Application of Particle Swarm Optimization in minimax D-optimal Design

Park, Sungmin

Department of Statistics
Seoul National University

10 MAY 19

Outline

Outline

Optimal design for nonlinear model is an optimization problem where analytical formula for the design is rarely available.

In this session,

- Locally D-optimal design is generalized to minimax D-optimal design.
- Particle swarm optimization is applied to minimax D-optimal design.
- Limitations of particle swarm optimization are discussed.

Outline

Review

Given

- $\mathcal{P}(\cdot)$: model
- θ: parameter
- k: $\#$ of distinct design points

find

- x_i: design points
- w_i: proportion of subjects assigned to design point x_i

  i.e. $\xi = \{(x_i, w_i) : i = 1, \ldots, k\}$: k-point approximate design

s.t. minimizes

$$\left| Var\left(\hat{\theta}(\xi)\right) \right| \quad (D\text{-optimality criterion})$$

Review

More specifically using MLE,

- $\mathcal{P}(\cdot)$: model
- θ: parameter
- k: $\#$ of distinct design points

find

- x_i: design points
- w_i: proportion of subjects assigned to design point x_i

  i.e. $\xi = \{(x_i, w_i) : i = 1, \dots, k\}$: k-point approximate design

s.t. minimizes

$$-log\ det\left(\mathcal{I}\left(\xi, \theta\right)\right), \quad \mathcal{I}\left(\xi, \theta\right): \text{Fisher information}$$

Minimax D-optimal design

In locally D-optimal design, nominal value of θ_0 is given.

$$\xi^* = \underset{\xi}{\text{argmin}} \left\{ -log\ det \left(\mathcal{I}\left(\xi, \theta_0\right)\right) \right\} \text{ for some fixed } \theta_0$$

In minimax D-optimal design, parameter space Θ is given.

Corresponding optimality criterion is defined as,

$$\xi^* = \underset{\xi}{\text{argmin}} \max_{\theta \in \Theta} \left\{ -log\ det \left(\mathcal{I}\left(\xi, \theta\right)\right) \right\}$$

Equivalence Theorem

Theorem (Berger et al., 2000)

[Notation]

- $\mathcal{X}$: *design space*, $\xi = \{(x_i, w_i) : i = 1, \ldots, k\}, x_i \in \mathcal{X}$
- Θ : *parameter space*, $\theta \in \Theta$
- $\mathcal{I}(x, \theta)$: *Fisher information at observation point* x
- $\mathcal{I}(\xi, \theta) = \int \mathcal{I}(x, \theta)\xi(dx)$: *Fisher information of design* ξ

  $*$ *Recall: design* ξ *is a probability measure.*

Equivalence Theorem

Theorem (Berger et al., 2000)

Suppose the design space $\mathcal{X}$ and parameter space Θ are known compact spaces and q is the number of parameters in the model. The following statements are equivalent:

1. $\xi^* = \underset{\xi}{\text{argmin}} \max_{\theta \in \Theta} \{-log\ det\,(\mathcal{I}\,(\xi, \theta))\}$
2. $\forall \xi \in \mathcal{X}, \min_{\theta \in A(\xi^*)} \int_{\Theta} tr\,\mathcal{I}^{-1}(\xi^*, \theta)\mathcal{I}(x, \theta)\xi(dx) - q \leq 0,$ *where*

   $$A(\xi) = \left\{\theta^* \in \Theta : -log\ det\,(\mathcal{I}\,(\xi, \theta^*)) = \max_{\theta \in \Theta} \{-log\ det\,(\mathcal{I}\,(\xi, \theta))\}\right\}$$
3. $\exists$ *probability measure* γ^* *on* $A(\xi^*)$ *s.t.*

   $$\int_{A(\xi^*)} tr\,\mathcal{I}^{-1}(\xi^*, \theta)\mathcal{I}(x, \theta)\gamma^*(d\theta) - q \leq 0, \forall \theta \in \Theta$$

Equivalence Theorem

Although equivalence theorem provides an alternative to check the optimality of the design, $A(\xi)$ and γ^* defined on $A(\xi^*)$ are unknown.

