A STUDY OF LOCALLY D-OPTIMAL DESIGNS FOR THE EMAX MODEL WITH HETEROSCEDASTICITY

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Xiao Zhang

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Statistics

November 2022

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

A STUDY OF LOCALLY D-OPTIMAL DESIGNS

FOR THE EMAX MODEL WITH HETEROSCEDASTICITY

By

Xiao Zhang

The supervisory committee certifies that this dissertation complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Gang Shen
Chair

Dr. Rhonda Magel

Dr. Mingao Yuan

Dr. Kenneth Magel

Approved:

November 14, 2022 — Date

Dr. Rhonda Magel — Department Chair

ABSTRACT

The classic theory of locally optimal designs is developed on the center+error model assuming Gaussianity and homoscedasticity for random error, in which, the Maximum Likelihood Estimator (MLE) turns out to be the most efficient in model parameter estimation. However, these assumptions are typically absent in practice. In this work, we study the locally D-optimal design based on our new oracle Second-order Least Square Estimator (SLSE). We compare asymptotic efficiency of locally D-optimal designs obtained via SLSE, the Maximum quasi-Likelihood Estimator (MqLE) and Maximum Gaussian Likelihood Estimator (MGLE), in the case where the underlying probability distribution of response is non-Gaussian and heteroscedastic. We find that even with less stringent assumptions, asymptotic efficiency of the locally D-optimal designs obtained via MqLE is comparable to oracle SLSE in some cases, albeit lesser in general. As a demonstration of how the locally D-optimal design is numerically found, we apply our feasibility-based particle swarm optimization algorithm to the locally D-optimal design based on the original SLSE.

ACKNOWLEDGEMENTS

Words cannot express my gratitude to my dear advisor and chair of my committee Dr. Gang Shen for his invaluable patience and instruction. He's a mentor and a friend. I also could not have undertaken this journey without my defense committee, including Drs. Rhonda Magel, Mingao Yuan, and Kenneth Magel, who generously provided their knowledge and expertise. I feel grateful to all the instructors in Department of Statistics at North Dakota State University (NDSU) who offered the valuable courses and help. Additionally, this endeavor would not have been possible without the financial aid from Department of Statistics at NDSU, who financed my study/research and employed me as a teaching assistant.

I am also grateful to other professors from Department of Mathematics at NDSU, especially Drs. Doğan Çömez, Mariá de los Ángeles Alfonseca-Cubero, Frankie Chan, and Jason Boynton, for their mathematical teaching and help. Thanks also go to other employers, Dr. Xuehui Li employed me as a research assistant in Department of Plant Sciences at NDSU, and Dr. Ping Yan retained me as statistics manager at JiangSu HengRui Pharmaceutical Co., Ltd.

Thanks also go to Drs. Gang Shen, Ping Yan, Zhaohui Wei, Yaping Tu, Yue Wang, and Daqing Li, who referred me during my application to the Ph.D. program of statistics. I also thank my friends, colleagues, classmates, who support me, to some extent.

Lastly, and possibly most importantly, I would be remiss in not mentioning my family, especially my father Guoming Zhang, wife Wei Pan, son Wei Zhang, and sister Yu Zhang. Their belief in me has kept my spirits and motivation high during this tough process.

I sincerely appreciate all.

DEDICATION

This dissertation is dedicated to my dear deceased mother, Ms. Baifen Yan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ED50 . . . . median effective dose

GA . . . . genetic algorithms

GMM . . . . generalized method of moments

MLE . . . . maximum likelihood estimator

MqLE . . . . maximum quasi-likelihood estimator

MGLE . . . . maximum Gaussian likelihood estimator

SLSE . . . . second-order least square estimator

$N(\cdot, \cdot)$ . . . . Gaussian distribution

$IG(\cdot, \cdot)$ . . . . inverse Gaussian distribution

OPEF . . . . one-parameter exponential family

PSO . . . . particle swarm optimization

YBT . . . . Yang-Biedermann-Tang algorithm

LIST OF SYMBOLS

$\boldsymbol{\theta}$ .... model parameter vector

$e_D(\xi_0^\wedge)$ .... relative D-efficiency of design $\xi_0^\wedge$

μ .... treatment mean

ν .... treatment variance

γ .... skewness of response measure

κ .... kurtosis of response measure

τ .... skewness index

ξ .... clinical design

1. INTRODUCTION

In pre-clinical research, scientists' aim is to identify a promising medicine from thousands of feasible compounds. In the following clinical study, dozens of recruited patients (subjects) are assigned to several treatment groups of different doses within a range of $[L, U]$ (log scale) for identification of the maximum tolerated dose (MTD) and recommended phase II dose (RP2D). Given the total number of subjects allowed by the budget in a clinical study, an optimal clinical design is to find the optimal doses and number of subjects assigned to each treatment group, which satisfies certain optimality criteria related to the precision with which model parameters of the response measure are estimated.

The classical theory of optimal designs is developed on the center+error model, in which the response measure of each subject is assumed to be Gaussian distributed and homoscedastic across all treatment groups. At any nominal value of the model parameters, precision of the maximum Gaussian likelihood estimator (MGLE) or Fisher information of the model parameters is locally optimal in the sense of Lowner ordering. (The word "locally" is henceforth omitted for simplicity.) The optimal design based on Fisher information or MGLE, therefore, possesses the maximum design efficiency among other designs.

A response variable in practice, however, is typically heteroscedastic and its probability distribution is often unknown, much less Gaussian. Disregarding the underlying distribution of the response measure, Gaviria and & López-Ríos (2014) suggested constructing D-optimal designs via Fisher information with respect to Gaussian distribution with heteroscedasticity. Their work inadvertently exemplified an inefficient information-based design in presence of probability model mis-specification. Assuming homoscedasticity of the error in the center+error model, Gao and Zhou (2014) and Yeh and Zhou (2021) studied optimal designs based on the second-order least square estimator (SLSE), a sort of generalized method of moments more robust and efficient than the least square estimator if the underlying distribution of the error is skewed. The SLSE based optimal designs in those works presume the skewness and kurtosis in addition to mean and variance of the error. This method allows skewness in the underlying distribution, but still requires homoscedasticity, which implicitly guarantees the resulted optimal design not relying on the variance of the error. In a view of robustness to model mis-specification possessed by the maximum quasi-likelihood estimator (MqLE), as argued by Nelder and Wedderburn (1972) and Wedderburn (1974), Shen et al. (2016) proposed their MqLE based optimal design, which requires only the structure of mean and variance of the response measure. No assumption of Gaussianity or homoscedasticity is needed in this method. The MqLE based optimal design turns out to be as efficient as the MLE based when the underlying distribution is in one-parameter exponential family.

In this work, we start from the asymptotic variance of MLE, MqLE, MGLE and oracle SLSE for the parameters in the $Emax$ model, with a focus on efficiency of D-optimal designs based on the precision of these estimators, in the case where Gaussianity and homoscedasticity are absent from the underlying probability model for the response measures. As far as we know, no such comparison has yet been done. In light of no analytic solution of the D-optimal designs for the $Emax$ model, some numerical algorithm must be applied in search of the optimal designs. In this work, we provide a brief review of some popular D-optimal design searching algorithms of diverse streams, including the sensitivity-based algorithms and population-based metaheuristic algorithms, as well as the algorithm based on disciplined convex programming. Theorem 2 by Yang (2010) and the Equivalence theorem proposed by Kiefer and Wolfowitz (1959, 1960) play a vital role in algorithmic search of optimal designs. The former limits the number of distinct dose levels in an optimal design, while the latter helps verify whether a derived clinical design is optimal.

We borrow the feasibility-based particle swarm optimization algorithm (PSO) from the field of engineering in search of the D-optimal designs. All the resulted D-optimal designs are verified by the Equivalence theorem.

This work is organized as follows. Chapter 2 introduces D-optimal designs based on the four estimators and presents the main results concerning their asymptotic variance, and this section also reports the locally D-optimal designs based on these estimators in the case of the three underlying distributions for the response measure and compares their efficiency. Chapter 3 reviews some classic searching algorithms for D-optimal designs and works out some examples for demonstration of the feasibility-based PSO. Chapter 4 provides a brief discussion and conclusion. The tedious technical details are deferred to Appendix.

2. MODELS, D-OPTIMAL DESIGNS & EFFICIENCY

A clinical design involves desirable dose levels of medication and the number of subjects assigned to each dose level, which is usually denoted as $\xi = \{(x_i, n_i)\}_{i=1}^d$, where $x_i \in [L, U]$ is the dose level in log scale and n_i is the number of subjects allocated to the treatment group, satisfying $\sum_{i=1}^{d} n_i = n$, the total number of recruited subjects is given by some specific sample size calculation methodology under budget. A design ξ could be alternatively denoted as $\xi \triangleq \{(x_i, w_i)\}_{i=1}^d$ (approximate design), where w_i is the weight of subjects allocated to the treatment group satisfying $\sum_{i=1}^{d} w_i = 1$ and not constrained to nw_i being an integer for some n. For convenience, we consider only approximate designs in this work, because any approximate design can be rounded to an exact design without losing much efficiency, see Pukelsheim (2006), §12.

2.1. Models in clinical design

The classical center+error model plays an important role in the statistical theory of clinical designs in which the response measure is modeled by

$$y_{ij} = \mu(x_i, \boldsymbol{\theta}) + \sigma\epsilon_{ij},\ j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, d$$

where y_{ij} is the response measure of subject j in treatment group i; $\mu(x_i, \boldsymbol{\theta})$ is the mean of treatment group at dose x_i, which depicts the functional relationship between dose x_i and mean response, given the model parameter $\boldsymbol{\theta} \in \mathbb{R}^k$; σ is the unknown dispersion parameter; and ϵ_{ij} is the random error satisfying $\epsilon_{ij} \overset{iid}{\sim} N(0, 1)$. This model implies the response measure $y_{ij} \overset{ind}{\sim} N(\mu_i, \sigma^2)$, where $\mu_i = \mu(x_i, \boldsymbol{\theta})$.

A popular choice of the treatment mean function $\mu(x, \boldsymbol{\theta})$ is the $Emax$ model for dose-response analyses in pharmacokinetics and pharmacodynamics, see Macdougall (2006), Li and Majumdar (2008) and Shen et al. (2016), among many others. Assuming the response measure $y_{ij} \overset{ind}{\sim} N(\mu_i, \sigma^2)$, Yang (2010) corroborated the $Emax$ model has minimally supported optimal designs. For convenience, we use the $Emax$ model in this study for the treatment mean. The $Emax$ model is defined as

$$\mu(x, \boldsymbol{\theta}) = \frac{\theta_1}{1 + e^{\theta_2 x + \theta 3}} + \theta_4, \tag{2.1}$$

where the model parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)' \in \mathcal{R}^4$.

- θ_1: > 0, $Emax$, where $Emax$ is the maximum effect attributable to the drug;
- θ_2: $-$Hill's coefficient;
- θ_3: Hill's coefficient $\times \log(ED_{50})$, where ED_{50} is the dose producing half of $Emax$;
- θ_4: The basal effect of response.

With the basal effect θ_4=0, Michaelis and Menten (1913) first proposed the reduced $Emax$ model, the $Michaelis - Menten$ model, which is defined as

$$\mu(x, \boldsymbol{\theta}) = \frac{\theta_1}{1 + e^{\theta_2 x + \theta 3}}, \tag{2.2}$$

where the model parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)' \in \mathcal{R}^3$.

For the classical center+error model, $\hat{\boldsymbol{\theta}}_{MGL}$, the MGLE of model parameter $\boldsymbol{\theta}$ enjoys the minimum asymptotic variance among all the estimators in the sense of Loewner ordering. Although the assumption of Gaussianity and homoscedasticity is frequently reported absent in practice and consequently the classical center+error model is still widely used with diverse choices of the treatment mean function $\mu(x_i, \boldsymbol{\theta})$, simply because of its mathematical convenience.

