Optimal Regression Design under Second-Order Least Squares Estimator: Theory, Algorithm and Applications

Chi-Kuang Yeh

Department of Mathematics and Statistics
University of Victoria

July 13th, 2018

Table of contents

Regression model

Consider a general regression model

$$y_i = g(\mathbf{x}_i; \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \cdots, n,$$

where

- y – response variable,
- $\mathbf{x} \in \mathbb{R}^p$ – independent (design) variable vector,
- $g(\mathbf{x}_i; \boldsymbol{\theta})$ – linear or nonlinear function of $\boldsymbol{\theta} \in \mathbb{R}^q$,
- ϵ_i's – i.i.d error with
  - mean 0,
  - variance σ^2

In design stage, we want to

- Find the best allocations of $\boldsymbol{x}$: $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_n}$
- Use θ from pilot studies

Optimal design depends on

- Estimator $\hat{\theta}$ of θ
- Measure of the covariance matrix $\mathbb{V}(\hat{\theta})$.

In design stage, we want to

- Find the best allocations of $\boldsymbol{x}$: $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_n}$
- Use $\boldsymbol{\theta}$ from pilot studies

Optimal design depends on

- Estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$
- Measure of the covariance matrix $\mathbb{V}(\hat{\boldsymbol{\theta}})$.

Question: Which estimator do we use?

- Ordinary least squares estimator (OLSE) is **BLUE.**

$$\hat{\theta} := \operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{n} (y_i - g(\boldsymbol{x_i}; \theta))^2.$$

- Second-order least squares estimator (SLSE) is **not linear.**

When the error distribution is asymmetric

SLSE is more efficient asymptotically than OLSE

Second-order least squares estimator (SLSE)

From Wang and Leblanc (2008), SLSE is defined as

$$(\hat{\boldsymbol{\theta}}^T, \hat{\sigma}^2)^T := \underset{\boldsymbol{\theta},\sigma^2}{\operatorname{argmin}} \sum_{i=1}^n \begin{pmatrix} y_i - g(\boldsymbol{x_i};\boldsymbol{\theta}) \\ y^2 - g^2(\boldsymbol{x_i};\boldsymbol{\theta}) - \sigma^2 \end{pmatrix}^T W(\boldsymbol{x_i}) \begin{pmatrix} y_i - g(\boldsymbol{x_i};\boldsymbol{\theta}) \\ y_i^2 - g^2(\boldsymbol{x_i};\boldsymbol{\theta}) - \sigma^2 \end{pmatrix}.$$

Define

$$\xi = \xi(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \dots & \boldsymbol{x_N} \\ p_1 & p_2 & \dots & p_N \end{bmatrix},$$

where $\boldsymbol{x_i} \in S \subset \mathbb{R}^p$.

Let θ_o and σ_o – true value of θ and σ.

Define, for any $\xi(\boldsymbol{x})$ of $\boldsymbol{x}$,

$$\boldsymbol{g_1} = \boldsymbol{g_1}(\xi, \boldsymbol{\theta_o}) = \mathbb{E}_\xi \left[\frac{\partial g(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_o}} \right]$$

and

$$\boldsymbol{G_2} = \boldsymbol{G_2}(\xi, \boldsymbol{\theta_o}) = \mathbb{E}_\xi \left[\frac{\partial g(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_o}} \right].$$

Covariance matrix

The asymptotic variance-covariance matrix of the SLSE, $\hat{\theta}$, is given by

$$\mathbb{V}(\hat{\theta}) = \sigma_o^2(1-t)(\boldsymbol{G_2} - t\boldsymbol{g_1}\boldsymbol{g_1}^T)^{-1},$$

where $t = \frac{\mu_3^2}{\sigma_o^2(\mu_4 - \sigma_o^4)}$.