Verifying the inequality in equivalence theorem requires solving yet another optimization problem.

Outline

Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a nature-inspired algorithm originating from research in fish and swarm movement behavior.

Benefits of PSO include the ability to find the optimal solution to a complex problem or get close to the optimal solution quickly **without requiring any assumption on the objective function**. $\Rightarrow$ **flexible**

However, the method still lacks a firm theoretical justification to date and is under-utilized in statistical literature. (Qui et al., 2014)

The idea of PSO is as follows,

1. A number of particles are scattered onto the search domain.
2. Each particle investigates the search domain and shares knowledge with the group.
3. Possible solution is obtained from the group's aggregated knowledge.

PSO Algorithm

Update Equation:

$$v_i^{t+1} = \tau_t v_i^t + \gamma_1 \beta_1 \odot (p_i^t - z_i^t) + \gamma_2 \beta_2 \odot (p_g^t - z_i^t), \quad (1)$$
$$z_i^{t+1} = z_i^t + v_i^{t+1}. \quad (2)$$

- $h(\cdot)$: objective function (fitness) to minimize
- z_i^t: position of the ith particle at time t
- v_i^t: velocity of the ith particle at time t
- p_i^t: $\mathrm{argmin}_{z_i^s, 1 \leq s \leq t} \{h(z_i^s)\}$, personal best position
- p_g^t: $\mathrm{argmin}_{z_m^s, 1 \leq m \leq n, 1 \leq s \leq t} \{h(z_m^s)\}$, global best position
- $\odot$: Hadamard product operator

PSO Algorithm

Update Equation:

$$
\begin{aligned}
v_i^{t+1} &= \tau_t v_i^t + \gamma_1 \beta_1 \odot (p_i^t - z_i^t) + \gamma_2 \beta_2 \odot (p_g^t - z_i^t), \\
z_i^{t+1} &= z_i^t + v_i^{t+1}.
\end{aligned}
$$

- τ_t: inertia wieght at time t, const. or decreasing between (0,1)
- γ_1: cognitive learning parameter
- γ_2: social learning parameter
- β_1, β_2: random vector

In this paper, learning parameter was $\gamma_1 = \gamma_2 = 2$ fixed and components of β_1, β_2 was sampled i.i.d from $U(0,1)$ at each iteration and particle.

PSO Algorithm (1) - Simple PSO

PSO pseudo-code for flock size n (i.e. n particles in the swarm)

(1) Initialize particles

(1.1) Initiate position x_i^0 and velocities v_i^0 for $i = 1, \ldots, n$

(1.2) Calculate the fitness values $h(x_i^0)$ for $i = 1, \ldots, n$

(1.3) Determine the personal best positions $p_i^0 = x_i^0$ and the global position p_g^0 for $i = 1, \ldots, n$

(2) Repeat until stopping criteria are satisfied,

(2.1) Calculate particle velocity according to Eq. (1)

(2.2) Update particle position according to Eq. (2)

(2.3) Project particle back to the design space

(2.4) Calculate the fitness values $h(x_i)$ for v_i for $i = 1, \ldots, n$

(2.5) Update personal and global best positions $p_i, (1 \leq i \leq n)$ and p_g

(3) Output $p_g = \underset{x}{\operatorname{argmin}} \{h(x)\}$ with $gbest = h(p_g)$

Minimax Optimization Problem

Let $g(u, v)$ be a given function defined on two compact spaces $\mathcal{U}$ and $\mathcal{V}$.

Minimax optimization probelms have the form:

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} g(u, v) = \min_{u \in \mathcal{U}} f_{outer}(u) = \min_{u \in \mathcal{U}} \left[\max_{v \in \mathcal{V}} f_{inner}(v; u) \right]$$

where $f_{outer}(u) = \max_{v \in \mathcal{V}} f_{inner}(v; u)$

and, for fixed u, $f_{inner}(v; u) = g(u, v)$

Minimax Optimization Problem

Let $g(u, v)$ be a given function defined on two compact spaces $\mathcal{U}$ and $\mathcal{V}$.

