

# ML-MCMC: Toy Problems

Juan Pablo Madrigal Cianci

April 18, 2018

## 1 ML with Sequential Resampling

## 2 Level Tempering

## 3 Law and Jasra's Algorithm

This is a write-up on the Implementation of the ML-MCMC algorithm of Jasra and Law, as well as some modifications to it. In general, for each pair of chains  $\chi_l, \chi_{l-1}$ , we generate proposals of either one these forms:

### 3.1 Method 1:

$$x_l^* = x_l^n + u_1, \quad (1)$$

$$x_{l-1}^* = x_{l-1}^n + u_1, \quad (2)$$

for  $u_1 \sim N(0, \Sigma)$ , where  $\Sigma$  is a properly chosen covariance matrix, and  $n$  denotes the  $n^{th}$  iteration of the MCMC sampler. This is the original method proposed on the paper.

### 3.2 Method 2:

(Madrigal). As a *trick* to keep the chains more correlated between levels, we let

$$x_l^* = x_l^n + u_1, \quad (3)$$

$$x_{l-1}^* = x_l^*. \quad (4)$$

This trick is chosen on a very *ad-hoc* way, however, the intuition is that it will provide a stronger correlation for the chains. This is also confirmed experimentally.

We now run some experiments for different toy problems which, even though simple, are useful to corroborate the validity of the method.

## 4 Scalar Gaussian

We begin with a simple, yet illustrative case. For each level  $l$ , we construct two chains  $\chi_l, \chi_{l-1}$  targeting  $\pi_l(x), \pi(x)_{l-1}$ , with  $\pi_l = N(2^{-l}, 1 + 2^{-l})$ . We are interested in studying the behaviour of chains for large number of samples  $N_l$ . In particular, we would like to see that the method is independent of the initial value. To study this, we set  $x_l^0 \neq x_{l-1}^0$  for  $l = 1, \dots, L$ , and run the MLMCMC sampler on each level for  $N_l = 10^5, l = 1, \dots, 4$ , iterations, with a burn-in of 50.000 samples. We obtain the results shown in Figure 1 for Method 1 and Figure 2 for Method 2. It seems that Method 2 is able to keep a stronger correlation between samples, as suggested. Note that as we increase  $l$ , the correlation coefficient increases between two contiguous levels, as we expected. **To Do: Experiment with appropriate  $N_l$ .**