In order to accommodate heteroscedasticity of the response measure, we may set $\operatorname{var} y_{ij} = \nu_i \triangleq \nu(\mu_i)$ for the treatment variance at dose x_i, where $\mu_i = \mu(x_i, \boldsymbol{\theta})$ is the treatment mean at dose x_i and $\nu(\cdot) > 0$ is some positive function. This setting actually includes the case of homoscedasticity with the choice of ν_i to be a universal constant. Gaviria and & López-Ríos (2014) considered modeling the response measure with $N(\mu_i, \nu_i)$, where $\nu(\mu_i) = \sigma^2 \mu_i^{2\tau}$ for some $\sigma^2, \tau > 0$. Shen et al. (2016) used Gaussian and Gamma distribution, but followed Cook and Weisberg (1983) and Atkinson and Cook (1995), setting $\nu(\mu_i) = \sigma^2 e^{h\mu_i}$ for some constant h and $\sigma^2 > 0$ instead.

2.2. Optimal designs

To obtain optimal designs based on an estimator $\hat{\boldsymbol{\theta}}$ for the model parameter $\boldsymbol{\theta}$, it is critical to find $\operatorname{var}_\xi \hat{\boldsymbol{\theta}}$ or its inverse, the precision of $\hat{\boldsymbol{\theta}}$, for any design ξ. In particular when the estimator $\hat{\boldsymbol{\theta}}$ is $\hat{\boldsymbol{\theta}}_{ML}$, the maximum likelihood estimator (MLE), the precision of $\hat{\boldsymbol{\theta}}_{ML}$ is simply the Fisher information of $\boldsymbol{\theta}$. Let χ be the collection of all the clinical designs $\xi = \{(x_i, w_i)\}_{i=1}^d$ satisfying $x_i \in [L, U]$, $w_i \geq 0$, $\sum_{i=1}^d w_i = 1$ and $d < +\infty$. In practice, the D-optimal design $\xi^* = \operatorname{argmin}_{\xi \in \chi} \det(\operatorname{var}_\xi \hat{\boldsymbol{\theta}})$ minimizing determinant of $\operatorname{var}_\xi \hat{\boldsymbol{\theta}}$, the E-optimal design $\xi^\dagger = \operatorname{argmin}_{\xi \in \chi} \lambda_{max}(\operatorname{var}_\xi \hat{\boldsymbol{\theta}})$ minimizing the maximum eigenvalue of $\operatorname{var}_\xi \hat{\boldsymbol{\theta}}$, and the A-optimal design $\xi^\star = \operatorname{argmin}_{\xi \in \chi} \operatorname{tr}(\operatorname{var}_\xi \hat{\boldsymbol{\theta}})$ minimizing the trace of $\operatorname{var}_\xi \hat{\boldsymbol{\theta}}$, are of great interest to statisticians among other alphabetic optimal designs. Indeed, the D-optimal design seeks to minimize volume of the confidence ellipsoid of $\hat{\boldsymbol{\theta}}$ and the E-optimal design aims to minimize the longest axis length of the confidence ellipsoid of $\hat{\boldsymbol{\theta}}$; while the A-optimal design intends to minimize sum of the variance of each component of $\hat{\boldsymbol{\theta}}$. Of all these optimal designs, only the D-optimal designs are invariant under reparametrization of $\boldsymbol{\theta}$.

The Equivalence theorem proposed by Kiefer and Wolfowitz (1959) asserts that if there exists a vector function $f(x, \boldsymbol{\theta})$ such that the precision of $\hat{\boldsymbol{\theta}}$ with respect to design ξ

$$M_{\wedge}(\xi, \boldsymbol{\theta}) \triangleq [\mathrm{var}_{\xi}\, \hat{\boldsymbol{\theta}}]^{-1} = \int_L^U f(x, \boldsymbol{\theta}) f'(x, \boldsymbol{\theta}) \xi(dx) \tag{2.3}$$

holds for any design ξ, where $\int_L^U \xi(dx)$ is the Lebesgue integral with respect to ξ as a probability measure, then ξ^* is the $\hat{\boldsymbol{\theta}}$-based D-optimal design iff $\max_x d(x, \xi^*) \leq k$, where $d(x, \xi^*) \triangleq f'(x, \boldsymbol{\theta}) M_{\wedge}^{-1}(\xi^*, \boldsymbol{\theta}) f(x, \boldsymbol{\theta})$ is the sensitivity function, k is the dimension of $\boldsymbol{\theta}$. Furthermore, according to Silvey (1980), Lemma 5.1.3, if the locally D-optimal design ξ^* is minimally supported, i.e., the number of supports of ξ^* equals to the dimension of $\boldsymbol{\theta}$, then ξ^* must be a uniform design (equally weighted).

2.3. D-efficiency

An estimator based D-optimal design seeks to minimize the determinant of the variance of the estimator. For the locally D-optimal design $\xi_0^{\wedge}$ based on an estimator $\hat{\boldsymbol{\theta}}$ at $\boldsymbol{\theta}_0$, its relative D-efficiency is

$$e_D(\xi_0^{\wedge}) \triangleq \left\{ \frac{\det[M_{\wedge}(\xi_0^{\wedge}, \boldsymbol{\theta}_0)]}{\det[M_{*}(\xi_0^{*}, \boldsymbol{\theta}_0)]} \right\}^{1/k} \tag{2.4}$$

where, $M_{\wedge}(\xi_0^{\wedge}, \boldsymbol{\theta}_0) = [\mathrm{var}_{\xi_0^{\wedge}}\, \hat{\boldsymbol{\theta}}]^{-1}$ is the precision of $\hat{\boldsymbol{\theta}}$ under its locally D-optimal design $\xi_0^{\wedge}$; $M_{*}(\xi_0^{*}, \boldsymbol{\theta}_0) = [\mathrm{var}_{\xi_0^{*}}\, \hat{\boldsymbol{\theta}}_{ML}]^{-1}$ is the precision of $\hat{\boldsymbol{\theta}}_{ML}$ under its locally D-optimal design ξ_0^{*};

and k is the dimension of parameter $\boldsymbol{\theta}$. Clearly, $e_D(\xi_0^{\wedge}) \in [0,1]$ and the higher the value, the larger the relative efficiency of the D-optimal design $\xi_0^{\wedge}$.

2.4. Estimators: SLSE, MqLE, MGLE

In principle, construction of D-optimal designs shall be based on precision of $\hat{\boldsymbol{\theta}}_{ML}$ or the Fisher information of model parameter $\boldsymbol{\theta}$ with respect to the underlying probability model which, however, could be mis-specified or even unavailable in practice. Choosing a reasonable estimator for model parameters then depends on how much information concerning the probability distribution of the response measure is available, e.g., the mean, variance, skewness and/or kurtosis. SLSE and MqLE are among other popular M-estimators beside MLE for estimation of model parameter $\boldsymbol{\theta}$, which either maximizes or minimizes the criterion function:

$$g_n(y, x, \boldsymbol{\theta}) \triangleq \frac{1}{n} \sum_{i=1}^{d} \sum_{j=1}^{n_i} g(y_{ij}, x_i, \boldsymbol{\theta}) \xrightarrow{p} \int_L^U E[g(y, x, \boldsymbol{\theta})] \xi(dx) \tag{2.5}$$

where $g(y_{ij}, x_i, \boldsymbol{\theta})$ is the kernel of the criterion function. For these estimators and the MGLE in particular, with the model specification $\mu_i = \mu(x_i, \boldsymbol{\theta})$ and $\nu_i = \nu(\mu_i)$,

1. MLE: $g(y_{ij}, x_i, \boldsymbol{\theta}) = \log f(y_{ij}, \mu_i, \nu_i)$, where $f(y_{ij}, \mu_i, \nu_i)$ is the pdf/pmf of y_{ij};

2. MGLE: $g(y_{ij}, x_i, \boldsymbol{\theta}) = \log \phi(y_{ij}, \mu_i, \nu_i)$, where $\phi(y_{ij}, \mu_i, \nu_i)$ is the pdf of $N(\mu_i, \nu_i)$;

3. MqLE: $g(y_{ij}, x_i, \boldsymbol{\theta}) = \int_{-\infty}^{\mu(x_i, \boldsymbol{\theta})} \frac{y_{ij} - u}{\nu(u)} du$;

4. SLSE: $g(y_{ij}, x_i, \boldsymbol{\theta}) = \rho'_{ij}(\boldsymbol{\theta}) W(x_i) \rho_{ij}(\boldsymbol{\theta})$, where $\rho_{ij}(\boldsymbol{\theta}) \triangleq (y_{ij} - \mu_i, y_{ij}^2 - \mu_i^2 - \nu_i)'$, and $W(x_i)$ is a user-specified 2×2 non-negative-definite matrix depending on x_i. According to Wang and Leblanc (2008), the best choice of the weighting matrix $W(x_i)$ that

yields the minimum variance of the estimator is $W_0(x_i) \triangleq \{E_{\boldsymbol{\theta}_0}[\rho_{ij}(\boldsymbol{\theta}_0)\rho'_{ij}(\boldsymbol{\theta}_0)|x_i]\}^{-1}$, where $\boldsymbol{\theta}_0$ is the nominal value of $\boldsymbol{\theta}$. It turns out $W_0(x_i) = \begin{pmatrix} \tilde{\nu}_i & \tilde{\gamma}_i+2\tilde{\mu}_i\tilde{\nu}_i \\ \tilde{\gamma}_i+2\tilde{\mu}_i\tilde{\nu}_i & \tilde{\kappa}_i-\tilde{\nu}_i^2+4\tilde{\mu}_i^2\tilde{\nu}_i+4\tilde{\mu}_i\tilde{\gamma}_i \end{pmatrix}^{-1}$, where $\tilde{\mu}_i = \mu(x_i, \boldsymbol{\theta}_0)$, $\tilde{\nu}_i = \nu(\tilde{\mu}_i)$, $\tilde{\gamma}_i \triangleq E_{\tilde{\mu}_i}(y_{ij} - \tilde{\mu}_i)^3$ and $\tilde{\kappa}_i \triangleq E_{\tilde{\mu}_i}(y_{ij} - \tilde{\mu}_i)^4$.

SLSE with the choice of $W_0(x_i)$ as the weighting matrix is an oracle, since it requires a priori knowledge of the function $\gamma(\mu) = E_\mu(y-\mu)^3$ and $\kappa(\mu) = E_\mu(y-\mu)^4$.

Given the nominal value of parameter $\boldsymbol{\theta}$, the asymptotic variance of M-estimators under design ξ as defined by Eq(2.5) have the well-known sandwich form,

$$\Big\{\int_L^U E_{\boldsymbol{\theta}}(\ddot{g})\xi(dx)\Big\}^{-1}\Big\{\int_L^U E_{\boldsymbol{\theta}}(\dot{g}\dot{g}')\xi(dx)\Big\}\Big\{\int_L^U E_{\boldsymbol{\theta}}(\ddot{g})\xi(dx)\Big\}^{-1}, \tag{2.6}$$

where, $E_{\boldsymbol{\theta}}$ is the expectation operator under the distribution parametrized with $\boldsymbol{\theta}$, $\dot{g} = \frac{\partial g}{\partial \boldsymbol{\theta}}(y, x, \boldsymbol{\theta})$ and $\ddot{g} = \frac{\partial^2 g}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(y, x, \boldsymbol{\theta})$. The asymptotic variance of $\hat{\boldsymbol{\theta}}_{ML}$ (MLE) under design ξ is therefore $\big\{\int_L^U [E_\mu \dot{l}^2]\dot{\mu}\dot{\mu}'\xi(dx)\big\}^{-1}$, where $\dot{\mu} \triangleq \frac{\partial \mu}{\partial \theta}$ and $\dot{l} \triangleq \frac{\partial}{\partial \mu}\log f(y, \mu, \nu(\mu))$ is the score function. Shen et al. (2016) corroborated the asymptotic variance of $\hat{\boldsymbol{\theta}}_{MqL}$ (MqLE) under design ξ is $\big\{\int_L^U \frac{1}{\nu}\dot{\mu}\dot{\mu}'\xi(dx)\big\}^{-1}$. Correspondingly,

Proposition 2.4.1. *The asymptotic variance of $\hat{\boldsymbol{\theta}}_{SLS}$ (oracle SLSE) under design ξ is*

$$\Big\{\int_L^U \big[\frac{1}{\nu} + \frac{(\gamma - \dot{\nu}\nu)^2}{\nu(\kappa\nu - \nu^3 - \gamma^2)}\big]\dot{\mu}\dot{\mu}'\xi(dx)\Big\}^{-1}, \tag{2.7}$$

where $\dot{\nu} \triangleq \frac{d\nu}{d\mu}$, $\dot{\mu} \triangleq \frac{\partial \mu}{\partial \theta}$, $\gamma \triangleq E_\mu(y-\mu)^3$ and $\kappa \triangleq E_\mu(y-\mu)^4$. Consequently, the asymptotic variance $\operatorname{var}_\xi \hat{\boldsymbol{\theta}}_{SLS} \preceq \operatorname{var}_\xi \hat{\boldsymbol{\theta}}_{MqL}$ in the sense of Loewner ordering.

