}, \tag{4}$$

where

$$\mathbf{B}_{11} = (\mathbb{I}_{n_1} - \mathbb{J}_{n_1}) \otimes \left[(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1 + \mathbf{D}_1^{-1})^{-1}\right] + \mathbb{J}_{n_1} \otimes \left[(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1}(\mathbb{I}_p - n_2 \mathbf{U}(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1})\right],$$

$$\mathbf{B}_{12} = (\mathbb{1}_{n_1} \mathbb{1}_{n_2}^\top) \otimes \left[(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1} \mathbf{U} (\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1}\right]$$

and

$$\mathbf{B}_{22} = (\mathbb{I}_{n_2} - \mathbb{J}_{n_2}) \otimes \left[(\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2 + \mathbf{D}_2^{-1})^{-1}\right] + \mathbb{J}_{n_2} \otimes \left[(\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1}(\mathbb{I}_p - n_1 \mathbf{U}(\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1})\right]$$

for

$$\mathbf{U} = \left[n_2((\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1)^{-1} + \mathbf{D}_1) + n_1((\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2)^{-1} + \mathbf{D}_2)\right]^{-1}$$

and $\mathbb{J}_{n_i} = \frac{1}{n_i} \mathbb{1}_{n_i} \mathbb{1}_{n_i}^\top$.

The MSE matrix of the BLUP $\hat{\boldsymbol{\mu}}$ can be established by Lemma 1 and the relation

$$\operatorname{Cov}\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right) = \left(\left(\frac{1}{n_1}\mathbb{1}_{n_1}^\top, -\frac{1}{n_2}\mathbb{1}_{n_2}^\top\right) \otimes \mathbb{I}_p\right) \operatorname{Cov}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \left(\left(\frac{1}{n_1}\mathbb{1}_{n_1}, -\frac{1}{n_2}\mathbb{1}_{n_2}\right) \otimes \mathbb{I}_p\right).$$

Corollary 2. *The MSE matrix of the BLUP $\hat{\boldsymbol{\mu}}$ is given by*

$$\operatorname{Cov}\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right) = \frac{1}{n_1 n_2}\left[\left(n_2\left(\tilde{\mathbf{F}}_1^\top \tilde{\mathbf{F}}_1\right)^{-1} + n_1\left(\tilde{\mathbf{F}}_2^\top \tilde{\mathbf{F}}_2\right)^{-1}\right)^{-1} + \mathbf{C}\right]^{-1}, \tag{5}$$

where $\mathbf{C} = (n_2 \mathbf{D}_1 + n_1 \mathbf{D}_2)^{-1}$.

The proofs of Theorem 1 and Lemma 1 are deferred to Appendix A.

3 Optimal Design

In this section we specify our model (1) in a more detailed way. In the two-groups model with t response variables, the t-dimensional observational vector for the ℓ-th observation of the j-th individual in the i-th group is given by

$$\mathbf{Y}_{ij\ell} = \mathbf{F}_{(i)}(x_{ij\ell})\boldsymbol{\beta}_{ij} + \boldsymbol{\varepsilon}_{ij\ell}, \quad x_{ij\ell} \in \mathcal{X}_i, \quad \ell = 1, \ldots, m_i, \tag{6}$$

where m_i is the number of observations per individual in group i, observational settings $x_{ij\ell}$ come from some experimental region $\mathcal{X}_i$. For analytical purposes, we assume the experimental settings (designs) to be the same for all individuals within one group: $x_{ij\ell} = x_{i\ell}$. However, the experimental regions - and consequently the designs - may differ between the two groups. Observational errors $\boldsymbol{\varepsilon}_{ij\ell}$ are assumed to be uncorrelated with each other and with all individual parameters and to have zero mean and a non-singular $(t \times t)$ covariance matrix Σ_i. In this work we allow for multivariate (t-variate) response and $\mathbf{F}_{(i)}$ denotes a $t \times p$ matrix of regression functions (regression matrix) in group i. In the particular case of univariate response: $t = 1$, the regression matrices $\mathbf{F}_{(i)}$ simplify to usual regression functions: $\mathbf{F}_{(i)} = \mathbf{f}_{(i)}^\top$. Note that the regression matrices may differ between the two groups. For $\mathbf{F}_i = (\mathbf{F}_{(i)}^\top(x_{i1}), ..., \mathbf{F}_{(i)}^\top(x_{im_i}))^\top$, $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij1}^\top, ..., \mathbf{Y}_{ijm_i}^\top)^\top$, $\boldsymbol{\varepsilon}_{ij} = (\boldsymbol{\varepsilon}_{ij1}^\top, ..., \boldsymbol{\varepsilon}_{ijm_i}^\top)^\top$ and $\mathbf{V}_i = \mathbb{I}_{m_i} \otimes \Sigma_i$, our model (6) is of form (1).

We define an individual exact design for individuals in group i as

$$\xi_i = \begin{pmatrix} x_{i1}, \ldots, x_{ik_i} \\ m_{i1}, \ldots, m_{ik_i} \end{pmatrix},$$

where $x_{i1}, \ldots, x_{ik_i}$ are the experimental settings in $\mathcal{X}_i$, $|\mathcal{X}_i| = k_i$ and $\sum_{r=1}^{k_i} m_{ir} = m_i$. For analytical purposes we also introduce approximate designs:

$$\xi_i = \begin{pmatrix} x_{i1}, ..., x_{ik_i} \\ w_{i1}, ..., w_{ik_i} \end{pmatrix},$$

where $w_{ir} \geq 0$, $r = 1, \ldots k_i$, and $\sum_{r=1}^{k_i} w_{ir} = 1$.

We will use the notation $\mathbf{M}_i$ for the moment matrix (in sense of Pukelsheim (1993)) in group i:

$$\mathbf{M}_i = \mathbf{M}_i(\xi_i) = \sum_{r=1}^{k_i} w_{ir}\, \tilde{\mathbf{F}}_{(i)}(x_{ir})^\top \tilde{\mathbf{F}}_{(i)}(x_{ir}), \tag{7}$$

where $\tilde{\mathbf{F}}_{(i)} = \Sigma_i^{-1/2} \mathbf{F}_{(i)}$. For exact designs we have $w_{ir} = m_{ir}/m_i$ and

$$\mathbf{M}_i = \frac{1}{m_i} \tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i,$$

where $\tilde{\mathbf{F}}_i$ is as defined after (2) for $\mathbf{F}_i$ as specified after (6).

We will also use the notation $\boldsymbol{\Delta}_i = m_i \mathbf{D}_i$ for the adjusted dispersion matrix of random effects in group i.

Further we extend the definition of the MSE matrix (4) with respect to approximate designs:

$$\mathrm{MSE}_\beta(\mathbf{M}_1, \mathbf{M}_2) = \begin{pmatrix} \tilde{\mathbf{B}}_{11} & \tilde{\mathbf{B}}_{12} \\ \tilde{\mathbf{B}}_{12}^\top & \tilde{\mathbf{B}}_{22} \end{pmatrix}, \tag{8}$$

where

$$\tilde{\mathbf{B}}_{11} = \frac{1}{m_1} \left\{ (\mathbb{I}_{n_1} - \mathbb{J}_{n_1}) \otimes \left[(\mathbf{M}_1 + \boldsymbol{\Delta}_1^{-1})^{-1} \right] + \mathbb{J}_{n_1} \otimes \left[\mathbf{M}_1^{-1} \left(\mathbb{I}_p - \frac{n_2}{m_1} \tilde{\mathbf{U}} \mathbf{M}_1^{-1} \right) \right] \right\}$$

$$\tilde{\mathbf{B}}_{12} = \frac{1}{m_1\, m_2} (\mathbb{1}_{n_1} \mathbb{1}_{n_2}^\top) \otimes \left[\mathbf{M}_1^{-1} \tilde{\mathbf{U}} \mathbf{M}_2^{-1} \right]$$

and

$$\tilde{\mathbf{B}}_{22} = \frac{1}{m_2} \left\{ (\mathbb{I}_{n_2} - \mathbb{J}_{n_2}) \otimes \left[(\mathbf{M}_2 + \boldsymbol{\Delta}_2^{-1})^{-1} \right] + \mathbb{J}_{n_2} \otimes \left[\mathbf{M}_2^{-1} \left(\mathbb{I}_p - \frac{n_1}{m_2} \tilde{\mathbf{U}} \mathbf{M}_2^{-1} \right) \right] \right\}.$$

for

$$\tilde{\mathbf{U}} = \left[\frac{n_2}{m_1}(\mathbf{M}_1^{-1} + \boldsymbol{\Delta}_1) + \frac{n_1}{m_2}(\mathbf{M}_2^{-1} + \boldsymbol{\Delta}_2)\right]^{-1}.$$

For the MSE matrix of the prediction $\hat{\boldsymbol{\mu}}$ for the group difference we obtain (neglecting the constant $\frac{1}{n_1 n_2}$) the following result:

$$\mathrm{MSE}_{\mu}(\mathbf{M}_1, \mathbf{M}_2) = \left[\left(\frac{n_2}{m_1}\mathbf{M}_1^{-1} + \frac{n_1}{m_2}\mathbf{M}_2^{-1}\right)^{-1} + \mathbf{C}\right]^{-1},$$

where $\mathbf{C}$ is defined after (5).

3.1 Optimal design for prediction of individual parameters

We define the linear (L-) criterion for the prediction of the individual parameters $\boldsymbol{\beta}$ as follows:

$$\phi_{L,\beta} = \sum_{i=1}^{2}\sum_{j=1}^{n_i} \mathrm{tr}\left\{\mathrm{E}\left[\left(\mathbf{L}_i^{\top}\hat{\boldsymbol{\beta}}_{ij} - \mathbf{L}_i^{\top}\boldsymbol{\beta}_{ij}\right)\left(\mathbf{L}_i^{\top}\hat{\boldsymbol{\beta}}_{ij} - \mathbf{L}_i^{\top}\boldsymbol{\beta}_{ij}\right)^{\top}\right]\right\}, \tag{9}$$

where $\mathbf{L}_i$ is a group-specific matrix for a linear transformation. Then we extend this definition for approximate designs and obtain the next result.

Theorem 2. *The L-criterion for the prediction of the individual parameters $\boldsymbol{\beta}$ is given by*

$$\begin{aligned}\phi_{L,\beta} \;=\; & \frac{1}{m_1}\mathrm{tr}\left\{\left[\mathbf{M}_1^{-1}\left(\mathbb{I}_p - \frac{n_2}{m_1}\tilde{\mathbf{U}}\mathbf{M}_1^{-1}\right) + (n_1-1)\left(\mathbf{M}_1 + \boldsymbol{\Delta}_1^{-1}\right)^{-1}\right]\mathbf{A}_1\right\} \\ & + \frac{1}{m_2}\mathrm{tr}\left\{\left[\mathbf{M}_2^{-1}\left(\mathbb{I}_p - \frac{n_1}{m_2}\tilde{\mathbf{U}}\mathbf{M}_2^{-1}\right) + (n_2-1)\left(\mathbf{M}_2 + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\right]\mathbf{A}_2\right\},\end{aligned} \tag{10}$$

where $\mathbf{A}_i = \mathbf{L}_i\mathbf{L}_i^{\top}$, $i = 1, 2$.

Well-known particular linear criteria are the A- and c-criteria with $\mathbf{L}_i = \mathbb{I}_p$, $i = 1, 2$, and $\mathbf{L}_i = \mathbf{c}_i$, $\mathbf{c}_i \in \mathbb{R}^p$, respectively. Another commonly used linear criterion is the IMSE-criterion. For the prediction of the individual parameters in two-groups model (1) we define this criterion as

$$\phi_{IMSE,\beta} = \sum_{i=1}^{2} a_i \sum_{j=1}^{n_i} \mathrm{tr}\left(\int_{\mathcal{X}_i} \mathrm{E}\left[\left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\beta}}_{ij} - \mathbf{F}_{(i)}(x)\boldsymbol{\beta}_{ij}\right)\left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\beta}}_{ij} - \mathbf{F}_{(i)}(x)\boldsymbol{\beta}_{ij}\right)^{\top}\right]\nu_i(\mathrm{d}x)\right), \tag{11}$$

where ν_i is a suitable measure on the experimental region $\mathcal{X}_i$ (typically uniform on $\mathcal{X}_i$) with $\nu_i(\mathcal{X}_i) = 1$ and a_i is a coefficient related to the i-th group, $a_1 + a_2 = 1$. The coefficients a_1 and a_2 may depend on the group sizes or, alternatively, equal weight may be given to each group. *IMSE*-criterion (21) may be represented in form

$$\phi_{IMSE,\beta} = \mathrm{tr}\left(\mathrm{Cov}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\mathrm{block\text{-}diag}\left(\mathbb{I}_{n_1}\otimes\mathcal{V}_1, \mathbb{I}_{n_2}\otimes\mathcal{V}_2\right)\right), \tag{12}$$

where $\mathcal{V}_i = a_i\int_{\mathcal{X}_i}\mathbf{F}_{(i)}(x)^{\top}\mathbf{F}_{(i)}(x)\,\nu_i(\mathrm{d}x)$, and results in the particular criterion (10) with $\mathbf{A}_i = \mathcal{V}_i$ for approximate designs.

The determinant (D-) criterion for the prediction of the individual parameters $\boldsymbol{\beta}$ may be defined as the logarithm of the determinant of the MSE matrix:

$$\phi_{D,\beta} = \ln\det\left(\mathrm{Cov}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right). \tag{13}$$

For approximate designs we obtain the next result.

Theorem 3. *The D-criterion for the prediction of the individual parameters $\boldsymbol{\beta}$ is given by*

$$\begin{aligned}\phi_{D,\beta} &= -\ln\det\left(\mathbf{M}_1\right) - \ln\det\left(\mathbf{M}_2\right) \\ &\quad - (n_1-1)\ln\det\left(\mathbf{M}_1+\boldsymbol{\Delta}_1^{-1}\right) - (n_2-1)\ln\det\left(\mathbf{M}_2+\boldsymbol{\Delta}_2^{-1}\right) \\ &\quad - \ln\det\left[\frac{n_2}{m_1}\left(\mathbf{M}_1^{-1}+\boldsymbol{\Delta}_1\right)+\frac{n_1}{m_2}\left(\mathbf{M}_2^{-1}+\boldsymbol{\Delta}_2\right)\right]. \end{aligned} \tag{14}$$

The proofs of Theorems 2 and 3 are deferred to Appendix B.

Note that in the particular case with only two individuals: $n_1 = n_2 = 1$, both *L*- and *D*-criteria significantly simplify:

$$\begin{aligned}\phi_{L,\beta} &= \frac{1}{m_1}\mathrm{tr}\left\{\left[\mathbf{M}_1^{-1}\tilde{\mathbf{U}}\left(\frac{1}{m_1}\boldsymbol{\Delta}_1+\frac{1}{m_2}(\mathbf{M}_2^{-1}+\boldsymbol{\Delta}_2)\right)\right]\mathbf{A}_1\right\} \\ &\quad + \frac{1}{m_2}\mathrm{tr}\left\{\left[\mathbf{M}_2^{-1}\tilde{\mathbf{U}}\left(\frac{1}{m_2}\boldsymbol{\Delta}_2+\frac{1}{m_1}(\mathbf{M}_1^{-1}+\boldsymbol{\Delta}_1)\right)\right]\mathbf{A}_2\right\},\end{aligned}$$

$$\phi_{D,\beta} = -\ln\det\left(\mathbf{M}_1\right) - \ln\det\left(\mathbf{M}_2\right) - \ln\det\left[\frac{n_2}{m_1}\left(\mathbf{M}_1^{-1}+\boldsymbol{\Delta}_1\right)+\frac{n_1}{m_2}\left(\mathbf{M}_2^{-1}+\boldsymbol{\Delta}_2\right)\right].$$

We will consider this case in more detail in Section 4.

Note also that in the case of identical groups: $\mathbf{F}_{(1)} = \mathbf{F}_{(2)}$, $m_1 = m_2$, $\mathcal{X}_1 = \mathcal{X}_2$, $\mathbf{D}_1 = \mathbf{D}_2$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, optimal designs in single-group models considered in Prus and Schwabe (2016) are optimal for the prediction of the individual parameters in model (1).

Further we verify the convexity of the proposed design criteria, to be able to formulate optimality conditions.

Lemma 2. *The L- and D-criteria for the prediction of the individual parameters are convex with respect to* $(\mathbf{M}_1, \mathbf{M}_2)$.

For the proof of Lemma 2 see Appendix C.

Note that the proposed *L*- and *D*-criteria depend on two designs simultaneously. Therefore, the general equivalence theorem (see Kiefer (1974)) cannot be used directly. We formulate optimality conditions using the results for multiple-design problems proposed in Prus (2021).

Theorem 4. *A pair of approximate designs $\xi^* = (\xi_1^*, \xi_2^*)$ is L-optimal for the prediction of the individual parameters $\boldsymbol{\beta}$ iff*

$$\begin{aligned}&\mathrm{tr}\left\{\tilde{\mathbf{F}}_{(1)}(x_1)\left[(n_1-1)\left(\mathbf{M}_1^*+\boldsymbol{\Delta}_1^{-1}\right)^{-1}\mathbf{A}_1\left(\mathbf{M}_1^*+\boldsymbol{\Delta}_1^{-1}\right)^{-1}\right.\right. \\ &\qquad \left.\left. + (\mathbf{M}_1^*)^{-1}\tilde{\mathbf{U}}^*\left(\mathbf{K}_1^*\mathbf{A}_1\mathbf{K}_1^* + \frac{n_1n_2}{m_2^2}(\mathbf{M}_2^*)^{-1}\mathbf{A}_2(\mathbf{M}_2^*)^{-1}\right)\tilde{\mathbf{U}}^*(\mathbf{M}_1^*)^{-1}\right]\tilde{\mathbf{F}}_{(1)}(x_1)^\top\right\} \\ &\le \mathrm{tr}\left\{\left[(n_1-1)\left(\mathbf{M}_1^*+\boldsymbol{\Delta}_1^{-1}\right)^{-1}\mathbf{A}_1\left(\mathbf{M}_1^*+\boldsymbol{\Delta}_1^{-1}\right)^{-1}\mathbf{M}_1^*\right.\right. \\ &\qquad \left.\left. + (\mathbf{M}_1^*)^{-1}\tilde{\mathbf{U}}^*\left(\mathbf{K}_1^*\mathbf{A}_1\mathbf{K}_1^* + \frac{n_1n_2}{m_2^2}(\mathbf{M}_2^*)^{-1}\mathbf{A}_2(\mathbf{M}_2^*)^{-1}\right)\tilde{\mathbf{U}}^*\right]\right\}, \quad x_1 \in \mathcal{X}_1, \end{aligned} \tag{15}$$

where

$$\tilde{\mathbf{U}}^* = \left[\frac{n_2}{m_1}((\mathbf{M}_1)^{*-1}+\boldsymbol{\Delta}_1)+\frac{n_1}{m_2}((\mathbf{M}_2^*)^{-1}+\boldsymbol{\Delta}_2)\right]^{-1}$$

and $\mathbf{K}_1^* = \frac{n_2}{m_1}\boldsymbol{\Delta}_1 + \frac{n_1}{m_2}((\mathbf{M}_2^*)^{-1} + \boldsymbol{\Delta}_2)$, *and*

$$\begin{aligned}
&\operatorname{tr}\left\{\tilde{\mathbf{F}}_{(2)}(x_2)\left[(n_2-1)\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\mathbf{A}_2\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\right.\right.\\
&\qquad\left.\left. + (\mathbf{M}_2^*)^{-1}\tilde{\mathbf{U}}^*\left(\mathbf{K}_2^*\mathbf{A}_2\mathbf{K}_2^* + \frac{n_1 n_2}{m_1^2}(\mathbf{M}_1^*)^{-1}\mathbf{A}_1(\mathbf{M}_1^*)^{-1}\right)\tilde{\mathbf{U}}^*(\mathbf{M}_2^*)^{-1}\right]\tilde{\mathbf{F}}_{(2)}(x_2)^\top\right\}\\
&\leq \operatorname{tr}\left\{\left[(n_2-1)\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\mathbf{A}_2\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\mathbf{M}_2^*\right.\right.\\
&\qquad\left.\left. + (\mathbf{M}_2^*)^{-1}\tilde{\mathbf{U}}^*\left(\mathbf{K}_2^*\mathbf{A}_2\mathbf{K}_2^* + \frac{n_1 n_2}{m_1^2}(\mathbf{M}_1^*)^{-1}\mathbf{A}_1(\mathbf{M}_1^*)^{-1}\right)\tilde{\mathbf{U}}^*\right]\right\}, \quad x_2 \in \mathcal{X}_2, \qquad (16)
\end{aligned}$$

where $\mathbf{K}_2^* = \frac{n_1}{m_2}\boldsymbol{\Delta}_2 + \frac{n_2}{m_1}((\mathbf{M}_1^*)^{-1} + \boldsymbol{\Delta}_1)$.

For support points of ξ_1^* *and* ξ_2^* *equality holds in* (15) *and* (16)*, respectively.*

Theorem 5. *A pair of approximate designs* $\xi^* = (\xi_1^*, \xi_2^*)$ *is D-optimal for the prediction of the individual parameters* $\boldsymbol{\beta}$ *iff*

$$\begin{aligned}
&\operatorname{tr}\left\{\tilde{\mathbf{F}}_{(1)}(x_1)\left((\mathbf{M}_1^*)^{-1} + (n_1-1)\left(\mathbf{M}_1^* + \boldsymbol{\Delta}_1^{-1}\right)^{-1} - \frac{n_2}{m_1}(\mathbf{M}_1^*)^{-1}\tilde{\mathbf{U}}^*(\mathbf{M}_1^*)^{-1}\right)\tilde{\mathbf{F}}_{(1)}(x_1)^\top\right\}\\
&\leq p + \operatorname{tr}\left\{(n_1-1)\left(\mathbf{M}_1^* + \boldsymbol{\Delta}_1^{-1}\right)^{-1}\mathbf{M}_1^* - \frac{n_2}{m_1}(\mathbf{M}_1^*)^{-1}\tilde{\mathbf{U}}^*\right\}, \quad x_1 \in \mathcal{X}_1 \qquad (17)
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{tr}\left\{\tilde{\mathbf{F}}_{(2)}(x_2)\left((\mathbf{M}_2^*)^{-1} + (n_2-1)\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1} - \frac{n_1}{m_2}(\mathbf{M}_2^*)^{-1}\tilde{\mathbf{U}}^*(\mathbf{M}_2^*)^{-1}\right)\tilde{\mathbf{F}}_{(2)}(x_2)^\top\right\}\\
&\leq p + \operatorname{tr}\left\{(n_2-1)\left(\mathbf{M}_2^* + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\mathbf{M}_2^* - \frac{n_1}{m_2}(\mathbf{M}_2^*)^{-1}\tilde{\mathbf{U}}^*\right\}, \quad x_2 \in \mathcal{X}_2. \qquad (18)
\end{aligned}$$

For support points of ξ_1^* *and* ξ_2^* *equality holds in* (17) *and* (18)*, respectively.*

Proofs of Theorems 4 and 5 have been deferred to appendix D.

3.2 Optimal designs for prediction of group difference

For the prediction of the group difference $\boldsymbol{\mu}$ we define the linear criterion as follows:

$$\phi_{L,\mu} = \operatorname{tr}\left\{\mathrm{E}\left[\left(\mathbf{L}^\top\hat{\boldsymbol{\mu}} - \mathbf{L}^\top\boldsymbol{\mu}\right)\left(\mathbf{L}^\top\hat{\boldsymbol{\mu}} - \mathbf{L}^\top\boldsymbol{\mu}\right)^\top\right]\right\}. \qquad (19)$$

For approximate designs we obtain the following result:

$$\phi_{L,\mu} = \operatorname{tr}\left\{\left[\left(\frac{n_2}{m_1}\mathbf{M}_1^{-1} + \frac{n_1}{m_2}\mathbf{M}_2^{-1}\right)^{-1} + \mathbf{C}\right]^{-1}\mathbf{A}\right\}, \qquad (20)$$

where $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$. Note that this criterion is not a particular linear criterion of form (10).

The *A*- and *c*-criteria for the prediction of the group difference are the particular linear criteria (of form (19)) with $\mathbf{L} = \mathbb{I}_p$ and $\mathbf{L} = \mathbf{c}$, $\mathbf{c} \in \mathbb{R}^p$, respectively. The related *IMSE*-criterion may be defined as

$$\phi_{IMSE,\mu} = \sum_{i=1}^{2} a_i \operatorname{tr}\left(\int_{\mathcal{X}_i} \mathrm{E}\left[\left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\mu}} - \mathbf{F}_{(i)}(x)\boldsymbol{\mu}\right)\left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\mu}} - \mathbf{F}_{(i)}(x)\boldsymbol{\mu}\right)^\top\right]\nu_i(\mathrm{d}x)\right), \qquad (21)$$

which results in the particular linear criterion with $\mathbf{A} = \sum_{i=1}^{2} a_i \int_{\mathcal{X}_i} \mathbf{F}_{(i)}(x)^\top \mathbf{F}_{(i)}(x)\, \nu_i(\mathrm{d}x)$.