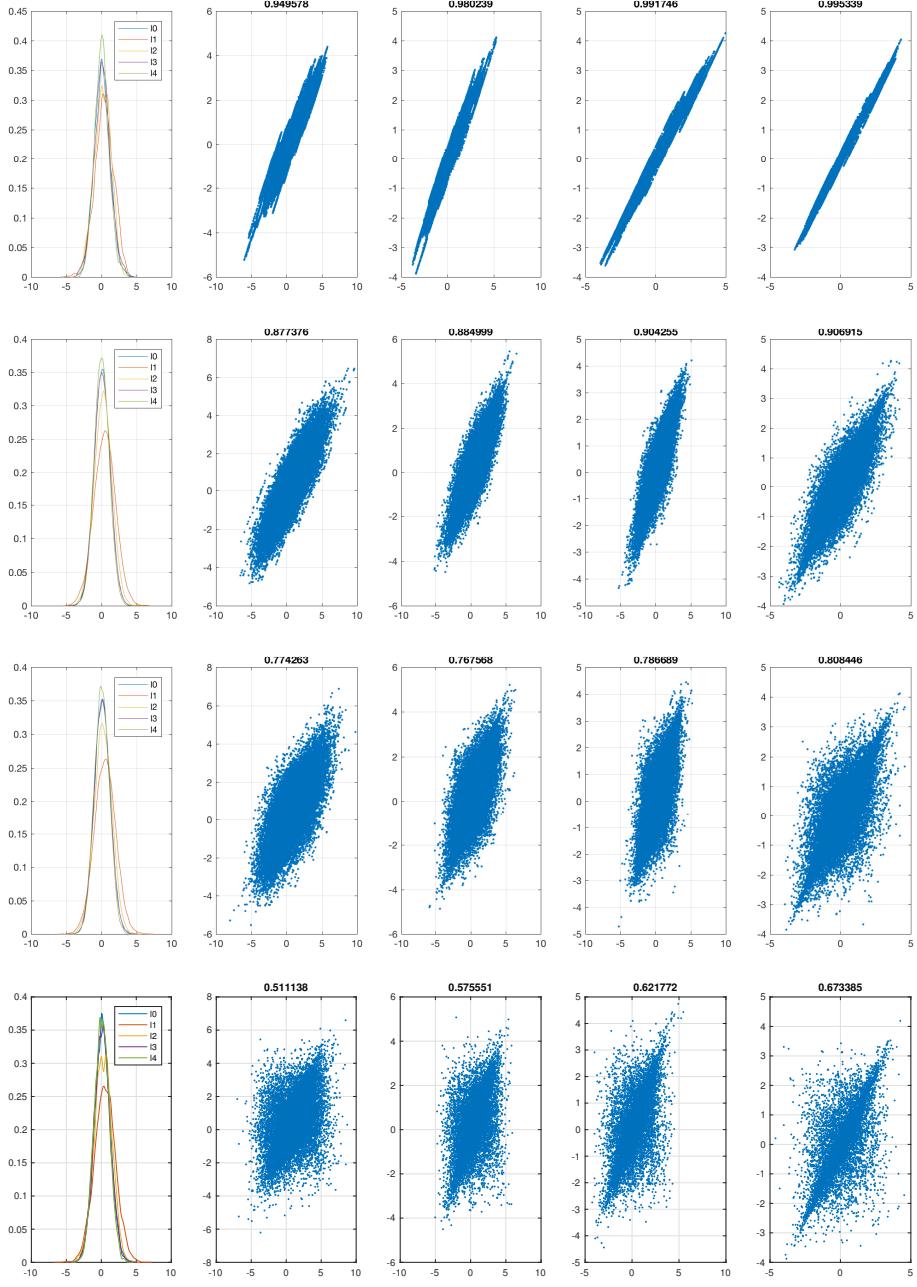


Figure 1: Results for 50.000 samples (after burn-in of 50.000 more) when samples are generated with Method 1. Results shown for  $\Sigma = 0.1, 1, 2, 10$ , in decreasing order of rows. From left to right: Leftmost: Estimated densities for each level. The other 4 figures are the correlation between samples at two consecutive levels, i.e correlation between  $\chi_l, \chi_{l-1}$ ,  $l = 1, \dots, 4$ . The correlation coefficient is shown as the title of each subplot.

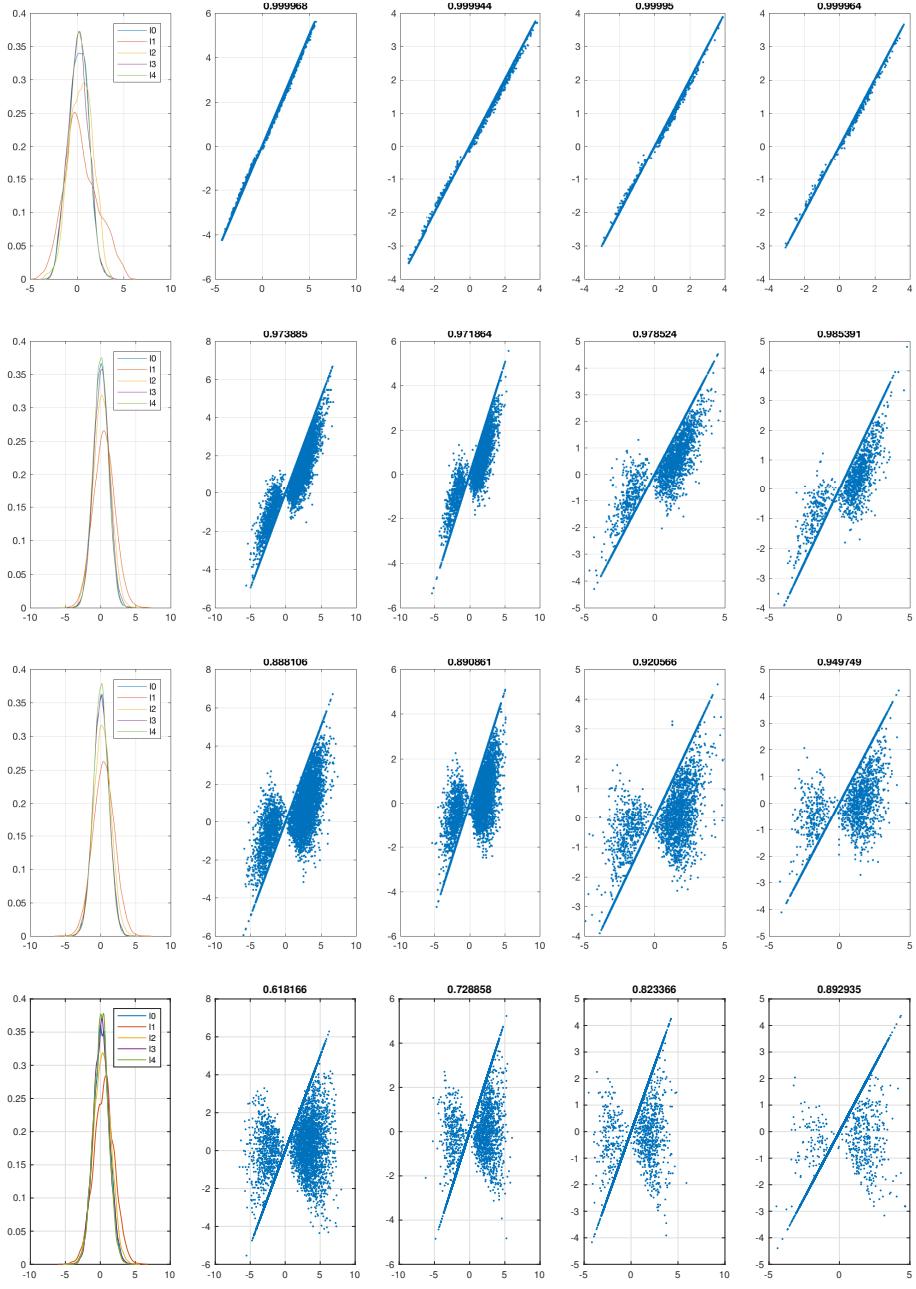


Figure 2: Results for 50.000 samples (after burn-in of 50.000 more) when samples are generated with Method 2. Results shown for  $\Sigma = 0.1, 1, 2, 10$  in decreasing order of rows. From left to right: Leftmost: Estimated densities for each level. The other 4 figures are the correlation between samples at two consecutive levels, i.e correlation between  $\chi_l, \chi_{l-1}$ ,  $l = 1, \dots, 4$ . The correlation coefficient is shown as the title of each subplot.

