Sparse Grids-Particle Swarm Optimization for finding Bayesian Optimal Designs

Yu Shi, Zizhao Zhang, Weng Kee Wong

Department of Biostatistics
University of California Los Angeles

Aug 9, 2017

Outline

Background

- Bayesian non-linear mixed effects models are widely used to analyze longitudinal data.
- Designs to estimate parameters in such models are often of interest but challenging.
- Such models are often employed to study HIV dynamics over time.

Bayesian Non-linear Mixed Effects Models

- A general set up of Non-linear Mixed Effects Model:

$$Y_i(\xi) = \eta(\xi; \beta_i) + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n)$$

where $\beta_i = \beta + b_i$ and $b_i \sim N(0, D)$. Then an n-run exact experiment ξ takes observations at $t_1, ..., t_n$ results in a Fisher information for β_i as

$$M(\beta_i; \xi) = \frac{1}{\sigma^2} \sum_{j}^{n} \nabla\eta(t_j; \beta_i) \nabla\eta(t_j; \beta_i)^T$$

Bayesian Optimal Design

- A Bayesian D-optimal design η maximizes the criterion function wrt a prior parameter distribution $\pi(\beta_i)$ as:

$$\xi_{BayesD} = \underbrace{\underset{\xi}{\text{argmax}} \underbrace{\int \{\log(\det[M(\xi, \beta_i)])\}\pi(\beta_i)\, d\beta_i}_{\text{Integration}}}_{\text{Optimization}}$$

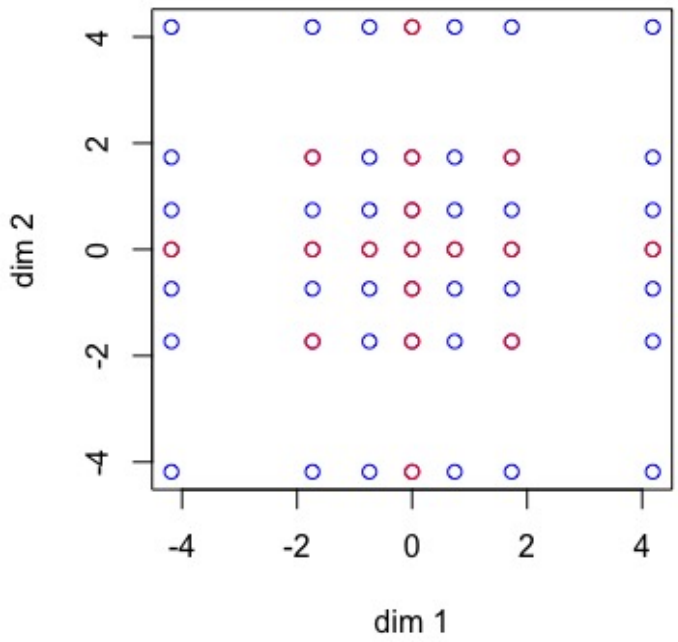
  where $M(\xi, \beta_i)$ is the information matrix.

- Numerical (multi-dimensional) integration results evaluated by optimization process.

Numerical Integration

- Monte Carlo: $\sim$100,000 samples (Nyberg, Karlsson, Hooker, 2009).
- MCMC: $\sim$10,000 iterations (Wakefield, 1996; Huang and Wu, 2008).
- Adaptive Gaussian quadrature (Ueckert and Mentre, 2016); Hamiltonian Monte Carlo (Riviere, Ueckert, and Mentre, 2016).
- Sparse Grids (Smolyak, 1963): the computational costs to achieve a negligible approximation error are considerably lower than with simulation techniques like pseudo-random Monte Carlo and quasi-random Monte Carlo (Heiss et. al., 2008). It has been used various fields like engineering and econometrics, and also used in computer experiment design (Plumlee, 2014).

Sparse Grids Method



- - Complexity: $O(N(logN)^{D-1})$ instead of $O(N^D)$
  - Error bound: $O(N^{-2}(logN)^{D-1})$ instead of $O(N^{-2})$.
  - Deterministic.

