Journal of Nonlinear Modeling and Analysis
Volume 3, Number 3, September 2021, 431–446
http://jnma.ca; http://jnma.ijournal.cn
DOI:10.12150/jnma.2021.431

2D Random Approximation Method*

Changan Liu[1,†] and Miao Ouyang[2]

Abstract H. Robbins and S. Monro studied the stochastic approximations of one-dimensional system. In this paper, we present the stochastic approximation method of 2D system.

Keywords 2D systems, Stochastic approximation, Mathematical expectation, Convergence.

MSC(2010) 62L20, 60B10.

1. Introduction

H. Robbins and S. Monro studied the stochastic approximation of one-dimensional system in [24]. However, there are a large number of 2D stochastic systems in stochastic fluid mechanics [2], especially in the diffusion of random electronic gas in magnetic region [29], random information flow [18,28] and other engineering fields. Recently, scholars have shown interest in the method of stochastic approximation, a lot of work has been done [3,7,8,14,30] and they are of great significance. The stochastic approximation method can be used to solve some random or random problems as well as some deterministic mathematical problems [16,23,31].

To a less extent, we investigate methods like stochastic average gradient [25], which performs well on objectives that are strongly-convex, and stochastic variance reduced gradient [11]. Fort et al. [6] establish results on the geometric ergodicity of hybrid samplers and in particular for the random-scan Gibbs sampler. Zhao and Wang [32], Jansen et al. [10] and Kriegesmann [13] estimate the statistical moments of the compliance by Monte Carlo approximations. Sparse polynomial chaos expansions [1,4,5,9] can be used to reduce the computational cost, but the computational cost associated with this approach becomes prohibitive for a large number of problems with uncertain inputs. [21,22,26,27] use the stochastic approximation algorithm of Polyak-Ruppert averaging to favor the performance of stochastic approximation. Convergence results for mini-batch EM and SAEM algorithms appear recently in [17,19] and [12] respectively.

Moreover, the standard method for stochastic root-finding problems is stochastic approximation [15,20,24]. In these 2D stochastic systems, take α as a constant, consider a region D on the plane R^2, and find the equation satisfied by the unknown

[†]the corresponding author.
Email address: changanliu@cuhk.edu.hk (C. Liu), mouyang@mail.sdu.edu.cn (M. Ouyang)
[1]Department of Medicine and Therapeutics Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China
[2]School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China
*The authors were supported by Key Program of National Natural Science Foundation of China (Nos. U1806203, 61533011).

measurable regression function $M(x)$ with error:

$$M(x) = \alpha. \tag{1.1}$$

The null point $x = \theta$ of the above equation is an ubiquitous and important problem in system identification, adaptive control, pattern recognition, adaptive filtering and neural network and other fields.

Generally, $M(x)$ represents a mathematical expected value at time x of a certain experiment, which is an unknown function. However, for any x, the value of $M(x)$ is measurable, and assume that $M(x)$ is a monotone function of x in an unknown experiment. To obtain the null point $x = \theta$ of (1.1), we need to design an algorithm to determine a series of values $\{x_{mn}\}_{m,n\geq 0}$ in the region D of the plane R^2, which are

$$\begin{gathered} x_{00}, x_{01}, , x_{02} \cdots\cdots, \\ x_{10}, x_{11}, x_{12} \cdots\cdots, \\ x_{20}, x_{21}, x_{22} \cdots\cdots, \\ \cdots\cdots\cdots\cdots\cdots\cdots, \end{gathered}$$

these values satisfy in a probabilistic way, $\lim\limits_{\substack{m\to\infty \\ n\to\infty}} x_{mn} = \theta$.

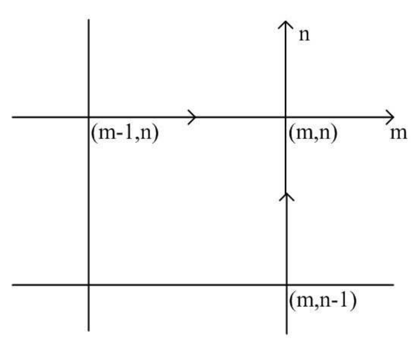
As mentioned above, for any x_{mn}, $M(x_{mn})$ can be measured, so it can provide information for the next measured value. Notice that there are two coordinate positions for the next measured value related to x_{mn}, $(m+1, n)$ and $(m, n+1)$, so there are two points,

$$x_{m+1,n}, x_{m,n+1}. \tag{1.2}$$

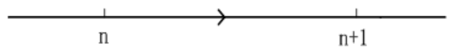
However, for the next measured value, considering the convenience of researching problem, we usually regard x_{mn} as the value to be measured in the next step. Therefore, there are two values related to x_{mn} directly, which are

$$x_{m-1,n}, x_{m,n-1}. \tag{1.3}$$

Define $D = \left\{ (m,n) / \begin{matrix} m \geq 0 \\ n \geq 0 \end{matrix} \right\}$, then $D \subset R^2$. Besides, for any $(m, n) \in D$, considering a mathematical sequence $\{x_{mn}\}_{m,n\geq 0}$, and the boundary values x_{m0}, x_{0n} are known. Therefore, $M(x_{m0})$ and $M(x_{0n})$ are determined. Thus, for any value x_{mn} in D, the points directly connected with x_{mn} have two coordinate positions, $(m-1, n)$ and $(m, n-1)$.



Therefore, the value $M(x_{mn})$ at the point x_{mn} is determined by the measured values $M(x_{m-1,n})$ and $M(x_{m,n-1})$, which are related to points $x_{m-1,n}$ and $x_{m,n-1}$ directly. As can be seen, this approach is similar to the one-dimensional measuring form, because in the one-dimensional random approximation, the measurements of $M(x_n)$ and $M(x_{n+1})$ from point x_n to point x_{n+1} are as follows.