The D-optimal designs based on MLE, MqLE and oracle SLSE could be easily obtained, since their precision under any design ξ all satisfy the assumption of the Equivalence theorem as quoted in Eq(2.3).

The asymptotic variance of $\hat{\boldsymbol{\theta}}_{MGL}$ (MGLE) under design ξ, in general, turns out to be

$$\left\{\int_L^U (\frac{1}{\nu}+\frac{\dot{\nu}^2}{2\nu^2})\dot{\mu}\dot{\mu}'\xi(dx)\right\}^{-1}\left\{\int_L^U (\frac{1}{\nu}-\frac{\dot{\nu}^2}{4\nu^2}+\frac{\dot{\nu}\gamma}{\nu^3}+\frac{\dot{\nu}^2\kappa}{4\nu^4})\dot{\mu}\dot{\mu}'\xi(dx)\right\}\left\{\int_L^U (\frac{1}{\nu}+\frac{\dot{\nu}^2}{2\nu^2})\dot{\mu}\dot{\mu}'\xi(dx)\right\}^{-1}. \tag{2.8}$$

In the case of the response measure $y_{ij} \overset{ind}{\sim} N(\mu_i, \nu_i)$, the asymptotic variance of $\hat{\boldsymbol{\theta}}_{MGL}$ is then reduced to $\left\{\int_L^U [\frac{1}{\nu}+\frac{\dot{\nu}^2}{2\nu^2}]\dot{\mu}\dot{\mu}'\xi(dx)\right\}^{-1}$.

In the absence of the knowledge of the underlying distribution, Gaviria and & López-Ríos (2014) suggested to use information $I_\xi \triangleq \int_L^U [\frac{1}{\nu}+\frac{\dot{\nu}^2}{2\nu^2}]\dot{\mu}\dot{\mu}'\xi(dx)$ to derive the D-optimal design. Actually, this I_ξ is the Fisher information of $\boldsymbol{\theta}$ with respect to the design ξ and postulated Gaussian distribution, which is completely detached from the underlying probability model. Indeed, their MGLE based method provided an excellent example of information based inefficient designs due to the model mis-specification when the underlying distribution is non-Gaussian. Following Wang and Leblanc (2008), Yin and Zhou (2017) and Gao and Zhou (2017) studied the SLSE based D-optimal design, which is essentially a generalized method of moments allowing asymmetric distribution. Although the assumption of Gaussianity is relaxed to some extent, it requires ν_i being some constant. In view of this, we set a varying $\nu_i = \nu(\mu_i)$ as well for the oracle SLSE in this study. Shen et al. (2016) proposed to obtain D-optimal designs based on MqLE, which obviates Gaussianity and homoscedasticity. The resulted D-optimal design is identical to that based on MLE when the underlying distribution belongs to the one-parameter exponential family (OPEF).

In our following numerical study of relative efficiency of D-optimal design for the $Emax$ model in the case where Gaussianity or homoscedasticity is absent, we consider the probability families as below with parameter ψ controlling model departure from OPEF as the underlying distribution for the response measure. Note all the probability models satisfy $Ey_{ij} = \mu_i$ and $\operatorname{var} y_{ij} = \nu_i$, and, intuitively, the larger the ψ value, the farther the departure from the OPEF.

1. $y_{ij} \overset{ind}{\sim} N(\mu_i, \nu_i)$ with $\nu_i = 300 + 60\psi\mu_i$, symmetric, in OPEF if $\psi = 0$;
2. $y_{ij} \overset{ind}{\sim} Gamma(\frac{\mu_i^2}{\nu_i}, \frac{\mu_i}{\nu_i})$ with $\nu_i = \mu_i^2 + 60\psi\mu_i$, asymmetric, in OPEF if $\psi = 0$;
3. $y_{ij} \overset{ind}{\sim} IG(\mu_i, \frac{\mu_i^3}{\nu_i})$ (Inverse Gaussian) with $\nu_i = \mu_i^3 + 60\psi\mu_i$, asymmetric, in OPEF if $\psi = 0$.

2.5. Numerical experiment

In our numerical study of relative efficiency of the D-optimal designs for the $Emax$ model as defined by Eq(2.1), the nominal value of $\boldsymbol{\theta}$ is taken at $\boldsymbol{\theta}_0 = (340, -1, 4.6741, 60)'$ and the range of dose x in log scale is set at $[L, U] = [\log 10^{-3}, \log 500]$, as implemented by Bretz et al. (2010) and Hyun et al. (2018). For the $Emax$ model, $[\operatorname{var}_\xi \hat{\boldsymbol{\theta}}_{ML}]^{-1} = \int_L^U [E_\mu \dot{l}^2] \dot{\mu}\dot{\mu}' \xi(dx)$, where $\dot{\mu} = \left(\frac{1}{1+e^{\theta_2 x + \theta 3}}, \frac{-\theta_1 x e^{\theta_2 x + \theta 3}}{(1+e^{\theta_2 x + \theta 3})^2}, \frac{-\theta_1 e^{\theta_2 x + \theta 3}}{(1+e^{\theta_2 x + \theta 3})^2}, 1\right)'$ and $E_\mu \dot{l}^2$ is the Fisher information of μ, with respect to the governing probability distribution for y. As to obtain the MLE based D-optimal design, we need an explicit expression of $E_\mu \dot{l}^2$ for each specified distribution of y below parametrized with $\mu = Ey$ and $\nu = \nu(\mu) = \operatorname{var} y$.

Corollary 2.5.1. *With $y \sim Gaussian(\mu, \nu)$, $E_\mu \dot{l}^2 = \frac{1}{\nu} + \frac{\dot{\nu}^2}{2\nu^2}$, where $\dot{\nu} = \frac{d\nu}{d\mu}$.*

Corollary 2.5.2. *With* $y \sim Gamma(\frac{\mu^2}{\nu}, \frac{\mu}{\nu})$, $E_\mu \dot{l}^2 = \dot{\alpha}^2 \frac{\partial^2}{\partial \alpha^2} \log \Gamma(\alpha) + \frac{\alpha \dot{\beta}^2}{\beta^2} - \frac{2\dot{\alpha}\dot{\beta}}{\beta}$*, where* $\alpha \triangleq \frac{\mu^2}{\nu}$, $\beta \triangleq \frac{\mu}{\nu}$, $\dot{\alpha} \triangleq \frac{d\alpha}{d\mu} = \frac{2\mu}{\nu} - \frac{\mu^2 \dot{\nu}}{\nu^2}$*, and* $\dot{\beta} \triangleq \frac{d\beta}{d\mu} = \frac{1}{\nu} - \frac{\mu\dot{\nu}}{\nu^2}$.

Corollary 2.5.3. *With* $y \sim IG(\mu, \frac{\mu^3}{\nu})$, $E_\mu \dot{l}^2 = \frac{1}{\nu} + \frac{1}{2}(\frac{\dot{\nu}}{\nu} - \frac{3}{\mu})^2$.

As for the SLSE based D-optimal design, since $[\mathrm{var}_\xi \, \hat{\boldsymbol{\theta}}_{SLS}]^{-1} = \int_L^U \left[\frac{1}{\nu} + \frac{(\gamma - \dot{\nu}\nu)^2}{\nu(\kappa\nu - \nu^3 - \gamma^2)}\right] \dot{\mu}\dot{\mu}' \xi(dx)$, where $\gamma \triangleq E_\mu(y-\mu)^3$ and $\kappa \triangleq E_\mu(y-\mu)^4$, we need an expression for γ and κ in terms of μ, ν for each specified probability distribution of y. It turns out

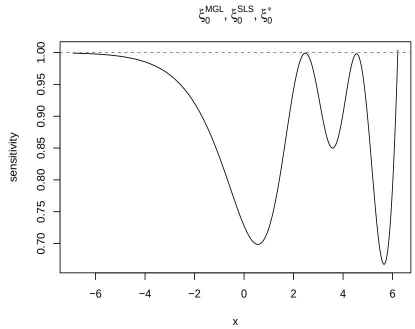
1. With $y \sim N(\mu, \nu)$, $\gamma = 0$ and $\kappa = 3\nu^2$;
2. With $y \sim Gamma(\frac{\mu^2}{\nu}, \frac{\mu}{\nu})$, $\gamma = \frac{2\nu^2}{\mu}$ and $\kappa = 3\nu^2 + \frac{6\nu^3}{\mu^2}$;
3. With $y \sim IG(\mu, \frac{\mu^3}{\nu})$, $\gamma = \frac{3\nu^2}{\mu}$, $\kappa = 3\nu^2 + \frac{15\nu^3}{\mu^2}$.

The D-optimal designs based on MLE, oracle SLSE, MqLE and MGLE of $\boldsymbol{\theta}$ for the $Emax$ model, denoted as ξ_0^*, ξ_0^{SLS}, ξ_0^{MqL} and ξ_0^{MGL}, respectively, are obtained in the case of the response measure governed by the following probability distribution:

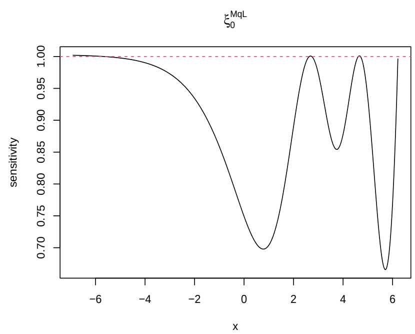
1. $N(\mu_i, \nu_i)$, with $\nu_i = 300 + 60\psi\mu_i$ and $\psi = 0, 1, 2, 3, 4$;
2. $Gamma(\frac{\mu_i^2}{\nu_i}, \frac{\mu_i}{\nu_i})$, with $\nu_i = \mu_i^2 + 60\psi\mu_i$ and $\psi = 0, 1, 2, 3, 4$;
3. $IG(\mu_i, \frac{\mu_i^3}{\nu_i})$, with $\nu_i = \mu_i^3 + 60\psi\mu_i$ and $\psi = 0, 1, 2, 3, 4$.

