[6G] Stochastic Analysis of Twice differtiable

Stochastic Analysis

Stochastic Analysis는 앞에서와 마찬가지의 증명 과정을 따른다. 그런데, Stochastic의 경우 는 보다 고차원의 미분값에 대한 생각을 하지 않을 수가 없다.

The stochastic analysis for an optimization algorithm follows the same procedure of a conventional proof . However, in the stochastic analysis, we should consider the high dimensional differential of an objective function, due to the properties of a random process.

Definition of Stochastic Model

일단 다음과 같이 생각한다.

Consider the following stochastic model

Model

Consider the random process $X_t \in \mathbf{R}^n$ with a Wiener process $W_t \in \mathbf{R}^n$ with a constant variance $\Sigma \in \mathbf{R}^{n \times n}$ which is symmetric matrix, such that

$$X_t = x_t^Q + \Sigma W_t \tag{1}$$

where $x_t^Q \in \mathbf{R}^n$ is a deterministic value.

Let the other randome process $Y_t \in \mathbf{R}^n$ such that

$$Y_t = X_{t+1} = x_{t+1}^Q - (\lambda_t h_t)^Q + \Sigma W_{t+1} \tag{2}$$

Let a random process $Z_t(s) \in \mathbf{R}^n$ for $s \in \mathbf{R}[0,1]$ such that

$$\begin{aligned} Z_t(s) &= X_t + s(Y_t - X_t) \\ &= x_t^Q + \Sigma W_t + s(-\lambda_t h_t) + s\Sigma(W_{t+1} - W_t) \end{aligned}$$

Let $\Delta W_t = W_{t+1} - W_t$, then

$$Z_t(s) = x_t^Q - s(\lambda_t h_t) + \Sigma(W_t + s\Delta W_t) \tag{3}$$

Considering the differntiate of $Z_t(s)$ to s, we obtain

$$\frac{dZ_t(s)}{ds} = -\lambda_t h_t + \Sigma \Delta W_t$$

By the definition of the stochastic differential, we define the differential of $Z_t(s)$ such that

$$dZ_t(s) = -\lambda_t h_t ds + \Sigma \Delta W_t ds \tag{4}$$

Consider the final term, i.e. $\Delta W_t ds$. The integration of $dZ_t(s)$ to s is as follows:

$$\int_0^s dZ_t(s) = Z_t(s) - Z_t(0) = -s\lambda_s h_t \int_0^s ds + \Sigma \int_0^s \Delta W_t ds. \tag{5}$$

Since the integration of $dZ_t(s)$ should be equal to (3), the integration of final term to ΔW_t is evaluated as follows.

$$\int_0^s \Delta W_t ds = s(W_{t+1} - W_t) = \int_t^{t+1} s dW_\tau \tag{6}$$

Thereby, when $s = 0$, $\tau = t$, and $s = 1$, $\tau = t + 1$, we obtain the following differential equation in the sense of (3)

$$\Delta W_t ds = s dW_\tau$$

In consequence, the stochastic differential equation of $Z_t(s)$ is same to the following:

$$dZ_t(s) = -\lambda_t h_t ds + s \cdot \Sigma dW_\tau \in \mathbf{R}^n \tag{7}$$

In (7), by the product rule of the stochastic differential, the dot product of the vector differential $dZ_t(s)$ is evaluated as

$$dZ_t(s)^2 = dZ_t(s)^T dZ_t(s) = s^2 dW_\tau \Sigma^T \Sigma dW_\tau = s^2 Tr(\Sigma\Sigma^T) d\tau \in \mathbf{R} \tag{8}$$

In (8), while $d\tau$ and ds contain the same domain, the scale of both are differnet. Since when s is increased from 0 to s, the τ is increased from 0 to 1. If let $\bar{s} = \max s, \ \forall s \in \mathbf{R}[0, 1]$ in the analysis, then

$$\tau = t + \frac{s}{\bar{s}}, \quad \text{for} \tau = \begin{cases} t+1 & s = \bar{s} \\ t & s = 0 \end{cases}$$

Considering the scale of both parameters we can obtain the following relation.

