VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF MECHANICAL ENGINEERING

MM Science Journal | www.mmscience.eu
ISSN 1803-1269 (Print) | ISSN 1805-0476 (On-line)
Special Issue | TEAM 2024
Transdisciplinary and Emerging Approaches to Multidisciplinary Science
11.9.2024 – 13.9.2024, Ostrava, Czech Republic
DOI: 10.17973/MMSJ.2024_11_2024086

MM SCIENCE JOURNAL

TEAM2024-00026

THE FIXED POINT ITERATIONS METHOD FOR THE CONTRACTION MAPPING

V. Novoselac[1]*, A. Katic[2], Z. Matijasevic[3], M. Culetic Condric[3]
[1]University of Slavonski Brod, Mechanical Engineering Faculty, Slavonski Brod, Croatia
[2]University of Slavonski Brod, Department of Social Sciences and Humanities, Slavonski Brod, Croatia
[3]University of Slavonski Brod, Technical Department, Slavonski Brod, Croatia
*Corresponding author e-mail: Vedran.Novoselac@unisb.hr

Abstract
This paper studies the fixed point iterations method that considers the contraction mapping and one of its applications for the parameter re-estimation of the normal (Gaussian) distribution, where the presence of outliers is considered. In order to study the observed method Banach's fixed point theorem is presented, where it is shown that the contraction property is directly related to the Jacobian of the observed mapping. Furthermore, convergence analysis is conducted in order to estimate the convergence order of the observed model.

Keywords:
fixed point iterations, Banach theorem, convergence analysis, data analysis

1 INTRODUCTION

This paper studies the problem of solving the fixed point equation

$$\boldsymbol{\theta} = F(\boldsymbol{\theta}), \qquad (1)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ presents n dimensional vector of parameters and $F: \Theta \to \Theta$ presents the vector-valued mapping of the parameter space $\Theta \subseteq \mathbb{R}^n$ such that $F(\Theta) \subseteq \Theta$, where $\Theta \subseteq \mathbb{R}^n$ is considered to be some closed, bounded, and convex set. The method to find the solution of (1) is known as the fixed point iteration method [Phillips 1996], [Traub 1982], which idea is to observe a sequence $(\boldsymbol{\theta}_k)$ where each element is defined iteratively as $\boldsymbol{\theta}_{k+1} = F(\boldsymbol{\theta}_k)$. In this situation, if the defined sequence $(\boldsymbol{\theta}_k)$ converge, then its limit $\boldsymbol{\theta}_k \to \boldsymbol{\theta}^*$ present the solution of the equation (1), and thus it holds that $\boldsymbol{\theta}^* = F(\boldsymbol{\theta}^*)$, where the solution $\boldsymbol{\theta}^* \in \Theta$ is called the fixed point of F. So, in order to ensure the convergence of $(\boldsymbol{\theta}_k)$ the sufficient conditions are presented which are presented in well known Banach's fixed point theorem. Moreover, the convergence analysis of $(\boldsymbol{\theta}_k)$ is also conducted by developing the estimation of the convergence order. In the last section, the application of the fixed point iterative method is presented in order to re-estimate the parameters of the normal distribution in a noisy environment, i.e. with the presence of outliers [Novoselac 2019], [Novoselac 2014], [Rousseeuw 2003]. The proposed model is developed in a way that includes conditional expectations that integrate the fine-tuning parameter which regulates a rejection process of noisy data, i.e. outliers. The numerical examples have shown that the determination of the optimal fine-tuning parameter presents the minimization problem which converges linearly to the solution.

2 CONTRACTION MAPPING PRINCIPLE

To ensure the existence and uniqueness of the fixed point, as well as the convergence of the corresponding iteration, the observed mapping F must be contractive.

Definition 1. (The contraction mapping) A mapping $F: \Theta \to \Theta$ is called the contraction if there exists $0 \le L < 1$ such that

$$\|F(\boldsymbol{\theta}) - F(\boldsymbol{\vartheta})\| \le L\|\boldsymbol{\theta} - \boldsymbol{\vartheta}\|$$

for all $\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \Theta$, where $\|\cdot\|$ is some norm.

