

"Uncertainty and Sensitivity Analysis"

Daniela Hanea

Delft University of Technology

Faculty of ITS

D.M.Hanea@ITS.TUdelft.nl

Outline of the talk

- Stochastic noise reaction algorithm
- Applications
- Novikov's Theorem
- Application in sensitivity analysis
- Conclusions

Stochastic noise reaction

$$\min_{x \in \mathbb{R}^n} f(x)$$

Gradient methods iteratively update the current solution x as:

$$x \leftarrow x + \mu \frac{\delta x}{|\delta x|}$$

where:

- $\delta x = -\nabla f(x)$
- μ is a step width determined by line search;
- δx is a decent direction;
- $\nabla f(x)$ is a gradient vector defined by:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T$$

- the stochastic noise reaction injects a Gaussian white noise sequence with zero mean and unit variance $\xi_i \sim N(0, 1)$ into a variable x_i as:

$$x_i^j = x_i + \xi_i^j, j = 1, 2, \dots, M$$

where ξ_i^j denotes the j -th noise in the noise sequence injected into the i -th variable.

- each component of a derivative is approximated by:

$$E \left[\frac{\partial f(x^j)}{\partial x_i^j} \right] = \frac{1}{M} \sum_{j=1}^M f(x^j) \xi_i^j$$

(1)

The algorithm proposed by Koda and Okano is:

1. Initialize the current solution $x = x^0$;

2. Initialize the best solution $x_{best} = x$;

3. For $k = 1, 2, \dots, N$ begin

- Initialize decent direction $\delta_x = 0$;

- For $j = 1, 2, \dots, M (= 100)$ do begin

- Generate a noise vector ξ^j

- $\delta x_i = \delta x_i - f(x_i) \xi_i^j$

- If $f(x_{best}) > f(x^j)$ then $x_{best} = x^j$

end;

4. $w = \max_i |\delta x_i|$

5. $s = \operatorname{argmin}_{s=1,2,\dots,100} f(x + 0.01s \frac{\delta x}{w})$

6. $x = x + \mu \frac{\delta x}{|\delta x|}$

7. If $f(x_{best}) > f(x)$ then $x^{best} = x$

8. If terminal condition is met then goto 10

9. end

10. Output x^{best}

Applications

- If we implement in Matlab the above algorithm, we obtain:

Function:	Result:	Simulation 1:	Simulation 2:	Simulation 3:
$x^2 - 2x + 3$	1	1.0000000000	1	1.0000000000
$-x^3 + 2x^2 + 9x$	-1	-1	-0.99691997	-1
$\frac{x^4}{4} - \frac{x^3}{3} - x^2$	2	2.012101195	1.998607329	1.991849
$\cos x$	$k\frac{\pi}{2}$	-3.27036718	3.171619298	-3.206294765
$x^2 + 10\cos(10x)$	-	-0.312250083	-0.311325948	-0.313530729

- **the travelling salesman problem (TSP):**

- $V = 1, 2, \dots, |V|$ is a set of vertices
- $d_{i,j}$ is the distance between each pair of vertices i and j ,
- **problem:** find an ordering π of vertices that minimizes a tour length defined by:

$$h(\pi) = \sum_{i=1}^{|V|} d_{\pi(i), \pi(i+1)} \tag{2}$$

where the index of π is defined modulo $|V|$, so that vertex $\pi(|V|)$ is adjacent in the tour to both $\pi(|V| - 1)$ and $\pi(1)$ (the vertices are mapped on a plane and the distance is Euclidian).

- the objective function of the TSP, (2), takes a discrete vector π and thus gradient methods cannot be applied directly.