The D-criterion for the prediction of the group difference may be defined as

$$\phi_{D,\mu} = \ln \det \left(\mathrm{Cov} \left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \right) \right).$$

For approximate designs we obtain

$$\phi_{D,\mu} = -\ln \det \left[\left(\frac{n_2}{m_1} \mathbf{M}_1^{-1} + \frac{n_1}{m_2} \mathbf{M}_2^{-1} \right)^{-1} + \mathbf{C} \right]. \tag{22}$$

Note that L- and D-optimal designs for the prediction of the group difference depend on the covariance matrices of the random effects only via their weighted sum $n_2\mathbf{D}_1 + n_1\mathbf{D}_2$. Note also that the designs depend on the group sizes only via the ratio n_1/n_2.

Lemma 3. *The L- and D-criteria for the prediction of the group difference are convex with respect to $(\mathbf{M}_1, \mathbf{M}_2)$.*

For the proof see Appendix C.

Now the optimality conditions for the linear and D-criteria for the prediction of the group difference can be formulated.

Theorem 6. *A pair of approximate designs $\xi^* = (\xi_1^*, \xi_2^*)$ is L-optimal for the prediction of the group difference $\boldsymbol{\mu}$ iff*

$$\begin{aligned}
&\mathrm{tr} \left\{ \tilde{\mathbf{F}}_{(i)}(x_i) (\mathbf{M}_i^*)^{-1} \left[\mathbb{I}_p + \mathbf{C} \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \right]^{-1} \mathbf{A} \right. \\
&\qquad \left. \left[\mathbb{I}_p + \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \mathbf{C} \right]^{-1} (\mathbf{M}_i^*)^{-1} \tilde{\mathbf{F}}_{(i)}(x_i)^\top \right\} \\
&\leq \mathrm{tr} \left\{ (\mathbf{M}_i^*)^{-1} \left[\mathbb{I}_p + \mathbf{C} \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \right]^{-1} \mathbf{A} \right. \\
&\qquad \left. \left[\mathbb{I}_p + \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \mathbf{C} \right]^{-1} \right\}, \quad \forall\, x_i \in \mathcal{X}_i, \quad i, i' = 1, 2, \quad i \neq i',
\end{aligned} \tag{23}$$

where $\mathbf{M}_i^* = \mathbf{M}_i(\xi_i^*)$.

For support points of ξ_i^* equality holds in (23).

Theorem 7. *A pair of approximate designs $\xi^* = (\xi_1^*, \xi_2^*)$ is D-optimal for the prediction of the group difference $\boldsymbol{\mu}$ iff*

$$\begin{aligned}
&\mathrm{tr} \left\{ \tilde{\mathbf{F}}_{(i)}(x_i) (\mathbf{M}_i^*)^{-1} \left[\mathbb{I}_p + \mathbf{C} \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \right]^{-1} \right. \\
&\qquad \left. \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right)^{-1} (\mathbf{M}_i^*)^{-1} \tilde{\mathbf{F}}_{(i)}(x_i)^\top \right\} \\
&\leq \mathrm{tr} \left\{ (\mathbf{M}_i^*)^{-1} \left[\mathbb{I}_p + \mathbf{C} \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right) \right]^{-1} \left(\frac{n_{i'}}{m_i} (\mathbf{M}_i^*)^{-1} + \frac{n_i}{m_{i'}} (\mathbf{M}_{i'}^*)^{-1} \right)^{-1} \right\},
\end{aligned} \tag{24}$$

for all $x_i \in \mathcal{X}_i$, $i, i' = 1, 2$, $i \neq i'$.

For support points of ξ_i^* equality holds in (24).

The proofs of Theorems 6 and 7 are deferred to Appendix E.

Further we consider the particular two-groups model of form (1) in which the regression matrices, the experimental regions and the covariance matrices of the errors are the same for both groups: $\mathbf{F}_{(1)} = \mathbf{F}_{(2)} =: \mathbf{F}_{(0)}$, $\Sigma_1 = \Sigma_2 =: \Sigma_0$ and $\mathcal{X}_1 = \mathcal{X}_2 =: \mathcal{X}_0$. In this case optimal approximate designs with respect to the Bayesian L- or D-criteria (see e. g. Gladitz and Pilz (1982))

$$\phi_{L,B} = \operatorname{tr}\left[\left(\mathbf{M} + \mathbf{D}^{-1}\right)^{-1}\mathbf{A}\right] \tag{25}$$

and

$$\phi_{D,B} = -\ln\det\left(\mathbf{M} + \mathbf{D}^{-1}\right),$$

where $\mathbf{M}$ denotes the moment matrix and $\mathbf{D}$ is the covariance matrix of random effects, are optimal for the prediction of the group difference in model (1).

Theorem 8. *Let the regression matrices, the experimental regions and the covariance matrices of the observational errors in model (1) be the same for both groups. Let ξ_b^* be the Bayesian L- or D-optimal approximate design for the covariance matrix of random effects given by*

$$\mathbf{D} = \frac{m_1 m_2}{n_1 m_1 + n_2 m_2}(n_2\mathbf{D}_1 + n_1\mathbf{D}_2).$$

Then the pair of approximate designs (ξ_1^*, ξ_2^*) with $\xi_i^* = \xi_b^*$, $i = 1, 2$, is, respectively, L- or D-optimal for the prediction of the group difference $\boldsymbol{\mu}$.

For the proof see Appendix E.

Note that the result of Theorem 8 holds also in cases where the covariance matrices of random effects, the group sizes and (or) the numbers of observations per individual are not the same in both groups: $\mathbf{D}_1 \neq \mathbf{D}_2$, $n_1 \neq n_2$ and (or) $m_1 \neq m_2$ (see the next section for numerical examples).

4 Examples

In this section we consider numerical examples for the straight line and bi-linear regression models. Because of complexity of the design criteria for the prediction of the individual parameters, we restrict ourselves to the case of two individuals: $n_1 = n_2 = 1$, and we focus on the D-criterion, which is the simplest one from the mathematical point of view. These group sizes may be suitable, for example, in the situation, where the observational units (that we for simplicity call "individuals") are some studies and m_i observations in study i are related to m_i animals/plants/etc. available for the study.

As it has been established in part 3.2, optimal designs for the prediction of the group difference depend on the group sizes only via the ratio n_1/n_2. For the group difference we consider more general situations with larger group sizes and we compute also A-optimal designs.

4.1 Linear regression

We consider the two-groups model of general form (1) with the regression functions $\mathbf{F}_{(i)}(x) = (1, x)$ on the design region $\mathcal{X}_i = [0, 1]$, $i = 1, 2$:

$$Y_{ij\ell} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2}x_{i\ell} + \boldsymbol{\varepsilon}_{ij\ell}, \quad \ell = 1, \ldots, m_i, \quad j = 1, \ldots, n_i. \tag{26}$$

The covariance structures of the random effects and observational errors are given by $\mathbf{D}_i = \operatorname{diag}(d_1, d_2)$ and $\Sigma_i = 1$ for both groups. Group sizes n_i and numbers observations per individual m_i are in general not the same for the first and the second group.

Table 1: D-optimal approximate designs for prediction of the individual parameters $\boldsymbol{\beta}$ and group difference $\boldsymbol{\mu}$ in straight line regression model for $d_1 = 0.1$, $d_2 = 1$, $n_1 = n_2 = 1$.

Numbers of observations			Optimal designs for $\hat{\boldsymbol{\beta}}$				Optimal designs for $\hat{\boldsymbol{\mu}}$	
m_1	m_2	m_1/m_2	w_1^*	$1-w_1^*$	w_2^*	$1-w_2^*$	w^*	$1-w^*$
6	24	1/4	0.772	0.228	0.514	0.486	≈ 1	≈ 0
9	21	3/7	0.669	0.331	0.535	0.465	0.897	0.103
12	18	2/3	0.616	0.384	0.557	0.443	0.847	0.153
15	**15**	**1**	**0.582**	**0.418**	**0.582**	**0.418**	**0.833**	**0.167**
18	12	3/2	0.557	0.443	0.616	0.384	0.847	0.153
21	9	7/3	0.535	0.465	0.669	0.331	0.897	0.103
24	6	4	0.514	0.486	0.772	0.228	≈ 1	≈ 0

For the predictions of the individual parameters and the group differences the L- and D-optimal approximate designs have the form

$$\xi_i = \begin{pmatrix} 0 & 1 \\ 1-w_i & w_i \end{pmatrix},$$

where w_i denotes the weight of observations in point 1 for the i-th group, and the moment matrices are given by

$$\mathbf{M}(\xi_i) = \begin{pmatrix} 1 & w_i \\ w_i & w_i \end{pmatrix}.$$

According to Theorem 8, optimal designs for the prediction of the group difference are the same for both groups:

$$\xi_i = \begin{pmatrix} 0 & 1 \\ 1-w & w \end{pmatrix}, \quad i = 1, 2.$$

Then only the optimal weight w^* has to be computed. Table 1 illustrates the behavior of D-optimal designs in both cases: for the prediction of individual parameters and the group difference. We fix the values of the variance parameters and the group sizes by $d_1 = 0.1$, $d_2 = 1$, $n_1 = n_2 = 1$.

As we can see in the table, optimal weights for the prediction of the individual parameters are in general different in the first and the second group and depend on the numbers of observations m_1 and m_2. In case $m_1 = m_2 = 15$ the optimal weight $w_1^* = w_2^*$ is in accordance with the optimal design in the one-group model considered in Prus and Schwabe (2016). Optimal design for the prediction of the group difference coincides for $m_1 = m_2$ with the Bayesian optimal design for the covariance matrix of random effects $\mathbf{D} = \mathbf{D}_1 = \mathbf{D}_2$. In case $m_1 = 6$ and $m_2 = 24$ (or $m_1 = 24$ and $m_2 = 6$) a singular design minimizes the D-criterion. Such designs, however, lead to a singular moment matrix and are, therefore, not admissible. The best exact design in this case assigns 5 and 23 observations at point 1 in the first and the second group, respectively. Note that in all cases considered in Table 1 optimal weights at point 1 (w_1^*, w_2^*, w^*) are larger than 0.5, which is the optimal weight in the fixed effects model ($d_1 = d_2 = 0$). This behavior may be explained by the fact that the slope variance d_2 is larger than the intercept variance d_1 and, therefore, more observations at point 1 are needed.

The optimal designs in Table 1 have been computed using the standard approach for determining the minimum of a convex function. The related equation / system of equations (gradient equals 0) has been solved numerically using software Maple 2020. The derivatives are not presented here because of their complexity.

Further we consider the A-criterion for the prediction of the group difference. To compute A-optimal designs we use "OptimalDesign" package in R proposed by Harman and Filová (2019). This package has been developed for classical design criteria and allows for additional constrains. We convert the Bayesian A-criterion to the standard A-criterion with specific constraints using the approach proposed in Harman and Prus (2018).

Figure 1 illustrates the behavior of the optimal weight in relation to the ratio $a = m_1/m_2$ of the numbers of observations (for fixed values of m_2) and the ratio $b = n_2/n_1$ of the group sizes. The variance parameters are fixed by $d_1 = 0.1$ and $d_2 = 1$. We consider the behavior for all values of a from the interval $[1, 5]$. As we can see on the graphics, the optimal weight decreases with increasing ratio a. Also for larger values of m_2 the optimal weight becomes smaller. Note that

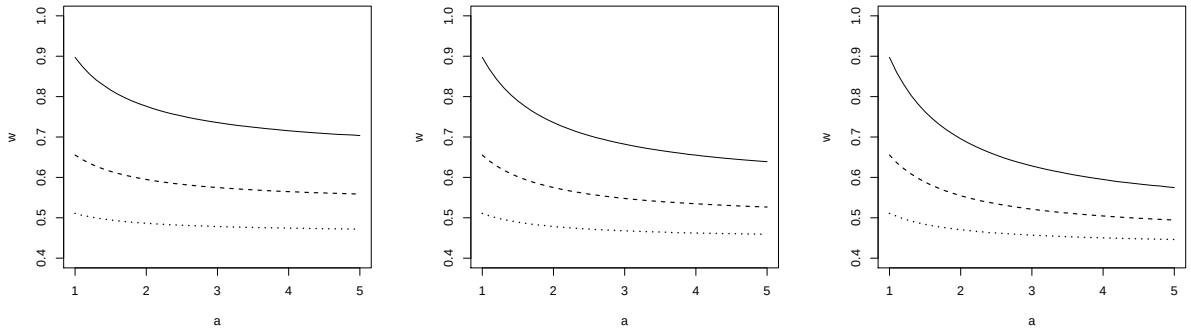

Figure 1: A-optimal weight w^* for prediction of group difference in straight line regression model in dependence on ratio a of numbers of observations for $m_2 = 8$ (solid line), $m_2 = 16$ (dashed line), $m_2 = 40$ (dotted line) and $b = 1$ (left panel), $b = 2$ (middle panel), $b = 5$ (right panel).

for the A-criterion (in contrast to the D-criterion) not all optimal weights at point 1 are higher than 0.5. However they are all higher than the optimal weight in the fixed effects model (≈ 0.41 for A-criterion). We can also observe that for larger values of both m_1 and m_2 (solid lines and larger values of a) optimal weights become more close 0.41, i. e. they become less sensitive with respect to variance parameters. An intuitive explanation may as follows: If many observations are possible, there is no need to require much more observations at point 1 to receive enough information about the slope.

4.2 Bi-linear regression

We consider the two-groups bi-linear model with the regression functions $\mathbf{F}_{(i)}(x) = (1, x_1, x_2)$ on the design region $\mathcal{X}_i = [-1, 1]^2$, $i = 1, 2$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih1} + \boldsymbol{\beta}_{ij3} x_{ih2} + \boldsymbol{\varepsilon}_{ijh}, \quad j = 1, \ldots, n_i, \quad h = 1, \ldots, m_i.$$

According to Theorem 8, Bayesian optimal designs are optimal for the prediction of the group difference. The left hand side of the optimality condition (34) for Bayesian linear criteria is a convex paraboloid. Therefore, the only admissible support points for optimal designs are $x_1 = (1, 1)$, $x_2 = (1, -1)$, $x_3 = (-1, 1)$, $x_4 = (-1, -1)$:

$$\xi_i = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ w_{i1} & w_{i2} & w_{i3} & w_{i4} \end{pmatrix},$$

where $\sum_{h=1}^{4} w_{ih} = 1$. The covariance structures of the random effects and observational errors are given by

$$\mathbf{D}_i = \begin{pmatrix} d_{11} & 0 & d_{13} \\ 0 & d_{22} & 0 \\ d_{13} & 0 & d_{33} \end{pmatrix} \tag{27}$$

and $\Sigma_i = 1$ for both groups. For this model the A-optimal designs turned out to have the following design structure:

$$\xi_i = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ w & \frac{1}{2}(1-2w) & w & \frac{1}{2}(1-2w) \end{pmatrix}.$$

Then only the optimal weight w^* at points $(1,1)$ and $(-1,1)$ has to be computed. This design structure can be explained by the equivariance property of the Bayesian A-criterion.

To compute the optimal designs we use "OptimalDesign" package. As it also has been done in section 4.1 for the A-criterion for linear regression, we convert the Bayesian A-criterion to the standard A-criterion with specific constraints.

Figure 2 illustrates the behavior of the A-optimal weight in relation to the ratio $a = m_1/m_2$ of the numbers of observations (for fixed values of m_2) and the ratio $b = n_2/n_1$ of the group sizes. The variance parameters are fixed by $d_{11} = d_{22} = d_{33} = 1$ and the values 0.9, 0.5, -0.5 or -0.9 of d_{13}. We consider the behavior for all values of a from the interval $[1, 5]$. Note that for $d_{13} = 0$ the balanced design $w^* = 0.25$ is optimal. For small values of d_{13} optimal designs are close to the balanced design. Note also that in case of a positive (negative) correlation d_{13} the optimal weight is decreasing (increasing) with increasing values of ratio a. We can also see on the graphics that for larger values of both m_1 and m_2 optimal weights become close to 0.25, which is the optimal weight in the fixed effects bi-linear regression model. The same behavior we observed for the linear regression model in part 4.1.

5 Discussion

In this paper we have considered optimal designs for the prediction of the individual parameters and the group difference in the two-groups mixed models with multivariate response. The solution for optimal approximate designs is given in the form of optimality conditions for the linear and D-criteria. Optimal designs for the prediction of the group differences coincide, in the particular case of the same regression function and the same design region for both groups, with the Bayesian optimal designs for an adjusted covariance matrix of random effects. For Bayesian linear criteria, optimal approximate and exact designs may be computed using "OptimalDesign" package in R. For the prediction of the individual parameters, the problem of computation remains, however, open. Also for determining optimal designs for the group difference in models where regression functions and / or design regions differ between groups, a new method has to be developed. If optimal approximate designs are already available and the number of weights (variables in the criterion function) is low, as for example in Section 4, the related optimal exact design can be determined by choosing the "best" of the neighbouring exact designs. Otherwise some more complicated rounding strategy should be used. Moreover, the proposed approach requires non-singular moment (or information) matrices, which is only possible if the numbers of observations per individual are large enough. However, in many practical situations the number of observations is restricted and may be even smaller than the number of parameters. For determining of optimal designs in such cases a new approach has to be developed, which may be a subject of future investigations. Furthermore, in this work we have restricted ourselves on the situation with two groups of observational units. An extension of the analytical results to a larger number of groups is (mathematically) challenging. However, for particular regression matrices the problem may simplify significantly. In our model we have assumed the population parameters to be the same for both groups. Although this assumption seems to be restrictive, statistical analysis for random effects in the case of different population parameters turned out to be rather simple as they may be performed separately in each group (see Prus (2015), ch. 6). Optimal designs for the prediction of the group difference may be, however, a reasonable direction for a

future research. Finally, it is worth-while mentioning that optimal designs obtained by using the proposed approach depend on the covariance matrix of the random effects, which is in practice unknown and has to be estimated. The quality of the resulting designs, however, depends on the accuracy of the estimation. Alternatively, robust design criteria, as maximin criterion for example, that are not sensitive with respect to variance parameters, may be considered in the future.

Acknowledgment

This research was supported by grant SCHW 531/16 of the German Research Foundation (DFG). The author is grateful to Norbert Gaffke for fruitful discussions on the optimality conditions and the proof of Lemma 2. The author thanks two referees and the Editor-in-chief for helpful comments which improved the presentation of the results.

A Proofs of Theorem 1 and Lemma 1

To make use of the theoretical results that are available in the literature (see e. g. Henderson (1975)) for the prediction of random parameters we will represent model (1) as a particular case of the general linear mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \tag{28}$$

with design matrices $\mathbf{X}$ and $\mathbf{Z}$ for the fixed effects and the random effects, respectively. In (28), $\boldsymbol{\beta}_0$ denotes the fixed effects and $\mathbf{b}$ are the random effects. The random effects and the observational errors $\boldsymbol{\varepsilon}$ are assumed to have zero mean and to be all uncorrelated with positive definite covariance matrices $\mathrm{Cov}\,(\mathbf{b}) = \mathbf{G}$ and $\mathrm{Cov}\,(\boldsymbol{\varepsilon}) = \mathbf{R}$, respectively.

Our model (1) may be rewritten in the following form:

$$\mathbf{Y} = \begin{pmatrix} \mathbb{1}_{n_1} \otimes \mathbf{F}_1 \\ \mathbb{1}_{n_2} \otimes \mathbf{F}_2 \end{pmatrix} \beta_0 + \begin{pmatrix} \mathbb{I}_{n_1} \otimes \mathbf{F}_1 & 0 \\ 0 & \mathbb{I}_{n_2} \otimes \mathbf{F}_2 \end{pmatrix} \mathbf{b} + \varepsilon, \tag{29}$$

where $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ for $\mathbf{Y}_i = (\mathbf{Y}_{i1}^\top, \ldots, \mathbf{Y}_{in_i}^\top)^\top$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \boldsymbol{\varepsilon}_2^\top)^\top$ for $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}^\top, \ldots, \boldsymbol{\varepsilon}_{in_i}^\top)^\top$, and $\mathbf{b} = \boldsymbol{\beta} - \mathbb{1}_n \otimes \boldsymbol{\beta}_0$. Then model (29) is of form (28) with the design matrices

$$\mathbf{X} = \begin{pmatrix} \mathbb{1}_{n_1} \otimes \mathbf{F}_1 \\ \mathbb{1}_{n_2} \otimes \mathbf{F}_2 \end{pmatrix}$$

and $\mathbf{Z} = \text{block-diag}\,(\mathbb{I}_{n_1} \otimes \mathbf{F}_1, \mathbb{I}_{n_2} \otimes \mathbf{F}_2)$ and the covariance matrices $\mathbf{G} = \text{block-diag}\,(\mathbb{I}_{n_1} \otimes \mathbf{D}_1, \mathbb{I}_{n_2} \otimes \mathbf{D}_2)$ and $\mathbf{R} = \text{block-diag}\,(\mathbb{I}_{n_1} \otimes \mathbf{V}_1, \mathbb{I}_{n_2} \otimes \mathbf{V}_2)$.

According to Henderson (1975) the MSE matrix of the BLUP $\left(\hat{\boldsymbol{\beta}}_0^\top, \hat{\mathbf{b}}\right)^\top$ is given by the block-matrix

$$\mathrm{Cov} \begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ \hat{\mathbf{b}} - \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\top & \mathbf{C}_{22} \end{pmatrix},$$

where

$$\mathbf{C}_{22} = \left(\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\right)^{-1},$$

$$\mathbf{C}_{12} = -\mathbf{C}_{11}\,\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}$$

and $\mathbf{C}_{11}$ corresponds to the covariance matrix of the BLUE $\hat{\boldsymbol{\beta}}_0$ and is given by formula (3). The BLUP for the random effects $\mathbf{b}$ is given by

$$\hat{\mathbf{b}} = \mathbf{G}\mathbf{Z}^\top(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R})^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0).$$

Then for any predictable linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta}_0 + \mathbf{S}\mathbf{b}$, where $\mathbf{K}$ and $\mathbf{S}$ are some matrices with suitable dimensions, the BLUP is equal to

$$\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}}_0 + \mathbf{S}\hat{\mathbf{b}} \tag{30}$$

and the MSE matrix of $\hat{\boldsymbol{\Psi}}$ can be computed as

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \mathbf{K}\,\mathbf{C}_{11}\,\mathbf{K}^\top + \mathbf{K}\,\mathbf{C}_{12}\,\mathbf{S}^\top + \mathbf{S}\,\mathbf{C}_{12}^\top\,\mathbf{K}^\top + \mathbf{S}\,\mathbf{C}_{22}\,\mathbf{S}^\top. \tag{31}$$

Then using the relation $\boldsymbol{\beta} = (\mathbb{1}_n \otimes \mathbb{I}_p)\boldsymbol{\beta}_0 + \mathbf{b}$ and the two latter formulae (30) and (31) we obtain, respectively, the results of Theorem 1 and Lemma 1.

B Proofs of Theorems 2 and 3

B.1 Proof of Theorem 2

The L-criterion (9) may be rewritten as the following function of the mean squared error matrix of the random effects:

$$\phi_{L,\beta} = \mathrm{tr}\left[\mathrm{Cov}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\,\text{block-diag}\left(\mathbb{I}_{n_1} \otimes \mathbf{A}_1, \mathbb{I}_{n_2} \otimes \mathbf{A}_2\right)\right], \tag{32}$$

where block-diag $(\mathbf{P}_1, \mathbf{P}_2)$ denotes the block-diagonal matrix with blocks $\mathbf{P}_1$ and $\mathbf{P}_2$. Then criterion (10) follows directly from formulae (32) and (8).

B.2 Proof of Theorem 3

To compute the determinant of the mean squared error matrix (8) we use the formula for the determinant of a block-matrix:

$$\det\left(\mathrm{MSE}_\beta(\mathbf{M}_1, \mathbf{M}_2)\right) = \det(\tilde{\mathbf{B}}_{11})\det(\tilde{\mathbf{B}}_{22} - \tilde{\mathbf{B}}_{12}^\top\tilde{\mathbf{B}}_{11}^{-1}\tilde{\mathbf{B}}_{12}).$$

After employing some linear algebra and some properties of the logarithm we obtain D-criterion (14).

