



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2021, 2021:89

<https://doi.org/10.28919/cmbn/6835>

ISSN: 2052-2541

A NEW APPROACH TO IMPROVE OPTIMIZER PERFORMANCE THROUGH ALGORITHMS DIVERSIFICATION FOR IMAGE RECONSTRUCTION IN DIFFUSE OPTICAL TOMOGRAPHY

NADA CHAKHIM*, MOHAMED LOUZAR, MOHAMMED ALAOUI

MISI Laboratory, Faculty of Science and Technology, University Hassan First, Settat 26000, Morocco

Copyright © 2021 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. In this work, we explore a different way to construct optimizer algorithms for solving the inverse problem of Diffuse Optical Tomography by using diversification of two stochastic gradient-based algorithms, namely NADAM and AMSGrad. We will study the speed of convergence of the proposed new breed of algorithms, also we will discuss the quality of reconstructed images in both cases of free of noise and noisy measurement data. For analysis and exploration of the potential of the proposed algorithm, we use statistical simulations and analysis approach.

Keywords: diffuse optical tomography; inverse problem; diversification; random switching algorithm; image reconstruction; statistical simulation.

2010 AMS Subject Classification: 78A46, 68W20.

1. INTRODUCTION

In the last two decades, the inverse problem of the Diffuse Optical Tomography (DOT), in a highly scattering medium, has attracted the attention of the scientific community due to its characteristics of being a safe, non-invasive, and a cheap medical imaging tool [1,2,3,4] Diffuse optical tomography uses Near-infrared light [5,6,7] to visualize the distribution of the optical

*Corresponding author

E-mail address: n.chakhim@uhp.ac.ma

Received September 28, 2021

parameters in the biological tissues such as the brain and breast [8,9,10,11,12]. Data generation system consists of a light source that delivers Near-infrared light at different locations of the studied domain and detectors located at domain boundary that measures the light transmitted through the tissue [7]. We should solve two mathematical problems to recover the optical properties of tissue in DOT: The forward and the inverse problem. The forward problem is to pursue the propagation of scattered light in biological tissue. The inverse problem is to recover the optical parameters from scattered light measurements obtained from the forward model [13]. Image reconstruction in DOT is a complex problem to solve due to the ill-posedness of the inverse problem, which is subject to the nonlinearity of the optical parameters. Throughout this paper, The light propagation within a tissue is modeled by the diffusion equation (DE), in the case of the continuous wave (CW) method, for the sake of its simplicity and its low computational cost [14].

This fairly interesting technique for medical imaging suffers nevertheless from the complexity of the related inverse problem as mentioned before, which hinders its broad adoption in medical imaging practice, compared to other readily available procedures like ultrasound and mammography, due to the time consuming numerical optimization solution. The most crucial factor that impacts the acceptance of this new promising procedure of DOT is the quality of the generated image of the biological tissue under-diagnosis and the speed at which we can get feedback from it in real-life configuration [15, 16]. Both of these issues have been studied in previous work [17] that aimed to enhance the solution of DOT problem practically by acting on the optimizer algorithm used to solve the underlying inverse problem. The conclusion reached can be briefly stated that adaptive moment gradient descent algorithms show statistically significant higher performance in solving numerically the inverse problem of DOT, which represents a significant step toward making DOT procedure more ready for practical application. However, with a caveat, that is the assessed algorithms, namely NADAM [18] and AMSGrad [19], performed individually well with respect to only one of the previously cited criteria of interest, namely quality of reconstructed image and speed of convergence while achieving inferior performance concerning the other. In this situation, NADAM is very efficient algorithm in providing superior quality of reconstructed images, but consumes in the way much larger number

of iterations than AMSGrad. Conversely, this latest algorithm is very fast in reaching convergence, yet the reconstructed images are far less clear and of low quality. In the present work, we devised a procedure to mix these two optimizers into a new brand of algorithms. This procedure of combining the two algorithms that we will refer to in this paper by "diversification" will receive the proper description and explanation in Section 3 hereafter. Statistical simulation results will show that, interestingly and significantly enough, the proposed brand of algorithms inherit the best of both worlds by accelerating the rate of convergence and achieving a good quality of the reconstructed images, which is a significant step toward accelerating and enhancing DOT medical imaging techniques to be adopted and used broadly in the practical medical context.