Minimax optimization probelms have the form:

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} g(u, v) = \min_{u \in \mathcal{U}} f_{outer}(u) = \min_{u \in \mathcal{U}} \left[\max_{v \in \mathcal{V}} f_{inner}(v; u) \right] \tag{3}$$

where

$$f_{outer}(u) = \max_{v \in \mathcal{V}} f_{inner}(v; u) \tag{4}$$

and, for fixed u,

$$f_{inner}(v; u) = g(u, v) \tag{5}$$

PSO Algorithm (2) - Nested PSO

Nested PSO pseudo-code for flock size n

(1) Initialize particles

  (1.1) Initiate position x_i^0 and velocities v_i^0 for $i = 1, \ldots, n$

  (1.2) Calculate $f_{outer}(x_i)$ via Algorithm (1) (Inner loop)

  (1.3) Determine personal and global best positions p_i, p_g

(2) Repeat until stopping criteria are satisfied, (Outer loop)

  (2.1) Calculate particle velocity with Eq. (1)

  (2.2) Update particle position with Eq. (2)

  (2.3) Project particle back to design space

  (2.4) Calculate $f_{outer}(x_i)$ via Algorithm (1) (Inner loop)

  (2.5) Update personal and global best positions p_i, p_g

(3) Output $p_g = \underset{u \in \mathcal{U}}{\operatorname{argmin}} \{f_{outer}(u)\}$ with $gbest = f_{outer}(p_g)$

Outline

Example 1: Compartment Model, locally D-optimal design (revisited)

Drug concentration(Y) is modeled as a function($\eta(\cdot,\theta)$) of time(x) with independent normal errors$\left(\varepsilon \sim \mathcal{N}(0,\sigma^2)\right)$.

$$Y \sim \mathcal{N}\left(\eta(x,\theta),\sigma^2\right), \sigma^2 : \textit{known}$$
$$\eta(x,\theta) = \theta_3\left\{\exp(-\theta_2 x) - \exp(-\theta_1 x)\right\}, x > 0$$

Compartment Model, locally D-optimal design

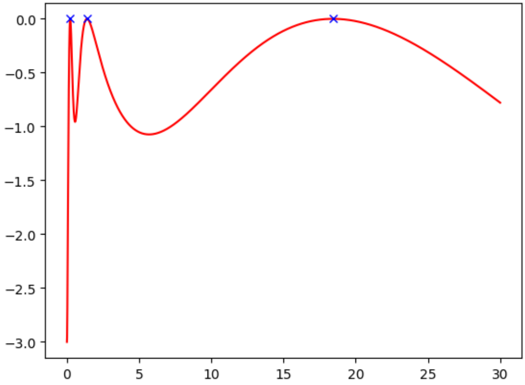
- $\theta_0 = (0.05884, 4.298, 21.8)$
- n (flock size) $= 100$
- max iteration $= 100$
- $k = 3$
- $\xi = \{(x_i, w_i) : i = 1, \ldots, k\}$,
  $x_i \in [0, 30], w_i \geq 0, \forall i$ and $\sum_{i=1}^{k} w_i = 1$

$\Rightarrow \xi^* = (0.228773, 1.38858, 18.4168, 0.333335, 0.333332, 0.333333)$

Compartment Model, locally D-optimal design

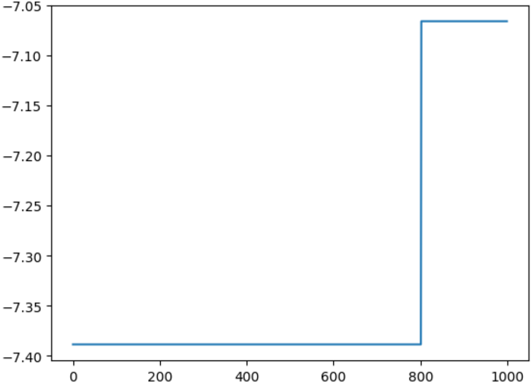
Equivalence plot for D-optimal design:

$$\left(\frac{\partial\eta(x;\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right)^T I(\xi^*;\theta_0)^{-1}\left(\frac{\partial\eta(x;\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right) - 3 \leq 0, \forall x \in [0,30]$$