Novikov's Theorem

Approximation (1) is based on Novikov's theorem, which in a simple form is:

$$E \left[\frac{\delta H(\xi)}{\delta \xi_i} \right] = E[H(\xi)\xi_i]$$

where:

- $H(\xi)$ is an arbitrary function of Gaussian stochastic sequence ξ_i , $i = 1, 2, \dots, n$;
- $\frac{\delta H(\xi)}{\delta \xi_i}$ denotes the functional derivative;
- ξ_i is a Gaussian noise with zero mean and unit variance, i.e. $E(\xi_i) = 0$, and $E[\xi_i^t \xi_j^s] = \delta_{ij} \delta_{ts}$, where δ_{ij} and δ_{ts} denote the Kronecker symbol and ξ_i^t denotes the i -th noise in the noise sequence injected into the i -th variable.

Proof:

- the n-dimensional Gaussian distribution:

$$p(x) = \frac{1}{(2\pi)^{n/2} \sqrt{|V|}} e^{-\frac{1}{2}(V^{-1}(x-m), x-m)}$$

with $V = (V_{i,j})_{i,j=\overline{1,n}}$, $V_{i,j} = \text{cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$

- since $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$, $\xi_i \sim N(0, \sigma_i^2)$, $i = \overline{1, n}$:

- $m = 0$

- $V_{i,j} = \begin{cases} 0, & \text{if } i \neq j \\ \sigma_i^2, & \text{if } i = j \end{cases}$

- $|V| = \prod_{i=1}^n \sigma_i^2$

- $(V^{-1}\xi, \xi) = \sum_{i=1}^n \frac{1}{\sigma_i^2} \xi_i^2$

- the Gaussian kernel is:

$$G(\xi) = \frac{1}{(2\pi)^{n/2} \sqrt{\prod_{i=1}^n \sigma_i^2}} e^{-\int_{-\infty}^{+\infty} \sum_i \frac{\xi_i^2(t)}{2\sigma_i^2} dt} \quad (3)$$

- in this case:

$$\begin{aligned} E \left[\frac{\delta H(\xi)}{\delta \xi_i} \right] &= \int \frac{\delta H(\xi)}{\delta \xi_i} G(\xi) \delta \xi = \\ &= \int \frac{\delta}{\delta \xi_i} [H(\xi) G(\xi)] \delta \xi - \int H(\xi) \frac{\delta}{\delta \xi_i} G(\xi) \delta \xi \\ &= - \int H(\xi) \frac{\delta}{\delta \xi_i} G(\xi) \delta \xi \end{aligned} \quad (4)$$

- using relation (3):

$$\begin{aligned} \frac{\delta}{\delta \xi_i} G(\xi) &= \frac{1}{(2\pi)^{n/2} \sqrt{\prod_{i=1}^n \sigma_i^2}} e^{-\int_{-\infty}^{+\infty} \sum_i \frac{\xi_i^2(t)}{2\sigma_i^2} dt} \left(-\frac{\xi_i}{\sigma_i^2} \right) \\ &= -\frac{\xi_i}{\sigma_i^2} G(\xi) \end{aligned}$$

(5)

- replacing in (4):

$$\begin{aligned} E \left[\frac{\delta H(\xi)}{\delta \xi_i} \right] &= \frac{1}{\sigma_i^2} \int H(\xi) \xi_i G(\xi) \delta \xi \\ &= \frac{1}{\sigma_i^2} E[H(\xi) \xi_i] \end{aligned}$$

(6)

Remarks regarding Novikov's Theorem:

- for $\sigma_i = 1$, the Novikov's theorem is:

$$E \left[\frac{\delta H(\xi)}{\delta \xi_i} \right] = E[H(\xi)\xi_i]$$

- the Novikov's theorem holds also for Gaussian noise with mean different of zero, as in this case, where $E[x^j] = E[x + \xi^j] = x + E[\xi^j] = x$.