The obtained ξ_0^*, ξ_0^{SLS}, ξ_0^{MqL} and ξ_0^{MGL} are reported in Fig.2.1 - 2.3 with the controlling parameter $\psi = 4$ in each of the governing probability distribution as specified above. All the resulted D-optimal designs are verified by the equivalence plot, in which the x-axis is dose x (log scale), the y-axis is the scaled sensitivity, and all the doses as

Figure 2.1. Locally D-optimal designs at $\boldsymbol{\theta}_0$ under Gaussian$(\mu_i, 300 + 240\mu_i)$

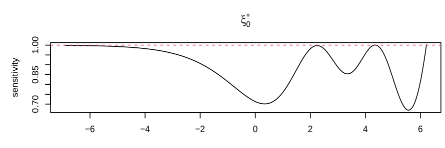


X	−6.91	2.48	4.54	6.21
W	0.25	0.25	0.25	0.25

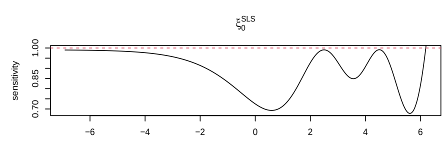


X	−6.91	2.69	4.66	6.21
W	0.25	0.25	0.25	0.25

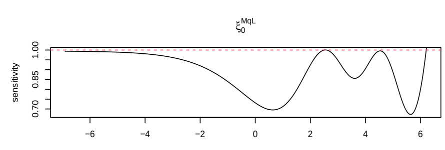
Figure 2.2. Locally D-optimal designs at $\boldsymbol{\theta}_0$ under Gamma$(\frac{\mu_i}{\mu_i+240}, \frac{1}{\mu_i+240})$



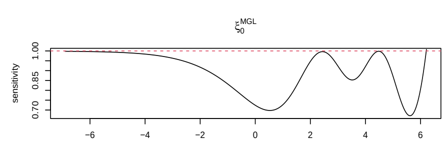
X	−6.91	2.25	4.35	6.21
W	0.25	0.25	0.25	0.25



X	−6.91	2.5	4.5	6.21
W	0.25	0.25	0.25	0.25

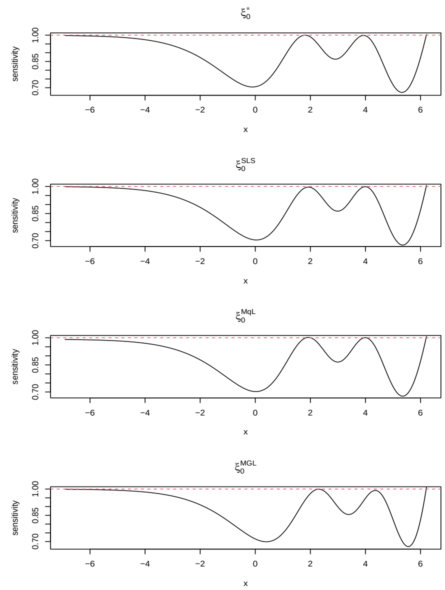


X	−6.91	2.54	4.53	6.21
W	0.25	0.25	0.25	0.25



X	−6.91	2.44	4.48	6.21
W	0.25	0.25	0.25	0.25

Figure 2.3. Locally D-optimal designs at $\boldsymbol{\theta}_0$ under IG$(\mu_i, \frac{\mu_i^2}{\mu_i^2+240})$



x	−6.91	1.8	3.94	6.21
w	0.25	0.25	0.25	0.25

x	−6.91	1.93	3.99	6.21
w	0.25	0.25	0.25	0.25

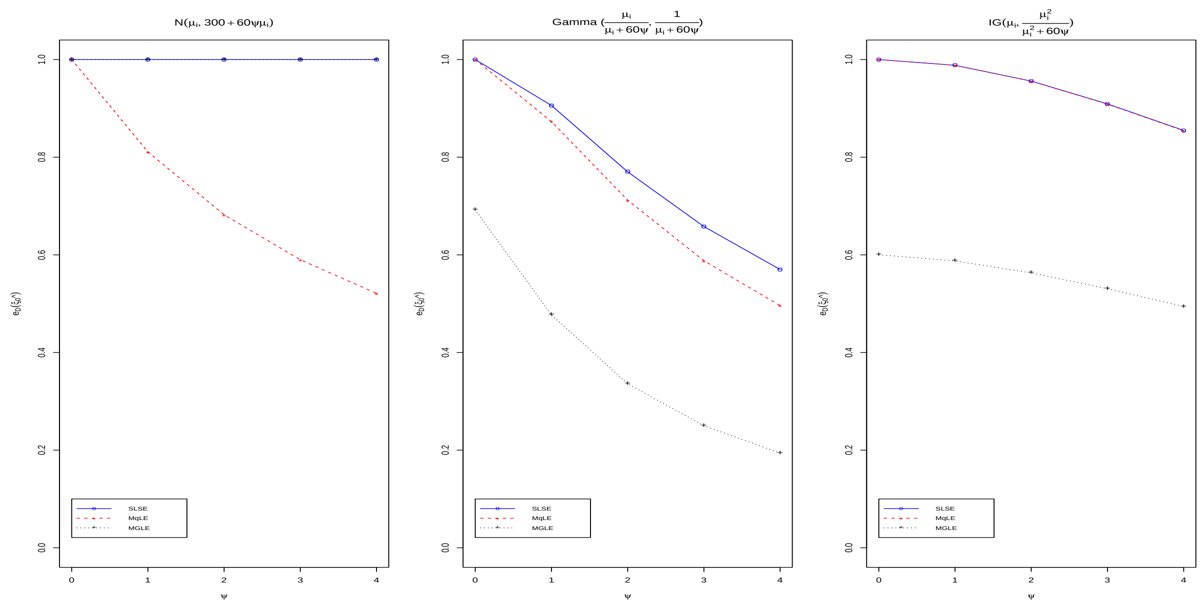
x	−6.91	1.93	3.99	6.21
w	0.25	0.25	0.25	0.25

x	−6.91	2.29	4.36	6.21
w	0.25	0.25	0.25	0.25

specified in the D-optimal design attain the maximum sensitivity. All the D-optimal designs are uniform. For the case of $N(\mu_i, \nu_i)$, $\xi_0^* = \xi_0^{SLS} = \xi_0^{MGL}$ except for ξ_0^{MqL}, because of $\text{var}_\xi\, \hat{\boldsymbol{\theta}}_{ML} = \text{var}_\xi\, \hat{\boldsymbol{\theta}}_{SLS} = \text{var}_\xi\, \hat{\boldsymbol{\theta}}_{MGL}$ for any design ξ in this case.

The relative efficiency curves of the resulted ξ_0^{SLS}, ξ_0^{MqL} and ξ_0^{MGL} (versus ξ_0^*) are plotted in Fig.2.4 with the controlling parameter $\psi = 0, 1, 2, 3, 4$ in each of the governing probability distribution as specified above. The x-axis of the efficiency plot is the ψ value and the y-axis is the relative D-efficiency $e_D(\xi_0^\wedge)$. The corresponding $e_D(\xi_0^\wedge)$ values are reported in Tab.2.1 for each case. At $\psi = 0$, all the governing distributions are in OPEF, thence $e_D(\xi_0^{SLS}) = e_D(\xi_0^{MqL}) = 1$. As the ψ value increases, all the $e_D(\xi_0^\wedge)$ value decreases. The left panel is for the case of $N(\mu_i, \nu_i)$, in which $e_D(\xi_0^{SLS}) = e_D(\xi_0^{MGL}) = 1$ at any ψ value due to $\text{var}_\xi\, \hat{\boldsymbol{\theta}}_{ML} = \text{var}_\xi\, \hat{\boldsymbol{\theta}}_{SLS} = \text{var}_\xi\, \hat{\boldsymbol{\theta}}_{MGL}$; while $e_D(\xi_0^{MqL})$ drops as ψ increases. The central and right panel are for the case of $Gamma(\frac{\mu_i^2}{\nu_i}, \frac{\mu_i}{\nu_i})$ and $IG(\mu_i, \frac{\mu_i^3}{\nu_i})$, respectively. Although

Figure 2.4. Relative D-efficiency curve of ξ_0^{SLS}, ξ_0^{MqL}, ξ_0^{MGL}



$e_D(\xi_0^{SLS})$ is as expected always higher than $e_D(\xi_0^{MqL})$, the two are quite comparable and significantly outperform $e_D(\xi_0^{MGL})$ in both cases.

Table 2.1. Relative D-efficiency of ξ_0^{SLS}, ξ_0^{MqL}, ξ_0^{MGL}

distribution	$\psi =$	0	1	2	3	4
	$e_D(\xi_0^{SLS})$	1.000	1.000	1.000	1.000	1.000
$N(\mu_i, 300 + 60\psi\mu_i)$	$e_D(\xi_0^{MqL})$	1.000	0.811	0.682	0.590	0.521
	$e_D(\xi_0^{MGL})$	1.000	1.000	1.000	1.000	1.000
	$e_D(\xi_0^{SLS})$	1.000	0.906	0.770	0.658	0.570
$Gamma(\frac{\mu_i}{\mu_i+60\psi}, \frac{1}{\mu_i+60\psi})$	$e_D(\xi_0^{MqL})$	1.000	0.873	0.712	0.588	0.496
	$e_D(\xi_0^{MGL})$	0.692	0.477	0.336	0.250	0.194
	$e_D(\xi_0^{SLS})$	1.000	0.988	0.956	0.909	0.855
$IG(\mu_i, \frac{\mu_i^2}{\mu_i^2+60\psi})$	$e_D(\xi_0^{MqL})$	1.000	0.988	0.956	0.909	0.854
	$e_D(\xi_0^{MGL})$	0.600	0.588	0.563	0.531	0.494

In our computation $e_D(\xi_0^{MGL})$, the actual asymptotic variance of $\hat{\boldsymbol{\theta}}_{MGL}$ with respect to the design ξ_0^{MGL} as quoted in Eq(2.8) is used. Indeed, the MGLE based D-optimal design provides an excellent example that a mis-specified underlying distribution of the response measure could lead to an inefficient MLE based D-optimal design.

3. D-OPTIMAL DESIGN SEARCHING ALGORITHMS

3.1. Review of D-optimal design searching algorithms

As discussed in §2, ξ_0, the locally D-optimal clinical design based on precision of an estimator $\hat{\boldsymbol{\theta}}$ is the one that minimizes the determinant of its asymptotic variance at the nominal value of model parameters $\boldsymbol{\theta}_0$, ie., $\xi_0 = \mathrm{argmin}_\xi \Psi(\xi, \boldsymbol{\theta}_0)$, where $\Psi(\xi, \boldsymbol{\theta}_0) \triangleq \det(\mathrm{var}(\hat{\boldsymbol{\theta}}|\xi, \boldsymbol{\theta}_0))$ is the criterion function, $\xi = \{(x_i, w_i)\}_{i=1}^d$ is an arbitrary clinical design satisfying $x_i \in [L, U]$, $d < +\infty$, $w_i \in (0, 1]$ and $\sum w_i = 1$. Evidently, on one hand, the estimability condition of the model parameter $\boldsymbol{\theta}$ requires the number of distinct dose of ξ_0, $d \geq k$, where k is dimension of the model parameter $\boldsymbol{\theta}$; while on the other hand, by Carathéodory's theorem, $d \leq k(k+1)/2$. Yang (2010) further provided a tighter upper bound of d in many cases including the Fisher-information based optimal designs for the $Emax$ model. Knowing the value of d being limited to a narrow range could help improve efficiency in a numerical search of the D-optimal design, because instead of examining innumerable candidate designs, one may work with a possibly much smaller complete class of designs which contains the D-optimal design, see Yang and Stufken (2012) or Kim (2017).

Motivated by the Equivalence theorem, Fedorov and Dubova (1968) and Fedorov (1972) introduced the V-algorithm based on the sensitivity function $d(x, \xi)$ to find D-optimal designs, a forerunner of many other D-optimal design searching algorithms in this stream. Yang et al. (2013) proposed their Yang-Biedermann-Tang (YBT) algorithm to find optimal designs of a broad class of optimality criteria and/or by multistage strategy. The YBT algorithm is considered as an extension of the Fedorov's algorithm by adding an optimization step for the design weight in the iterations. Hyun and Wong (2015) further modified the YBT algorithm by applying the V-Algorithm for a better set of initial doses in the iterative search of the optimal designs. Hyun et al. (2018) further developed the VNM R-package for implementation of the modified YBT algorithm for the multiple-objective optimal designs. As the surface of the criterion function $\Psi(\xi, \boldsymbol{\theta}_0)$ is pretty rugged in the case of the $Emax$ model, numerical search of the D-optimal designs by the algorithms in this stream often fails, if the initial design used in iterations is far away from the optimal design.