Geometrically, the contraction means that the distance between maps of $\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \Theta$ under F is strictly less than the distance between $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$. This means that each step in the fixed point iteration process will iteratively contract the distance from the current iterative step, which finally leads to the solution of the fixed point equation. So, these results are presented in the well-known Banach's fixed point theorem for the contraction.

Theorem 1. (The Banach's fixed point theorem) Let $F: \Theta \to \Theta$ be the contraction mapping. Then it holds

1. there exists the unique fixed point $\boldsymbol{\theta}^* \in \Theta$ such that $\boldsymbol{\theta}^* = F(\boldsymbol{\theta}^*)$;
2. for any initial guess $\boldsymbol{\theta}_0 \in \Theta$ the fixed point iterates $\boldsymbol{\theta}_{k+1} = F(\boldsymbol{\theta}_k)$ converge to $\boldsymbol{\theta}^* \in \Theta$, i.e. $\boldsymbol{\theta}_k \to \boldsymbol{\theta}^*$;
3. $(\boldsymbol{\theta}_k)$ satisfies a-priori error estimate
$$\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\| \le \frac{L^k}{1-L}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\|;$$
4. $(\boldsymbol{\theta}_k)$ satisfies a-posteriori error estimate
$$\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\| \le \frac{L}{1-L}\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|.$$

By considering statement 3. from the Banach's fixed point theorem, the estimation of the error $\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\| \le \varepsilon$, where $\varepsilon > 0$ denotes some small predefined value, by considering the number of iteration steps that can be easily conducted. In this situation, it can be concluded that

$$\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\| \le \frac{L^k}{1-L}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\| \le \varepsilon \quad \Rightarrow \quad k \ge \frac{\log \dfrac{\varepsilon(1-L)}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\|}}{\log L}. \tag{2}$$

2.1 Criterion for contraction

The aim of this section is to present the criterion that provides the contraction property for an observed mapping. To achieve this statement generalization of Lagrange's mean value formula which conducts the relation between the Jacobian and contraction properties is observed. So, the induced matrix norm which is generated by the observed vector norm is first presented and described [Phillips 1996].

Definition 2. (Induced matrix norm) If $\|\cdot\|$ is a vector norm in $\mathbb{R}^n$, then we can define the corresponding function $\|\cdot\|$ on $\mathbb{R}^{n\times n}$ for every real squared matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ by
$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{\theta}\|=1}\|\boldsymbol{A}\boldsymbol{\theta}\|.$$
The function $\boldsymbol{A} \to \|\boldsymbol{A}\|$ is called subordinate matrix norm or induced matrix norm.

Lemma 1. An induced matrix norm satisfies the following properties.

1. $\|\boldsymbol{A}\| \ge 0$;
2. $\|\alpha\boldsymbol{A}\| = |\alpha|\cdot\|\boldsymbol{A}\|$, $\alpha \in \mathbb{R}$;
3. $\|\boldsymbol{A}+\boldsymbol{B}\| \le \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$; (triangle inequality)
4. $\|\boldsymbol{A}\boldsymbol{\theta}\| \le \|\boldsymbol{A}\|\cdot\|\boldsymbol{\theta}\|$;
5. $\|\boldsymbol{A}\boldsymbol{B}\| \le \|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|$; (sub-multiplicative property)

Definition 3. (The matrix spectrum) Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be any given squared matrix and $\boldsymbol{I} \in \mathbb{R}^{n\times n}$ identity matrix, then the polynomial
$$P(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A})$$
is called the characteristic polynomial of $\boldsymbol{A}$. The n (not necessarily distinct) roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic polynomial are all the eigenvalues of $\boldsymbol{A}$ and constitute the spectrum of $\boldsymbol{A}$. Let
$$\rho(\boldsymbol{A}) = \max_{1\le i\le n}|\lambda_i|$$
be the largest absolute value of eigenvalues of $\boldsymbol{A}$, called the spectrum of $\boldsymbol{A}$.

Theorem 2. Let $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{n\times n}$ be a real squared matrix, then it follows that some norms of practical interest, namely the induced 1-norm, induced 2-norm, and induced ∞-norm:

1. the induced 1-norm $\|\boldsymbol{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|$;
2. the induced 2-norm $\|\boldsymbol{A}\|_2 = \rho(\boldsymbol{A})$;
3. the induced ∞-norm $\|\boldsymbol{A}\|_\infty = \max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|$;

The next theorem presents the generalization of Lagrange's mean value formula for vector-valued mapping [Hall 1979], [Marjanovic 2017].

