Efficient global optimization: testing, reliability and efficiency

Julian Hall, Ken McKinnon and Theresia Mayer

School of Mathematics

University of Edinburgh

14^{th} October 2003

Overview

- Statistical background
- Methods for black-box global optimization
- Efficient global optimization (EGO): method of expected improvement
  - Testing
  - Reliability
  - Efficiency
- Conclusions

Expensive black-box objective

- Scalar function of d parameters.

$$y(\boldsymbol{x}), \;\; \boldsymbol{x} \in [0, 1]^d$$

- Search space dimension d is small.
- $y(\boldsymbol{x})$ is deterministic.
- $\boldsymbol{\nabla} y(\boldsymbol{x})$ may be available.
- Examples
  - Automotive/aeronautical design
  - Molecular modelling
  - and many, many more...

Stochastic model of $y(\boldsymbol{x})$

- Model $y(\boldsymbol{x})$ as
$$Y(\boldsymbol{x}) = \boldsymbol{f}_{\boldsymbol{x}}^T \boldsymbol{\beta} + \epsilon(\boldsymbol{x})$$
where
  - $\boldsymbol{f}_{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ are m known regression functions;
  - $E(\epsilon(\boldsymbol{x})) = 0$;
  - $Cov(\epsilon(\boldsymbol{x}), \epsilon(\bar{\boldsymbol{x}})) = \sigma_\epsilon^2 R(\boldsymbol{x}, \bar{\boldsymbol{x}})$, using
$$R(\boldsymbol{x}, \bar{\boldsymbol{x}}) = \exp\left(-\sum_{k=1}^{d} \theta_k |x_k - \bar{x}_k|^{p_k}\right)$$
- True values of parameters $\boldsymbol{\beta}$, σ_ϵ and $\boldsymbol{\theta}$ are not known.
Estimation procedure derived later
- For $y(\boldsymbol{x})$ smooth take $p_k = 2$, $k = 1, \ldots, d$.

Conditional model of $y(\boldsymbol{x})$

- Approximate $Y(\boldsymbol{x})$, conditional on values $\{y_i\}_{i=1}^n$ of $y(\boldsymbol{x})$ at **sample points** $\{\boldsymbol{x}^{(i)}\}_{i=1}^n$.
- Define $\boldsymbol{y} = [\, y_1 \quad \ldots \quad y_n \,]^T$.
- Approximate $Y(\boldsymbol{x})$ by best linear unbiased predictor (BLUP)

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{c_x}^T \boldsymbol{y}.$$

- BLUP is the linear predictor which minimizes mean squared error (MSE)

$$\mathrm{MSE}[\hat{y}(\boldsymbol{x})] = E([\boldsymbol{c_x}^T\boldsymbol{y} - Y(\boldsymbol{x})]^2) = (\boldsymbol{c_x}^T F\boldsymbol{\beta} - \boldsymbol{f_x}^T\boldsymbol{\beta})^2 + \sigma^2(1 - 2\boldsymbol{c_x}^T\boldsymbol{r_x} + \boldsymbol{c_x}^T R\boldsymbol{c_x}),$$

where

$$F = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}^{(1)}) \\ \vdots \\ \boldsymbol{f}(\boldsymbol{x}^{(n)}) \end{bmatrix} \in \mathbb{R}^{n\times m}, \boldsymbol{r_x} = \begin{bmatrix} R(\boldsymbol{x}, \boldsymbol{x}^{(1)}) \\ \vdots \\ R(\boldsymbol{x}, \boldsymbol{x}^{(n)}) \end{bmatrix} \in \mathbb{R}^n \quad \text{and} \quad R \in \mathbb{R}^{n\times n}$$

with

$$R_{ij} = R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \exp\left(-\sum_{k=1}^{d} \theta_k (x_k^{(i)} - x_k^{(j)})^2\right).$$

Best linear unbiased predictor (BLUP)

- Squared term in $\mathrm{MSE}[\hat{y}(\boldsymbol{x})]$ is zeroed for all $\boldsymbol{\beta}$ by imposing the **unbiasedness constraint**

$$F^T \boldsymbol{c_x} = \boldsymbol{f_x}$$

- $\boldsymbol{c_x}$ now found by solving

$$\begin{array}{lrcl} \min & \mathrm{MSE} & = & \sigma^2(1 - 2\boldsymbol{c_x}^T \boldsymbol{r_x} + \boldsymbol{c_x}^T R \boldsymbol{c_x}) \\ \text{subject to} & F^T \boldsymbol{c_x} & = & \boldsymbol{f_x} \end{array}$$

- Stationary point of the Lagrangian is given by solution of

$$\begin{bmatrix} 0 & F^T \\ F & R \end{bmatrix} \begin{bmatrix} -\boldsymbol{\lambda} \\ \boldsymbol{c_x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f_x} \\ \boldsymbol{r_x} \end{bmatrix},$$

for which $\boldsymbol{c_x}$ gives

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{c_x}^T \boldsymbol{y} = \begin{bmatrix} -\boldsymbol{\lambda} \\ \boldsymbol{c_x} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f_x} \\ \boldsymbol{r_x} \end{bmatrix}^T \begin{bmatrix} 0 & F^T \\ F & R \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{y} \end{bmatrix},$$

whence

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{f_x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{r_x}^T R^{-1}(\boldsymbol{y} - F\hat{\boldsymbol{\beta}}) \quad \text{where} \quad \hat{\boldsymbol{\beta}} = (F^T R^{-1} F)^{-1} F^T R^{-1} \boldsymbol{y}.$$

Mean squared error (MSE) for BLUP

- For the BLUP,

$$
\begin{aligned}
s^2(\boldsymbol{x}) &= \sigma^2(1-2\boldsymbol{c_x}^T\boldsymbol{r_x}+\boldsymbol{c_x}^T R\boldsymbol{c_x}) \\
&= \sigma^2\left(1-\begin{bmatrix}-\boldsymbol{\lambda} \\ \boldsymbol{c_x}\end{bmatrix}^T\begin{bmatrix}\boldsymbol{f_x} \\ \boldsymbol{r_x}\end{bmatrix}\right) \\
&= \sigma^2\left(1-\begin{bmatrix}\boldsymbol{f_x} \\ \boldsymbol{r_x}\end{bmatrix}^T\begin{bmatrix}0 & F^T \\ F & R\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{f_x} \\ \boldsymbol{r_x}\end{bmatrix}\right)
\end{aligned}
$$

Special case $f = 1$

- Common to use 1-dimensional basis of regression function $f(\boldsymbol{x}) = 1$ so
  - $F = \boldsymbol{e} = [\,1 \quad \ldots \quad 1\,]^T$.
  - $\hat{\boldsymbol{\beta}} = \hat{\mu}$.
- BLUP is then

$$\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{r}^T\boldsymbol{w} \quad \text{where} \quad \boldsymbol{w} = R^{-1}(\boldsymbol{y} - \hat{\mu}\boldsymbol{e}) \quad \text{and} \quad \hat{\mu} = \frac{\boldsymbol{e}^T R^{-1}\boldsymbol{y}}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}},$$

- MSE is

$$s^2(\boldsymbol{x}) = \sigma^2\left[1 - \boldsymbol{r}^T R^{-1}\boldsymbol{r} + \frac{(1 - \boldsymbol{e}^T R^{-1}\boldsymbol{r})^2}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}}\right] = \sigma^2\left(1 + \frac{1}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}}\right) + \boldsymbol{g}^T\boldsymbol{r} + \frac{1}{2}\boldsymbol{r}^T G\boldsymbol{r}$$

where

$$\boldsymbol{g} = -2\frac{R^{-1}\boldsymbol{e}}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}} \quad \text{and} \quad G = -2R^{-1} + 2\frac{(R^{-1}\boldsymbol{e})(R^{-1}\boldsymbol{e})^T}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}}$$

with

$$\boldsymbol{r} = [\,r_1 \quad \ldots \quad r_n\,]^T \quad \text{for} \quad r_i = \exp\left(-\sum_{k=1}^{d}\theta_k(x_k - x_k^{(i)})^2\right).$$

Parameter estimation

- The parameters σ and $\boldsymbol{\theta}$ are estimated by maximizing the likelihood function

$$L(\boldsymbol{y}) = |2\pi\sigma^2 R|^{-1/2} \exp\left(-\frac{(\boldsymbol{y}-\mu\boldsymbol{e})^T R^{-1}(\boldsymbol{y}-\mu\boldsymbol{e})}{2\sigma^2}\right)$$

or equivalently, and more conveniently, the log-likelihood function

$$\ln L(\boldsymbol{y}) = -\frac{1}{2}\left[n\ln(2\pi\sigma^2) + \ln(|R|) + \frac{(\boldsymbol{y}-\mu\boldsymbol{e})^T R^{-1}(\boldsymbol{y}-\mu\boldsymbol{e})}{\sigma^2}\right]$$

from whose derivatives come the following maximum likelihood estimates (MLE) of μ and σ^2:

$$\hat{\mu} = \frac{\boldsymbol{e}^T R^{-1}\boldsymbol{y}}{\boldsymbol{e}^T R^{-1}\boldsymbol{e}} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y}-\mu\boldsymbol{e})^T R^{-1}(\boldsymbol{y}-\mu\boldsymbol{e}).$$

- For these values of μ and σ^2, $\ln L(\boldsymbol{y})$ simplifies to

$$\ln L(\boldsymbol{y}) = -\frac{1}{2}\left[n\ln(\hat{\sigma}^2) + \ln(|R|)\right] + \text{constant terms}$$

- The optimal value of $\boldsymbol{\theta}$ is estimated by maximizing this function.

A general algorithmic structure

The BLUP $\hat{y}(\boldsymbol{x})$ and MSE $s^2(\boldsymbol{x})$ have been used in various ways to derive optimization methods with the following general structure

1. Sample according to an initial design $\{\boldsymbol{x}^{(i)}\}_{i=1}^{n}$
2. Estimate parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$ then form $\hat{y}(\boldsymbol{x})$ and $s^2(\boldsymbol{x})$.
3. Optimize a merit function $V(\hat{y}(\boldsymbol{x}), s^2(\boldsymbol{x}))$ to get 'best' new sample point $\boldsymbol{x}^{(n+1)}$.
4. Set $n := n + 1$. Go to 2.

A stopping criterion can be based on the value of $V(\boldsymbol{x}^{(n+1)})$.

Merit functions

Metit functions which have been maximized are

- $V(\boldsymbol{x}) = -\hat{y}(\boldsymbol{x})$.
- $V(\boldsymbol{x}) = -\hat{y}(\boldsymbol{x}) + ws(\boldsymbol{x})$, some w. Cox and John [1997]

  Torczon and Trosset [1999]
- $V(\boldsymbol{x}) = \Phi\left(\dfrac{y^* - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right)$, some y^*. Žilinskas [1985]

  Probability of exceeding a target value y^*.
- $V(\boldsymbol{x}) = (y_{\min} - \hat{y}(\boldsymbol{x}))\Phi\left(\dfrac{y_{\min} - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right) + s(\boldsymbol{x})\phi\left(\dfrac{y_{\min} - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right)$.

  Jones *et al.* [1998]

  Expected improvement (EI) on the current best function value $y_{\min}$.

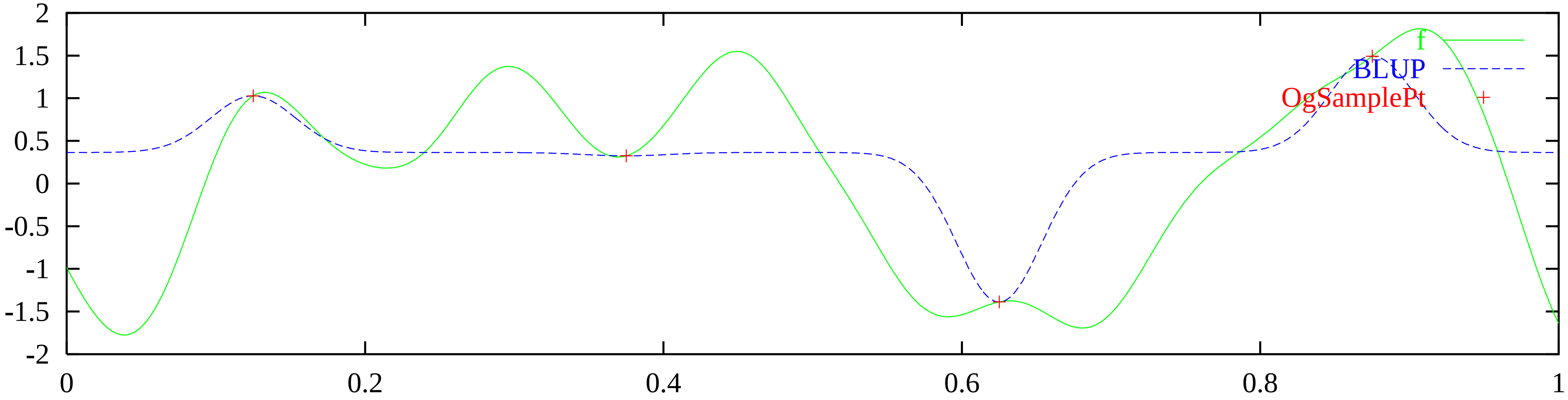
$\Phi(.)$ and $\phi(.)$ are the standard normal probability density and cumulative distribution functions.