Let

$$\boldsymbol{J} = \boldsymbol{J}(\xi, \boldsymbol{\theta_o}, t) = \boldsymbol{G_2} - t\boldsymbol{g_1}\boldsymbol{g_1}^T$$

and

$$\boldsymbol{B} = \boldsymbol{B}(\xi, \boldsymbol{\theta_o}, t) = \begin{pmatrix} 1 & \sqrt{t}\boldsymbol{g_1}^T \\ \sqrt{t}\boldsymbol{g_1} & \boldsymbol{G_2} \end{pmatrix}.$$

Optimal design problem

Minimization problem

$$
\begin{aligned}
\min_{p_1,\ldots,p_N} \quad & \phi(\xi) \\
\text{s.t.} \quad & \sum_{i=1}^{N} p_i = 1, \\
& -p_i \leq 0, \ \text{for } i = 1, 2, ..., N,
\end{aligned}
$$

where $\phi(\xi)$ – loss function depending on the design criterion.

Loss function

Loss functions for D-, A- and c-optimality criteria under SLSE can be expressed as

$$\begin{aligned}\phi_D(\xi, \boldsymbol{\theta_o}, t) &= \log(det(\boldsymbol{J}^{-1}(\xi, \boldsymbol{\theta_o}, t))),\\ \phi_A(\xi, \boldsymbol{\theta_o}, t) &= \mathrm{tr}(\boldsymbol{J}^{-1}(\xi, \boldsymbol{\theta_o}, t)),\\ \phi_c(\xi, \boldsymbol{\theta_o}, t) &= \boldsymbol{c_1}^T \boldsymbol{J}^{-1}(\xi, \boldsymbol{\theta_o}, t)\boldsymbol{c_1},\end{aligned}$$

where $\boldsymbol{c_1}$ is a given vector in $\mathbb{R}^q$.

If $\boldsymbol{J}$ is singular, all the three loss functions are defined to be $+\infty$.

Properties

Lemma 1

If $\boldsymbol{J}$ is invertible, then

$$\phi_D(\xi, \boldsymbol{\theta_o}, t) = \log(\det(\boldsymbol{J}^{-1})) = \log(\det(\boldsymbol{B}^{-1})).$$

Lemma 2

If $\boldsymbol{J}$ is invertible, then

$$\phi_A(\xi, \boldsymbol{\theta_o}, t) = \mathrm{tr}(\boldsymbol{J}^{-1}) = \mathrm{tr}(\boldsymbol{C}^T \boldsymbol{B}^{-1} \boldsymbol{C}),$$

where $\boldsymbol{C} = 0 \oplus \boldsymbol{I_q}$, $\boldsymbol{I_q}$ denotes for the $q \times q$ identity matrix, and $\oplus$ denotes for matrix direct sum operator.

Loss functions for D-, A- and c-optimality criteria are given as

$$\phi_D(\xi, \boldsymbol{\theta_o}, t) = \log(\det(\boldsymbol{B}^{-1}(\xi, \boldsymbol{\theta_o}, t))),$$
$$\phi_A(\xi, \boldsymbol{\theta_o}, t) = \mathrm{tr}(\boldsymbol{C}^T \boldsymbol{B}^{-1}(\xi, \boldsymbol{\theta_o}, t)\boldsymbol{C}),$$
$$\phi_c(\xi, \boldsymbol{\theta_o}, t) = \boldsymbol{c}^T \boldsymbol{B}^{-1}(\xi, \boldsymbol{\theta_o}, t)\boldsymbol{c},$$

where

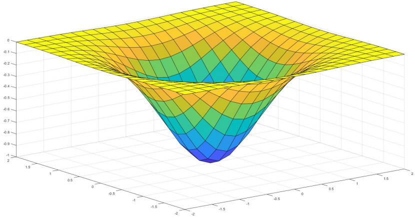
- $\boldsymbol{C} = 0 \oplus \boldsymbol{I_q}$,
- $\boldsymbol{c_1} \in \mathbb{R}^q$,
- $\boldsymbol{c}^T = (0, \boldsymbol{c_1}^T)$

Note: If $\boldsymbol{B}$ is singular, all the three loss functions are defined to be $+\infty$.