Theorem 3. (The Lagrange's theorem) Let $F: \Theta \to \Theta$ be differentiable mapping. Then it holds that
$$\|F(\boldsymbol{\theta}) - F(\boldsymbol{\vartheta})\| \le \max_{0\le t\le 1}\|F'(t\boldsymbol{\theta} + (1-t)\boldsymbol{\vartheta})\|\cdot\|\boldsymbol{\theta} - \boldsymbol{\vartheta}\|,$$
where $F'(\boldsymbol{\theta}) = (\partial F_i/\partial\theta_j) \in \mathbb{R}^{n\times n}$ is the Jacobian matrix.

The next theorem presents a simple criterion for a continuously differentiable function to be the contraction.

Theorem 4. (Criterion for the contraction) Let $F: \Theta \to \Theta$ be differentiable mapping. If
$$\|F'(\boldsymbol{\theta})\| \le L, \text{ for all } \boldsymbol{\theta} \in \Theta$$
where $F'(\boldsymbol{\theta}) = (\partial F_i/\partial\theta_j) \in \mathbb{R}^{n\times n}$ is the Jacobian matrix and $0 \le L < 1$, then F is the contraction on $\Theta \subseteq \mathbb{R}^n$ by considering the given norm $\|\cdot\|$.

Proof. By considering the statement that $\|F'(\boldsymbol{\theta})\| \le L$ for all $\boldsymbol{\theta} \in \Theta$, where $0 \le L < 1$, it may be conclude that for all $\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \Theta$ it holds by Lagrange's theorem that
$$\|F(\boldsymbol{\theta}) - F(\boldsymbol{\vartheta})\| \le \max_{0\le t\le 1}\underbrace{\|F'(t\boldsymbol{\theta} + (1-t)\boldsymbol{\vartheta})\|}_{\le L}\cdot\|\boldsymbol{\theta} - \boldsymbol{\vartheta}\|$$
$$\le L\|\boldsymbol{\theta} - \boldsymbol{\vartheta}\|,$$
what directly shows that F is the contraction on $\Theta \subseteq \mathbb{R}^n$.

Finally, by considering the previous theorem of the simple criterion of the observed differentiable mapping F to be the contraction, and the previously mentioned Banach's fixed point theorem, it can be directly concluded that the sufficient condition for the existence of the fixed point is that the norm of the Jacobian matrix in the observed fixed point must be less than one, i.e. $\|F'(\boldsymbol{\theta}^*)\| < 1$.

3 CONVERGENCE ANALYSIS

One of the main discussions for the iterative processes is the order of convergence, i.e. how fast the iterative method converges to the fixed point. So, the next definition is presented.

Definition 4. (Order of convergence) Let the sequence $(\boldsymbol{\theta}_k)$ converges to $\boldsymbol{\theta}^* \in \Theta$. If there exist $L \geq 0$ and $q \geq 1$ such that

$$\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\| \leq L\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|^q,$$

then it is said that $q \geq 1$ is the order of convergence of the sequence. The limit value $L \geq 0$ is the rate of convergence or the asymptotic constant.

In particular, it is said for the order of convergence that: if $q = 1$ then the order of convergence is linear; if $1 < q < 2$ then is super-linear; if $q = 2$ then is quadratic; if $q = 3$ then is called cubic convergence etc. By considering the fixed point iteration for the contraction mapping when the fixed point is not stationary, i.e. $F'(\boldsymbol{\theta}^*) \neq 0$, it can be concluded that at least linear convergence is ensured. In that case, associated iterates converge linearly to the fixed point $\theta^* \in \Theta$ with the rate of convergence $0 < L \leq \max_{\boldsymbol{\theta}\in\Theta}\|F'(\boldsymbol{\theta})\|$.

Theorem 5. (Higher order of convergence) Let $F: \Theta \to \Theta$ be q –times continuously differentiable and let $\boldsymbol{\theta}^* \in \Theta$ be the fixed point of F such that

$$F^{(i)}(\boldsymbol{\theta}^*) = 0 \quad \text{for all} \quad i = 1, 2, \ldots, q-1.$$

Then it holds that the iterates associated to F converge with qth order to $\boldsymbol{\theta}^*$.