## 5 Update to include information requested

### 5.1 Update Method 1

We run the method for 18 levels, with  $\Sigma = 5, N_l = 10^5, l = 1, \dots, L$ , and a burn in of 50k and obtain the results shown in the figures below. We define the *divergence* as the amount of times that  $\chi_l$  accepts and  $\chi_{l-1}$  rejects, or vice-versa. Method 1 seems to introduce variance and divergence that are  $O(1)$ . There seems to be a small bias on the ml-estimator.

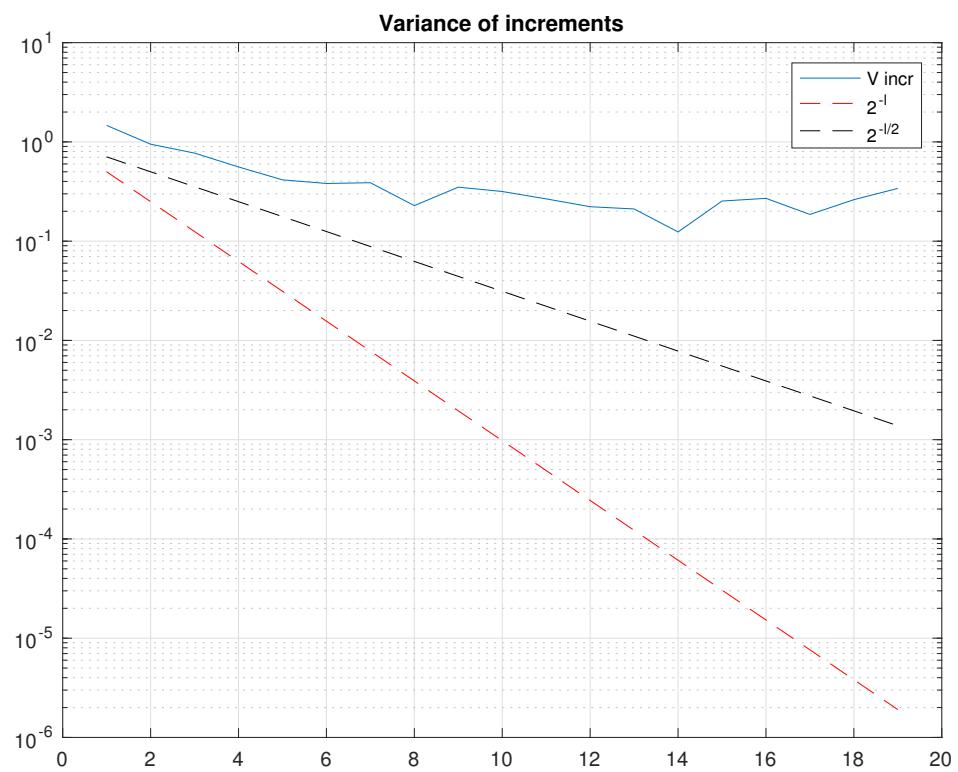


Figure 3: O(1)?