Application in sensitivity analysis

- $\left. \frac{\partial E[X_i|Z=z_0]}{\partial z} \right|_{z=z_0}$, where $Z = g(X_1, X_2, \dots, X_n)$.
- **Example:** $Z=Z(X,Y)=2X+Y$, with $X, Y \sim U(0, 1)$, independent variables;

Z	z_0	$\frac{\partial E(X/z_0)}{\partial z_0}$	simulation 1	simulation 2	simulation 3
2X+Y	0.25	0.25	0.0585	0.036	0.0537
	0.5	0.25	0.0675	0.0343	0.0171
	1.5	0.5	0.0838	0.1602	0.1652
	2.5	0.25	0.0150	0.0250	-0.0259

- The above simulations don't give the correct results
- We have to verify the relation (1) for some simple examples.
- Consider a function of only one variable and add to this variable a noise

$$\xi \sim N(0, 1)$$

- relation (1) becomes:

$$\frac{\partial f(x)}{\partial x} \approx E \left[\frac{\partial f(\tilde{x})}{\partial \tilde{x}} \right] = f(\tilde{x})\xi$$

or, more correct:

$$\frac{\partial f(x)}{\partial x} \approx E \left[\frac{\partial f(\tilde{x})}{\partial \tilde{x}} \right] = E[f(\tilde{x})\xi] \tag{7}$$

where $\tilde{x} = x + \xi$.

Example 1:

$$f(x) = x^2 \implies \frac{\partial f}{\partial x} = 2x$$

$$\begin{aligned} E[f(\tilde{x})\xi] &= E[(x + \xi)^2\xi] = \\ &= E[x^2\xi + 2x\xi^2 + \xi^3] \\ &= x^2E[\xi] + 2xE[\xi^2] + E[\xi^3] \end{aligned}$$

(8)

Since $\xi \sim N(0, 1)$, we have: $E[\xi^n] = \begin{cases} 0, & \text{if } n = 2r + 1 \\ \frac{(2r)!}{2^r r!}, & \text{if } n = 2r \end{cases}$

So:

$$E[f(\tilde{x})\xi] = 2x$$

Example 2:

$$f(x) = x^3 \implies \frac{\partial f}{\partial x} = 3x^2$$

$$\begin{aligned} E[f(\tilde{x})\xi] &= E[(x + \xi)^3\xi] = E[x^3\xi + 3x^2\xi^2 + 3x\xi^3 + \xi^4] = \\ &= x^3E[\xi] + 2x^2E[\xi^2] + 3xE[\xi^3] + E[\xi^4] = \\ &= 3x^2 + \frac{4!}{2^22!} = 3x^2 + 3 \end{aligned}$$

(9)

Example 3:

$$f(x) = e^x \implies \frac{\partial f}{\partial x} = e^x$$

$$\begin{aligned} E[f(\tilde{x})\xi] &= E[e^{x+\xi}\xi] = e^x E[\xi e^\xi] = \\ &= e^x E\left[\xi \sum_{n=0}^{\infty} \frac{\xi^n}{n!}\right] = \\ &= e^x \sum_{n=0}^{\infty} \frac{1}{n!} E[\xi^{n+1}] = \\ &= e^x \sum_{r=0}^{\infty} \frac{1}{(2r+1)!} \frac{(2r+1)!}{2^r r!} = \\ &= e^x \sum_{r=0}^{\infty} \frac{\left(\frac{1}{2}\right)^r}{r!} = e^x e^{1/2} = e^{x+1/2} \end{aligned}$$

(10)

We obtain the same conclusion using UNICORN.

The result for the third example is:

$$f(x) = \frac{1}{17e^x} \sum_{i=1}^{17} \xi_i e^{x+\xi_i}$$

For $x = 1$, the result is:

Nr.	Variable	Mean	Variance
1	sum	1.56E+000	1.22E+000

The series of approximations used in article:

$$\frac{\partial f(x)}{\partial x_i} \approx E \left[\frac{\partial f(x^j)}{\partial \xi_i} \right] = E[f(x^j \xi_i)] \approx \frac{1}{M} \sum_{j=1}^M f(x^j) \xi_i^j$$

where we use :

$$\frac{\partial f(x^j)}{\partial x_i^j} = \frac{\frac{\partial f(x^j)}{\partial \xi_i}}{\frac{\partial x_i^j}{\partial \xi_i}} = \frac{\partial f(x^j)}{\partial \xi_i}$$

- to obtain $E \left[\frac{\partial f(x^j)}{\partial \xi_i} \right] = E[f(x^j)\xi_i]$, we can apply the Novikov's theorem for ξ with mean x , because: $E[x^j] = E[x + \xi^j] = x + E[\xi^j] = x$.