Proof. Let $f(t) = F(t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^*)$, $t \in [0,1]$, for some $\boldsymbol{\theta} \in \Theta$. Then, by considering Taylor's expansion of f on some interval $[a,b]$ to the qth order ([Phillips 1996], [Skala 2021]), there exists $c \in (a,b)$ such that

$$f(b) = \sum_{i=0}^{q-1} \frac{f^{(i)}(a)}{i!}(b-a)^i + \frac{f^{(q)}(c)}{q!}(b-a)^q.$$

So, taking the expansion of f to the qth order on interval $[0,1]$, and taking into account the statement that $F^{(i)}(\boldsymbol{\theta}^*) = f^{(i)}(0) = 0$ for all $i = 1, 2, \ldots, q-1$, it can be written that

$$\begin{aligned}
\|F(\boldsymbol{\theta}) - F(\boldsymbol{\theta}^*)\| &= \|f(1) - f(0)\| \\
&= \left\| \underbrace{\sum_{i=1}^{q-1} \frac{f^{(i)}(0)}{i!}}_{=0} + \frac{f^{(q)}(c)}{q!} \right\| \\
&= \left\| \frac{f^{(q)}(c)}{q!} \right\| \\
&\leq \|f^{(q)}(c)\| \\
&= \|F^{(q)}(t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^q\| \\
&\leq \underbrace{\|F^{(q)}(t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^*)\|}_{\leq L} \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^q \\
&\leq L\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^q,
\end{aligned}$$

where $L \geq 0$. Therefore it holds that

$$\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\| = \|F(\boldsymbol{\theta}_k) - F(\boldsymbol{\theta}^*)\| \leq L\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|^q,$$

for all $k \in \mathbb{N}$, what proves the statement of the theorem.

3.1 Estimating the order of convergence

In this subsection, the estimation of the order of convergence is presented. For that purpose let suppose that a sequence $(\boldsymbol{\theta}_k)$ converges to $\boldsymbol{\theta}^* \in \Theta$ with the qth order, where kth error of an iterative process is denoted as

$$\boldsymbol{e}_k = \boldsymbol{\theta}_k - \boldsymbol{\theta}^*. \tag{3}$$

Because $(\boldsymbol{\theta}_k)$ converges with the qth order, it is clear that if $k \to \infty$ it holds that

$$\|\boldsymbol{e}_{k+1}\| \approx L\|\boldsymbol{e}_k\|^q \quad \text{and} \quad \|\boldsymbol{e}_k\| \approx L\|\boldsymbol{e}_{k-1}\|^q \tag{4}$$

what directly implies that

$$\frac{\|\boldsymbol{e}_{k+1}\|}{\|\boldsymbol{e}_k\|} \approx \left(\frac{\|\boldsymbol{e}_k\|}{\|\boldsymbol{e}_{k-1}\|}\right)^q, \tag{5}$$

and thus

$$q \approx \frac{\log(\|\boldsymbol{e}_{k+1}\| / \|\boldsymbol{e}_k\|)}{\log(\|\boldsymbol{e}_k\| / \|\boldsymbol{e}_{k-1}\|)}. \tag{6}$$

In order to use this formula as an estimator of the order convergence, the fixed point must be omitted from observation because it is not known. For that purpose let's expand the mapping F to the qth order by the Taylor series ([Phillips 1996], [Skala 2021]) around the fixed point $\boldsymbol{\theta}^* \in \Theta$, i.e.

$$F(\boldsymbol{\theta}_k) = \sum_{i=0}^{q-1} \frac{F^{(i)}(\boldsymbol{\theta}^*)}{i!}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^*)^i + \frac{F^{(q)}(c)}{q!}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^*)^q. \tag{7}$$

Noting that $\boldsymbol{\theta}_{k+1} = F(\boldsymbol{\theta}_k)$ and $\boldsymbol{\theta}^* = F(\boldsymbol{\theta}^*)$ we can obtain by (7) that

$$\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^* = \sum_{i=1}^{q-1} \frac{F^{(i)}(\boldsymbol{\theta}^*)}{i!}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^*)^i + \frac{F^{(q)}(c)}{q!}(\boldsymbol{\theta}_k - \boldsymbol{\theta}^*)^q. \tag{8}$$