Example

Using an initial design

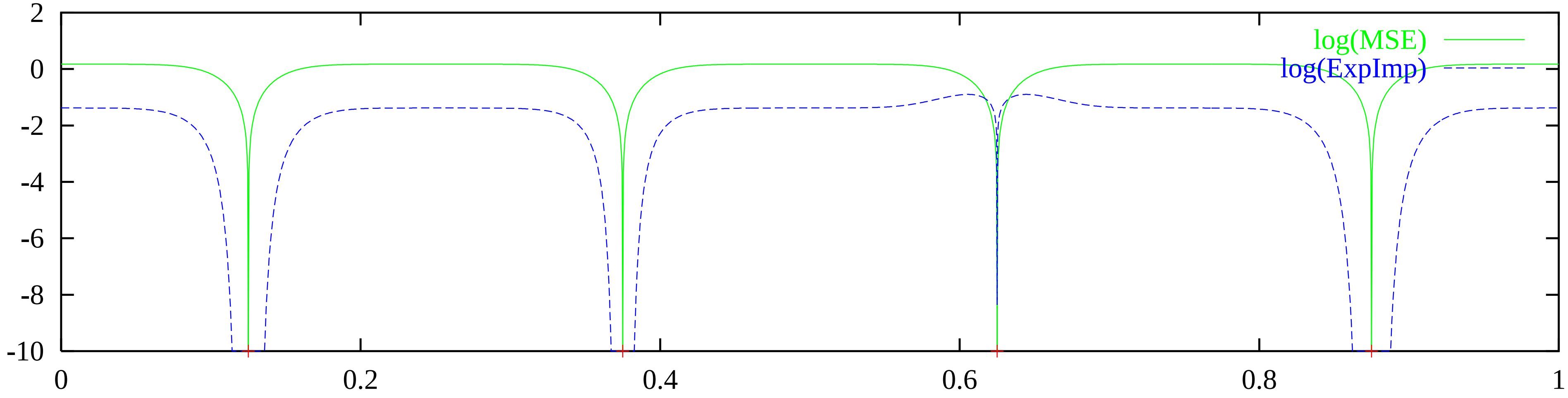
$$x^{(1)} = \frac{1}{8}, \quad x^{(2)} = \frac{3}{8}, \quad x^{(3)} = \frac{5}{8}, \quad x^{(4)} = \frac{7}{8}$$

the objective and BLUP are



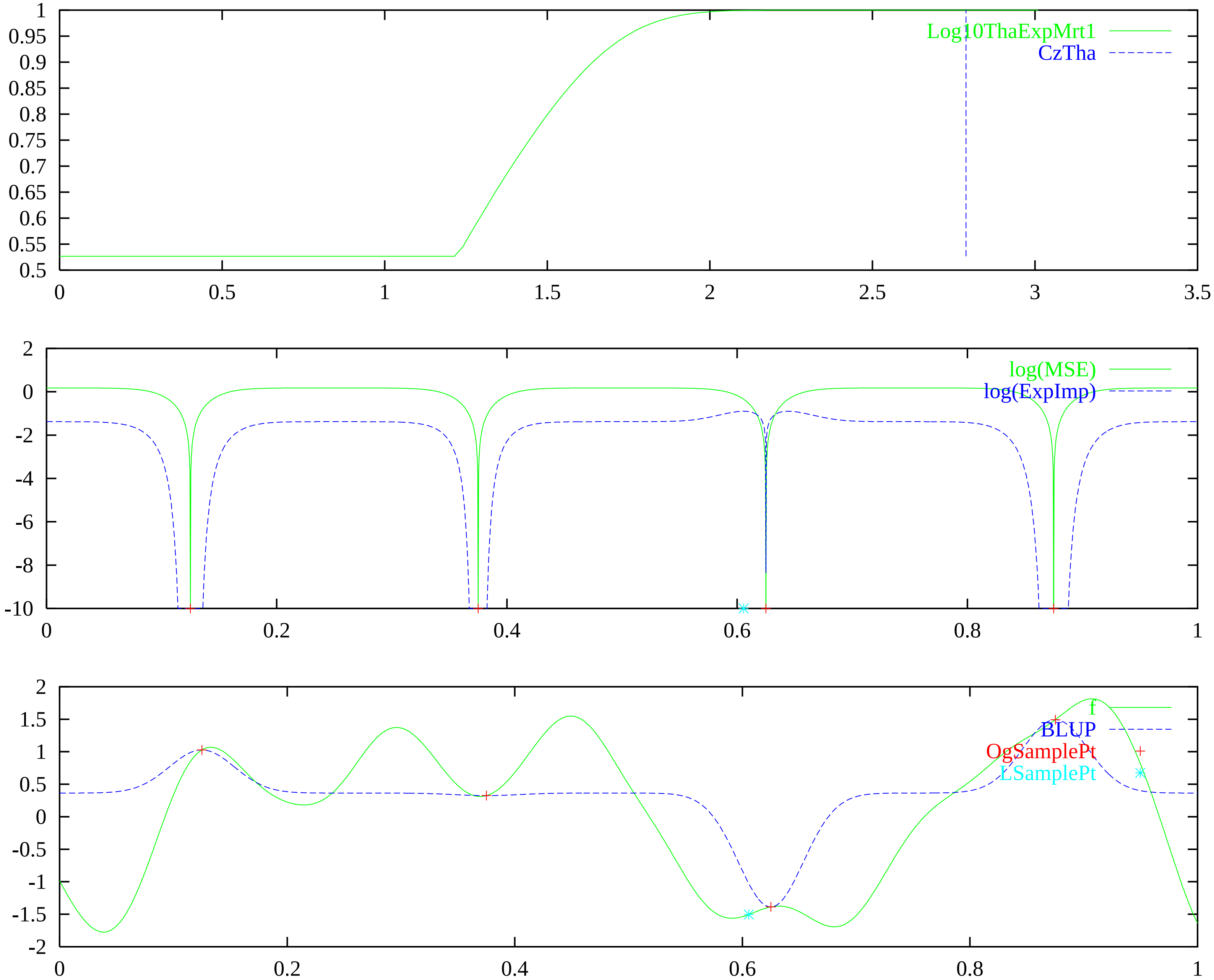
- Note the large MLE $\theta = 614$;
- Observe $\hat{y}(x^{(i)}) = y_i$, $i = 1, \ldots, 4$.

The corresponding (log) MSE and EI are

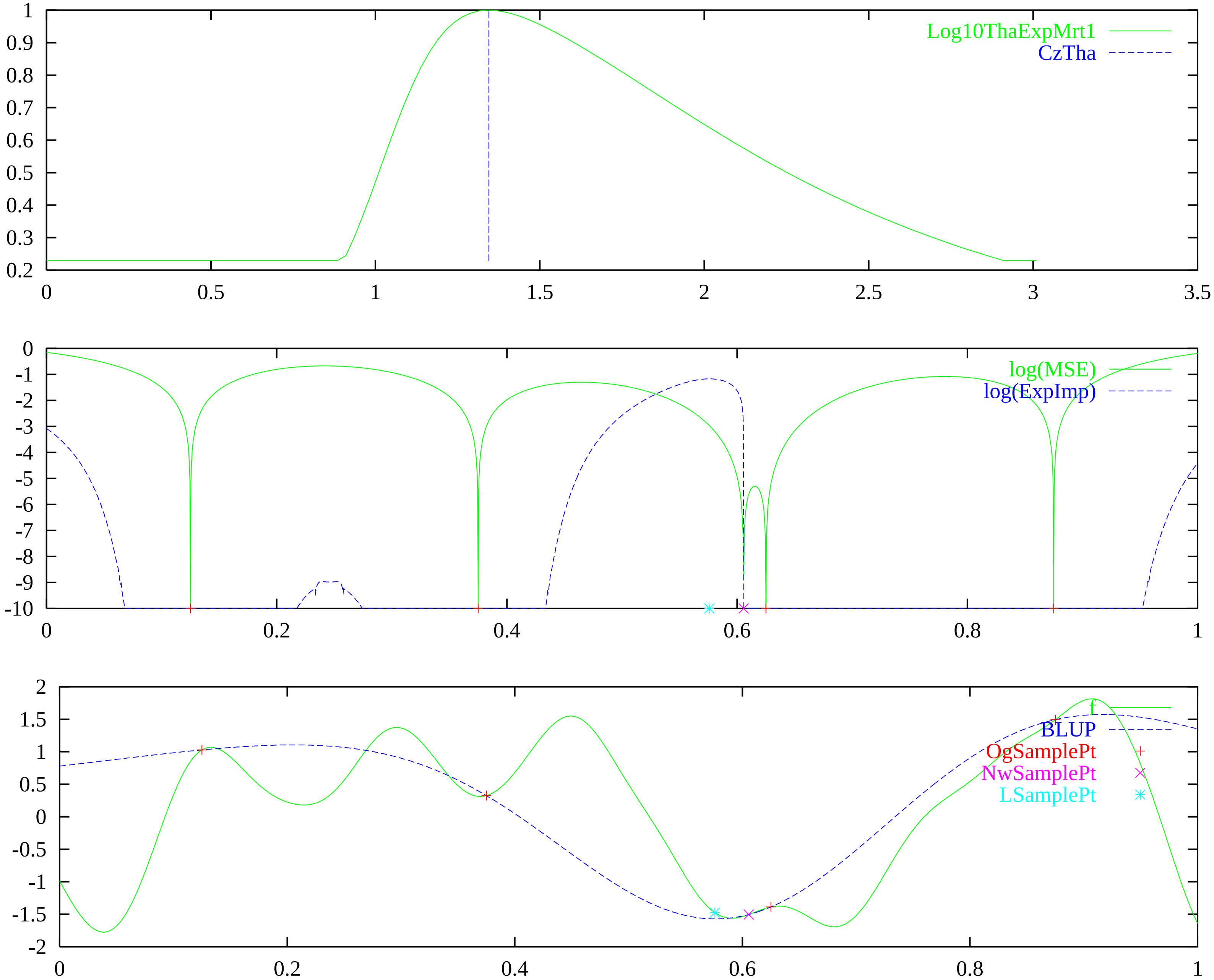


- Observe that $s^2(x^{(i)}) = EI(x^{(i)}) = 0$, $i = 1, \ldots, 4$.