Stability

$-log\ det\left(\mathcal{I}\left(\xi, \theta_0\right)\right)$ from 1000 simulation arranged in ascending order



* Relative Efficiency : $RE(\xi) = \frac{|Cov(\hat{\beta}(\xi^*))|}{|Cov(\hat{\beta}(\xi))|}$
- 20% of simulation failed
- RE (failed case) $= 0.71$

Example 2: Two-parameter Logistic Regression Model
Minimax D-optimal design

$$Y \sim Ber\left(p(x,\theta)\right)$$
$$logit(p(x,\theta)) = -b(x-a), \quad \theta = (a,b)^T$$

Find $$\xi^* = \underset{\xi \in \mathcal{X}}{\operatorname{argmin}} \max_{\theta \in \Theta} \left\{ -log\ det\left(\mathcal{I}\left(\xi,\theta\right)\right) \right\}$$

Two-parameter Logistic Regression Model

$$\mathcal{I}(\xi,\theta) = \int \begin{pmatrix} b^2 & -b(x-a) \\ -b(x-a) & (x-a)^2 \end{pmatrix} p(x,\theta)(1-p(x,\theta))d\xi(x)$$

$$\begin{aligned} |\mathcal{I}(\xi,\theta)| = & \sum_{i=1}^{k} w_i p(x_i,\theta)(1-p(x_i,\theta)) \\ & \times \sum_{i=1}^{k} w_i \left\{b(x_i-a)\right\}^2 p(x_i,\theta)(1-p(x_i,\theta)) \\ & - \left\{ \sum_{i=1}^{k} w_i b(x_i-a)^2 p(x_i,\theta)(1-p(x_i,\theta)) \right\} \end{aligned}$$

$$*p(x_i,\theta) = 1 - p(2a - x_i,\theta)$$

Two-parameter Logistic Regression Model

Above equations have useful implications when $\Theta = [a_L, a_U] \times [b_L, b_U]$.

Let $a_M = \frac{1}{2}(a_L + a_U)$.

Given a design $\xi = \{(x_i, w_i) : i = 1, \ldots, k\}$, consider a mirrored design $\xi^s = \{(2a_M - x_i, w_i) : i = 1, \ldots, k\}$.

$|\mathcal{I}(\xi, \theta)| = |\mathcal{I}(\xi^s, \theta)|$, $\frac{1}{2}(\xi + \xi^s)$ is also optimal for minimax D-optimal ξ.

$\therefore$ Minimax D-optimal design ξ^* is symmetric about a_M.

Two-parameter Logistic Regression Model

Consider two cases,

(a) $\Theta = [0, 2.5] \times [1, 3], \mathcal{X} = [-1, 4], k = 4$
(b) $\Theta = [0, 3.5] \times [1, 3.5], \mathcal{X} = [-5, 5], k = 6$

Minimax D-optimal design are presented in King and Wong (2000).

(a) support: -0.429, 0.629, 1.871, 2.929
weight: 0.245, 0.255, 0.255, 0.245
(b) support: -0.35, 0.62, 1.39, 2.11, 2.88, 3.85
weight: 0.18, 0.21, 0.11, 0.11, 0.21, 0.18

Note the designs are symmetric.

Nested PSO for minimax D-optimal design

We use nested PSO to find minimax D-optimal design for these two cases.

For case (a),

- Outer loop # of particles: 32
  # of iterations: 100
- Inner loop # of particles: 64
  # of iterations: 50

Nested PSO for minimax D-optimal design

For case (b),

- Outer loop # of particles: 512
  # of iterations: 200
- Inner loop # of particles: 256
  # of iterations: 100

* Particle velocity, position and $-log\ det(\cdot)$ is updated 2.6×10^9 times.

