Computing T-optimal Designs via Nested Semi-Infinite Programming and Twofold Adaptive Discretization

David Mogalle[1], *Philipp Seufert*[1], *Jan Schwientek*[1], *Michael Bortz*[1], *Karl-Heinz Küfer*[1]

[1]*Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany*

Model Discrimination

Setting

We want to describe a phenomenon by a model $f(x, \theta)$ with *design variables* x and *parameters* θ.

Issue

We have two plausible models $f_1(x, \theta_1)$ and $f_2(x, \theta_2)$.

We want more experiments to distinguish/discriminate f_1 and f_2

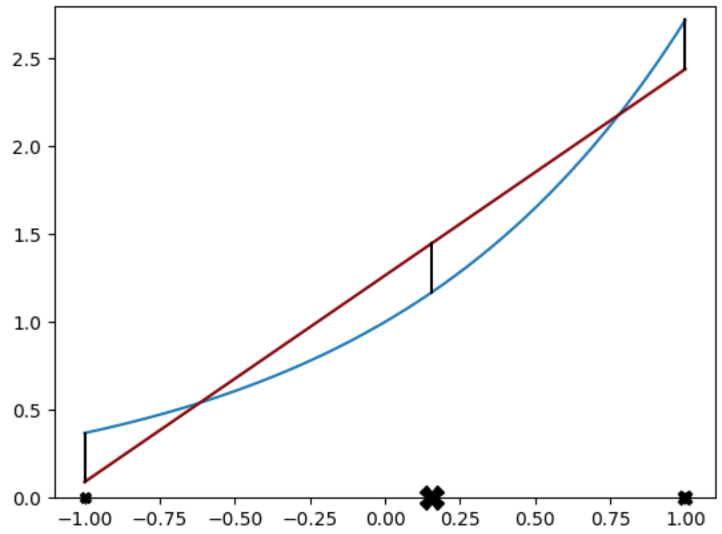
Goal

- Determine a *experimental design* $\xi = \begin{Bmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{Bmatrix}$ where
  - the models should *differ the most* at the chosen design points $x_i \in X \subseteq \mathbb{R}^d$.
  - the weights $w_i \in [0,1]$ should state the *importance* of the corresponding design point.

T-Criterion

Idea

- Fix one model → true model $f_1(x) = f_1(x, \theta_t)$
- Fit alternative model $f_2(x, \theta_2)$ to f_1 as well al possible
- Take design points where models have the largest difference



Example: linear model against true exponential model

T-criterion

$$\xi^* = \underset{\xi \in \Xi}{\operatorname{argmax}} \min_{\theta_2 \in \Theta_2} \int_X [f_1(x) - f_2(x, \theta_2)]^2 \ \xi(dx)$$

For multi-response models, use $\|f_1(x) - f_2(x, \theta_2)\|_2^2$

The following ideas work analogously for the KL-criterion

Fraunhofer ITWM

Semi-Infinite Programming

Alternatingly solving upper- and lower-level problem solves SIP

→ **Blankenship & Falk algorithm** (adaptive discretization)

Semi-infinite program (SIP)

$$\max_{x \in \mathbb{R}^d} \; f(x)$$
$$s.t. \;\; g(x,y) \geq 0, \qquad \text{for all } y \in Y$$

Upper-level problem: Compute new approximate (infeasible) solution $\bar{x}$ by solving the discretized problem given a finite discretization $\dot{Y} \subset Y$:

$$SIP(\dot{Y}): \; \max_{x \in \mathbb{R}^d} \; f(x)$$
$$s.t. \;\; g(x,y) \geq 0, \qquad \text{for all } y \in \dot{Y}$$

Lower-level problem: Determine constraint, or *index variable* y, with highest violation for a fixed solution $\bar{x}$:

$$Q(\bar{x}): \; \min_{y \in Y} \; g(\bar{x}, y)$$

T-Criterion as Semi-Infinite Program

Maximin problems as SIP

A problem of the form $\max_{x\in X} \min_{y\in Y} f(x,y)$ can be formulated as a semi-infinite program (*epigraph formulation*)

$$\max_{\substack{t\in\mathbb{R}\\ x\in X}} t$$
$$s.t. \quad f(x,y) \geq t, \ \forall y \in Y$$

