Comments on “Generalization of the gradient method with fractional order gradient direction”

Abdul Wahab* † Shujaat Khan‡

September 14, 2020

Abstract

In this paper, a detrimental mathematical mistake is pointed out in the proof of Theorem 1 presented in the paper [Generalization of the gradient method with fractional order gradient direction, J. Franklin Inst., 357 (2020) 2514-2532]. It is highlighted that the way the authors prove the convergence of the fractional extreme points of a real valued function to its integer order extreme points lacks correct and valid mathematical argument. Rest of the theorems contained in the paper are mostly announced without any proof relaying on that of Theorem 1.

Key words. Fractional least mean squares; fractional gradient descent; Fractional calculus; Fractional learning algorithm.

1 Introduction

In [1], a generalized fractional gradient descent scheme is presented and three fractional least means squares algorithms are introduced. The main aim of [1] is to tackle the problem faced by the fractional order gradient methods in converging to real extreme points. Towards this end, the most important part of the study is the mathematical convergence analysis presented in [1, Theorem 1] that concerns Algorithm 1 [1, Eq. (11)]. Then, two derived algorithms (Algorithm 2 [1, Eq. (19)] and Algorithm 3 [1, Eq. (29)]) are presented and corresponding convergence results are furnished in [1, Theorems 2 and 3] (without any proof by relying on the similarity to that of [1, Theorem 1]). However, there are some trivial mathematical errors in the proof of [1, Theorem 1] which are detrimental to the correctness of the entire framework. The main objective of this note is to indicate those mathematical errors.

Remark: The symbols, notations and equation numbers used in this comment are consistent with [1].

*Corresponding Author. E-mail address: abdul.wahab@sns.nust.edu.pk.
†Department of Mathematics, School of Natural Sciences, National University of Sciences and Technology (NUST), Sector H-12, 44000, Islamabad, Pakistan (abdul.wahab@sns.nust.edu.pk).
‡Bio-Imaging, Signal Processing, and Learning Lab., Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, 34141, Daejeon, South Korea (shujaat@kaist.ac.kr).
2 Mathematical Errors

In order to facilitate ensuing discussion, let us recall [1, Eq. (4)] and [1, Eq. (11)]:

\[
C^i D_x^\alpha f(x) = \sum_{i=n}^{+\infty} \frac{(\alpha-n)}{i-n} \frac{f(i)(x)}{\Gamma(i+1-\alpha)} (x-c)^i-\alpha,
\]

\[
x_{k+1} = x_k - \mu \left| C^i D_x^\alpha f(x) \right|_{x=x_k} \quad (\mu > 0, K \in \mathbb{Z}_+, 0 < \alpha < 1),
\]

where

\[
\Gamma(\alpha) = \int_0^{+\infty} e^{-t^{\alpha-1}} dt \quad (\alpha > 0) \quad \text{and} \quad \left( \frac{p}{q} \right) = \frac{\Gamma(p+1)}{\Gamma(q+1)\Gamma(p-q+1)} \quad (p \in \mathbb{R}, q \in \mathbb{N}).
\]

2.1 Main Remark

We show that the proof of [1, Theorem 1] has a detrimental flaw. The statement of [1, Theorem 1] is the following.

**Theorem 1.** When the algorithm in (11) is convergent, it will converge to the real extreme point of \( f(x) \).

The method of contradiction is invoked for proof, however, the contraction is obtained through incorrect mathematical argument. We establish our claim below.

It is assumed that \( x^* \) is the real extreme point of \( f(x) \) and that the sequence \( (x_k) \) converges to a point \( X \neq x^* \). Thus, for \( 0 < \varepsilon < |x^* - X| \) there exists \( N \) such that \( |x_k - X| < \varepsilon < |x^* - X| \) for all \( k > N \). Then, by combining (4) and (11), the following inequality [1, Eq.(12)] is obtained:

\[
\left| x_{k+1} - x_k \right| = \mu \left| C^i D_x^\alpha f(x) \right|_{x=x_k}
\]

\[
\geq \mu \sigma \sum_{i=0}^{\infty} \left| x_k - x_{k-K} \right|^i \left| x_k - x_{k-K} \right|^{1-\alpha}
\]

where

\[
\sigma := \sup_{k>N,i\in\mathbb{N}} \left( \frac{\alpha-1}{i} \right) \frac{f(i+1)(x_k)}{\Gamma(i+2-\alpha)}, \quad \text{and} \quad d := \frac{\mu \sigma}{1-\varepsilon}.
\]