Nested PSO, case (a)

$$-log\ det\left(\mathcal{I}\left(\xi^*, \theta^*\right)\right) = \min_{\xi \in \mathcal{X}} \max_{\theta \in \Theta} \left\{ -log\ det\left(\mathcal{I}\left(\xi, \theta\right)\right) \right\}$$

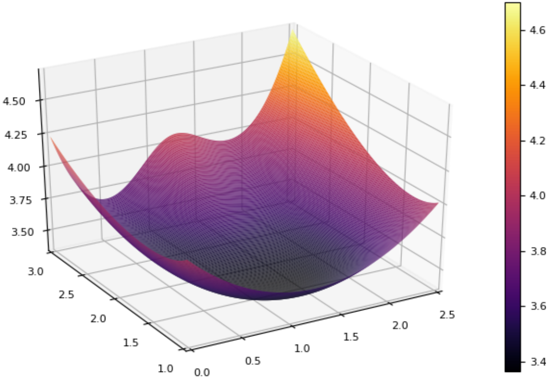
Results, case (a) :

- Outer loop support(ξ_i^*): -0.128, 1.126, 2.598, 4.0
  weight(w_i^*): 0.235, 0.553, 0.212, 0
- Inner loop parameter(θ^*): 2.5, 1
- Optimum $-log\ det\left(\mathcal{I}\left(\xi^*, \theta^*\right)\right) = 4.014$

Design is not symmetric.

Nested PSO, case (a)

Plotting $-log\ det\left(\mathcal{I}\left(\xi^{*},\theta\right)\right)$ as a function of θ



* Max: 4.700
* Min: 3.365

- PSO result

  Nested PSO: 4.014

  Simple PSO: 4.700

Nested PSO, case (b)

$$-log\ det\left(\mathcal{I}\left(\xi^{*},\theta^{*}\right)\right)=\min_{\xi\in\mathcal{X}}\max_{\theta\in\Theta}\left\{\ -log\ det\left(\mathcal{I}\left(\xi,\theta\right)\right)\ \right\}$$

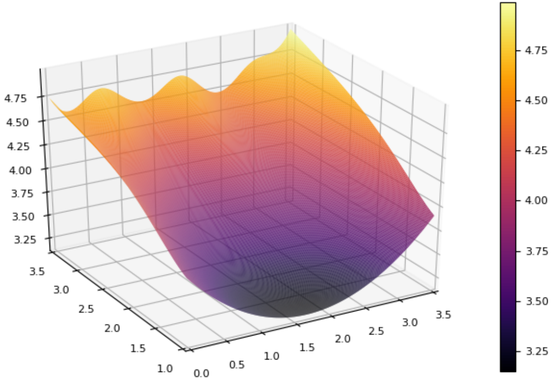
Results, case (b) :

- Outer loop support(ξ_i^*): -0.28, 0.66, 1.93, 2.88, 3,79, 5
  weight(w_i^*): 0.19, 0.25, 0.22, 0.18, 0.16, 0
- Inner loop parameter(θ^*): 0, 3.5
- Optimum $-log\ det\left(\mathcal{I}\left(\xi^{*},\theta^{*}\right)\right)=4.74$

Design is also not symmetric.

Nested PSO, case (b)

Plotting $-log\ det\left(\mathcal{I}\left(\xi^{*},\theta\right)\right)$ as a function of θ



* Max: 4.989
* Min: 3.149

- PSO result

  Nested PSO: 4.740

  Simple PSO: 4.989

Symmetric Design Constraint, case (a)

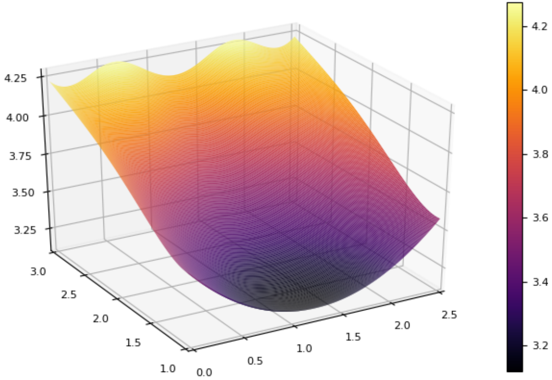
Include symmetric design constraint in the PSO algorithm