There is only one point n directly associated with the point $n+1$, so $2D$ is a natural extension of $1-D$. To determine the root $x=\theta$ of (1.1), we usually select the sequence values,

$$\left.\begin{array}{l} x_{00}, x_{01}, x_{02}\dots x_{0r} \\ x_{10}, x_{11}, x_{12}\dots x_{1r} \\ x_{20}, x_{21}, x_{22}\dots x_{2r} \\ \dots\dots\dots\dots\dots\dots \\ x_{s0}, x_{s1}, x_{s2}\dots x_{sr} \end{array}\right\}, \tag{1.4}$$

and a series of new observations on D can be obtained continuously,

$$\{\ x_{mn}\ \}_{m>s,n>r}. \tag{1.5}$$

From these new values, these function values in (1.4) can be determined, which are

$$\left.\begin{array}{cccc} M(x_{00}) & M(x_{01}) & \cdots & M(x_{0r}) \\ M(x_{10}) & M(x_{11}) & \cdots & M(x_{1r}) \\ \cdots & \cdots & \cdots & \cdots \\ M(x_{s0}) & M(x_{s1}) & \cdots & M(x_{sr}) \end{array}\right\}, \tag{1.6}$$

and the derivatives of these functions in (1.6),

$$\{M'(x_{ij})\}_{i,j=0,0}^{s,r}. \tag{1.7}$$

Without considering arbitrary initial values $\{x_{ij}\}_{i,j=0,0}^{s,r}$, for the special function $M(x)$ and the value α, satisfy

$$\lim_{\substack{i\to\infty \\ j\to\infty}} x_{ij} = \theta, \tag{1.8}$$

this method is valid, and convergence rate of (1.8) can be obtained by practical application and specific calculation about $\{x_{ij}\}_{i,j=0,0}^{s,r}$.

For the above problem, we consider the following random method. The character of the function $M(x)$ is an unknown experiment. Suppose each value x corresponds to a random variable $Y=Y(x)$, and the distribution function is $\Pr\left[Y(x)\le y\right] = H\left(y|x\right)$, which enable

$$M(x) = \int_{-\infty}^{\infty} y dH\left(y|x\right) \tag{1.9}$$

to be the expected value of the random variable Y for which $\{x_{ij}\}_{i,j=0}^{s,r}$ is given. For the experimenters, and the exact properties of $H\,(y|x)$ and $M(x)$ are unknown, but there is a unique solution θ of $M(x) = \alpha$. By obtaining the continuous observational data at the point $\{x_{ij}\}_{i,j=0}^{s,r}$, and utilizing some explicit experimental methods, the value of θ can be estimated. For any initial value $\{x_{ij}\}_{i,j=0}^{s,r}$, the probability method is used, if (1.8) holds, for the given $H\,(y|x)$ and the value α, we usually refer it to the approximation process of random field. Next, we will give a detailed method to estimate the value of θ under the premise of some constraints on the properties of $H\,(y|x)$. These constraints are strict and cannot be weakened arbitrarily, but they generally meet the actual needs. Here, the method is not required for obtaining optimal properties, but the results given at least point out that stochastic approximation is useful and worthy for further study.

2. Convergence theorem

For each $x_{m-1,n}$ and $x_{m,n-1}$, $M(x_{m-1,n})$, $M(x_{m,n-1})$ and $M(x_{m-1,n} + \omega x_{m,n-1})$ can be measured, where ω is a regulatory parameter. In this way, they can provide information for determining the next x_{mn}. Considering the values of $M(x)$ measured at different times, set the expectation of measured error at any time be zero, and it depends on the value of x. If y_{mn} represents the measurement at the point (m, n), then the actual values of $M(x_{m-1,n})$, $M(x_{m,n-1})$ and $M(x_{m-1,n} + \omega x_{m,n-1})$ are obtained, they can be expressed as:

$$y_{mn} = M(x_{m-1}, n) + r_0 M(x_{m,n-1}) + r_1 \varepsilon_{mn}, \tag{2.1}$$

where r_0 and r are the regulation parameters.

Since the measured value of position (m, n) in D is

$$x_{mn} - c_0(x_{m-1,n} + \omega x_{m,n-1}) = a_{mn}(\alpha - y_{mn}), \tag{2.2}$$

where a_{mn} is called the gain coefficient. It is noted that the directly connected points associated with y_{mn} are $y_{m-1,n}$ and $y_{m,n-1}$, so y_{mn} can be replaced by $y_{m-1,n} + \omega y_{m,n-1}$. Thus, from (2.2), we have

$$x_{mn} = c_0(x_{m-1,n} + \omega x_{m,n-1}) + a_{mn}[\alpha - c_1(y_{m-1,n} + \omega y_{m,n-1})]. \tag{2.3}$$

For any x, suppose $H\,(y|x)$ is a distribution function related to y, and there is a positive constant C satisfy

$$Pr\,(\;|\;Y(x)\;|\;\leq C\;)\;=\;\int_{-C}^{C} dH(y|x) = 1. \tag{2.4}$$

Particularly, the expected values defined by (1.9) exist for any x. Suppose that there is a finite α and θ,

$$M(x) \leq \alpha, when \quad x < \theta, \quad M(x) \geq \alpha, when \quad x > \theta, \tag{2.5}$$

and importantly, whether there is $M(\theta) = \alpha$.

Let $\{ a_{mn} \}$ be a positive constant sequence of two indexes, and satisfy

$$0 < \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn}^2 = A < \infty. \tag{2.6}$$