Brief Intro to Sparse Grids

- Define the difference of the approximation when increasing the accuracy level from $i-1$ to i as

$$\Delta_i[g] = V_i[g] - V_{i-1}[g] \quad \forall i \in \mathbb{N}$$

with $i = [i_1, ..., i_D]$. For any non negative integer q, define non empty set

$$\mathbb{N}_q^D = \{i \in \mathbb{N}^D : \sum_d^D i_d = D + q\}$$

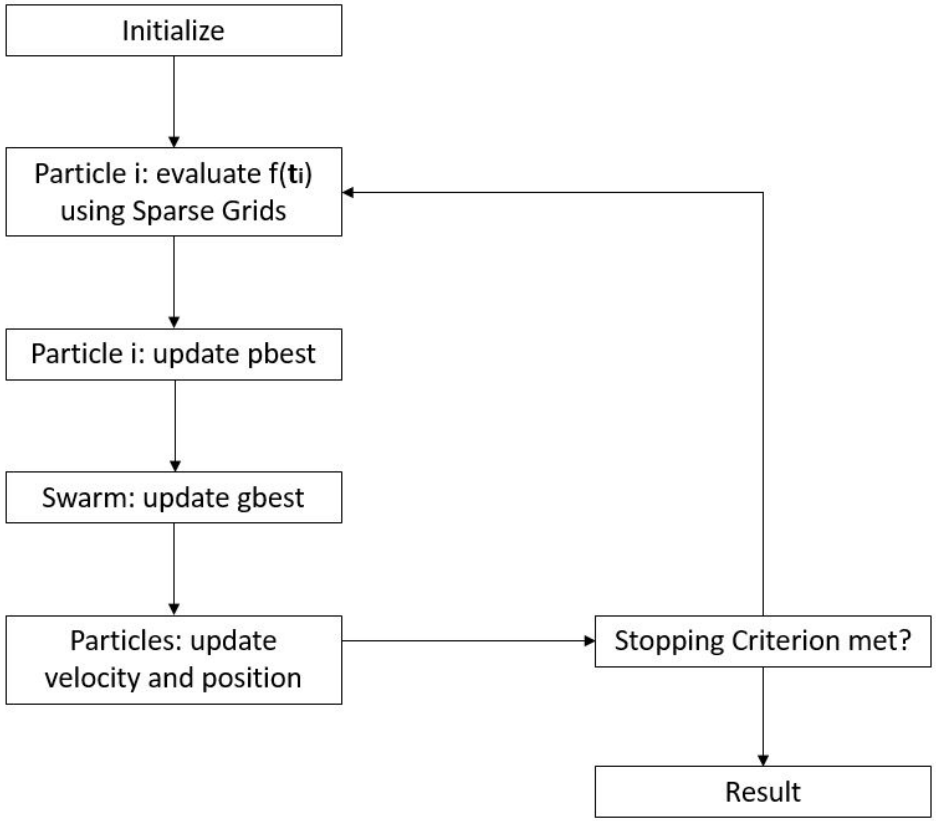
For instance, $\mathbb{N}_2^2 = \{[1,3],[2,2],[3,1]\}$. Then the sparse grids quadrature rule for D-dimensional integration at accuracy level k is defined as

$$A_{D,k}[g] = \sum_{q=0}^{k-1} \sum_{i \in \mathbb{N}_q^D} (\Delta_{i_1} \otimes ... \otimes \Delta_{i_D})[g]$$

Optimization

- Compare a small number of candidate designs (Huang and Wu, 2008).
- Fedorov algorithm (Fedorov, 1972; Retout, Mentre, Bruno, 2002).
- Coordinate exchange algorithm (Meyer and Nachtsheim, 1995; Gotwalt, Jones, Steinberg, 2009).
- Particle Swarm Optimization (Eberhart and Kennedy, 1995; Wong et al., 2015; Chen et al., 2015; Phoa et al., 2016. Kim and Wong, 2017).