C Proofs of Lemmas 2 and 3

C.1 Proof of Lemma 2

The inverse of the MSE matrix (8) is given by

$$\left(\mathrm{MSE}_\beta(\mathbf{M}_1, \mathbf{M}_2)\right)^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix}, \tag{33}$$

where

$$\mathbf{A}_{11} = m_1\left\{\mathbb{I}_{n_1} \otimes \mathbf{M}_1 + (\mathbb{I}_{n_1} - \mathbb{J}_{n_1}) \otimes \boldsymbol{\Delta}_1^{-1} + \mathbb{J}_{n_1} \otimes \left(\boldsymbol{\Delta}_1 + \frac{n_1 m_1}{m_2 n_2}\boldsymbol{\Delta}_2\right)^{-1}\right\},$$

$$\mathbf{A}_{12} = -(\mathbb{1}_{n_1}\mathbb{1}_{n_2}^\top) \otimes \left(\frac{n_2}{m_1}\mathbf{\Delta}_1 + \frac{n_1}{m_2}\mathbf{\Delta}_2\right)^{-1}$$

and

$$\mathbf{A}_{22} = m_2 \left\{ \mathbb{I}_{n_2} \otimes \mathbf{M}_2 + (\mathbb{I}_{n_2} - \mathbb{J}_{n_2}) \otimes \mathbf{\Delta}_2^{-1} + \mathbb{J}_{n_2} \otimes \left(\mathbf{\Delta}_2 + \frac{n_2 m_2}{m_1 n_1}\mathbf{\Delta}_1\right)^{-1} \right\}.$$

This matrix is linear in $\mathbf{M}_1$ and $\mathbf{M}_2$, which implies the convexity of the both criteria.

C.2 Proof of Lemma 3

The function $\phi(\mathbf{N}) = \mathbf{N}^{-1}$ is non-increasing in Loewner ordering and matrix-convex for any positive definite matrix $\mathbf{N}$. Then $\phi_1(\mathbf{M}_1, \mathbf{M}_2) = \left(a_1\mathbf{M}_1^{-1} + a_2\mathbf{M}_2^{-1}\right)^{-1} + \mathbf{C}$ is matrix-concave with respect to $(\mathbf{M}_1, \mathbf{M}_2)$ for all positive definite $\mathbf{M}_1$ and $\mathbf{M}_2$, $\mathbf{C} \in \mathbb{R}^{p\times p}$, $a_i > 0$, $i = 1, 2$. The functions $\phi_2(\mathbf{N}) = -\ln\det(\mathbf{N})$ and $\phi_3(\mathbf{N}) = \operatorname{tr}\left(\mathbf{N}^{-1}\mathbf{A}\right)$ are non-increasing in Loewner ordering and convex for any positive definite matrix $\mathbf{N}$ and any positive semi-definite matrix $\mathbf{A}$ as the standard D- and L-criteria (see e. g. Pázman (1986), ch. 4). Then the functions $\phi_2 \circ \phi_1$ and $\phi_3 \circ \phi_1$ are convex.

D Proofs of Theorems 4 and 5

According to Theorem 2 in Prus (2021) a pair of designs $\xi^* = (\xi_1^*, \xi_2^*)$ is optimal with respect to a convex criterion $\phi(\mathbf{M}_1, \mathbf{M}_2)$ if and only if the directional derivative of ϕ at $(\mathbf{M}_i, \mathbf{M}_{i'})$ in the direction of $(\tilde{\mathbf{M}}_i, \mathbf{M}_{i'})$, given by the formula

$$\mathbf{\Phi}_{M_{i'}}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) = \lim_{\alpha \searrow 0} \frac{1}{\alpha}\left(\phi((1-\alpha)\mathbf{M}_i + \alpha\tilde{\mathbf{M}}_i, \mathbf{M}_{i'}) - \phi(\mathbf{M}_i, \mathbf{M}_{i'})\right),$$

is larger than or equal to zero for $\mathbf{M}_i = \mathbf{M}_i(\xi_i^*)$, $\tilde{\mathbf{M}}_i = \tilde{\mathbf{F}}_{(i)}(x_i)^\top\tilde{\mathbf{F}}_{(i)}(x_i)$, all points x_i of the experimental region $\mathcal{X}_i$, and $i, i' = 1, 2$, $i \neq i'$. For all support points x_i of ξ_i^* the derivative has to be equal to zero.

D.1 Proof of Theorem 4

For L-criterion (10) the directional derivatives of $\phi_{L,\beta}$ at $(\mathbf{M}_i, \mathbf{M}_{i'})$ in the direction of $(\tilde{\mathbf{M}}_i, \mathbf{M}_{i'})$ for $i, i' = 1, 2$, $i \neq i'$, are given by

$$\begin{aligned}\mathbf{\Phi}_{M_2}(\mathbf{M}_1, \tilde{\mathbf{M}}_1) \;=\; -\frac{1}{m_1}\operatorname{tr}\Big\{(\tilde{\mathbf{M}}_1 - \mathbf{M}_1)\Big[(n_1-1)\left(\mathbf{M}_1 + \mathbf{\Delta}_1^{-1}\right)^{-1}\mathbf{A}_1\left(\mathbf{M}_1 + \mathbf{\Delta}_1^{-1}\right)^{-1} \\ + \mathbf{M}_1^{-1}\tilde{\mathbf{U}}\left(\mathbf{K}_1\mathbf{A}_1\mathbf{K}_1 + \frac{n_1 n_2}{m_2^2}\mathbf{M}_2^{-1}\mathbf{A}_2\mathbf{M}_2^{-1}\right)\tilde{\mathbf{U}}\mathbf{M}_1^{-1}\Big]\Big\}\end{aligned}$$

and

$$\begin{aligned}\mathbf{\Phi}_{M_1}(\mathbf{M}_2, \tilde{\mathbf{M}}_2) \;=\; -\frac{1}{m_2}\operatorname{tr}\Big\{(\tilde{\mathbf{M}}_2 - \mathbf{M}_2)\Big[(n_2-1)\left(\mathbf{M}_2 + \mathbf{\Delta}_2^{-1}\right)^{-1}\mathbf{A}_2\left(\mathbf{M}_2 + \mathbf{\Delta}_2^{-1}\right)^{-1} \\ + \mathbf{M}_2^{-1}\tilde{\mathbf{U}}\left(\mathbf{K}_2\mathbf{A}_2\mathbf{K}_2 + \frac{n_1 n_2}{m_1^2}\mathbf{M}_1^{-1}\mathbf{A}_1\mathbf{M}_1^{-1}\right)\tilde{\mathbf{U}}\mathbf{M}_2^{-1}\Big]\Big\},\end{aligned}$$

where $\mathbf{K}_1 = \frac{n_2}{m_1}\mathbf{\Delta}_1 + \frac{n_1}{m_2}(\mathbf{M}_2^{-1} + \mathbf{\Delta}_2)$ and $\mathbf{K}_2 = \frac{n_1}{m_2}\mathbf{\Delta}_2 + \frac{n_2}{m_1}(\mathbf{M}_1^{-1} + \mathbf{\Delta}_1)$. Then the optimality condition follows directly from Theorem 2 in Prus (2021).

D.2 Proof of Theorem 5

For D-criterion (14) the directional derivatives of $\phi_{D,\beta}$ at $(\mathbf{M}_i, \mathbf{M}_{i'})$ in the direction of $(\tilde{\mathbf{M}}_i, \mathbf{M}_{i'})$ for $i, i' = 1, 2$, $i \neq i'$, are given by

$$\begin{aligned}\boldsymbol{\Phi}_{M_2}(\mathbf{M}_1, \tilde{\mathbf{M}}_1) &= p + \operatorname{tr}\left\{(n_1 - 1)\left(\mathbf{M}_1 + \boldsymbol{\Delta}_1^{-1}\right)^{-1}\mathbf{M}_1 - \frac{n_2}{m_1}\mathbf{M}_1^{-1}\tilde{\mathbf{U}}\right\} \\ &\quad -\operatorname{tr}\left\{\left(\mathbf{M}_1^{-1} + (n_1 - 1)\left(\mathbf{M}_1 + \boldsymbol{\Delta}_1^{-1}\right)^{-1} - \frac{n_2}{m_1}\mathbf{M}_1^{-1}\tilde{\mathbf{U}}\mathbf{M}_1^{-1}\right)\tilde{\mathbf{M}}_1\right\}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Phi}_{M_1}(\mathbf{M}_2, \tilde{\mathbf{M}}_2) &= p + \operatorname{tr}\left\{(n_2 - 1)\left(\mathbf{M}_2 + \boldsymbol{\Delta}_2^{-1}\right)^{-1}\mathbf{M}_2 - \frac{n_1}{m_2}\mathbf{M}_2^{-1}\tilde{\mathbf{U}}\right\} \\ &\quad -\operatorname{tr}\left\{\left(\mathbf{M}_2^{-1} + (n_2 - 1)\left(\mathbf{M}_2 + \boldsymbol{\Delta}_2^{-1}\right)^{-1} - \frac{n_1}{m_2}\mathbf{M}_2^{-1}\tilde{\mathbf{U}}\mathbf{M}_2^{-1}\right)\tilde{\mathbf{M}}_2\right\}.\end{aligned}$$

E Proofs of Theorems 6, 7 and 8

To verify the results of Theorems 6 and 7, we use the same argumentation as in the proofs of Theorem 4 and 5.

E.1 Proof of Theorem 6

For linear criterion (20) the directional derivatives are given by

$$\begin{aligned}\boldsymbol{\Phi}_{M_{i'}}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) &= \frac{n_{i'}}{m_i}\operatorname{tr}\left\{\left[\mathbb{I}_p + \left(\frac{n_{i'}}{m_i}\mathbf{M}_i^{-1} + \frac{n_i}{m_{i'}}\mathbf{M}_{i'}^{-1}\right)\mathbf{C}\right]^{-1}\mathbf{M}_i^{-1}(\mathbf{M}_i - \tilde{\mathbf{M}}_i)\mathbf{M}_i^{-1}\right. \\ &\qquad \left.\left[\mathbb{I}_p + \mathbf{C}\left(\frac{n_{i'}}{m_i}\mathbf{M}_i^{-1} + \frac{n_i}{m_{i'}}\mathbf{M}_{i'}^{-1}\right)\right]^{-1}\mathbf{A}\right\}.\end{aligned}$$

E.2 Proof of Theorem 7

For D-criterion (22) the directional derivative of $\phi_{D,\mu}$ at $(\mathbf{M}_i, \mathbf{M}_{i'})$ in the direction of $(\tilde{\mathbf{M}}_i, \mathbf{M}_{i'})$ is given by

$$\begin{aligned}\boldsymbol{\Phi}_{M_{i'}}(\mathbf{M}_i, \tilde{\mathbf{M}}_i) &= \frac{n_{i'}}{m_i}\operatorname{tr}\left\{\left[\mathbb{I}_p + \left(\frac{n_{i'}}{m_i}\mathbf{M}_i^{-1} + \frac{n_i}{m_{i'}}\mathbf{M}_{i'}^{-1}\right)\mathbf{C}\right]^{-1}\mathbf{M}_i^{-1}(\mathbf{M}_i - \tilde{\mathbf{M}}_i)\mathbf{M}_i^{-1}\right. \\ &\qquad \left.\left(\frac{n_{i'}}{m_i}\mathbf{M}_i^{-1} + \frac{n_i}{m_{i'}}\mathbf{M}_{i'}^{-1}\right)^{-1}\right\}.\end{aligned}$$

E.3 Proof of Theorem 8

We will prove this theorem for the linear criterion only. For the D-criterion the result can be easily verified using the same approach.

Let the group-designs be the same for both groups: $\xi_1 = \xi_2 =: \xi_0$. Then we obtain $\mathbf{M}_1 = \mathbf{M}_2 =: \mathbf{M}_0$ and inequality (23) from Theorem 6 simplifies to

$$\begin{aligned}&\operatorname{tr}\left\{\tilde{\mathbf{F}}_{(0)}(x)\left(\mathbf{M}_0^* + \mathbf{D}^{-1}\right)^{-1}\mathbf{A}\left(\mathbf{M}_0^* + \mathbf{D}^{-1}\right)^{-1}\tilde{\mathbf{F}}_{(0)}(x)^\top\right\} \\ &\leq \operatorname{tr}\left\{\mathbf{M}_0^*\left(\mathbf{M}_0^* + \mathbf{D}^{-1}\right)^{-1}\mathbf{A}\left(\mathbf{M}_0^* + \mathbf{D}^{-1}\right)^{-1}\right\}, \quad \forall\, x \in \mathcal{X}_0, \qquad (34)\end{aligned}$$

which coincides with the optimality condition for the Bayesian designs (see e. g. Gladitz and Pilz (1982)). Consequently, optimal designs with respect to Bayesian linear criterion (25) are optimal for the prediction of the group difference.

References

Dumont, C., Lestini, G., Nagard, H. L., Mentré, F., Comets, E., and Nguyen, T. T. (2018). PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Computer methods and programs in biomedicine*, **156**, 217–229.

Fedorov, V. and Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248.

Filová, L. and Prus, M. (2021). Computational aspects of experimental designs in multiple-group mixed models. Preprint available under www.iam.fmph.uniba.sk/ospm/Filova/papers/RCR12.pdf.

Gladitz, J. and Pilz, J. (1982). Construction of optimal designs in random coefficient regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, **13**, 371–385.

Harman, R. and Filová, L. (2019). Package 'OptimalDesign'. https://cran.r-project.org/web/packages/OptimalDesign/index.html.

Harman, R. and Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. *Statistics and Probability Letters*, **137**, 135–141.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–477.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, **2**, 849–879.

Lemme, F., van Breukelen, G. J. P., and Berger, M. P. F. (2015). Efficient treatment allocation in two-way nested designs. *Statistical Methods in Medical Research*, **24**, 494–512.

Liu, X., Yue, R.-X., and Wong, W. K. (2019). D-optimal designs for multi-response linear mixed models. *Metrika*, **82**, 87–98.

Pázman, A. (1986). *Foundations of optimum Experimental Design*. D. Reidel Publishing Company, New York.

Prus, M. (2015). *Optimal Designs for the Prediction in Hierarchical Random Coefficient Regression Models*. Ph.D. thesis, Otto-von-Guericke University, Magdeburg.

Prus, M. (2021). Equivalence theorems for multiple-design problems with application in mixed models. *Journal of Statistical Planning and Inference*, **217**, 153–164.

Prus, M. and Piepho, H.-P. (2021). Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing. *Journal of Agricultural, Biological and Environmental Statistics*, **26**, 267–288.

Prus, M. and Schwabe, R. (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B*, **78**, 175–191.

Prus, M., Benda, N., and Schwabe, R. (2020). Optimal design in hierarchical random effect models for individual prediction with application in precision medicine. *Journal of Statistical Theory and Practice*, **14**, 24.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Schmelter, T. (2007). Considerations on group-wise identical designs for linear mixed models. *Journal of Statistical Planning and Inference*, **137**, 4003–4010.

Schwabe, R. (1996). Optimal design of experiments for multivariate response in two-factor linear models. In A. K. Gupta and V. L. Girko, editors, *Multidimensional Statistical Analysis and Theory of Random Matrices. Proceedings of the Sixth Eugene Lukacs Symposium*, pages 235–242. Bowling Green, OH. VSP, Utrecht.

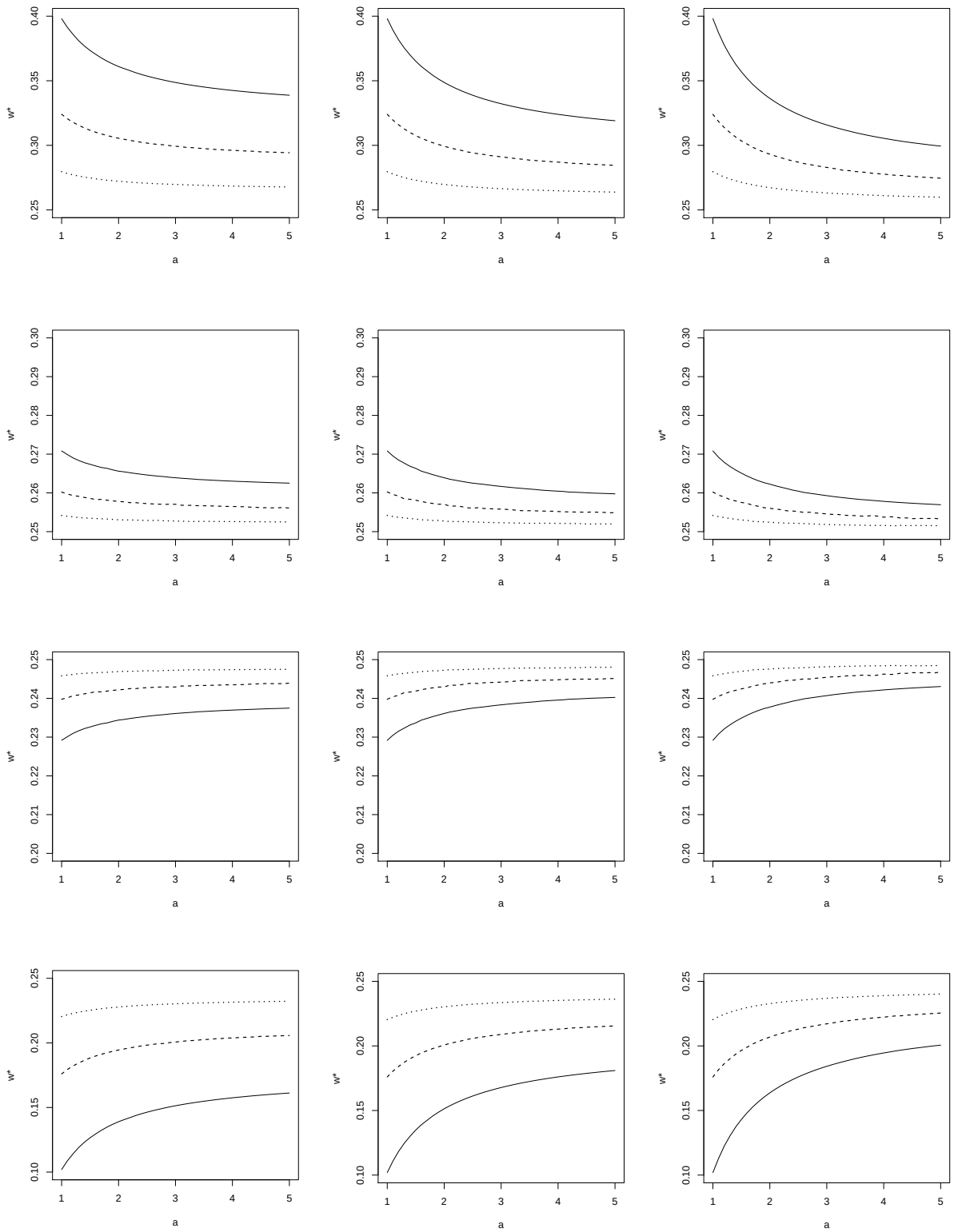

Figure 2: A-optimal weight w^* for prediction of group difference in bi-linear regression model in dependence on ratio a of numbers of observations for $m_2 = 8$ (solid line), $m_2 = 16$ (dashed line), $m_2 = 40$ (dotted line) and $b = 1$ (left panel), $b = 2$ (middle panel), $b = 5$ (right panel), and for values 0.9 (first row), 0.5 (second row), -0.5 (third row) and -0.9 (fourth row) of d_{13}.

Statistical Papers
https://doi.org/10.1007/s00362-023-01416-1

REGULAR ARTICLE

Computational aspects of experimental designs in multiple-group mixed models

Maryna Prus[1,2] · **Lenka Filová**[3]

Received: 12 April 2022 / Revised: 9 December 2022

Abstract
We extend the equivariance and invariance conditions for construction of optimal designs to multiple-group mixed models and, hence, derive the support of optimal designs for first- and second-order models on a symmetric square. Moreover, we provide a tool for computation of D- and L-efficient exact designs in multiple-group mixed models by adapting the algorithm of Harman et al. (Appl Stoch Models Bus Ind, 32:3–17, 2016). We show that this algorithm can be used both for size-constrained problems and also in settings that require multiple resource constraints on the design, such as cost constraints or marginal constraints.

Keywords Optimal design · Exact design · Random coefficient regression · Equivariance · Invariance · Resource constraints

1 Introduction

The aim of this work is computation of highly efficient experimental designs in multiple-group random coefficient regression models. Analytical approach for determining optimal approximate designs for this type of models has been discussed, i.e., in Fedorov and Jones (2005), Schmelter (2007) and Prus (2022). In Fedorov and Jones (2005), optimal designs were obtained for specific regression functions. Schmelter (2007) proposed optimality conditions in the particular case of group-wise identical designs for commonly used linear and determinant criteria. In Prus (2022), equivalence theorems for the general form of multiple-group models have been formulated.

✉ Maryna Prus
maryna.prus@uni-hohenheim.de

1 Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

2 Division of Statistics and Machine Learning, Linköping University, Linköping, Sweden

3 Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovakia

Published online: 13 March 2023

For computing optimal designs in mixed-effect models, several solutions are available. However, most of them focus on computing *approximate* designs for several traditional criteria, and we are not aware of any work that considers additional constraints besides the size of the design.

Namely, Dumont et al. (2018) created an R package for computing designs in mixed-effects model that is predominantly focused on nonlinear models that are used in drug development. However, they only consider the D-optimality criterion without any additional constraints that may arise in the experiment (such as budget, material or other types of resources). The software package (Aliev et al. 2012) is aimed at computing *approximate* designs, mainly for mixed-effects models arising in pharmacokinetic applications. Finally, the software solution of Nyberg et al. (2012) seems to be the most versatile, as it admits user-defined criteria, including their Bayesian versions, but the focus is on approximate designs in nonlinear mixed-effect models and no additional constraints can be included.

In our paper, we propose to use the algorithm of Harman et al. (2016), originally developed for computing D-efficient *exact* designs in the linear regression model with possibly multiple *resource constraints*, for the general form of multiple-group models. We also propose analytical solutions based on equi- and invariance properties of optimal designs for several particular models.

The paper has the following structure: In Sect. 2, we shortly introduce the multiple group model, the design problem and the optimality conditions that are subsequently used in Sect. 3 to show the equivariance and invariance properties of D- and L-optimal designs. In Sect. 4, we show that the problem of computing optimal designs in the multiple-group mixed model can be reformulated as a problem of computing optimal designs with respect to a monotonous criterion function with resource constraints on weights, and, hence, a modification of a recent algorithm for computing resource constrained designs can be used to obtain efficient exact designs in our model. In Sects. 5 and 6, we compute the D- and $IMSE$-optimal designs in bilinear and quadratic models and show that we can easily solve problems with additional constraints that cannot be solved analytically.

2 Multiple-group RCR model

2.1 Model specification

In multiple-group random coefficient regression model the h-th observation of the j-th observational unit in the i-th group is given by

$$\begin{aligned} \mathbf{Y}_{ijh} &= \mathbf{F}_{(i)}(x_{ih})\boldsymbol{\beta}_{ij} + \boldsymbol{\varepsilon}_{ijh}, \quad x_{ih} \in \mathcal{X}_i, \quad i = 1, \ldots, s, \quad j = 1, \ldots, n_i, \\ &\quad h = 1, \ldots, m_i, \end{aligned} \tag{1}$$

where n_i is the number of observational units in group i, m_i is the number of observations per unit in group i, observational settings x_{ih} come from some experimental region $\mathcal{X}_i$. In this work we allow for multivariate (l-variate) response and $\mathbf{F}_{(i)}$ denotes a group-specific $(l \times p)$ matrix of known regression functions in group i. In the particular

case of univariate response we deal with "classical" regression functions: $\mathbf{F}_{(i)} = \mathbf{f}_{(i)}^\top$, and $l = 1$. Unit-specific random parameters $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, \ldots, \beta_{ijp})^\top$ have unknown mean $\boldsymbol{\beta}_0$ and given $(p \times p)$ covariance matrix $\mathbf{D}_i$, $\boldsymbol{\varepsilon}_{ijh}$ denote observational errors with zero mean and non-singular $(l \times l)$ covariance matrix Σ_i. All observational errors and all random parameters are assumed to be uncorrelated.

The covariance matrix of the best linear unbiased estimator for $\boldsymbol{\beta}_0$ is given by

$$\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}_0\right) = \left[\sum_{i=1}^{s} n_i((\tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i)^{-1} + \mathbf{D}_i)^{-1}\right]^{-1}, \tag{2}$$

where $\tilde{\mathbf{F}}_i = (\tilde{\mathbf{F}}_{(i)}^\top(x_{i1}), \ldots, \tilde{\mathbf{F}}_{(i)}^\top(x_{im_i}))^\top$ for $\tilde{\mathbf{F}}_{(i)}(x_{ih}) = \Sigma_i^{-1/2}\mathbf{F}_{(i)}(x_{ih})$, $h = 1, \ldots, m_i$, and the symmetric positive definite matrix $\Sigma_i^{1/2}$ with the property $\Sigma_i = \Sigma_i^{1/2}\Sigma_i^{1/2}$.

