Complex-step derivative approximation in noisy environment

Filip Nikolovski [*] Irena Stojkovska[†]

May 26, 2017

Abstract

The complex-step derivative approximation is a powerful method for derivative approximations which has been successfully implemented in deterministic numerical algorithms. We explore and analyze its implementation in noisy environment through examples, error analysis and application to optimization methods. Numerical results show a promising performance of the complex-step gradient approximation in noisy environment.
Key words. derivative approximation, complex-step derivative approximation, nonmonotone line-search methods, noisy environment.
AMS subject classification. 65D25, 30E10, 90C56

1 Introduction

Approximations of derivatives of functions are widely used in many areas such as chemical, biomedical and mechanical engineering, physics and finance, in solving differential equations or optimization. It might happen that there exists an underlying function which should be differentiated, but only its values at a sampled data set are known, without knowing the function itself;

[*]NOVA International Schools, Skopje, Macedonia, e-mail: filipnikolovski@gmail.com
[†]Department of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Arhimedova 3, 1000 Skopje, Macedonia, e-mail: irenatra@pmf.ukim.mk

or the exact formulas of derivatives are available, but the exact computation of the derivatives might require a lot of function evaluations. So, in those and all similar cases, approximations of derivatives are recommended.

The most basic methods for approximating the first derivative of a function are the various finite difference methods. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. Depending on the required precision and computational cost, the most used finite difference approximations of the first derivative of f at a point x are the *forward* and the *centered finite difference approximation* given by:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \tag{1}$$

and

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h} \tag{2}$$

respectively, where h is a small positive real step. It can be easily verified, by using the Taylor series expansion, that the order of the approximation error of the forward finite difference approximation (1) is $\mathcal{O}(h)$, while the order of the approximation error of centered finite difference approximation (2) is $\mathcal{O}(h^2)$, [1]. Here $\mathcal{O}(\cdot)$ is the order notation defined for any nonnegative sequences of scalars $\{a_k\}$ and $\{b_k\}$, for which we write $a_k = \mathcal{O}(b_k)$, if there is a constant $C > 0$ such that $|a_k| \leq C|b_k|$, for all k sufficiently large, [21].

From the perspective of computational cost, the forward finite difference requires one additional function evaluation to make the approximation (assuming that the value of $f(x)$ is already available), while the centered finite difference requires two additional function evaluations to make the same approximation. While using the finite difference, the "step-size dilemma" is also present, that is using a small step size to minimize the truncation error versus avoiding to use very small values that will lead to the subtractive cancellation error, [17].

Another approach to derivative approximation uses complex variables, and was firstly proposed by Lyness and Moler in [10]. It was later used by Squire and Trap in [23] for obtaining a very simple expression for estimating the first derivative. This procedure uses an imaginary step to approximate the first derivative and avoids subtractive error cancellation. The *complex-step derivative approximation* in [23] is obtained as follows.

Let f be an analytic function of a complex variable z, and also assume that f is real on the real axis. Then f has Taylor series expansion about

$x \in \mathbb{R}$ given by:

$$f(x+ih) = f(x) + ihf'(x) - h^2\frac{f''(x)}{2!} - ih^3\frac{f'''(x)}{3!} + \dots,$$

where h is a small positive real step and i is the imaginary unit ($i^2 = -1$). Taking imaginary parts on both sides, we obtain:

$$\mathrm{Im}\big(f(x+ih)\big) = hf'(x) - h^3\frac{f'''(x)}{3!} + \dots, \tag{3}$$

from where we can finally write:

$$f'(x) = \frac{\mathrm{Im}\big(f(x+ih)\big)}{h} + h^2\frac{f'''(x)}{3!} + \dots = \frac{\mathrm{Im}\big(f(x+ih)\big)}{h} + \mathcal{O}(h^2).$$

Thus, the first derivative of f at x can be approximated by the following expression:

$$f'(x) \approx \frac{\mathrm{Im}\big(f(x+ih)\big)}{h} \tag{4}$$

with error of order $\mathcal{O}(h^2)$. The error of this approximation is of same order as the error of the central finite difference approximation (2), but in the approximation given by (4) there is no possibility of subtractive cancellation error. The step h can be chosen arbitrary small, the approximation is robust and achieves great accuracy for any h below 10^{-8}, although in double precision values below 10^{-308} results in underflow, [17].

Very often when physical system measurements or computer simulations are used for approximations, a noise is present. The noise can also arise from some uncertainty in the system under consideration. In that case, it is common to note that in the absence of the true (noiseless) functional values $f(x)$, we are forced to use noisy measurements defined by

$$F(x) = f(x) + \xi(x), \tag{5}$$

where $\xi(x)$ is a general state-dependent deterministic or stochastic noise. See, for example, [12, 13, 14, 22].

If we have to approximate the first derivative of the function f at a point x using the finite difference approximation in noisy environment, the approximation will be

$$f'(x) \approx \frac{F(x+h) - F(x)}{h} \tag{6}$$

and

$$f'(x) \approx \frac{F(x+h) - F(x-h)}{2h}, \tag{7}$$

for the *forward* and the *centered finite difference approximation in presence of noise*, respectively, where $F(x)$ is defined by (5) and h is a positive real step, see [19].

Similarly, noise can occur when complex-valued functions are used for calculations. In that case, we have a complex noise $\zeta(z) = \xi_1(z) + i\xi_2(z)$, where $\xi_1(z)$ and $\xi_2(z)$ are real valued deterministic or stochastic functions. So, instead of a true function value $f(z)$, we will use a noisy functional value defined by

$$F(z) = f(z) + \zeta(z) = f(z) + \xi_1(z) + i\xi_2(z). \tag{8}$$

Complex-valued functions and various types of complex noise can be found in certain problems in statistics, signal processing, digital communications and electrical engineering, see for example [3, 8].

In this work we define, explore and analyze the complex-step derivative approximation in presence of noise and implement it to the nonmonotone line-search optimization algorithms.

Our paper is organized as follows. In Section 2, some known results for finite difference approximations in absence and presence of noise are given. The complex-step derivative approximation in presence of noise is defined and explored in Section 3. Section 4 contains the results of numerical experiments with the new complex-step gradient approximation in noisy environment used in the nonmonotone line-search optimization algorithms, while the conclusions and the final remarks are given in Section 5.

2 Preliminaries: Finite difference approximations

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real valued function of vector argument. Finite difference approximations (1) and (2) can be modified to approximate the gradient ∇f of the function f at a point x:

$$\left[\nabla f(x)\right]_j \approx \frac{f(x+he_j) - f(x)}{h}, \ j = 1, 2, ..., n \tag{9}$$

and

$$\left[\nabla f(x)\right]_j \approx \frac{f(x+he_j) - f(x-he_j)}{2h}, \ j = 1, 2, ..., n \tag{10}$$

respectively, where h is a small positive real step and e_j is the j-th coordinate vector, [21].

Note that, the overall computational cost for approximating the gradient using the forward finite difference approximation is $n+1$, while using the centered difference approximation is $2n$. This noticeable increase in function evaluations results in decrease of error. Namely, the forward finite difference approximation error is of first-order i.e. $\mathcal{O}(h)$, while the centered finite difference approximation error is of second-order i.e. $\mathcal{O}(h^2)$, [21]. The centered finite difference approximation, if implemented carefully, has satisfactory performance, however given the nature of the calculations required, finite difference approximations are susceptible to subtractive cancellation errors in the numerator. For very small steps h, the subtraction in the numerator gives zero which results with an incorrect approximation to the gradient components. So, a value of $h = \sqrt{\text{mach. eps}} \approx 10^{-8}$ is recommended, [21].

