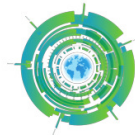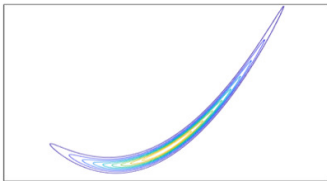# Notes on Affine Invariant Ensemble Sampler

Author: Adolphus Lye

Institute for Risk and Uncertainty

Email: adolphus.lye@liverpool.ac.uk
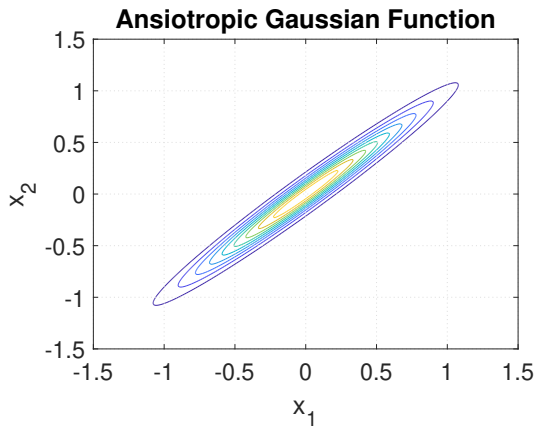
Problem

- Suppose we have a poorly-scaled target distribution $\pi(\boldsymbol{x})$, for $\boldsymbol{x} \in \mathbb{R}^2$:
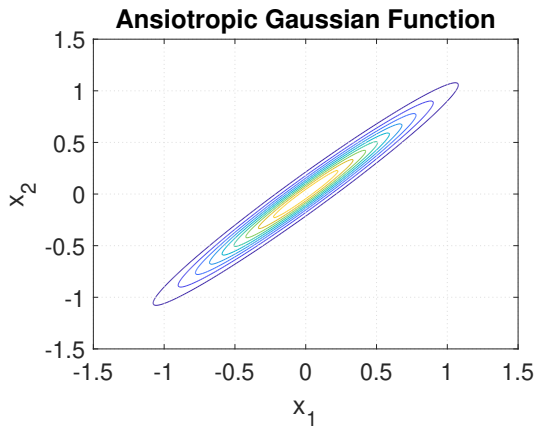
$$\pi(\boldsymbol{x}) \propto exp\left[-\frac{(x_1 - x_2)^2}{2 \cdot \epsilon} - \frac{(x_1 + x_2)^2}{2}\right]$$

whereby ϵ is the scaling parameter set as 0.01 here.

Ansiotropic Gaussian Function

Problem

- Anisotropic functions would pose difficulties for MCMC strategies such as the Metropolis-Hastings (MH) [1] (*will be proven later*).

- One solution would be to perform Affine Transformation [2].

Ansiotropic Gaussian Function

Affine Transformation

- **Affine Transformation** is the invertible $\mathbb{R}^d \mapsto \mathbb{R}^d$ space linear mapping in the form [2]:

$$\boldsymbol{x} \mapsto \boldsymbol{y} = \hat{A}\boldsymbol{x} + \boldsymbol{b}$$

whereby $\hat{A}$ is a $n \times n$ non-singular scaling matrix, and $\boldsymbol{b}$ is a $n \times 1$ scaling vector.

- The target distribution demonstrates **Affine Invariance** if:

$$\pi(\boldsymbol{x}) \mapsto \pi_T(\boldsymbol{y}) = \pi_T(\hat{A}\boldsymbol{x} + \boldsymbol{b}) \propto \pi(\boldsymbol{x})$$
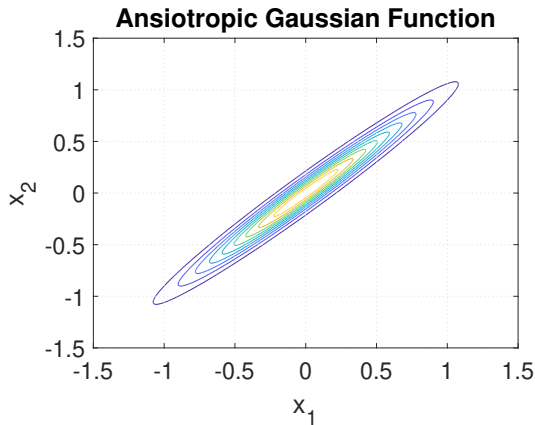
Problem