Results, case (a) :

- Outer loop support(ξ_i^*): -0.446282, 0.59195, 1.90805, 2.946282
  weight(w_i^*): 0.249949, 0.250051, 0.250051, 0.249949
- Inner loop parameter(θ^*): 2.5, 3
- Optimum $-log\ det\left(\mathcal{I}\left(\xi^*, \theta^*\right)\right) = 4.225$

Symmetric Design Constraint, case (a)

Plotting $-log\ det\left(\mathcal{I}\left(\xi^*, \theta\right)\right)$ as a function of θ



- Max: 4.275
- Min: 3.120

- PSO result

  Nested PSO: 4.225

  Simple PSO: 4.275*

  $*$: with 80% chance

Symmetric Design Constraint, case (a)

Relative efficiency against the worst case scenario:

$$RE(\xi) = \frac{\max_{\theta\in\Theta}\left\{|Cov(\hat{\beta}(\xi^*,\theta))|\right\}}{\max_{\theta\in\Theta}\left\{|Cov(\hat{\beta}(\xi,\theta))|\right\}}$$

Symmetric Design Constraint, case (b)

$$-log\ det\left(\mathcal{I}\left(\xi^{*},\theta^{*}\right)\right)=\min_{\xi\in\mathcal{X}}\max_{\theta\in\Theta}\left\{\ -log\ det\left(\mathcal{I}\left(\xi,\theta\right)\right)\ \right\}$$

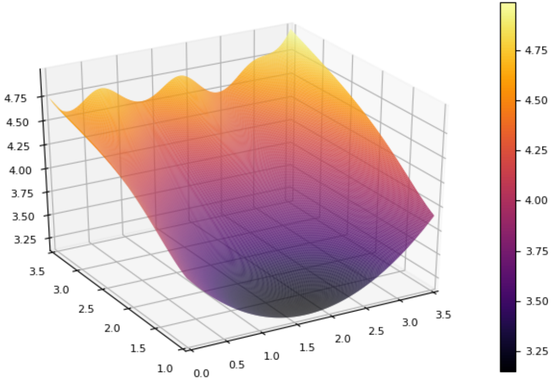
Results, case (b) :

- Outer loop support(ξ_i^*): -0.28, 0.66, 1.93, 2.88, 3,79, 5
  weight(w_i^*): 0.19, 0.25, 0.22, 0.18, 0.16, 0
- Inner loop parameter(θ^*): 0, 3.5
- Optimum $-log\ det\left(\mathcal{I}\left(\xi^{*},\theta^{*}\right)\right)=4.74$

Design is also not symmetric.

Symmetric Design Constraint, case (b)

Plotting $-log\ det\left(\mathcal{I}\left(\xi^{*},\theta\right)\right)$ as a function of θ



* Max: 4.989
* Min: 3.149

- PSO result

  Nested PSO: 4.740

  Simple PSO: 4.989

Learning Parameter

Fixed learning paramter $\gamma_1, \gamma_2 = 2$ might not fit every problem.

Compartment model example: $\mathcal{X} = [0, 30]$
Two-parameter logistic regression example: $\mathcal{X} = [-1, 4]$ and $[-5, 5]$.

The scale of w in probability simplex does not vary.

Since γ determines the diameter of particle movement in each iteration, learning parameter should be chosen according to the design space.

Inadequately large γ will make the particles move only along the boundary.

False Positive Error

Example 1 showed PSO algorithm may fail to find the global optimum.

Suppose, inner loop PSO algorithm fails and returns false $f_{outer}(u^x)$.

If $f_{outer}(u^x)$ is smaller than the current best position,
u_F and $f_{outer}(u_F)$ is saved in the solution path through the entire iteration.