When noise is present, and true functional values $f(x)$ are not known, but only their noisy measurements $F(x)$ given by (5), forward and centered finite difference gradient approximations in noisy environment are given by

$$[\hat{g}^{FFD}(x)]_j = \frac{F(x+he_j) - F(x)}{h}, \ j = 1, 2, ..., n \tag{11}$$

and

$$[\hat{g}^{CFD}(x)]_j = \frac{F(x+he_j) - F(x-he_j)}{2h}, \ j = 1, 2, ..., n \tag{12}$$

respectively, where h is a positive real step and e_j is the j-th coordinate vector, [13, 14].

The choice of h is crucial for the accuracy of the approximation with finite difference in presence of noise. It has to be related to the noise level, see [19] for the approximation of the first derivative with forward finite difference and see [14] for the approximation of the gradient with centered finite difference. Actually, it has been empirically established that the value $h = 3\sigma$, where σ is the noise level (standard deviation of the simulated noise), suits the best when gradient is approximated with centered finite difference and it is used in derivative-free nonmonotone line-search optimization methods, [14]. The importance of h not being small in noisy gradient approximations used in line-search optimization methods is also pointed out in [9], where it is hinted that small steps h can lead to failure of the line-search procedure.

Finite difference gradient approximations have been theoretically analyzed and successfully implemented in various optimization frameworks, both in absence of noise (e.g. [5, 20]), or when noise is present (e.g. [9, 13, 14, 19]).

3 Complex-step approximations

In this section we extend the complex-step approximation of the first derivative of f established in [23] and given by formula (4), to complex-step approximation of the first derivative of f in presence of noise. When noise is present, instead of true function values $f(z)$, we use the noisy function measurements $F(z)$, defined by (8). So, we define the *complex-step first derivative approximation in presence of noise* of an analytic function f at $x \in \mathbb{R}$ by:

$$f'(x) \approx \frac{\operatorname{Im}\big(F(x+ih)\big)}{h}, \tag{13}$$

where h is a positive real step.

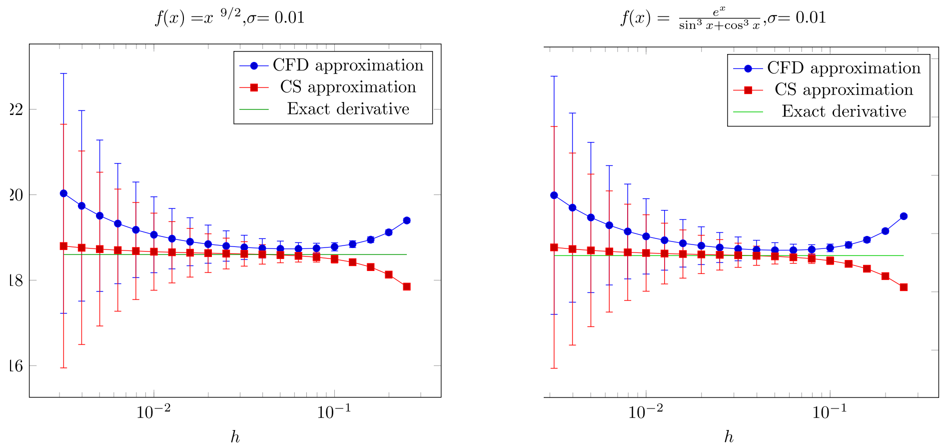
In [23], a comparison has been made between two approximation formulas for the first derivative: the centered finite difference approximation (2) and the complex-step derivative approximation (4), in absence of noise. Results show that the approximation (4) does not suffer from the subtractive cancelation error as h decreases, which is not the case with the approximation (2). We want to find out how the complex-step approximation acts in presence of noise. On the next two examples, the same ones used in [23], we are going to explore the sensitivity of the complex-step gradient approximation (13) on the step h and compare it to the centered finite difference approximation (7), in noisy environment.

Example 3.1 ([23]) *Let* $f(x) = x^{9/2}$*, then the true value of the first derivative of* f *at* $x_0 = 1.5$ *is* $f'(1.5) \approx 18.60081$.

Example 3.2 ([11]) *Let* $f(x) = e^x/(\sin^3 x + \cos^3 x)$*, then the true value of the first derivative of* f *at* $x_0 = 1.5$ *is* $f'(1.5) \approx 3.62203$.

For the purpose of obtaining noisy function evaluations $F(x)$, used in the centered finite difference approximation (7), simulated *white Gaussian noise* $\xi \sim \mathcal{N}(0, \sigma^2)$ is used, and the noisy values $F(x)$ are obtained as $F(x) = f(x) \cdot (1 + \xi)$. For noisy function evaluations $F(z)$, used in the complex-step approximation (13), simulated *white circular noise* of the form $\zeta = \xi_1 + i \cdot \xi_2$, where ξ_1 and ξ_2 are independent random variables with $\xi_1, \xi_2 \sim \mathcal{N}(0, \frac{\sigma^2}{2})$, is used. The noisy values $F(z)$ are obtained as $F(z) = f(z)\cdot(1+\xi_1+i\cdot\xi_2)$. Both noises ξ and ζ are comparable in a sense that the both have zero expectation, they are normally distributed on real parts and both have the noise level equal to σ. Testing has been performed at three noise levels $\sigma = 0.01, 0.1, 0.5$ and different values for the step h.

Each derivative has been evaluated 100 times for each tested value of the step h and 95% confidence intervals for the values of the derivative are presented on Figures 1-3 for the noise levels $\sigma = 0.01, 0.1, 0.5$ respectively.

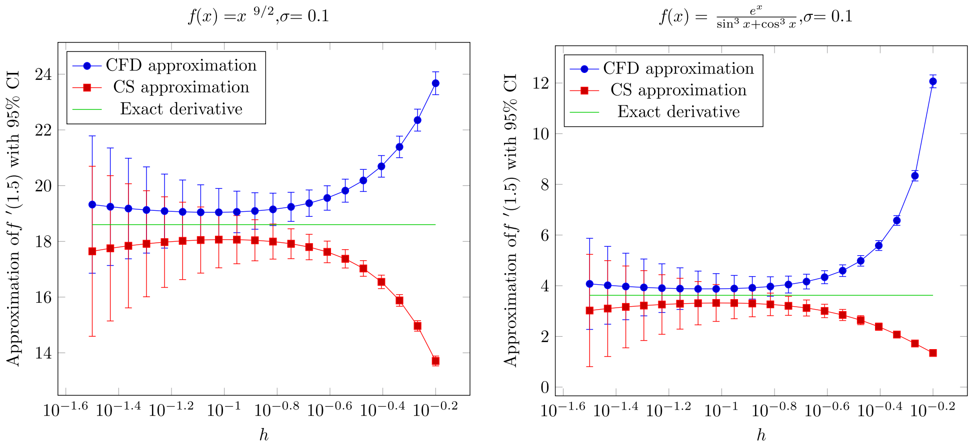


(a) Example 3.1. (b) Example 3.2.

Figure 1: 95% confidence intervals for the first derivative approximated by the centered finite difference (CFD) approximation (7) and by the complex-step (CS) approximation (13), at noise level $\sigma = 0.01$.