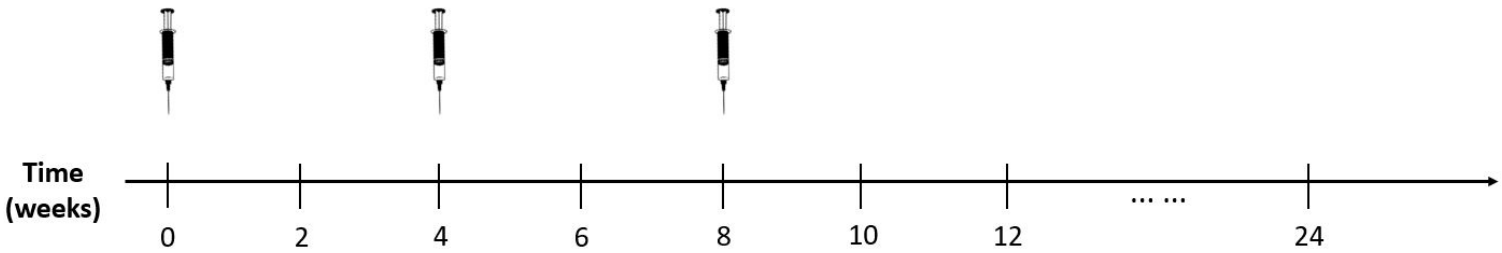
SG-PSO for Finding Bayesian Optimal Designs



$f(t_i) = \int_\beta log(det[M(\xi, \beta_i)])\pi(\beta_i)d\beta_i$ evaluated at SG samples. PSO parameters: 1000 iter, 40 particles, inertia $\frac{1}{2ln\ 2}$, $c_1 = c_2 = 0.5 + ln\ 2$.

Application in HIV Studies

- HIV experiments often requires to get plasma samples from patients to quantify viral load at specific time points. The longitudinal change is important to study the pathological process involved in HIV dynamics.



- In this work we focus on deciding the best sampling schedule.

Application in HIV Studies I

- Model Structure (Hu and Ding, 1999): the log-transformation of the total viral load for the i-th subject at j-th time point is modeled as

$$y_{ij} = log(V(t_{ij}, \beta_i)) + e_{ij} = log(e^{\theta_i - \delta_i t_j} + e^{\tau_i - \lambda_i t_j}) + e_{ij}$$

  assume isotropic errors across time and subjects $e_{ij}|\beta_i \sim N(0, \sigma^2)$, $\beta_i = (\theta_i, \tau_i, \delta_i, \lambda_i)' = (\theta, \tau, \delta, \lambda)' + (b_{1i} + b_{2i} + b_{3i} + b_{4i})' = \beta + b_i$; $b_i \sim N(\underline{0}, D)$. θ, τ represent initial viral production rate and δ, λ represent exponential decay rate of virus in compartments.

- Prior Distribution: estimates from 46 patients' data: $\hat{\beta} = (12.142, 7.624, 0.442, 0.032)'$, $\hat{\sigma}^2 = 0.267^2$, $\hat{D} = diag(1.397^2, 1.545^2, 0.137^2, 0.015^2)$.

Application in HIV Studies I

- Want to find an exact design $\xi = (t_1, ..., t_n)$ that gives best parameter estimations in response curve.

- Design Space and Number of Support Points:

  - Schedule 1 (9 points in [0, 168]); ACTG Protocol A5055: 0, 7, 14, 28, 56, 84, 112, 140, 168.

  - Schedule 2 (8 points in [2, 84]); ACTG Protocol 315: 2, 7, 12, 13, 21, 27, 55, 84.

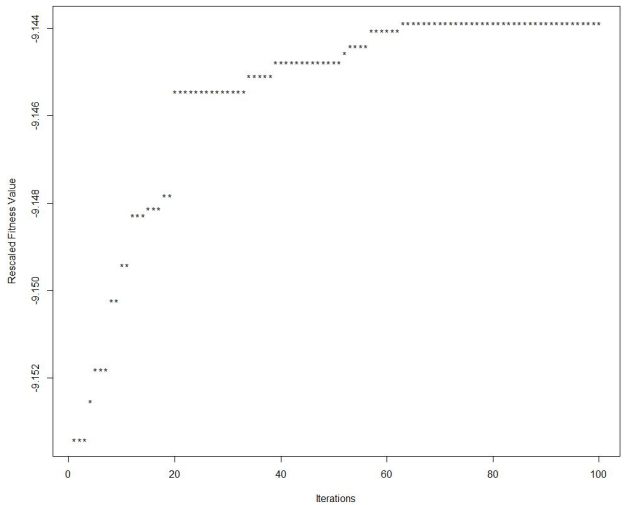
  - Schedule 3 (6 points in [0, 168]); Standard sampling schedule currently used in many AIDS clinical trials: 0, 28, 56, 84, 112, 168.

