Computational Exercise #2

Type of Activity: Stability analysis of the solution in the presence of noise through a practical procedure**.**

In the computational exercise #1, you know the true parameters (a and b) because you have simulated synthetic data. In this case, you can compare the estimated parameters ($\hat{a}$ and $\hat{b}$) with their true values (a and b).

Probably, you have done this comparison, haven't you?

However, the knowledge of the true parameters is not realistic. In a real world scenario, the **true parameters are unknown**. How can I analysis the stability of our solution? This is the main goal of this homework.

We adopted the following practical procedure to assess if a solution is stable.

1) Generate L pseudorandom sequences of noise with Gaussian distribution with zero mean ($\mu = 0.0$) and standard deviation σ. Each sequence of noise has N values, where N is the number of geophysical observations.

   For example, if N = 3, Seed = 3, $\mu = 0.0, \sigma = 0.3$ and L = 5 we can get:

$\boldsymbol{\varepsilon}^1$	$\boldsymbol{\varepsilon}^2$	$\boldsymbol{\varepsilon}^3$	$\boldsymbol{\varepsilon}^4$	$\boldsymbol{\varepsilon}^5$
0.53658854	0.13095296	0.02894924	-0.55904781	-0.08321646
-0.10642769	-0.02482244	-0.1881002	-0.01314545	-0.14316541
-0.39415943	0.26538671	0.26439541	0.51287192	0.01501009

   where $\boldsymbol{\varepsilon}^{\boldsymbol{k}}$, $k = 1, \dots, L$, is a N-dimensional vector which contains the kth Gaussian pseudorandom noise sequence with zero mean and standard deviation σ.

   An example of python code to create five pseudorandom sequences of noise with Gaussian distribution with zero mean ($\mu = 0.0$) and standard deviation $\sigma = 0.3$.

```
N     =  3
L     = 5
sigma = 0.3
mu     = 0.
np.random.seed(seed_0)
noise_matrix = np.random.normal(mu, sigma, (N,L))
```

2) Generate L sets of noise-corrupted data by adding L different Gaussian pseudo-random noise sequences with zero mean and standard deviation σ. Hence, we generate L vectors of noise-corrupted data each one with N elements. For example, if the geophysical data is the N-dimensional vector **y**. Then, we generate L sets of noise-corrupted data calculating:

$$\mathbf{yo}^{\mathrm{k}} = \mathbf{y} + \boldsymbol{\varepsilon}^{\boldsymbol{k}}, \qquad \mathrm{k} = 1,\dots, \mathrm{L},$$

   where $\mathbf{yo}^{\mathrm{k}}$, $k = 1, \dots, L$, is a N-dimensional vector which represents the kth noise-corrupted data .

For example, if the N-dimensional vector $\mathbf{y}$ $(5.\ 10.\ 15)^T$, where the superscript means transposition, by using the L sequences of noise (item 1), we get the following L sets of noise-corrupted data:

yo^1	yo^2	yo^3	yo^4	yo^5
5.53658854	5.13095296	5.02894924	4.44095219	4.91678354
9.89357231	9.97517756	9.8118998	9.98685455	9.85683459
14.60584057	15.26538671	15.26439541	15.51287192	15.01501009

3) Invert the L sets of noise-corrupted data ($\mathbf{yo}^{\mathrm{k}}$, $k = 1, \dots, L$) to obtain the L sets of solution $\hat{\mathbf{p}}^{\mathrm{k}}$, $k = 1, \dots, L$, where $\hat{\mathbf{p}}^{\mathrm{k}}$ is a M-dimensional vector which contains the estimated parameters by inverting noise-corrupted data $\mathbf{yo}^{\mathrm{k}}$.

   For example, if the number of parameters M is equal to 2 and we invert L sets of noise-corrupted data, where L is equal to 5, we obtain L solutions:

	$\hat{p}^1$	$\hat{p}^2$	$\hat{p}^3$	$\hat{p}^4$	$\hat{p}^5$
p_1	1.10	1.12	1.02	1.05	0.99
p_2	4.90	5.15	4.85	5.01	5.07

4) Compute sample standard deviation of the jth estimated parameter by:

$$\tilde{s}(\hat{p}_j) = \left[\frac{\sum_{k=1}^{L}\left(\hat{p}_j^k - \tilde{p}_j\right)^2}{L-1}\right]^{\frac{1}{2}}, \qquad j = 1, \dots, M,$$

   where $\tilde{p}_j$ is the sample mean of the jth parameter

$$\tilde{p}_j = \frac{\sum_{k=1}^{L}\hat{p}_j^k}{L}, \qquad j = 1, \dots, M,$$

   By using the L solutions (item 3), we get:

- the sample standard deviation of the first parameter $\tilde{s}(\hat{p}_1) = 0.054$
- the sample standard deviation of the second parameter $\tilde{s}(\hat{p}_2) = 0.122$

5) Compare the sample standard deviation of the jth estimated parameter $\tilde{s}(\hat{p}_j)$ with the user-specified threshold for the uncertainties of the jth parameter (τ_j).

   The estimated jth parameter is stable if the following inequality is satisfied:

$$\tilde{s}(\hat{p}_j) \leq \tau_j, \qquad j = 1, \dots, M.$$

6) If the sample standard deviation of the jth estimated parameter $\tilde{s}(\hat{p}_j)$ is less or equal to the prior threshold for the uncertainties of the jth parameter (τ_j); hence, the jth estimated parameter is stable. Note that the set of τ_j , $j = 1, \dots, M$, are specified by the interpreter (you).

Here, you can apply this practical procedure to analyze the stability of the solution in the presence of noise in the first and second synthetic tests of the Computational Exercise #1.