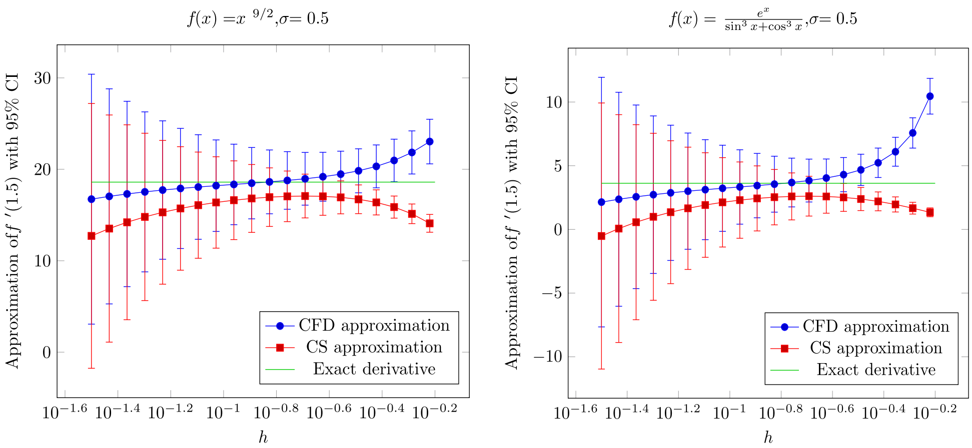
As it can be seen from the plots at Figures 1-3, derivative approximations with small values for the step h have bigger confidence intervals regardless the noise level. The approximations with higher values for the step h may have smaller confidence intervals, but are less accurate. It seems that for small values of the step h, complex-step approximations are more accurate than centered finite difference approximations when the noise level is low, but with the increase of the noise level the opposite happens. For bigger values of the step h we have different situation, both approximations have similar accuracy, but the complex-step approximation has smaller confidence intervals.

In Section 2, it was mentioned that the choice of h is crucial for the accuracy of the approximation with the finite difference formula in presence of



(a) Example 3.1. (b) Example 3.2.

Figure 2: 95% confidence intervals for the first derivative approximated by the centered finite difference (CFD) approximation (7) and by the complex-step (CS) approximation (13), at noise level $\sigma = 0.1$.



(a) Example 3.1. (b) Example 3.2.

Figure 3: 95% confidence intervals for the first derivative approximated by the centered finite difference (CFD) approximation (7) and by the complex-step (CS) approximation (13), at noise level $\sigma = 0.5$.

noise and that h has to be related to the noise level, [19]. Similar conclusions can be made for the choice of h when the first derivative is approximated by the complex-step derivative approximation formula in presence of noise

(13). For that purpose we define *the least squares error of the complex-step approximation* by

$$\mathcal{SE}(h) = \left(\frac{1}{h}\mathrm{Im}(F(x+ih)) - f'(x)\right)^2, \tag{14}$$

and we seek a step $h > 0$ that minimizes the expected value $E[\mathcal{SE}(h)]$, the same approach as in [19] for finite difference approximations in presence of noise.

We also assume that the noise $\xi(x)$ defined by (5) is a random variable with expectation $E[\xi(x)] = 0$ and variance $\mathrm{Var}[\xi(x)] = \sigma^2$, and that the real and the imaginary part of the complex noise $\zeta(z)$ defined by (8) i.e. $\xi_1(z)$ and $\xi_2(z)$ respectively, are independent and identically distributed random variables with $E[\xi_1(z)] = E[\xi_2(z)] = 0$ and $\mathrm{Var}[\xi_1(z)] = \mathrm{Var}[\xi_2(z)] = \sigma^2/2$. So, both noises $\xi(x)$ and $\zeta(z)$ have the same noise level $\sigma > 0$.

The approximation problem is formulated in terms of the derivative of the expected value $E[F(x)]$, since the assumptions about the noise imply that the derivatives of $f(x)$ and $E[F(x)]$ agree. Choosing a step h that minimizes the expected value $E[\mathcal{SE}(h)]$ yields an optimal approximation to the derivative of the expected value $E[F(x)]$, hence it yields an optimal approximation to the first derivative of $f(x)$, [19]. As we will see, the numerical results of complex-step approximations also show a good performance of the step h that is chosen in this way.

Before analyzing errors of complex-step derivative approximation in presence of noise, let us rewrite the expansion (3) as

$$\mathrm{Im}\big(f(x+ih)\big) = hf'(x) - h^3\frac{f'''(u)}{3!}, \tag{15}$$

for some $u \in (x, x+h)$. Then, using (8) and (15), we can derive the following expression for the difference between the complex-step approximation and the true value of the first derivative, in presence of noise:

$$\begin{aligned}
&\frac{1}{h}\mathrm{Im}(F(x+ih)) - f'(x) \\
&= \frac{1}{h}\mathrm{Im}\Big(f(x+ih) + \xi_1(x+ih) + i\xi_2(x+ih)\Big) - f'(x) \\
&= \frac{1}{h}\mathrm{Im}\Big(f(x+ih)\Big) + \frac{1}{h}\xi_2(x+ih) - f'(x) \\
&= \frac{1}{h}\left(hf'(x) - \frac{h^3}{6}f'''(u)\right) + \frac{1}{h}\xi_2(x+ih) - f'(x) \\
&= f'(x) - \frac{h^2}{6}f'''(u) + \frac{1}{h}\xi_2(x+ih) - f'(x) \\
&= -\frac{h^2}{6}f'''(u) + \frac{1}{h}\xi_2(x+ih), \qquad (16)
\end{aligned}$$

fore some $u \in (x, x+h)$. We will need the last equation (16) for analyzing the errors of complex-step derivative approximation in presence of noise. The following lemma gives bounds of the expected value $E[\mathcal{SE}(h)]$.

Lemma 3.1 *Assume that $f : \mathbb{R} \to \mathbb{R}$ is an analytic function, and m and M are minimum and maximum of $|f'''|$ on $(x, x+h)$, respectively. Then,*

$$\frac{\sigma^2}{2h^2} + \frac{h^4}{36}m^2 \le E[\mathcal{SE}(h)] \le \frac{\sigma^2}{2h^2} + \frac{h^4}{36}M^2. \qquad (17)$$

Proof. Using (16) and the assumptions about the noise we have

$$\begin{aligned}
E[\mathcal{SE}(h)] &= E\left[\frac{1}{h}\mathrm{Im}(F(x+ih)) - f'(x)\right]^2 \\
&= E\left[-\frac{h^2}{6}f'''(u) + \frac{1}{h}\xi_2(x+ih)\right]^2 \\
&= \mathrm{Var}\left[\frac{1}{h}\xi_2(x+ih)\right] + \left(-\frac{h^2}{6}f'''(u)\right)^2 \\
&= \frac{1}{h^2}\mathrm{Var}\left[\xi_2(x+ih)\right] + \frac{h^4}{36}\left(f'''(u)\right)^2 \\
&= \frac{\sigma^2}{2h^2} + \frac{h^4}{36}\left(f'''(u)\right)^2, \qquad (18)
\end{aligned}$$

since for a random variable X such that $E[X] = 0$ we have $E[X + \alpha]^2 = \mathrm{Var}[X]+\alpha^2$. Then, directly from (18), for m and M, minimum and maximum of $|f'''|$ on $(x, x + h)$, respectively, we get the bounds (17). ■

Let us note that the resulting bounds on $E[\mathcal{SE}(h)]$ in Lemma 3.1 coincide with the bounds on the expected least squares error of the centered difference approximation in presence of noise (see Lemma 4.1 in [19]), under the assumptions about the real and the complex noises made in this Section, with same noise level $\sigma > 0$. So, a similar discussion (as in [19]) follows. Namely, if we denote by $\phi(h, \mu) = \frac{\sigma^2}{2h^2} + \frac{h^4}{36}\mu^2$, then ϕ is uniformly convex and

$$\min_h \phi(h, \mu) = \frac{3^{1/3}}{4}\mu^{2/3}\sigma^{4/3}.$$

Then, the global minimizer of $\phi(h, \mu)$ for $\mu = M$ is

$$h_M = \left(\frac{3\sigma}{M}\right)^{1/3},$$

which plays an important role in the behavior of $E[\mathcal{SE}(h)]$, as the following theorem shows.