T-criterion as SIP

$$\max_{\substack{t\in\mathbb{R}\\ \xi\in\Xi}} t$$
$$s.t. \quad \int_X [f_1(x) - f_2(x,\theta_2)]^2 \ \xi(dx) \geq t, \ \forall \theta_2 \in \Theta_2$$

Kuczewski (2006) Computational aspects of discrimination between models of dynamic systems. PhD thesis

Duarte et al. (2014) A semi-infinite programming based algorithm for determining T-optimum designs for model discrimination

Computing T-optimal Designs

Previous algorithms to compute T-optimal designs

- Vertex Direction Method

  Atkinson and Fedorov (1975) The design of experiments for discriminating between two rival models.

  > Very robust but slow convergence

- Semi-infinite programming

  Kuczewski (2006) Computational aspects of discrimination between models of dynamic systems. PhD thesis

  Duarte et al. (2014) A semi-infinite programming based algorithm for determining T-optimum designs for model discrimination

  - Adaptively discretizes the parameter space

  > Robust, but solving upper level to global optimality is very hard

- Dette et al.

  Dette et al. (2015) Bayesian T-optimal discriminating designs

  - Adaptively discretizes the design space
  - Uses linearization to compute optimal weights in each iteration

  > Very fast (if applicable) but unreliable

New SIP Ansatz to Compute T-optimal Designs

Simulatneous optimization w.r.t. x and w

General T-criterion

$$\max_{\xi \in \Xi} \min_{\theta_2 \in \Theta_2} \int_X [f_1(x_i) - f_2(x_i, \theta_2)]^2 \, \xi(dx) = \max_{\substack{x_1,\ldots,x_N \in X \\ w_1,\ldots,w_N \in [0,1] \\ \|w\|_1 = 1}} \min_{\theta_2 \in \Theta_2} \sum_{i=1}^{N} w_i [f_1(x) - f_2(x, \theta_2)]^2$$

Discretized T-criterion

$$\max_{\substack{w \geq 0 \\ \|w\|_1 = 1}} \min_{\theta_2 \in \Theta_2} \sum_{i=1}^{N} w_i [f_1(x_i) - f_2(x_i, \theta_2)]^2, \quad \text{for fixed } X_N = \{x_1, \ldots, x_N\}$$

Linear w.r.t w

2-ADAPT-MD Algorithm

Given an initial grid of design points $X_N = \{x_1, \dots, x_N\}$ and an initial discretized parameter set $\dot{\Theta}_2^{(0)}$.

2-ADAPT-MD: WHILE (convergence criterion not met) DO

- **Find new design on $X_N \subset X$ (DISC_MD):** WHILE (convergence criterion not met) DO
  - Compute new weights: $$w^{(s)} = \underset{\substack{w \geq 0 \\ \|w\|_1 = 1}}{\operatorname{argmax}} \min_{\theta_2 \in \dot{\Theta}_2^{(s)}} \sum_{i=1}^N w_i [f_1(x_i) - f_2(x_i, \theta_2)]^2$$
  - Update parameter discretization: $$\hat{\theta}_2^{(s)} = \underset{\theta_2 \in \Theta_2}{\operatorname{argmin}} \sum_{i=1}^N w_i^{(s)} [f_1(x_i) - f_2(x_i, \theta_2)]^2, \quad \dot{\Theta}_2^{(s+1)} := \dot{\Theta}_2^{(s)} \cup \left\{\hat{\theta}_2^{(s)}\right\}$$
- **Find new design point:** $$x^{(N+1)} = \underset{x \in X}{\operatorname{argmax}} \left[f_1(x) - f_2\left(x, \hat{\theta}_2^{(N)}\right)\right]^2, \quad X_{N+1} := X_N \cup \left\{x^{(N+1)}\right\}$$

Return: (approximate) optimal design ξ^* on X.

Linear program

Least-squares problem

(General) non-linear program

Convergence of the 2-ADAPT-MD

For more details, see Mogalle et al. (2023) *Computing T-optimal Designs via Nested Semi-Infinite Programming and Twofold Adaptive Discretization*

It converges

Assumptions on m… …sume single-response models $f: X \times \Theta \to \mathbb{R}$ with

- $X \neq \emptyset$ is comp…
- $\Theta \neq \emptyset$ is com…
- f is continu…
- K such that $T\left((1 - \alpha)…\right.$

ε-T-optimal Designs

We call a design ξ ε-T-optimal if $\max_{x \in X} \psi(x, \xi) \leq$ … …al design ξ^*.