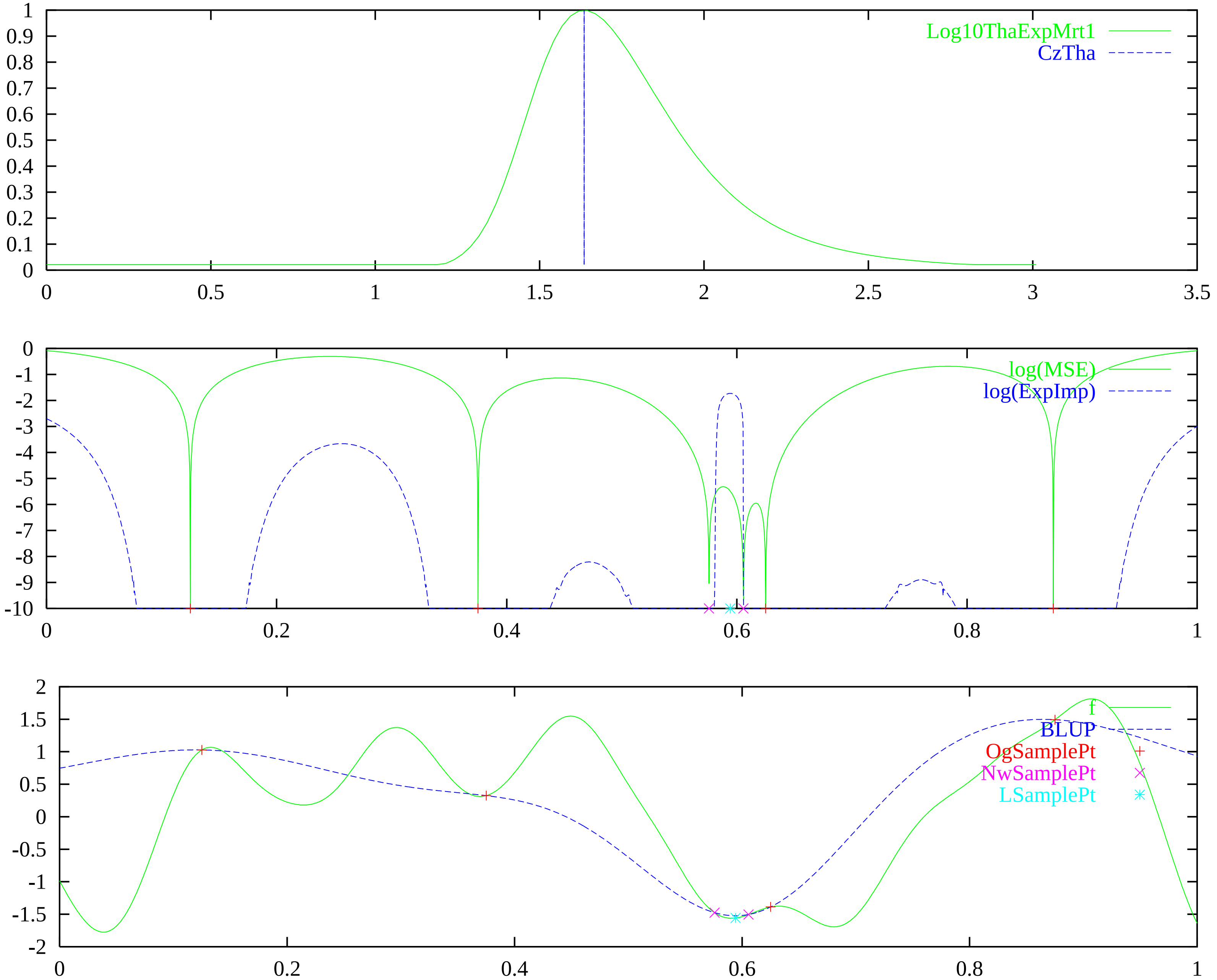
Iteration 1: $\theta = 614$



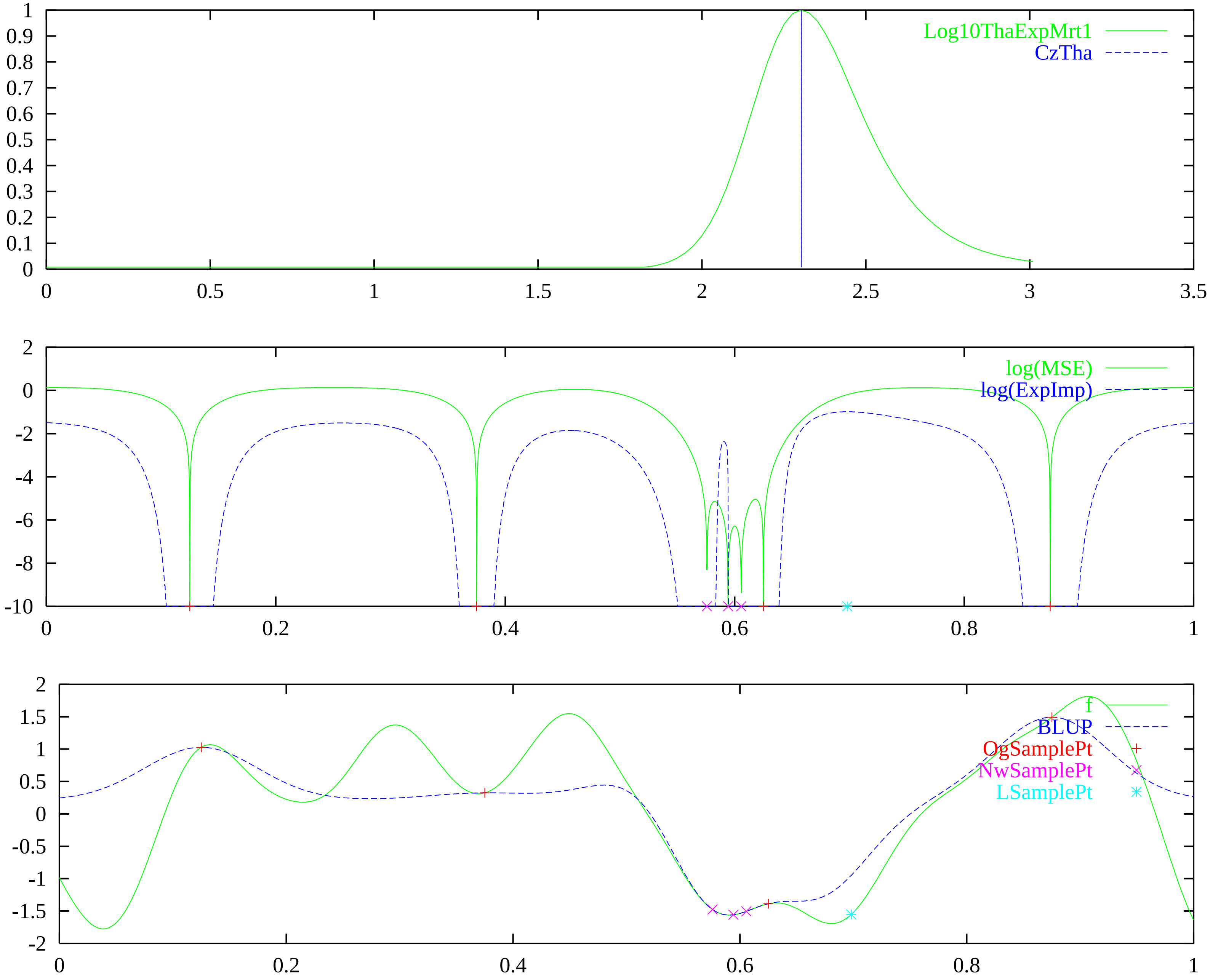
Iteration 2: $\theta = 22$



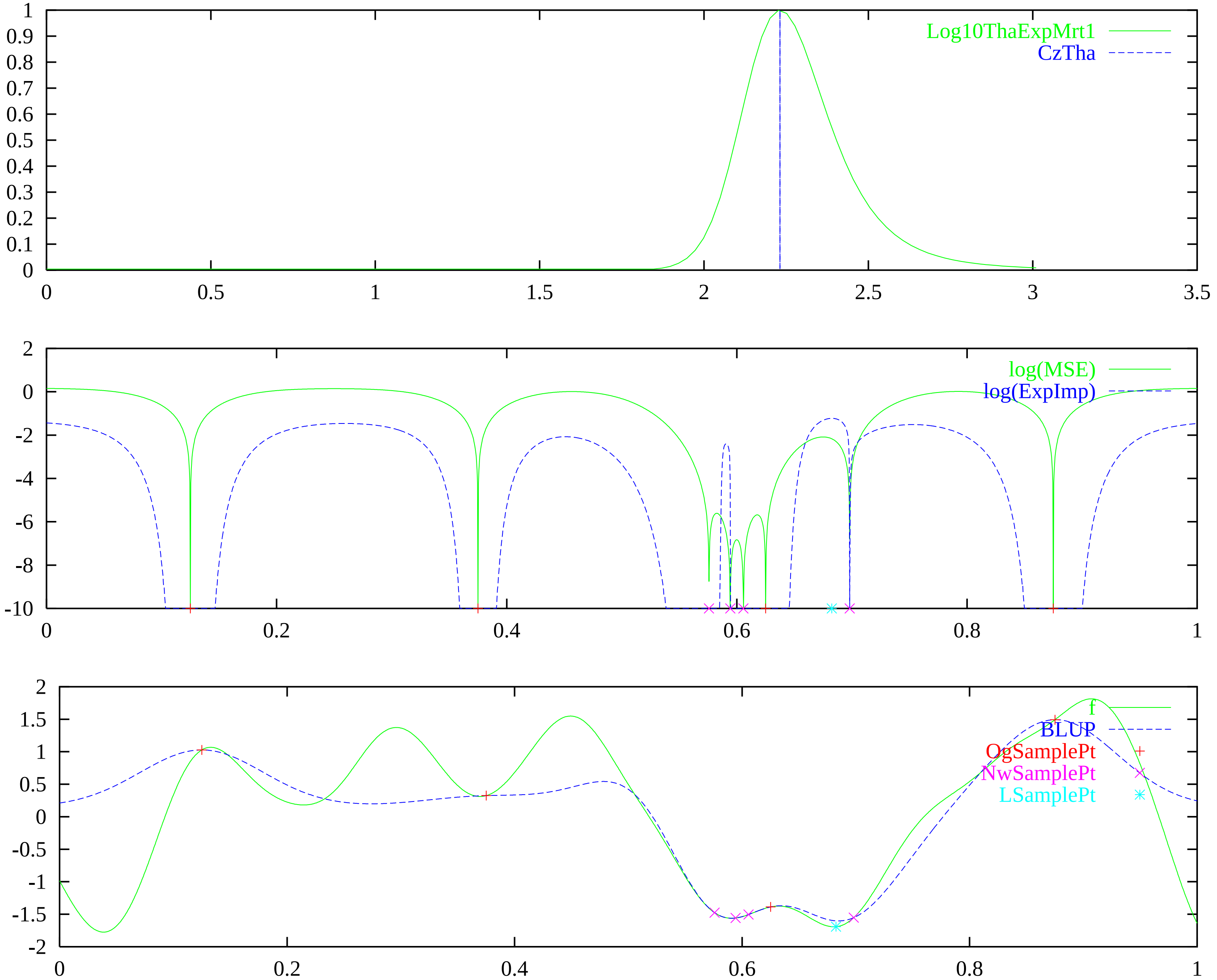
Iteration 3: $\theta = 43$



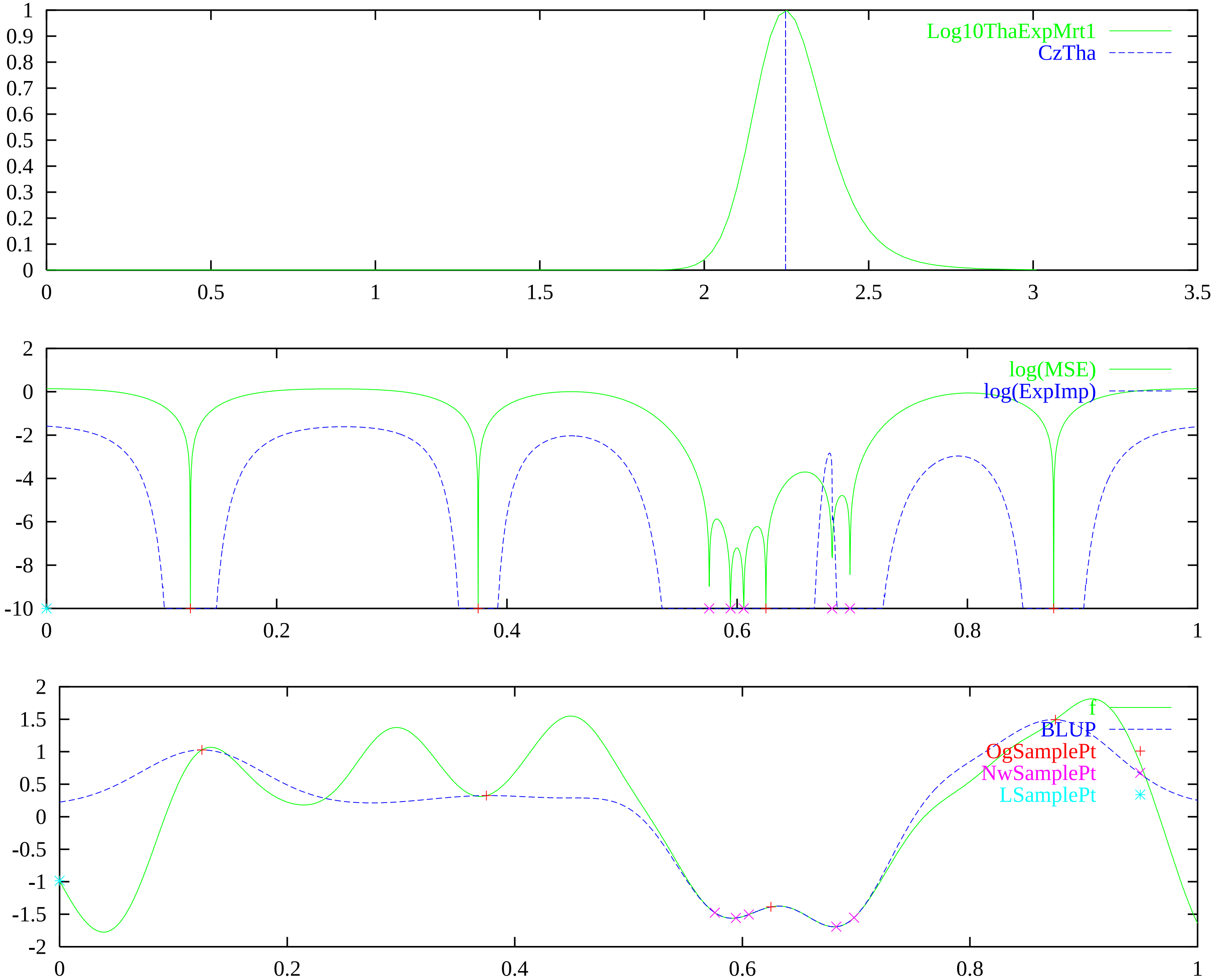
Iteration 4: $\theta = 201$



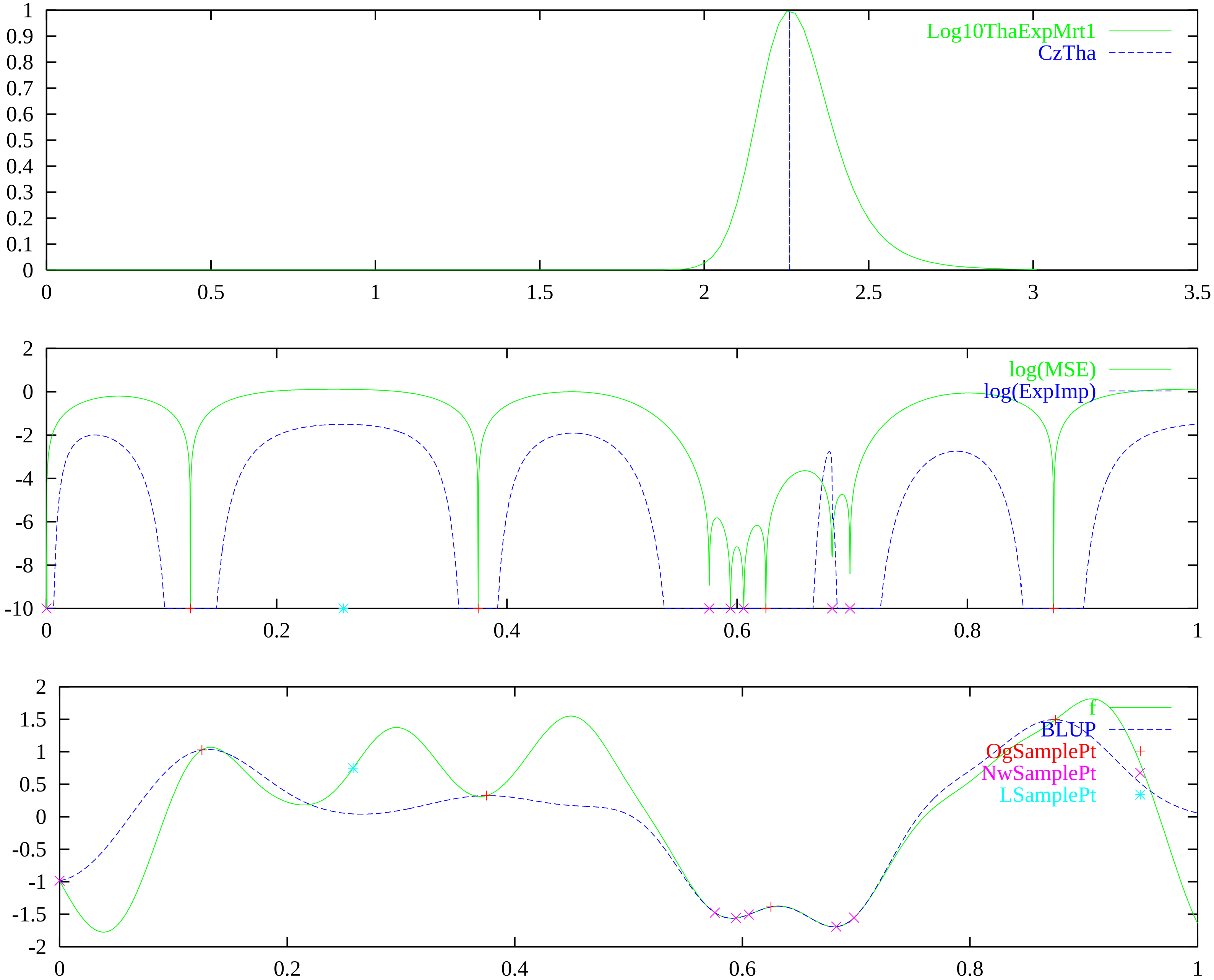
Iteration 5: $\theta = 170$



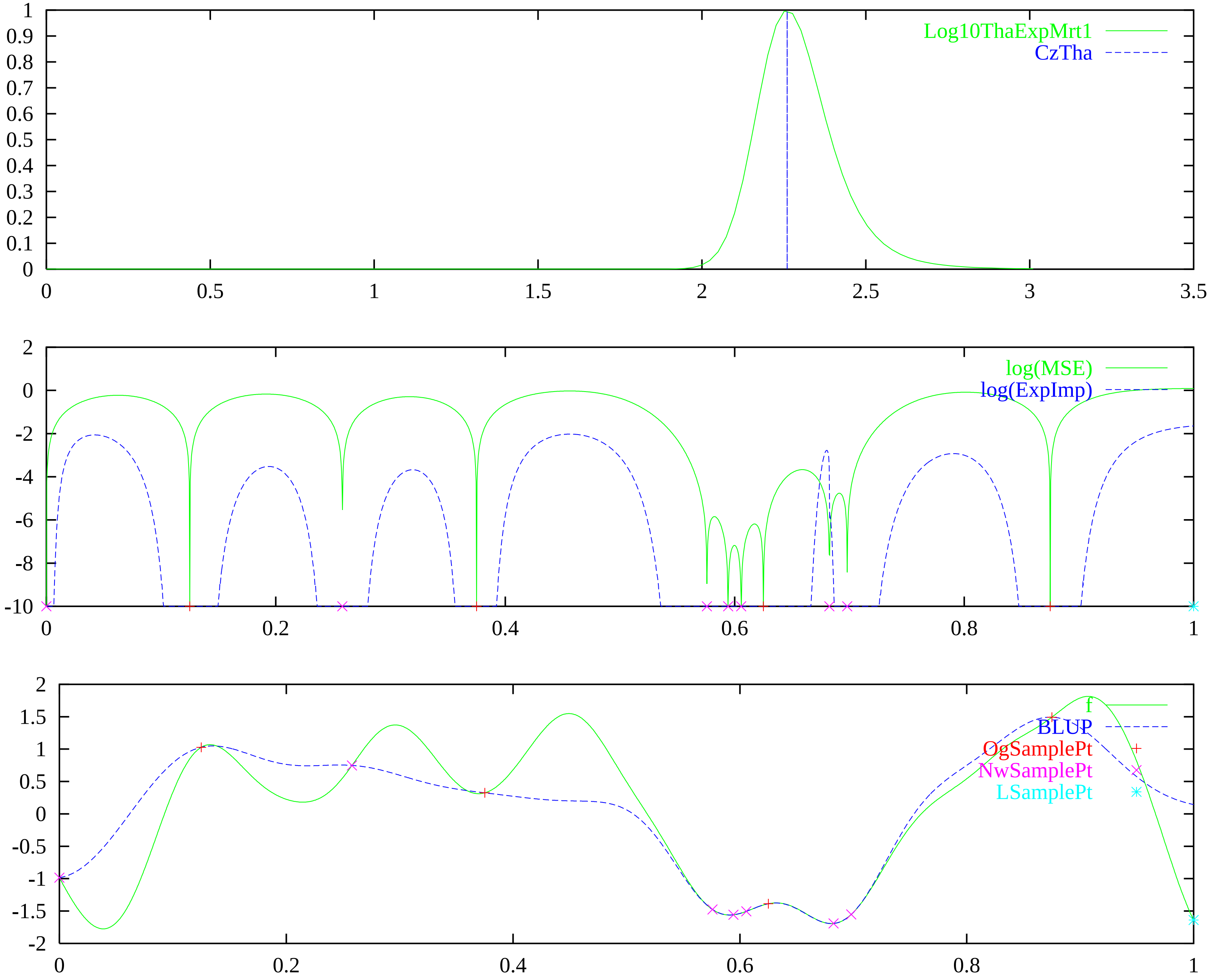
Iteration 6: $\theta = 177$



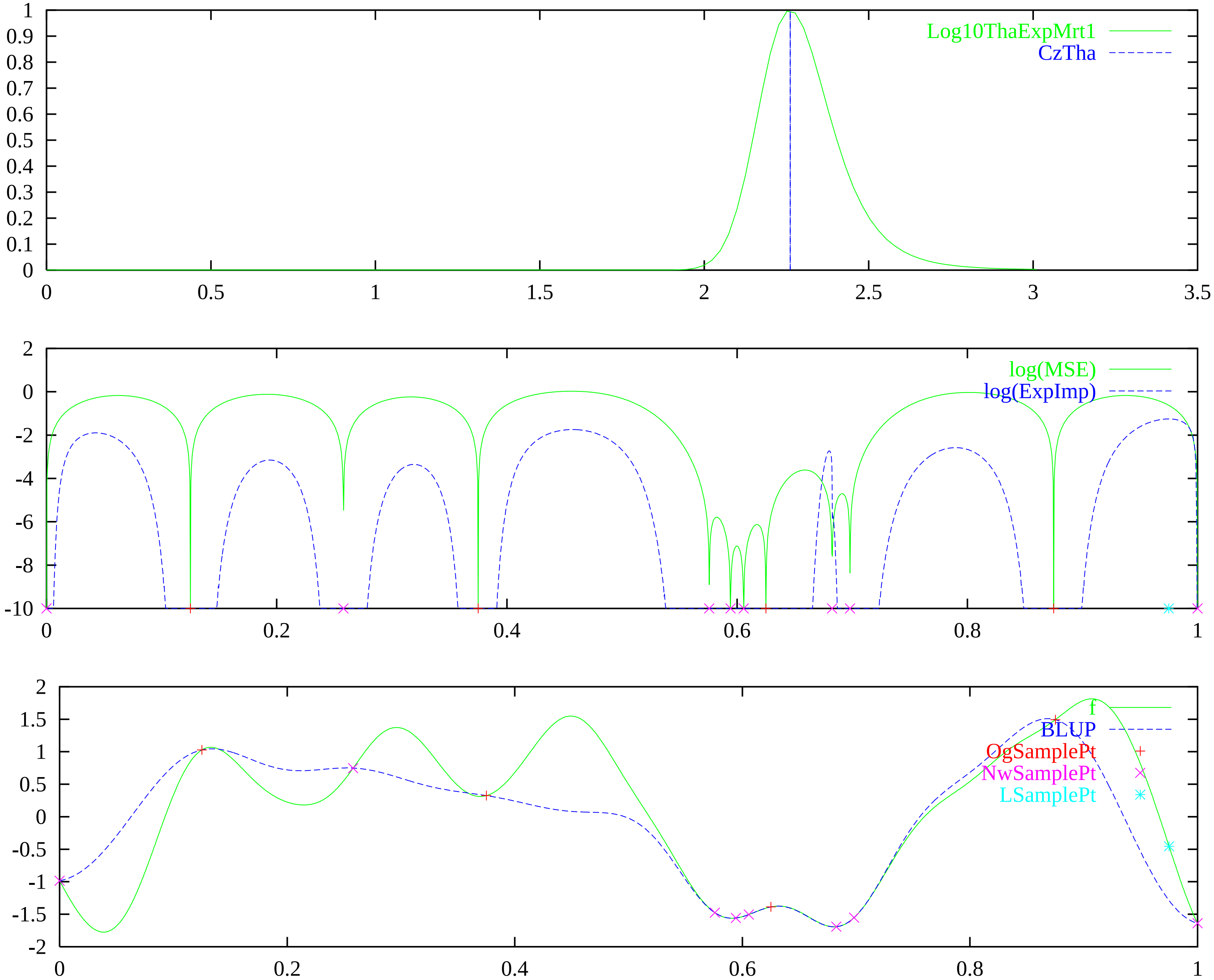
Iteration 7: $\theta = 182$



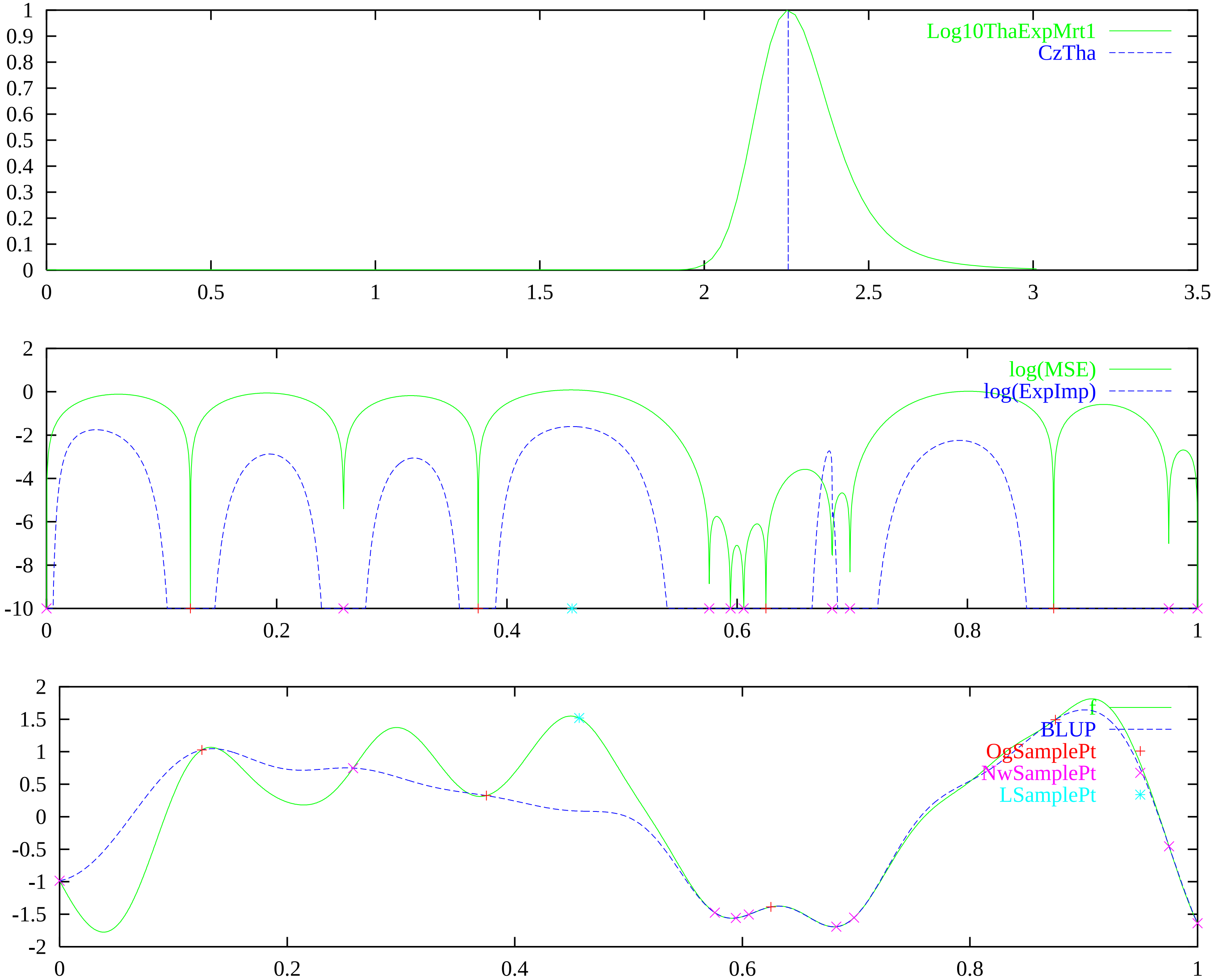
Iteration 8: $\theta = 182$



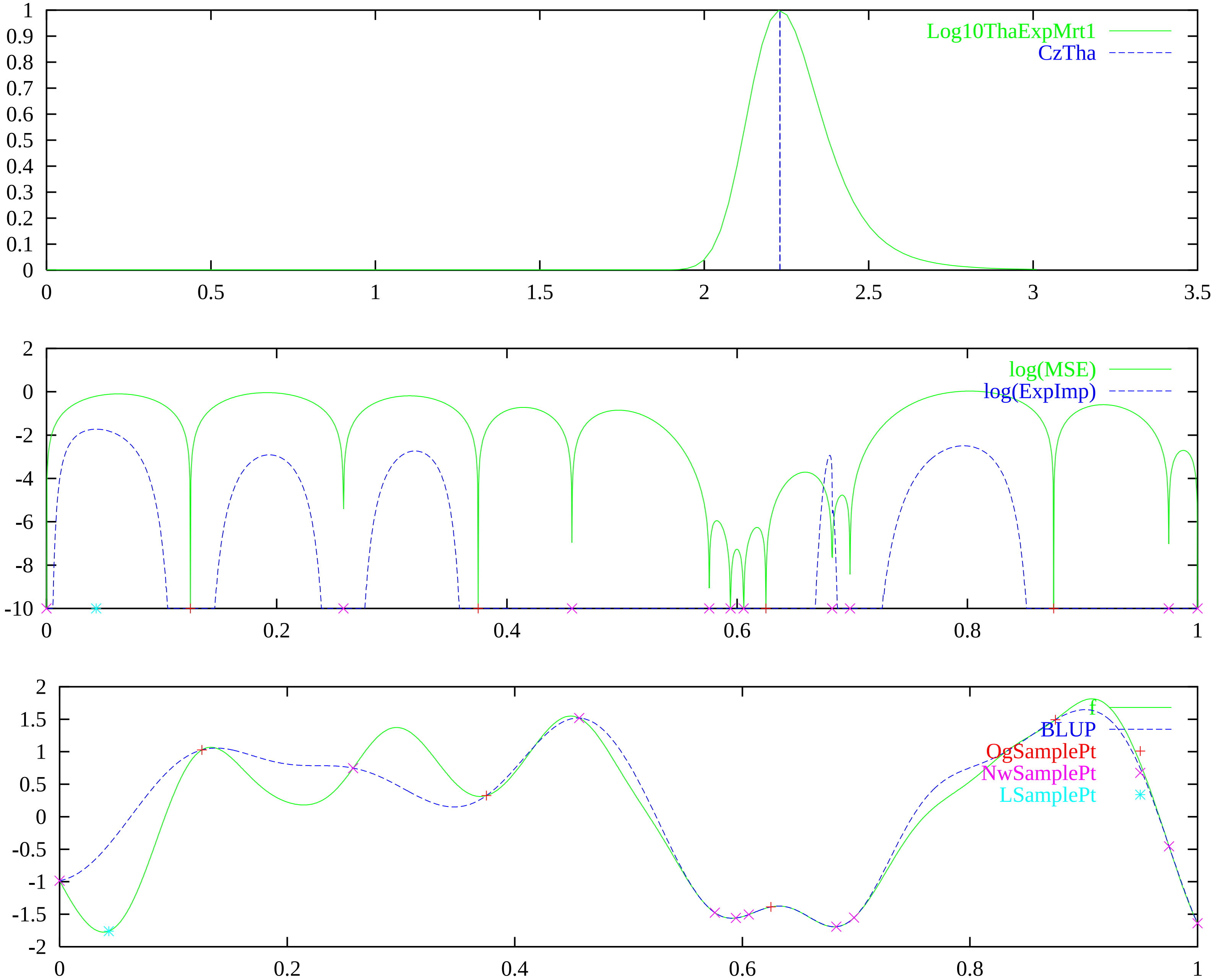
Iteration 9: $\theta = 182$



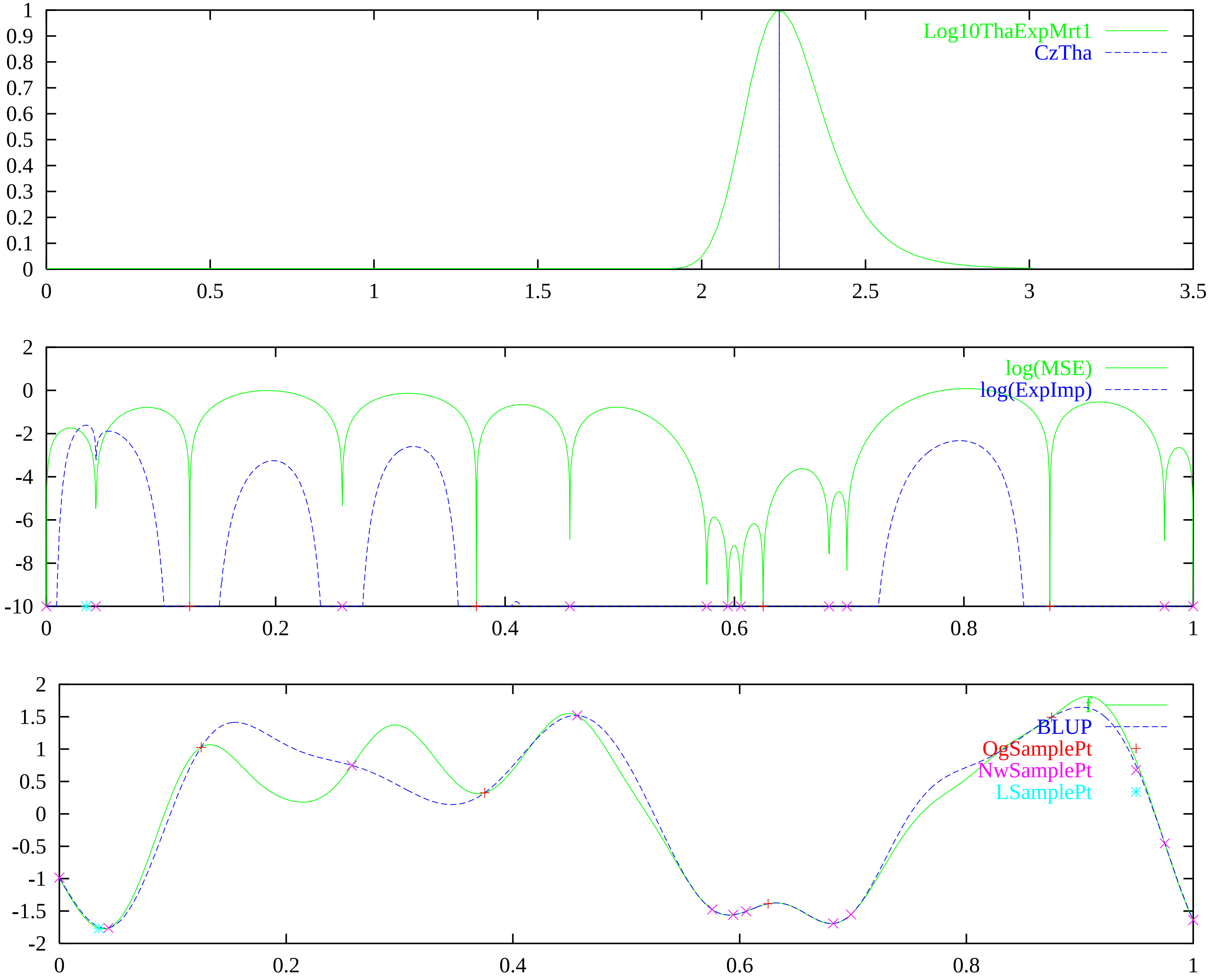
Iteration 10: $\theta = 180$



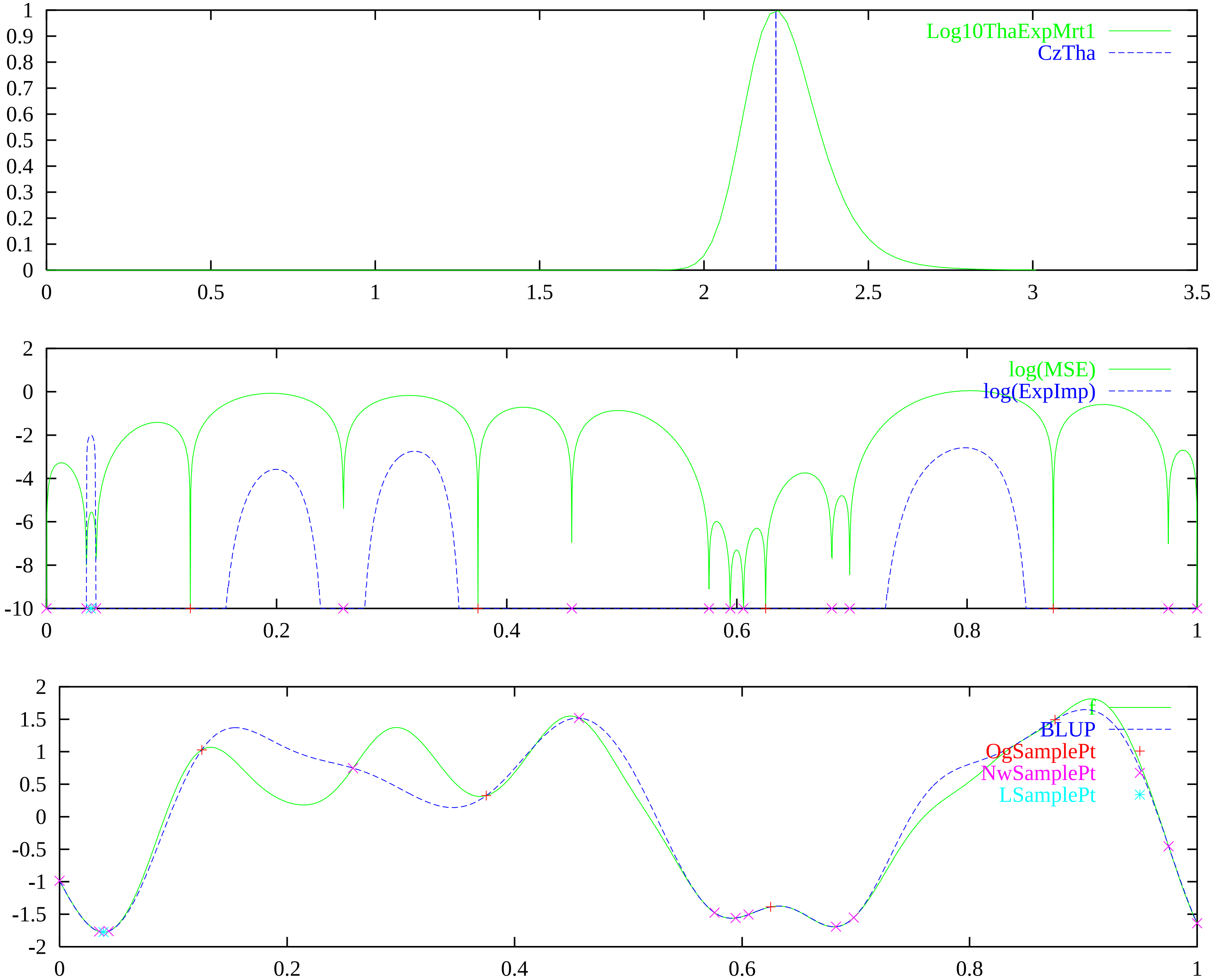
Iteration 11: $\theta = 170$



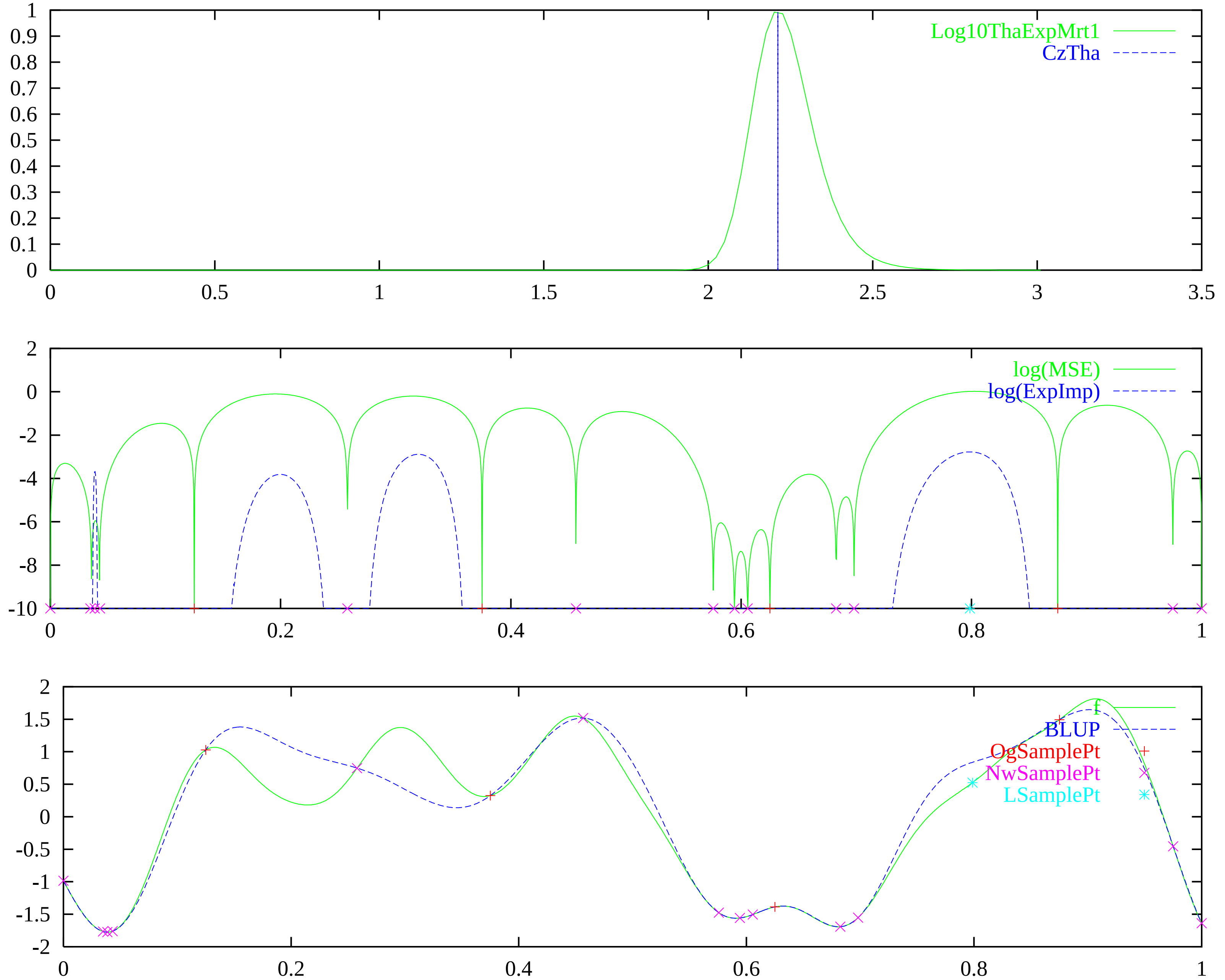
Iteration 12: $\theta = 173$



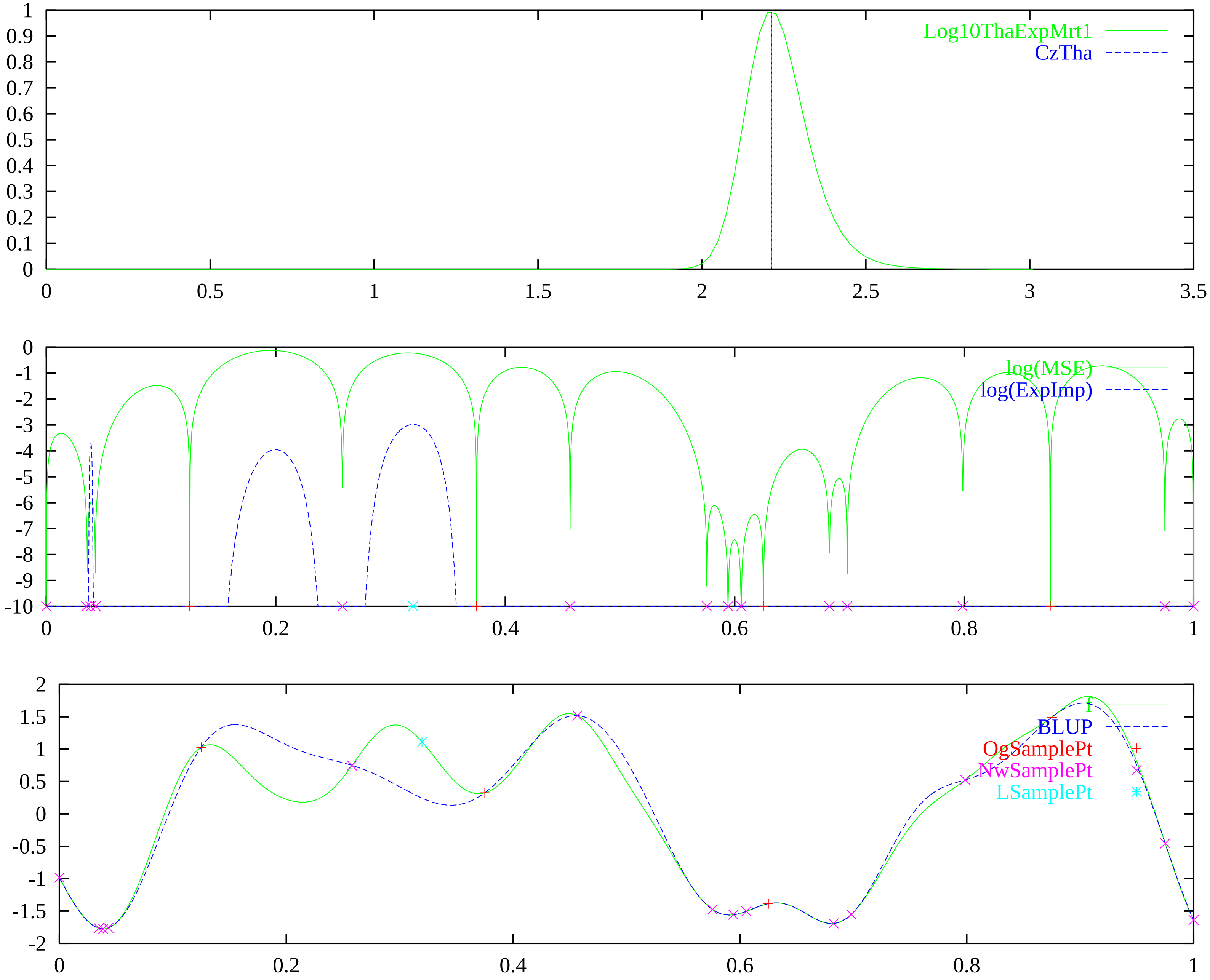
Iteration 13: $\theta = 165$



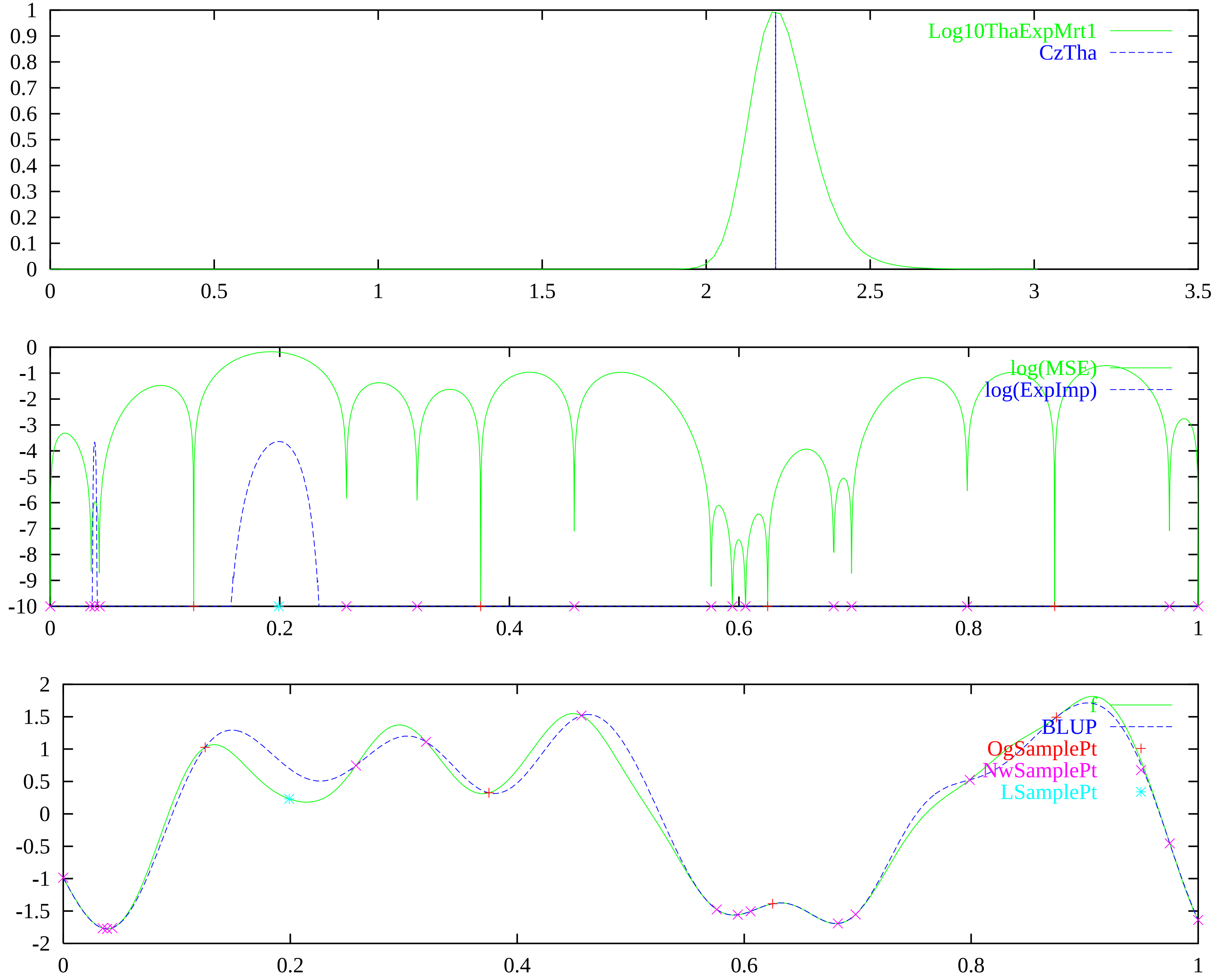
Iteration 14: $\theta = 163$



Iteration 15: $\theta = 163$