The equations (12a), (12b), and (12c) are derived from (11), (4), and (12b) (by changing the dummy index \( i \)), respectively. The inequality (12d) is derived from (12c) by distributing the absolute value over individual terms in the infinite series and then by replacing \( \left( \frac{\alpha-1}{i} \right) \frac{f(i+1)(x_k)}{\Gamma(i+2-\alpha)} \)
by $\sigma$. The argument there to arrive at (12d) is incorrect and the inequality is in the other sense. Indeed, it is trivial to note that by the triangular inequality and Eq. (12c),

$$
\mu \sum_{i=0}^{+\infty} \binom{\alpha - 1}{i} \frac{f^{(i+1)}(x_k)}{\Gamma(i + 2 - \alpha)} |x_k - x_{k-K}|^{i+1-\alpha} \\
\leq \mu \sum_{i=0}^{+\infty} \binom{\alpha - 1}{i} \frac{f^{(i+1)}(x_k)}{\Gamma(i + 2 - \alpha)} |x_k - x_{k-K}|^{i+1-\alpha} \\
\leq \mu \sup_{k>N, i \in \mathbb{N}} \left| \binom{\alpha - 1}{i} \frac{f^{(i+1)}(x_k)}{\Gamma(i + 2 - \alpha)} \sum_{j=0}^{+\infty} |x_k - x_{k-K}|^j \right| |x_k - x_{k-K}|^{1-\alpha}.
$$

Note also that an absolute is missing in the definition of $\sigma$, otherwise it may be negative despite being supremum for a general function $f$, for example, $f(x) = -1/(1 - x)$ with $x \in (0, 1)$.

In addition to that (12e) is derived from (12d) by using the sum of infinite geometric series whose convergence is ensured by $|x_k - x_{k-K}| < 1$. This tacit assumption is not justified as well for all $k, K \in \mathbb{N}$ even if the series is convergent.

Finally, inequality (12f) is derived from (12e) with an underlying assumption that $|x_k - x_{k-K}| < \varepsilon$ that also may be true for specific values of $k$ and $K$ but not in general for all $k, K \in \mathbb{N}$ even when the series is convergent.

In a nutshell, the contradiction derived in [1, Eqs (13)-(14)] using inequality [1, Eq. (12)] is mathematically incorrect.

### 2.2 Minor Remarks

We have following minor comments

1. The result announced in [1, Theorem 2] concerning [1, Algorithm 2] is not proved. It is mentioned that "This theorem can be proved in the similar method like Theorem 1". As we have indicated that the proof of [1, Theorem 1] is flawed, it is unclear whether [1, Theorem 2] is valid or not.

2. A similar remark is valid for [1, Theorem 3], which is the counterpart of [1, Theorem 1] corresponding to [1, Algorithm 3].

3. In Eq. (29) of [1], given by

$$
x_{k+1} = x_k - \mu \sum_{i=1}^{+\infty} \binom{\alpha(x) - 1}{i-1} \frac{f^{(i)}(x_k)}{\Gamma(i + 1 - \alpha(x))} (x_k - c)^{i-\alpha(x)},
$$

it is unclear at which $x$ the exponent $\alpha(x)$ will be evaluated. Moreover, based on the design of $\alpha(x)$ (as discussed in [1, Fig. 2]) $\alpha(x) - 1 \leq 0$ for all values of $x$ and, thus

$$
\binom{\alpha(x) - 1}{i-1} = \frac{\Gamma(\alpha(x))}{\Gamma(i)\Gamma(\alpha(x) - i + 1)},
$$

is undefined for all $i \geq 2$. Indeed, for $0 \leq \alpha(x) \leq 1$ and $i \geq 2$,

$$
\alpha(x) - (i-1) \leq \alpha(x) - 1 \leq 0.
$$

but the Gamma function $\Gamma(x)$ is defined only for $x \in \mathbb{R}$ such that $x > 0$. 

3
Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Y. Wei, Y. Kang, W. Yin, and Y. Wang, Generalization of the gradient method with fractional order gradient direction, *J. Franklin. Inst.*, 357(4): (2020), pp. 2514-2532.