Convergence of 2-ADAPT-MD

- *DISC-MD converges to T-optimal design on $X^{(n)} \subset X$ → stops with ε-T-optima…*
- *2-ADAPT-MD* converges to *a* ***$2\sqrt{K\varepsilon}$-T-optimal design*** *$\hat{\xi}$ on X*

Fraunhofer ITWM

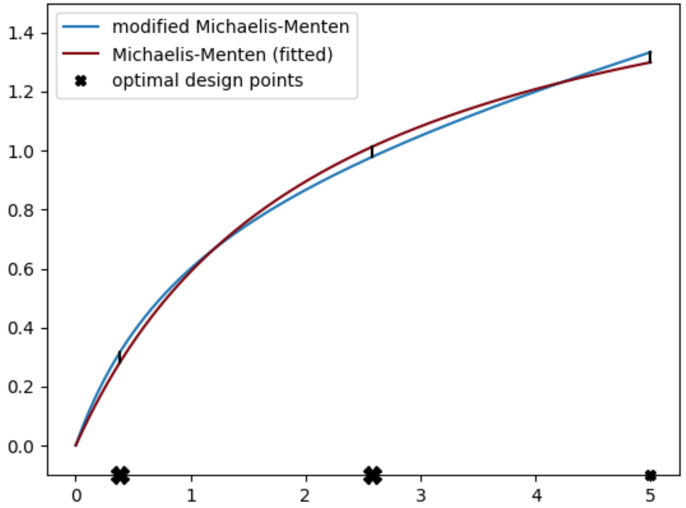
Numerical Example 1: Michaelis-Menten Model

Modified-Michaelis-Menten-Modell (MMM) vs. Michaelis-Menten-Modell (MM)

$$f_1(x) = \frac{x}{1+x} + 0.1x \qquad \text{vs.} \qquad f_2(x,(V,K)) = \frac{Vx}{K+x}$$

- $X = [10^{-3}, 5.0]$
- $(V,K) \in \Theta_2 = [10^{-3}, 5] \times [10^{-3}, 5]$

Algorithm	Accuracy	T-Criterion	Run Time
VDM	$9.64 \cdot 10^{-6}$	$1.1805 \cdot 10^{-3}$	41.31 s
Dette et al	$5.02 \cdot 10^{-6}$	$1.1853 \cdot 10^{-3}$	1.40 s
SIP	$4.75 \cdot 10^{-6}$	$1.1820 \cdot 10^{-3}$	16511.88 s
2-ADAPT-MD	$3.46 \cdot 10^{-6}$	$1.1854 \cdot 10^{-3}$	6.91 s



MMM reference model vs. fitted MM alternative model

bad performance due to global optimization in upper level

Numerical Example 2: Consecutive Reaction

Partially reversible vs. irreversible consecutive reaction

- Irreversible reaction: $A \rightarrow^{k_1} B \rightarrow^{k_2} C$
- Partially reversible reaction: $A \rightleftharpoons^{k_1}_{k_3} B \rightarrow^{k_2} C$

- $([A]_0, [B]_0, [C]_0, t) \in X$ → grid with **135** design points
- $(k_1, k_2, n_1, n_2) \in \Theta_2$

- ODE system:

$$\frac{d[A]}{dt} = -k_1[A]^{n_1} + k_3[B]^{n_3}$$
$$\frac{d[B]}{dt} = k_1[A]^{n_1} - k_2[B]^{n_2} - k_3[B]^{n_3}$$
$$\frac{d[C]}{dt} = k_2[B]^{n_2}$$

Algorithm	Accuracy	T-Criterion	Run Time
VDM	$9.90 \cdot 10^{-6}$	$1.9275 \cdot 10^{-3}$	2818.71 s
Dette et al	–	–	–
SIP	$4.04 \cdot 10^{-6}$	$1.9322 \cdot 10^{-3}$	12042.51 s
2-ADAPT-MD	$1.31 \cdot 10^{-6}$	$1.9322 \cdot 10^{-3}$	345.30 s