Iteration 16: $\theta = 163$



Testing

- Want to test the following
  - Validity of assumption that objective can be modelled as a realisation of a stochastic process
  - Accuracy of parameter estimation
  - Reliability of the method
- Other authors restrict testing to a very small number of problems
- **Aim:** Generate a large number of test problems corresponding to known parameter values.

  Realisations of stationary stochastic process with mean zero, unit variance and autocorrelation function

$$\rho_G(\boldsymbol{x}, \overline{\boldsymbol{x}}) = \prod_{j=1}^{d} \exp(-\theta_j (x_j - \overline{x}_j)^2)$$

  for particular values of $\theta_j \geq 0$, $j = 1, \ldots, d$.

Generating $Y(x)$ for $x \in I\!R$ and $\theta = 1$

- Observe that for $x \in I\!R$ and $\theta = 1$,

$$\rho_G(x, \overline{x}) = \exp(-(x - \overline{x})^2) = c(x - \overline{x}) \quad \text{where} \quad c(h) = e^{-h^2}$$

- Can show that, for any positive definite symmetric function $c(h)$,

$$Y(x) = \frac{1}{\sqrt{\pi}} \int_0^\infty \sqrt{\tilde{c}(\omega)} [N_1(\omega) \cos(x\omega) + N_2(\omega) \sin(x\omega)] d\omega,$$