Kim (2017) applied the linear constraint optimization approach for numerical search of the D-optimal designs within the completed class of designs, which minimizes the criterion function $\Psi(\xi, \boldsymbol{\theta}_0)$ with respect to $\xi = \{(x_i, w_i)\}_{i=1}^d$ subjected to the box constraints $x_i \in [L, U]$, $w_i \in (0, 1]$, $i = 1, \cdots, d$, and the linear constraint $\sum_{i=1}^{d-1} w_i \leq 1$. Although this approach also suffers from an unwitty choice of initial design, it could be easily implemented in many computational software with some built-in function. In a view of the convexity of criterion function $\Psi(\xi, \boldsymbol{\theta}_0)$ in weight $\{w_i\}_{i=1}^d$ for any design ξ given dose $\{x_i\}_{i=1}^d$, Gao and Zhou (2017) and Wong and Zhou (2019) utilized the CVX programs in MATLAB (vide Grant and Boyd (2008) for details) for search of the optimal designs. Yeh and Zhou (2021) partitioned the space of dose $[L, U]$ with a mesh of tiny size and then applied the CVX program to the criterion function $\Psi(\xi, \boldsymbol{\theta}_0)$ with design ξ exhaustively taking all the mesh nodes as its target doses. They claimed the CVX program quite efficient in finding the optimal designs though evaluation of the asymptotic variance at all the mesh nodes for $\Psi(\xi, \boldsymbol{\theta}_0)$ is computationally expensive.

Inspired by Darwin's natural selection theory, Holland (1975) developed the genetic algorithms (GA), an iterative optimization procedure that repeatedly apply the evolving mechanism: recombination, mutation and selection to encode the solutions until the convergence criterion is satisfied. Heredia-Langner et al. (2003) demonstrated GA for finding nearly-optimal in highly constrained regions. Their method is possibly suitable for finding the exact D-optimal designs only, where the total size of subjects allowed in the clinical study is fixed.

Kennedy and Eberhart (1995) introduced particle swarm optimization (PSO) algorithm, another nature-inspired metaheuristic algorithm, which could simulate the graceful but unpredictable choreography of a bird flock or fish school to find optimum in parameter space for a user-specified criterion function. As in GA, PSO exploits a population of potential solutions to probe the search space. Although as a technique appearing not long time, PSO algorithm has received wide attentions and tons of literature have popped up in recent years on the topic of its implementation, enhancement and applications, vide Zhang et al. (2015) and Wang et al. (2017) for a review and overview of the PSO. Masuda et al. (2010) introduced a penalty approach to handle inequality constraints in PSO, in which a constrained optimization problem is transformed to an unconstrained problem by adding an additional penalty term in the criterion function. Kaur and Kaur (2015) discussed several mechanisms to deal with boundary constraints violation in PSO, which includes bringing the infeasible solution back into the feasible search space, clamping and re-initializing particles velocity in search of the solution.

Application of PSO variants to optimal design problem recently appears hot in litertature, see Qiu et al. (2014) and Shi et al. (2019) among many other research works. Chen et al. (2022) provided a comprehensive review of the works in this line. As discussed in the beginning of this section, in a numerical search of the optimal design, both types of the constraints: the box constraints on the dose x and the linear constraint on weight w of the design, need to be taken care of. While the box constraints could be easily handled beforehand by a preset search space of x, the linear constraint on w needs to be dealt with dynamically in the iterations. Although Qiu et al. (2014) tangentially mentioned of pulling the particles which move outside of the searching space back to the boundaries of the feasible space, unfortunately, both Qiu et al. (2014) and Shi et al. (2019) lack of implementation details on how the linear constraint $\sum w_i = 1$ is handled in particular. Motivated by Coello and Montes (2002), we apply a feasibility-based PSO approach here to find locally D-optimal designs. The pseudo code of our feasibility-based PSO algorithm for the optimal designs is given in next section.

3.2. Feasibility-based PSO algorithm for D-optimal designs

3.2.1. Vanilla PSO

For the minimization problem of finding $\xi_0 \in \mathcal{R}^d$ such that ξ_0 minimizes the criterion function $\Psi(\xi)$, ie., $\xi_0 = \operatorname{argmin}_{\xi \in \mathcal{R}^d} \Psi(\xi)$, the vanilla PSO algorithm introduced by Kennedy and Eberhart (1995) is implemented by iterations of the following procedure:

At iteration $(t+1)$, the movement of the k^{th} particle, $k = 1, \cdots, N$, is updated by:

$$\begin{cases} v_k^{(t+1)} = w^{(t)} \cdot v_k^{(t)} + c_1 \cdot u_{1,k}^{(t)} \odot (\zeta_k^{(t)} - \xi_k^{(t)}) + c_2 \cdot u_{2,k}^{(t)} \odot (\xi^{(t)} - \xi_k^{(t)}), \\ \xi_k^{(t+1)} = \xi_k^{(t)} + v_k^{(t+1)}. \end{cases}$$

where,

t: index of iterations ($t = 0, 1, \cdots$);

k: index of particles ($k = 1, 2, \cdots, N$); N: user-specified size of the swarm;

$w^{(t)} \in (0, 1)$ user-specified inertia weight at iteration t;

c_1, c_2: user-specified cognitive, social learning factor;

$\xi_k^{(t)}$, $v_k^{(t)}$: position and velocity of particle k at iteration t;

$u_{1,k}^{(t)}, u_{2,k}^{(t)} \overset{iid}{\sim} U[0, 1]$;

$\zeta_k^{(t)} = \underset{\xi_k^{(s)},\ 0 \leq s \leq t}{\operatorname{argmin}} \Psi(\xi_k^{(s)}, \boldsymbol{\theta}_0)$, $\xi^{(t)} = \underset{\zeta_k^{(t)},\ 0 \leq k \leq N}{\operatorname{argmin}} \Psi(\zeta_k^{(t)}, \boldsymbol{\theta}_0)$;

"$\odot$": component-wise product.

3.2.2. Feasiblility-based PSO

In the problem of finding a locally optimal design ξ_0 at a nominal value of the model parameter $\boldsymbol{\theta}_0$, let $\Psi(\xi, \boldsymbol{\theta}_0)$ be the optimality criterion function, then

$$\xi_0 = \operatorname{argmin}_{\xi \in \chi} \Psi(\xi, \boldsymbol{\theta}_0),$$

where $\xi \triangleq \{(x_i, w_i)\}_{i=1}^{d}$ is an arbitrary design of d doses, $d < +\infty$,
and $\chi \triangleq \{\xi | x_i \in [L, U], w_i \geq 0, \sum\limits_{i=1}^{d} w_i = 1\}$ is the design space of ξ.

Under the one-one mapping φ: $\varphi(\xi') = \xi$, where $\xi' \triangleq (x_1, \cdots, x_d, w_1, \cdots, w_{d-1})$, one has $\varphi(\chi') = \chi$, where $\chi' \triangleq \left\{\xi' | x_i \in [L, U], w_i \geq 0, \sum\limits_{i=1}^{d-1} w_i \leq 1\right\}$. Then the problem is equivalent to find $\xi'_0 = \operatorname{argmin}_{\xi' \in \chi'} \Psi(\varphi(\xi'), \boldsymbol{\theta}_0)$. For ease of notation, we still use ξ instead of ξ' in the pseudo code outlined below for our feasibility-based PSO algorithm for optimal designs. Note that the optimization problem is implicitly subjected to the box constraints, $x_i \in [L, U]$, $w_i \in [0, 1]$ and the linear constraint $\sum\limits_{i=1}^{d-1} w_i \leq 1$.

Step 1. At $t=0$, initialize $\xi_k^{(0)}$ and $v_k^{(0)} \in \chi'$.

Step 2. At iteration $t+1$, update

$$\begin{cases} v_k^{(t+1)} = w^{(t)} \cdot v_k^{(t)} + c_1 \cdot u_{1,k}^{(t)} \odot (\zeta_k^{(t)} - \xi_k^{(t)}) + c_2 \cdot u_{2,k}^{(t)} \odot (\xi^{(t)} - \xi_k^{(t)}), \\ \xi_k^{(t+1)} = \xi_k^{(t)} + v_k^{(t+1)}. \end{cases}$$

where $\zeta_k^{(t+1)} = \underset{\xi_k^{(s)},\ 0\le s\le t+1}{\arg\min} \Psi_{new}(\varphi(\xi_k^{(s)}), \boldsymbol{\theta}_0)$, $\xi^{(t+1)} = \underset{\zeta_k^{(t+1)},\ 0\le k\le N}{\arg\min} \Psi_{new}(\varphi(\zeta_k^{(t+1)}), \boldsymbol{\theta}_0)$, and

$$\Psi_{new}(\varphi(\xi_k^{(t+1)}), \boldsymbol{\theta}_0) = \begin{cases} \Psi(\varphi(\xi_k^{(t+1)}), \boldsymbol{\theta}_0), & \text{if } \xi_k^{(t+1)} \in \chi', \\ +\infty, & \text{otherwise.} \end{cases}$$

Step 3. Repeat Step 2 till $\|\xi^{(t+1)} - \xi^{(t)}\| < \epsilon$, where ϵ is the pre-specified error tolerance.

Step 4. Report $\xi_0 = \xi^{(t+1)}$ and $\Psi_{new}(\varphi(\xi_0), \boldsymbol{\theta}_0)$.

3.3. An application of feasibility-based PSO to locally optimal designs

As demonstrated in §2.5, the locally D-optimal designs for the $Emax$ model at nominal value $\boldsymbol{\theta}_0 = (340, -1, 4.6741, 60)'$ are all uniform with possibly different supports for different estimators. Applying the feasibility-based PSO, we work out here one more example of the regular-SLSE-based D-optimal design for the Michaelis-Menten model, where the response measure is modeled by the homoscedastic center+error model, i.e., $y_{ij} = \mu(x_i, \boldsymbol{\theta}) + \sigma\epsilon_{ij}$ with error $\epsilon_{ij} \overset{iid}{\sim} \text{WN}(0,1)$, and σ is the unknown dispersion parameter. The Michaelis-Menten model is a special case of the Emax model in Eq(2.1) with $\theta_4 = 0$, i.e., $\mu(x, \boldsymbol{\theta}) = \frac{\theta_1}{1+e^{\theta_2 x + \theta_3}}$. Clearly, the dimension of the model parameter $\boldsymbol{\theta} \triangleq (\theta_1, \theta_2, \theta_3)'$,

$k = 3$ in this case. The locally D-optimal design below is derived at nominal value $\boldsymbol{\theta}_0 = (340, -1, 4.6741)'$ with the range of support (for dose x), $[L, U] = [\log 10^{-3}, \log 500]$.

According to Wang and Leblanc (2008), the asymptotic precision of the regular SLSE $\tilde{\boldsymbol{\theta}}_{SLS}$ with respect to design ξ under this setting, is

$$[\text{var}_\xi\, \tilde{\boldsymbol{\theta}}_{SLS}]^{-1} = \frac{1}{(1-\tau)\sigma^2}\left\{ \int_L^U \dot{\mu}\dot{\mu}'\xi(dx) - \tau \int_L^U \dot{\mu}\xi(dx) \int_L^U \dot{\mu}'\xi(dx) \right\},$$

where $\dot{\mu} = \left(\frac{1}{1+e^{\theta_2 x+\theta 3}}, \frac{-\theta_1 x e^{\theta_2 x+\theta 3}}{(1+e^{\theta_2 x+\theta 3})^2}, \frac{-\theta_1 e^{\theta_2 x+\theta 3}}{(1+e^{\theta_2 x+\theta 3})^2}\right)'$, $\gamma = E_\mu(y-\mu)^3$, $\kappa = E_\mu(y-\mu)^4$, and $\tau = \frac{\gamma^2}{\sigma^2(\kappa-\sigma^4)}$ is the skewness index. Note that in this setting, σ^2, γ and κ are all universally constant across different levels of dose, and $\tau \in [0, 1)$, particularly, $\tau = 0$ for symmetric errors. The regular-SLSE-based D-optimality criterion for the Michaelis-Menten model is $\Psi(\xi, \boldsymbol{\theta}_0) = \log(\det(\text{var}_\xi\, \tilde{\boldsymbol{\theta}}_{SLS}))$. In a view that the regular-SLSE-based locally D-optimal design for Michaelis-Menten model depends on value of the skewness index τ, we exemplify the resulting optimal design with $\tau = 0.15$ and $\tau = 0.85$ here.

Referring to Theorem 2 in Yang (2010), we have the following result regarding the number of distinct doses x in the optimal design for this case, which helps confine our numerical search of the optimal design to the complete class of designs $\xi = \{(x_i, w_i)\}_{i=1}^4$.