2.2 Design criteria

The experimental settings $x_{i1}, \ldots, x_{im_i}$ in formula 1 are not necessarily all distinct. We define an exact design in group i as

$$\xi_i = \begin{pmatrix} x_{i1}, \ldots, x_{ik_i} \\ m_{i1}, \ldots, m_{ik_i} \end{pmatrix}, \tag{3}$$

where $x_{i1}, \ldots, x_{ik_i}$ are the distinct support points in $\mathcal{X}_i$ with the corresponding numbers of observations $m_{i1}, \ldots, m_{ik_i} \in \mathbb{N}$, $\sum_{k=1}^{k_i} m_{ik} = m_i$.

For analytical purposes we also introduce approximate designs:

$$\xi_i = \begin{pmatrix} x_{i1}, \ldots, x_{ik_i} \\ w_{i1}, \ldots, w_{ik_i} \end{pmatrix},$$

where $w_{ik} \geq 0$ denotes the weight of observations at x_{ik}, $k = 1, \ldots k_i$, and $\sum_{k=1}^{k_i} w_{ik} = 1$.

We will use the following notation for the moment (or information) matrix in group i:

$$\mathbf{M}_i(\xi_i) = m_i \sum_{k=1}^{k_i} w_{ik}\, \tilde{\mathbf{F}}_{(i)}(x_{ik})^\top \tilde{\mathbf{F}}_{(i)}(x_{ik}). \tag{4}$$

For exact designs we have $w_{ik} = m_{ik}/m_i$ and

$$\mathbf{M}_i(\xi_i) = \tilde{\mathbf{F}}_i^\top \tilde{\mathbf{F}}_i,$$

which follows from formula (4) and the definition of $\tilde{\mathbf{F}}_i$ below formula (2).

We will also use the notation $\boldsymbol{\xi}$ for the tuple of all group-designs ξ_i: $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_s)$.

Further we extend the definition of the variance–covariance matrix (2) with respect to approximate designs:

$$\mathrm{Cov}_{\xi} = \left[\sum_{i=1}^{s} n_i \left(\mathbf{M}_i(\xi_i)^{-1} + \mathbf{D}_i \right)^{-1} \right]^{-1}. \tag{5}$$

We generally search for the designs which minimize the variance-covariance matrix. Instead of the minimization of the matrix itself (which is in general not possible), we instead minimize suitable functions of this matrix which we call optimality criteria. We focus on the commonly used linear (L-) and determinant (D-) criteria for the estimation of the population parameters $\boldsymbol{\beta}_0$, which are given by

$$\phi_L(\boldsymbol{\xi}) = \mathrm{tr} \left(\left[\sum_{i=1}^{s} n_i \left(\mathbf{M}_i(\xi_i)^{-1} + \mathbf{D}_i \right)^{-1} \right]^{-1} \mathbf{V} \right), \tag{6}$$

where $\mathbf{V}$ is some non-negative definite $(p \times p)$ matrix, and

$$\phi_D(\boldsymbol{\xi}) = -\ln \det \left(\sum_{i=1}^{s} n_i \left(\mathbf{M}_i(\xi_i)^{-1} + \mathbf{D}_i \right)^{-1} \right), \tag{7}$$

respectively (see Prus 2022). (Note that matrix $\mathbf{M}_i(\xi_i)$ here differs from that in Prus Prus 2022 by constant m_i.)

Frequently used particular cases of the L-criterion are the c- and A-criterion, which are of the form (6) with $\mathbf{V} = \mathbf{c}\mathbf{c}^{\top}$, $\mathbf{c} \in \mathbb{R}^p$, and $\mathbf{V} = \mathbb{I}_p$, where $\mathbb{I}_p$ is the $p \times p$ identity matrix, respectively. Another frequently used linear criterion is the $IMSE$-criterion. For the estimation of the mean parameters $\boldsymbol{\beta}_0$ in multiple-group model (1) we define this criterion as follows:

$$\begin{aligned} &\phi_{IMSE}(\boldsymbol{\xi}) \\ &= \sum_{i=1}^{s} a_i \, \mathrm{tr} \left(\int_{\mathcal{X}_i} \mathrm{E} \left[\left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\beta}}_0 - \mathbf{F}_{(i)}(x)\boldsymbol{\beta}_0 \right) \left(\mathbf{F}_{(i)}(x)\hat{\boldsymbol{\beta}}_0 - \mathbf{F}_{(i)}(x)\boldsymbol{\beta}_0 \right)^{\top} \right] \nu_i(\mathrm{d}x) \right), \end{aligned} \tag{8}$$

where ν_i is some suitable measure on the experimental region $\mathcal{X}_i$ (typically uniform on $\mathcal{X}_i$) with $\nu_i(\mathcal{X}_i) = 1$ and a_i is a coefficient related to group i, $\sum_{i=1}^{s} a_i = 1$. The coefficients $a_1, \ldots, a_s$ may depend on the group sizes or, alternatively, equal weight may be given to each group. $IMSE$-criterion (8) may be rewritten in form

$$\phi_{IMSE}(\boldsymbol{\xi}) = \mathrm{tr} \left(\mathrm{Cov} \left(\hat{\boldsymbol{\beta}}_0 \right) \sum_{i=1}^{s} a_i \int_{\mathcal{X}_i} \mathbf{F}_{(i)}(x)^{\top} \mathbf{F}_{(i)}(x) \, \nu_i(\mathrm{d}x) \right). \tag{9}$$

Then we extend it for approximate designs by using the extended variance-covariance matrix (5) and we obtain the particular linear criterion with $\mathbf{V} = \sum_{i=1}^{s} a_i \int_{\mathcal{X}_i} \mathbf{F}_{(i)}(x)^\top \mathbf{F}_{(i)}(x)\, \nu_i(\mathrm{d}x)$, which simplifies to $\mathbf{V} = \int_{\mathcal{X}_1} \mathbf{F}_{(1)}(x)^\top \mathbf{F}_{(1)}(x)\, \nu_1(\mathrm{d}x)$ if the regression matrices $\mathbf{F}_{(i)}$, the experimental regions $\mathcal{X}_i$ and the weighting measures ν_i are the same among all groups: $\mathbf{F}_{(i)} = \mathbf{F}_{(1)}$, $\mathcal{X}_i = \mathcal{X}_1$ and $\nu_i = \nu_1$ for $i = 1, \ldots, s$.

2.3 Optimality conditions

The optimality conditions for the L- and D-criteria are provided by the following theorems (see Prus 2022):

Theorem 1 *Approximate designs $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ are L-optimal for estimation of the mean parameters $\boldsymbol{\beta}_0$ iff*

$$
\begin{aligned}
& m_i \operatorname{tr}\left\{ \tilde{\mathbf{F}}_{(i)}(x_i) \left[\mathbf{M}_i(\xi_i^*)^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \mathbf{V} \right.\right. \\
& \left.\left. \cdot \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \mathbf{M}_i(\xi_i^*)^{-1} \right] \tilde{\mathbf{F}}_{(i)}(x_i)^\top \right\} \\
& \leq \operatorname{tr}\left\{ \mathbf{M}_i(\xi_i^*)^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \mathbf{V} \right. \\
& \left. \cdot \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \right\} \qquad (10)
\end{aligned}
$$

for $x_i \in \mathcal{X}_i$, $i = 1, \ldots, s$.

For support points of ξ_i^* equality holds in (10).

Theorem 2 *Approximate designs $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_s^*)$ are D-optimal for estimation of the mean parameters $\boldsymbol{\beta}_0$ iff*

$$
\begin{aligned}
& m_i \operatorname{tr}\left\{ \tilde{\mathbf{F}}_{(i)}(x_i) \left[\mathbf{M}_i(\xi_i^*)^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \right.\right. \\
& \left.\left. \cdot \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \mathbf{M}_i(\xi_i^*)^{-1} \right] \tilde{\mathbf{F}}_{(i)}(x_i)^\top \right\} \\
& \leq \operatorname{tr}\left\{ \mathbf{M}_i(\xi_i^*)^{-1} \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \left[\sum_{r=1}^{s} n_r \left(\mathbf{M}_r(\xi_r^*)^{-1} + \mathbf{D}_r \right)^{-1} \right]^{-1} \right. \\
& \left. \cdot \left(\mathbf{M}_i(\xi_i^*)^{-1} + \mathbf{D}_i \right)^{-1} \right\} \qquad (11)
\end{aligned}
$$

for $x_i \in \mathcal{X}_i$, $i = 1, \ldots, s$.

For support points of ξ_i^* equality holds in (11).

Example 1 We consider the two-groups model of general form (1) with the regression functions $\mathbf{F}_{(i)}(x) = (1, x), x \in \mathcal{X}_i$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih} + \boldsymbol{\varepsilon}_{ijh}, \quad j = 1, \ldots, n_i, \quad h = 1, \ldots, m_i, \quad i = 1, 2, \tag{12}$$

on the design regions $\mathcal{X}_i = [0, 1]$. The covariance structures of random effects and observational errors are given by $\mathbf{D}_i = \mathrm{diag}(d_{i1}, d_{i2})$ and $\Sigma_i = 1$ for both groups. For this model the left hand sides of the optimality conditions (10) and (11) are parabolas with positive leading terms. Therefore, D- and L-optimal approximate group-designs have the form

$$\xi_i = \begin{pmatrix} 0 & 1 \\ 1 - w_{i1} & w_{i1} \end{pmatrix}, \tag{13}$$

where w_{i1} denotes the weight of observations in point 1 for the i-th group and may depend on the choice of the design criterion as well as on model parameters. The moment matrices are given by

$$\mathbf{M}_i(\xi_i) = \begin{pmatrix} m_i & m_{i1} \\ m_{i1} & m_{i1} \end{pmatrix}, \tag{14}$$

where $m_{i1} = w_{i1} m_i$. Optimal designs for random intercept and random slope have been considered in more detail in Prus (2022).

3 Equi- and invariance considerations for construction of optimal designs

Equi- and invariance of design criteria play an important role for determining optimal designs in fixed effects models (see e.g. Heiligers 1992 or Schwabe 1996, ch. 3). Prus and Schwabe (2016) investigated the related properties of designs, which are optimal for prediction of individual random parameters in single-group mixed effects models. Here we extend those results to multiple-group models.

We consider a one-to-one transformation g of the experimental regions $\mathcal{X}_i$ for all $i = 1, \ldots s$ simultaneously with $g(\mathcal{X}_i) = \mathcal{X}_i^g$. We assume the regression matrices $\mathbf{F}_{(i)}$ to be defined on both $\mathcal{X}_i$ and $\mathcal{X}_i^g$. We also assume the existence of a non-singular $p \times p$ matrix $\mathbf{Q}_g$ such that

$$\tilde{\mathbf{F}}_{(i)}(g(x)) = \mathbf{Q}_g \, \tilde{\mathbf{F}}_{(i)}(x), \quad \forall x \in \mathcal{X}_i, \quad i = 1, \ldots, s, \tag{15}$$

i.e. all $\tilde{\mathbf{F}}_{(i)}$ are linearly equivariant with respect to the transformation g (see e.g. Schwabe 1996, ch. 3). We denote by ξ_i^g the following transformation of an approximate design ξ_i:

$$\xi_i^g = \begin{pmatrix} g(x_{i1}), \ldots, g(x_{ik_i}) \\ w_{i1}, \ldots, w_{ik_i} \end{pmatrix}, \tag{16}$$

where the weight w_{ik} is the same for both ξ_i and ξ_i^g and only the design points x_{ik} are transformed. Then we obtain the next property of the moment matrices:

$$\mathbf{M}_i(\xi_i^g) = \mathbf{Q}_g \, \mathbf{M}_i(\xi_i) \, \mathbf{Q}_g^\top, \quad i = 1, \ldots, s. \tag{17}$$

Further we use the notations $\mathcal{D} = (\mathbf{D}_1, \ldots, \mathbf{D}_s)$ and $\mathbb{X} = \times_{i=1}^s \mathcal{X}_i$ for the tuple of covariance matrices and the Cartesian product of the experimental regions, respectively, in all groups. For the covariance matrix (5) the following relation can be easily verified:

$$\mathrm{Cov}_{\xi^g}(\mathcal{D}^g) = \mathbf{Q}_g^{-\top} \mathrm{Cov}_\xi(\mathcal{D}) \mathbf{Q}_g^{-1}, \tag{18}$$

where $\boldsymbol{\xi}^g = (\xi_1^g, \ldots, \xi_s^g)$, $\mathcal{D}^g = (\mathbf{D}_1^g, \ldots, \mathbf{D}_s^g)$, $\mathbf{D}_i^g = \mathbf{Q}_g^{-\top} \mathbf{D}_i \mathbf{Q}_g^{-1}$ and $\mathbf{Q}_g^{-\top} = (\mathbf{Q}_g^\top)^{-1}$. We use the notation $\mathrm{Cov}_\xi(\mathcal{D})$ [instead of Cov_ξ as in formula (5)] to emphasize the dependence on the covariance matrices $\mathcal{D}$ of random effects.

Then the equivariance of the D- and L-criteria with respect to a transformation g can be established.

Theorem 3 *If the approximate designs $\boldsymbol{\xi}^*$ are D-optimal for the estimation of $\boldsymbol{\beta}_0$ on the experimental regions $\mathbb{X}$ under the dispersion matrices $\mathcal{D}$, then the induced approximate designs $\boldsymbol{\xi}^g$ are D-optimal for the estimation of $\boldsymbol{\beta}_0$ on the experimental regions $\mathbb{X}^g = \times_{i=1}^s \mathcal{X}_i^g$ under the induced dispersion matrices $\mathcal{D}^g$.*

Proof From the definition of the D-criterion for the estimation of $\boldsymbol{\beta}_0$ and formula (18) we obtain

$$\phi_D(\boldsymbol{\xi}^g, \mathcal{D}^g) = -2 \ln |\det(\mathbf{Q}_g)| + \phi_D(\boldsymbol{\xi}, \mathcal{D}),$$

which proves the optimality of $\boldsymbol{\xi}^g$ on $\mathbb{X}^g$ for $\boldsymbol{\xi}$ optimal on $\mathbb{X}$. □

Theorem 4 *If the approximate designs $\boldsymbol{\xi}^*$ are L-optimal for the estimation of $\boldsymbol{\beta}_0$ on the experimental regions $\mathbb{X}$ under the dispersion matrices $\mathcal{D}$ with respect to the transformation matrix $\mathbf{V}$, then the induced approximate designs $\boldsymbol{\xi}^{*g}$ are L-optimal for the estimation of $\boldsymbol{\beta}_0$ on the experimental regions $\mathbb{X}^g$ under the induced dispersion matrices $\mathcal{D}^g$ with respect to the induced transformation matrix $\mathbf{V}_g = \mathbf{Q}_g \mathbf{V} \mathbf{Q}_g^\top$.*

Proof Using formulas (6) and (18) it can be easily verified that

$$\phi_L(\boldsymbol{\xi}^g, \mathcal{D}^g, \mathbf{V}_g) = \phi_L(\boldsymbol{\xi}, \mathcal{D}, \mathbf{V}),$$

which proves the optimality of $\boldsymbol{\xi}^g$ on $\mathbb{X}^g$. □

Corollary 1 *The A-criterion for the estimation of $\boldsymbol{\beta}_0$ is equivariant with respect to a transformation g if $\mathbf{Q}_g$ is orthogonal, i.e.:*

$$\mathbf{Q}_g \mathbf{Q}_g^\top = \mathbf{Q}_g^\top \mathbf{Q}_g = \mathbb{I}_p. \tag{19}$$

To verify the equivariance of the $IMSE$-criterion we assume, besides the transformed regression matrices $\tilde{\mathbf{F}}_{(i)}$, the original regression matrices $\mathbf{F}_{(i)}$ to be linearly equivariant with respect to the transformation g:

$$\mathbf{F}_{(i)}(g(x)) = \mathbf{Q}_g \mathbf{F}_{(i)}(x), \quad \forall x \in \mathcal{X}_i, \quad i = 1, \ldots, s. \tag{20}$$

Then if the measure ν_i is transformed to its image ν_i^g, we obtain

$$\mathbf{V}_g = \sum_{i=1}^{s} a_i \int_{\mathcal{X}_i^g} \mathbf{F}_{(i)}(x)^\top \mathbf{F}_{(i)}(x)\, \nu_i^g(\mathrm{d}x) = \mathbf{Q}_g \mathbf{V} \mathbf{Q}_g^\top.$$

Corollary 2 *The $IMSE$-criterion for the estimation of $\boldsymbol{\beta}_0$ is equivariant with respect to a transformation g if condition (20) is satisfied.*

Example 1 (continued). *We consider again the two-groups linear regression model (12) on $\mathcal{X}_i = [0, 1]$ with diagonal covariance structure of random effects. For the $IMSE$-criterion we chose the uniform weighting $\nu_i = \lambda_{[0,1]}$, $i = 1, 2$, where $\lambda_{[c_1,c_2]}$ denotes the Lebesgue measure on $[c_1, c_2]$.* Let ξ_i^* be *D-, A- or $IMSE$-optimal group-designs of form (13) with the optimal weight of observations w_{i1}^* (which generally depends on the choice of the design criterion).*

Now we consider the linear transformation $g(x) = ax, a > 0$, for which we obtain $\mathbf{Q}_g = \mathrm{diag}(1, a)$. Then the D-, A- or $IMSE$-optimal group-designs in model (12) on $\mathcal{X}_i^g = [0, a]$ for $\mathbf{D}_i^g = \mathrm{diag}(d_{i1}, d_{i2}/a^2)$ and $\nu_i^g = \frac{1}{a}\lambda_{[0,a]}$ are given by

$$\xi_i^{*g} = \begin{pmatrix} 0 & a \\ 1 - w_{i1}^* & w_{i1}^* \end{pmatrix}. \tag{21}$$

Same behavior of optimal designs has been established for the prediction of random effects in single-group model in Prus and Schwabe (2016).

Further we consider a finite group G of transformations $g : \mathcal{X}_i \to \mathcal{X}_i$ of the experimental regions $\mathcal{X}_i$ onto themselves for all $i = 1, \ldots s$ simultaneously. We assume the equivariance condition (15) to be satisfied and the dispersion matrices to be invariant: $\mathbf{D}_i^g = \mathbf{D}_i$, for all $g \in G, i = 1, \ldots, s$. For the linear criteria we additionally assume the invariance of the transformation matrices: $\mathbf{V}_g = \mathbf{V}$. Then the D- and L-criteria are invariant with respect to all $g \in G$ and the following statement can be formulated:

Theorem 5 *If the approximate designs $\boldsymbol{\xi}^*$ are D- or L-optimal for the estimation of $\boldsymbol{\beta}_0$, then the symmetrized designs $\bar{\boldsymbol{\xi}}^* = (\bar{\xi}_1^*, \ldots, \bar{\xi}_s^*)$ for $\bar{\xi}_i^* = \frac{1}{\#G} \sum_{g \in G} \xi_i^{*g}$ are also D- or L-optimal for the estimation of $\boldsymbol{\beta}_0$.*

Proof Let the designs $\boldsymbol{\xi}^*$ be D-optimal for the estimation of $\boldsymbol{\beta}_0$. Then it follows from Theorem 3 and the invariance of the dispersion matrices that the induced designs $\boldsymbol{\xi}^{*g}$ are also D-optimal, i.e.

$$\phi_D(\boldsymbol{\xi}^{g*}, \mathcal{D}) = \phi_D(\boldsymbol{\xi}^*, \mathcal{D}), \quad \forall g \in G.$$

From the convexity of the criterion we obtain

$$\phi_D(\bar{\boldsymbol{\xi}}^*, \mathcal{D}) \leq \phi_D(\boldsymbol{\xi}^*, \mathcal{D}),$$

which implies the D-optimality of the designs $\bar{\boldsymbol{\xi}}^*$.

For the linear criterion the proof is similar. □

The invariance of the A-criterion is straightforward if condition (19) is satisfied for all $g \in G$.

Corollary 3 *If the approximate designs $\boldsymbol{\xi}^*$ are A-optimal for the estimation of $\boldsymbol{\beta}_0$ and condition (19) is satisfied for all $g \in G$, then the symmetrized designs $\bar{\boldsymbol{\xi}}^*$ are also A-optimal for the estimation of $\boldsymbol{\beta}_0$.*

For the $IMSE$-criterion we require the invariance of the weighting measures: $\nu_i^g = \nu_i$, which leads to $\mathbf{V}_g = \mathbf{V}$.

Corollary 4 *If the approximate designs $\boldsymbol{\xi}^*$ are $IMSE$-optimal for the estimation of $\boldsymbol{\beta}_0$ and condition (20) is satisfied for all $g \in G$, then the symmetrized designs $\bar{\boldsymbol{\xi}}^*$ are also $IMSE$-optimal for the estimation of $\boldsymbol{\beta}_0$.*

Example 2 We consider the multiple-group model of the form (1) with the regression functions $\mathbf{F}_{(i)}(x) = (1, x, x^2)$ on a symmetric design region $\mathcal{X}_i = [-a, a]$, $a > 0$, $i = 1, \ldots, s$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih} + \boldsymbol{\beta}_{ij3} x_{ih}^2 + \boldsymbol{\varepsilon}_{ijh}, \quad j = 1, \ldots, n_i, \quad h = 1, \ldots, m_i. \tag{22}$$

For this model the left hand sides of the optimality conditions (10) and (11) for the L- and D-criterion, respectively, are polynomial functions of degree four. Consequently, the corresponding optimal group-designs ξ_i are supported by not more than three design points including the two endpoints of the experimental region:

$$\xi_i^* = \begin{pmatrix} -a & o_i & a \\ w_{i1}^* & 1 - w_{i1}^* - w_{i2}^* & w_{i2}^* \end{pmatrix}, \tag{23}$$

where $o_i \in (-a, a)$ may differ for different design criteria or for different groups. We assume covariance structures of random effects and observational errors to be given by

$$\mathbf{D}_i = \begin{pmatrix} d_{i11} & 0 & d_{i13} \\ 0 & d_{i22} & 0 \\ d_{i13} & 0 & d_{i33} \end{pmatrix} \tag{24}$$

and $\Sigma_i = 1$ for all groups. For the $IMSE$-criterion we chose the uniform weighting measure $\nu_i = \frac{1}{2a}\lambda_{[-a,a]}$ for all $i = 1, \ldots, s$.

Further we consider the group of transformations $G = \{g_1, g_2\}$ with $g_1(x) = -x$ and $g_2(x) = x$. Then we obtain $\mathbf{Q}_{g_1} = \mathrm{diag}(1, -1, 1)$ and $\mathbf{Q}_{g_2}$ is equal to the identity matrix. Hence, the dispersion matrices $\mathbf{D}_i$ and the measures ν_i are invariant and conditions (19) and (20) are satisfied for both g_1 and g_2. Then by Theorem 5 and Corollary 3 group-designs of the general form

$$\bar{\xi}_i^* = \begin{pmatrix} -a & 0 & a \\ w_{i1}^* & 1 - 2w_{i1}^* & w_{i1}^* \end{pmatrix} \tag{25}$$

are D-, A- and $IMSE$-optimal for the estimation of the mean parameters $\boldsymbol{\beta}_0$. The optimal weights of observations w_{i1}^* at points $x = a$ and $x = -a$ generally depend on the design criterion, the variance parameters, the group sizes, the numbers of observations and the length of the interval (see Sect. 6 for examples of the designs).

Further we consider some examples of multiple polynomial regression. For models without random effects optimal designs for multiple polynomial regression have been discussed, e.g., in Galil and Kiefer (1977) and Heiligers (1992).

Example 3 We consider the multiple-group bi-linear model with the regression functions $\mathbf{F}_{(i)}(x) = (1, x_1, x_2)$ on a design region $\mathcal{X}_i = [-a, a]^2$, $a > 0$, $i = 1, \ldots, s$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih1} + \boldsymbol{\beta}_{ij3} x_{ih2} + \boldsymbol{\varepsilon}_{ijh}, \quad j = 1, \ldots, n_i, \quad h = 1, \ldots, m_i. \tag{26}$$

For this model the left hand sides of the optimality conditions (10) and (11) for the L- and D-criterion, respectively, are convex paraboloids. Therefore, the only admissible support points for optimal designs are $x_{i1} = (a, a)$, $x_{i2} = (a, -a)$, $x_{i3} = (-a, a)$, $x_{i4} = (-a, -a)$:

$$\xi_i^* = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} & x_{i4} \\ w_{i1}^* & w_{i2}^* & w_{i3}^* & w_{i4}^* \end{pmatrix}, \tag{27}$$

where $\sum_{k=1}^4 w_{ik} = 1$. For the $IMSE$-criterion we use the product measure $\nu_i = \frac{1}{2a}\lambda_{[-a,a]} \times \frac{1}{2a}\lambda_{[-a,a]}$ for all $i = 1, \ldots, s$.