Results from SGPSO

	logdet $M(\xi)$	SGPSO Design	logdet $M(\xi^{SGPSO})$	RE	time[1]
1	14.83	0, 0, 4.7, 8.2, 15.0, 27.0, 45.5, 168, 168	15.80	78.5%	18.9
2	12.27	2, 2, 6.1, 10.9, 20.3, 34.7, 84, 84	12.73	88.9%	15.6
3	2.46	0, 4.7, 10.2, 21.6, 43.0, 168	14.01	5.6%	10.9

Relative Efficiency:

$$\left(\frac{|M(\xi, \beta_i)|}{|M(\xi^{SGPSO}, \beta_i)|}\right)^{1/4}$$



[1]CPU time. Computer configuration: x86_64−win64 (64 bit) 3.6GHz, 16GB RAM, Intel i7−4790 CPU on Windows 10 Enterprise OS. R version: 3.3.0.

Application in HIV Studies I: Other criteria

Another criterion is to maximize the determinant of FIM whose (k, l) term is denoted as (Riviere, Ueckert, Mentre, 2016)

$$M(\beta,\xi)_{k,l} = E_y[\frac{\partial log(L(y|\beta))}{\partial \beta_k}\frac{\partial log(L(y|\beta))}{\partial \beta_l})]$$

$$= E_y[E_{b|y}(\frac{\partial log(p(y|b,\beta)p(b))}{\partial \beta_k})E_{b|y}(\frac{\partial log(p(y|b,\beta)p(b))}{\partial \beta_l})]$$

	logdet $M(\xi)$	SGPSO Design	logdet $M(\xi^{SGPSO})$	RE	time
1	37.32	0, 8.6, 14.3, 18.9, 30.3, 157.9, 164.5, 167.8, 168	46.46	10.17%	1983
2	37.96	2, 2.3, 15.8, 16.2, 34.1, 37.7, 65.1, 83.0	40.84	48.6%	1847
3	36.67	0, 29.3, 29.4, 32.3, 168, 168	44.62	13.7%	1688

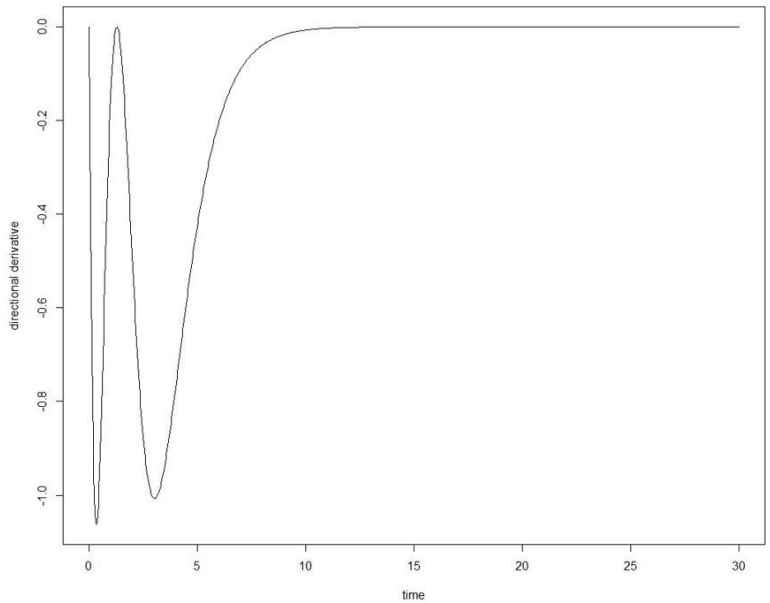
Application in HIV Studies II: Approximate Design