Theorem 3.1 *Assume that $f : \mathbb{R} \to \mathbb{R}$ is an analytic function, and m and M are minimum and maximum of $|f'''|$ on $(x, x + h)$, respectively. Then,*

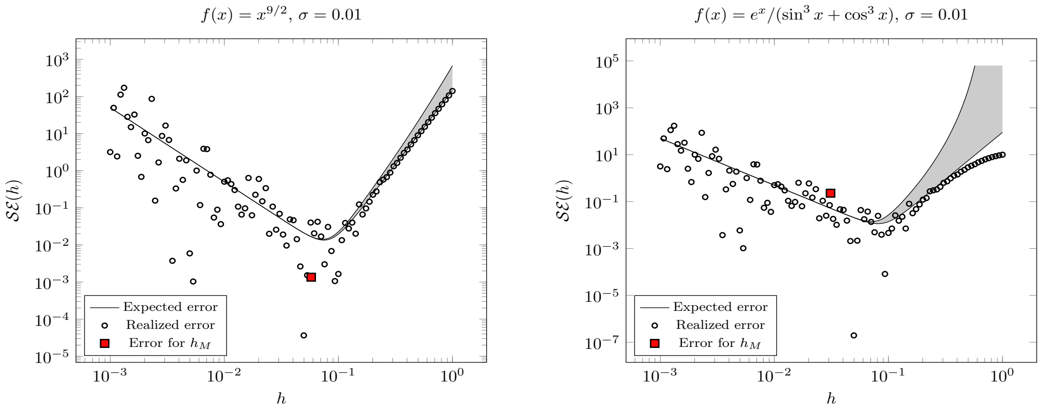
$$\frac{3^{1/3}}{4}m^{2/3}\sigma^{4/3} \le \min_h E[\mathcal{SE}(h)] \le \frac{3^{1/3}}{4}M^{2/3}\sigma^{4/3}. \tag{19}$$

Proof. See the proof of Theorem 4.2 in [19]. ■

Theorem 3.1 shows that with a complex-step derivative approximation in presence of noise we can expect an error of order $\sigma^{4/3}$. It can be also concluded that h_M is nearly optimal in the sense that $E[\mathcal{SE}(h_M)]$ satisfies the bounds in Theorem 3.1 i.e.

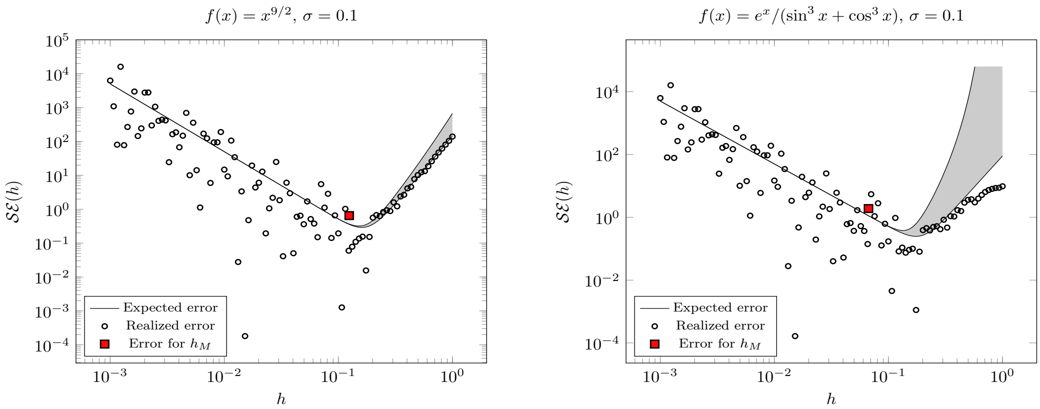
$$\frac{3^{1/3}}{4}m^{2/3}\sigma^{4/3} \le E[\mathcal{SE}(h_M)] \le \frac{3^{1/3}}{4}M^{2/3}\sigma^{4/3}.$$

Before discussing the derived theoretical results, in terms of their similarity to the centered finite difference approximation results in [19], we are going to illustrate the bounds for the expected least squares error for complex-step approximation, on Example 3.1 and Example 3.2.



(a) Example 3.1. (b) Example 3.2.

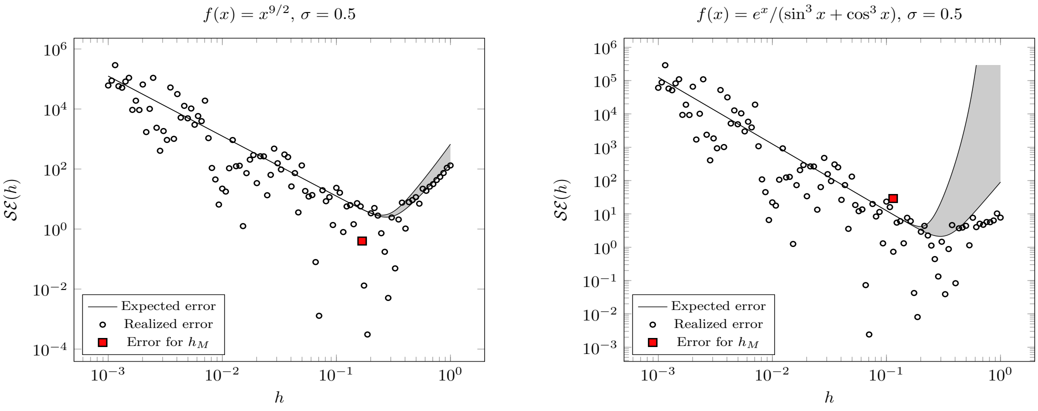
Figure 4: Realized least squares error $\mathcal{SE}(h)$ for the complex-step approximation, along with the expected error, and error for h_M, at noise level $\sigma = 0.01$.



(a) Example 3.1. (b) Example 3.2.

Figure 5: Realized least squares error $\mathcal{SE}(h)$ for the complex-step approximation, along with the expected error, and error for h_M, at noise level $\sigma = 0.1$.

Same as earlier in this Section, we use simulated white circular noise for noisy function evaluations $F(z)$. Three different noise levels are tested $\sigma = 0.01, 0.1, 0.5$. The log-log plots of the realizations of the least squares error $\mathcal{SE}(h)$ for different values of h, along with the expected least squares error and the realized least squares error for h_M are shown on Figures 4-6.



(a) Example 3.1. (b) Example 3.2.

Figure 6: Realized least squares error $\mathcal{SE}(h)$ for the complex-step approximation, along with the expected error, and error for h_M, at noise level $\sigma = 0.5$.

Numerical results shown on Figures 4-6 confirm the theoretical results that the complex-step approximation with $h = h_M$ results with nearly optimal (minimal) error.

Although the experimental and the theoretical results show almost similar behavior of both approximations, the centered finite difference approximation and the complex-step approximation, in presence of noise, there is an expected advantage of the complex-step approximation over the centered finite difference approximation when applied to gradient approximations. Namely it is expected that the noise will have smaller influence on the complex-step approximation, since there is only one noisy function evaluation per (partial) derivative approximation versus two noisy function evaluations in the centered finite difference approximation. This also results in lower computational cost when complex-step approximation is used for approximating the gradient, which is expected to improve the efficiency of the optimization algorithms that employ complex-step derivative approximations. At the end of this Section, we give the complex-step gradient approximations in order to analyze the gradient approximation errors.

The complex-step approximation of the first derivative of f established in [23] and given by formula (4), has been extended for complex-step approximation of the gradient of an analytic function f at a point $x \in \mathbb{R}^n$, given by

$$\left[\nabla f(x)\right]_j \approx \frac{\mathrm{Im}\big(f(x+ihe_j)\big)}{h},\ j=1,2,...,n, \tag{20}$$

where h is a small positive real step and e_j is j-th coordinate vector, [16]. Generalizations of the complex-step approximation have also been explored in [1, 2].