Further we assume the same covariance structures of random effects and observational errors as in Example 2 of quadratic regression and we consider the group of transformations $G = \{g_1, g_2\}$ with $g_1(x) = (-x_1, x_2)^\top$ and $g_2(x) = (x_1, x_2)^\top$. We obtain $\mathbf{Q}_{g_1} = \mathrm{diag}(1, -1, 1)$ and $\mathbf{Q}_{g_2}$ is equal to the 3×3 identity matrix. Then the dispersion matrices $\mathbf{D}_i$ and the weighting measures ν_i are invariant and conditions (19) and (20) are satisfied for both g_1 and g_2 and, consequently, group-designs of the general form

$$\bar{\xi}_i^* = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} & x_{i4} \\ w_{i1}^* & w_{i2}^* & w_{i1}^* & w_{i2}^* \end{pmatrix} \tag{28}$$

with $w_{i2}^* = \frac{1}{2}(1 - 2w_{i1}^*)$ are D-, A- and $IMSE$-optimal for the estimation of the mean parameters $\boldsymbol{\beta}_0$. Only the optimal weights of observations w_{i1}^* have to be determined.

Note that besides the choice of the design criterion these numbers may also depend on the model parameters (see Sect. 5 for illustrative examples).

For the particular case with diagonal covariance structure of random effects: $d_{i13} = 0$, we consider the group of transformations $G = \{g_1, g_2, g_3, g_4\}$ with $g_3(x) = (x_1, -x_2)^\top$ and $g_4(x) = (-x_1, -x_2)^\top$, for which we obtain $\mathbf{Q}_{g_3} = \text{diag}(1, 1, -1)$ and $\mathbf{Q}_{g_4} = \text{diag}(1, -1, -1)$. The dispersion matrices $\mathbf{D}_i$ and the measures ν_i are invariant and conditions (19) and (20) are satisfied for all transformation in G. Then the balanced group-designs

$$\bar{\xi}_i^* = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} & x_{i4} \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \tag{29}$$

are D-, A- and $IMSE$-optimal.

Example 4 We consider the multiple-group bi-quadratic model with the regression functions $\mathbf{F}_{(i)}(x) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$ on a design region $\mathcal{X}_i = [-a, a]^2, a > 0$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2}x_{ih1} + \boldsymbol{\beta}_{ij3}x_{ih2} + \boldsymbol{\beta}_{ij4}x_{ih1}x_{ih2} + \boldsymbol{\beta}_{ij5}x_{ih1}^2 + \boldsymbol{\beta}_{ij6}x_{ih2}^2 + \boldsymbol{\varepsilon}_{ijh} \tag{30}$$

for $j = 1, \ldots, n_i$, $h = 1, \ldots, m_i$ and $i = 1, \ldots, s$. For this model the left hand sides of the optimality conditions (10) and (11) are quadric surfaces in (x_1, x_2), for which the projections on both $x_1 = 0$ and $x_2 = 0$ are polynomials of degree four. Then the only admissible support points for L- and D-optimal designs are $x_{i1} = (a, a)$, $x_{i2} = (a, -a), x_{i3} = (-a, a), x_{i4} = (-a, -a), x_{i5} = (o_{i1}, a), x_{i6} = (o_{i2}, -a), x_{i7} = (a, o_{i3}), x_{i8} = (-a, o_{i4})$ and $x_{i9} = (o_{i5}, o_{i6})$, where $o_{il} \in (-a, a)$, $l = 1, \ldots, 6$. For the $IMSE$-criterion we use the same weighting measures as in Example 3.

Further we assume the following simple covariance structure of the random effects and the observational errors: $\mathbf{D}_i = \text{diag}(d_{i1}, \ldots d_{i6})$ and $\Sigma_i = 1, i = 1, \ldots s$. Then we consider the same group of transformations G as in the previous example and we obtain $\mathbf{Q}_{g_1} = \text{diag}(1, -1, 1, -1, 1, 1)$, $\mathbf{Q}_{g_2}$ is equal to the 6×6 identity matrix, $\mathbf{Q}_{g_3} = \text{diag}(1, 1, -1, -1, 1, 1)$ and $\mathbf{Q}_{g_4} = \text{diag}(1, -1, -1, 1, 1, 1)$. Then conditions (19) and (20) are satisfied and dispersion matrices $\mathbf{D}_i$ and the measures ν_i are invariant for all $g_i \in G$. Therefore, D-, A- and $IMSE$-optimal designs have the general form

$$\bar{\xi}_i^* = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} & x_{i4} & x_{i5} & x_{i6} & x_{i7} & x_{i8} & x_{i9} \\ w_{i1}^* & w_{i1}^* & w_{i1}^* & w_{i1}^* & w_{i2}^* & w_{i2}^* & w_{i3}^* & w_{i3}^* & w_{i4}^* \end{pmatrix}, \tag{31}$$

where $w_{i4}^* = 1 - 4w_{i1}^* - 2w_{i2}^* - 2w_{i3}^*$ and all $o_{il} = 0$, i. e. $x_{i5} = (0, a)$, $x_{i6} = (0, -a)$, $x_{i7} = (a, 0)$, $x_{i8} = (-a, 0)$ and $x_{i9} = (0, 0)$. The weights of observations w_{i2}^*, w_{i3}^* and w_{i4}^* depend on the choice of the design criterion and on the model parameters and have to be optimized.

Then we additionally assume the conditions $d_{i2} = d_{i3}$ and $d_{i5} = d_{i6}$ to be satisfied and consider the extended group of transformations $G_1 = G \cup \{g_5\}$ with $g_5(x) = (x_2, x_1)^\top$, for which we obtain $\mathbf{Q}_{g_5} = \text{block-diag}(1, \mathbf{P}, 1, \mathbf{P})$, where $\mathbf{P}$ is the (2×2)

permutation matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The dispersion matrices $\mathbf{D}_i$ and the measures ν_i are also invariant with respect to g_5 and conditions (19) and (20) are satisfied. Then the general form (31) of optimal designs simplifies to

$$\xi_i^* = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} & x_{i4} & x_{i5} & x_{i6} & x_{i7} & x_{i8} & x_{i9} \\ w_{i1}^* & w_{i1}^* & w_{i1}^* & w_{i1}^* & w_{i2}^* & w_{i2}^* & w_{i2}^* & w_{i2}^* & 1 - 4(w_{i1}^* + w_{i2}^*) \end{pmatrix}. \tag{32}$$

Note that similar behavior has been established for optimal designs for Kiefer's Φ_p-criteria in fixed effects models (see Galil and Kiefer 1977). However, designs obtained in that work depend on the choice of the design criterion only. In the model under investigation optimal designs may also depend on the variance parameters, the group sizes and the numbers of observations per observational unit.

4 Computing the multiple-group mixed models designs

In this Section, we will show how to compute efficient exact designs for model (1). To this end, let's discretize each (possibly continuous) experimental region $\mathcal{X}_i$, $i = 1, \ldots, s$, into k_i points $x_{i1}, \ldots, x_{ik_i}$ and denote the corresponding numbers of measurements in these points by $m_{i1}, \ldots, m_{ik_i} \in \mathbb{N}_0$, as is customary in optimal design algorithms. Similarly to the notation adopted in Sect. 2, we define the k_i-dimensional vectors $m_i = (m_{i1}, \ldots, m_{ik_i})$ and the $\sum_{i=1}^{s} k_i = u$-dimensional vector $\mathbf{m} = (m_1, \ldots, m_s)$.

Now, consider the optimization problem presented in Harman et al. (2016):

$$\begin{aligned} &\min_{\mathbf{m}} \Phi(\mathbf{m}) \\ &\text{subject to } \mathbf{Am} \leq \mathbf{b}. \end{aligned}$$

Here, we minimize the function Φ on the set of permissible designs determined by the linear inequality $\mathbf{Am} \leq \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{k \times u}$ and $\mathbf{b} \in \mathbb{R}^k$ are such that the elements of $\mathbf{A}$ are nonnegative and the elements of $\mathbf{b}$ are positive. These kinds of constraints are called *resource constraints*, i.e., we can view each measurement as consuming some amount of each of the k resources, limit on which are given by the vector $\mathbf{b}$.

The method described in Harman et al. (2016) is related to the Detmax procedure, employing a tabu search principle. The algorithm is based on excursions in the set of all feasible designs. More precisely, from a design ξ we can either make a forward step to one of its upper neighbours or a backward step to one of its lower neighbours. These excursions are directed by the attribute of each design (which can be, e.g., its criterion value), a tabu list of the attributes of already visited designs and a local heuristic evaluation of the design that roughly estimates how promising a design is as a part of an excursion leading to an efficient design.

Note that although the algorithm is primarily developed for D-optimality in the standard linear regression model, it can be easily adapted for different criteria that are monotonous on the set of all approximate designs, which enables us to compute D- and L-efficient exact designs in model (1).

To show this, we rewrite the covariance matrix (5) in the following form:

$$\mathrm{Cov}_\xi = \left[\left(\mathbb{1}_s^\top \otimes \mathbb{I}_p \right) \left(\mathbf{M}_\xi^{-1} + \mathbf{D} \right)^{-1} \left(\mathbb{1}_s \otimes \mathbb{I}_p \right) \right]^{-1}, \tag{33}$$

where $\mathbf{M}_\xi = \mathrm{diag}\left(\tilde{\mathbf{M}}_1(\xi_1), \ldots, \tilde{\mathbf{M}}_s(\xi_s) \right)$ is the block-diagonal matrix with the blocks $\tilde{\mathbf{M}}_i(\xi_i) = n_i \, \mathbf{M}_i(\xi_i)$ and $\mathbf{D} = \mathrm{diag}(\tilde{\mathbf{D}}_1, \ldots, \tilde{\mathbf{D}}_s)$ is the block-diagonal matrix with the blocks $\tilde{\mathbf{D}}_i = \frac{1}{n_i} \mathbf{D}_i$, $\mathbb{1}_s$ is the vector of length s with all entries equal to 1 and "$\otimes$" denotes the Kronecker product.

Then the L- and D-criteria defined by (6) and (7) can be written as the function of the design vector w in the following way:

$$\phi_L(\boldsymbol{\xi}) = \mathrm{tr} \left(\left[\left(\mathbb{1}_s^\top \otimes \mathbb{I}_p \right) \left(\mathbf{M}_\xi^{-1} + \mathbf{D} \right)^{-1} \left(\mathbb{1}_s \otimes \mathbb{I}_p \right) \right]^{-1} \mathbf{V} \right) \tag{34}$$

and

$$\phi_D(\boldsymbol{\xi}) = -\ln \det \left[\left(\mathbb{1}_s^\top \otimes \mathbb{I}_p \right) \left(\mathbf{M}_\xi^{-1} + \mathbf{D} \right)^{-1} \left(\mathbb{1}_s \otimes \mathbb{I}_p \right) \right]. \tag{35}$$

Note that both criteria (34) and (35) are monotonically decreasing with respect to $\mathbf{M}_\xi$.

Further, model (1) can be viewed as a one-group model on $\mathcal{X} = \times_{i=1}^s \mathcal{X}_i$ with marginal constraints (see, e.g., Cook and Thibodeau 1980) that constrict the number of observations in each group to $\sum_{h=1}^{k_i} m_{ih} = m_i$. This can be formulated in the form of resource constraints by putting $\mathbf{A} = diag(\mathbb{1}_{k_1}^\top, \ldots, \mathbb{1}_{k_s}^\top)$ and $\mathbf{b} = (m_1, \ldots, m_s)^\top$.

Hence, the optimization problem to solve is

$$\begin{aligned} &\min_{\mathbf{m}} \phi(\mathbf{m}) \\ &\text{subject to } \mathrm{diag}(\mathbb{1}_{k_1}^\top, \ldots, \mathbb{1}_{k_s}^\top)\mathbf{m} \leq \mathbf{b}, \end{aligned} \tag{36}$$

where by ϕ we denote either of the optimality criteria in (34) or (35).

Note that the algorithm used here is heuristic, i.e., it does not guarantee that the resulting design is optimal, although it is demonstrated in Harman et al. (2016) that it is usually highly efficient. Therefore, in the following sections, we will call the designs obtained by the algorithm as *efficient exact designs*.

Further, we will demonstrate that it is of great practical use that the matrix $\mathbf{A}$ and the vector b can be modified so that they incorporate additional linear resource constraints on the weights, such as the limit on the number of measurements in particular points or cost constraints (see Sect. 6 for an example of such constraints), simply by adding suitable rows to the matrix $\mathbf{A}$ and elements to the vector b.

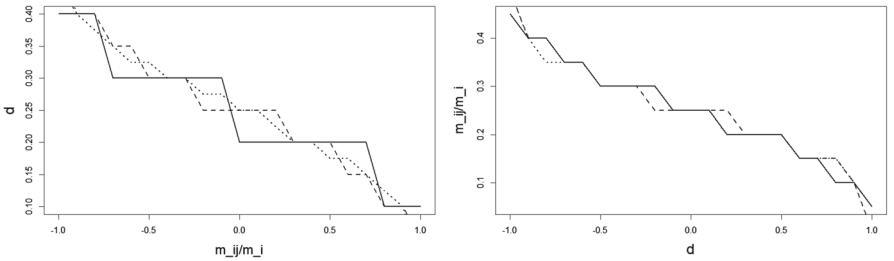

Fig. 1 The dependence of the number of observations in the point (1, 1) on the parameter d in the exact D-efficient designs in bilinear model (26) on $[-1, 1]^2$ with total numbers of observations in the groups given by $m = (10, 20, 40)$ (left) and $m = (20, 20, 20)$ (right). The three lines denote the number of observations in the point (1, 1) normalized by m_i for the first (full line), second (dashed line) and third (dotted line) group

5 Bi-linear regression

Let's consider the model of bi-linear regression (26) with three groups, $\mathcal{X}_i = [-1, 1]^2$, $\Sigma_i = 1$, $n = (1, 1, 1)$ and

$$D_i = \begin{pmatrix} 1 & 0 & d \\ 0 & 1 & 0 \\ d & 0 & 1 \end{pmatrix}, \ i = 1, 2, 3.$$

As the analytical results in Example 3 show, the approximate optimal designs are supported on the four vertices of the square $[-1, 1]^2$ and the number of observations is identical in the points $(1, 1)$, $(-1, 1)$ and in the points $(-1, -1)$, $(1, -1)$. This phenomenon was confirmed also for the exact D-efficient designs by our algorithm.

In this example, we will numerically illustrate the dependence of efficient designs on the parameter d in the matrices D_i, $i = 1, 2, 3$. To this end, let's consider that in all three groups, the parameter d is the same. Figure 1 shows how the numbers of observation in the point (1, 1) change with d varying from -1 to 1 for two different settings: $m = (10, 20, 40)$ (left) and $m = (20, 20, 20)$ (right). We can see that in both cases, the number of observations in (1, 1) decreases with increasing d.

Now, suppose that

$$D_i = \begin{pmatrix} 1 & 0 & d_i \\ 0 & 1 & 0 \\ d_i & 0 & 1 \end{pmatrix}, \ i = 1, 2, 3, \tag{37}$$

where $d_i \in \{-0.5, 0, 0.5\}$ are not necessarily the same between groups. In Table 1 we show the behavior of the numbers of observation for several selected d_1, d_2, d_3 in the case $m = (20, 20, 20)$.

Table 1 Exact D-efficient designs in bilinear model (26) on $[-1, 1]^2$ with total numbers of observations in the groups given by $m = (20, 20, 20)$ with $d_i \in \{-0.5, 0, 0.5\}$

(d_1, d_2, d_3)	$(-1, -1)$	$(1, 1)$	$(-1, -1)$	$(1, 1)$	$(-1, -1)$	$(1, 1)$
$(0, 0, 0)$	5	5	5	5	5	5
$(0, 0, 0.5)$	4	6	4	6	7	3
$(0, 0.5, 0.5)$	4	6	7	3	7	3
$(0.5, 0.5, 0.5)$	6	4	6	4	6	4
$(0, 0, -0.5)$	6	4	6	4	3	7
$(0, 0.5, -0.5)$	5	5	7	3	3	7
$(-0.5, 0.5, -0.5)$	3	7	8	2	3	7

6 Quadratic regression on a symmetric interval

Consider the two-groups model of the form (1) with the regression functions $\mathbf{F}_{(i)}(x) = (1, x, x^2)^\top$, $x \in \mathcal{X}_i$, and the design region $\mathcal{X}_i = [-1, 1]$, $i = 1, 2$:

$$Y_{ijh} = \boldsymbol{\beta}_{ij1} + \boldsymbol{\beta}_{ij2} x_{ih} + \boldsymbol{\beta}_{ij3} x_{ih}^2 + \boldsymbol{\varepsilon}_{ijh}, \quad j = 1, \ldots, n_i, \quad h = 1, \ldots, m_i. \tag{38}$$

The covariance structures of random effects and observational errors are given by $\mathbf{D}_i = \mathrm{diag}(d_{i1}, d_{i2}, d_{i3})$ and $\Sigma_i = 1$ for both groups.

The corresponding approximate optimal group-designs ξ_i are supported by three design points -1, 0, 1 (see Sect. 3):

$$\xi_i^* = \begin{pmatrix} -1 & 0 & 1 \\ w_{i1}^* & 1 - 2w_{i1}^* & w_{i1}^* \end{pmatrix}. \tag{39}$$

This result was heuristically confirmed to hold also for the exact designs: we discretized the design region into q points $-1 = x_1 < x_2 < \cdots < x_q = 1$ and confirmed that for all cases considered below, the support points are indeed -1, 0 and 1.

The exact D- and $IMSE$-efficient designs for this case and several particular $m = (m_1, m_2)$ are given in Table 2 in the Appendix.

We can see that for the criterion of D-optimality and the values of the diagonal of the matrix D_i equal to either $(1, 1, 1)$ or $(1, 1, 0)$, half of the measurements is in the point 0 and the remaining half is distributed equally among the points -1 and 1. For (d_1, d_2, d_3) equal either to $(0, 1, 1)$ or $(1, 0, 0)$, the measurements are heavily concentrated in the support point 0, but the designs are still nonsingular. The case $(1, 0, 1)$ shows opposite phenomenon with the measurements being concentrated in the points -1 and 1. For the remaining cases, the pattern is not so clear and the weights depend more on m, sometimes even resulting in singular designs.

For the $IMSE$ criterion, we get results identical to D-optimality if the diagonal of D_i is $(1, 1, 1)$. For the rest of the cases, the situation is more varied and we refer reader to Table 2 for details.

From practical point of view, it may not be desirable to only have three support points for each group. Therefore, additional constraints on the design were suggested,

where it is prescribed that, for each group, maximum one half of the measurements can be taken at $-1, 0$ or 1. Formally, these constraints can be written in the form $\mathbf{A}^{(1)} w \leq b^{(1)}$ (see Sect. 4 for details), where

$$\mathbf{A}^{(1)} = \begin{pmatrix} c_q & 0_q^\top \\ 0_q^\top & c_q \end{pmatrix} \in \mathbb{R}^{2 \times 2q}, \; b^{(1)} = \begin{pmatrix} m_1/2 \\ m_2/2 \end{pmatrix}, \tag{40}$$

where $c_q = (1, 0, \ldots, 0, 1, 0, \ldots, 0, 1) \in \mathbb{R}^q$ with 1 on the positions corresponding to the points $-1, 0, 1$. The D- and $IMSE$-efficient designs for the discretization $\mathcal{X}_i = \{-1, -0.8, \ldots, 0.8, 1\}$ of the interval $[-1, 1]$ with the step 0.2 (i.e. $q = 11$) are given in Tables 3 and 4. Note that for both criteria, the tendency is to distribute the measurements as close as possible to the original support points $-1, 0$ and 1.

Another type of constraint that is often used in practical situations, is the cost constraint: this is natural, for example, in clinical trials, where taking a measurement at a point x consumes a certain number of time, personal or material resources and the total cost of the experiment is limited. In our case, let the measurement at the point x cost $|x| + 0.1$ units, and, for group j, let the maximum admissible cost be $m_j/4$. This leads to adding the constraints $\mathbf{A}^{(2)} w \leq b^{(2)}$ with the following $\mathbf{A}^{(2)}, b^{(2)}$ to the problem (36):

$$\mathbf{A}^{(2)} = \begin{pmatrix} u_q + 0.1 \mathbb{1}_q^\top & 0 \\ 0 & u_q + 0.1 \mathbb{1}_q^\top \end{pmatrix} \in \mathbb{R}^{2 \times 2q}, \; b^{(2)} = \frac{1}{4} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \tag{41}$$

where $u_q = (1, 0.8, \ldots, 0.8, 1)$.

Again, we computed D- and $IMSE$-efficient designs with respect to this constraint for the discretization $\mathcal{X}_i = \{-1, -0.8, \ldots, 0.8, 1\}$. Now, the designs are supported on $-1, 0$ and 1, but, compared to the unconstrained designs, much more measurements are made at the point 0, which is 'cheap': the results are summarized in Table 5.

Finally, it is also feasible and possible to consider both types of constraints together, resulting in $\mathbf{A}^{(3)} w \leq b^{(3)}$ with

$$\mathbf{A}^{(3)} = \begin{pmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \end{pmatrix} \in \mathbb{R}^{4 \times 4q}, \; b^{(3)} = \begin{pmatrix} b^{(1)} \\ b^{(2)} \end{pmatrix}. \tag{42}$$

The resulting D- and $IMSE$-efficient designs for this constraint are given in Tables 6 and 7.

Note that in some cases, the additional constraints on the designs were saturated for a number of measurements that is lower than the maximum attainable number of measurements given by $(m_1, m_2)^\top$. This is demonstrated in a more detailed way in Fig. 2, where we again consider the D-efficient design with $(m_1, m_2) = (20, 40)$ and the cost constraints (41), but now the cost $b^{(2)}$ can vary between 0 and m_i for the i-th group. All the designs are supported in the points $-1, 0$ and 1 and the figure shows that when the maximum allowed cost is too low, the total maximum number of measurements is (sometimes significantly) lower than the corresponding m_i.

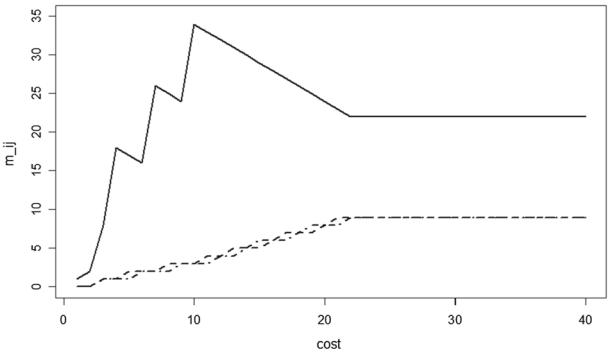

Fig. 2 The numbers of measurements in the point 0 (full line), −1 (dashed line) and 1 (dot-dashed line) for the second group in the D-efficient design in model (38) with $(m_1, m_2) = (20, 40)$ and constraints of the type (41) with the maximum cost $b^{(2)}$ in the second group varying between 0 and 40

7 Discussion

In the paper, we have considered equi- and invariance properties of approximate optimal designs in multiple-group mixed models. We have used these properties to fix the support points and, consequently, to reduce the number of unknown variables in first- and second-order models on a symmetric square. As we currently have no universal computational tool for approximate designs, these results can be used to determine optimal designs analytically in a few isolated and easy cases, as shown in the examples in Sect. 3.

However, from practical point of view, it is more important to be able to compute efficient exact designs, possibly even with some additional constraints given by the experimental conditions. We have shown a modified version of the algorithm of Harman et al. (2016) is a useful tool for such computations, even in the cases where there are several nontrivial constraints on the design.

In the models considered here, covariance matrix of random effects is assumed to be known. A natural question that arises while reading this work is how to perform in the situation where no prior knowledge about variances and covariances is available. In this case an estimation can be used. However, the quality of obtained designs depends on the accuracy of the estimation. For some particular structures of the covariance matrix it may happen that optimal designs turn out to be independent on the variance parameters (consider, for example, compound symmetry structure in Prus and Piepho 2021).

Acknowledgements This research was partially supported by Grant SCHW 531/16 of the German Research Foundation (DFG) and by Grant Number 1/0341/19 and 1/0362/22 from the Slovak Scientific Grant Agency (VEGA). A part of the research has been performed as the second author was visiting Comenius University of Bratislava in framework of the postdoc program (PKZ 91672803) of German Academic Exchange Service (DAAD). The authors are grateful to Radoslav Harman for fruitful discussions. Discussions with Rainer Schwabe on design criteria in multiple group-models were also very helpful.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix

See Tables 2, 3, 4, 5, 6 and 7.