For any $(m, n) \in D$, take x_{m0} and x_{0n} as arbitrary constants, and define a double index $\{ x_{mn} \}$. For the convenience of solving this problem, take $c_0 = c_1 = \frac{1}{2}$ and $\omega = 1$ in (2.3), then

$$x_{mn} - \frac{1}{2}(x_{m-1,n} + x_{m,n-1}) = a_{mn} \left[\alpha - \frac{1}{2}(y_{m-1,n} + y_{m,n-1}) \right], \tag{2.7}$$

y_{mn} is a random field, and satisfy

$$Pr\left[y_{mn} \leq y | x_{mn} \right] \; = \; H(y|x_{mn}). \tag{2.8}$$

Let s and t be integer variables greater than or equal to zero, and regard m and n as integer parameters temporarily. Besides, make $s \leq m$ and $t \leq n$, take μ as a real parameter, then make the following transformation,

$$\begin{aligned} b_{st} = \mu^{(t-n)} \Big\{ & \tfrac{1}{8} E \left[\tfrac{1}{2} {x^2}_{st} + 2 x_{m-1,n} x_{m,n-1} + \left(\tfrac{1}{2} \right)^{-(t-n)} {x^2}_{m,n-1} \right. \\ & \left. + \left(\tfrac{1}{2} \right)^{-(s-m)} {x^2}_{m-1,n} \right]^{(m+n)-(s+t)} - \tfrac{1}{2} \left[\theta E \left(x_{m-1,n} + x_{m,n-1} \right) \right]^{(m+n)-(s+t)} \\ & + \tfrac{1}{2} \left(E\theta^2 \right)^{(m+n)-(s+t)} + \omega \left(\tfrac{\delta}{2r} \right)^{(m+n)-(s+t)} E(x_{mn} - \theta)^{\,2} \Big\}, \end{aligned} \tag{2.9}$$

where ω, r, δ are the real parameters, and from (2.9), when $s = m-1, \ t = n$, we have

$$\begin{aligned} b_{m-1,n} = & \tfrac{1}{8} E(x_{m-1,n} + \ x_{m,n-1})^2 - \tfrac{1}{2} \theta E \left(x_{m-1,n} + x_{m,n-1} \right) \\ & + \tfrac{1}{2} (E\theta)^2 + \tfrac{\omega\delta}{2r} E(x_{mn} - \theta)^2. \end{aligned}$$

Moreover, when $s = m, \ t = n-1$, we have

$$\begin{aligned} b_{m,n-1} = \tfrac{1}{\mu} \Big\{ & \tfrac{1}{8} E(x_{m-1,n} + x_{m,n-1})^2 - \tfrac{1}{2} \theta E \left(x_{m-1,n} + x_{m,n-1} \right) \\ & + \tfrac{1}{2} (E\theta)^2 + \tfrac{\omega\delta}{2r} E(x_{mn} - \theta)^2 \Big\}. \end{aligned}$$

Therefore,

$$b_{m-1,n} + \mu b_{m,n-1} = E \left[\frac{(x_{m-1,n} + x_{m-1,n})}{2} - \theta \right]^2 + \frac{\omega\delta}{r} E(x_{mn} - \theta)^2. \tag{2.10}$$

In the meantime, we can get

$$b_{m-1,n} = \mu b_{m,n-1}. \tag{2.11}$$

Similarly, when $s = m, \ t = n$, we have

$$b_{mn} = \ \omega \ E(x_{mn} - \theta)^2. \tag{2.12}$$

From (2.12), for $(m, n) \in N_0^2$, take arbitrary initial values $x_{m0}, \ x_{0n}$, we will find the condition enable

$$\lim_{m, \ n \to \infty} \ b_{mn} = 0. \tag{2.13}$$

Obviously, (2.13) means that $\{ x_{mn} \}$ converges to θ in a probabilistic way.

$$
\begin{aligned}
& b_{m-1,n} + \mu b_{m,n-1} \\
&= E\Big[\tfrac{(x_{m-1,n}+x_{m-1,n})}{2} - \theta \Big]^2 + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= E[x_{mn} - \theta + a_{mn}(y-\alpha)]^2 + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= E\{ \int_{-\infty}^{+\infty} [(x_{mn}-\theta) + a_{mn}(y-\alpha)]^2 \; dH(y/x_{mn})]\} + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= E[\int_{-\infty}^{+\infty} (x_{mn}-\theta)^2 dH(y/x_{mn}) + \int_{-\infty}^{+\infty} a^2{}_{mn}(y-\alpha)^2 dH(y/x_{mn}) \\
&\quad +2a_{mn} \int_{-\infty}^{+\infty} (x_{mn}-\theta)(y-\alpha) dH(y/x_{mn})] + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= E(x_{mn}-\theta)^2 + a^2{}_{mn} E[\int_{-\infty}^{+\infty} (y-\alpha)^2 dH(y/x_{mn})] \\
&\quad +2a_{mn} E(x_{mn}-\theta)[\int_{-\infty}^{+\infty} y dH(y/x_{mn}) - \int_{-\infty}^{+\infty} \alpha dH(y/x_{mn})] + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= E(x_{mn}-\theta)^2 + a^2{}_{mn} E[\int_{-\infty}^{+\infty} (y-\alpha)^2 dH(y/x_{mn})] + 2a_{mn} E(x_{mn}-\theta)[M(x)-\alpha] \\
&\quad + \tfrac{\omega\delta}{r} E(x_{mn}-\theta)^2 \\
&= (\tfrac{1}{\omega} + \tfrac{\delta}{r}) b_{mn} + a^2{}_{mn} E[\int_{-\infty}^{+\infty} (y-\alpha)^2 dH(y/x_{mn})] + 2a_{mn} E(x_{mn}-\theta)[M(x)-\alpha]
\end{aligned}
\tag{2.14}
$$

Make

$$d_{mn} = E(x_{mn} - \theta)[M(x) - \alpha], \tag{2.15}$$

$$e_{mn} = E\left[\int_{-\infty}^{\infty} (y-\alpha)^2 dH\left(y \left| x_{mn}\right.\right) \right], \tag{2.16}$$

then from (2.14), we have

$$(b_{m-1,n} + \mu b_{m,n-1}) - (\frac{1}{\omega} + \frac{\delta}{r}) \, b_{mn} = a_{mn}^2 e_{mn} + 2a_{mn} d_{mn}. \tag{2.17}$$

Note that from (2.5), we have

$$d_{mn} \geq 0, \tag{2.18}$$

and from (2.4), we can get

$$0 \leq e_{mn} \leq [\; c + |\alpha| \;]^2 < \infty, \tag{2.19}$$

combine (2.19) and (2.5), the series of positive terms $\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 e_{ij}$ converge.