Due to the analysis of the corresponding derivative approximation in the one dimensional case (4), the gradient approximation (20) allows us to avoid subtractive cancellation errors when approximating the gradient of f. This property becomes very important when working in presence of noise, since in case of subtractive cancellation errors, the noise will have greater influence on the accuracy of the approximation.

Using the noisy function measurements $F(z)$, defined by (8), we extend the complex-step gradient approximation (20) and define the *complex-step gradient approximation in presence of noise* of an analytic function f at $x \in \mathbb{R}^n$ by:

$$\left[\hat{g}^{CS}(x)\right]_j = \frac{\mathrm{Im}\big(F(x+ihe_j)\big)}{h},\ j=1,2,...,n, \tag{21}$$

where h is a positive real step and e_j is the j-th coordinate vector.

Analyzing *the absolute error of the complex-step approximation* defined by

$$\mathcal{AE}(h) = \left|\frac{1}{h}\mathrm{Im}(F(x+ih)) - f'(x)\right|, \tag{22}$$

we can derive bounds on the error of the complex-step gradient approximation in presence of noise and show that the value of h in the complex-step gradient approximation (21) is related to the noise level, similarly as in [14] for the centered finite difference gradient approximation in presence of noise.

Assuming that $|f'''|$ has an upper bound $M > 0$, the complex noise ζ is bounded by $D > 0$ i.e. $|\zeta(z)| \leq D$ for all $z \in \mathbb{C}$, where $|\zeta(z)|$ is the modulus of ζ, and by using (16), we obtain the following expression:

$$\begin{aligned}\mathcal{AE}(h) &= \left|\frac{1}{h}\mathrm{Im}(F(x+ih)) - f'(x)\right| \\ &= \left|-\frac{h^2}{6}f'''(u) + \frac{1}{h}\xi_2(x+ih)\right| \\ &\le \frac{h^2}{6}\left|f'''(u)\right| + \frac{1}{h}\left|\xi_2(x+ih)\right| \\ &\le \frac{M}{6}h^2 + \frac{D}{h}. \end{aligned} \tag{23}$$

Now, using (23), and assuming that the third partial derivatives of f and the noise ζ are bounded, i.e. assuming that there exist $M, D > 0$ such that $\left|\frac{\partial^3 f}{\partial x_j^3}(x)\right| \le M$, $j = 1, 2, ..., n$ for all $x = (x_1, x_2, ..., x_n)^T \in \mathbb{R}^n$ and $|\zeta(z)| \le D$ for all $z \in \mathbb{C}^n$, where $|\zeta(z)|$ is the modulus of the complex noise ζ, we can obtain an estimate for the approximation error when using the complex-step gradient approximation (21) i.e.

$$\begin{aligned}||\hat{g}^{CS}(x) - \nabla f(x)|| &= \left[\sum_{j=1}^{n}\left|\left[\hat{g}^{CS}(x)\right]_j - \left[\nabla f(x)\right]_j\right|^2\right]^{1/2} \\ &\le \left[\sum_{j=1}^{n}\left(\frac{M}{6}h^2 + \frac{D}{h}\right)^2\right]^{1/2} \\ &= \left(\frac{M}{6}h^2 + \frac{D}{h}\right)\sqrt{n}, \end{aligned} \tag{24}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$. Note that, the right hand side in (24) achieves minimum for $h = \sqrt[3]{3D/M}$, so small values for h are not recommended for complex-step gradient approximation in noisy environment.

In the next section, we implement the complex-step gradient approximation (21) to existing optimization algorithms in presence of noise.

4 Application to line-search optimization algorithms

Let us consider the unconstrained minimization problem in noisy environment:

$$\min_{x \in \mathbb{R}^n} f(x) \tag{25}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ has continuous partial derivatives and only noisy function evaluations $F(x)$ defined by (5) are available.

Several approaches have been suggested for solving the problem (25) assuming that noisy function evaluations (5) are available. Some of them are: random search method [22], coordinate search method [12], or nonmonotone line search methods [13, 14].

Nonmonotone line search methods have several advantages over monotone methods and have been successfully used in noise free environment, [5, 7, 15, 20, 24]. They have also been implemented in noisy environment, [13, 14]. These methods could accept search directions that are not necessarily descent directions, which is a frequent occurrence in presence of noise. Further, in noise free environment, these methods tend to converge to a global solution to problems with multiple local and global solutions, which is desirable property in noisy environment since the presence of noise may induce many false local solutions.

In this Section we are going to compare the performances of nonmonotone line-search methods presented in [14] which use centered finite difference gradient approximation (12), to the same methods that use complex-step gradient approximation, defined here by (21).

Let us briefly go through the nonmonotone line-search methods for solving the problem (25) that are presented in [14]. Having the current iterate x_k, the next iterate is defined by $x_{k+1} = x_k + \alpha_k d_k$, where $\alpha_k > 0$ is a positive step size, and d_k is a search direction. The step size α_k is chosen to satisfy the line-search rule of the form:

$$F(x_k + \alpha_k d_k) \leq \overline{F}_k + \eta_k - \alpha_k^2 \beta_k. \tag{26}$$

where $F(x)$ is a noisy functional value defined by (5). The sequences $\{\eta_k\}$ and $\{\beta_k\}$ are sequences of positive numbers such that:

$$\sum_{k=0}^{\infty} \eta_k = \eta < \infty, \tag{27}$$

while β_k is bounded and

$$\lim_{k\in K} \beta_k = 0 \quad \Rightarrow \quad \lim_{k\in K} \nabla f(x_k) = 0 \tag{28}$$

for some infinite set of indices $K \subseteq \mathbb{N}$. The choice of $\overline{F}_k$ in (26) determines different line search strategies. These are:

(LS1) $\overline{F}_k = F(x_k)$ and $\eta_k \equiv 0$ for all $k \in \mathbb{N}$, a monotone line search.

(LS2) $\overline{F}_k = F(x_k)$.

(LS3) $\overline{F}_k = \max\{F(x_k), \ldots, F(x_{\max\{k-\overline{M}+1,0\}})\}$, for some $M \in \mathbb{N}$.

(LS4) $\overline{F}_{k+1} = \frac{r_k Q_k(\overline{F}_k+\eta_k)+F(x_{k+1})}{Q_{k+1}}$, where $Q_{k+1} = r_k Q_k + 1$ and $r_k \in [r_{\min}, r_{\max}]$, $0 \le r_{\min} \le r_{\max} \le 1$, $\overline{F}_0 = F(x_0)$ and $Q_0 = 1$.

In our experiments, we use the same values of the parameters in the line-search procedures as in [14] i.e. for the sequences η_k and β_k defined by (27) and (28), we set

$$\eta_k = |F(x_0)|/k^{1.1} \quad \text{and} \quad \beta_k \equiv 1,$$

where x_0 is the initial iterate. The initial step length in the non-monotone line-search procedure (26) is set to $\alpha = 1$. If the line-search rule (26) is not satisfied then a smaller step is computed using the safeguard quadratic interpolation, [4]. The maximum number of 1000 line-search trials are allowed, and if no step is found within these attempts, the algorithm stops declaring the total number of function evaluations equal to $400n$. The rest of the parameters in the line-search strategies LS3 and LS4 are set to $\overline{M} = 10$ and $r_k = 0.85$.