All design points have to be evaluated in every iteration

Take smaller subset of design points

Benefits of New Formulation

Advantages of 2-ADAPT-MD algorithm

- Run times are competitive with state-of-the-art algorithms
- Increased numerical stability
  - Our (subjective) experience: Our algorithm works reliably, while others have their little problems
  - Bad points in a discretization do not hurt
- Better structure
  - Each sub-problem is solved whenever needed
  - One could use heuristics to augment the discretizations

Thank you for your attention

Fraunhofer
ITWM

References

- Mogalle D, Seufert P, Schwientek J, Bortz M, Küfer KH (2023) *Computing T-optimal designs via nested semi-infinite programming and twofold adaptive discretization.*
- Kuczewski B (2006) *Computational aspects of discrimination between models of dynamic systems*. PhD thesis
- Dette H, Melas VB, Guchenko R (2015) *Bayesian T-optimal discriminating designs.*
- Braess D, Dette H (2013) *Optimal discriminating designs for several competing regression models.*
- López M, Still G (2007) *Semi-infinite programming.*
- Atkinson AC, Fedorov VV (1975a) *The design of experiments for discriminating between two rival models.*
- Duarte BP, Wong WK, Atkinson AC (2015) *A semi-infinite programming based algorithm for determining T-optimum designs for model discrimination.*

Thank you for your attention

Computing T-optimal Designs

How to order these sub-problems efficiently?

ALGORITHM: WHILE (convergence criterion not met) DO

- Update parameters:

$$\hat{\theta}_2^{(s)} = \underset{\theta_2 \in \Theta_2}{\operatorname{argmin}} \int_X [f_1(x) - f_2(x, \theta_2)]^2 \, \xi_s(dx)$$

- Find new design point:

$$x^{(s+1)} = \underset{x \in X}{\operatorname{argmax}} \left[f_1(x) - f_2\left(x, \hat{\theta}_2^{(s)}\right) \right]^2$$

- Update design:
  - Vertex Direction Method:

Atkinson and Fedorov (1975) The design of experiments for discriminating between two rival models.

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi\left(x^{(s+1)}\right)$$

  - Dette et al:

Dette et al (2015) Bayesian T-optimal discriminating designs.

$$w_{s+1} = \underset{\substack{w \geq 0 \\ \|w\|_1 = 1}}{\operatorname{argmax}} \sum_{i=1}^{n} w_i \left[f_1(x_i) - f_2\left(x_i, \hat{\theta}_2^{(s)}\right) - \beta(w)^T D_{\theta_2} f_2\left(x_i, \hat{\theta}_2^{(s)}\right) \right]^2$$

Linearization around $\hat{\theta}_2^{(s)}$
→ $\beta(w)$ constructed to enforce good fit

Important Properties of the T-Criterion

First-order methods converge to global optimum

Concavity

The T-criterion is concave, i.e. for $\alpha \in [0,1]$:

$$T\left((1-\alpha)\xi + \alpha\bar{\xi}\right) \geq (1-\alpha)\,T(\xi) + \alpha\,T\left(\bar{\xi}\right)$$

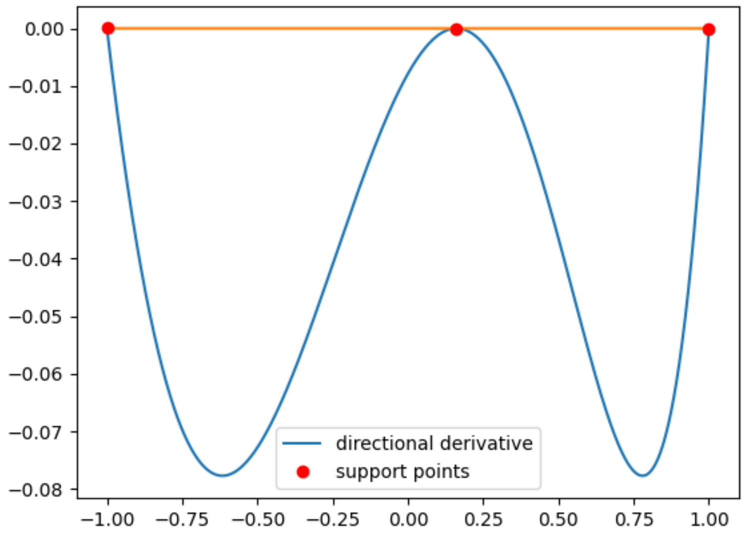
Equivalence Theorem (assuming $\hat{\theta}_2(\xi^*)$ is a unique minimizer)