From (2.11) and (2.17),

$$2\mu b_{m,n-1} - (\frac{1}{\omega} + \frac{\delta}{r}) b_{mn} = a^2{}_{mn} e_{mn} + 2a_{mn} d_{mn}. \tag{2.20}$$

Therefore,

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(\frac{1}{\omega}+\frac{\delta}{r})b_{ij}-\sum_{i=1}^{m}\sum_{j=1}^{n}2\mu b_{i,j-1}=-\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}e_{ij}-2\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij}. \tag{2.21}$$

After proper decomposition and simplification,

$$\begin{aligned}
&\sum_{i=1}^{m}\sum_{j=1}^{n-1}(\tfrac{1}{\omega}+\tfrac{\delta}{r})b_{ij}+\sum_{i=1}^{m}(\tfrac{1}{\omega}+\tfrac{\delta}{r})b_{in}-\sum_{i=1}^{m}\sum_{j=1}^{n-1}2\mu b_{i,j}-\sum_{i=1}^{m}2\mu b_{i0}\\
&\qquad\qquad=-\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}e_{ij}-2\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij},\\
&\sum_{i=1}^{m}\sum_{j=1}^{n-1}[(\tfrac{1}{\omega}+\tfrac{\delta}{r})-2\mu]b_{ij}+\sum_{i=1}^{m}(\tfrac{1}{\omega}+\tfrac{\delta}{r})b_{in}+\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}e_{ij}\\
&\qquad\qquad+2\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij}=\sum_{i=1}^{m}2\mu b_{i0}.
\end{aligned} \tag{2.22}$$

Make $(\frac{1}{\omega}+\frac{\delta}{r})>2\mu>0$ and $\omega>0$, then $b_{ij}\geq 0$, considering $\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2 e_{ij}$ is convergent and positive, $b_{io}=\omega\, E(x_{i0}-\theta)^2$ is positive constant, we can get

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij}<\sum_{i=1}^{m}\mu b_{i0}<\infty.$$

Therefore, the series of positive terms $\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij}$ converge.

From (2.22),

$$\begin{aligned}
&\sum_{i=1}^{m}\sum_{j=1}^{n-1}[(\tfrac{1}{\omega}+\tfrac{\delta}{r})-2\mu]b_{ij}+\sum_{i=1}^{m-1}(\tfrac{1}{\omega}+\tfrac{\delta}{r})b_{in}+(\tfrac{1}{\omega}+\tfrac{\delta}{r})b_{mn}\\
&\qquad+\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}e_{ij}+2\sum_{i=1}^{m}\sum_{j=1}^{n}a^2{}_{ij}d_{ij}=\sum_{i=1}^{m}2\mu b_{i0}.
\end{aligned}$$

Then,

$$(\frac{1}{\omega}+\frac{\delta}{r})b_{mn}<\sum_{i=1}^{m}2\mu b_{i0}.$$

Therefore,

$$0\leq\lim_{m,n\to\infty}(\frac{1}{\omega}+\frac{\delta}{r})b_{mn}\leq\sum_{i=1}^{\infty}2\mu b_{i0}=b \tag{2.23}$$

exists. Moreover, $b\geq 0$.

Now, let's prove $b\to 0$.

Suppose there is a non-zero constant sequence $\{\ k_{mn}\ \}$ that satisfy

$$d_{mn}\geq k_{mn}b_{mn},\ \sum_{i=1}^{m}\sum_{j=1}^{n}a_{mn}k_{mn}=\infty. \tag{2.24}$$

Using the first part of (2.24) and the convergence of (2.21), we get

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{mn}k_{mn}b_{mn}<\infty. \tag{2.25}$$

From (2.25) and the second part of (2.24), for $\forall\varepsilon > 0$, there is an infinite number of mn that satisfy $b_{mn} < \varepsilon$. Therefore, $\lim\limits_{m,n\to\infty} b_{mn} = b = 0$.

Thus, the following lemma is obtained.

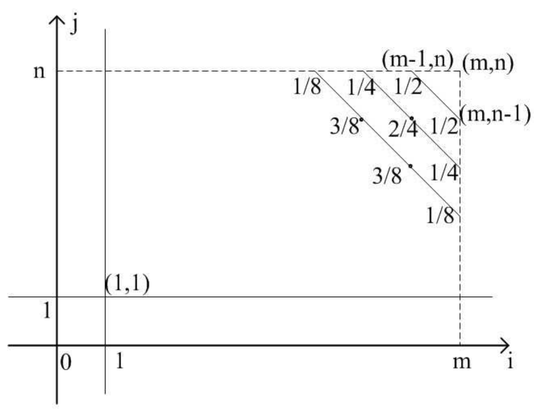
Lemma 2.1. *If there is a non-zero constant sequence* $\{ k_{mn} \}$ *that satisfies* (2.24), *then* $b = 0$.

Make

$$\begin{aligned} A_{mn} = &\left|\sum_{i=1}^{m} \frac{C_{m+n-(i+1)}^{n-1}}{2^{m+n-(i+1)+1}} x_{i0} + \sum_{j=1}^{n} \frac{C_{m+n-(1+j)}^{m-1}}{2^{m+n-(1+j)+1}} x_{0j} - \theta\right| \\ &+\left(\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{C_{m+n-(i+j)}^{m-i}}{2^{m+n-(i+j)}} a_{ij}\right)(|\alpha| + C), \end{aligned} \tag{2.26}$$

from $x_{mn} - \frac{1}{2}(x_{m-1,n} + x_{m,n-1}) = a_{mn}\left[\alpha - \frac{1}{2}(y_{m-1,n} + y_{m,n-1})\right]$, we can get

$$\begin{aligned} &|x_{mn} - \theta| \\ \le &\left|\tfrac{1}{2}(x_{m-1,n} + x_{m,n-1}) - \theta\right| + a_{mn}(|\alpha| + C) \\ \le &\left|\tfrac{1}{4}x_{m-2,n} + \tfrac{2}{4}x_{m-1,n-1} + \tfrac{1}{4}x_{m,n-2} - \theta\right| + (a_{mn} + \tfrac{1}{2}a_{m-1,n} + \tfrac{1}{2}a_{m,n-1})(|\alpha| + C) \\ \le &\left|\tfrac{1}{8}x_{m-3,n} + \tfrac{3}{8}x_{m-2,n-1} + \tfrac{3}{8}x_{m-1,n-2} + \tfrac{1}{8}x_{m,n-3} - \theta\right| \\ &+[a_{mn} + (\tfrac{1}{2}a_{m-1,n} + \tfrac{1}{2}a_{m,n-1}) + (\tfrac{1}{4}a_{m-2,n} + \tfrac{2}{4}a_{m-1,n-1} + \tfrac{1}{4}a_{m,n-2})](|\alpha| + C) \\ \le &\cdots\cdots. \end{aligned}$$