Nonmonotone line-search methods presented in [14] are derivative-free methods, which means that they use an approximation of the gradient based on function values. Let $\hat{g}_k$ denote the approximation of the gradient $g = \nabla f$ at x_k. We will consider the following search directions d_k used in [14]:

(SGR) The spectral gradient search direction, $d_k = -\hat{g}_k/\sigma_k$, where $0 < \sigma_{\min} < \sigma_k < \sigma_{\max} < \infty$ is the spectral coefficient obtained recursively by:

$$\sigma_{k+1} = \max\left\{\sigma_{\min}, \min\left\{\sigma_{\max}, \frac{(\hat{g}_{k+1} - \hat{g}_k)^T(x_{k+1} - x_k)}{\|x_{k+1} - x_k\|^2}\right\}\right\}. \tag{29}$$

(BFGS) The BFGS search direction, $d_k = -H_k \hat{g}_k$, where H_k is the inverse Hessian approximation and is updated by:

$$H_{k+1} = \left(I_k - \rho_k s_k y_k^T\right) H_k \left(I_k - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T \tag{30}$$

with $s_k = x_{k+1} - x_k$, $y_k = \hat{g}_{k+1} - \hat{g}_k$ and $\rho_k = 1/(y_k^T s_k)$.

We use the same values of the parameters used in the above search directions, as in [14], i.e. $\sigma_0 = 1$, $\sigma_{\min} = 10^{-10}$ and $\sigma_{\max} = 10^{10}$ in SGR direction, and a rescaling the initial inverse Hessian approximation $H_0 = I$ in BFGS direction into:

$$H_0 \leftarrow \frac{y_k^T s_k}{y_k^T y_k} I.$$

In [14], the centered finite difference formula is used for approximating the gradient of f at x_k i.e.

$$[\hat{g}_k^{CFD}]_j = \frac{F(x_k + he_j) - F(x_k - he_j)}{2h}, \; j = 1, 2, ..., n, \tag{31}$$

where h is a positive real step and e_j is the j-th coordinate vector.

Additionally, assuming that the function f in (25) is an analytic function, we propose an implementation of the complex-step gradient approximation to the gradient of f at x_k, defined by

$$\left[\hat{g}_k^{CS}\right]_j = \frac{\mathrm{Im}\big(F(x_k + ihe_j)\big)}{h}, \; j = 1, 2, ..., n, \tag{32}$$

where h is a positive real step and e_j is the j-th coordinate vector. Let us note that if the function f is not an analytic function, one possible way to implement the complex-step approximation in practice, is first to approximate the function f by an analytic function (such as polynomial approximation), then to implement the complex-step approximation to the gradient of that analytic function. The same approach can also be taken if only real discrete values of f are known.

We have compared the performance of the complex-step gradient approximation (32) to the centered finite difference gradient approximation (31) implemented in the above described eight nonmonotone line-search algorithms (four line-search rules for each of two search directions). They have been tested on the set of 18 standard test problems from [18]. All test problems are least squares problems of the form $f(x) = \sum_{j=1}^{m} f_j^2(x)$ and are listed in Table 1 in [14].

For the purpose of obtaining noisy function evaluations $F(x)$ and $F(z)$, used in (31) and (32), we use simulated white Gaussian noise ξ and white circular noise $\zeta = \xi_1 + i \cdot \xi_2$, the both with a noise level $\sigma > 0$, as it is described in Section 3. Testing has been performed at three noise levels $\sigma = 0.01, 0.1, 0.5$.

As mentioned before, after evaluation of the approximation error, it was empirically established that the value $h = 3\sigma$, where σ is the noise level of the noise ξ, best suits the approximation with centered finite differences (31), see [14]. Here we have also empirically established that the same value $h = 3\sigma$, where σ is the noise level of the complex noise ζ, best suits the complex-step approximation (32).

The comparison of the gradient approximations (31) and (32) for implementation in nonmonotone line-search algorithms described above is presented through performance profiles defined in [6]. For that purpose, 50 independent test-runs have been performed for each of 18 problems, for each of eight algorithms and two gradient approximations. Each run stops if the maximum of $400n$ function evaluations have been reached, where n is the dimensionality of the problem, or when the objective function has been reduced "sufficiently" i.e $|F(x_k)| < (1 + 2\sigma) \cdot |F(x_0)| \cdot 10^{-3}$, where x_0 and x_k are the initial and the current iterates respectively. If a test-run stopped according to either of these criteria, we consider that run to be *successful*. Otherwise we count it as unsuccessful run, [14].

As it is defined in [6], performance profile is a performance measurement and comparison tool for optimization algorithms that uses a cumulative distribution function as a performance metric. Let $\mathbb{P}$ be the set of problems we are testing on, and let $\mathbb{S}$ be the set of solvers we use to solve the problems in $\mathbb{P}$. Let size$\{\mathbb{P}\} = n_p$. For each pair of a problem $p \in \mathbb{P}$ and solver $s \in \mathbb{S}$, we define $\varphi_{p,s}$ as the median number of function evaluations per dimension needed for solver s to solve problem p, and $\mathcal{M}_{p,s}$ as the mean absolute deviation from the median that corresponds to $\varphi_{p,s}$. These are then combined in the performance measure:

$$\pi_{p,s} = \varphi_{p,s} + \mathcal{M}_{p,s}. \tag{33}$$

We have chosen the median as a representative measure since the distribution of the number of function evaluations in the successful runs was highly nonsymmetric. As a baseline for comparison we use the performance ratio:

$$r_{p,s} = \frac{\pi_{p,s}}{\min\{\pi_{p,s} \mid s \in \mathbb{S}\}}$$

which measures how well a solver s performs on problem p compared to the performance of the best solver for that problem. Finally, we define the cumulative distribution function ρ_s for the performance ratio of solver s as a measure of the overall performance of the solver with:

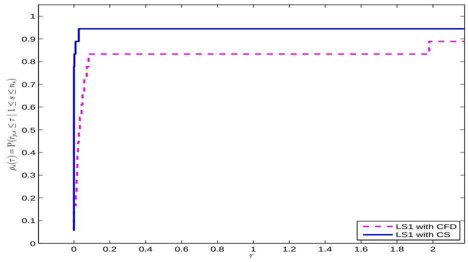
$$\rho_s(\tau) = \frac{1}{n_p} \text{size} \left\{ p \in \mathbb{P} \mid r_{p,s} \leq \tau \right\}.$$

The function $\rho_s(\tau)$ gives the probability for a solver $s \in \mathbb{S}$ that a performance ratio $r_{s,p}$ is within a factor of $\tau \in \mathbb{R}$ of the best possible ratio. The results are summarized on Figures 7- 9.

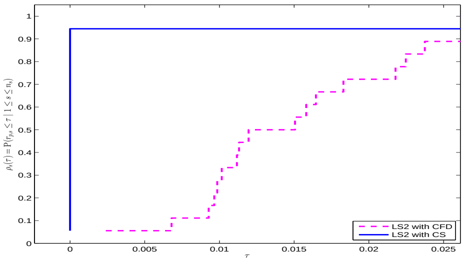
As it can be seen from the performance profiles in Figures 7- 9, the complex-step (CS) approximation has better performance and is more robust compared to the centered finite difference (CFD) approximation when the search direction is the spectral gradient (SGR), regardless the noise level and the line-search strategy. An exception is the monotone line-search LS1 at the noise level $\sigma = 0.1$, for which the CFD approximation results in an algorithm which is more robust at the end compared to CS approximation. When the search direction is BFGS, we have a different situation. For smaller noise levels $\sigma = 0.01$ and $\sigma = 0.1$, the CS approximation outperforms the CFD approximation, regardless the the line-search strategy, but at the greater noise level $\sigma = 0.5$, CFD approximation gives better performance profiles at the beginning for each line-search strategy, that end with the same robustness as the performance profiles obtained with CS approximation.