- A design ξ^* is T-optimal if and only if

$$\psi(x,\xi^*) \coloneqq \varphi\left(x, \hat{\theta}_2(\xi^*)\right) - T(\xi^*) \leq 0, \qquad \text{for all } x \in X$$

- $\varphi\left(x, \hat{\theta}_2(\xi^*)\right) - T(\xi^*) = 0, \quad \text{for all } x \in \text{supp}(\xi^*)$

convergence criterion



Example: directional derivative ψ of linear model against true exponential model

Kuczewski (2006) Computational aspects of discrimination between models of dynamic systems. PhD thesis

Fraunhofer ITWM

Convergence of DISC-MD Subroutine

Convergence of DISC-MD Procedure

Let

- $\dot{\Theta}_2^{(0)}$ *be a finite discretization of* Θ_2,
- $\xi^{(s)}$ *be a regular design in each iteration (least-squares problem has unique minimizer).*

$\Longrightarrow$ *Then, every accumulation point of the sequence of solutions* $\left(\xi^{(s)}\right)_{s\in\mathbb{N}}$ *of the DISC-MD routine is a T-optimal solution on* X_n.

Proof (Sketch)

Show that the initial feasible region

$$\mathcal{F}\left(\dot{\Theta}_2^{(0)}\right) = \left\{ (w,t) \in \mathbb{R}^{N+1} \;\middle|\; t \leq \sum_{i=1}^{N} w_i \varphi(x_i, \theta_2)\,, \forall \theta_2 \in \dot{\Theta}_2^{(0)}, \quad \sum_{i=1}^{N} w_i = 1, \quad w_i \geq 0 \right\}$$

is compact (Lopez and Still, 2007).

Convergence of 2-ADAPT-MD Algorithm

Assumption

Choose K such that $T\left((1-\alpha)\xi\ +\alpha\bar{\xi}\ \right) \geq T(\xi) + \alpha \int_X \psi(x,\xi)\, d\bar{\xi} - \alpha^2 K$.

K is bound on second (directional) derivative

$\boldsymbol{\varepsilon}$**-**$\boldsymbol{T}$-optimal Designs

We call a design ξ ε-T-optimal if $\max_{x\in X} \psi(x,\xi)\ \leq \varepsilon$. Then, also $T(\xi) \geq T(\xi^*) - \varepsilon$ for a T-optimal design ξ^*.

Convergence of 2-ADAPT-MD

Let

- *DISC-MD returns ε-T-optimal design $\xi^{(n)}$ on $X^{(n)} \subset X$*

$\Longrightarrow$ *Then, a subsequence of (intermediate) solutions $\left(\xi^{(n_j)}\right)_{j\in\mathbb{N}}$ converges to a* ***$2\sqrt{K\varepsilon}$-T-optimal design*** *$\hat{\xi}$ on X.*

Convergence of 2-ADAPT-MD Algorithm

Convergence of 2-ADAPT-MD

Let

- *DISC-MD returns ε-T-optimal design $\xi^{(n)}$ on $X^{(n)} \subset X$*

$\Longrightarrow$ *Then, a subsequence of (intermediate) solutions $\left(\xi^{(n_j)}\right)_{j\in\mathbb{N}}$ converges to a* ***$2\sqrt{K\varepsilon}$-T-optimal design*** *$\hat{\xi}$ on X.*

Proof (Sketch)

- There is a subsequence of solutions which converges weakly to some $\hat{\xi}$ (due to convergence of objective value)
- The weak limit $\hat{\xi}$ is a $2\sqrt{K\varepsilon}$-T-optimal solution.
  - Assume contrary, i.e. $\max_{x\in X} \psi(x, \xi_s) \geq 2\sqrt{K\varepsilon} + \varepsilon'$ from some iteration onwards.
  - Consider optimal line search → we improve the (optimal) objective value by at least $(\varepsilon')^2/4K$ in every iteration
  - Contradiction to $T(\xi^*)$ to being bounded.