For $\quad \pi(\boldsymbol{x}) \quad \propto \quad exp\left[-\frac{(x_1-x_2)^2}{2\cdot\epsilon} - \frac{(x_1+x_2)^2}{2}\right]$
whereby $\epsilon = 0.01$:

- Set the Affine Transformation as follows:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\epsilon}} & -\frac{1}{\sqrt{\epsilon}} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

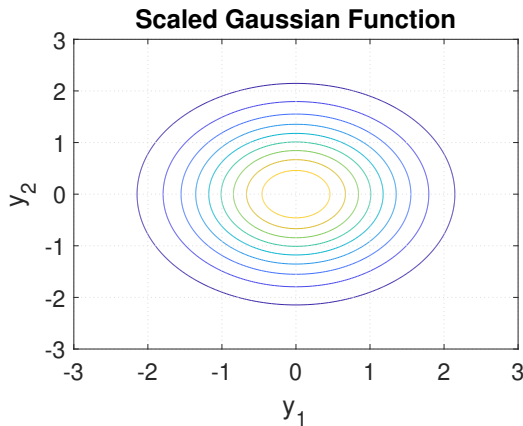- This gives: $y_1 = \frac{x_1-x_2}{\sqrt{\epsilon}}$; $y_2 = x_1 + x_2$

Ansiotropic Gaussian Function

Problem

Now, $\pi(\boldsymbol{x}) \mapsto \pi_T(\boldsymbol{y})$:

- $\pi_T(\boldsymbol{y})$ becomes simply:

$$\pi_T(\boldsymbol{y}) \propto exp\left[-\frac{(y_1^2 + y_2^2)}{2}\right]$$

- π_T is independent ϵ and becomes a simpler target distribution to sample from.
- Samples of y_1 and y_2 can be obtained directly and and inverse transformation can be done from there to obtain samples of x_1 and x_2.


Scaled Gaussian Function

Problem

- However, not all target distributions can be easily scaled like in the case of $\pi(\boldsymbol{x})$;

- E.g. Rosenbrock Banana function $r(\boldsymbol{x})$:

$$r(\boldsymbol{x}) \propto exp\left[-\frac{100 \cdot (x_2 - x_1^2)^2 + (1 - x_1)^2)}{20}\right]$$

- To sample directly from such poorly-scaled target distributions in general, Goodman and Weare (2010) proposed the Ensemble Sampling with Affine Invariance (ESAI) technique [1].

The Rosenbrock Banana
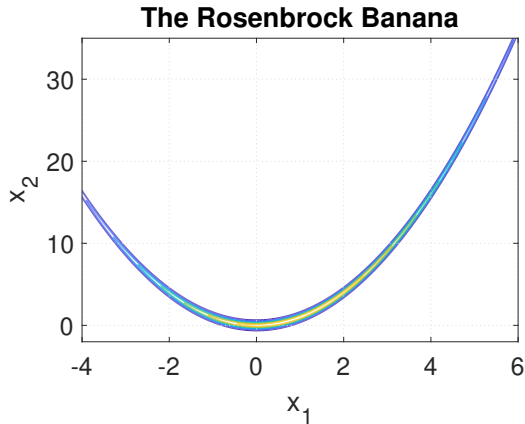
Ensemble Sampling with Affine Invariance (ESAI)

Some basic description of the ESAI technique [1, 3]:

- Analogous to the MH sampler but involves an ensemble of L walkers (or chains);

- Each kth chain (for $k = 1, ..., L$) represented as $\boldsymbol{\theta}_i^k$ (for $i = 1, ..., N$), where N is the sample size per chain and $\boldsymbol{\theta}$ is the $d \times 1$ vector of estimated parameter(s);

- Condition: $L \geq d + 1$; In practice, it is recommended to have $L \geq 2d$;

- An ensemble is defined as a set of all L walkers: $\vec{\boldsymbol{\theta}}_i = \{\boldsymbol{\theta}_i^1, \boldsymbol{\theta}_i^2, ..., \boldsymbol{\theta}_i^{L-1}, \boldsymbol{\theta}_i^L\}$;

- Target distribution for the ensemble now becomes:

$$\pi(\vec{\boldsymbol{\theta}}) = \prod_{k=1}^{L} \pi(\boldsymbol{\theta}^k)$$

Ensemble Sampling with Affine Invariance (ESAI)

Workings of the ESAI technique [1, 3]:

- Updating is done at the ensemble level: $\vec{\boldsymbol{\theta}}_i \rightarrow \vec{\boldsymbol{\theta}}_{i+1}$;

- For each k, $\boldsymbol{\theta}^j$ is randomly chosen from $\vec{\boldsymbol{\theta}}^{[k]} \in \{\boldsymbol{\theta}_{i+1}^1, ..., \boldsymbol{\theta}_{i+1}^{k-1}, \boldsymbol{\theta}_i^{k+1}, ..., \boldsymbol{\theta}_i^L\}$;

- Proposal sample, $\boldsymbol{\theta}^*$, is defined as: $\boldsymbol{\theta}^* = \boldsymbol{\theta}^j + z \cdot (\boldsymbol{\theta}_i^k - \boldsymbol{\theta}^j)$;

- z is a variable that is randomly sampled from $g(z)$ for $z \in [\frac{1}{a}, a]$:

$$g(z) \propto \frac{1}{\sqrt{z}}$$

- Acceptance probability α is then calculated:

$$\alpha = min\left[1, z^{d-1} \cdot \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}_i^k)}\right]$$

Workings of the ESAI technique [1, 3] (cot'd):

- Sample uniformly r between 0 and 1;

- If $\alpha > r$, set $\boldsymbol{\theta}_{i+1}^k \rightarrow \boldsymbol{\theta}^*$. Else, set $\boldsymbol{\theta}_{i+1}^k \rightarrow \boldsymbol{\theta}_i^k$

- Repeat this procedure for all k^{th} chains.

- Repeat this procedure for all i until $i = N$.

Ensemble Sampling with Affine Invariance (ESAI)

Algorithm 1 Algorithm for ESAI to update the ensemble state $\vec{\theta}_i \rightarrow \vec{\theta}_{i+1}$

1: Initiate L chains: $\vec{\theta}_1 = \{\boldsymbol{\theta}_1^1, \boldsymbol{\theta}_1^2, ..., \boldsymbol{\theta}_1^{L-1}, \boldsymbol{\theta}_1^L\}$; each $\boldsymbol{\theta}_1^k$ generated from prior
2: **for** $i = 1 : N$ **do**
3:     **for** $k = 1 : L$ **do**
4:         Choose randomly $\boldsymbol{\theta}^j$ from $\vec{\theta}^{[k]} \in \{\boldsymbol{\theta}_{i+1}^1, ..., \boldsymbol{\theta}_{i+1}^{k-1}, \boldsymbol{\theta}_i^{k+1}, ..., \boldsymbol{\theta}_i^L\}$
5:         Sample $z \sim g(z)$
6:         Generate $\boldsymbol{\theta}^* = \boldsymbol{\theta}^j + z \cdot (\boldsymbol{\theta}_i^k - \boldsymbol{\theta}^j)$
7:         Calculate acceptance probability α and sample uniformly $r \in [0, 1]$
8:         **if** $\alpha > r$ **then**
9:             Set $\boldsymbol{\theta}_{i+1}^k = \boldsymbol{\theta}^*$
10:        **else**
11:            Set $\boldsymbol{\theta}_{i+1}^k = \boldsymbol{\theta}_i^k$
12:        **end if**
13:     **end for**
14: **end for**

Methodology

To highlight the Affine Invariance property of the Ensemble MCMC (EMCMC) technique [1] and its strength in sampling from highly anisotropic and poorly-scaled target distributions, we will compare its performance against the Metropolis-Hastings (MH) technique:

- For both techniques, samples are generated with 4 chains ($N = 2000$ per chain);
- Samples of x_1 and x_2 will first be obtained from $\pi(\boldsymbol{x})$ using both techniques;
- Samples of y_1 and y_2 will then be obtained from $\pi_T(\boldsymbol{y})$ using both techniques;
- Perform an inverse Affine Transformation for samples of y_1 and y_2 obtained from both techniques;

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{\epsilon}}{2} & \frac{1}{2} \\ -\frac{\sqrt{\epsilon}}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