$$d\tau = \frac{1}{\bar{s}} ds \tag{9}$$

Therefore, from (8), we obtain the dot product of the vector differential $dZ_t(s)$ is using (9) when $s = \bar{s}$,

$$dZ_t(s)^2 = \frac{s^2}{\bar{s}} Tr(\Sigma\Sigma^T) ds = s \cdot Tr(\Sigma\Sigma^T) ds \tag{10}$$

Deduction of an exact form of the Taylor expansion with the twice differentiable form

The deterministic version of the exact Taylor expansion with the twice differentiable form for a objective function $f(x) : \mathbf{R}^n \to \mathbf{R}$ is

$$f(y) - f(x) = \langle \nabla f(x), y - x \rangle + \int_0^1 (1-s) \langle y - x, H(x + s(y-x))(y-x) \rangle ds$$

where $H(x) \in \mathbf{R}^{n\times n}$ is a Hessian of $f(x)$.

For evaluation of the stochastic version, we let a function $g(s) = f(Z_t(s))$. The first order differentiation to s is

$$
\begin{aligned}
\frac{dg(s)}{ds} &= \frac{1}{ds}(dg(s)) \\
&= \frac{1}{ds}\left(\frac{\partial f(Z_t(s))}{\partial Z_t(s)} dZ_t(s) + \frac{1}{2}\frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)} dZ_t(S)^2\right) \\
&= \left(\frac{\partial f(Z_t(s))}{\partial Z_t(s)}\frac{dZ_t(s)}{ds} + \frac{1}{2}\frac{1}{ds}\frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)} dZ_t(S)^2\right)
\end{aligned} \tag{11}
$$

Substituting $(???)$ and (10) to the (11), we can obtain

$$
\begin{aligned}
\frac{dg(s)}{ds} &= \nabla f(X_t)^T \frac{dZ_t(s)}{ds} + \frac{1}{2}\frac{1}{ds} s \cdot Tr\left(\Sigma \frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)}\Sigma^T\right) ds \\
&= \langle \nabla f(X_t), -(\lambda_t h_t)^Q + \Sigma \Delta W_t\rangle + \frac{1}{2} s \cdot Tr\left(\Sigma \frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)}\Sigma^T\right).
\end{aligned} \tag{12}
$$

For the second order diiferetiation of $g(s)$, let $y(Z_t(s)) = \frac{dg(s)}{ds}$. Then

$$
\frac{d^2 g(s)}{ds^2} = \frac{dy}{ds} = \left(\frac{\partial y(Z_t(s))}{\partial Z_t(s)} \cdot \frac{dZ_t(s)}{ds} + \frac{1}{2} s \cdot Tr\left(\Sigma \frac{\partial^2 y(Z_t(s))}{\partial Z_t(s)^2}\Sigma^T\right)\right) \tag{13}
$$

For the first term of (13), we span it to the differtial of $f(X)$ such that

$$
\begin{aligned}
\frac{\partial y(Z_t(s))}{\partial Z_t(s)} \cdot \frac{dZ_t(s)}{ds} &= \frac{\partial}{\partial Z_t(s)}\left(\frac{\partial f(Z_t(s))}{\partial Z_t(s)} \cdot \frac{dZ_t(s)}{ds}\right) \cdot \frac{dZ_t(s)}{ds} \\
&= \frac{\partial^2 f(Z_t(s))}{\partial Z_t(s)^2} \cdot \left(\frac{dZ_t(s)}{ds}\right)^2 + \frac{\partial f(Z_t(s))}{\partial Z_t(s)} \cdot \frac{\partial^2 f(Z_t(s))}{\partial Z_t(s) \partial s} \cdot \frac{\partial Z_t(s)}{\partial s}
\end{aligned} \tag{14}
$$

Subsequently, by the definition of vedtor valued differetiation, the first term of (14) is

$$
\frac{\partial^2 f(Z_t(s))}{\partial Z_t(s)^2} \cdot \left(\frac{dZ_t(s)}{ds}\right)^2 = \langle \frac{dZ_t(s)}{ds}, \frac{\partial^2 f(Z_t(s))}{\partial Z_t(s)^2}\frac{dZ_t(s)}{ds}\rangle.
$$

In addition, for the analysis of the second term, we evaluate the following differentiation as follows.