Table 2 Exact D- and $IMSE$-efficient designs in quadratic model on the interval $[-1, 1]$ with respect to the numbers of observations m_i for $D_i = diag(d_1, d_2, d_3), i = 1, 2$

criterion			D						IMSE					
(d_1, d_2, d_3)	m_1	m_2	-1	0	1	-1	0	1	-1	0	1	-1	0	1
(1, 1, 1)	20	80	5	10	5	20	40	20	5	10	5	20	40	20
	50	50	12	25	13	13	25	12	13	25	12	12	25	13
	80	20	20	40	20	5	10	5	20	10	20	5	10	5
	40	160	10	20	10	40	80	40	10	20	10	40	80	40
	100	100	25	50	25	25	50	25	25	50	25	25	50	25
	160	40	40	80	40	10	20	10	40	80	40	10	20	10
(1, 1, 0)	20	80	5	10	5	20	40	20	9	2	9	15	49	16
	50	50	12	25	13	13	25	12	13	25	12	12	25	13
	80	20	20	40	20	5	10	5	15	49	16	9	2	9
	40	160	10	20	10	40	80	40	19	2	19	31	98	31
	100	100	25	50	25	25	50	25	25	50	25	25	50	25
	160	40	40	80	40	10	20	10	31	98	31	19	2	19
(0, 1, 1)	20	80	2	16	2	1	78	1	10	0	10	13	54	13
	50	50	1	48	1	1	48	1	12	26	12	13	24	13
	80	20	1	78	1	2	16	2	13	54	13	10	0	10
	40	160	2	36	2	1	158	1	20	0	20	27	106	27
	100	100	1	98	1	1	98	1	25	50	25	25	50	25
	160	40	1	158	1	2	36	2	27	106	27	20	0	20
(1, 0, 1)	20	80	8	4	8	40	1	39	4	12	4	25	30	25
	50	50	25	1	24	24	1	25	13	25	12	12	25	13
	80	20	39	1	40	8	4	8	25	30	25	4	12	4
	40	160	18	4	18	80	1	79	8	25	7	49	61	50
	100	100	50	1	49	49	1	50	25	50	25	25	50	25
	160	40	80	1	79	18	4	18	49	61	50	8	25	7
(1, 0, 0)	20	80	2	16	2	1	78	1	1	19	0	26	28	26
	50	50	1	48	1	1	48	1	13	25	12	12	25	13
	80	20	1	78	1	2	16	2	26	28	26	1	19	0
	40	160	2	36	2	1	158	1	1	38	1	53	54	53
	100	100	1	98	1	1	98	1	25	50	25	25	50	25
	160	40	1	158	1	2	36	2	53	54	53	1	38	1

Table 2 continued

criterion			D						IMSE					
(d_1, d_2, d_3)	m_1	m_2	-1	0	1	-1	0	1	-1	0	1	-1	0	1
(0, 1, 0)	20	80	10	0	10	19	42	19	10	0	10	16	48	16
	50	50	12	26	12	13	24	13	25	12	13	13	25	12
	80	20	19	42	19	10	0	10	16	48	16	10	0	10
	40	160	18	4	18	38	84	38	20	0	20	32	96	32
	100	100	25	50	25	25	50	25	25	50	25	25	50	25
	160	40	38	84	38	18	4	18	32	96	32	20	0	20
(0, 0, 1)	20	80	9	2	9	19	42	19	10	0	10	17	46	17
	50	50	12	26	12	13	24	13	12	26	12	13	24	13
	80	20	19	42	19	9	2	9	17	46	17	10	0	10
	40	160	18	4	18	38	84	38	20	0	20	35	91	34
	100	100	25	50	25	25	50	25	25	50	25	25	50	25
	160	40	38	84	38	18	4	18	34	91	35	20	0	20

Table 3 Exact D-efficient designs in quadratic model on the interval $[-1, 1]$ with constraints given by 40 and numbers of observations m_i for D_i given by 37 with $d_i = 0.5, i = 1, 2$

m_1	m_2	-1	-0.2	0	0.2	1	-1	-0.2	0	0.2	1
20	40	4	5	2	5	4	8	10	4	10	8
40	20	8	10	4	10	8	4	5	2	5	4
25	100	5	7	2	6	5	21	26	9	24	20
100	25	21	27	9	23	20	5	6	2	7	5

Table 4 Exact $IMSE$-efficient designs in quadratic model on the interval $[-1, 1]$ with constraints given by (40) and numbers of observations m_i for D_i given by (37) with $d_i = 0.5, i = 1, 2$

m_1	m_2	-1	-0.2	0	0.2	1	-1	-0.2	0	0.2	1
20	40	5	5	0	5	5	9	10	1	10	10
40	20	9	10	1	10	10	5	5	0	5	5
25	100	6	6	0	7	6	23	26	3	24	24
100	25	24	25	3	25	23	6	6	0	7	6

Table 5 Exact D- (left) and $IMSE$-efficient (right) designs in quadratic model on the interval $[-1, 1]$ with constraints given by (41) and numbers of observations m_i for D_i given by (37) with $d_i = 0.5, i = 1, 2$

(m_1, m_2)	-1	0	1	-1	0	1	-1	0	1	-1	0	1
(20,40)	2	17	1	3	34	3	2	6	2	4	23	3
(40,20)	3	34	3	2	17	1	4	23	3	2	6	2
(25,100)	3	14	3	8	74	8	3	14	3	9	52	9
(100,25)	8	74	8	3	14	3	8	63	9	3	14	3

Table 6 Exact D-efficient designs in quadratic model on the interval $[-1, 1]$ with constraints given by (42) and numbers of observations m_i for D_i given by (37) with $d_i = 0.5$, $i = 1, 2$

(m_1, m_2)	-1	-0.2	0	0.2	1	-1	-0.2	0	0.2	1
(20,40)	2	3	7	0	1	3	3	14	3	3
(40,20)	3	0	13	3	4	2	0	6	0	2
(25,100)	3	5	7	1	2	8	7	34	6	8
(100,25)	8	7	34	6	8	2	4	8	5	2

Table 7 Exact $IMSE$-efficient designs in quadratic model on the interval $[-1, 1]$ with constraints given by (42) and numbers of observations m_i for D_i given by (37) with $d_i = 0.5, i = 1, 2$

(m_1, m_2)	-1	-0.2	0	0.2	1	-1	-0.2	0	0.2	1
(20,40)	2	0	6	0	2	4	0	12	3	3
(40,20)	4	0	12	0	4	2	0	6	0	2
(25,100)	3	1	7	2	3	8	2	38	0	12
(100,25)	9	1	35	0	11	3	2	7	1	3

References

Aliev A, Fedorov V, Leonov S, McHugh B, Magee M (2012) Pkstamp library for constructing optimal population designs for pk/pd studies. Commun Stat Simul Comput 41(6):717–729

Cook RD, Thibodeau L (1980) Marginally restricted D-optimal designs. J Am Stat Assoc 75(370):366–371

Dumont C, Lestini G, Nagard HL, Mentré F, Comets E, Nguyen TT (2018) PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. Comput Methods Programs Biomed 156:217–229

Fedorov V, Jones B (2005) The design of multicentre trials. Stat Methods Med Res 14:205–248

Galil Z, Kiefer J (1977) Comparison of designs for quadratic regression on cubes. J Stat Plan Inference 1:121–132

Harman R, Bachratá A, Filová L (2016) Construction of efficient experimental designs under multiple resource constraints. Appl Stoch Models Bus Ind 32:3–17

Heiligers B (1992) Admissible experimental designs in multiple polynomial regression. J Stat Plan Inference 31:219–233

Nyberg J, Ueckert S, Strömberg EA, Hennig S, Karlsson MO, Hooker AC (2012) Poped: an extended, parallelized, nonlinear mixed effects models optimal design tool. Comput Methods Programs Biomed 108(2):789–805

Prus M (2022) Equivalence theorems for multiple-design problems with application in mixed models. J Stat Plan Inference 217:153–164

Prus M, Piepho H-P (2021) Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing. J Agric Biol Environ Stat 26:267–288

Prus M, Schwabe R (2016) Optimal designs for the prediction of individual parameters in hierarchical models. J R Stat Soc B 78:175–191
Schmelter T (2007) Considerations on group-wise identical designs for linear mixed models. J Stat Plan Inference 137:4003–4010
Schwabe R (1996) Optimum designs for multi-factor models. Springer, New York

Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing

Maryna Prus and Hans-Peter Piepho

Abstract

New crop varieties are extensively tested in multi-environment trials in order to obtain a solid empirical basis for recommendations to farmers. When the target population of environments is large and heterogeneous, a division into sub-regions is often advantageous. When designing such trials, the question arises how to allocate trials to the different sub-regions. We consider a solution to this problem assuming a linear mixed model. We propose an analytical approach for computation of optimal designs for best linear unbiased prediction of genotype effects and their pairwise linear contrasts and illustrate the obtained results by a real data example from Indian nation-wide maize variety trials. It is shown that, except in simple cases such as a compound symmetry model, the optimal allocation depends on the variance-covariance structure for genotypic effects nested within sub-regions.

Keywords: Target population of environments, multi-environment trials, mixed models, optimal design

1 Introduction

New crop varieties are usually evaluated for their performance in a target population of environments (TPE), where environments correspond to locations in specific years. This evaluation requires conducting randomized field trials at several environments sampled from the TPE. Such trials are called multi-environment trials (MET). Analysis of MET is routinely performed using linear mixed models comprising effects for genotypes, environments and their interaction (Isik *et al.* (2017)).

If the TPE is large and can be suitably stratified along geographical borders or agro-ecological zonations, it may be advantageous to subdivide the TPE into sub-regions. If the same set of genotypes is tested at a number of locations in each of the sub-regions, a linear mixed model may be fitted with random genotype-within-sub-region effects that allows estimating a genotype's average performance in each sub-region using best linear unbiased prediction (BLUP) (Atlin *et al.* (2000); Piepho and Möhring (2005)). If a covariance is assumed between a genotype's performance in different sub-regions, the model allows borrowing strength across sub-regions, meaning that estimates of mean performance in a sub-region become more accurate than when based on data from the sub-region alone (Kleinknecht *et al.* (2013)).

Whereas analysis of sub-divided TPE data has received some attention in the recent past, to the best of our knowledge the design of MET when a sub-division is envisaged has not been considered. The design of such trials has gained interest recently in endeavours to integrate trial networks across country borders (e.g., Horizon 2020 project INVITE = INnovations in plant VarIety Testing in Europe). The efficiency, and hence the optimal design is of great importance for the performance of trialling systems. The opportunity to integrate trials across sub-regions, potentially even across national borders, promises substantial efficiency gains but at the same time complicates both the design and analysis, because efficient estimates are still needed at the sub-region level. Our objective in this paper is to tackle this important design problem.

The design of MET for a sub-divided TPE involves two decisions: (1) The total number of environments at which to conduct the trials and (2) the allocation of this total number of environments to the different sub-regions. This paper is devoted to the second decision.

2 Model Specification and Prediction

In this work we use for design purposes a hierarchical linear mixed model (LMM) with K varieties (genotypes), P sub-regions, J locations and L replications per genotype in each location. Locations are nested lexicographically within sub-regions, i. e. there are J_i locations within sub-region i and $J = \sum_{i=1}^{P} J_i$. In this model observation l of genotype k in location j within the i-th sub-region is given by the random variable

$$Y_{ijkl} = \mu_i + \alpha_{ik} + \lambda_{ij} + \gamma_{ijk} + b_{ijl} + \varepsilon_{ijkl} \tag{1}$$

for $l = 1, \ldots, L$, $k = 1, \ldots, K$, $i = 1 \ldots P$ and $j = 1, \ldots, J_i$, where μ_i denotes the mean (fixed) effect of the i-th sub-region, α_{ik} is the interaction effect of genotype k in sub-region i, λ_{ij} is the effect of the j-th location within the i-th sub-region, γ_{ijk} denotes the effect of the k-th genotype in the j-th location within the i-th sub-region, b_{ijl} is the effect of the l-th replication in location j in sub-region i and ε_{ijkl} denotes the observational error. All effects besides μ_i are random. All random effects and observational errors are assumed to have zero mean. The variances are given by $\mathrm{var}(\varepsilon_{ijkl}) = \sigma^2$, $\mathrm{var}(\lambda_{ij}) = \sigma_\lambda^2 = \sigma^2 v_1$, $\mathrm{var}(\gamma_{ijk}) = \sigma_\gamma^2 = \sigma^2 v_2$ and $\mathrm{var}(b_{ijl}) = \sigma_b^2 = \sigma^2 v_3$ and the covariance matrix of the genotype effects $\boldsymbol{\alpha}_k = (\alpha_{1k}, \ldots, \alpha_{Pk})^\top$ is $\mathrm{Cov}(\boldsymbol{\alpha}_k) = \sigma^2 \mathbf{D}$, where $\mathbf{D}$ is some positive definite matrix. We note that the error variance typically varies among locations and hence our assumption of a homogeneous error variance constitutes an approximation. The approximation is a necessity at the design stage, as the error variance of individual locations is unpredictable. We also note that there is a justification in randomization theory for the use this model, despite the empirical evidence that error variances vary between locations (Caliński *et al.* (2009)).

Our main focus is the prediction of the genotype effects $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_K^\top)^\top$ and their pairwise linear contrasts $\boldsymbol{\theta}^{k,k'} = \boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k'}$, $k, k' = 1, \ldots, K$, $k \neq k'$. For a given total number of locations J we search for the numbers of locations $J_1, \ldots, J_P$ within the sub-regions, which are optimal for the prediction. Optimal designs for the estimation of fixed effects in LMM are well discussed in the literature (see e.g. Fedorov and Jones (2005) or Entholzner *et al.* (2005)). Less has been done for the prediction of random effects: the most general case - hierarchical random coefficient regression models - has been considered by Prus and Schwabe (2016). However, due to its more complicated covariance structure, model (1) is not a particular case of those models. Therefore, the proposed approach cannot be used here. Also, in the recently published work of Prus (2019) a simpler covariance structure has been assumed.

The next Lemma provides the BLUPs for the genotype effects $\boldsymbol{\alpha}$ and for their pairwise linear contrasts $\boldsymbol{\theta}^{k,k'}$.

Lemma 1. *a) The BLUP $\hat{\boldsymbol{\alpha}}_k$ of the genotype effects $\boldsymbol{\alpha}_k$ is given by*

$$\hat{\boldsymbol{\alpha}}_k = \left(diag(J_1, \ldots, J_P) + \frac{Lv_2 + 1}{L} \mathbf{D}^{-1} \right)^{-1} \left(J_1 \left(\bar{Y}_{1\cdot k\cdot} - \bar{Y}_{1\cdots} \right), \ldots, J_P \left(\bar{Y}_{P\cdot k\cdot} - \bar{Y}_{P\cdots} \right) \right)^\top, \tag{2}$$

where $\bar{Y}_{i\cdot k\cdot} = \frac{1}{J_i L} \sum_{j=1}^{J_i} \sum_{l=1}^{L} Y_{ijkl}$ and $\bar{Y}_{i\cdots} = \frac{1}{J_i K L} \sum_{j=1}^{J_i} \sum_{k=1}^{K} \sum_{l=1}^{L} Y_{ijkl}$.

b) The BLUP $\hat{\boldsymbol{\theta}}^{k,k'}$ of the pairwise linear contrasts $\boldsymbol{\theta}^{k,k'}$ is given by

$$\hat{\boldsymbol{\theta}}^{k,k'} = \left(diag(J_1, \ldots, J_P) + \frac{Lv_2 + 1}{L} \mathbf{D}^{-1} \right)^{-1} \left(J_1 \left(\bar{Y}_{1\cdot k\cdot} - \bar{Y}_{1\cdot k'\cdot} \right), \ldots, J_P \left(\bar{Y}_{P\cdot k\cdot} - \bar{Y}_{P\cdot k'\cdot} \right) \right)^\top. \tag{3}$$

We measure the performance of the prediction in terms of the mean squared error (MSE) matrix. The MSE matrices of the BLUPs for the genotype effects $\boldsymbol{\alpha}$ and for their pairwise linear contrasts $\boldsymbol{\theta}^{k,k'}$ are provided by the next lemma.

Lemma 2. *a) The MSE matrix of the BLUP* $\hat{\boldsymbol{\alpha}}$ *of* $\boldsymbol{\alpha}$ *is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \sigma^2 \left[\frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \otimes \mathbf{D} + (\mathbb{I}_K - \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top) \otimes \left(\frac{1}{Lv_2 + 1} \mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1} \right)^{-1} \right], \tag{4}$$

where $\mathbf{F} = \textit{block-diag}(\mathbb{1}_{LJ_1}, \ldots, \mathbb{1}_{LJ_P})$, $\mathbb{1}_s$ *is the vector of length* s *with all entries equal to* 1, $\mathbb{I}_s$ *is the* $s \times s$ *identity matrix and* $\otimes$ *denotes the Kronecker product.*

b) The MSE matrix of the BLUP $\hat{\boldsymbol{\theta}}^{k,k'}$ *of* $\boldsymbol{\theta}^{k,k'}$ *is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}^{k,k'} - \boldsymbol{\theta}^{k,k'}) = 2\sigma^2 \left(\frac{1}{Lv_2 + 1} \mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1} \right)^{-1}. \tag{5}$$

The proofs of the Lemmas 1 and 2 are deferred to the appendix.

Note that in Lemma 2 the MSE matrix (5) is the same for all k, k'. Therefore, we can fix k and k' and use the simplified notation $\boldsymbol{\theta}$ instead of $\boldsymbol{\theta}^{k,k'}$. Note also that both MSE matrices depend on the numbers of locations J_i only through the design matrix $\mathbf{F}$. Moreover, $\mathbf{F}$ coincides with the design matrix in a one-way ANOVA for P groups with sample sizes LJ_i under a group-mean parametrization.

Further, for a given total number of locations J, we search for the numbers of locations within sub-regions J_i, which minimize the MSE matrix (4) or (5) of the prediction for the genotype effects or for their pairwise linear contrasts, respectively.

3 Optimal Design

For the present optimization problem, we define (exact) designs as follows:

$$\xi := \begin{pmatrix} x_1 & \ldots & x_P \\ J_1 & \ldots & J_P \end{pmatrix}, \tag{6}$$

where $x_1, \ldots, x_P$ denote the sub-regions.

For analytical purposes we also introduce approximate designs (see e. g. Kiefer (1974)):

$$\xi := \begin{pmatrix} x_1 & \ldots & x_P \\ w_1 & \ldots & w_P \end{pmatrix}, \tag{7}$$

where $w_i = J_i/J$ is the weight of locations within sub-region i. For these designs the requirement of integer values of J_i is dropped and only the conditions $\sum_{i=1}^P w_i = 1$ and $w_i \geq 0$ have to be satisfied.

We define the information matrix as

$$\mathbf{M}(\xi) = \mathrm{diag}(w_1, \ldots, w_P) \tag{8}$$

and note that for exact designs the following condition is satisfied:

$$\mathbf{M}(\xi) = \frac{1}{LJ} \mathbf{F}^\top \mathbf{F}. \tag{9}$$

Then we extend the definitions of MSE matrices (4) and (5) with respect to approximate designs and obtain

$$\mathrm{MSE}_\alpha(\xi) = \sigma^2 \left[\frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top \otimes \mathbf{D} + (\mathbb{I}_K - \frac{1}{K} \mathbb{1}_K \mathbb{1}_K^\top) \otimes \left(\frac{LJ}{Lv_2 + 1} \mathbf{M}(\xi) + \mathbf{D}^{-1} \right)^{-1} \right] \tag{10}$$

and

$$\mathrm{MSE}_\theta(\xi) = 2\sigma^2 \left(\frac{LJ}{Lv_2 + 1} \mathbf{M}(\xi) + \mathbf{D}^{-1} \right)^{-1}. \tag{11}$$

3.1 A-optimal designs

The A-criterion for prediction may be defined as the trace of the MSE matrix (see e. g. Prus and Schwabe (2016)). For approximate designs this definition can be generalized using the extended MSE matrices (10) and (11). Then we evaluate (neglecting the constant factor $\frac{2\sigma^2}{LJ}(Lv_2+1)$) for the pairwise linear contrasts $\boldsymbol{\theta}$ the criterion

$$\Phi_A(\xi) = \operatorname{tr}\left(\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1}\right)^{-1}, \tag{12}$$

where $\boldsymbol{\Delta} = \frac{LJ}{Lv_2+1}\mathbf{D}$. Note that A-optimal designs depend on the number of replications L through the dispersion matrix Δ. This number needs to be fixed by the experimenter and is not subject to the optimization. A common number of replications per location in multi-environment trials is $L = 2$.

The A-criterion for the genotype effects $\boldsymbol{\alpha}$ differs from (12) only by the constant term $\sigma^2\operatorname{tr}(\mathbf{D})$ and the multiplicator $\frac{\sigma^2}{LJ}(K-1)(Lv_2+1)$, both of which have no influence on the solution to the optimization. Therefore, optimal designs for the prediction of the genotype effects and the linear contrasts are the same. The next theorem provides the optimality condition for approximate designs.

Theorem 1. *An approximate design ξ^* is A-optimal for the prediction of the genotype effects $\boldsymbol{\alpha}$ and their pairwise linear contrasts $\boldsymbol{\theta}$ iff*

$$\operatorname{tr}\left(\mathbf{M}(\xi^*)\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-2}\right) \geq \mathbf{e}_i^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-2} \mathbf{e}_i, \quad i = 1, \ldots, P, \tag{13}$$

where $\mathbf{e}_i$ is the vector of length P with the i-th entry equal to 1 and all other entries equal to 0.

For all i with $w_i^* > 0$ equality holds in (13).

Proof. The A-criterion (12) may be recognized as a particular Bayesian A-criterion. The optimality condition follows from Theorem 1 in Gladitz and Pilz (1982) for the linear transformation matrix $\mathbf{H} = \mathbb{I}_P$, the regression functions $\mathbf{f}(x_i) = \mathbf{e}_i$ and the design region $\mathcal{X} = \{x_1, \ldots, x_p\}$. (The notations $\mathbf{H}$, $\mathbf{f}$ and $\mathcal{X}$ are those used in Gladitz and Pilz (1982)). □

Corollary 1. *Let ξ^* be an A-optimal design for the prediction of the genotype effects $\boldsymbol{\alpha}$ and their pairwise linear contrasts $\boldsymbol{\theta}$. Let x_i and $x_{i'}$ be support points of ξ^* ($w_i > 0$ and $w_{i'} > 0$). Then the following equality holds:*

$$\mathbf{e}_i^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-2} \mathbf{e}_i = \mathbf{e}_{i'}^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-2} \mathbf{e}_{i'}. \tag{14}$$

Note that the terms on both sides of equation (14) denote the i-th and i'-th diagonal entries of the matrix $(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-2}$.

Example 1: Compound symmetry (CS) model. We consider a (CS) model with the particular covariance structure of genotype effects $\mathbf{D} = a\,\mathbb{1}_P\mathbb{1}_P^\top + b\,\mathbb{I}_P$ with positive b and real valued a with $a > -b/P$. For this model some optimal designs can be obtained explicitly.

Theorem 2. *In the CS model the (balanced) design ξ_P with $w_i = \frac{1}{P}$, $i = 1, \ldots, P$, is A-optimal for the prediction of the genotype effects and their pairwise linear contrasts.*

Proof. For the balanced design ξ_P the information matrix is given by $\mathbf{M}(\xi_P) = \frac{1}{P}\,\mathbb{I}_P$. Then it can be easily verified that all diagonal entries of the matrix $\left(\mathbf{M}(\xi_P) + \boldsymbol{\Delta}^{-1}\right)^{-2}$ are the same. For ξ_P the trace on the left-hand side of (13) is the average of the diagonal elements of $\left(\mathbf{M}(\xi_P) + \boldsymbol{\Delta}^{-1}\right)^{-2}$. Then we obtain equalities in (13) for all $i = 1, \ldots, P$. □

Note that the result of Theorem 2 can also be verified using symmetry considerations with respect to permutations of the sub-regions.

3.2 Optimal designs with respect to weighted A-criterion

In this and the following sections we focus on the prediction of the pairwise contrasts $\boldsymbol{\theta}$. We define the weighted A-criterion as the weighted sum across all sub-regions of the variances of the differences between the predicted and the real contrasts:

$$\Phi_{A_w} = \sum_{i=1}^{P} \ell_i \mathrm{var}(\hat{\theta}_i - \theta_i), \tag{15}$$

where $\ell_1, \ldots, \ell_P > 0$ denote coefficients, which are related to the sub-regions. One possible choice is the size of the sub-regions. Alternatively, equal weight may be given to each sub-region, meaning that $\ell_i = \frac{1}{P}$ for all i. (In this case the weighted A-criterion coincides with the standard A-criterion (12)). Then we extend this definition with respect to approximate designs and obtain the following criterion:

$$\Phi_{A_w}(\xi) = \mathrm{tr}\left(\mathbf{L}\left(\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1}\right)^{-1}\right), \tag{16}$$

where $\mathbf{L} = \mathrm{diag}(\ell_1, \ldots, \ell_P)$. The next theorem presents the optimality condition for approximate designs.

Theorem 3. *A design ξ^* is optimal for the prediction of pairwise linear contrasts $\boldsymbol{\theta}$ with respect to the weighted A-criterion iff*

$$\begin{aligned} &\mathrm{tr}\left(\mathbf{M}(\xi^*)\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{L}\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\right) \\ &\geq \mathbf{e}_i^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{L}\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{e}_i, \quad i = 1, \ldots, P. \end{aligned} \tag{17}$$

For all i with $w_i^* > 0$ equality holds in (17).