- For each technique, compare the distribution of the samples of x_1 and x_2 obtained directly from $\pi(\boldsymbol{x})$ and those obtained indirectly from $\pi_T(\boldsymbol{y})$.
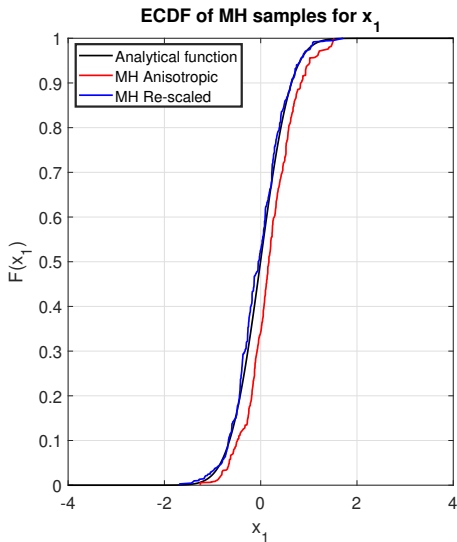
RESULTS

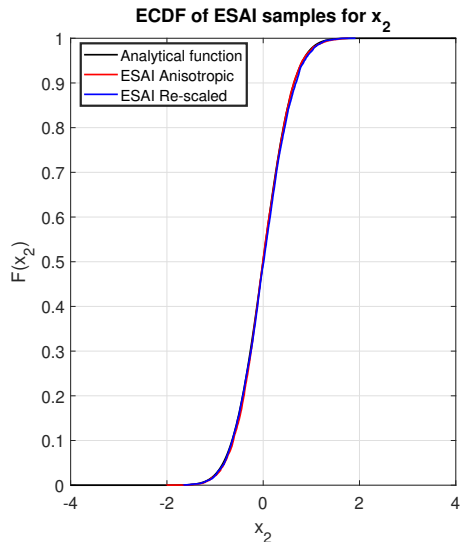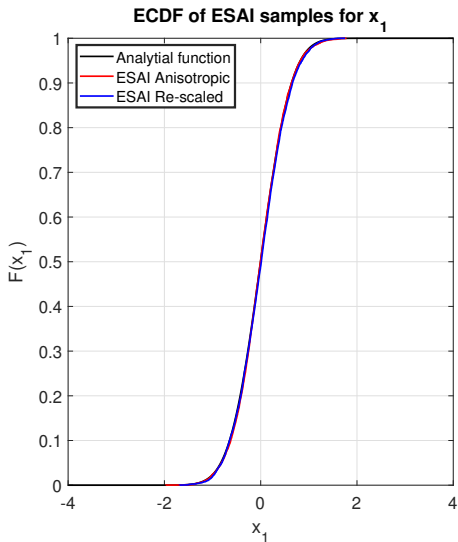# Results: Scatterplot Profile vs Analytical Function
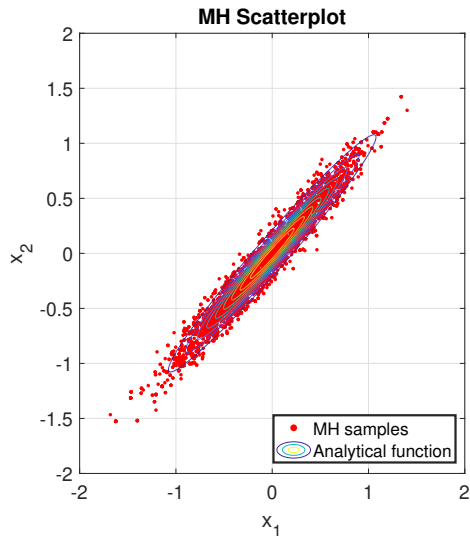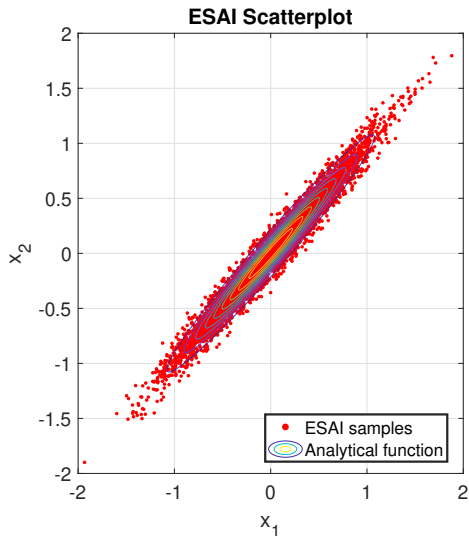
Results: Scatterplot Profile vs Analytical Function
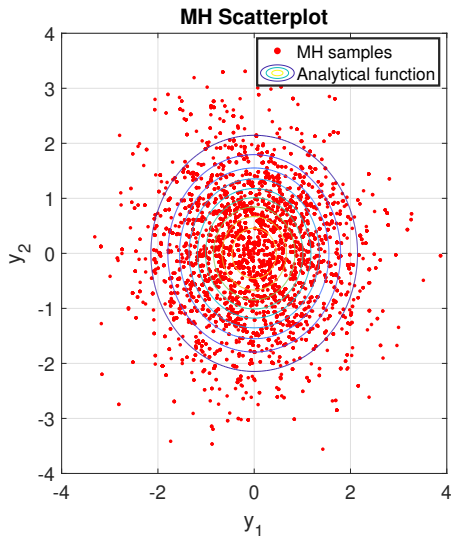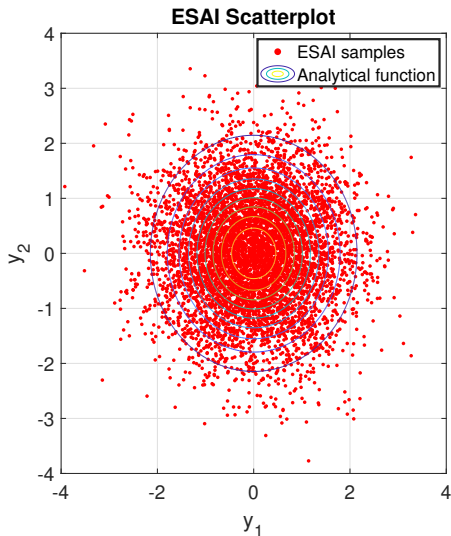
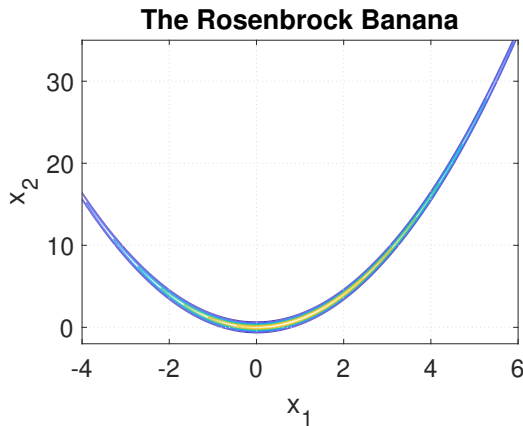# Results for MH Technique
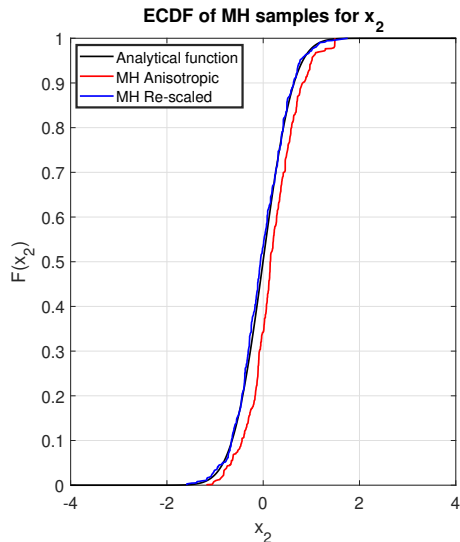
Results for ESAI Technique