Corollary 3.3.1. *The regular-SLSE-based locally D-optimal design for the Michaelis-Menten model has at most 4 support points (distinct doses).*

Note that a D-optimal design of the MM model with 3 distinct doses is minimally supported, and hence, by Silvey (1980), Lemma 5.1.3, must be uniform.

Following Kennedy and Eberhart (1995), we set the number of particles $N = 100$, the cognitive learning factor $c_1 = 2$ and the social learning factor $c_2 = 2$. Instead of adopting a linearly decreasing inertia weight $w^{(t)}$ at iteration t, we set $w^{(t)} = 0.4 + 0.5t^{-1}$ and the error tolerance $\epsilon = 10^{-6}$ in our feasibility-based PSO algorithm. To kick off, we initialize all the $v_k^{(0)}$ and $\xi_k^{(0)}$, $k = 1, \cdots, N$, with $\{(x_i, w_i)\}_{i=1}^4$, where $x_i \overset{iid}{\sim}$ Unif (L, U), $i = 1, 2, 3, 4$ and $(w_1, w_2, w_3, w_4) \sim$ Dirichlet$(1, 1, 1, 1)$. The algorithm stopping criterion is successfully met after about 120 iterations.

The resulting D-optimal design of both cases are presented in Table 3.1. In the case of $\tau = 0.15$, it is a uniform design with 3 doses, while in the case of $\tau = 0.85$, it is a non-uniform design with 4 doses. They are both verified by the equivalence plot, see Figure 3.1. The trace plot of the criterion function $\Psi(\xi^{(t)}, \boldsymbol{\theta}_0)$ by Figure 3.2 shows how fast the search converges to the optimum. The intermittent stagnation of the value of $\Psi(\xi^{(t)}, \boldsymbol{\theta}_0)$ is due to the nature of vanilla PSO and probably rejection of infeasible solutions $\xi^{(t)}$ in the iterations.

Table 3.1. Locally D-optimal designs for Michaelis-Menten model based on regular SLSE

τ	locally D-optimal design ξ_0
0.15	$\begin{pmatrix} 3.10 & 4.90 & 6.21 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$
0.85	$\begin{pmatrix} -6.91 & 3.10 & 4.90 & 6.21 \\ 0.13 & 0.29 & 0.29 & 0.29 \end{pmatrix}$

Figure 3.1. Equivalence plot for regular-SLSE-based locally D-optimal designs

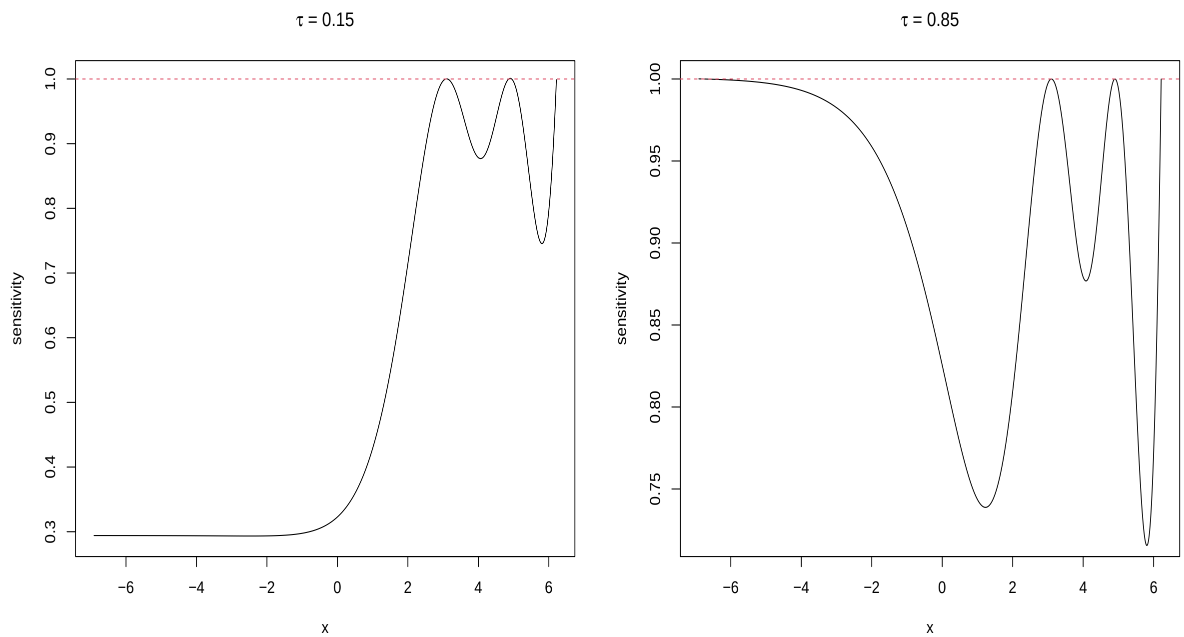
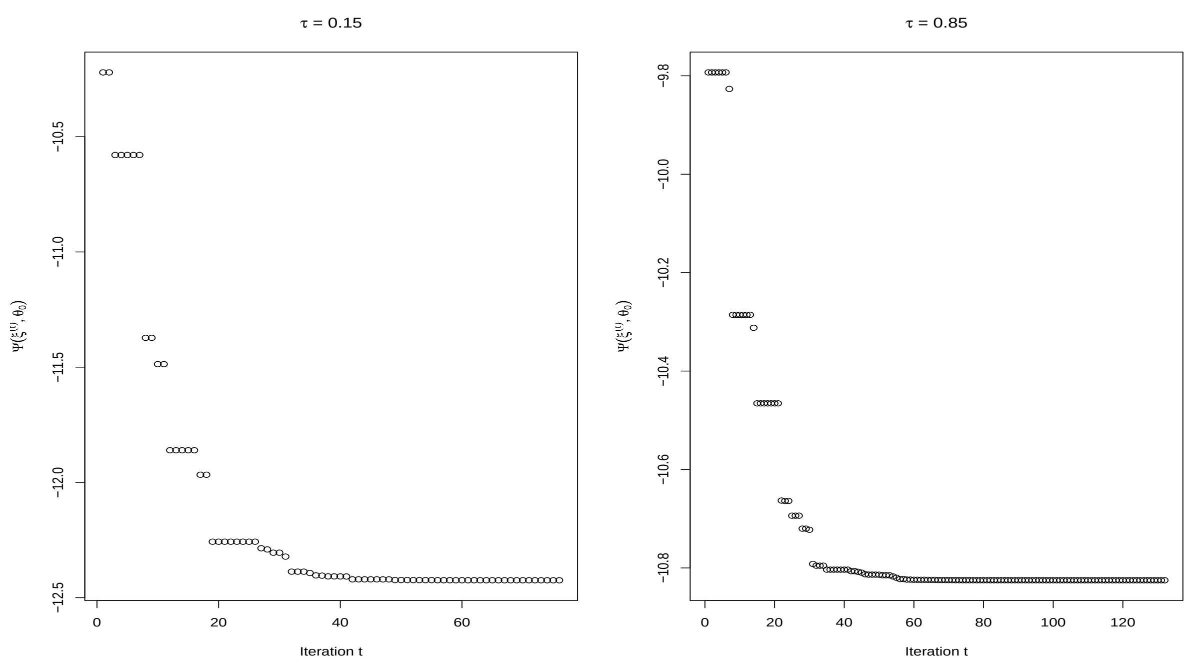


Figure 3.2. Trace plot of criterion function $\Psi(\xi^{(t)}, \boldsymbol{\theta}_0)$



4. CONCLUSION & DISCUSSION

The Fisher information or MLE based optimal designs require full knowledge of the governing probability distribution for the response measure which, nonetheless, is often absent or mis-specified in practice. Our numerical experiment in §2.5 exemplifies that the MGLE based D-optimal design under the mis-specified Gaussianity could be very inefficient as compared with the oracle SLSE or MqLE based D-optimal design, yet both the mean and variance of the Gaussian distribution are correctly structured. The MqLE based D-optimal design, on the other hand, requires only partial knowledge of the probability distribution: the structure of mean and variance. It demonstrates a fairly comparable efficiency performance to our novel oracle-SLSE based D-optimal design, but the latter relies on additional knowledge of the structure of skewness and kurtosis of the distribution. In literature, heteroscedasticity is typically modeled by $\operatorname{var} y_{ij} \triangleq \nu_i = \sigma^2 v(\mu_i)$, where σ^2 is the dispersion parameter of no interest, and $v(\cdot) > 0$ is some positive function of μ. This covers the well-known Tweedie distribution family with $v(\mu) = \mu^p$, which itself embraces the distribution of $N(\mu, \sigma^2)$ when $p = 0$, $quasi\text{-}Poisson(\mu)$ when $p = 1$, $Gamma(\sigma^{-2}, \sigma^{-2}\mu^{-1})$ when $p = 2$, and $IG(\mu, \sigma^{-2})$ when $p = 3$. Our consideration of the three probability models for the response measure is thus motivated in this study. Indeed, including σ^2 in $\operatorname{var} y_{ij}$ would not change the MqLE based D-optimal design; however, it would make the SLSE based D-optimal design depend on the nuisance parameter σ^2, i.e., even with the same nominal value of the model parameter $\boldsymbol{\theta}$, a different σ^2 value would result in a different SLSE based D-optimal design. Instead, for a fair comparison of both we adopt the one-parameter model $\nu_i = \nu(\mu_i)$ for heteroscedasticity in this study to get rid of the impact of the undesirable σ^2 on the oracle SLSE based optimal design.

On the other hand, under the homoscedastic center+error model with unknown dispersion parameter σ^2 for the response measure, the regular SLSE based locally D-optimal design relies only on the skewness index τ which is fully determined by the ancillarity of error, instead of magnitude of σ^2. We apply our feasibility-based PSO algorithm proposed in this work to demonstrate numeric search of the locally D-optimal designs based on the regular SLSE. The resulting D-optimal designs are also verified by the Equivalence theorem.

Although this study is focused on the D-optimal designs in the case of the $Emax$ model with respect to the three underlying probability distributions for the response measure, our work can be extended to other optimality criteria and other scenarios.

REFERENCES

Atkinson, A. C. and Cook, R. D. (1995). D-optimum designs for heteroscedastic linear models. *Journal of the American Statistical Association*, 90(429):204–212.

Bretz, F., Dette, H., and Pinheiro, J. C. (2010). Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*, 29(7):731–742.

Chen, P. Y., Chen, R. B., and Wong, W. K. (2022). Particle swarm optimization for searching efficient experimental designs: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(5):e1578.

Coello, C. A. C. and Montes, E. M. (2002). Constraint-handling in genetic algorithms through the use of dominance-based tournament selection. *Advanced Engineering Informatics*, 16(3):193–203.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1):1–10.

Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic Press.

Fedorov, V. V. and Dubova, I. (1968). Methods for constructing optimal designs in regression experiments. (4).

Gao, L. L. and Zhou, J. (2014). New optimal design criteria for regression models with asymmetric errors. *Journal of Statistical Planning and Inference*, 149:140–151.

Gao, L. L. and Zhou, J. (2017). D-optimal designs based on the second-order least squares estimator. *Statistical Papers*, 58(1):77–94.

Gaviria, J. A. and & López-Ríos, V. I. (2014). Locally d-optimal designs with heteroscedasticity: A comparison between two methodologies. *Statistical Papers*, 58(1):77–94.

Grant, M. and Boyd, S. (2008). Graph implementations for nonsmooth convex programs. In Blondel, V., Boyd, S., and Kimura, H., editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited. http://stanford.edu/~boyd/graph_dcp.html.

Heredia-Langner, A., Carlyle, W. M., Montgomery, D., Borror, C. M., and Runger, G. (2003). Genetic algorithms for the construction of d-optimal designs. *Journal of Quality Technology*, 35(1):28–46.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. MIT Press.

Hyun, S. W. and Wong, W. K. (2015). Multiple objective optimal designs to study the interesting features in a dose-response relationship. *International Journal of Biostatistics*, 11(2):253–271.