  where

  - $\tilde{c}(\omega)$ is the Fourier transform of $c(h)$
  - $N_1(\omega)$ and $N_2(\omega)$ are independent standard normal processes

  is a strongly stationary process with zero mean and autocovariance function c.

Transformation using the Box-Muller method

- The Box-Muller method yields the transformation

$$\begin{aligned} N_1(\omega) &= L(\omega)\cos(V(\omega)2\pi) \\ N_2(\omega) &= -L(\omega)\sin(V(\omega)2\pi), \end{aligned}$$

  where

$$L(\omega) = \sqrt{-2\ln(U(\omega))},$$

  and $U(\omega)$ and $V(\omega)$ are independent and uniformly distributed on $[0,1]$.

- Applying the transformation gives

$$Y(x) = \frac{1}{\sqrt{\pi}} \int_0^\infty \sqrt{\tilde{c}(\omega)} L(\omega) \cos(x\omega + V(\omega)2\pi) d\omega.$$

- Test functions (realisations of $Y(x)$) correspond to realisations of $L(\omega)$ and $V(\omega)$.

Discretisation

- Discrete approximation to integral is obtained using points $\omega_n \in I\!R^+$

$$Y(x) = \frac{1}{\sqrt{\pi}} \sum_{n=0}^{\infty} \sqrt{a_n} \sqrt{\tilde{c}(\omega_n)} L_n \cos(x\omega_n + V_n 2\pi),$$

  where a_n is length of interval associated with ω_n.

- Using equally-spaced points, $\omega = nh$,

$$a_n = \begin{cases} h/2 & n = 0 \\ h & n > 0 \end{cases}$$

  so

$$Y(x) = \sqrt{\frac{h}{\pi}} \left[\frac{\sqrt{\tilde{c}(0)} L_0}{\sqrt{2}} + \sum_{n=1}^{\infty} \sqrt{\tilde{c}(nh)} L_n \cos(xnh + V_n 2\pi) \right].$$

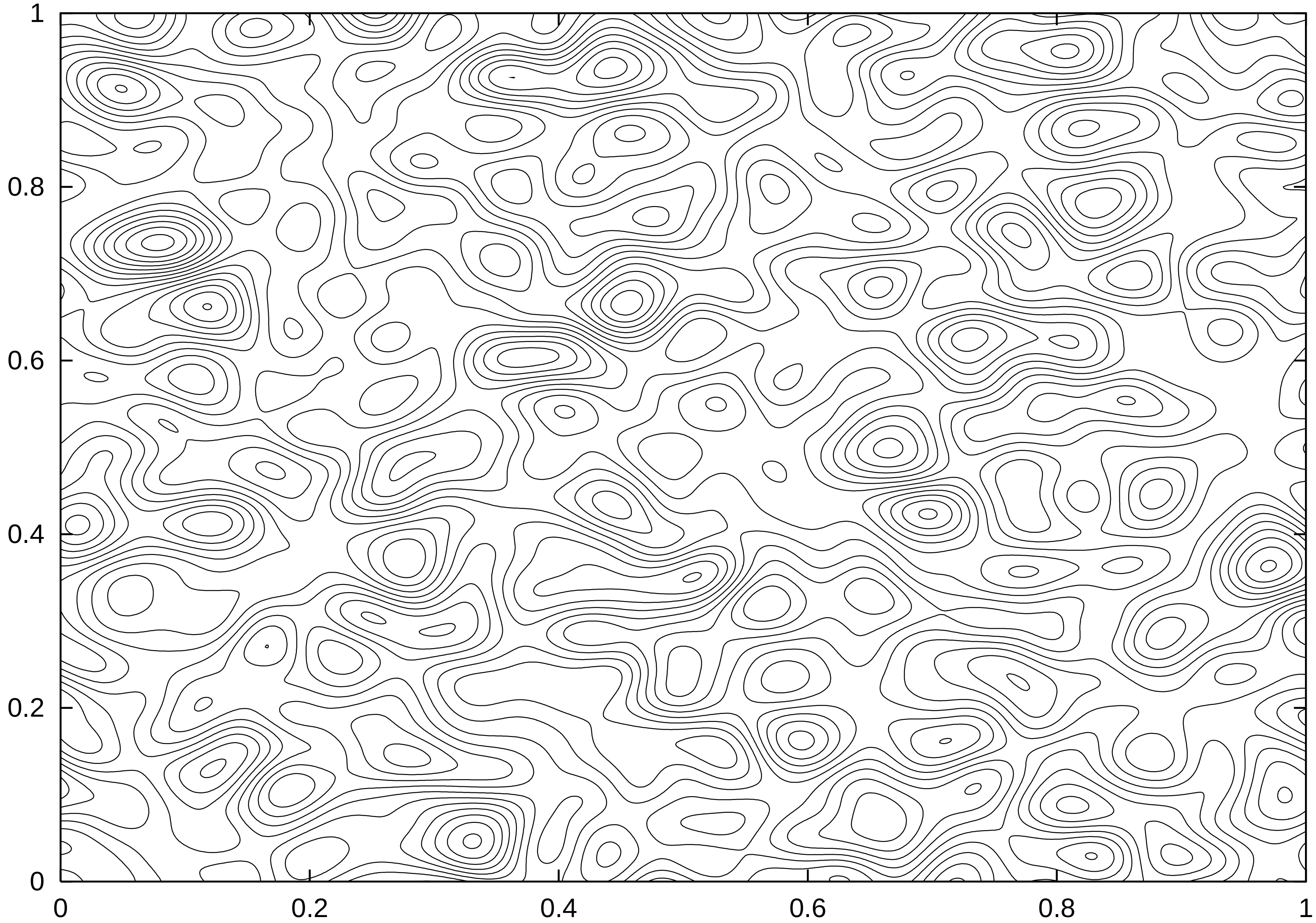
- Discrete expression for $Y(x)$ when $c(v - x) = e^{-(v-x)^2}$ has associated autocovariance function

$$c_D(v - x) = cov(Y(v), Y(x)) = \sum_{n=-\infty}^{\infty} exp\left(- \left(\frac{2\pi n}{h} + v - x \right)^2 \right).$$

A practical procedure

- $\tilde{c}(\omega) \to 0$ as $\omega \to \infty$ so restrict $\omega \leq \omega_{ub} = 10$ (for which $\tilde{c} = 1.2 \times 10^{-11}$).
- Fix h (and hence n) to minimize dominant term in error $c_D(v-x) - c(v-x)$.
- Analysis extends to general θ and $\boldsymbol{x} \in I\!R^n$.
- Practical techniques (currently) extended to $n = 2$.