$$
\frac{\partial^2 f(Z_t(s))}{\partial Z_t(s) \partial s} = \frac{\partial}{\partial Z_t(s)}\left(\frac{\partial f(Z_t(s))}{\partial s}\right) = \frac{\partial}{\partial Z_t(s)}\left(-(\lambda_t h_t)^Q + \Sigma \Delta W_t\right) = 0
$$

For the verification, of the above equation, changing the order of differentiation , we obtain

$$
\frac{\partial}{\partial s}\frac{\partial Z_t(s)}{\partial Z_t(s)} = 0.
$$

Therefore, the first term of (14) is

$$
\frac{\partial y(Z_t(s))}{\partial Z_t(s)} \cdot \frac{dZ_t(s)}{ds} = \langle \frac{dZ_t(s)}{ds}, \frac{\partial^2 f(Z_t(s))}{\partial Z_t(s)^2}\frac{dZ_t(s)}{ds}\rangle. \tag{15}
$$

For the second term of (13), we differentiate twice $y(Z_t(s))$ with respect to $Z_t(s)$ as follows.

$$
\begin{aligned}
\frac{\partial^2 y(Z_t(s))}{\partial Z_t^2(s)} &= \frac{\partial^2}{\partial Z_t^2(s)}\left(\frac{dg(s)}{ds}\right) \\
&= \frac{\partial^2}{\partial Z_t^2(s)}\left(\frac{\partial f(Z_t(s))}{\partial Z_t(s)}\frac{dZ_t(s)}{\partial ds} + \frac{1}{2}Tr\left(\Sigma\frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)}\Sigma^T\right)\right)
\end{aligned} \tag{16}
$$

In (16), the first term of is evaluated as

$$
\begin{aligned}
&\frac{\partial^2}{\partial Z_t^2(s)}\left(\frac{\partial f(Z_t(s))}{\partial Z_t(s)}\frac{dZ_t(s)}{ds}\right) \\
&= \frac{\partial}{\partial Z_t(s)}\left(\frac{\partial^2 f(Z_t(s))}{\partial Z_t^2(s)}\frac{dZ_t(s)}{ds} + \frac{\partial f(Z_t(s))}{\partial Z_t(s)}\frac{\partial^2 Z_t(s)}{\partial Z_t(s)\partial s}\right) \\
&= \frac{\partial^3 f(Z_t(s))}{\partial Z_t^3(s)}\frac{dZ_t(s)}{ds} \in \mathbf{R}^{n\times n}, \quad \because \frac{\partial^2 Z_t(s)}{\partial Z_t(s)\partial s} = 0
\end{aligned} \tag{17}
$$

Additionally, the second term is

$$
\frac{\partial^2}{\partial Z_t^2(s)}H(Z_t(s)) = \frac{\partial^4}{\partial Z_t^4(s)}f(Z_t(s)) \tag{18}
$$

where $\frac{\partial^4}{\partial Z_t^4(s)}f(Z_t(s))$ is a rank-4 tensor such that

$$
\Sigma\frac{\partial^4}{\partial Z_t^4(s)}f(Z_t(s))\Sigma^T \in \mathbf{R}^{n\times n}
$$

Finally, since the exact expansion of the twice differential form is evaluated such that

$$
f(Y_t) - f(X_t) = g(1) - g(0) = \frac{dg}{ds}(0) + \int_0^1 (1-s)\frac{d^2 g}{ds^2}(s)ds, \tag{19}
$$

from (11) to (18), we obtain the following exact expansion of the twice differentiable form.

$$
\begin{aligned}
f(Y_t) - f(X_t) &= \langle \nabla f(Z_t(s)), \frac{dZ_t(s)}{ds}\rangle\Big|_{s=0} + \frac{1}{2}s \cdot Tr\left(\Sigma H(Z_t(s))\Sigma^T\right)\Big|_{s=0} \\
&+ \int_0^1 (1-s)(\langle\frac{dZ_t(s)}{ds}, H(Z_t(s))\frac{dZ_t(s)}{ds}\rangle \\
&+ \frac{1}{2}Tr\left(\Sigma, \left(\frac{\partial^3 f(Z_t(s))}{\partial Z_t^3(s)}\frac{dZ_t(s)}{ds} + \frac{1}{2}\Sigma\frac{\partial^4 f(Z_t(s))}{\partial Z_t^4(s)}\Sigma\right)\Sigma^T\right))ds
\end{aligned} \tag{20}
$$