Suppose a point (a, b) on the coordinate axis, where a and b are integers greater than zero, then we can get the coefficient before x_{ab} is $\frac{C_{m+n-(a+b)}^{m-a}}{2^{m+n-(a+b)}}$. At any point $(1, j)$ on $i = 1$, the coordinate is $\frac{C_{m+n-(1+j)}^{m-1}}{2^{m+n-(1+j)}}$, and $x_{1j} - \frac{1}{2}(x_{0,j} + x_{1,j-1}) = a_{1j}\left[\alpha - \frac{1}{2}(y_{0,j} + y_{1,j-1})\right]$. Therefore, the coefficient at the boundary point x_{0j} is 1/2 of x_{1j}. That is, the coefficient of x_{0j} is $\frac{C_{m+n-(1+j)}^{m-1}}{2^{m+n-(1+j)+1}}$. Similarly, the coefficient of x_{i0} is $\frac{C_{m+n-(i+1)}^{n-1}}{2^{m+n-(i+1)+1}}$. Furthermore, the coefficient of a_{ij} is consistent with x_{ij}, which is $\frac{C_{m+n-(i+j)}^{m-i}}{2^{m+n-(i+j)}}$.

After determining the coefficients, we can get

$$
\begin{aligned}
&|x_{mn}-\theta| \\
&\le \left|\tfrac{1}{2}(x_{m-1,n}+x_{m,n-1})-\theta\right| + a_{mn}(|\alpha|+C) \\
&\le \left|\tfrac{1}{4}x_{m-2,n}+\tfrac{2}{4}x_{m-1,n-1}+\tfrac{1}{4}x_{m,n-2}-\theta\right| + (a_{mn}+\tfrac{1}{2}a_{m-1,n}+\tfrac{1}{2}a_{m,n-1})(|\alpha|+C) \\
&\le \left|\tfrac{1}{8}x_{m-3,n}+\tfrac{3}{8}x_{m-2,n-1}+\tfrac{3}{8}x_{m-1,n-2}+\tfrac{1}{8}x_{m,n-3}-\theta\right| \\
&\quad +[a_{mn}+(\tfrac{1}{2}a_{m-1,n}+\tfrac{1}{2}a_{m,n-1})\ +(\tfrac{1}{4}a_{m-2,n}+\tfrac{2}{4}a_{m-1,n-1}+\tfrac{1}{4}a_{m,n-2})](|\alpha|+C) \\
&\le \cdots\cdots \\
&\le \left|\sum_{i=1}^{m}\tfrac{C^{n-1}_{m+n-(i+1)}}{2^{m+n-(i+1)+1}}x_{i0}+\sum_{j=1}^{n}\tfrac{C^{m-1}_{m+n-(1+j)}}{2^{m+n-(1+j)+1}}x_{0j}-\theta\right| + (\sum_{i=1}^{m}\sum_{j=1}^{n}\tfrac{C^{m-i}_{m+n-(i+j)}}{2^{m+n-(i+j)}}a_{ij})(|\alpha|+C).
\end{aligned}
$$

Then, from (2.1), (2.4) and (2.24), we obtain

$$\Pr(|\ x_{mn}-\theta\ |\le A_{mn})=1. \tag{2.27}$$

For $0<|\ x_{mn}-\theta\ |\le A_{mn}$, make

$$\bar{k}_{mn}=\inf\left(\frac{M(x_{mn})-\alpha'}{x_{mn}-\theta}\right). \tag{2.28}$$

Then, from (2.2), $\bar{k}_{mn}\ge 0$.

$P_{mn}(x)$ represents the probability distribution of x_{mn}, so we gain

$$
\begin{aligned}
d_{mn} &= \int\limits_{|\ x_{mn}-\theta\ |\ \le\ A_{mn}} (x_{mn}-\theta\ )\ \ (M(x_{mn})-\alpha)\,dP_{mn}(x) \\
&\ge \int\limits_{|\ x_{mn}-\theta\ |\ \le\ A_{mn}} k_{mn}\,|x_{mn}-\theta\ |\ ^2 d\ p_{mn}(x) \\
&= \bar{k}_{mn}\ b_{mn}.
\end{aligned} \tag{2.29}
$$

Thus, the sequence defined by (2.28) satisfies the first part of (2.24) in order to satisfy the second part of (2.24), set

$$\bar{k}_{mn}\ge\frac{k}{A_{mn}}. \tag{2.30}$$

For sufficiently large m,n and a constant $K>0$, and

$$\sum_{m=2}^{\infty}\sum_{n=2}^{\infty}\frac{a_{mn}}{\sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n-1}a_{ij}}=\infty, \tag{2.31}$$

from (2.31),

$$\sum_{m=2}^{\infty}\sum_{n=2}^{\infty}\ a_{mn}=\infty. \tag{2.32}$$

Therefore, for sufficiently large m,n,

$$2\,(C+|\ \alpha\ |)\ \left(\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}a_{ij}\right)\ge A_{mn}, \tag{2.33}$$

and from (2.30), it means

$$a_{mn}\bar{k}_{mn} \geq a_{mn}\frac{K}{A_{mn}} \geq \frac{a_{mn}\ K}{2(C+|\alpha|)\left(\sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n-1} a_{ij}\right)}. \tag{2.34}$$

From (2.30), (2.31) and the second part of (2.34), we have proved it.