The loss functions

- $\phi_D(\xi, \boldsymbol{\theta_o}, t)$
- $\phi_A(\xi, \boldsymbol{\theta_o}, t)$
- $\phi_c(\xi, \boldsymbol{\theta_o}, t)$

are convex functions of $p_1, p_2, \ldots, p_N$.



Define $\boldsymbol{f}(x, \boldsymbol{\theta_o}) = \frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_o}}$, and

$$\boldsymbol{M}(\boldsymbol{x}) = \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta_o}, t) = \begin{pmatrix} 1 & \sqrt{t}\boldsymbol{f}^T(\boldsymbol{x}, \boldsymbol{\theta_o}) \\ \sqrt{t}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta_o}) & \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta_o})\boldsymbol{f}^T(\boldsymbol{x}, \boldsymbol{\theta_o}) \end{pmatrix}_{(q+1)\times(q+1)},$$

then $\boldsymbol{B}(\xi, \boldsymbol{\theta_o}, t) = \mathbb{E}_\xi[\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{\theta_o}, t)]$.

Define functions

$$\begin{aligned} d_D(\boldsymbol{x}, \xi, t) &= \mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{M}(\boldsymbol{x})) - (q+1) \\ d_A(\boldsymbol{x}, \xi, t) &= \mathrm{tr}(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{B}^{-1}\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{B}^{-1}) - \mathrm{tr}(\boldsymbol{C}^T\boldsymbol{B}^{-1}\boldsymbol{C}) \\ d_c(\boldsymbol{x}, \xi, t) &= \boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{M}(\boldsymbol{x})\boldsymbol{B}^{-1}\boldsymbol{c} - \boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{c} \end{aligned}$$

Optimal condition based on Equivalence theorem

Theorem 1

Let ξ_D^*, ξ_A^* and ξ_c^* be the optimal probability measures for D-, A- and c-optimality, respectively. Then $\boldsymbol{B}$ is invertible and for any $\boldsymbol{x} \in S$,

$$d_D(\boldsymbol{x}, \xi_D^*, t) \leq 0,$$
$$d_A(\boldsymbol{x}, \xi_A^*, t) \leq 0,$$
$$d_c(\boldsymbol{x}, \xi_c^*, t) \leq 0.$$

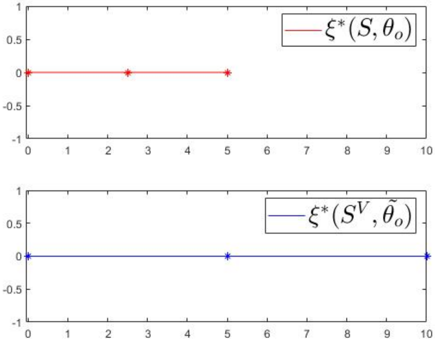
Note

Equality holds at the **support points** x_i! (i.e. $p_i > 0$)

Scale Invariance

For D-optimal design only.
Suppose we have $S, \boldsymbol{\theta}_o$

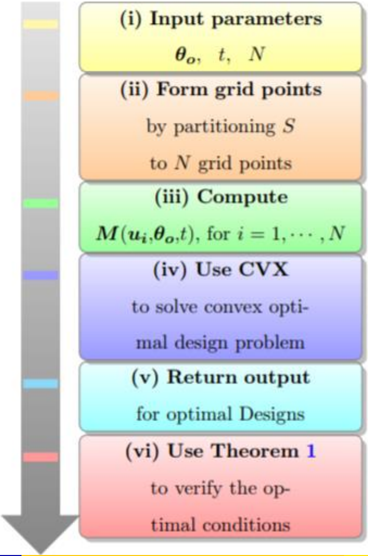
- Can obtain a design $\xi^*_D(S, \boldsymbol{\theta}_o)$
- Sometimes, easier to compute in scaled design space $S^{\boldsymbol{V}}$