Now it can be easily seen that if we divide a norm of (8) by $\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|$, i.e. $\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\| / \|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|$, and consider when $k \to \infty$, it can be easily concluded, because $\boldsymbol{\theta}_k \to \boldsymbol{\theta}^*$, that

$$\lim_{k\to\infty} \frac{\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\|}{\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|} \approx \|F'(\boldsymbol{\theta}^*)\|. \tag{9}$$

Likewise, by considering (8), let subtract $\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*$ from $\boldsymbol{\theta}_k - \boldsymbol{\theta}^*$ and obtained $\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k$, and then afterward divide the norm of the subtraction $\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\|$ by $\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|$. Then by considering $\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\| / \|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|$ when $k \to \infty$ it can be concluded that

$$\lim_{k\to\infty} \frac{\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\|}{\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|} \approx \|F'(\boldsymbol{\theta}^*)\|. \tag{10}$$

So, by considering (9) and (10) for suitably large values of $k \to \infty$, it holds that

$$\frac{\|\boldsymbol{e}_{k+1}\|}{\|\boldsymbol{e}_k\|} = \underbrace{\frac{\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\|}{\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\|}}_{\|F'(\boldsymbol{\theta}^*)\|} \approx \underbrace{\frac{\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\|}{\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|}}_{\|F'(\boldsymbol{\theta}^*)\|}. \tag{11}$$

Finally, statement (11) allows to approximate the order of convergence (6) without taking the fixed point into account as

$$q \approx \frac{\log(\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\| / \|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|)}{\log(\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\| / \|\boldsymbol{\theta}_{k-1} - \boldsymbol{\theta}_{k-2}\|)}. \tag{12}$$

4 APPLICATION

The parameter re-estimation problem presents significant topics in a wide area of interest, such as statistics, mathematical modeling, cluster analysis, image processing, etc. [Gonzalez 2008], [Novoselac 2019], [Novoselac 2014], [Rousseeuw 2003]. Thereby, in this section one iterative model that re-estimates the parameters of a normal (Gaussian) distribution $X \sim N(\mu, \sigma^2)$, where μ denotes expectation and σ standard deviation with a corresponding PDF (probability density function) defined as

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{13}$$

is presented. The proposed method is defined by considering conditional expectations

$$\begin{aligned} \mu_{k+1} &= E\left[X \mid I_k\right]; \\ \sigma_{k+1}^2 &= E\left[\left(X - \mu_{k+1}\right)^2 \mid I_k\right], \end{aligned} \tag{14}$$

where $I_k = [\mu_k - h\sigma_k, \mu_k + h\sigma_k]$ denotes an interval that is centered around the corresponding expectation. Interval I_k is constructed in such a way that it is considered the fine-tuning parameter $h > 0$ that regulates a rejection of tails of a low probability [Novoselac 2019], [Rousseeuw 2003]. So, in Fig. 1 it is shown the relation between the fine-tuning parameter $h > 0$ and the standard deviations σ that creates the regions of corresponding probabilities.

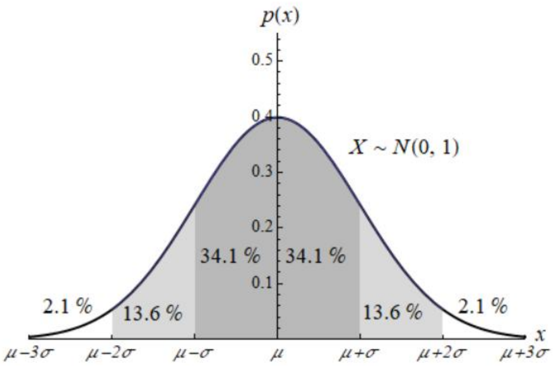

Fig. 1: The probability regions.

Furthermore, by taking into account the fact that each normal distribution $X \sim N(\mu, \sigma^2)$ can be transformed to the standard normal distribution $N(0,1)$, which can be obtained by scaling by σ and shifting by μ, i.e. $X = \mu + \sigma N(0,1)$, the study of the proposed method is conducted only for the standard form when $X \sim N(0,1)$.