Lemma 2.2. *If* (2.30) *and* (2.31) *hold, then* $b = 0$.

In the assumptions of (2.2) and (2.31), if $a_{mn} = \frac{1}{mn}$ is taken, and due to $\sum\limits_{m=1}^{\infty}\sum\limits_{n=1}^{\infty} a_{mn}^2 = \sum\sum \frac{1}{mn^2} = \left(\sum\limits_{m=1}^{\infty}\frac{1}{m^2}\right)\left(\sum\limits_{n=1}^{\infty}\frac{1}{n^2}\right) = \left(\frac{\pi^2}{6}\right)^2 = \frac{\pi^4}{36}$, then

$$\sum_{m=2}^{\infty}\sum_{n=2}^{\infty}\frac{a_{mn}}{\sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n-1} a_{ij}} = \sum_{m=2}^{\infty}\sum_{n=2}^{\infty}\left(\frac{1}{mn\sum\limits_{i=1}^{m-1}\sum\limits_{j=1}^{n-1}\frac{1}{ij}}\right) = \infty.$$

More generally, for a sequence $\{ a_{mn} \}$, if there are positive constants $c^{'}$ and $c^{''}$ enabling

$$\frac{c^{'}}{mn} \leq a_{mn} \leq \frac{c^{''}}{mn} \tag{2.35}$$

to satisfy (2.2) and (2.21), otherwise (2.35) is satisfied, then such a sequence $\{ a_{mn} \}_{m,\, n=1}^{\infty}$ is called a sequence of type $\frac{1}{mn}$.

If a sequence $\{ a_{mn} \}_{m,\, n=1}^{\infty}$ is a sequence of type $\frac{1}{mn}$, it is easy to obtain that the function $M(x)$ satisfies (2.2) and (2.30). For example, we suppose $M(x)$ meets the following condition of (2.2). That is, for $\delta > 0$,

$$M(x) \leq \alpha - \delta, \quad \textit{when} \quad x < 0, \qquad M(x) \geq \alpha + \delta, \quad \textit{when} \quad x > 0, \tag{2.36}$$

then for $0 < |\ x - \theta\ | \leq A_{mn}$,

$$\frac{M(x) - \alpha}{x - \theta} \geq \frac{\delta}{A_{mn}}. \tag{2.37}$$

Thus,

$$\bar{k}_{mn} \geq \frac{\delta}{A_{mn}}, \tag{2.38}$$

and this is exactly the form when $k = \delta$ in (2.30). Therefore, by using Lemma 2.2, we get the following theorem.

Theorem 2.1. *If* $\{ a_{mn} \}$ *is a sequence of type* $\frac{1}{mn}$, (2.1) *holds and* $M(x)$ *satisfies* (2.26), *then* $b = 0$.

Here, we notice that if take m or n as constants, the results obtained are consistent with those obtained by the $1 - D$ stochastic approximation method. It is also noted that if $M(x)$ meets the following conditions

$$\left.\begin{array}{r} M(x) \text{ is an unsubtracted function,} \\ M(\theta) = \alpha, \\ M'(\theta) > 0\ . \end{array}\right\}, \tag{2.39}$$

then we will prove that (2.30) also holds under the above conditions. In fact, from (2.29),

$$M(x) - \alpha = (x - \theta)\ \left(M'(\theta) + \varepsilon(x - \theta)\ \right), \tag{2.40}$$

where $\varepsilon(t)$ is a function satisfy

$$\lim_{t \to 0} \varepsilon(t) = 0. \tag{2.41}$$

Therefore, there is a constant $\delta \to 0$, which enable

$$\varepsilon(t) \geq -\frac{1}{2} M'(\theta) \tag{2.42}$$

for $|\ t\ | \leq \delta$. Therefore, for $|\ x - \theta\ | < \delta$,

$$\frac{M(x) - \alpha}{x - \theta} \geq \frac{1}{2} M'(\theta) > 0, \tag{2.43}$$

since $M(x)$ is non-subtractive, for $\theta + \delta \leq \theta \leq A_{mn}$, we have

$$\frac{M(x) - \alpha}{x - \theta} \geq \frac{M(\theta + \delta) - \alpha}{A_{mn}} \geq \frac{\delta M'(\theta)}{2A_{mn}}. \tag{2.44}$$

Moreover, for $\theta - A_{mn} \leq x \leq \theta - \delta$, we obtain

$$\frac{M\,(x) - \alpha}{x - \theta} = \frac{\alpha - M\,(x)}{\theta - x} \geq \frac{\alpha - M\,(\theta - \delta)}{A_{mn}} \geq \frac{\delta\ M'\,(\theta)}{2A_{mn}}. \tag{2.45}$$

Thus, without loss of generality, suppose $\frac{\delta}{A_{mn}} \leq 1$, for $0 < |\ x - \theta\ | \leq A_{mn}$, we get

$$\frac{M\,(x) - \alpha}{x - \theta} \geq \frac{\delta\ M'\,(\theta)}{2A_{mn}}. \tag{2.46}$$

Therefore, when $K = \frac{\delta\ M'(\theta)}{2} > 0$, (2.30) holds. Hence, the following results are received.

Theorem 2.2. *If* $\{\ a_{mn}\ \}$ *is a sequence of type* $\frac{1}{mn}$, (2.1) *holds and* $M(x)$ *satisfies* (2.39)*, then* $b = 0$.