An example



Test function with $\theta_1 = \theta_2 = 400$

Efficiency

- Cost of finding next sample point determined by
  - Cost of parameter estimation
  - Cost of maximizing $EI(\boldsymbol{x})$
- Cost of these operations *does* matter!
- Major cost is global optimization of $EI(\boldsymbol{x})$ by branch-and-bound
- Questionable whether it is worth optimizing $EI(\boldsymbol{x})$ globally

Global optimization of the expected improvement

- Use branch-and-bound, subdividing $[0,1]^d$ into boxes.
- For box $X \subseteq [0,1]^d$
  lower bound on $EI(X)$ is $EI(\boldsymbol{x})$, some $\boldsymbol{x} \in X$;
  upper bound on $EI(X)$ is available since
  - $EI(\hat{y}, s^2)$ decreases monotonically with $\hat{y}$;
  - $EI(\hat{y}, s^2)$ increases monotonically with s^2.

  so, if
  - $\hat{y}_L$ is lower bound on $\hat{y}(X)$;
  - s_U^2 is upper bound on $s^2(X)$;

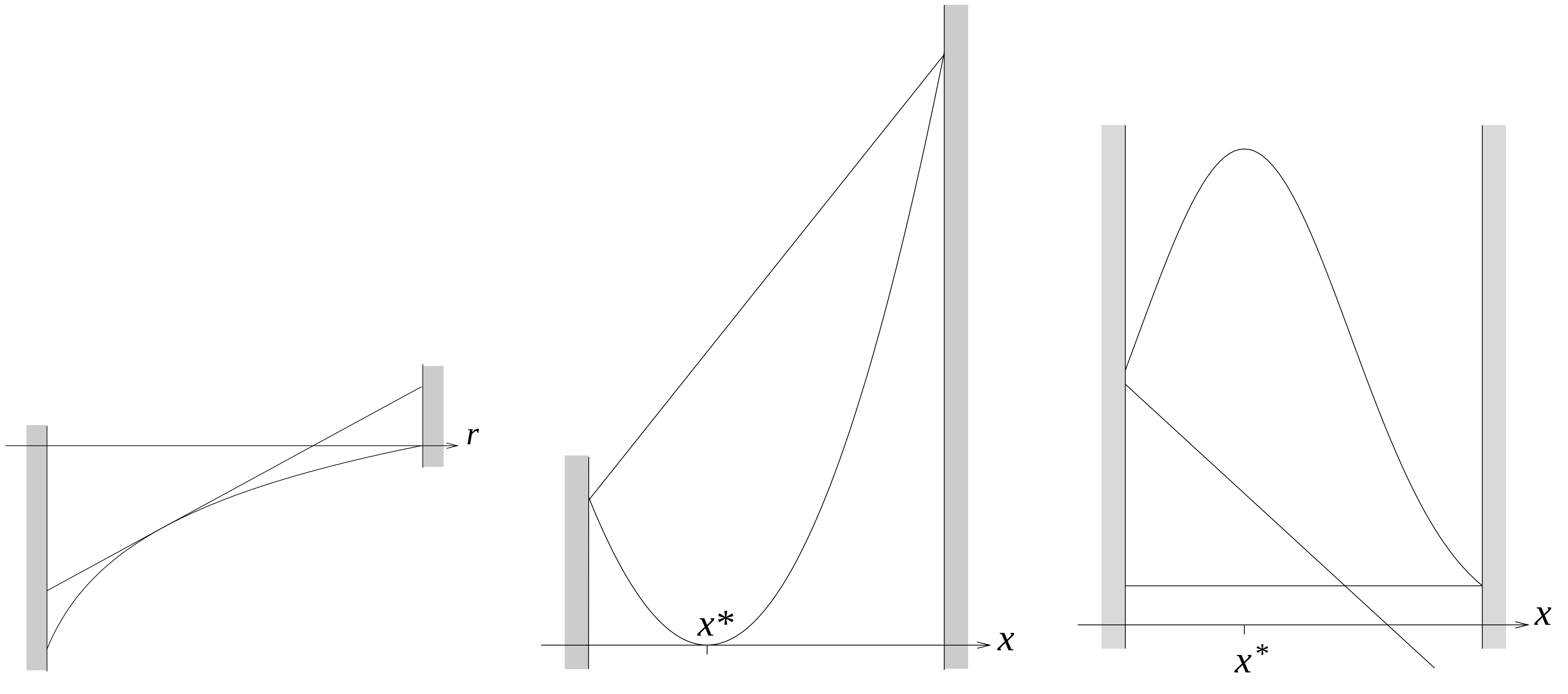
  $EI(\hat{y}_L, s_U^2)$ is upper bound on $EI(X)$.

Bounding $\hat{y}(X)$ and $s^2(X)$

- $\hat{y}$ is a linear function of $\boldsymbol{r}$.
- s^2 is a concave quadratic function of $\boldsymbol{r}$.
- However
$$r_i = \exp\left(-\sum_{k=1}^{d} \theta_k (x_k - x_k^{(i)})^2\right)$$
- Local methods cannot guarantee to find $\min \hat{y}(X)$ or $\max s^2(X)$.
- Lower bound on $\hat{y}(X)$ and upper bound on $s^2(X)$ can be found by linearising the nonlinear equations and solving separate LP and QP problems.

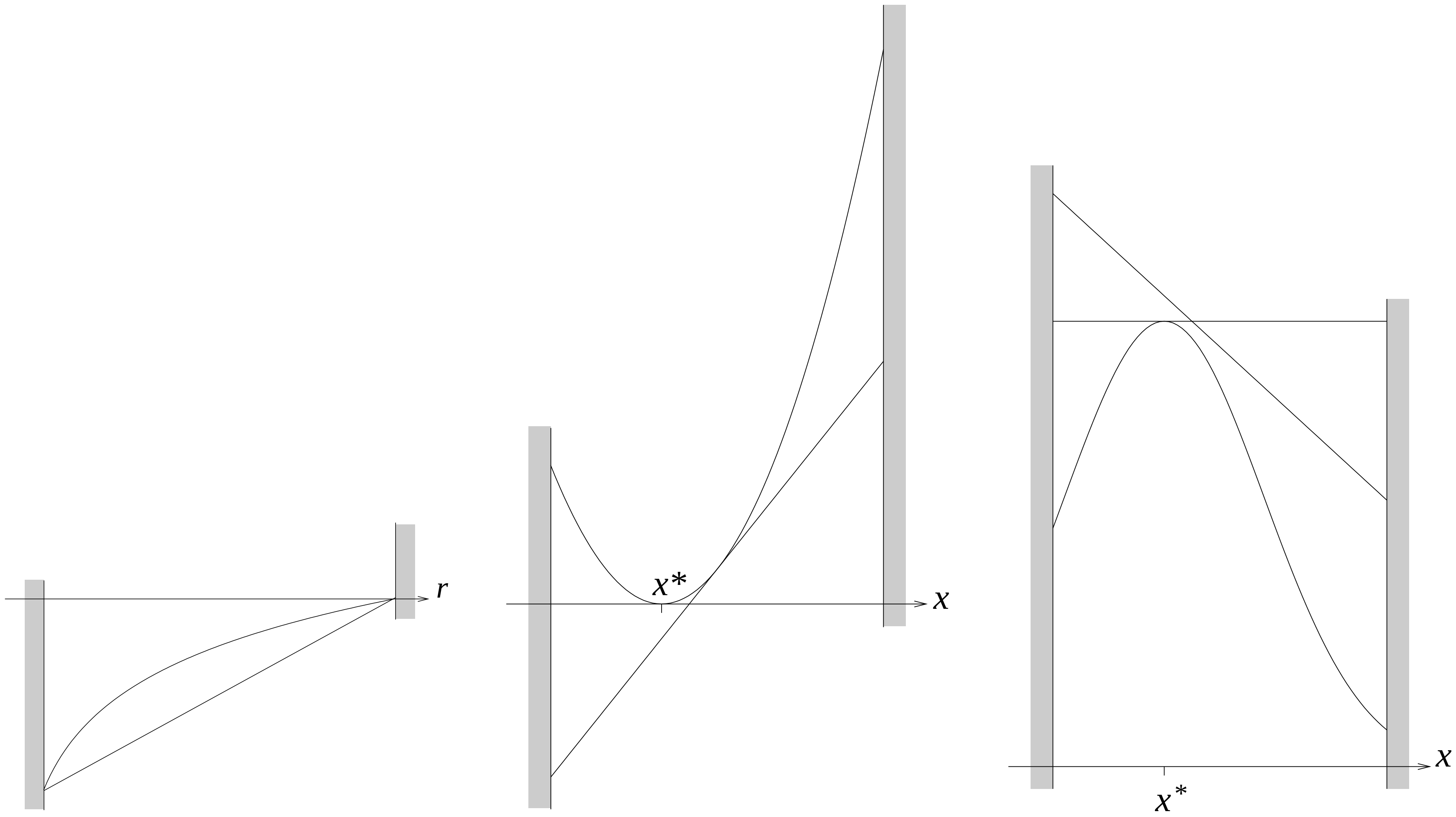
Linearisation of nonlinear equations

For $r = \exp\left(-\theta(x - x^*)^2\right)$, Jones *et al.* use a tangent to $\ln(r_i)$ and a chord across $\theta(x - x^*)^2$ to construct a linear underestimator



Linearisation of nonlinear equations (cont.)

For $r = \exp\left(-\theta(x - x^*)^2\right)$, Jones *et al.* use a chord across $\ln(r_i)$ and a tangent to across $\theta(x - x^*)^2$ to construct a linear overestimator



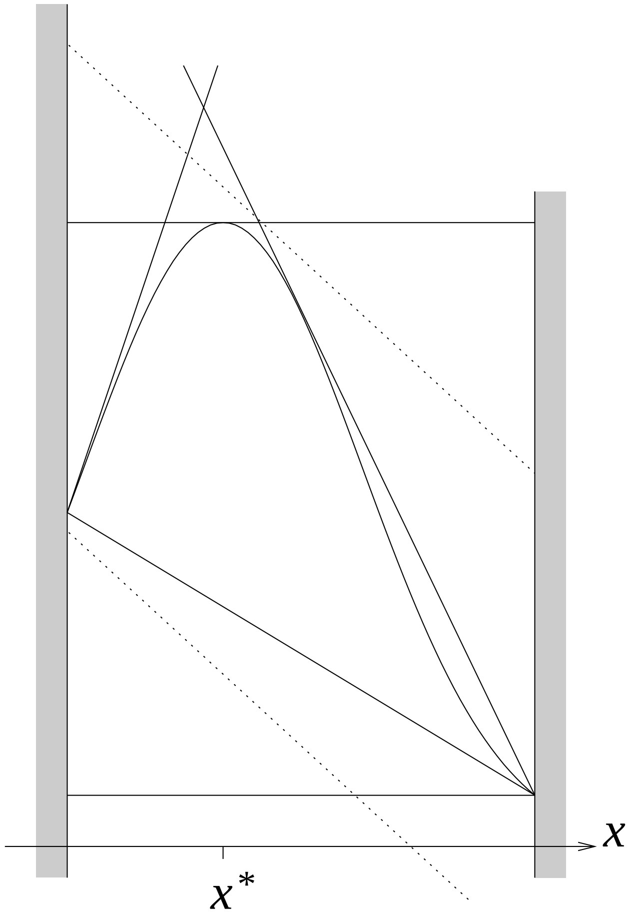
Drawbacks

Lower bound on $\hat{y}(X)$ and upper bound on $s^2(X)$, and hence upper bound on $EI(X)$, are typically poor.

- Bounds on $\hat{y}(X)$ and $s^2(X)$ typically cannot be achieved at the same point in X, even with the relaxation of the nonlinear equations.
- Much branching may be required before upper bounds are valuable.
- Need to address following issues
  - Slackness of Jones' linearisations.
  - Independence of bounding procedures for $\hat{y}(X)$ and $s^2(X)$.

Improving the linearisations of nonlinear equations

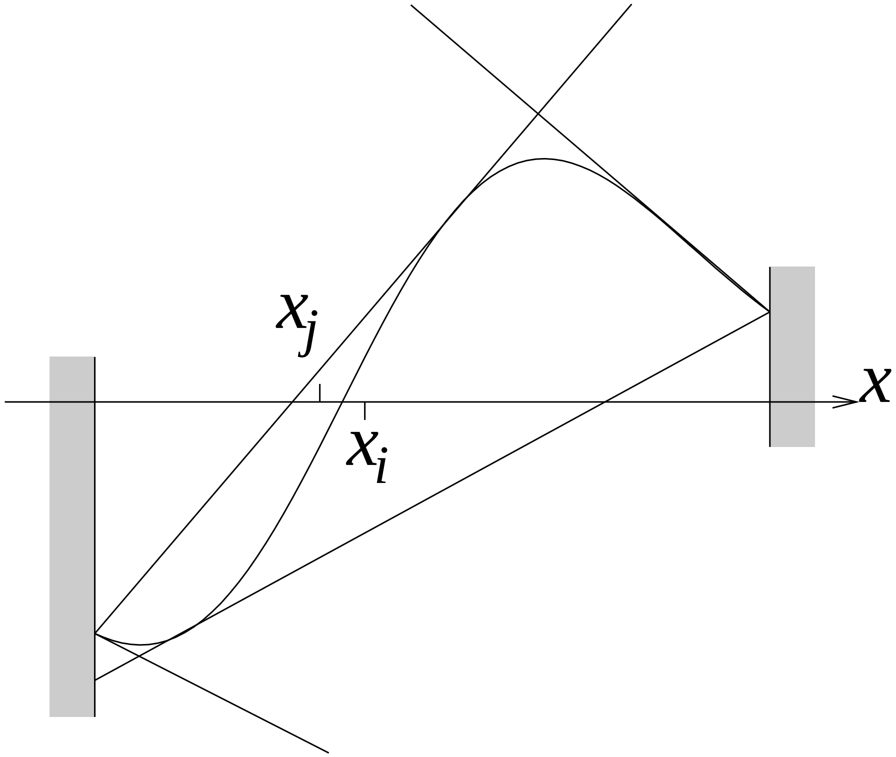
We construct use explicit tangents/chords for $r = \exp\left(-\theta(x - x^*)^2\right)$ to construct (up to) two linear overestimators and two linear underestimators.