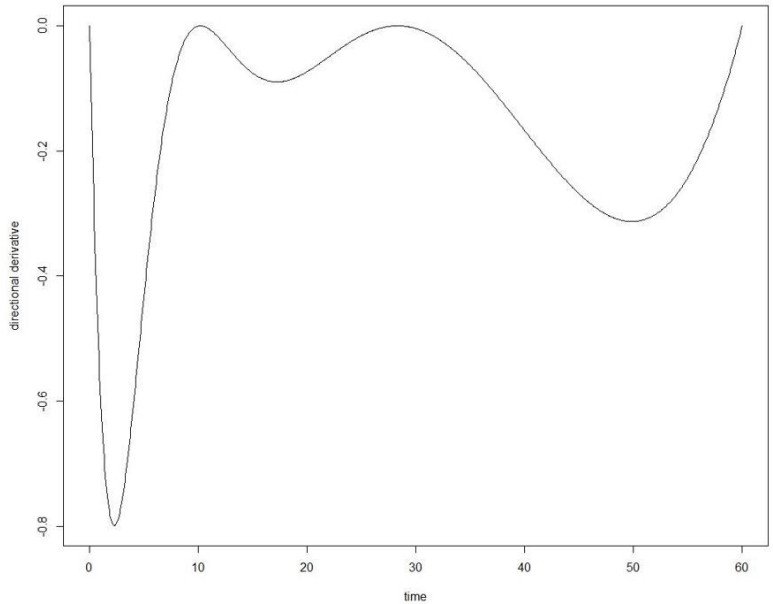
- An exponential model which describes the trajectory of plasma HIV RNA level under antiviral treatment (Han and Chaloner, 2003)

$$Y_j = log(P_0 + P_1 e^{-\delta t_j}) + \epsilon_j$$

where $\epsilon_j \overset{iid}{\sim} N(0, \sigma^2)$, $t_j \in [0, 60]$, P_0, $P_1 \sim$ unif(0.5, 1.5).

- $\delta \sim$u(0.9, 1.1): $\begin{pmatrix} 0 & 1.27 & 14 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$; $\delta \sim$u(0, 0.2): $\begin{pmatrix} 0 & 10.17 & 28.30 & 60 \\ 0.32 & 0.28 & 0.10 & 0.30 \end{pmatrix}$





(Preliminary) Conclusion

- The proposed SGPSO algorithm is promising for finding Bayesian optimal designs with mixed effects models.
- Fast and convenient: softwares and packages widely available.
- Flexible: deals with different priors, design space constrains, number of support points with minimal adjustment; good for practical use.
- Works for approximate designs, also seems to work for exact designs.
- Deals with various criteria.

Advanced Topics: Hybrid Algorithm

- Advanced versions: Quantum PSO (Sun et. al, 2004), Competitive Swarm Optimizer (Cheng and Jin, 2015), Galactic Swarm Optimization (Muthiah-Nakarajan and Noel, 2016), etc.

Algorithm 1 Hybrid Swarm Optimization (Zhang et al., 2017+)

Lower Level Initialization: $\mathbf{x}$, $\mathbf{v}_L$.
Upper Level Initialization: $\mathbf{v}_U$.
for EP in 1 to EP_{max} **do**
    for subswarm 1 to M **do**
        CSO m_1 iterations by updating $\mathbf{x}$, $\mathbf{v}_L$;
        Pick k_1 winner particles;
    end for
    do CSO among winners by updating $\mathbf{x}$, $\mathbf{v}_U$;
end for

Advanced Topics: Hybrid Algorithm

- The hybrid swarm optimization algorithm has been tested for finding locally optimal design for a logistic model with 4 factors with all pairwise interaction terms (80-dimensional). Ten different sets of coefficients have been tested and the average efficiency lower bound obtained from 20 experiments show as follows:

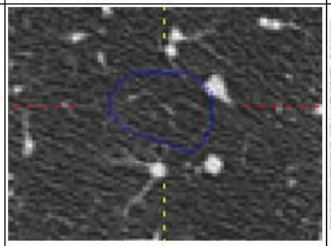
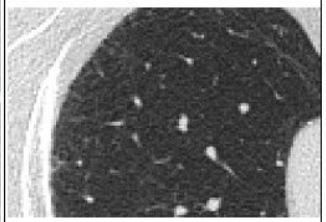
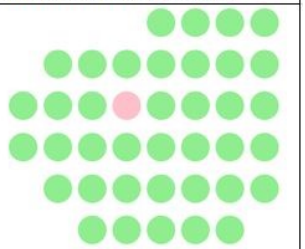
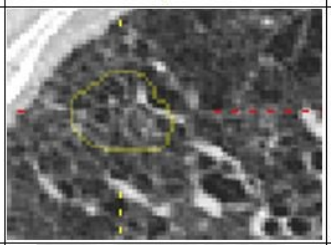
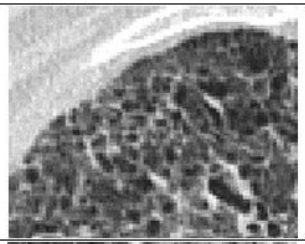
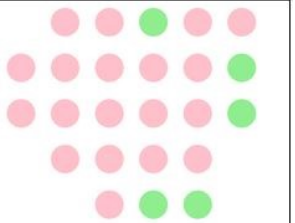
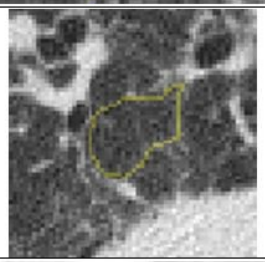
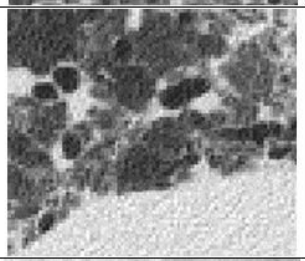
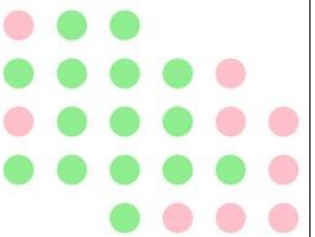
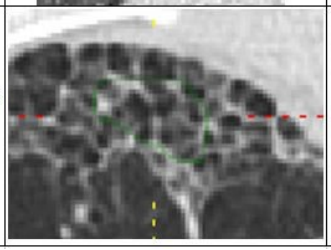
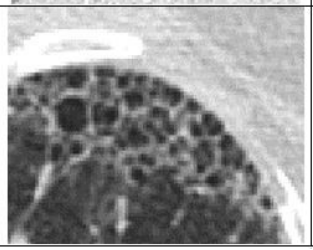
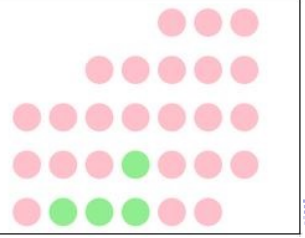
	HSO	CSO	GSO	PSO
Average ELB	81.2%	43.4%	14.56%	6.7%

Advanced Topics: Application in Other Fields

- Biomedical imaging study: using QPSO to maximize the lung cancer status prediction.
- Idiopathic Pulmonary Fibrosis (IPF) is a rare and ultimately fatal lung disease of unknown cause with a median survival of 2$\sim$5 years. The disease natural history is unpredictable at the time of diagnosis; as a result, some patients miss the best time to receive transplant. We want to predict the progression status so that the timely treatment can be delivered to increase patients' survival.

Advanced Topics: Application in Other Fields

- LASSO: sensitivity 63.04%, specificity 75.18%, accuracy 73.04% (51 features). QPSO: 77.27% sensitivity, 85.15% specificity, 83.74% accuracy (19 features).

Case	Classfication type	ROI on CT-baseline	ROI on CT-followup	classification at voxel level
1	non-progression ROI classified as non-progression			
2	progression ROI classified as progression			
3	progression ROI classified as non-progression			
4	non-progression ROI classified as progression			

Remarks and Future Work

- Sparse Grid rule usually works better for normal priors.
- No free lunch theorem (Wolpert and Macready, 1997).
- Future work:
  - Explore advanced swarm algorithms; hybridized algorithms.
  - Apply the SGPSO algorithm to other design criteria for non-linear mixed effect models and some other models.
  - ...etc.

Acknowledgement

- National Institutes of Health Grant R01GM107639.
- UCLA Graduate Research Summer Mentorship Award.
- UCLA-CVIB group; UCLA CTD-ILD program; NIH R01 HL089758; R21 HL123477-01A1.