4.1 Continuous case random variable

In order to conduct the convergence analysis a continuous random variable case is observed. Thus (14) can be presented as

$$\begin{aligned} \mu_{k+1} &= \frac{\int_{\mu_k - h\sigma_k}^{\mu_k - h\sigma_k} x e^{-\frac{1}{2}x^2} dx}{\int_{\mu_k - h\sigma_k}^{\mu_k - h\sigma_k} e^{-\frac{1}{2}x^2} dx}; \\ \sigma_{k+1} &= \sqrt{\frac{\int_{\mu_k - h\sigma_k}^{\mu_k - h\sigma_k} (x - \mu_{k+1})^2 e^{-\frac{1}{2}x^2} dx}{\int_{\mu_k - h\sigma_k}^{\mu_k - h\sigma_k} e^{-\frac{1}{2}x^2} dx}}. \end{aligned} \tag{15}$$

The Fig. 2 presents the three following iteration steps of the proposed method. The filling under the standard normal PDF in Fig. 2 presents a conditional expectation area that acts on $I_k = [\mu_k - h\sigma_k, \mu_k + h\sigma_k]$ of the current iterative step which is defined by $X \sim N(\mu_k, \sigma_k^2)$. Consequently, the next step is obtained by (15), i.e. (14), which generate $X \sim N(\mu_{k+1}, \sigma_{k+1}^2)$, and the whole process is again iteratively repeated in order to calculate $X \sim N(\mu_{k+2}, \sigma_{k+2}^2)$.

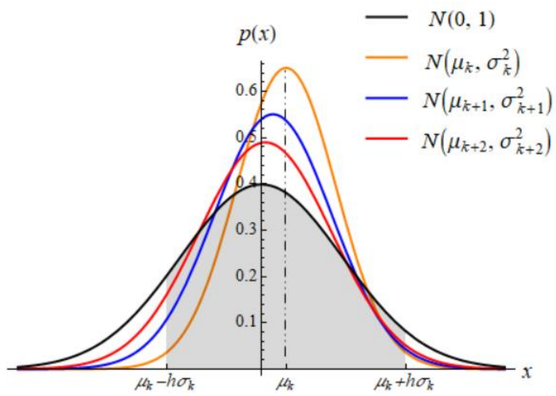

Fig. 2: The proposed method.

The experimental results presented in Fig. 2 have shown that the proposed method converges for the different fine-tuning parameters. The results show that the observed error function, which is defined as the difference between two steps $\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\|$ in sense of 2-norm such that $\boldsymbol{\theta}_k = (\mu_k, \sigma_k)$, tends to zero for each presented $h > 0$. The presented experimental results are conducted with the same starting initialization point $\boldsymbol{\theta}_0 = (\mu_0 = 0.5, \sigma_0 = 0.5)$.

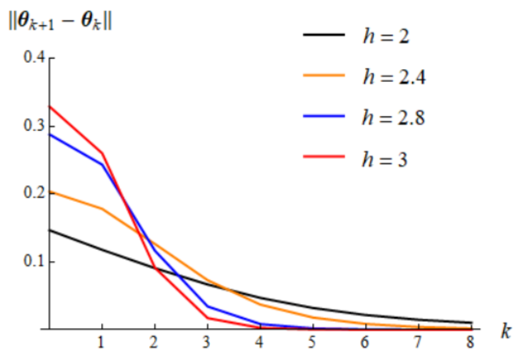

Fig. 3: The error function.

Furthermore, by considering the estimation of the convergence order (12), Fig. 4 presents that the proposed method converges linearly, i.e. $q \approx 1$ as $\boldsymbol{\theta}_k \to \boldsymbol{\theta}^*$.

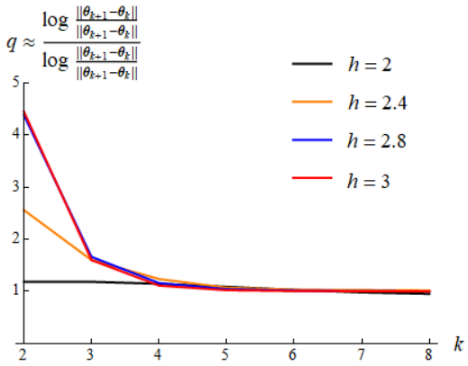

Fig. 4: The order estimator.

In Fig. 4 the final results of the proposed method are presented.