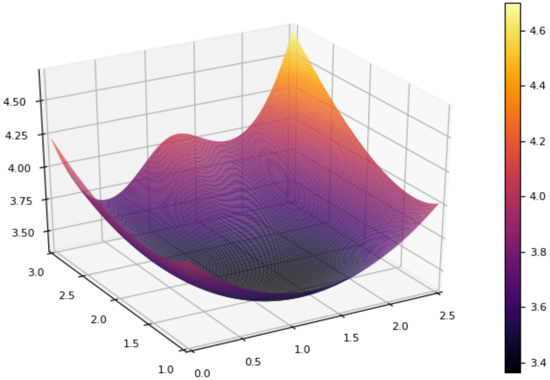
The output we observe is a nonexistent minimax point.

Consider a simple setting where such failure occurs with probability p.
Probability of a obtaining a solution free of $f_{outer}(u^x)$ after n iteration is $(1-(1-p)^n)$.

Large number of iteration in nested PSO guarantees failure.

Nested PSO algorithm is not suitable for minimax optimization problem.

Case (a): 1000 inner loop PSO simulation at ξ^*



(1) global max.
4.700 $[2.5, 3]$

(2) local max:
4.233 $[0, 3]$

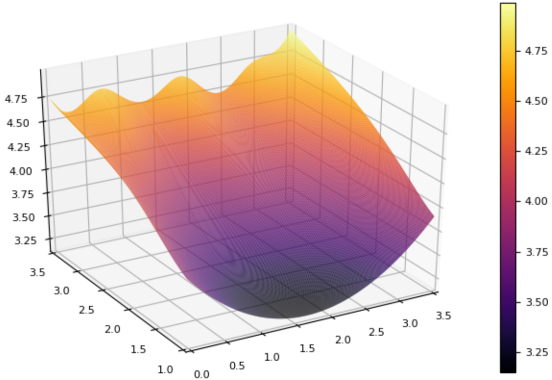
PSO Result

(1) 95%

(2) 5%

Stability Inside the Inner Loop

Case (b): 1000 inner loop PSO simulation at ξ^*



(1) global max.
4.989 $[3.5, 3.5]$
(2) local max:
4.740 $[0, 3.5]$

PSO Result
(1) 99.8%
(2) 0.2%

Modified Nested PSO

Since instability of simple PSO can raise a false alarm,
nested PSO that double checks before solution update may perform better.

(1) Initialize particles

(2) Repeat until stopping criteria are satisfied,

(2.1) Calculate particle velocity with Eq. (1)

(2.2) Update particle position with Eq. (2)

(2.3) Project particle back to design space

(2.4) Calculate $f_{outer}(x_i)$ via Algorithm (1)

(2.5) Check for failure and update best positions p_i, p_g

(3) Output $p_g = \underset{u \in \mathcal{U}}{\operatorname{argmin}} \{f_{outer}(u)\}$ with $gbest = f_{outer}(p_g)$

Modified Nested PSO

Results, Example 2 case (a):

Double check before global update

- Outer loop support(ξ_i^*): -1, 0.51, 1.70, 3.22
  weight(w_i^*): 0.21, 0.14, 0.54, 0.11
- Inner loop parameter(θ^*): 2.5, 1

Double check before personal and global update

- Outer loop support(ξ_i^*): -0.22, 0.52, 1.89, 2.82
  weight(w_i^*): 0.31, 0.15, 0.32, 0.22
- Inner loop parameter(θ^*): 1.77, 3

As long as "double check" is stochastic, modified nested PSO also fails.

Reference

- Berger, M. P., & Wong, W. K. (2009). An introduction to optimal designs for social and biomedical research (Vol. 83). John Wiley & Sons.
- Chen, R. B., Chang, S. P., Wang, W., Tung, H. C., & Wong, W. K. (2015). Minimax optimal designs via particle swarm optimization methods. Statistics and Computing, 25(5), 975-988.
- Berger, M. P., King, C. J., & Wong, W. K. (2000). Minimax D-optimal designs for item response theory models. Psychometrika, 65(3), 377-390.
- King, J., & Wong, W. K. (2000). Minimax D-optimal designs for the logistic model. Biometrics, 56(4), 1263-1267.
- Qiu, J., Chen, R. B., Wang, W., & Wong, W. K. (2014). Using animal instincts to design efficient biomedical studies via particle swarm optimization. Swarm and evolutionary computation, 18, 1-10.