Further Investigation: Rosenbrock Banana Function

Further investigation is being performed to illustrate the ability of ESAI technique in sampling from highly anisotropic distributions.
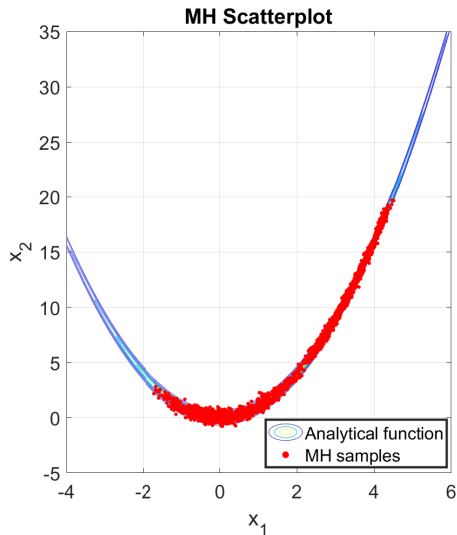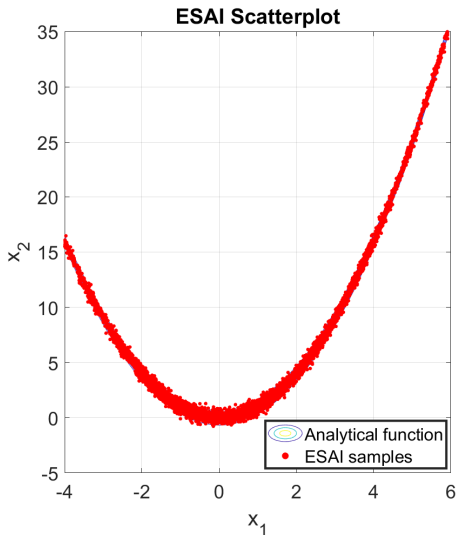
- Case Study: Rosenbrock Banana function, $r(\boldsymbol{x})$;