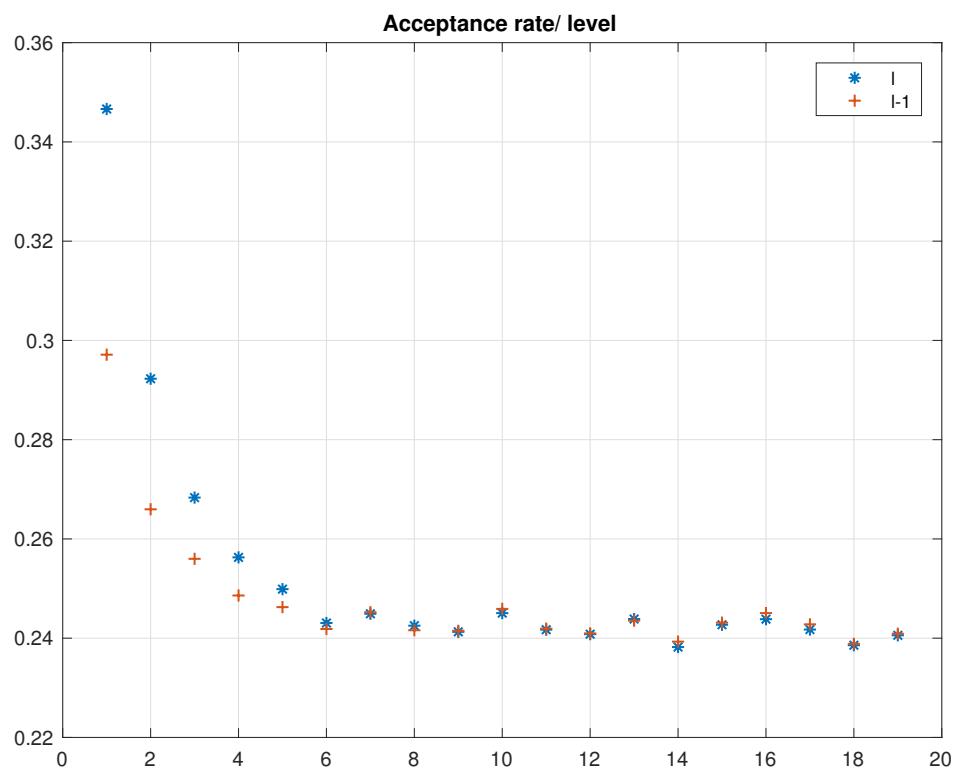


Figure 4

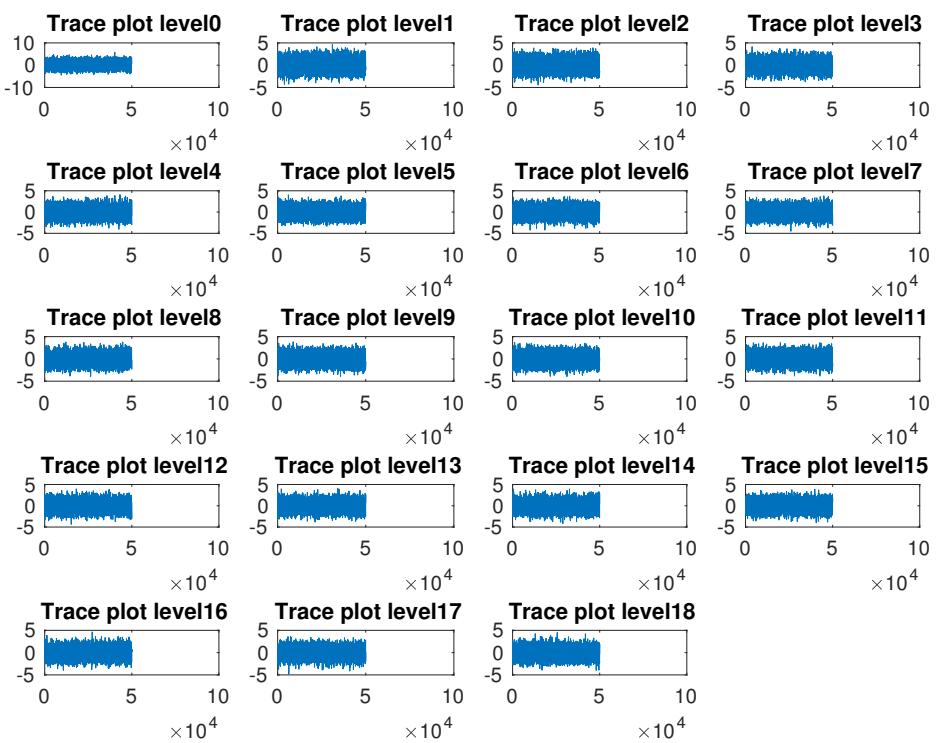


Figure 5

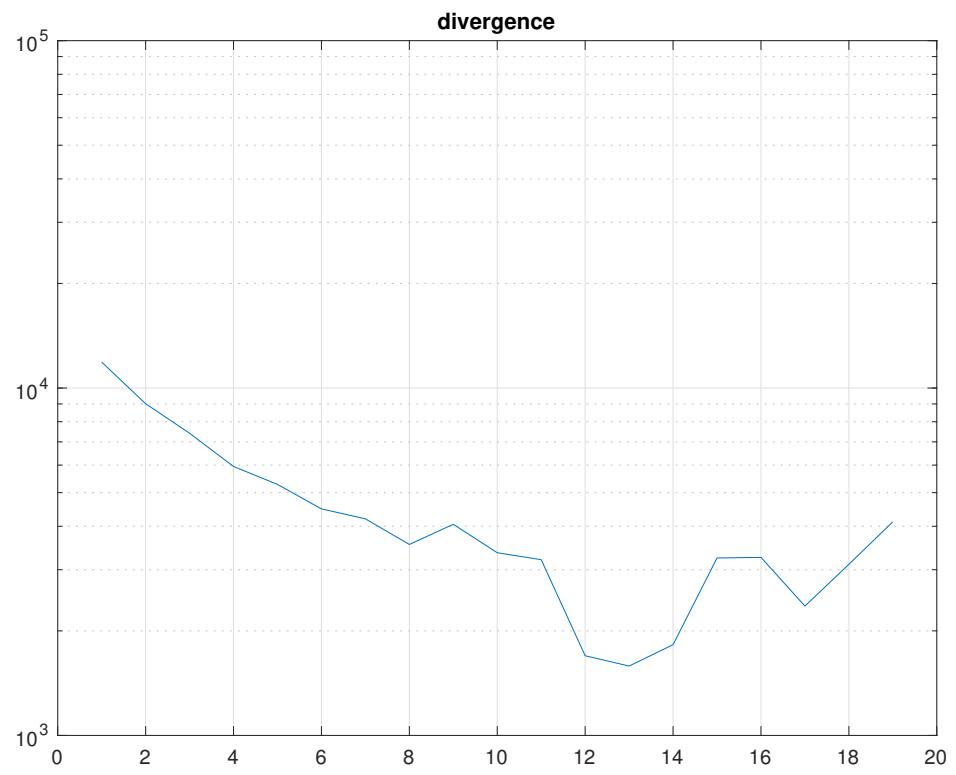


Figure 6

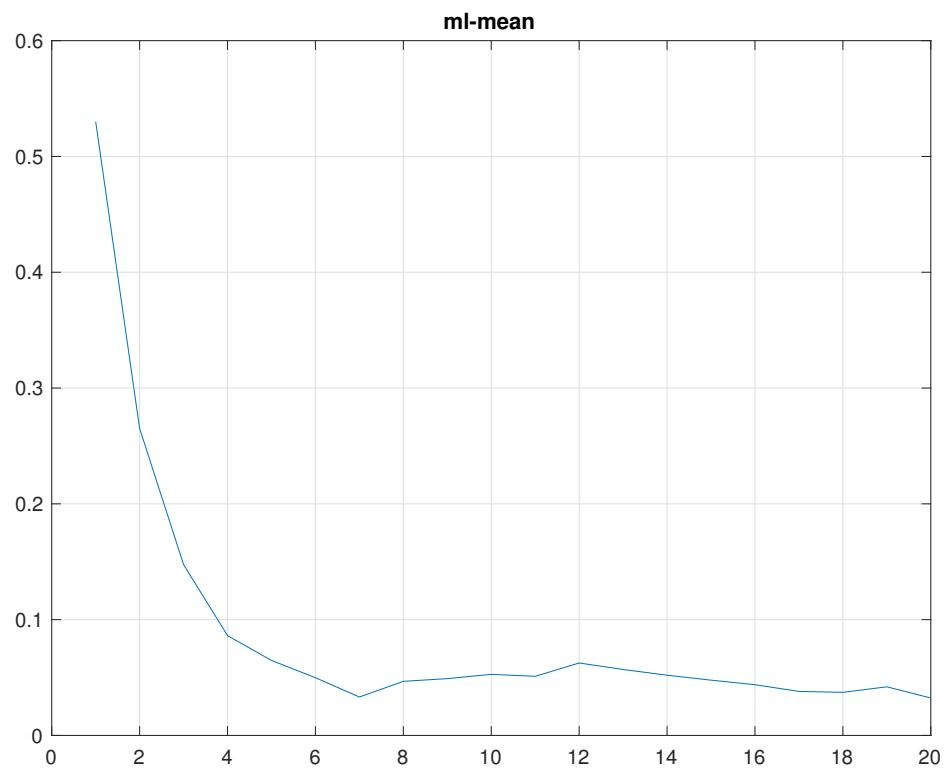