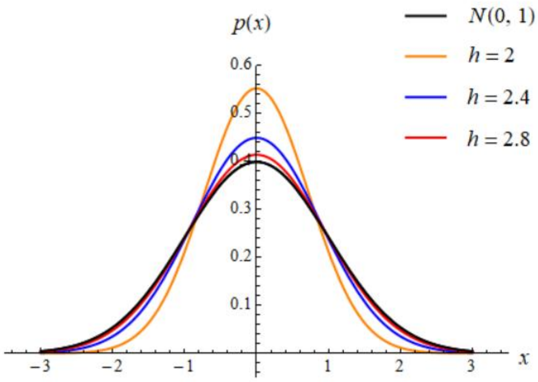

Fig. 5: The re-estimation results.

By considering the final results in Fig. 5, it can be concluded from (15) that when $h \to \infty$, re-estimated parameters tend to the standard normal distribution, i.e. $N(\mu_k, \sigma_k^2) \to N(0,1)$. Situation when $h \to 0$ implies stoppage of the iterative process because of the dividing zero case. The Fig. 6 presents the error function with respect to the fine-tuning parameter which confirms the final results. The error function considers the 2-norm difference between the final result $\boldsymbol{\theta}^* = (\mu^*, \sigma^*)$ which is obtained by the corresponding $h > 0$, and the parameters of the standard normal distribution, which are denoted as $\widetilde{\boldsymbol{\theta}} = (\tilde{\mu} = 0, \tilde{\sigma} = 1)$.

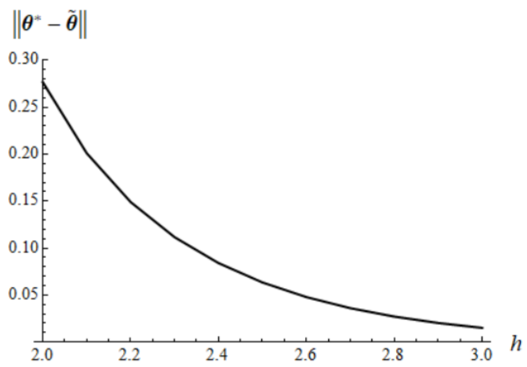

Fig. 6: The fine-tuning parameter.

4.2 Discrete case random variable

Finally, the proposed method is also illustrated on the discrete random variable with the presence of noisy data, i.e. outliers. In this situation $X \sim N(0,1)$ generates data set $S = \{x_i : i = 1, \dots, m\}$ which also contains outliers, i.e. aberrant data that may lead to model misspecification, biased parameter estimation, and incorrect results, what can lead to a suspicion that they are generated by a different mechanism. It is therefore important to identify them prior to modeling and analysis [Rousseeuw 2003]. So, by considering the discrete case, the iterative process (14) can be written as

$$\begin{aligned} \mu_{k+1} &= \mathrm{Mean}\ S_k; \\ \sigma_{k+1} &= \mathrm{StdDev}\ S_k, \end{aligned} \tag{16}$$

where $S_k = \{x_i \in S : x_i \in I_k\}$, $I_k = [\mu_k - h\sigma_k, \mu_k + h\sigma_k]$. In this situation Mean denotes the arithmetic mean of data set S_k, while StdDev denotes the standard deviation. So, for that purpose the Fig. 7 presents the numerical example where the data set S is denoted on the abscissa. The black data points present the data which are generated by the standard normal distribution, while the red data points denotes noisy data, i.e. outliers. Experimental results are conducted for the same initialization point $\boldsymbol{\theta}_0 = (\mu_0 = 0.5, \sigma_0 = 0.5)$, where is shown that the proposed method achieves different results for the different fine-tuning parameters $h > 0$.

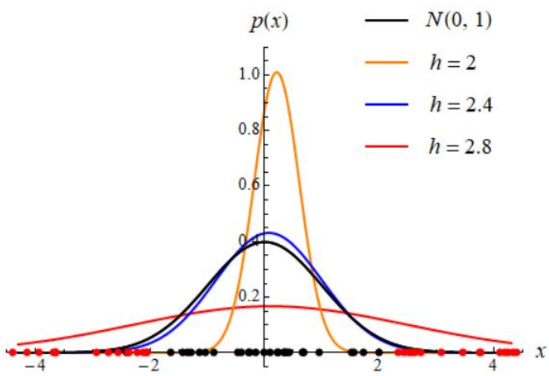

Fig. 7: The discrete noisy case.