References I

[1] Wu, Hulin, and A. Adam Ding. "Population HIV1 Dynamics In Vivo: Applicable Models and Inferential Tools for Virological Data from AIDS Clinical Trials." Biometrics 55.2 (1999): 410-418.

[2] Huang, Yangxin, and Hulin Wu. "Bayesian experimental design for long-term longitudinal HIV dynamic studies." Journal of statistical planning and inference 138.1 (2008): 105-113.

[3] Wakefield, J.C., 1996. The Bayesian analysis to population pharmacokinetic models. J. Amer. Statist. Assoc. 91, 6275.

[4] Nguyen, Thu Thuy, and France Mentr. "Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature." Computational Statistics Data Analysis 80 (2014): 57-69.

References II

[5] S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR, 148:10421043, 1963. Russian, Engl. Transl.: Soviet Math. Dokl. 4:240243, 1963.

[6] Heiss, Florian, and Viktor Winschel. "Likelihood approximation by numerical integration on sparse grids." Journal of Econometrics 144.1 (2008): 62-80.

[7] Plumlee M. Fast prediction of deterministic functions using sparse grid experimental designs[J]. Journal of the American Statistical Association, 2014, 109(508): 1581-1591.

[8] Retout, Sylvie, France Mentr, and Rene Bruno. "Fisher information matrix for nonlinear mixedeffects models: Evaluation and application for optimal design of enoxaparin population pharmacokinetics." Statistics in medicine 21.18 (2002): 2623-2639.

References III

[9] Gotwalt, Christopher M., Bradley A. Jones, and David M. Steinberg. "Fast computation of designs robust to parameter uncertainty for nonlinear settings." Technometrics 51.1 (2009): 88-95.

[10] Eberhart, Russell, and James Kennedy. "A new optimizer using particle swarm theory." Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE, 1995.

[11] Wong, Weng Kee, et al. "A modified particle swarm optimization technique for finding optimal designs for mixture models." PloS one 10.6 (2015): e0124720.

[12] Chen, R. B., Wang, W.C., Chang, S. P. and Wong, W. K. (2015). Minimax Optimal Designs via Particle Swarm Optimization Methods. Statistics and Computing, Vol. 25, 975-988.

References IV

[13] Phoa, K. H. F., Chen, R. B., Wang, W. C. and Wong, W. K. (2016). Optimizing Two-level Supersaturated Designs using Swarm Intelligence Techniques. Technometrics, Vol.58, 1, 43-49.

[14] Kim, S. and Wong, W. K. (2017). Extended Two-stage Adaptive Designs for Phase II Clinical Trials using Particle Swarm Optimization. Statistical Methods in Medical Research. In press.

[15] Chen, R. B., Chen P. Y., Tung, H. C. and Wong, W. K. (2015). Exact D-optimal Designs for Michaelis-Menten Model with Correlated Observations by Particle Swarm Optimization. Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday, (editor: Ellinor Fackle-Fornius), 60-73. Department of Statistics, Stockholm University.

[16] Riviere, Marie-Karelle, Sebastian Ueckert, and France Mentr. "An MCMC method for the evaluation of the Fisher information matrix for non-linear mixed effect models." Biostatistics 17.4 (2016): 737-750.

References V

[17] Sun, Jun, Wenbo Xu, and Bin Feng. "A global search strategy of quantum-behaved particle swarm optimization." Cybernetics and Intelligent Systems, 2004 IEEE Conference on. Vol. 1. IEEE, 2004.

[18] Cheng, Ran, and Yaochu Jin. "A competitive swarm optimizer for large scale optimization." IEEE transactions on cybernetics 45.2 (2015): 191-204.

[19] Muthiah-Nakarajan, Venkataraman, and Mathew Mithra Noel. "Galactic Swarm Optimization: A new global optimization metaheuristic inspired by galactic motion." Applied Soft Computing 38 (2016): 771-787.

[20] Wolpert, David H., and William G. Macready. "No free lunch theorems for optimization." IEEE transactions on evolutionary computation 1.1 (1997): 67-82.

References VI

[21] Han, Cong, and Kathryn Chaloner. "D-and c-optimal designs for exponential regression models used in viral dynamics and other applications." Journal of statistical planning and inference 115.2 (2003): 585-601.