Figure 7

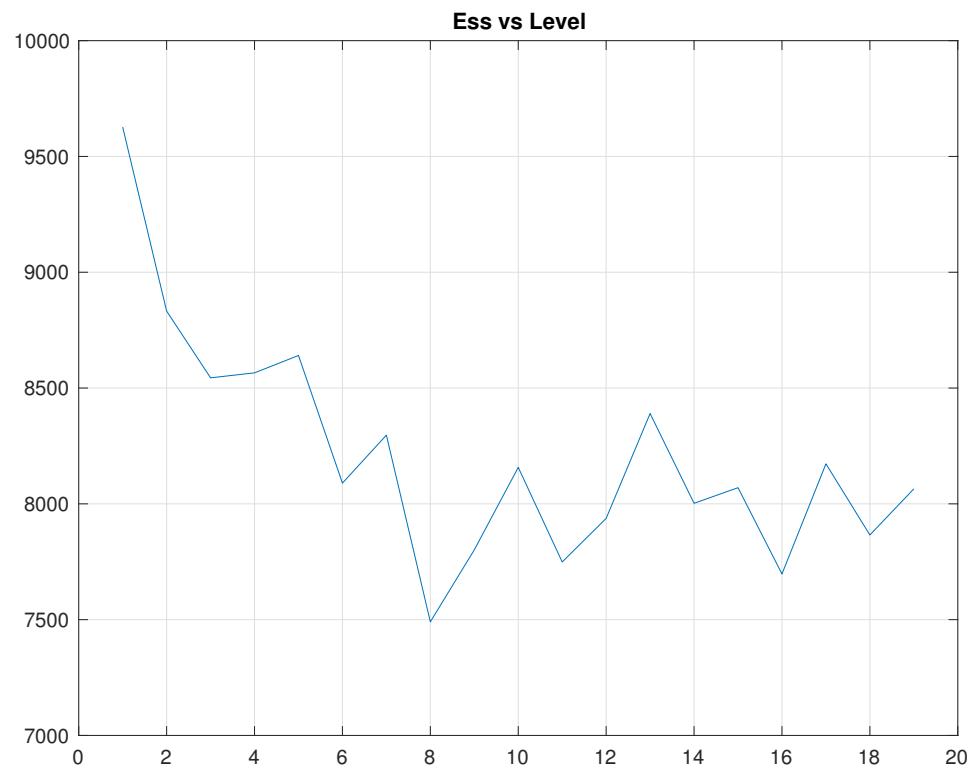


Figure 8

## 5.2 Update Method 2

We run the same experiment as above for Method 2. Here the variance nor the divergence are  $O(1)$ . results below.