Here, we notice that if take m or n as constants, the results obtained are consistent with those obtained by the $1-D$ stochastic approximation method. In addition, weakening the condition (2.1) can not affect the correctness of Theorems 2.1 and 2.2, which is an important problem worth considering. The following condition is a suitable weakening for (2.1). For all x, we have

$$|\ M\,(x)\ | \leq C, \int_{-\infty}^{\infty} (y - M\,(x))^2 dH\,(y\,|x) \leq \sigma^2 < \infty. \tag{2.47}$$

We don't know whether Theorems 2.1 and 2.2 still holds or not if we replace (2.1) with (2.47). Similarly, the condition (2.39) in Theorem 2.2 can also be weakened. That is,

$$\left.\begin{array}{ll} M\,(x) < \alpha, & when \quad x < \theta \\ M\,(x) > \alpha, & when \quad x > \theta \end{array}\right\}. \tag{2.48}$$

3. Estimation applied to information

Take $F(x)$ as an unknown distribution function. Besides,

$$F(\theta) = \alpha,\ (0 < \alpha < 1),\ F'(\theta) > 0, \tag{3.1}$$

and let $\{ z_{mn} \}$ be a double index sequence of function $\Pr(z_{mn} \le x) = F(x)$ independent of the distribution of each random variable x, and by using the sequence $\{ z_{mn} \}$, we estimate θ. However, in some practical applications (such as biology and photosensitive information), it is not allowed to know the values of $\{ z_{mn} \}$. For the value x_{mn} of each pair m, n, we can freely replace it, given $\{ y_{mn} \}$, where

$$y_{mn} = \begin{cases} 1: \ if\ z_{mn} \le x_{mn},\ (\text{"}response\text{"}) \\ 0: \ otherwise\ ,\ (\text{"}nonresponse\text{"}) \end{cases}. \tag{3.2}$$

How to select the value $\{ x_{mn} \}$ and apply sequences $\{ y_{mn} \}$ to estimate the value of θ ? We will proceed as follows.

For all $m, n \in N_0$, we choose the initial value $x_{m0},\ x_{0n}$ as the best guess of estimating θ. Let $\{ a_{mn} \}$ be a constant sequence of type $\frac{1}{mn}$, and take

$$\left. \begin{matrix} x_{00}\ x_{01}\ x_{02} \cdots \\ x_{10}\ x_{11}\ x_{12} \cdots \\ \cdots\ \cdots\ \cdots\ \cdots \end{matrix} \right\}, \tag{3.3}$$

and continue to iterate according to random field sequence

$$x_{mn} - \frac{1}{2}(x_{m-1,n} + x_{m,n-1}) = a_{mn}\left[\alpha - \frac{1}{2}(y_{m-1,n} + y_{m,n-1})\right]. \tag{3.4}$$

(2.1) holds because of

$$\left. \begin{matrix} \Pr\ [y_{mn} = 1\,|x_{mn}] = F(x_{mn}) \\ \Pr\ [y_{mn} = 0\,|x_{mn}] = 1 - F(x_{mn}) \end{matrix} \right\}. \tag{3.5}$$

Moreover,

$$M(x) = F(x). \tag{3.6}$$

Therefore, all of the assumptions of Theorem 2.1 and Theorem 2.2 are satisfied. Thus,

$$\lim_{m,n\to\infty} x_{mn} = \theta. \tag{3.7}$$

Therefore, $\{ x_{mn} \}$ is a suitable estimate for θ. Efficiency of $\{ x_{mn} \}$ depends on the selection of initial values $x_{m0},\ x_{0n}$ and sequences $\{ a_{mn} \}$ for all $m, n \in N_0$. As the best property of $F(x)$, for any given $F(x)$, there are undoubtedly more efficient estimates for θ of type $\{ x_{mn} \}$ defined by (3.4), but $\{ x_{mn} \}$ has the essential advantage of free distribution.

In the application, it is convenient to obtain some observation data sets, which are produced before the generation of the next position in one position, while the

$m \times n$ observation data set is

$$\left.\begin{array}{cccc} y_{mr-r+1,\, ns+s+1}, & y_{mr-r+2,\, ns-s+1} & \cdots & y_{mr,\, ns-s+1} \\ y_{mr-r+1,\, ns-s+2} & y_{mr-r+2,\, ns-s+2} & \cdots & y_{mr,\, ns-s+2} \\ \cdots & \cdots & \cdots & \cdots \\ y_{mr-r+1,\, ns} & y_{mr-r+2,\, ns} & \cdots & y_{mr,\, ns} \end{array}\right\}. \tag{3.8}$$

Apply the mark of (3.2) and set the value of (3.8) as $\bar{y}_{mn}$, and then from

$$x_{mn} - \frac{1}{2}(x_{m-1,n} + x_{m,n-1}) = a_{mn}\left[\alpha - \frac{1}{2}(\bar{y}_{m-1,n} + \bar{y}_{m,n-1})\right], \tag{3.9}$$

we obtain $M(x) = F(x)$, (3.7) still holds.

4. A more general regression problem

An obvious problem is that it is a special case of a more general regression problem in Section 3. In fact, applying the notation from Section 1, consider a random variable Y. For an observed value of x, its joint conditional distribution function is $H(y\,|x)$, and the function $M(x)$ is a regression function Y of x.

According to the hypothesis of regression analysis, $M(x)$ is an unknown parameter form. That is,

$$M(x) = \beta_{00} + (\beta_{10} + \beta_{01})\, x + \beta_{11} x^2, \tag{4.1}$$

and the observed data corresponding to the observed value

$$\left.\begin{array}{ccccc} x_{00} & x_{01} & x_{02} & \cdots & x_{0n} \\ x_{10} & x_{11} & x_{12} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m0} & x_{m1} & x_{m2} & \cdots & x_{mn} \end{array}\right\} \tag{4.2}$$

is

$$\left.\begin{array}{cccc} y_{00} & y_{01} & \cdots & y_{0n} \\ y_{10} & y_{11} & \cdots & y_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{m0} & y_{m1} & \cdots & y_{mn}. \end{array}\right\}.$$

When dealing with β_{ij} of one or two parameters, the least square method is used. For example, to obtain the estimation of β_{ij}, which can be expressed by the least square method as

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_{ij} - \left[\beta_{00} + (\beta_{10} + \beta_{01})\, x + \beta_{11}\,(1+\beta)\, x_{ij}\right]\,\right)^{\,2},$$

given that $M(x)$ is a linear function of x, to estimate the parameter $\beta_{ij}\,(i,\ j\geq 0,\ 1)$ of $M(x)$, we try to estimate the value of θ to enable $M\,(\theta)=\alpha$, where α is given, there are no assumptions about the form of $M(x)$. If we only assume that $H\,(y\,|x\,)$ satisfies the hypothesis of Theorem 2.2, then the sequence $\{\ (1+\beta)\,x_{mn}\ \}$ to estimate values of θ defined by (2.3) will be at least compatible. This indicates that the free distribution sequence values defined by (2.3) and generated by observations are worth studying from the actual situation of the observed regression problem.