5 Conclusion

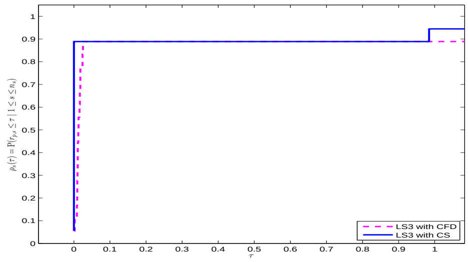
In this paper we explored and analyzed the complex-step derivative approximation in noisy environment. Different noise levels have been tested and results have been compared to the centered finite difference approximations. An analysis of the complex-step derivative approximation errors has been conducted. The complex-step gradient approximation is applied to the non-monotone line-search optimization algorithms. In the future, it would be



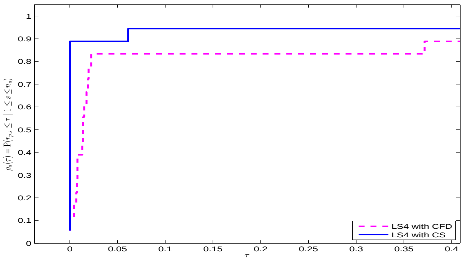
(a) SGR LS1



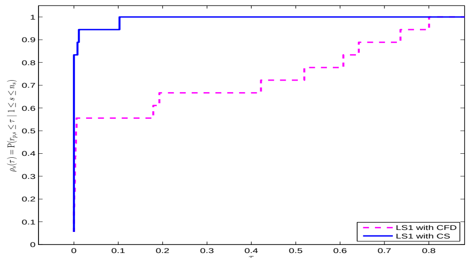
(b) SGR LS2



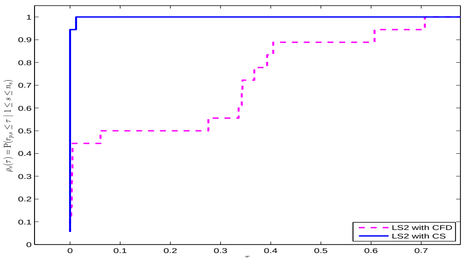
(c) SGR LS3



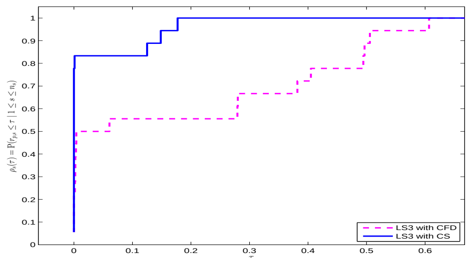
(d) SGR LS4



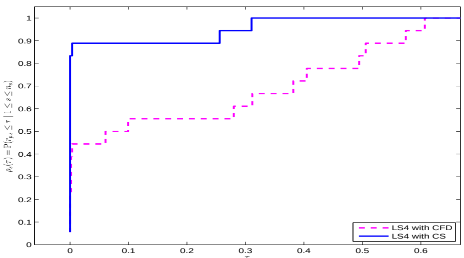
(e) BFGS LS1



(f) BFGS LS2

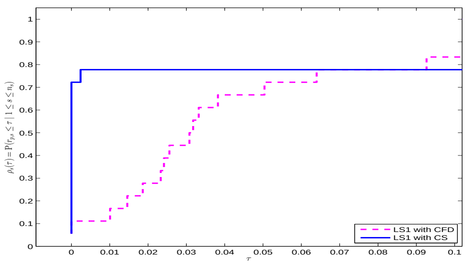


(g) BFGS LS3

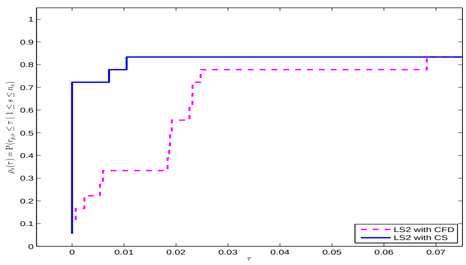


(h) BFGS LS4

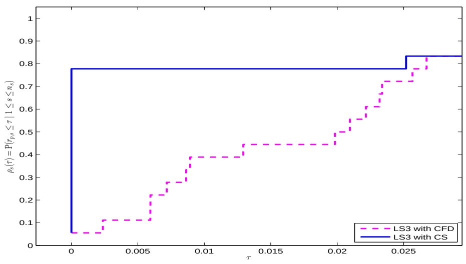
Figure 7: Performance profiles of the optimization algorithms that use centered finite difference (CFD) gradient approximation versus complex-step (CS) gradient approximation, at noise level $\sigma = 0.01$