Scale transform $S^{\boldsymbol{V}}, \tilde{\boldsymbol{\theta}}_o$

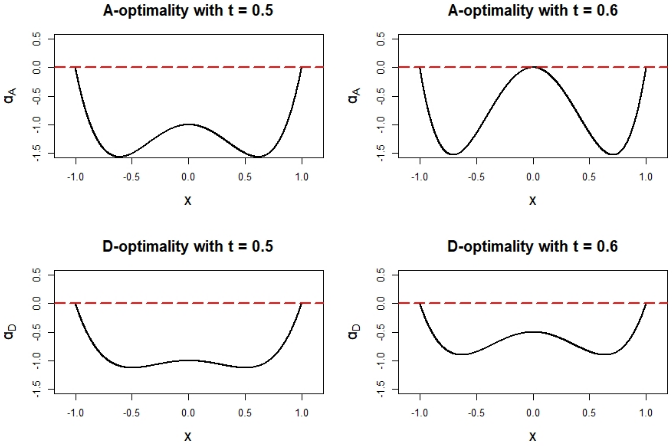
- Can obtain another design $\xi^*_D(S^V, \tilde{\boldsymbol{\theta}}_o)$

Computing algorithm



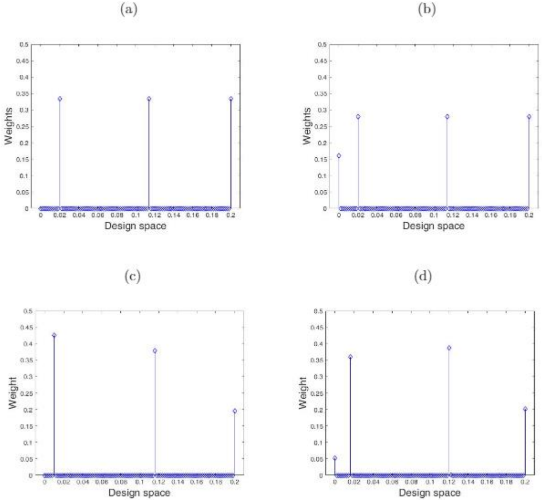
Example: Polynomial model of order 2

Model: $y_i = \theta_1 x_i + \theta_2 x_i^2 + \epsilon_i$, $x_i \in S = [-1, +1]$



Example: A fractional polynomial model

Model: $y = \theta_1 x + \theta_2 x^{\frac{1}{2}} + \theta_3 x^2 + \epsilon, \quad x \in (0.0, 0.2]$.



Number of support points

Parameter t plays an important role!

- Michaelis-Menton model:
  $y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \epsilon_i, \;\; x_i \in [0, k_o], \;\; \theta_1, \theta_2 \geq 0.$

  $n_D = 2$ or $3, \quad n_A = 2$ or $3.$

- Peleg model:
  $y_i = y_o + \frac{x_i}{\theta_1 + \theta_2 x_i} + \epsilon_i, \;\; x_i \in [0, b].$
  $n_D = 2$ or $3, \quad n_A = 2$ or $3.$

- Trigonometric model:
  $y_i = \sum_{j=1}^{k} [cos(jx_i)\theta_{1j} + sin(jx_i)\theta_{2j}] + \epsilon_i, \;\; S = [-b, b]$ where $b \in (0, \pi)$.
  $n_D = k$ or $k + 1, \quad n_A = k$ or $k + 1.$

Example – Mixture model

Consider an experiment by mixing three solutions x_1, x_2 and x_3. The linear model is given by

$$y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_3^2 + \theta_7 x_1 x_2 + \theta_8 x_1 x_3 + \epsilon,$$

where
$(x_1,\ x_2,\ x_3) \in S = \{(x_1,\ x_2,\ x_3) | x_1 + x_2 + x_3 \leq 1,\ x_i \geq 0,\ i = 1, 2, 3\}$.

When $N_1 = N_2 = N_3 = 21$, there are $N = 1771$ points in S_N.

When $N_1 = N_2 = N_3 = 51$, there are $N = 23424$ points in S_N.