- the proof of this for only one noise:

$$\begin{aligned}
 E \left[\frac{\partial f(x(\xi))}{\partial \xi} \right] &= \int \frac{\partial f}{\partial \xi}(x + \xi)p(\xi)d\xi = \\
 &= \int \frac{\partial}{\partial \xi}[f(x + \xi)p(\xi)]d\xi - \int f(x + \xi)\frac{\partial p}{\partial \xi}d\xi = \\
 &= \int f(x + \xi)\xi p(\xi)d\xi = E[f(x + \xi)\xi]
 \end{aligned}$$

(11)

- this is true for a function f with polynomial growth at infinity.

- the first approximation:

$$\frac{\partial f}{\partial x_i} \approx E \left[\frac{\partial f(x^j)}{\partial x_i^j} \right]$$

- to simplify the notation, we take f with only one variable and we inject only one noise $\xi \sim N(0, 1)$.

- if g is a nonlinear function ($f''(x) \neq 0$) and $\xi \sim N(0, 1)$, then:

$$\begin{aligned} E[g(x + \xi)] &= E[g(x) + \frac{\partial g}{\partial x}\xi + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\xi^2 + \dots] = \\ &= g(x)\frac{\partial g}{\partial x}E[\xi] + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}E[\xi^2] + \frac{1}{3!}\frac{\partial^3 g}{\partial x^3}E[\xi^3] + \dots \\ &= \sum_{r=0}^{\infty} \frac{\partial^r g}{\partial x^r} E[\xi^{2r}] = \\ &= \sum_{r=0}^{\infty} \frac{(2r)!}{2^r r!} \frac{\partial^r g}{\partial x^r} \end{aligned}$$

(12)

- for $g(x) = \frac{\partial f(x)}{\partial x}$, we obtain:

$$\begin{aligned}
 E \left[\frac{\partial f(x + \xi)}{\partial(x + \xi)} \right] &= \frac{\partial f(x)}{\partial x} + \frac{1}{2} \frac{\partial^3 f(x)}{\partial x^3} E[\xi^3] + \frac{1}{4!} \frac{\partial^5 f}{\partial x^5} E[\xi^4] + \dots \\
 &= \frac{\partial f(x)}{\partial x} + \frac{1}{2} \frac{\partial^3 f(x)}{\partial x^3} + \frac{1}{4!} \frac{\partial^5 f}{\partial x^5} \frac{4!}{2^2 2!} + \dots
 \end{aligned}$$

(13)

- the error is:

$$\begin{aligned}
 error &= E \left[\frac{\partial f(x + \xi)}{\partial(x + \xi)} \right] - \frac{\partial f(x)}{\partial x} \\
 &= \sum_{n=1}^{\infty} \frac{1}{(2n)!} \frac{\partial^{2n+1} f(x)}{\partial x^{2n+1}} E[\xi^{2n}] \\
 &= \sum_{n=1}^{\infty} \frac{\partial^{2n+1} f(x)}{\partial x^{2n+1}} \frac{1}{2^n n!}
 \end{aligned}$$

(14)

Conclusions:

- The approximation (1) is not good for any function f ;
- The results for optimization problems presented in the first table are good, even if the approximation is not good. But in algorithm, the error is reduced. This can be seen in the following graph, which presents every x^{best} from simulation.
- Unfortunately, the approximation of the derivative proposed by Okano and Kodama can not be used in sensitivity analysis to calculate: $\left. \frac{\partial E[X_i|Z=z_0]}{\partial z} \right|_{z=z_0}$, where $Z = g(X_1, X_2, \dots, X_n)$.