Proof. The weighted A-criterion (16) may be recognized as a particular Bayesian linear criterion. Optimality condition (17) follows from Theorem 1 in Gladitz and Pilz (1982) for the linear transformation matrix $\mathbf{H} = \mathbf{L}$. □

Corollary 2. *Let ξ^* be an optimal design with respect to the weighted A-criterion for the prediction of the pairwise linear contrasts $\boldsymbol{\theta}$. Let x_i and $x_{i'}$ be support points of ξ^* ($w_i > 0$ and $w_{i'} > 0$). Then the following equality holds:*

$$\mathbf{e}_i^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{L}\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{e}_i = \mathbf{e}_{i'}^\top \left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{L}\left(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\right)^{-1}\mathbf{e}_{i'}. \tag{18}$$

For the weighted A-criterion the optimal designs are not as easy to guess as for the standard A-criterion. Specifically, it is worth mentioning that the design ξ_ℓ with $w_i = \ell_i/\ell$, $\ell = \sum_{i=1}^{P} \ell_i$, which intuitively could be a solution of the optimization problem, is in general not optimal (see Example 2 below and the real data example in Section 4).

Example 2. We consider a simple example with $P = 2$ sub-regions, $L = 1$ replications and the following values of variance parameters: $v_2 = 1$ and $\mathbf{D} = \mathrm{diag}(1, 1)$. Then the optimal approximate designs are of the form

$$\xi := \begin{pmatrix} 1 & 2 \\ w & 1-w \end{pmatrix},$$

where $w = J_1/J$ is the weight of locations in the first sub-region, and the information matrix is given by $\mathbf{M} = \mathrm{diag}(w, 1-w)$. The present model is of the CS type and, therefore, the weight

Table 1: Numbers of locations J_1^ℓ in the first sub-region with respect to the weighted design ξ_ℓ and the optimal numbers of locations J_1^* in dependence on total number of locations J

J	15	30	45	60
J_1^*	11	21	31	41
J_1^ℓ	12	24	36	48

$w^* = 0.5$ is optimal for the A-criterion (12). For the weighted A-criterion we use the coefficients matrix $\mathbf{L} = \mathrm{diag}(4, 1)$ and assume $J \geq 5$. Then the weighted design is given by

$$\xi_\ell := \begin{pmatrix} 1 & 2 \\ 0.8 & 0.2 \end{pmatrix}.$$

For the weighted A-criterion (16) we obtain the formula

$$\Phi_{A_w}(\xi) = \frac{4}{w + 2/J} + \frac{1}{1 - w + 2/J}$$

and the resulting optimal weight w^* is

$$w^* = \frac{2J + 2}{3J}.$$

Table 1 summarizes the results for the numbers of locations $J_1^\ell = 0.8J$ in the first sub-region with respect to the weighted design ξ_ℓ and the optimal numbers of locations $J_1^* = w^*J$ (rounded) for different values of the total number of locations J. As we can see in the table the weighted design ξ_ℓ is in general not optimal.

3.3 Enforcing the same efficiency in each sub-region

For some studies it is required that the variances of the differences between the real and the predicted effects are the same for all sub-regions. For the model under investigation (model (1)) this condition is given by

$$\mathbf{e}_i^\top \left(\mathbf{M}(\xi^*) + \mathbf{\Delta}^{-1}\right)^{-1} \mathbf{e}_i = \mathbf{e}_{i'}^\top \left(\mathbf{M}(\xi^*) + \mathbf{\Delta}^{-1}\right)^{-1} \mathbf{e}_{i'}, \quad i, i' = 1, \ldots, P. \tag{19}$$

Under this condition the numbers of locations $J_1, \ldots, J_P$ can be obtained as a solution of a system of $P-1$ equations (for example fix $i = 1$ and $i' = 2, \ldots, P$) with $P-1$ unknown variables ($J_P = J - \sum_{i=1}^{P-1} J_i$) and no further optimization is needed.

Remark 1. *For the CS model the balanced design is a solution of* (19).

Note that the latter statement can be easily verified by symmetry considerations with respect to permutations of sub-regions or (more technically) by the fact that all diagonal entries of the matrix $\left(\mathbf{M}(\xi_P) + \mathbf{\Delta}^{-1}\right)^{-1}$ are the same for the balanced design ξ_P. Note also that not all A-optimal designs for the prediction of the genotype effects and their pairwise linear contrasts satisfy condition (19).

4 Real Data Example

We here consider variance components from a study on maize variety trials in India with $P = 5$ agroecological sub-regions and $L = 2$ replications per genotype in each location. The dataset

Table 2: Variance components used in this example (column "Variance in Model (1)") and how they are derived from the variance parameter estimates in Kleinknecht *et al.* (2013)

Effect	Model (1)	Model in Kleinknecht *et al.* (2013)	Variance in Model (1)	Variance in Kleinknecht *et al.* (2013)
Zone + mean	μ_i	$\mu + z_h + za_{hk} + a_k$	fixed	426+107+153
Genotype×zone	α_{ik}	$g_{i(h)} + gza_{ihk} + ga_{ik}$	$\sigma^2\mathbf{D}$ for α_k	$\mathbf{V} + 31\,\mathbb{1}_5\mathbb{1}_5^\top + 18\,\mathbb{I}_5$
Location×zone	$\lambda_{ij} + b_{ijl}$	$l_{jh} + la_{jhk}$	$\sigma^2_\lambda + \sigma^2_b = \sigma^2(v_1 + v_3)$	1129+1000
Gen×loc×zone + Obs errors	$\gamma_{ijk} + \varepsilon_{ijkl}$	$gl_{ijh} + e_{hijk}$	$\sigma^2_\gamma + \sigma^2 = \sigma^2(v_2 + 1)$	160+333

comprises four maturity groups of maize. Here, we consider only the extra-early maturity group (Kleinknecht *et al.* (2013), Tables 6 and 7). Based on the variance components reported in the paper, we derived variance components to be used in our design problems.

In Table 2 we summarize the variance components in the model under investigation (Model (1)) and how we determined these from the parameter estimates for the model considered by Kleinknecht *et al.* (2013). The matrix $\mathbf{V} = \mathrm{Cov}(g_{i(h)})$ denotes the covariance structure of genotype×zone effects in that paper. As we can see the values of σ^2 and v_2 are not given explicitly. Therefore, we consider different values of σ^2 and consequently of v_2 determined from the formula $\sigma^2(v_2 + 1) = 493$. Then the adjusted covariance matrix $\mathbf{\Delta}$ from formula (12) may be computed as

$$\mathbf{\Delta} = \frac{J}{493 - \sigma^2/2}(\mathbf{V} + 31\,\mathbb{1}_5\mathbb{1}_5^\top + 18\,\mathbb{I}_5). \tag{20}$$

We consider both the standard and the weighted A-criterion for first-order factor-analytic (FA) and compound symmetry (CS) variance-covariance structures, which were discussed in Kleinknecht *et al.* (2013) (see also Piepho (1997) for FA models).

4.1 Standard A-criterion

For the first-order factor-analytic model we take the covariance matrix $\mathbf{V}$ from Table 6 in Kleinknecht *et al.* (2013) (right part):

$$\mathbf{V} = \begin{pmatrix} 567 & 254 & 239 & 485 & 328 \\ 254 & 155 & 118 & 240 & 162 \\ 239 & 118 & 155 & 226 & 153 \\ 485 & 240 & 226 & 488 & 310 \\ 328 & 162 & 153 & 310 & 215 \end{pmatrix}. \tag{21}$$

Table 3 summarizes results for optimal designs in the first-order factor-analytic model. Note that optimal designs may assign zero locations in some sub-regions. Such designs are, however, not suitable when sub-region specific variances and covariances need to be estimated as is the case for the factor-analytic model underlying (21). Therefore, from here on we use the additional constraints of at least one location per sub-region: $J_i \geq 1$, $i = 1, \ldots, 5$. For computing efficiencies of the balanced design compared to the optimal designs we used standard formulae

$$\mathrm{Eff}_a = \frac{\Phi_A(\xi_a^*)}{\Phi_A(\xi_P)} \tag{22}$$

and

$$\mathrm{Eff}_e = \frac{\Phi_A(\xi_e^*)}{\Phi_A(\xi_P)}, \tag{23}$$

Table 3: Optimal numbers of locations per sub-region and efficiency of balanced design compared to optimal designs
with respect to the standard A-criterion in the FA model for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					Eff_a [1]	Exact design ξ_e^*					Eff_e [2]
		w_1	w_2	w_3	w_4	w_5		J_1	J_2	J_3	J_4	J_5	
10	50	0.37	0.10	0.10	0.33	0.10	0.91	4	1	1	3	1	0.91
	200	0.37	0.10	0.10	0.33	0.10	0.92	4	1	1	3	1	0.92
	400	0.34	0.10	0.15	0.31	0.10	0.94	3	1	2	3	1	0.94
20	50	0.33	0.13	0.18	0.31	0.04	0.95	7	3	3	6	1	0.95
	200	0.31	0.15	0.19	0.29	0.06	0.96	6	3	4	6	1	0.96
	400	0.29	0.16	0.20	0.27	0.09	0.97	6	3	4	5	2	0.97
40	50	0.27	0.17	0.20	0.25	0.10	0.97	11	7	8	10	4	0.98
	200	0.26	0.18	0.20	0.24	0.12	0.98	10	7	8	10	5	0.98
	400	0.25	0.19	0.21	0.23	0.13	0.98	10	8	8	9	5	0.99
100	50	0.23	0.19	0.21	0.22	0.15	0.99	23	19	21	22	15	0.99
	200	0.23	0.19	0.21	0.21	0.16	0.99	23	19	21	21	16	0.99
	400	0.22	0.20	0.21	0.21	0.17	0.99	22	20	20	21	17	0.99

[1] Computed using formula (22)

[2] Computed using formula (23)

where ξ_a^* and ξ_e^* denote optimal approximate and exact designs, respectively. As we can see in Table 3, optimal designs depend on both the total number of locations J and the error variance σ^2. In particular, the optimal proportion of allocation to sub-regions changes as the total number of environments changes. This behavior is typical for designs in mixed models and can be explained by the dependence of the covariance matrix: As we can see by formula (12), the influence of the covariance matrix $\mathbf{D}$ decreases with increasing total number of locations J. Note also that optimal designs generally tend to the balanced design if σ^2 or J increase. For designs without constraints this fact can be explained by formulae (12) and (20): According to (20) the adjusted covariance matrix $\mathbf{\Delta}$ increases with increasing σ^2 and J. For large $\mathbf{\Delta}$ the A-criterion (12) tends to the A-criterion in the fixed effects model, for which the balanced design is optimal. For constrained designs the same behavior can be observed with respect to σ^2. However, it does not necessarily hold in relation to J. For the compound symmetry model the covariance matrix $\mathbf{V}$ is taken from Table 6 in Kleinknecht *et al.* (2013) (left part, CS model):

$$\mathbf{V} = \begin{pmatrix} 308 & 270 & 270 & 270 & 270 \\ 270 & 308 & 270 & 270 & 270 \\ 270 & 270 & 308 & 270 & 270 \\ 270 & 270 & 270 & 308 & 270 \\ 270 & 270 & 270 & 270 & 308 \end{pmatrix}. \tag{24}$$

For the CS model we obtain optimal designs $J_i = J/5$, $i = 1, \ldots, 5$, which is in accordance with Theorem 2.

4.2 Weighted A-criterion

For the weighted A-criterion we used the coefficients $\ell_1 = 813685$, $\ell_2 = 432716$, $\ell_3 = 477365$, $\ell_4 = 995298$, $\ell_5 = 1174818$, which correspond to the areas of the sub-regions, respectively, as determined from a digitized version of the map shown in Kleinknecht *et al.* (2013).

Table 4: Optimal numbers of locations per sub-region and efficiencies of balanced and weighted designs compared to optimal designs with respect to the weighted A-criterion for the FA model for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$ [1]	$\text{Eff}_{a,\ell}$ [2]	Exact design ξ_e^*					$\text{Eff}_{e,P}$ [3]
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5	
10	50	0.35	0.10	0.10	0.35	0.10	0.88	0.92	3	1	1	4	1	0.88
	200	0.34	0.10	0.10	0.36	0.10	0.89	0.93	3	1	1	4	1	0.89
	400	0.34	0.10	0.10	0.36	0.10	0.90	0.94	3	1	1	4	1	0.90
20	50	0.34	0.05	0.09	0.37	0.15	0.91	0.95	7	1	2	7	3	0.91
	200	0.33	0.05	0.11	0.35	0.16	0.92	0.96	7	1	2	7	3	0.92
	400	0.30	0.08	0.12	0.32	0.18	0.94	0.97	6	2	2	6	4	0.94
40	50	0.28	0.09	0.13	0.30	0.19	0.94	0.97	11	4	5	12	8	0.95
	200	0.27	0.10	0.14	0.29	0.20	0.95	0.98	11	4	6	11	8	0.95
	400	0.27	0.10	0.14	0.29	0.20	0.96	0.98	10	5	6	11	8	0.96
100	50	0.24	0.13	0.15	0.26	0.22	0.97	0.98	24	13	15	26	22	0.97
	200	0.24	0.13	0.15	0.25	0.22	0.97	0.98	24	13	15	25	23	0.97
	400	0.23	0.14	0.16	0.25	0.23	0.97	0.98	23	14	15	25	23	0.97

[1] Computed using formula (25)

[2] Computed using formula (28)

[3] Computed using formula (26)

Table 4 summarizes the results for optimal designs under the factor-analytic model. The optimal designs for the compound symmetry model are presented in Table 5. For computing efficiencies of the balanced design ξ_P compared to optimal approximate and exact designs we used, respectively, formulae

$$\text{Eff}_{a,P} = \frac{\Phi_{A_w}(\xi_a^*)}{\Phi_{A_w}(\xi_P)} \tag{25}$$

and

$$\text{Eff}_{e,P} = \frac{\Phi_{A_w}(\xi_e^*)}{\Phi_{A_w}(\xi_P)}. \tag{26}$$

Efficiencies of the weighted design ξ_ℓ with weights $w_i = \ell_i/\ell$ for $\ell = \sum_{i=1}^{P} \ell_i$:

$$\xi_\ell = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.21 & 0.11 & 0.12 & 0.26 & 0.30 \end{pmatrix} \tag{27}$$

compared to optimal approximate designs have been obtained by formula

$$\text{Eff}_{a,\ell} = \frac{\Phi_{A_w}(\xi_a^*)}{\Phi_{A_w}(\xi_\ell)}. \tag{28}$$

We do not compare optimal exact designs with weighted design ξ_ℓ since ξ_ℓ is not in the class of exact designs.

As we can see in the tables the results for the factor-analytic and compound symmetry models are different, illustrating that the optimal designs depend on the variance-covariance structure of genotypic effects within sub-regions. In case of compound symmetry optimal designs are less sensitive to J and σ^2 than for the factor-analytic model. The efficiency of the weighted design for the CS model turned out to be very high. However, the weighted design is in general not optimal. To illustrate this fact, we used the (rather unrealistic for practical applications but

Table 5: Optimal numbers of locations per sub-region and efficiencies of balanced and weighted designs compared to optimal designs with respect to the weighted A-criterion for the CS model for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$	$\text{Eff}_{a,\ell}$	Exact design ξ_e^*					$\text{Eff}_{e,P}$
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5	
20	50	0.21	0.11	0.12	0.26	0.30	0.97	0.99	4	2	3	5	6	0.97
	200	0.21	0.11	0.13	0.26	0.29	0.97	0.99	4	2	3	5	6	0.97
	400	0.21	0.12	0.13	0.25	0.29	0.97	0.99	4	2	3	5	6	0.97
40	50	0.21	0.12	0.14	0.25	0.28	0.97	0.99	9	5	5	10	11	0.97
	200	0.21	0.13	0.14	0.25	0.27	0.97	0.99	8	5	6	10	11	0.97
	400	0.21	0.13	0.14	0.25	0.27	0.97	0.99	8	5	6	10	11	0.97
100	50	0.21	0.14	0.15	0.24	0.26	0.96	0.99	21	14	15	24	26	0.96
	200	0.21	0.14	0.15	0.24	0.26	0.96	0.99	21	14	15	24	26	0.96
	400	0.21	0.14	0.15	0.24	0.26	0.96	0.99	21	14	15	24	26	0.96
400	50	0.21	0.15	0.15	0.23	0.26	0.96	0.98	84	59	62	93	102	0.96
	200	0.21	0.15	0.16	0.23	0.25	0.96	0.97	84	59	63	93	101	0.96
	400	0.21	0.15	0.16	0.23	0.25	0.96	0.97	83	60	63	93	101	0.96

considered here to study the design properties) case $J = 400$, for which the efficiency goes down to 0.97.

All computations were performed using the procedures *od.SOCP* and *od.MISOCP* from the package *OptimalDesign* in *R* (see Harman and Filová (2016)) for optimal approximate and exact designs, respectively. The package has been originally developed for determining optimal designs in fixed-effects models. Harman and Prus (2018) proposed an approach for using it for compound Bayes risk criteria (CBRC) and, in particular, the Bayesian linear criteria. We used this approach here. Note that the exact designs obtained using *od.MISOCP* are optimal in the class of exact designs for the model under investigation for the given data. *R* code for our computations is given in the Supporting Information to this article. Note that the code produces the optimal allocation directly. All that is required from the user is the input of the variance values, the total number of locations, the number of replications per trial and the number of sub-regions.

4.3 Enforcing the same efficiency in each sub-region

When using the CS structure (24) in $\mathbf{D}$, we obtained the trivial solution $w_i = 0.2$, $i = 1, \ldots, 5$. When using the factor-analytic structure in (21), the solutions were as shown in Table 6. The exact designs were obtained by efficient rounding (see Pukelsheim and Rieder (1992)). We used the function *nlphqn* in *SAS/IML* to solve the nonlinear system of equations in (19). Since the presented results were obtained without using any optimization algorithm, we do not compare them with other designs.

5 Discussion

In crop research, the design of experiments is mainly considered in the context of a single environment and assuming that treatment effects are fixed (John and Williams (1995); Mead *et al.* (2012)). Recently, there has been an increased interest in design for experiments when treatments are modeled as correlated random effects using kinship or pedigree information (Bueno Filho and Gilmour (2003), Bueno Filho and Gilmour (2007), Cullis *et al.* (2006), Butler *et al.* (2014), Cullis *et al.* (2020), Heslot and Feoktistov (2020)). Also, the design of multi-environment trials has

Table 6: Optimal numbers of locations enforcing the same efficiency in each sub-region for FA model for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design					Exact design				
		w_1	w_2	w_3	w_4	w_5	J_1	J_2	J_3	J_4	J_5
20	50	0.342	0.148	0.205	0.302	0.003	6	3	4	6	1
	200	0.320	0.158	0.209	0.284	0.029	6	3	4	6	1
	400	0.291	0.172	0.211	0.260	0.065	6	3	4	5	2
40	50	0.274	0.179	0.212	0.247	0.088	11	7	8	10	4
	200	0.262	0.183	0.212	0.239	0.104	10	7	9	10	4
	400	0.247	0.189	0.211	0.228	0.125	10	8	8	9	5
100	50	0.231	0.194	0.209	0.217	0.150	23	19	21	22	15
	200	0.226	0.195	0.208	0.214	0.157	22	20	21	21	16
	400	0.219	0.197	0.206	0.210	0.167	22	20	20	21	17

been considered in a few papers, most notably in the context of partially replicated (p-rep) trials (Williams *et al.* (2014)), but also in broader contexts (González-Barrios *et al.* (2019)). To the best of our knowledge, however, the problem of allocation of location numbers in subdivided TPE has never been considered in any detail. The problem is reminiscent of optimal allocation in stage-wise sampling based on a nested random-effects model (Snedecor and Cochran (1967), p. 529) but the approach needed is more complex due to the linear mixed model involving several fixed and random effects and the optimization being targeted to the prediction of random effects. There is also some relation to small-area estimation in surveys where mixed models are used for estimation (Jiang and Lahiri (2006), Torabi and Jiang (2020)), but design in that context is not usually targeted at individual domains or small areas, and there is no notion of a larger number of treatments as in MET.

Our main focus was the optimal allocation of locations for different sub-regions with respect to the estimation of genotype effects and their pairwise linear contrasts for A- and particular linear (weighted A-) criteria. The proposed approach is based on the method of best linear unbiased prediction (BLUP). For our problem Bayesian optimal designs for a transformed covariance matrix of genotype effects turn out to be optimal. In the example we considered two kinds of models with respect to the covariance structure: first-order factor-analytic and compound symmetry. The resulting designs in both cases depend on the covariance structure, observational errors variance and the total number of locations in all sub-regions. The only exception is the standard A-criterion for compound symmetry: in this case balanced designs are optimal.

A general problem in the design of experiments under an assumed mixed model is that the optimal solution depends on the variance values. Our application here is no exception. For the design to be robust, it is therefore necessary to have reliable variance estimates, based on a sufficiently large database. Where there is uncertainty about the variance components, it is useful to try a range of values and inspect the sensitivity of the design. We have done this by considering four sets of variance values, originating from different maturity groups.

In a similar vein, all of our efficiency equations assume that at the analysis stage the variance components are known and are based on the optimal statistical properties of BLUP for this case (Searle *et al.* (1992)). In practice, variance components need to be estimated, in which case BLUP loses its optimal properties, and the efficiency realized with a design will usually be slightly diminished relative to the efficiencies reported here for known variance components. The Achilles' Heel in this context will usually be the estimation of genetic variances and covariances for the sub-regions (Piepho and Möhring (2005)). When the number of varieties is large, which is often the case, the loss in efficiency will be small. The loss in efficiency from the use of variance

component estimates in place of known values can be accounted for in analysis by using suitable small-sample inference methods (Kackar and Harville (1984), McLean and Sanders (1988)).

We have assumed that the trials at each location are laid out according to a randomized complete block design. While such designs are still very popular in variety testing, including the Indian maize trials we considered, it is common for trials involving a large number of varieties to be laid out in incomplete blocks (John and Williams (1995)). Our optimization approach can be adapted for these kinds of design, which typically display more complex variance-covariance structures with heterogeneity of variance as well as covariance among the BLUPs. We conjecture, however, that the optimal allocation of locations will be dominated by the variance components for the genotype-by-sub-region effects and hence an approximation for the within-trial variance structure would suffice for most practical purposes. Specifically, a resolvable incomplete block design could simply be analyses as a randomized complete blocking design, for which there is backing in randomization theory (Speed *et al.* (1985)). Alternatively, an effective error mean square based on a full analysis as a resolvable incomplete block design (Cochran and Cox (1957)) could be used as the error variance of an approximating optimization assuming a randomized complete block design.

Our criterion integrates the efficiencies for BLUPs of interest across sub-regions. There are three variations to this approach. Two of them take a weighted or unweighted average across sub-regions, and optimization typically leads to allocations that imply unequal efficiency between sub-regions. The third approach imposes the additional restriction that efficiency be the same for each sub-region. We think this latter approach is particularly relevant when several administrative entities (federal states or countries) join forces to link up their trialling networks for cross-boundary analyses. For such efforts to be successful it is vital that the benefit, in terms of efficiency gain compared to independent analysis, can be split equally between the administrative entities involved.

For the compound symmetry model the balanced design is optimal and the weighted design turned out to be highly efficient for the standard and weighted A-criterion, respectively. However, one should be more careful with the choice of design in case of the factor-analytic covariance structure, especially if there are large differences between variances of genotype effects for different sub-regions. Under a factor-analytic model, for the designs to be used in practice it is important to have accurate estimates of the variance structure, most importantly that of $\mathbf{D}$. The good news from a practical point of view is that with an increasing number of environments (J), the optimal designs under a factor-analytic model depend less on the exact variance values and approach that under a compound symmetry model. Hence, if reliable parameter estimates of the factor-analytic model are not available, deriving the design under a compound symmetry model seems a viable option.

A Proof of Lemmas 1 and 2

To make use of the theoretical results that are available in the literature (see e. g. Henderson (1975)) for the prediction of random parameters we will represent the model (1) as a particular case of the general LMM

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\zeta} + \boldsymbol{\epsilon} \tag{29}$$

with design matrices $\mathbf{X}$ and $\mathbf{Z}$ for the fixed effects and the random effects, respectively. In (29), $\boldsymbol{\beta}$ denotes the fixed effects and $\boldsymbol{\zeta}$ are the random effects. The random effects and the observational errors $\boldsymbol{\epsilon}$ are assumed to have zero mean and to be all uncorrelated with positive definite covariance matrices $\mathrm{Cov}(\boldsymbol{\zeta}) = \mathbf{G}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}) = \mathbf{R}$, respectively. Random effects and observational errors are assumed to be uncorrelated.