$$r(\boldsymbol{x}) \propto exp\left[-\frac{100 \cdot (x_2 - x_1^2)^2 + (1 - x_1)^2}{20}\right]$$

- Based on the Rosenbrock function proposed by Rosenbrock (1960) which serves as a performance test problem for optimization algorithms [4].
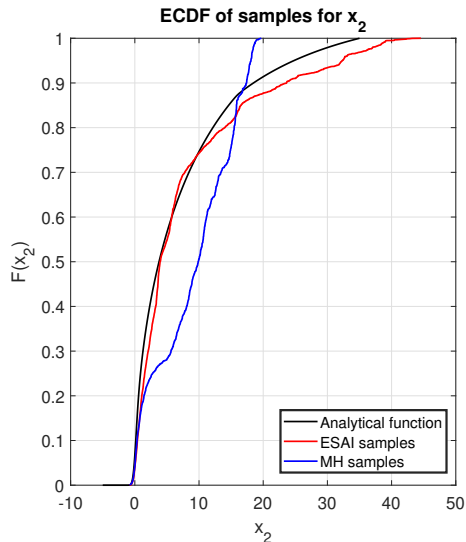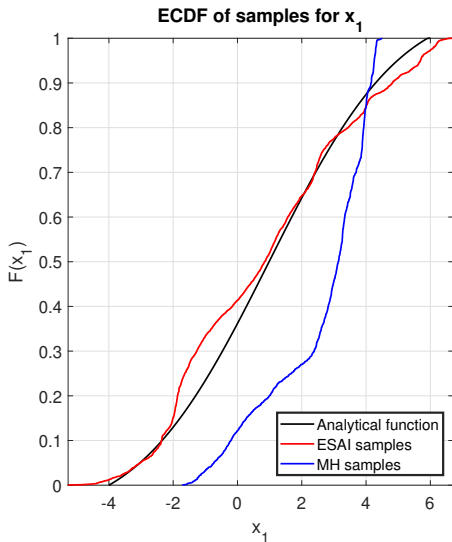
The Rosenbrock Banana

RESULTS

Conclusion

- The distribution of samples obtained from ESAI technique have consistently been relatively closer to that of the analytical function compared to that obtained from MH technique;

- ESAI technique serves as a relatively more effective sampler in sampling from highly anisotropic distributions;

- The effectiveness of ESAI technique comes at the cost of its computation time. For the same number of samples and chains, the ESAI technique has consistently took relatively longer time to generate samples compared to the MH sampler:

	Anisotropic	**Isotropic**	**Rosenbrock**
ESAI	2.578s	0.873s	0.897s
MH	0.112s	0.132s	0.074s

The End

Thank you so much for your Undivided attention!

References

J. Goodman and J. Weare, "Ensemble Samplers with Affine Invariance," *Communications in Applied Mathematics and Computational Science*, vol. 5, no. 1, p. 65–80, 2010.

R. Hartley and A. Zisserman, *Multiple view Geometry in Computer Vision.* New York: Cambridge University Press, 2004. ISBN: 0521540518.

F. Hou, J. Goodman, D. W. Hogg, J. Weare, and C. Schwab, "An Affine-Invariant Sampler for Exoplanet Fitting and discovery in Radial Velocity data," *The Astrophysical Journal*, vol. 745, no. 2, p. 198, 2012.

H. H. Rosenbrock, "An Automatic Method for Finding the Greatest or Least Value of a Function," *The Computer Journal*, vol. 3, no. 3, pp. 175–184, 1960.