In the numerical example which is presented in Fig. 7, the determination of the optimal fine-tuning parameter $h > 0$ presents the optimization problem of the error function $\|\boldsymbol{\theta}^* - \widetilde{\boldsymbol{\theta}}\|$, where $\boldsymbol{\theta}^* = (\mu^*, \sigma^*)$ presents the final result of the proposed method for the discrete case with the corresponding $h > 0$, and the parameters of the standard normal distribution, which are denoted as $\widetilde{\boldsymbol{\theta}} = (\tilde{\mu} = 0, \tilde{\sigma} = 1)$. In Fig. 8 it is shown that the optimal $h > 0$ is attended at $h = 2.4$, which minimizes the observed error function in order to re-estimate the standard normal distribution.

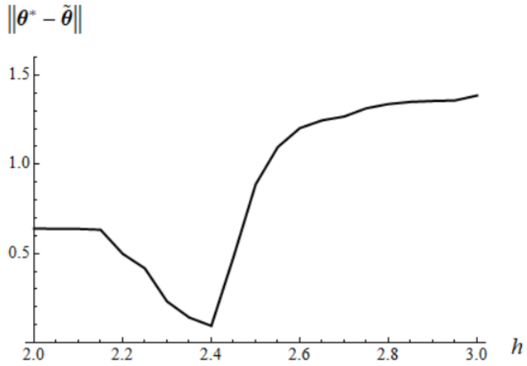

Fig. 8: The fine-tuning parameter.

5 SUMMARY

It is shown in the presented investigation that the fixed point iteration method can be effectively managed and modeled. To study the fixed point iteration model and conduct the convergence analysis, a wide range of applicable numerical analysis is presented and used. Furthermore, experimental research has shown that the proposed iteration model for parameter re-estimation of the normal distribution converges linearly, where different fine-tuning parameters balanced a restriction of tails of low probabilities. For that purpose, the discrete random variable case is observed with the presence of outliers, where it is shown that in this case, a determination of the optimal fine-tuning parameter presents minimization of the observed error function in order to re-estimate the standard normal distribution.

6 ACKNOWLEDGMENTS

We are grateful to dr. sc. Zeljka Rosandic who has linguistically corrected and refined this paper.

7 REFERENCES

[Gonzalez 2008] Gonzalez, R. C. and Woods, R. E., Digital Image Processing, Pearson Prentice Hall, New Jersey, 2008.

[Hall 1979] Hall, W. S. and Newell, M. L., The Mean Value Theorem for Vector Valued Functions: A Simple Proof. Mathematics Magazine, 1979, Vol.52, No.3, pp 157-158.

[Marjanovic 2017] Marjanovic, M. M. and Kadelburg, Z., Lagrange's Formula for Vector-Valued Functions. The Teaching of Mathematics, 2017, Vol.20, No.2, pp 81-88.

[Novoselac 2019] Novoselac, V., Investigation of the optimal number of clusters by the adaptive EM algorithm. Croatian operational research review, 2019, Vol.10, No.1, pp 1-12.

[Novoselac 2014] Novoselac, V. and Pavic, Z., Outlier detection in experimental data using a modified expectation maximization algorithm. In: Adamne Major, A.; Kovacs, L.; Csaba Johanyak, Z.; Pap-Szigeti, R., ed. Proceedings of 6th International Scientific and Expert Conference of the International TEAM Society, Kecskemet: Faculty of Mechanical Engineering and Automation, 2014. pp 112-115.

[Phillips 1996] Philips, G. M. M. and Taylor, P. J., Theory and Applications of Numerical Analysis, Elsevier Science & Technology Books, 1996.

[Rousseeuw 2003] Rousseeuw, P. J. and Leroy, A. M., Robust Regression and Outlier Detection, New York: Wiley, 2003.

[Skala 2021] Skala, V., Efficient Taylor expansion computation of multidimensional vector functions on GPU. Annales Mathematicae et Informaticae, 2021, Vol.54, pp 83-95.

[Traub 1982] Traub, J. F., Iterative methods for the solution of equations, Chelsea Publishing Company, New York, 1982.