Table: D-optimal designs for mixture model with $N = 23424$ grid points for t = 0.0 and 0.7.

Support points	Weights	
(x_1, x_2, x_3)	t = 0.0	t = 0.7
(0.00, 0.00, 0.50)	0.083	0.085
(0.00, 0.00, 1.00)	0.125	0.124
(0.00, 0.50, 0.00)	0.083	0.085
(0.00, 0.50, 0.50)	0.083	0.085
(1.00, 0.00, 0.00)	0.125	0.124
(0.50, 0.00, 0.00)	0.125	0.124
(0.50, 0.50, 0.00)	0.125	0.124
(0.50, 0.00, 0.50)	0.125	0.124
(1.00,0.00, 0.00)	0.125	0.124
ϕ_D	30.211	31.350

Example – Piecewise polynomial regression using knots

Model (Dette el al., 2008)
$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 + \theta_5 (x - \lambda)_+^3 + \epsilon, \; x \in [0, b]$

1. $S = [0, 10]$ with $\lambda = 8$, $\boldsymbol{B}(\xi, \boldsymbol{\theta_o}, t)$ is ill-conditioned. Failed!
2. Instead, using $S^V = [0, 1]$ with $\lambda = 0.8$ and obtain

$$\xi_D^*(S^V, \tilde{\boldsymbol{\theta_o}}) = \begin{bmatrix} 0.000 & 0.2250 & 0.5900 & 0.8200 & 0.9350 & 1.0000 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}.$$

Consistent with that in Dette et al. (2008).

4. Then we scale back and obtain

$$\xi_D^*(S, \boldsymbol{\theta_o}) = \begin{bmatrix} 0.000 & 2.2500 & 5.9000 & 8.2000 & 9.3500 & 10.0000 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}.$$

Table: Comparsion between 2 design spaces

$S = [0, 10],\ \lambda = 8$		$S^V = [0, 1],\ \lambda = 0.8$	
Support points	Weights	Support points	Weights
0.000	0.1667	0.000	0.1664
2.230	0.0002	0.2250	0.1664
2.240	0.0012	0.5900	0.1665
2.250	0.1413	0.8200	0.1669
2.260	0.0220	0.9350	0.1673
5.920	0.0003	1.0000	0.1666
5.930	0.0001		
8.200	0.1635		
8.210	0.0030		
9.340	0.0002		
9.350	0.1663		
10.000	0.1667		
$\log(\phi_D)$	-11.6064	Loss function	39.0607
Solved?	Inaccurate/Solved	Solved?	Solved

Conclusion and remarks

1. For simple models, we may find the results analytically.

2. For complicated models, we apply computational algorithms to find numerical results. We applied **CVX program in MATLAB** to solve the design problems.

3. The results can be easily extended to higher dimensional space S.

4. We have new results for c-optimality under the SLSE.
5. We have obtained equivalence theorem for the A-, c- and D-optimality criteria.
6. We have investigated the number of support points for various regression models.
7. We have an efficient and effective computing algorithm based on the CVX program for finding optimal designs.

THANK YOU!

- Polynomial model:
  $y_i = \theta_1 x_i + \theta_2 x_i^2 + ... + \theta_q x_i^q + \epsilon_i,$
  $x_i \in S = [-1, +1].$
  $n_D =$ q or q+1, $\quad n_A =$ q or q+1.

- A fractional polynomial model:
  $y = \theta_1 x + \theta_2 x^{\frac{1}{2}} + \theta_3 x^2 + \epsilon, \quad x \in (0.0, 0.2].$
  $n_D =$ 3, 4 or 5, $\quad n_A =$ 3, 4 or 5.

Application – Gompertz growth model

A biological model

Source: First Forest

The model

$$y_i = \theta_1 e^{-\theta_2 e^{-\theta_3 x_i}} + \epsilon_i,\ i = 1, 2, ..., n,\ \boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T,\ x_i \in S = [0, 5],$$

1. θ_1 – maximum growing capacity
2. θ_2 – initial status of the subject
3. θ_3 – growth rate
4. y is the overall growth at the current time point x

Table: D-optimal designs for Gompertz growth model with $t = 0.0$, $S = [0, 5]$, $\boldsymbol{\theta_o} = (1, 2, 3)^T$ and various values of N.