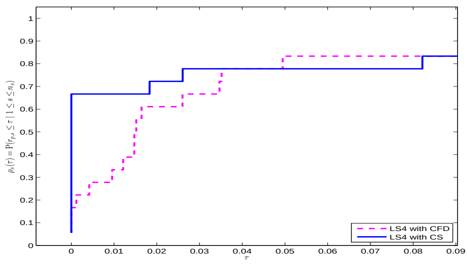
(a) SGR LS1



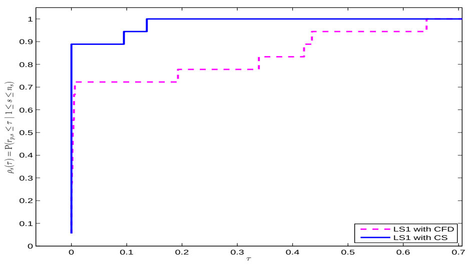
(b) SGR LS2



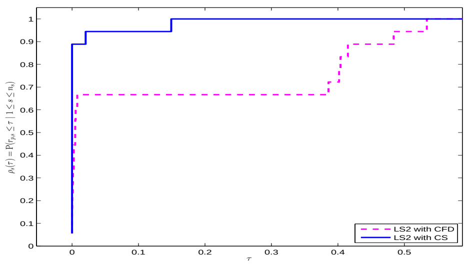
(c) SGR LS3



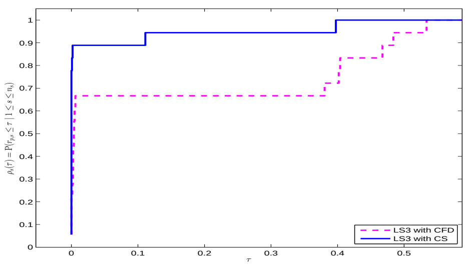
(d) SGR LS4



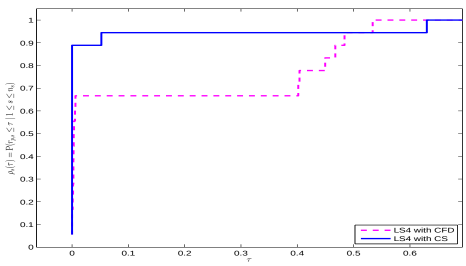
(e) BFGS LS1



(f) BFGS LS2

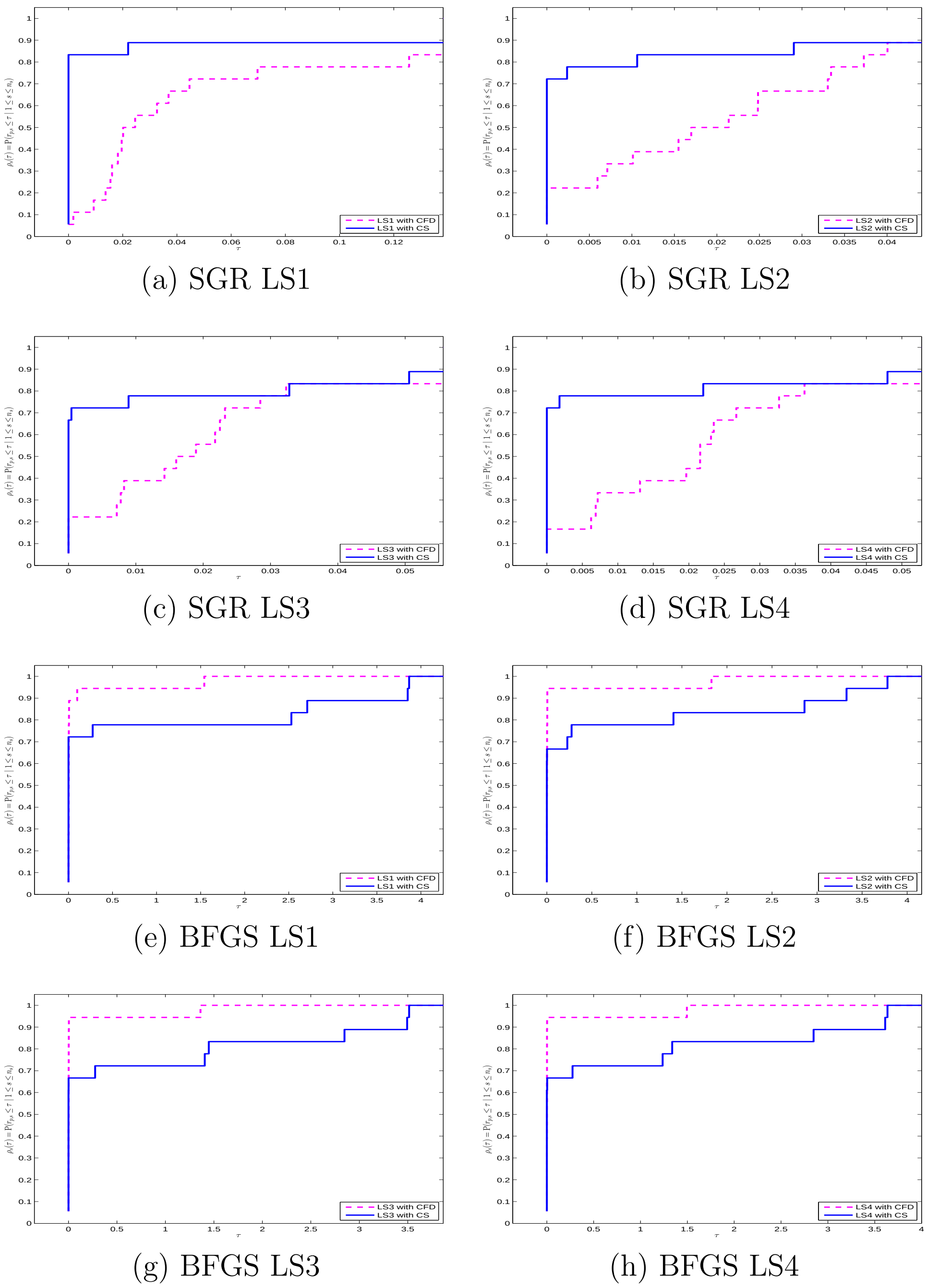


(g) BFGS LS3



(h) BFGS LS4

Figure 8: Performance profiles of the optimization algorithms that use centered finite difference (CFD) gradient approximation versus complex-step (CS) gradient approximation, at noise level $\sigma = 0.1$



(a) SGR LS1

(b) SGR LS2

(c) SGR LS3

(d) SGR LS4

(e) BFGS LS1

(f) BFGS LS2

(g) BFGS LS3

(h) BFGS LS4

Figure 9: Performance profiles of the optimization algorithms that use centered finite difference (CFD) gradient approximation versus complex-step (CS) gradient approximation, at noise level $\sigma = 0.5$

interesting to test different extensions of the complex-step derivative approximations that might be more suitable for implementation in noisy environment.

Acknowledgements. We are grateful to the anonymous referees whose comments and suggestions helped us to improve the paper.

References

[1] R. Abreu, D. Stich, J. Morales *On the generalization of the Complex Step Method*, J. Comput. Appl. Math. 241 (2013), 84-102

[2] R. Abreu, D. Stich, J. Morales *The Complex-Step-Finite-Difference method*, Geophys. J. Int. 202(1) (2015), 72-93

[3] P. Ciblat, M. Ghogho, *Harmonic Retrieval in Non-circular Complex-valued Multiplicative Noise: Cramer-Rao Bound*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 2 (2004), 489-492

[4] J. E. Dennis Jr., R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia (1996)

[5] M. A. Diniz-Ehrhardt, J. M. Martínez, M. Raydan, *A derivative-free nonmonotone line-search technique for unconstrained optimization*, J. Comput. Appl. Math. 219 (2) (2008), 383-397

[6] E. D. Dolan, J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Progr. Ser. A 91 (2002), 201-213

[7] L. Grippo, F. Lampariello, S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM. J. Numer. Anal. 23 (1986), 707-716

[8] S. Javidi, D. P. Mandic, A. Cichocki, *Complex Blind Source Extraction From Noisy Mixtures Using Second-Order Statistics*, IEEE Trans. Circuits Syst. I: Reg. Papers 57 (7) (2010), 1404-1416

[9] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, 1999

[10] J. N. Lyness, C. B. Moler, *Numerical differation of analytic functions*, SIAM J. Numer. Anal. 4 (2) (1967), 202-210

[11] J. N. Lyness, G. Sande, *Algorithm 413-ENTCAF and ENTCRE: Evaluation of normalized Taylor coefficients of an analytic function*, Comm. ACM 10 (1971), 669-675

[12] S. Lucidi, M. Sciandrone, *A derivative-free algorithm for bound constrained optimization*, Comput. Optim. App. 21 (2002), 119-142

[13] N. Krejić, Z. Lužanin, F. Nikolovski, I. Stojkovska, *A line search method with memory for optimization of noisy functions*, Proceedings of the V Congress of Mathematicians of Macedonia, September 24-27, 2014, Ohrid, R. Macedonia, Vol. 2 (2015), 15-25

[14] N. Krejić, Z. Lužanin, F. Nikolovski, I. Stojkovska, *A nonmonotone line search method for noisy minimization*, Optim. Lett. 9 (7) (2015), 1371-1391

[15] D. H. Li, M. Fukushima, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Optim. Methods Softw. 13 (2000), 181-201

[16] K. L. Lai, J. L. Crassidis, *Extensions of the first and second complex-step derivative approximations*, J. Comput. Appl. Math. 219 (2008), 276-293

[17] J. R. R. A. Martins, P. Sturdza, J. J. Alonso, *The Complex-Step Derivative Approximation*, ACM Trans. Math. Software 29 (3) (2003), 245-262

[18] J. J. Moré, B. S. Garbow, K. E. Hillström, *Testing unconstrained optimization software*, ACM Trans. Math. Software7 (1) (1981), 17-41

[19] J. J. Moré, S. M. Wild, *Estimating derivatives of noisy sumulations*, ACM Trans. Math. Software 38 (3) (2012), 19:1-19:21

[20] F. Nikolovski, I. Stojkovska, *New derivative-free nonmonotone line search methods for unconstrained minimization*, Proceedings of the Fifth International Scientific Conference - FMNS2013, Vol.1 Mathematics and Informatics (2013), 47-53

[21] J. Nocedal, S. J. Wright, *Numerical Optimization, 2nd edition*, Springer-Verlag, New York, 2006

[22] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2003

[23] W. Squire, G. Trapp, *Using complex variables to estimate derivatives of real functions*, SIAM Rev. 40 (1) (1998), 110-112

[24] H. Zhang, W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim. 14 (4) (2004), 1043-1056