To present model (1) in form (29) we follow the next steps:

$$\mathbf{Y}_{ijk} = \mathbb{1}_L\,\mu_i + \mathbb{1}_L\,\alpha_{ik} + \mathbb{1}_L\,\lambda_{ij} + \mathbb{1}_L\,\gamma_{ijk} + \mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ijk}, \quad i = 1 \dots P, \quad k = 1, \dots, K, \quad j = 1, \dots, J_i,$$

where $\mathbf{b}_{ij} = (b_{ij1}, \dots, b_{ijL})^\top$.

$$\mathbf{Y}_{ik} = \mathbb{1}_{LJ_i}\,\mu_i + \mathbb{1}_{LJ_i}\,\alpha_{ik} + (\mathbb{I}_{J_i} \otimes \mathbb{1}_L)\,\boldsymbol{\lambda}_i + (\mathbb{I}_{J_i} \otimes \mathbb{1}_L)\,\boldsymbol{\gamma}_{ik} + \mathbf{b}_i + \boldsymbol{\varepsilon}_{ik}, \quad i = 1 \dots P, \quad k = 1, \dots, K,$$

where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{iJ_i})^\top$ and $\boldsymbol{\gamma}_{ik} = (\gamma_{i1k}, \dots, \gamma_{iJ_ik})^\top$.

$$\mathbf{Y}_k = \mathbf{F}\boldsymbol{\mu} + \mathbf{F}\boldsymbol{\alpha}_k + \mathbf{H}\boldsymbol{\lambda} + \mathbf{H}\boldsymbol{\gamma}_k + \mathbf{b} + \boldsymbol{\varepsilon}_k, \quad k = 1, \dots, K,$$

where $\mathbf{H} = (\mathbb{I}_J \otimes \mathbb{1}_L)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_P)^\top$, $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \dots, \boldsymbol{\lambda}_P^\top)^\top$ and $\boldsymbol{\gamma}_k = (\boldsymbol{\gamma}_{1k}^\top, \dots, \boldsymbol{\gamma}_{Pk}^\top)^\top$.

$$\mathbf{Y} = (\mathbb{1}_K \otimes \mathbf{F})\boldsymbol{\mu} + (\mathbb{I}_K \otimes \mathbf{F})\boldsymbol{\alpha} + (\mathbb{1}_K \otimes \mathbf{H})\boldsymbol{\lambda} + (\mathbb{I}_K \otimes \mathbf{H})\boldsymbol{\gamma} + (\mathbb{1}_K \otimes \mathbb{I}_{LJ})\mathbf{b} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \dots, \boldsymbol{\gamma}_K^\top)^\top$.

The latter equation may alternatively be written as

$$\mathbf{Y} = (\mathbb{1}_K \otimes \mathbf{F})\boldsymbol{\mu} + (\mathbb{I}_K \otimes \mathbf{F})\boldsymbol{\alpha} + \tilde{\boldsymbol{\varepsilon}}, \tag{30}$$

where $\tilde{\boldsymbol{\varepsilon}} := (\mathbb{1}_K \otimes \mathbf{H})\boldsymbol{\lambda} + (\mathbb{I}_K \otimes \mathbf{H})\boldsymbol{\gamma} + (\mathbb{1}_K \otimes \mathbb{I}_{LJ})\mathbf{b} + \boldsymbol{\varepsilon}$. Model (30) is of form (29) with $\mathbf{X} = (\mathbb{1}_K \otimes \mathbf{F})$, $\mathbf{Z} = (\mathbb{I}_K \otimes \mathbf{F})$, $\mathbf{G} = \mathrm{Cov}(\boldsymbol{\alpha}) = \sigma^2\mathbb{I}_K \otimes \mathbf{D}$ and

$$\mathbf{R} = \mathrm{Cov}(\tilde{\boldsymbol{\varepsilon}}) = \sigma^2((v_1\mathbb{1}_K\mathbb{1}_K^\top + v_2\mathbb{I}_K) \otimes \mathbb{I}_J \otimes (\mathbb{1}_L\mathbb{1}_L^\top) + v_3(\mathbb{1}_K\mathbb{1}_K^\top) \otimes \mathbb{I}_{LJ} + \mathbb{I}_{LJK}).$$

According to Henderson (1975) the BLUP of the random effects $\boldsymbol{\zeta}$ (which corresponds to $\boldsymbol{\alpha}$ in our model (30)) is given by

$$\begin{aligned} \hat{\boldsymbol{\zeta}} \;=\; & \left(\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^-\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\right)^{-1} \\ & \cdot \left(\mathbf{Z}^\top\mathbf{R}^{-1} - \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^-\mathbf{X}^\top\mathbf{R}^{-1}\right)\mathbf{Y}. \end{aligned} \tag{31}$$

Using this formula we obtain the BLUP for the genotype effects $\boldsymbol{\alpha}$, which results in formula (2). The MSE matrix of the BLUP of the random effects $\boldsymbol{\zeta}$ is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) = \left(\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^-\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}, \tag{32}$$

where $\mathbf{A}^-$ denotes a generalized inverse of $\mathbf{A}$. By this formula we obtain MSE matrix (4). Then using the relation $\boldsymbol{\theta}^{k,k'} = \boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k'} = ((\mathbf{e}_k - \mathbf{e}_{k'})^\top \otimes \mathbb{I}_P)\,\boldsymbol{\alpha}$ between the genotype effects and their pairwise contrasts we obtain formulae (3) and (5).

B Sensitivity Analysis

B.1 Standard A-criterion

We take values of the covariance matrix $\mathbf{V}$ from Tables 3, 4 and 5 in Kleinknecht *et al.* (2013) for late, medium and early maturity. Tables 7, 8 and 9 summarize the results for optimal designs for the standard A-criterion in FA model for late, medium and early maturity, respectively.

Table 7: Optimal numbers of locations per sub-region and efficiency of balanced design compared to optimal designs with respect to the standard A-criterion in the FA model (late maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					Eff_a	Exact design ξ_e^*					Eff_e
		w_1	w_2	w_3	w_4	w_5		J_1	J_2	J_3	J_4	J_5	
10	50	0.31	0.10	0.19	0.25	0.15	0.96	3	1	2	3	1	0.96
	200	0.31	0.10	0.20	0.24	0.15	0.96	3	1	2	2	2	0.96
	400	0.30	0.10	0.21	0.23	0.16	0.96	3	1	2	2	2	0.96
20	50	0.30	0.12	0.21	0.21	0.16	0.97	6	3	4	4	3	0.97
	200	0.29	0.13	0.21	0.21	0.16	0.97	6	3	4	4	3	0.97
	400	0.28	0.16	0.21	0.19	0.16	0.98	6	3	4	4	3	0.98
40	50	0.27	0.17	0.21	0.19	0.16	0.98	11	7	8	8	6	0.98
	200	0.26	0.18	0.21	0.19	0.16	0.98	11	7	8	8	6	0.98
	400	0.25	0.18	0.21	0.19	0.17	0.99	10	7	9	7	7	0.99
100	50	0.24	0.19	0.21	0.19	0.17	0.99	24	19	21	19	17	0.99
	200	0.23	0.20	0.21	0.19	0.17	0.99	23	20	21	19	17	0.99
	400	0.22	0.20	0.21	0.19	0.18	0.99	22	20	21	19	18	0.99

Table 8: Optimal numbers of locations per sub-region and efficiency of balanced design compared to optimal designs with respect to the standard A-criterion in the FA model (medium maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					Eff_a	Exact design ξ_e^*					Eff_e
		w_1	w_2	w_3	w_4	w_5		J_1	J_2	J_3	J_4	J_5	
10	50	0.30	0.10	0.21	0.29	0.10	0.91	3	1	2	3	1	0.91
	200	0.30	0.10	0.22	0.28	0.10	0.91	3	1	2	3	1	0.91
	400	0.29	0.10	0.24	0.27	0.10	0.91	3	1	2	3	1	0.91
20	50	0.31	0.05	0.27	0.32	0.05	0.90	6	1	5	7	1	0.90
	200	0.31	0.05	0.27	0.32	0.05	0.90	6	1	6	6	1	0.90
	400	0.30	0.05	0.27	0.29	0.09	0.92	6	1	5	6	2	0.92
40	50	0.29	0.06	0.27	0.27	0.11	0.93	11	2	11	11	5	0.93
	200	0.28	0.08	0.26	0.26	0.12	0.94	11	3	11	10	5	0.94
	400	0.27	0.11	0.25	0.24	0.13	0.95	11	4	10	10	5	0.95
100	50	0.25	0.14	0.24	0.22	0.15	0.97	25	14	24	22	15	0.97
	200	0.24	0.15	0.23	0.22	0.16	0.98	24	15	23	22	16	0.98
	400	0.23	0.16	0.23	0.21	0.17	0.98	23	16	23	21	17	0.98

Table 9: Optimal numbers of locations per sub-region and efficiency of balanced design compared to optimal designs with respect to the standard A-criterion in the FA model (early maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					Eff_a	Exact design ξ_e^*					Eff_e
		w_1	w_2	w_3	w_4	w_5		J_1	J_2	J_3	J_4	J_5	
10	50	0.42	0.10	0.15	0.23	0.10	0.93	4	1	2	2	1	0.93
	200	0.40	0.10	0.17	0.23	0.10	0.94	4	1	2	2	1	0.94
	400	0.37	0.10	0.20	0.23	0.10	0.95	4	1	2	2	1	0.95
20	50	0.35	0.09	0.22	0.23	0.11	0.96	7	2	4	5	2	0.96
	200	0.33	0.11	0.22	0.22	0.12	0.96	7	2	4	4	3	0.96
	400	0.31	0.14	0.22	0.20	0.13	0.97	6	3	4	4	3	0.97
40	50	0.29	0.16	0.22	0.20	0.13	0.98	12	6	9	8	5	0.98
	200	0.28	0.17	0.22	0.19	0.14	0.98	11	7	9	8	5	0.98
	400	0.26	0.18	0.22	0.19	0.15	0.98	11	7	9	7	6	0.98
100	50	0.24	0.19	0.22	0.19	0.16	0.99	24	19	22	19	16	0.99
	200	0.24	0.19	0.22	0.19	0.16	0.99	24	19	22	19	16	0.99
	400	0.23	0.20	0.21	0.19	0.17	0.99	23	20	21	19	17	0.99

Table 10: Optimal numbers of locations per sub-region and efficiencies of balanced and weighted designs compared to optimal designs with respect to the weighted A-criterion for the FA model (late maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$	$\text{Eff}_{a,\ell}$	Exact design ξ_e^*					$\text{Eff}_{e,P}$
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5	
10	50	0.30	0.10	0.10	0.28	0.22	0.93	0.97	3	1	1	3	2	0.93
	200	0.31	0.10	0.11	0.26	0.22	0.93	0.97	3	1	1	3	2	0.93
	400	0.30	0.10	0.12	0.25	0.23	0.93	0.97	3	1	1	3	2	0.94
20	50	0.31	0.05	0.14	0.26	0.24	0.93	0.97	6	1	3	5	5	0.93
	200	0.31	0.06	0.15	0.26	0.24	0.94	0.97	6	1	3	5	5	0.94
	400	0.29	0.08	0.15	0.24	0.24	0.95	0.97	6	2	3	5	4	0.95
40	50	0.29	0.10	0.15	0.23	0.23	0.95	0.97	11	4	6	9	10	0.95
	200	0.28	0.11	0.15	0.23	0.23	0.96	0.97	11	4	6	9	10	0.96
	400	0.27	0.12	0.16	0.23	0.23	0.95	0.97	11	5	6	9	9	0.95
100	50	0.25	0.14	0.16	0.22	0.23	0.97	0.97	25	14	16	22	23	0.97
	200	0.25	0.14	0.16	0.22	0.23	0.97	0.97	25	14	16	22	23	0.97
	400	0.24	0.14	0.16	0.22	0.24	0.97	0.97	24	14	16	22	24	0.97

Table 11: Optimal numbers of locations per sub-region and efficiencies of balanced and weighted designs compared to optimal designs with respect to the weighted A-criterion for the FA model (medium maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$	$\text{Eff}_{a,\ell}$	Exact design ξ_e^*					$\text{Eff}_{e,P}$
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5	
10	50	0.31	0.10	0.13	0.36	0.10	0.89	0.92	3	1	1	4	1	0.89
	200	0.31	0.10	0.14	0.35	0.10	0.89	0.92	3	1	1	4	1	0.90
	400	0.30	0.10	0.16	0.34	0.10	0.90	0.93	3	1	2	3	1	0.90
20	50	0.32	0.05	0.18	0.37	0.08	0.88	0.91	6	1	4	7	2	0.89
	200	0.31	0.05	0.19	0.34	0.11	0.89	0.91	6	1	4	7	2	0.89
	400	0.30	0.05	0.19	0.32	0.14	0.90	0.92	6	1	4	6	3	0.90
40	50	0.29	0.03	0.20	0.30	0.18	0.90	0.92	12	1	8	12	7	0.90
	200	0.29	0.03	0.20	0.39	0.19	0.91	0.93	12	1	8	12	7	0.91
	400	0.28	0.04	0.19	0.28	0.21	0.93	0.94	11	2	8	11	8	0.93
100	50	0.26	0.07	0.19	0.26	0.22	0.94	0.95	26	7	19	26	22	0.94
	200	0.25	0.09	0.18	0.26	0.22	0.95	0.96	25	9	18	26	22	0.95
	400	0.24	0.10	0.18	0.25	0.23	0.96	0.96	24	10	18	25	23	0.96

Table 12: Optimal numbers of locations per sub-region and efficiencies of balanced and weighted designs compared to optimal designs with respect to the weighted A-criterion for the FA model (early maturity) for different values of the total number of locations J and the error variance σ^2

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$	$\text{Eff}_{a,\ell}$	Exact design ξ_e^*					$\text{Eff}_{e,P}$
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5	
10	50	0.42	0.10	0.10	0.28	0.10	0.91	0.95	4	1	1	3	1	0.91
	200	0.39	0.10	0.10	0.28	0.13	0.91	0.95	4	1	1	3	1	0.91
	400	0.39	0.10	0.10	0.28	0.13	0.93	0.96	4	1	1	3	1	0.93
20	50	0.35	0.05	0.13	0.27	0.20	0.92	0.96	7	1	3	5	4	0.92
	200	0.34	0.05	0.13	0.27	0.21	0.93	0.96	7	1	3	5	4	0.93
	400	0.32	0.05	0.15	0.26	0.22	0.94	0.97	6	1	3	5	5	0.94
40	50	0.30	0.08	0.15	0.25	0.22	0.95	0.97	12	3	6	10	9	0.95
	200	0.29	0.09	0.16	0.24	0.22	0.96	0.98	12	3	6	10	9	0.96
	400	0.28	0.10	0.16	0.24	0.22	0.96	0.98	11	4	6	10	9	0.96
100	50	0.26	0.12	0.16	0.23	0.23	0.97	0.98	26	12	16	23	23	0.97
	200	0.25	0.13	0.16	0.23	0.23	0.97	0.98	25	13	16	23	23	0.97
	400	0.24	0.14	0.16	0.23	0.23	0.97	0.98	24	14	16	23	23	0.97

B.2 Weighted A-criterion

Tables 10, 11 and 12 summarize the results optimal designs with respect to the weighted A-criterion in the FA model for late, medium and early maturity, respectively.

Figure 1 illustrates the behavior of efficiencies of balanced and weighted designs with respect to optimal approximate and exact designs ($\text{Eff}_{a,P}$, $\text{Eff}_{a,\ell}$ and $\text{Eff}_{e,P}$ as in Section 4) in dependence on the total number of allocations J for weighted A-criterion in the CS model. For J we considered all multiples of 5 between 15 and 200. The error variance is fixed at $\sigma^2 = 50$, $\sigma^2 = 200$ and $\sigma^2 = 400$.

Acknowledgment

This research was partially supported by grant SCHW 531/16 of the German Research Foundation (DFG). The authors are grateful to Waqas Malik (University of Hohenheim) for determining the areas of the five breeding zones for maize in India based on a digitized map. The authors thank three referees, the Associate Editor and the Editor-in-Chief for helpful comments which improved the presentation of the results.

References

Atlin, G., Baker, R. J., McRae, K. B., and Lu., X. (2000). Selection response in subdivided target regions. *Crop Science*, **40**, 7–13.

Bueno Filho, J. S. D. S. and Gilmour, S. (2003). Planning incomplete block experiments when treatments are genetically related. *Biometrics*, **59**, 375–381.

Bueno Filho, J. S. D. S. and Gilmour, S. (2007). Block designs for random treatment effects. *Journal of Statistical Planning and Inference*, **137**, 1446–1451.

Butler, D. G., Smith, A. B., and Cullis, B. R. (2014). On the design of field experiments with correlated treatment effects. *Journal of Agricultural, Biological and Environmental Statistics*, **19**, 539–555.

Caliński, T., Czajka, S., Kaczmarek, Z., Krajewski, P., and Pilarczyk, W. (2009). Analysing the genotype by environment interactions under a randomization-derived mixed model. *Journal of Agricultural, Biological and Environmental Statistics*, **14**, 224–241.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*. John Wiley & Sons, New York.

Cullis, B. R., Smith, A., and Coombes, N. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 381–393.

Cullis, B. R., Smith, A., Cocks, N. A., and Butler, D. G. (2020). The design of early-stage plant breeding trials using genetic relatedness. *Journal of Agricultural, Biological and Environmental Statistics*, **25**, 553–578.

Entholzner, M., Benda, N., Schmelter, T., and Schwabe, R. (2005). A note on designs for estimating population parameters. *Biometrical Letters - Listy Biometryczne*, **42**, 25–41.

Fedorov, V. and Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248.

Gladitz, J. and Pilz, J. (1982). Construction of optimal designs in random coefficient regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, **13**, 371–385.

González-Barrios, P., Díaz-García, L., and Gutiérrez, L. (2019). Mega-environmental design: Using genotype × environment interaction to optimize resources for cultivar testing. *Crop Science*, **59**, 1899–1915.

Harman, R. and Filová, L. (2016). Package 'OptimalDesign'. https://cran.r-project.org/web/packages/OptimalDesign/index.html.

Harman, R. and Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. *Statistics and Probability Letters*, **137**, 135–141.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–477.

Heslot, N. and Feoktistov, V. (2020). Optimization of selective phenotyping and population design for genomic prediction. *Journal of Agricultural, Biological and Environmental Statistics*, **25**, 579–600.

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding*. Springer, New York.

Jiang, J. and Lahiri, P. (2006). Estimation of finite population domain means: A model-assisted empirical best prediction approach. *Journal of the American Statistical Association*, **101**, 301–311.

John, J. and Williams, E. (1995). *Cyclic and computer generated designs*. Chapman and Hall, London.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Am. Statist. Assoc.*, **79**, 853–862.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, **2**, 849–879.

Kleinknecht, K., Möhring, J., Singh, K., Zaidi, P., Atlin, G., and Piepho, H.-P. (2013). Comparison of the performance of blue and blup for zoned indian maize data. *Crop Science*, **53**, 1384–1391.

McLean, R. A. and Sanders, W. L. (1988). Approximating Degrees of Freedom for Standard Errors in Mixed Linear Models. In *Proceedings of the Statistical Computing Section*, pages 50–59. Alexandria, VA: American Statistical Association.

Mead, R., Gilmour, S., and Mead, A. (2012). *Statistical principles for the design of experiments*. Cambridge Univ. Press, Cambridge.

Piepho, H.-P. (1997). Analyzing genotype-environment data by mixed models with multiplicative effects. *Biometrics*, **53**, 761–766.

Piepho, H.-P. and Möhring, J. (2005). Best linear unbiased prediction of cultivar effects for subdivided target regions. *Crop Science*, **45**, 1151–1159.

Prus, M. (2019). Optimal designs in multiple group random coefficient regression models. *TEST*. https://doi.org/10.1007/s11749-019-00654-6.

Prus, M. and Schwabe, R. (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B*, **78**, 175–191.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, **79**, 763–770.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.

Snedecor, G. W. and Cochran, W.G. Press, A. (1967). *Statistical Methods*. Iowa State University Press, Ames.

Speed, T. P., Williams, E. R., and Patterson, H. D. (1985). A note on the analysis of resolvable block designs. *Journal of the Royal Statistical Society B*, **47**, 357–361.

Torabi, M. and Jiang, J. (2020). Estimation of mean squared prediction error of empirically spatial predictor of small area means under a linear mixed model. *Journal of Statistical Planning and Inference*, **208**, 82–93.

Williams, E., John, J. A., and Whitaker, D. (2014). Construction of more flexible and efficient p-rep designs. *Australian and New Zealand Journal of Statistics*, **56**, 89–96.

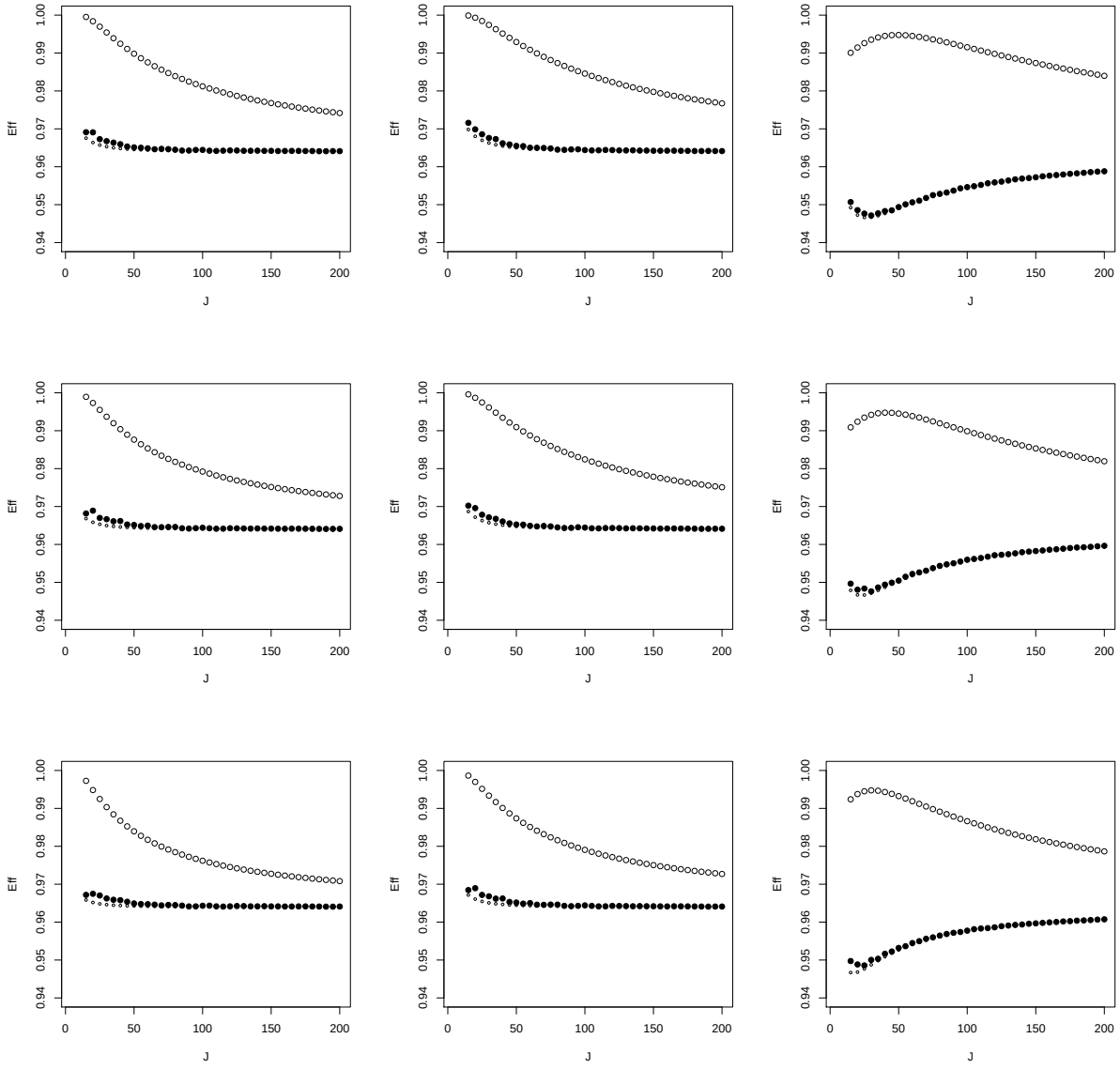

Figure 1: Efficiencies $\mathrm{Eff}_{a,P}$ (dark dots) $\mathrm{Eff}_{a,\ell}$ (big light dots) and $\mathrm{Eff}_{e,P}$ (small light dots) in dependence on the total number of allocations J for weighted A-criterion in the CS model for late (left panel), medium (middle panel) and early (right panel) maturity for $\sigma^2 = 50$ (first row), $\sigma^2 = 200$ (second row) and $\sigma^2 = 400$ (third row)