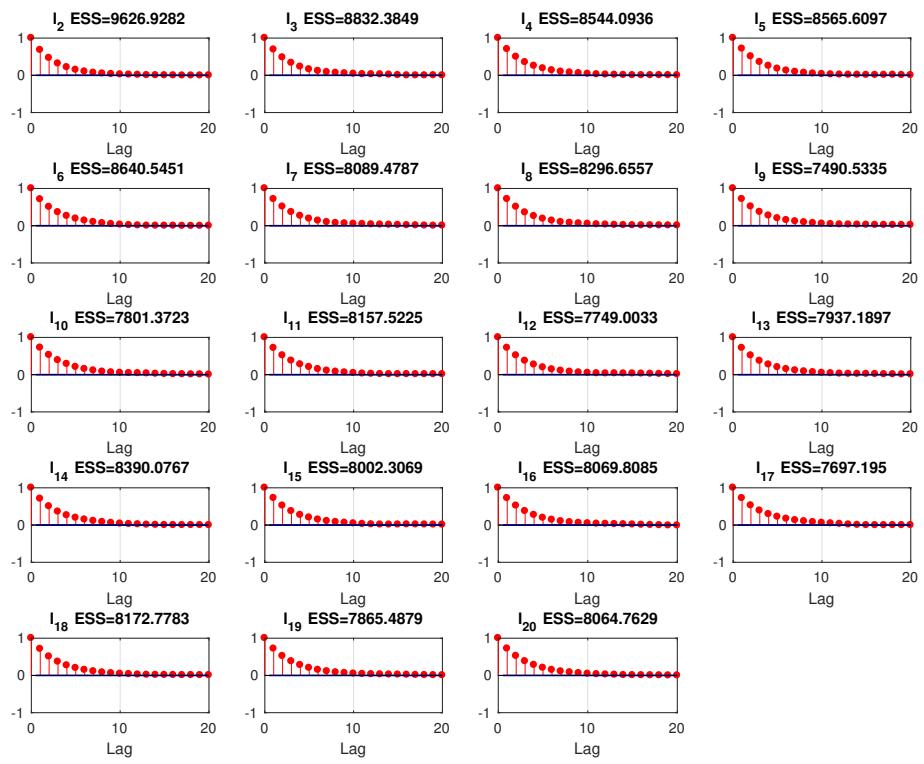


Figure 9

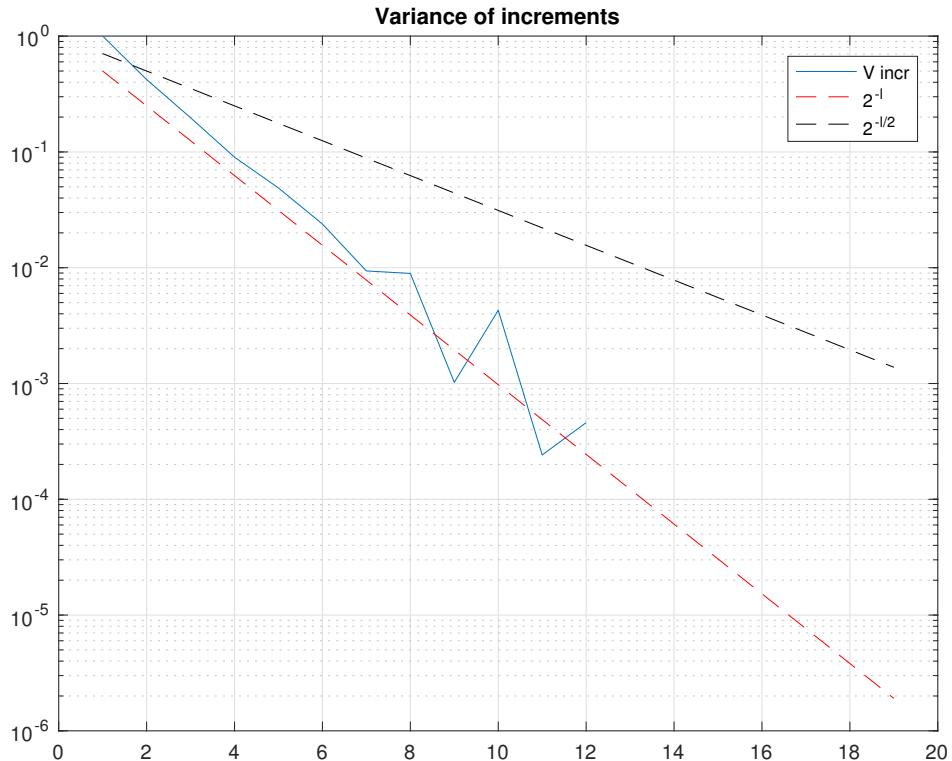


Figure 10

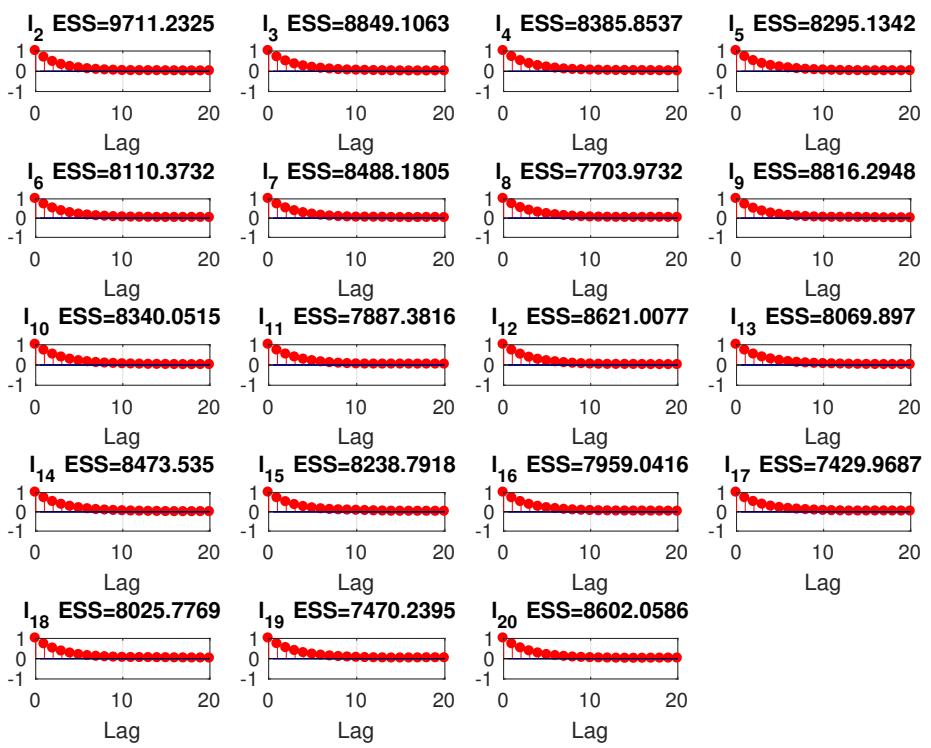


Figure 11

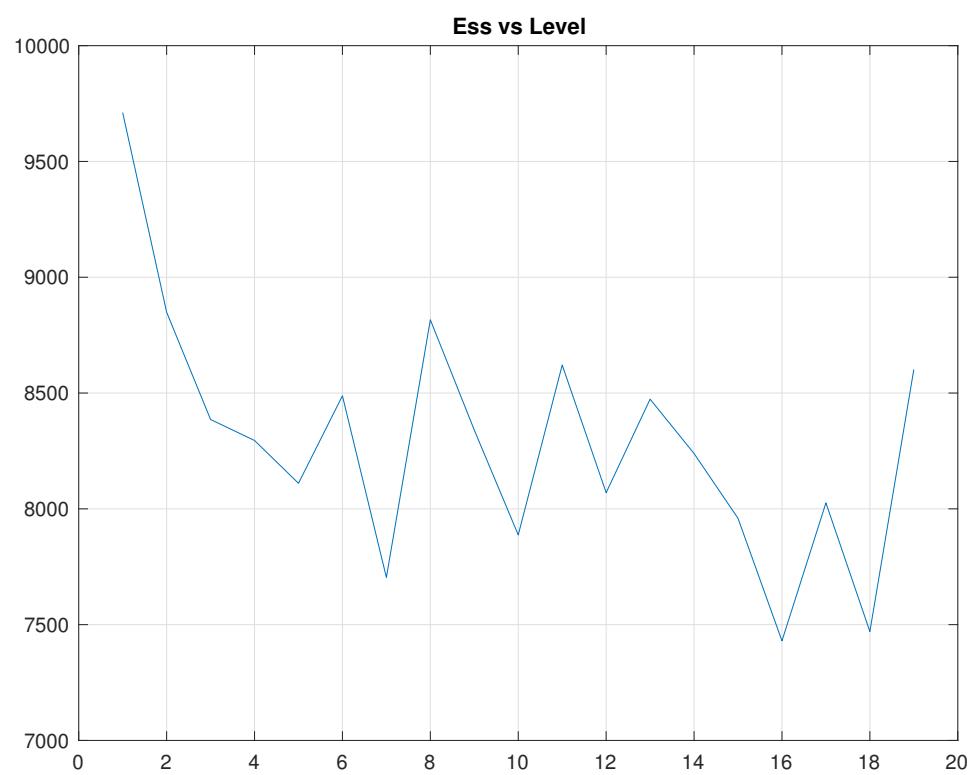


Figure 12

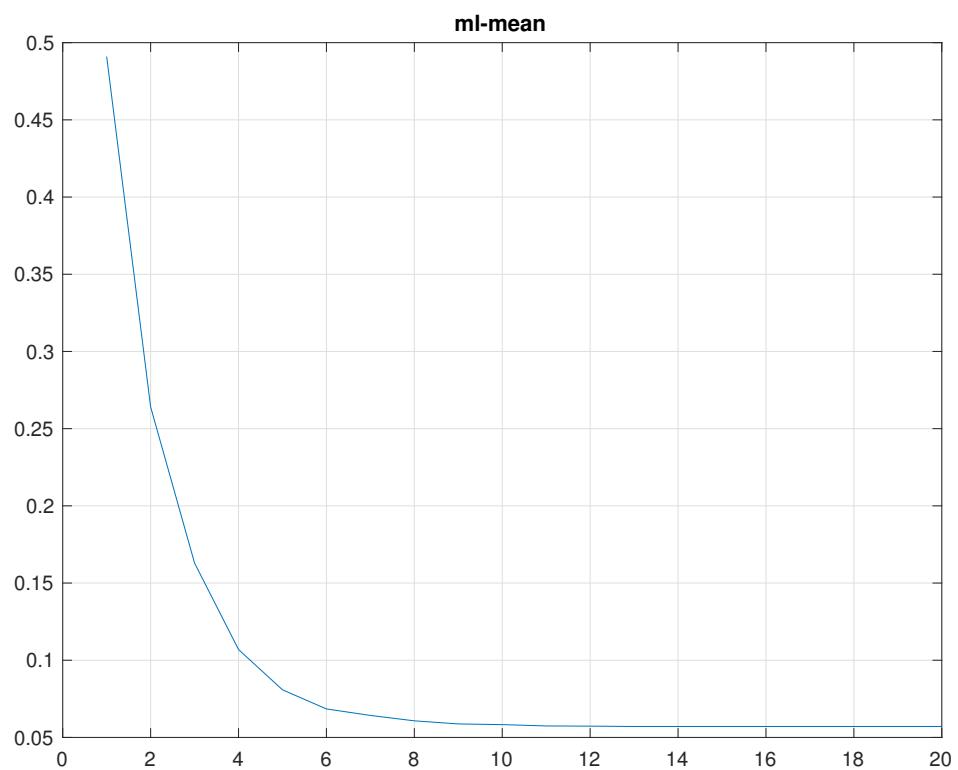


Figure 13

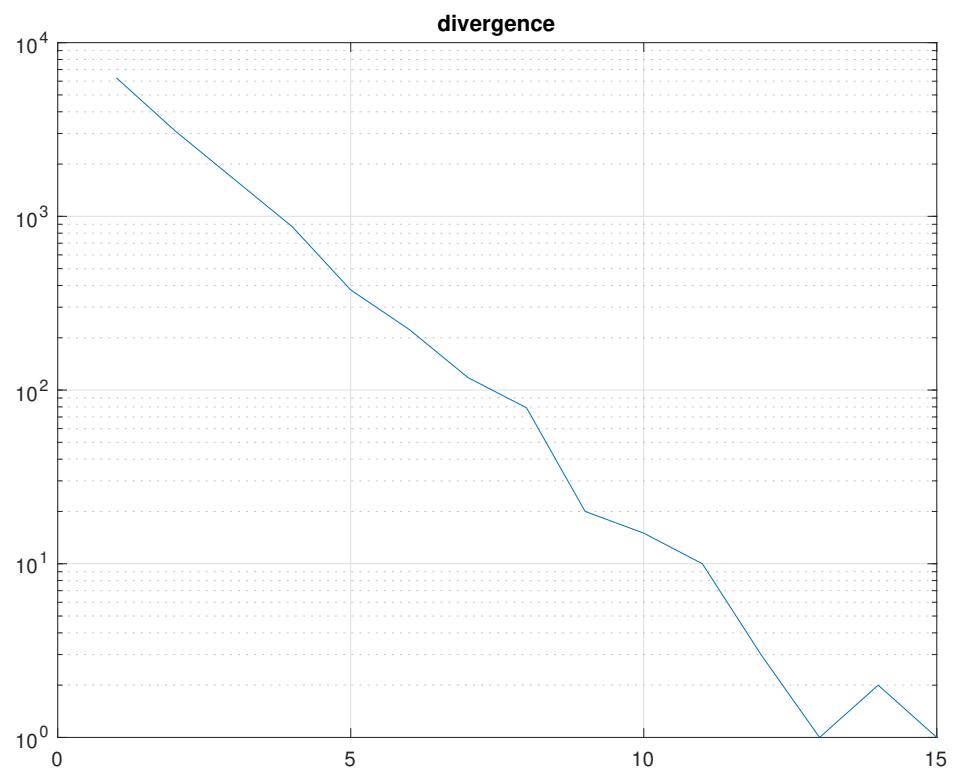


Figure 14

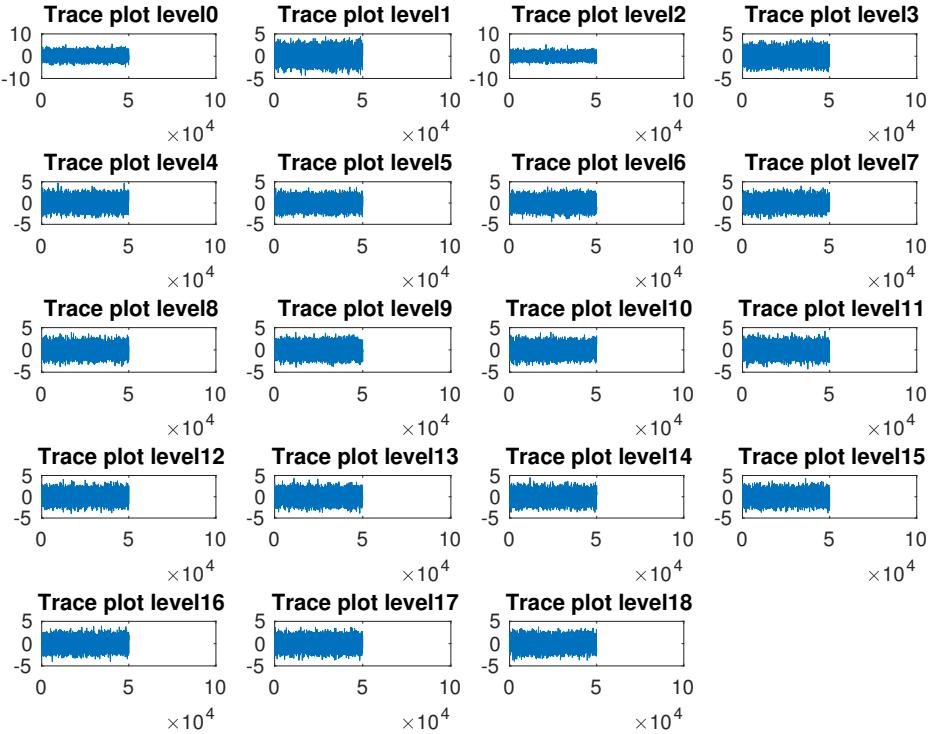


Figure 15

## 6 1D Subsurface Flow.

We now study a test-case involving the following one-dimensional elliptic equation

$$\begin{cases} -(\alpha(x, \xi)u(x))' = 1, & \forall x \in (0, 1) \\ u(0) = u(1) = 0. \end{cases} \quad (5)$$

where

$$\log(\alpha(x, \xi)) = \sum_{i=1}^{\infty} \frac{\cos(\pi xi)}{i^2} \xi_i, \quad \xi_i \sim N(0, 1). \quad (6)$$

Suppose that we can measure noisy data  $d$  at  $x = 0.25, x = 0.5$ , and  $x = 0.75$ . Based on these measurements we would like to infer properties of a quantity of interest  $Q$ . Suppose that our quantity of interest (QoI) is  $\xi_1$ . Data is generated at a much finer grid (to avoid the inverse crime). The equation is solved using a piecewise linear finite difference approximation and by truncating (6) at  $P = 10$ . We consider a hierarchy of discretization parameters  $h_l$  with  $h_l = 2^{-l}$ ,  $l = 0, \dots, 11$ .

## 7 Wave Equation 1D.

Should have by tomorrow or so.

## 8 Rejection-Free: ZZ Method.

Same as above.

## **References**