Hyun, S. W., Wong, W. K., and Yang, Y. (2018). Vnm: An r package for finding multiple-objective optimal designs for the 4-parameter logistic model. *Journal of Statistical Software*, 83(5):1–19.

Kaur, A. and Kaur, M. (2015). Dealing with boundary constraint violations in particle swarm optimization with aging leader and challengers (alc-pso). *International Journal of Computer Applications*, 121(11):13–19.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, pages 1942–1948, Perth, WA, Australia.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30:271–294.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Kim, S. (2017). *Optimal Experimental Designs for Mixed Categorical and Continuous Responses*. PhD thesis, Arizona State University. https://keep.lib.asu.edu/_flysystem/fedora/c7/184344/Kim_asu_0010E_17267.pdf.

Li, G. and Majumdar, D. (2008). D-optimal designs for logistic models with three and four parameters. *Journal of Statistical Planning and Inference*, 138(7):1950–1959.

Macdougall, J. (2006). *Analysis of Dose–Response Studies—Emax Model*, chapter 9, pages 127–145. Springer New York.

Masuda, K., Kurihara, K., and Aiyoshi, E. (2010). A penalty approach to handle inequality constraints in particle swarm optimization. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 2520–2525, Istanbul, Turkey.

Michaelis, L. and Menten, M. M. L. (1913). Die kinetik der invertinwirkung. *Biochemische Zeitschrift*, 49(37):333–369.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384.

Pukelsheim, F. (2006). *Optimal design of experiments*. Society for Industrial and Applied Mathematics (SIAM).

Qiu, J., Chen, R.-B., Wang, W., and Wong, W. K. (2014). Using animal instincts to design efficient biomedical studies via particle swarm optimization. *Swarm and Evolutionary Computation*, 18:1–10.

Shen, G., Hyun, S. W., and Wong, W. K. (2016). Optimal designs based on the maximum quasi-likelihood estimator. *Journal of Statistical Planning and Inference*, 178:128–139.

Shi, Y., Zhang, Z., and Wong, W. K. (2019). Particle swarm based algorithms for finding locally and bayesian d-optimal designs. *Journal of Statistical Distributions and Applications*, 6(1):1–17.

Silvey, S. D. (1980). *Optimal Design*. Springer.

Wang, D., Tan, D., and Liu, L. (2017). Particle swarm optimization algorithm: an overview. *Soft Computing*, (22):387–408.

Wang, L. and Leblanc, A. (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics*, 60(4):883–900.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61(3):439–447.

Wong, W. K. and Zhou, J. (2019). Cvx-based algorithms for constructing various optimal regression designs. *Canadian Journal of Statistics*, 47(3):374–391.

Yang, M. (2010). On the de la garza phenomenon. *The Annals of Statistics*, 38(4):2499–2524.

Yang, M., Biedermann, S., and Tang, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association*, 108(504):1411–1420.

Yang, M. and Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics*, 40(3):1665–1681.

Yeh, C. K. and Zhou, J. (2021). Properties of optimal regression designs under the second-order least squares estimator. *Statistical Papers*, 62:75–92.

Yin, Y. and Zhou, J. (2017). Optimal designs for regression models using the second-order least squares estimator. *Statistica Sinica*, 27(4):1841–1856.

Zhang, Y., Wang, S., and Ji, G. (2015). A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering*, pages 1–38.

APPENDIX A. TECHNICAL DETAILS

1. Proof of Proposition 2.4.1 for the asymptotic variance of oracle SLSE, $\mathrm{var}_\xi\, \hat{\boldsymbol{\theta}}_{SLS}$.

Proof. For $\hat{\boldsymbol{\theta}}_{SLS}$, the kernel of criterion function in Eq(2.5) is $g(y_{ij}, x_i, \boldsymbol{\theta}) = \rho'_{ij}(\boldsymbol{\theta}) W_0(x_i) \rho_{ij}(\boldsymbol{\theta})$, where $\rho_{ij}(\boldsymbol{\theta}) \triangleq (y_{ij} - \mu_i, y_{ij}^2 - \mu_i^2 - \nu_i)'$ and $W_0(x_i) = \begin{pmatrix} \tilde{\nu}_i & \tilde{\gamma}_i + 2\tilde{\mu}_i\tilde{\nu}_i \\ \tilde{\gamma}_i + 2\tilde{\mu}_i\tilde{\nu}_i & \tilde{\kappa}_i - \tilde{\nu}_i^2 + 4\tilde{\mu}_i^2\tilde{\nu}_i + 4\tilde{\mu}_i\tilde{\gamma}_i \end{pmatrix}^{-1}$. As demonstrated by Wang and Leblanc (2008),

$$
\begin{aligned}
\dot{g}(y_{ij}, x_i, \boldsymbol{\theta}) &= 2\Big\{ \frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \rho_{ij}(\boldsymbol{\theta}) \Big\}, \\
\ddot{g}(y_{ij}, x_i, \boldsymbol{\theta}) &= 2\Big\{ \frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} + (\rho'_{ij}(\boldsymbol{\theta}) W_0(x_i) \otimes I_4) \frac{\partial}{\partial \boldsymbol{\theta}'} [\mathrm{vec}(\frac{\partial \rho'_{ij}}{\partial \boldsymbol{\theta}})] \Big\},
\end{aligned}
$$

where $\otimes$ is the Kronecker product, vec is the vec operator and I_4 is the identity matrix of dimension 4. At $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, note $E_{\boldsymbol{\theta}}[\rho_{ij}(\boldsymbol{\theta})] = \mathbf{0}$ and $E_{\boldsymbol{\theta}}[\rho_{ij}(\boldsymbol{\theta}) \rho'_{ij}(\boldsymbol{\theta})] = W_0^{-1}(x_i)$, then

$$
\begin{aligned}
E_{\boldsymbol{\theta}}[\ddot{g}(y_{ij}, x_i, \boldsymbol{\theta})] &= 2\big[\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \big], \\
E_{\boldsymbol{\theta}}[\dot{g}(y_{ij}, x_i, \boldsymbol{\theta}) \dot{g}'(y_{ij}, x_i, \boldsymbol{\theta})] &= 4E_{\boldsymbol{\theta}}\big[\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \rho_{ij}(\boldsymbol{\theta}) \rho'_{ij}(\boldsymbol{\theta}) W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \big] \\
&= 4\big[\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) W_0^{-1}(x_i) W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \big] \\
&= 4\big[\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \big].
\end{aligned}
$$

Observe $\frac{\partial \rho'_{ij}}{\partial \boldsymbol{\theta}} = \frac{\partial \rho'_{ij}}{\partial \mu_i} \otimes \dot{\mu}_i$, then at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$
\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = (\frac{\partial \rho'_{ij}}{\partial \mu_i} \otimes \dot{\mu}_i)(W_0(x_i) \otimes 1)(\frac{\partial \rho_{ij}}{\partial \mu_i} \otimes \dot{\mu}'_i) = \big[\frac{\partial \rho'_{ij}}{\partial \mu_i} W_0(x_i) \frac{\partial \rho_{ij}}{\partial \mu_i} \big] \dot{\mu}_i \dot{\mu}'_i.
$$

Plug in $\frac{\partial \rho'_{ij}}{\partial \mu_i} = (-1, -2\mu_i - \dot{\nu}_i)$ and $W_0(x_i) = \begin{pmatrix} \nu_i & \gamma_i+2\mu_i\nu_i \\ \gamma_i+2\mu_i\nu_i & \kappa_i-\nu_i^2+4\mu_i^2\nu_i+4\mu_i\gamma_i \end{pmatrix}^{-1}$, then

$$\frac{\partial \rho'_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} W_0(x_i) \frac{\partial \rho_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = [\frac{1}{\nu_i} + \frac{(\gamma_i - \dot{\nu}_i\nu_i)^2}{\nu_i(\kappa_i\nu_i - \nu_i^3 - \gamma_i^2)}]\dot{\mu}_i\dot{\mu}'_i.$$

Applying the design weights and then using the sandwich form of the asymptotic variance of M-estimators in general, it follows $\mathrm{var}_\xi\, \hat{\boldsymbol{\theta}}_{SLS} = \big\{ \int_L^U \big[\frac{1}{\nu} + \frac{(\gamma-\dot{\nu}\nu)^2}{\nu(\kappa\nu-\nu^3-\gamma^2)}\big]\dot{\mu}\dot{\mu}'\xi(dx)\big\}^{-1}$.

For the claim of superiority of the oracle SLSE to MqLE, observe that

$\kappa\nu - \nu^3 - \gamma^2 = \nu(\kappa - \nu^2) - \gamma^2 = E_\mu(y-\mu)^2 E_\mu[(y-\mu)^2 - \nu]^2 - \{E_\mu(y-\mu)^3\}^2$

$\geq E_\mu\{(y-\mu)[(y-\mu)^2 - \nu]\}^2 - \{E_\mu(y-\mu)^3\}^2 = \{E_\mu(y-\mu)^3\}^2 - \{E_\mu(y-\mu)^3\}^2 = 0.$

It follows $\mathrm{var}_\xi\, \hat{\boldsymbol{\theta}}_{SLS} = \big\{ \int_L^U \big[\frac{1}{\nu} + \frac{(\gamma-\dot{\nu}\nu)^2}{\nu(\kappa\nu-\nu^3-\gamma^2)}\big]\dot{\mu}\dot{\mu}'\xi(dx)\big\}^{-1} \preceq \big\{ \int_L^U \frac{1}{\nu}\dot{\mu}\dot{\mu}'\xi(dx)\big\}^{-1} = \mathrm{var}_\xi\, \hat{\boldsymbol{\theta}}_{MqL}$.

□

2. Proof of Eq(2.8), the asymptotic variance of MGLE, $\mathrm{var}_\xi(\hat{\boldsymbol{\theta}}_{MGL})$.

Proof. For $\hat{\boldsymbol{\theta}}_{MGL}$, the kernel of criterion function in Eq(2.5), $g(y_{ij}, x_i, \boldsymbol{\theta}) = \log\phi(y_{ij}, \mu_i, \nu_i)$, where $\mu_i = \mu(x_i, \boldsymbol{\theta})$, $\nu_i = \nu(\mu_i)$. Let $\dot{\mu}_i = \frac{\partial \mu_i}{\partial \boldsymbol{\theta}}$, $\ddot{\mu}_i = \frac{\partial^2 \mu_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$, $\dot{\nu}_i = \frac{d\nu_i}{d\mu_i}$ and $\ddot{\nu}_i = \frac{d^2\nu_i}{d\mu_i^2}$. Note $\log\phi(y_{ij}, \mu_i, \nu_i) = -\frac{1}{2}\log(2\pi\nu_i) - \frac{(y_{ij}-\mu_i)^2}{2\nu_i}$, then

$$\begin{aligned}
\dot{g}(y_{ij}, x_i, \boldsymbol{\theta}) &= -\frac{1}{2}\left[\frac{\dot{\nu}_i}{\nu_i} + \frac{2(\mu_i - y_{ij})}{\nu_i} - \frac{(y_{ij}-\mu_i)^2\dot{\nu}_i}{\nu_i^2}\right]\dot{\mu}_i, \\
\ddot{g}(y_{ij}, x_i, \boldsymbol{\theta}) &= -\frac{1}{2}\left[\frac{\ddot{\nu}_i}{\nu_i} - \frac{\dot{\nu}_i^2}{\nu_i^2} + \frac{2}{\nu_i} + \frac{4(y_{ij}-\mu_i)\dot{\nu}_i}{\nu_i^2} - \frac{(y_{ij}-\mu_i)^2\ddot{\nu}_i}{\nu_i^2} + \frac{2(y_{ij}-\mu_i)^2\dot{\nu}_i^2}{\nu^3}\right]\dot{\mu}\dot{\mu}' \\
&\quad -\frac{1}{2}\left[\frac{\dot{\nu}_i}{\nu_i} - \frac{2(y_{ij}-\mu_i)}{\nu_i} - \frac{(y_{ij}-\mu_i)^2\dot{\nu}_i}{\nu_i^2}\right]\ddot{\mu}_i.
\end{aligned}$$

It follows

$$E_{\boldsymbol{\theta}}\big[-\ddot{g}(y_{ij},x_i,\boldsymbol{\theta})\big] = (\frac{1}{\nu_i}+\frac{\dot{\nu}_i^2}{2\nu_i^2})\dot{\mu}_i\dot{\mu}_i',$$
$$E_{\boldsymbol{\theta}}\big[\dot{g}(y_{ij},x_i,\boldsymbol{\theta})\dot{g}'(y_{ij},x_i,\boldsymbol{\theta})\big] = (\frac{1}{\nu_i}-\frac{\dot{\nu}_i^2}{4\nu_i^2}+\frac{\dot{\nu}_i\gamma_i}{\nu_i^3}+\frac{\dot{\nu}_i^2\kappa_i}{4\nu_i^4})\dot{\mu}_i\dot{\mu}_i'.$$

Applying the design weights and then using the sandwich form of the asymptotic variance of M-estimators in general, Eq(2.8) holds. □

3. Proof of Corollary 2.5.2, the Fisher information of μ w.r.t $Gamma(\frac{\mu^2}{\nu},\frac{\mu}{\nu})$ distribution.

Proof. With $y\sim Gamma(\alpha,\beta)$, where $\alpha\triangleq\frac{\mu^2}{\nu}$ is the shape and $\beta\triangleq\frac{\mu}{\nu}$ is the scale of the distribution, the log-likelihood of μ is $l(\mu)=\alpha\log\beta+(\alpha-1)\log y-\log\Gamma(\alpha)-\beta y$.