N	Support point (weight)	ϕ_D
51	0.100 (0.333), 0.600 (0.333), 5.000 (0.333)	11.120
101	0.050 (0.333), 0.550 (0.213), 0.600 (0.121), 5.000 (0.333)	11.106
501	0.060 (0.189), 0.070 (0.144), 0.580 (0.333), 5.000 (0.333)	11.101
1001	0.065 (0.333), 0.580 (0.333), 5.000 (0.333)	11.100
2001	0.065 (0.333), 0.578 (0.027), 0.580 (0.306), 5.000 (0.333)	11.100
5001	0.064 (0.129), 0.578 (0.204), 0.580 (0.333), 5.000 (0.333)	11.100
10001	0.065 (0.333), 0.579 (0.333), 5.000 (0.333)	11.100
20001	0.065 (0.333), 0.579 (0.333), 5.000 (0.333)	11.100

Table: A- and c-optimal design for Gompertz growth model with $\boldsymbol{\theta_o} = (1, 2, 3)^T$, $S = [0, 5]$, $\boldsymbol{c_1} = (2, 0.5, 1)^T$, $N = 2001$ and various values of t.

t		Support points (weights)	ϕ
0.0	c-	0.000 (0.204), 0.598 (0.575), 5.000 (0.221)	148.53
	A-	0.000 (0.471), 0.563 (0.331), 5.000 (0.198)	441.29
0.3	c-	0.000 (0.206), 0.595 (0.575), 5.000 (0.219)	149.98
	A-	0.000 (0.021), 0.003 (0.447), 0.568 (0.333), 5.000 (0.200)	464.52
0.5	c-	0.000 (0.211), 0.590 (0.574), 5.000 (0.216)	151.84
	A-	0.015 (0.457), 0.575 (0.340), 5.000 (0.204)	493.83
0.7	c-	0.000 (0.219) , 0.580 (0.572), 5.000 (0.209)	156.05
	A-	0.038 (0.438), 0.593 (0.351), 5.000 (0.211)	554.73
0.9	c-	0.000 (0.256), 0.538 (0.239), 0.540 (0.324), 5.000 (0.182)	174.18
	A-	0.095 (0.072), 0.098 (0.323), 0.638 (0.379), 5.000 (0.227)	791.57

Application – Peleg model

Often used in food studies

$$y_i = y_o + \frac{x_i}{\theta_1 + \theta_2 x_i} + \epsilon_i, \quad i = 1, 2, \cdots, n, \quad \boldsymbol{\theta_o} = (\theta_1, \theta_2)^T,$$

1. y_o – initial moisture of the food
2. θ_1 (≥ 0) – Peleg's moisture rate constant
3. θ_2 (≥ 0) – asymptotic moisture
4. y_i is the current level of moisture at current time x_i

Table: Loss functions and D-optimal designs for Peleg model with $\boldsymbol{\theta_o} = (0.5, 0.05)^T$, $S = [0, 180]$ and $N = 1001$ for $t = 0.0$, 0.3, 0.5, 0.7 and 0.9.

t	Support points (weights)	$\log(\phi_D)$	$d_D^{\max}$
0.0	9.000 (0.500), 180.00 (0.500)	-14.8774	8.4365e-7
0.3	9.000 (0.500), 180.00 (0.500)	-14.5207	1.2224e-6
0.5	9.000 (0.500), 180.00 (0.500)	-14.1843	3.7681e-6
0.7	0.000 (0.048), 9.000 (0.476), 180.00 (0.476)	-13.6812	5.2797e-6
0.9	0.000 (0.259), 9.000 (0.370), 180.00 (0.370)	-13.1786	1.6091e-5

Note: Peleg has GSI property.