- Close to capturing the convex hull of $r(x)$.
- All linearisations extend to $\boldsymbol{x} \in [0,1]^d$.

Cuts on $r_i - r_j$

- When $\theta(x^{(i)} - x^{(j)})^2$ is small, $r_i - r_j$ is also small.
- We construct additional tangents/chords for $r_i - r_j$.



A feasible bound on $EI(X)$

Consider maximizing

$$EI(\hat{y}, s^2) = (y_{\min} - \hat{y})\Phi\left(\frac{y_{\min} - \hat{y}}{s}\right) + s\phi\left(\frac{y_{\min} - \hat{y}}{s}\right)$$

as a function of $\hat{y}$ and s^2.

- Points $\left[\, \hat{y}(\boldsymbol{x}) \quad s^2(\boldsymbol{x}) \,\right]^T$, for $\boldsymbol{x} \in X$ form a feasible region S:

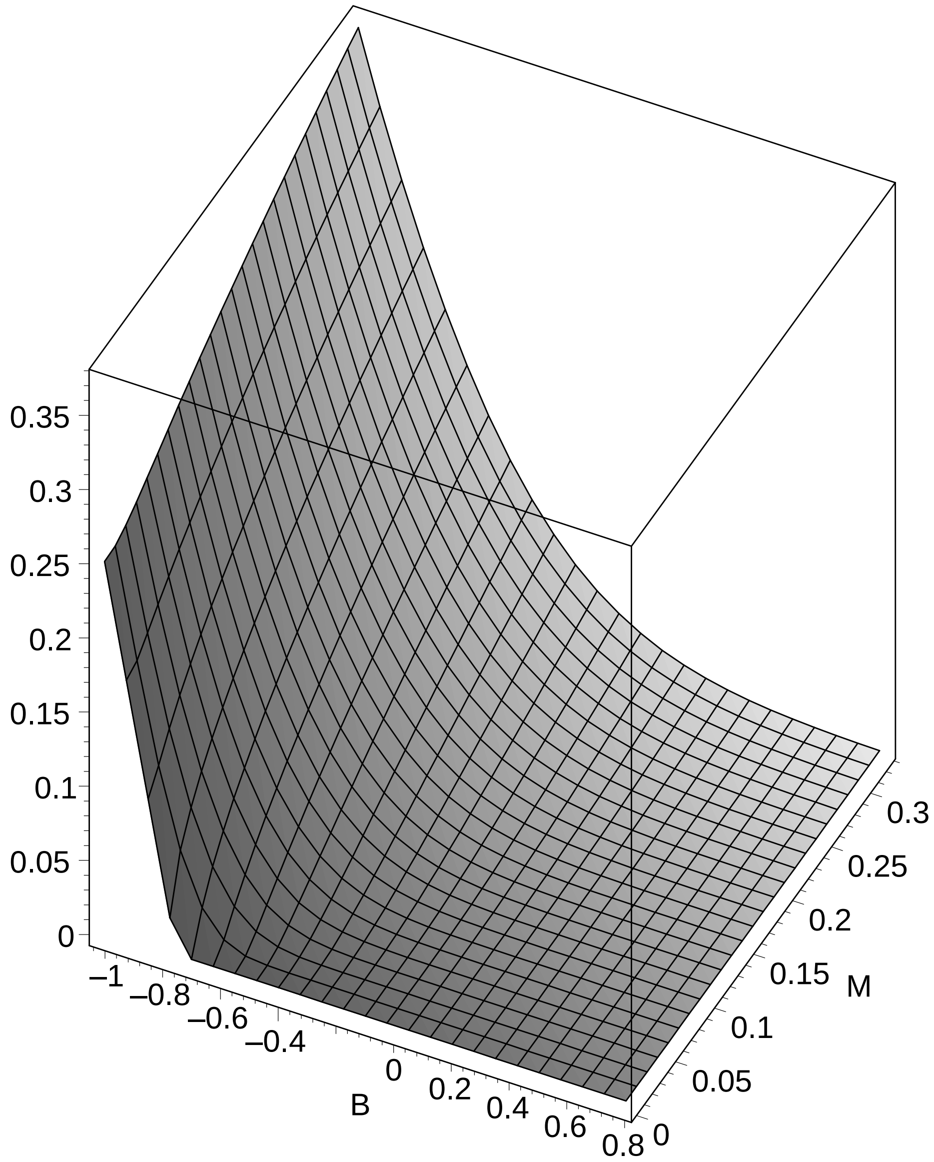
$$\max EI(\hat{y}, s^2), \quad \left[\, \hat{y}(\boldsymbol{x}) \quad s^2(\boldsymbol{x}) \,\right]^T \in S \qquad \equiv \qquad \max EI(\boldsymbol{x}), \quad \boldsymbol{x} \in X.$$

- S is bounded by $\hat{y}_L \leq \hat{y}$ and $s^2 \leq s_U^2$.
- Solving a further LP problem gives a bound $\hat{y} \leq \hat{y}_U$ and, clearly, $s^2 \geq 0$.
- For any constants b and $m \geq 0$, solving the concave QP problem

$$\max b\hat{y}(\boldsymbol{x}) + ms^2(\boldsymbol{x})$$

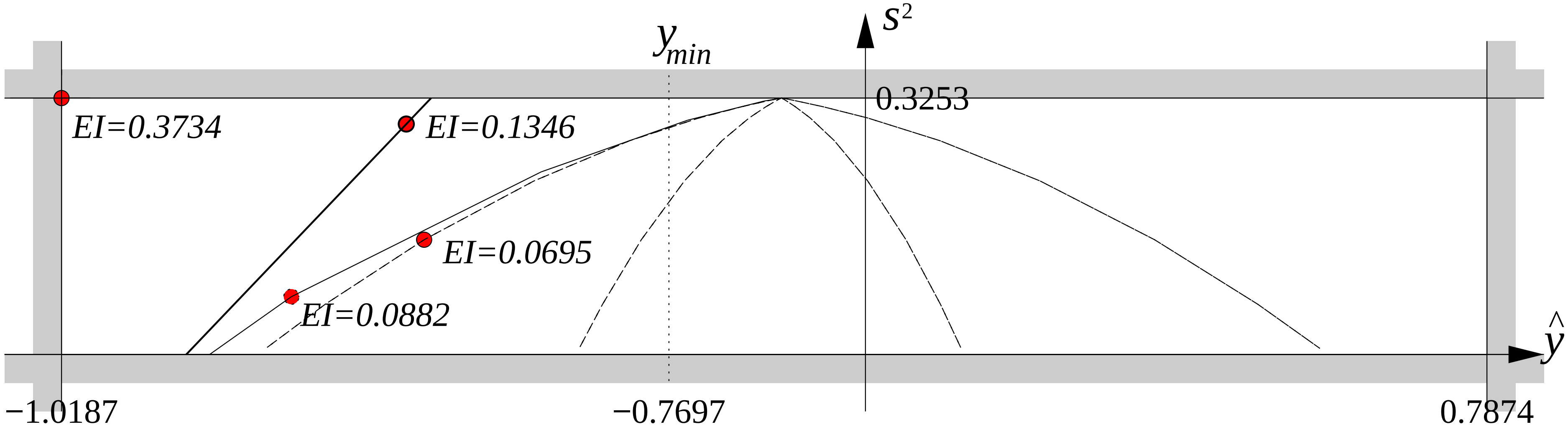
subject to the linearisations of the nonlinear equations yields a bound on S.

$EI(\hat{y}, s^2)$


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
–1
–0.8
–0.6
–0.4
0
0.2
0.4
0.6
0.8
B
0
0.05
0.1
0.15
0.2
0.25
0.3
M

A feasible bound on $EI(X)$ cont.

For the example function and initial design above, when $X = [0, 1]$, S and (approximate) envelope of bounds on S are



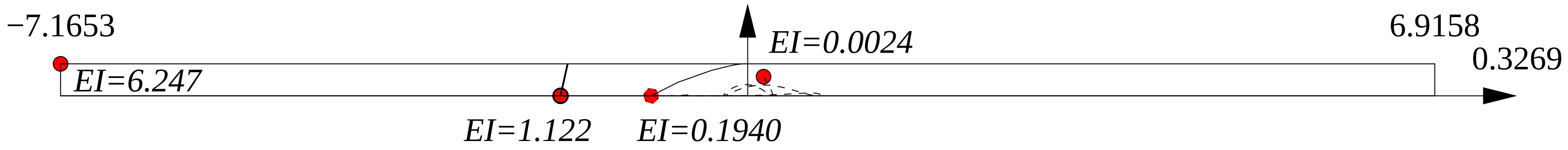
- Maximizing $EI(\hat{y}, s^2)$ over (approximate) envelope of bounds on S typically yields a much better bound on $EI(X)$ than Jones' approach.

Results

k	$x^{(k)}$	$f(x^{(k)})$	Number of boxes in B&B with EI upper bound from			
1	0.8750	-0.7697				
2	0.6250	0.5862				
3	0.1250	0.1247				
4	0.3750	-0.3666				
			$[\hat{y}_L, s_U^2]$	Approx. envelope	Exact envelope	$\max EI(X)$
5	0.8521	-0.6117	127	47	51	31
6	0.9281	-0.9188	100	50	46	29
7	0.0000	0.0000	142	115	108	25
8	1.0000	-0.5366	173	125	115	21
9	0.9222	-0.9199	86	61	59	20
10	0.9238	-0.9201	164	133	125	20
11	0.9237	-0.9201	179	151	145	

Deterioration of bounds

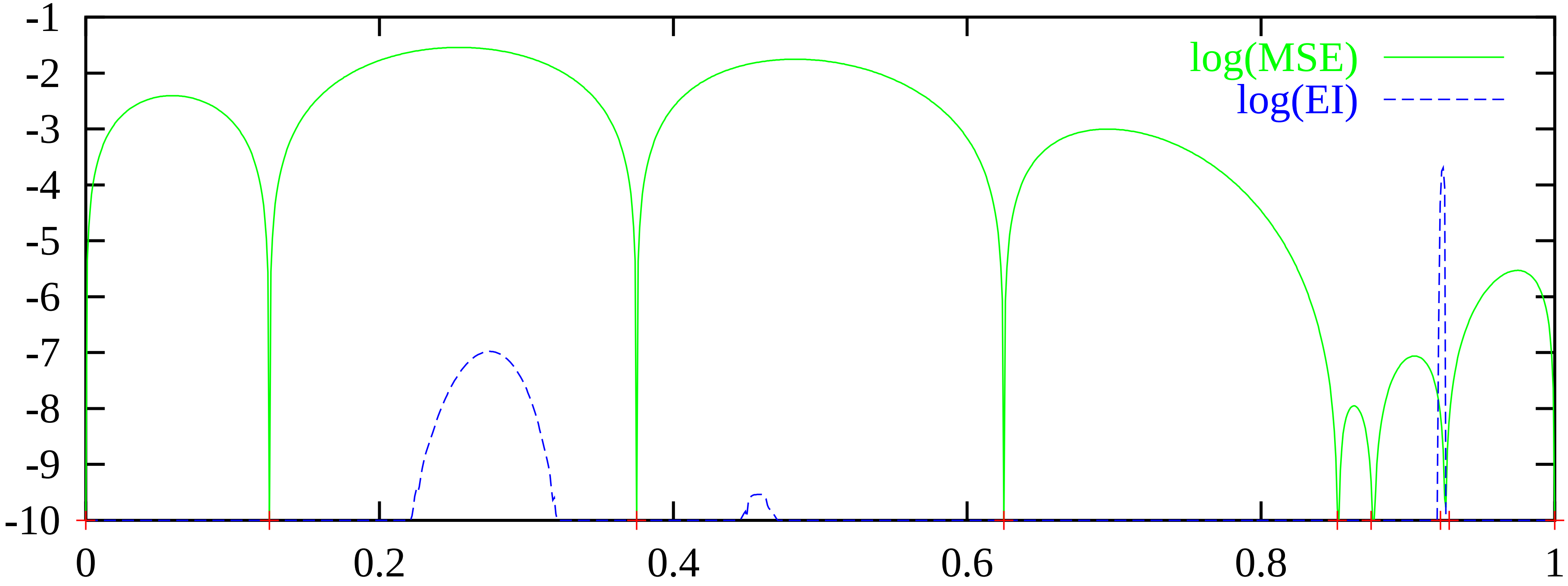
After two iterations, for $X = [0, 1]$, S and (approximate) envelope of bounds are



- Bounds over (approximate) envelope of bounds on S still much better but now well short of $\max EI(X)$.
- So much slack in (Jones') linearisations that bounds on $\hat{y}(\boldsymbol{x})$ and $s^2(\boldsymbol{x})$ are very poor.
- Need to combine tighter linearisations and techniques for tighter bounding of $EI(X)$.

Sample point 10

With 9 sample points, (log) MSE and EI are



- Grid search for max $EI(X)$ fails: yields $x^{(9)} = 0.2751$ not 0.9238.

Conclusions

EGO is a most appealing method of global optimization.

- Tested on significant number of problems
- Reliability has been improved
- Efficiency via bounding of $EI(X)$ has been improved