Let $\dot{l}=\frac{dl}{d\mu}$, $\ddot{l}=\frac{d^2l}{d\mu^2}$, $\dot{\alpha}=\frac{d\alpha}{d\mu}$ and $\dot{\beta}=\frac{d\beta}{d\mu}$, then

$$\dot{l} = \dot{\alpha}\big[\log y+\log\beta-\frac{d}{d\alpha}\log\Gamma(\alpha)\big]-\dot{\beta}(y-\frac{\alpha}{\beta}),$$
$$\ddot{l} = \ddot{\alpha}\big[\log y+\log\beta-\frac{d}{\partial\alpha}\log\Gamma(\alpha)\big]-\ddot{\beta}(y-\frac{\alpha}{\beta})+\frac{2\dot{\alpha}\dot{\beta}}{\beta}-\frac{\alpha\dot{\beta}^2}{\beta^2}-\dot{\alpha}^2\big[\frac{d^2}{d\alpha^2}\log\Gamma(\alpha)\big],$$

where $\ddot{\alpha}=\frac{d^2\alpha}{d\mu^2}$ and $\ddot{\beta}=\frac{d^2\beta}{d\mu^2}$. Note $\forall\mu$, $E_\mu\dot{l}$ =0 implies $\dot{\alpha}E_\mu\big[\log y+\log\beta-\frac{d}{d\alpha}\log\Gamma(\alpha)\big]=0$, then $\ddot{\alpha}E_\mu\big[\log y+\log\beta-\frac{d}{d\alpha}\log\Gamma(\alpha)\big]=0$. So $E_\mu\dot{l}^2=-E_\mu\ddot{l}=\dot{\alpha}^2\big[\frac{d^2}{d\alpha^2}\log\Gamma(\alpha)\big]+\frac{\alpha\dot{\beta}^2}{\beta^2}-\frac{2\dot{\alpha}\dot{\beta}}{\beta}$. $\dot{\alpha}=\frac{2\mu}{\nu}-\frac{\mu^2\dot{\nu}}{\nu^2}$ and $\dot{\beta}=\frac{1}{\nu}-\frac{\mu\dot{\nu}}{\nu^2}$ result from the definition of α and β as given above. □

4. Proof of Corollary 2.5.3, the Fisher information of μ w.r.t $IG(\mu,\frac{\mu^3}{\nu})$ distribution.

Proof. With $y\sim IG(\mu,\lambda)$, where μ is the mean and $\lambda=\frac{\mu^3}{\nu}$ is the shape of the distribution, the log-likelihood of μ is $l(\mu)=\frac{3}{2}\log\mu-\frac{1}{2}\log\nu-\frac{1}{2}\log 2\pi y^3-\frac{\mu(y-\mu)^2}{2\nu y}$. Let $\dot{l}=\frac{dl}{d\mu}$, $\ddot{l}=\frac{d^2l}{d\mu^2}$, then

$$
\begin{aligned}
\dot{l} &= \frac{3}{2\mu} - \frac{\dot{\nu}}{2\nu} - \frac{(y-\mu)^2}{2y\nu} + \frac{\mu(y-\mu)}{y\nu} + \frac{\mu\dot{\nu}(y-\mu)^2}{2y\nu^2}, \\
\ddot{l} &= (-\frac{3}{2\mu^2} - \frac{\ddot{\nu}}{2\nu} + \frac{\dot{\nu}^2}{2\nu^2} - \frac{\mu^2\ddot{\nu}}{\nu^2} + \frac{2\mu^2\dot{\nu}^2}{\nu^3} + \frac{2}{\nu} - \frac{4\mu\dot{\nu}}{\nu^2}) \\
&\quad +(\frac{\dot{\nu}}{\nu^2} + \frac{\mu\ddot{\nu}}{2\nu^2} - \frac{\mu\dot{\nu}^2}{\nu^3})y + (\frac{3\mu^2\dot{\nu}}{\nu^2} + \frac{\mu^3\ddot{\nu}}{2\nu^2} - \frac{\mu^3\dot{\nu}^2}{\nu^3} - \frac{3\mu}{\nu})y^{-1}.
\end{aligned}
$$

Note $Ey = \mu$ and $E(y^{-1}) = \frac{1}{\mu} + \frac{\nu}{\mu^3}$, then $E_\mu \dot{l}^2 = -E_\mu \ddot{l} = \frac{1}{\nu} + \frac{1}{2}(\frac{\dot{\nu}}{\nu} - \frac{3}{\mu})^2$. □

5. Proof of Corollary 3.3.1, the upper-bound of number of support points in the regular-SLSE-based D-optimal designs for the Michaelis-Menten model.

Proof. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$, $\mu(x, \boldsymbol{\theta}) = \frac{\theta_1}{1+e^{\theta_2 x+\theta_3}}$ and $c = e^{\theta_2 x+\theta_3}$, then

$$
\begin{aligned}
\dot{\mu} &\triangleq \frac{\partial\mu}{\partial\boldsymbol{\theta}} = (\frac{1}{1+e^{\theta_2 x+\theta_3}}, \frac{-\theta_1 x e^{\theta_2 x+\theta_3}}{(1+e^{\theta_2 x+\theta_3})^2}, \frac{-\theta_1 e^{\theta_2 x+\theta_3}}{(1+e^{\theta_2 x+\theta_3})^2})' \\
&= (\frac{1}{1+c}, -\frac{\theta_1}{\theta_2}\frac{c(\log c - \theta_3)}{(1+c)^2}, \frac{-\theta_1 c}{(1+c)^2})'.
\end{aligned}
$$

The regular-SLSE-based locally D-optimal design equivalently maximizes $\frac{\det(E_\xi I_{\boldsymbol{\theta}})}{(1-t)\sigma^2}$, where

$$
I_{\boldsymbol{\theta}} = \begin{pmatrix} 1 & \sqrt{t}\dot{\mu}' \\ \sqrt{t}\dot{\mu} & \dot{\mu}\dot{\mu}' \end{pmatrix} = \begin{pmatrix}
1 & \frac{\sqrt{t}}{1+c} & -\frac{\sqrt{t}\theta_1}{\theta_2}\frac{c(\log c-\theta_3)}{(1+c)^2} & -\frac{\sqrt{t}\theta_2 c}{(1+c)^2} \\
\frac{\sqrt{t}}{1+c} & \frac{1}{(1+c)^2} & -\frac{\theta_1}{\theta_2}\frac{c(\log c-\theta_3)}{(1+c)^3} & -\frac{\theta_2 c}{(1+c)^3} \\
-\frac{\sqrt{t}\theta_1}{\theta_2}\frac{c(\log c-\theta_3)}{(1+c)^2} & -\frac{\theta_1}{\theta_2}\frac{c(\log c-\theta_3)}{(1+c)^3} & \frac{\theta_1^2}{\theta_2^2}\frac{c^2(\log c-\theta_3)^2}{(1+c)^4} & \frac{\theta_1^2}{\theta_2}\frac{c^2(\log c-\theta_3)}{(1+c)^4} \\
-\frac{\sqrt{t}\theta_1 c}{(1+c)^2} & -\frac{\theta_1 c}{(1+c)^3} & \frac{\theta_1^2}{\theta_2}\frac{c^2(\log c-\theta_3)}{(1+c)^4} & \frac{\theta_1^2 c^2}{(1+c)^4}
\end{pmatrix}.
$$

It turns out $I_{\boldsymbol{\theta}} = P(\boldsymbol{\theta})C(\boldsymbol{\theta}, c)P'(\boldsymbol{\theta})$, where

$$P(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{\theta_1\theta_3}{\theta_2} & -\frac{\theta_1}{\theta_2} & -\frac{\theta_1\theta_3}{\theta_2} \\ 0 & -\theta_1 & 0 & \theta_1 \end{pmatrix},$$

and

$$C(\boldsymbol{\theta}, c) = \begin{pmatrix} 1 & \frac{\sqrt{t}}{1+c} & \frac{\sqrt{t}c\log c}{(1+c)^2} & \frac{\sqrt{t}}{(1+c)^2} \\ \frac{\sqrt{t}}{1+c} & \frac{1}{(1+c)^2} & \frac{c\log c}{(1+c)^3} & \frac{1}{(1+c)^3} \\ \frac{\sqrt{t}c\log c}{(1+c)^2} & \frac{c\log c}{(1+c)^3} & \frac{c^2 log^2 c}{(1+c)^4} & \frac{c\log c}{(1+c)^4} \\ \frac{\sqrt{t}}{(1+c)^2} & \frac{1}{(1+c)^3} & \frac{c\log c}{(1+c)^4} & \frac{1}{(1+c)^4} \end{pmatrix}.$$

Note $C(\boldsymbol{\theta}, c)$ can be transformed into the matrix $(\Psi_{ij})_{i,j=1}^4$ in Yang (2010), Eq(2.2) by using the $k = 8$ functions, namely, $\Psi_1(c) \triangleq \frac{1}{(1+c)^4}$, $\Psi_2(c) \triangleq \frac{1}{(1+c)^3}$, $\Psi_3(c) \triangleq \frac{c\log c}{(1+c)^4}$, $\Psi_4(c) \triangleq \frac{1}{(1+c)^2}$, $\Psi_5(c) \triangleq \frac{c\log c}{(1+c)^3}$, $\Psi_6(c) \triangleq \frac{1}{1+c}$, $\Psi_7(c) \triangleq \frac{c\log c}{(1+c)^2}$, and $\Psi_8(c) \triangleq \frac{c^2\log^2 c}{(1+c)^4}$. The $f_{l,t}$ functions in Eq(3.2) thereof may be tediously achieved using Eq(3.1) and the $\Psi(\cdot)$ functions above, which results in $f_{1,1} = -\frac{4}{(1+c)^5}$, $f_{2,2} = \frac{3}{4}$, $f_{3,3} = \frac{3c+1}{3c^2}$, $f_{4,4} = \frac{4c(3c+2)}{(3c+1)^2}$, $f_{5,5} = \frac{9c^3+15c^2+7c+1}{c^2(3c+2)^2}$, $f_{6,6} = \frac{9c(3c+2)}{9c^2+6c+1}$, $f_{7,7} = \frac{3c+1}{3c^2}$, $f_{8,8} = \frac{2}{3c^2}$. Therefore, $F(c) \triangleq \prod_{l=1}^{8} f_{l,l}(c) = -\frac{8(9c^3+15c^2+7c+1)}{(1+c)^5c^6(9c^2+6c+1)} < 0$, for $c > 0$. By Yang (2010), Theorem 2, case (d), the upper-bound of number of supports in the optimal design is $k/2 = 4$. □