References

[1] G. Blatman and B. Sudret, *An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis*, Probabilistic Engineering Mechanics, 2010, 25(2), 183–197.

[2] I. Chueshov and A. Millet, *Stochastic two-dimensional hydrodynamical systems: Wong-Zakai approximation and Support theorem*, Journal Stochastic Analysis and Applications, 2011, 29, 570–611.

[3] S. De, J. Hampton, K. Maute et al., *Topology optimization under uncertainty using a stochastic gradient-based approach*, Structural and Multidiplinary Optimization, 2020, 62, 2255–2278.

[4] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of PDEs with stochastic inputs*, Journal of Computational Physics, 2011, 230(8), 3015–3034.

[5] A. Doostan, H. Owhadi, A. Lashgari and G. Iaccarino, *Non-adapted sparse approximation of PDEs with stochastic inputs*, Technical Report Annual Research Brief, Center for Turbulence Research, Stanford University, Stanford, 2009.

[6] G. Fort, E. Moulines, G. O. Roberts and J. S. Rosenthal, *On the geometric ergodicity of hybrid samplers*, Journal of Applied Probability, 2003, 40(1), 123–146.

[7] P. I. Frazier, S. G. Henderson and R. Waeber, *Probabilistic Bisection Converges Almost as Quickly as Stochastic Approximation*, Mathematics of Operations Research, 2019, 44(2), 651–667.

[8] S. Ghadimi, A. Ruszczyński and M. Wang, *A Single Time-Scale Stochastic Approximation Method for Nested Stochastic Optimization*, SIAM Journal on Optimization, 2020, 30(1), 960–979.

[9] J. Hampton and A. Doostan, *Compressive sampling methods for sparse polynomial chaos expansions*, Handbook of Uncertainty Quantification, 2017, 827–855.

[10] M. Jansen, G. Lombaert and M. Schevenels, *Robust topology optimization of structures with imperfect geometry based on geometric nonlinear analysis*, Computer Methods in Applied Mechanics and Engineering, 2015, 285, 452–467.

[11] R. Johnson and T. Zhang, *Accelerating stochastic gradient descent using predictive variance reduction*, News in physiological sciences, 2013, 1(3), 315–323.

[12] B. Karimi, *Non-Convex Optimization for Latent Data Models: Algorithms, Analysis and Applications*, Ph. D. Thesis, l Université Paris-Saclay, Paris, 2019. DOI: 10.13140/RG.2.2.31417.72803

[13] B. Kriegesmann, *Robust design optimization with design dependent random input variables*, Structural and Multidisciplinary Optimization, 2020, 61, 661–674.

[14] E. Kuhn, C. Matias and T. Rebafka, *Properties of the Stochastic Approximation EM Algorithm with Mini-batch Sampling*, Statistics and Computing, 2020, 30, 1725–1739.

[15] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, New York, 2003.

[16] C. Li, M. Wang, H. Liu and T. Zhang, *Near-optimal stochastic approximation for online principal component estimation*, Mathematical Programming, 2018, 167(1), 75–97.

[17] P. Liang and D. Klein, *Online EM for unsupervised models*, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2009, 611–619.

[18] X. Liang, *Information flow within stochastic dynamical system*, Physical Review E-Statistical, Nonlinear and Soft Matter Physics, 2008, 78, 3, Part 1.

[19] H. Nguyen, F. Forbes and G. Mclachlan, *Mini-batch learning of exponential family finite mixture models*, Statistics and Computing, 2020, 1–18.

[20] R. Pasupathy and S. Kim, *The stochastic root-finding problem: Overview, solutions, and open questions*, ACM Transactions on Modeling and Computer Simulation, 2011, 21(3), 1–23.

[21] B. T. Polyak, *New method of stochastic approximation type*, Automation and Remote Control, 1990, 51(7), 937–946.

[22] B. T. Polyak and A. B. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM Journal on Control and Optimization, 1992, 30(4), 838–855.

[23] G. Qu and A. Wierman, *Finite-Time Analysis of Asynchronous Stochastic Approximation and Q-Learning*, Conference on Learning Theory, Proceedings of Machine Learning Research, 2020, 125, 3185–3205.

[24] H. Robbins and S. Monro, *A stochastic approximation method*, Annals of Mathematical Statistics, 1951, 22, 400–407.

[25] N. L. Roux, M. Schmidt and F. R. Bach, *A stochastic gradient method with an exponential convergence rate for finite training sets*, Advances in neural information processing systems, 2012, 4, 2663–2671.

[26] D. Ruppert, *Efficient estimators from a slowly convergent robbins-monro process*, Technical Report, Cornell University, Ithaca, New York, 1988.

[27] D. Ruppert, *Stochastic approximation*, Handbook of Sequential Analysis, Marcel-Dekker, New York, 1991, 503–529.

[28] S. Song, S. Napel, N. J. Pelc and G. H. Glover, *Phase unwrapping of MR phase images using Poisson equation*, IEEE Transactions on Image Processing, 1995, 4, 667–676.

[29] N. M. S. Choque and Z. D. Kvon, *Stochastic Dynamics of 2D Electrons in Antidot Lattice in the Presence of an in-Plane Magnetic Field*, Brazilian Journal of Physics, 1999, 29.

[30] J. Warrington and D. Ruchti, *Two-stage stochastic approximation for dynamic rebalancing of shared mobility systems*, Transportation Research, 2019, 104(7), 110–134.

[31] P. Yi, J. Lei and Y. Hong, *Distributed resource allocation over random networks based on stochastic approximation*, Systems & Control Letters, 2018, 114, 44–51.

[32] J. Zhao and C. Wang, *Robust topology optimization of structures under loading uncertainty*, American Institute of Aeronautics and Astronautics Journal, 2014, 52(2), 398–407.