The rest of this paper is organized as follows: In Section 2, we give a brief overview of the mathematical formulation of the diffusion approximation in the continuous wave (CW) case and describe the inverse problem. In section 3, we describe the proposed new algorithm that we used to reconstruct the absorption coefficient of DOT. In Section 4, we discuss the results obtained by our algorithm in the case of free noise and noisy measurement data. Finally, in Section 5, we summarize some conclusions.

2. FORWARD AND INVERSE PROBLEM

2.1. Forward problem. In this section we describe the mathematical formulation of the diffusion approximation (DA).

Let $\Omega \subset R^n, n = 2, 3$, be our domain of interest, and $\partial\Omega$ the boundary of Ω . Then the DA inside the domain Ω satisfies the partial differential equation

$$(1) \quad -\nabla[D(r)\nabla\Psi(r)] + \mu_a(r)\Psi(r) = 0 \quad r \in \Omega$$

with the Robin-boundary condition

$$(2) \quad \Psi(r) + 2aD(r)\frac{\partial\Psi(r)}{\partial\hat{n}} = S(r) \quad r \in \partial\Omega$$

where $\Psi(r)$ is the photon density, $D(r)$ is the diffusion coefficient defined by $D(r) = \frac{1}{3(\mu_a + \mu'_s)}$. a is the Fresnel reflection coefficient, which depends on the mismatch between the refractive indices, μ_a and μ_s the absorption and scattering coefficient respectively, and μ'_s the reduced

scattering coefficient expressed as $\mu'_s = (1 - g)\mu_s$, where g is the anisotropic factor. $S(r)$ describes the boundary condition for the incoming radiation and \hat{n} is the outward normal vector to Ω .

2.2. Inverse problem. The main goal of the inverse problem of DOT, is to determine the optical parameters μ_a and μ_s based on the boundary measured data y_i such that

$$(3) \quad F_i(\mu_a, \mu_s) = y_i \quad 1 \leq i \leq s$$

where F_i is the forward operator which is assumed to be Fréchet differentiable, and y_i the approximate measured data. through out this paper, we will focus our attention on the reconstruction of the absorption coefficient μ_a , and we will consider that the scattering parameter μ_s is known. Then, the inverse problem of DOT can be written as follows

$$(4) \quad J(\mu_a) = \frac{1}{2} \sum_{i=1}^s (F_i(\mu_a) - y_i)^2$$

Then this problem can be stated in term of an optimization problem with an additive regularization term

$$(5) \quad \mu_a^* = \operatorname{argmin} J(\mu_a) + \lambda R(\mu_a)$$

where $R(\mu_a)$ is the regularization operator that enforces smoothness conditions in the solution, and λ is the regularization parameter.

3. PROPOSED ALGORITHM

This section will describe our proposed approach to construct the new brand of algorithms to solve the inverse problem of DOT, using the method that we denounce by " Algorithms Diversification ". The main idea behind this technique of algorithm construction procedure is to benefit from the adversarial advantages of Nadam and AMSGrad. In our previous work [17], statistical numerical results have shown that when the learning rate is in the range $[0.001, 0.2]$ NADAM algorithm takes more iterations to converge than AMSGrad. However, always in the same learning rate range, the NADAM algorithm achieves a significantly better-reconstructed image quality relative to AMSGrad. To exploit the adversarial advantages of NADAM and AMSGrad, we propose a diversified algorithm, called NADA-p (Nadam-AMSGrad Diversified

Algorithm with ratio p), that combines NADAM and AMSGrad algorithms in a stochastic fashion according to a probability ratio p . Literally speaking, we combine the two gradient-based algorithms with incompatible advantages (speed and reconstruction quality), hoping to design an algorithm that inherits the benefits of both, since, as statistically established in the previous article, NADAM is significantly more efficient in terms of reconstruction quality. However, the better quality it achieves, the poorer it performs in terms of the rate of convergence. In contrast, AMSGrad evolves in the opposite direction, giving a better convergence speed than NADAM, but a low quality of reconstructed images. For performance assessment, we will compare the proposed algorithm with AMSGrad and NADAM. For more details about NADAM and AMSGrad algorithms, we refer the reader to [20]. The implementation of the NADA-p algorithm is outlined below in algorithm 1. Also, we empirically note that the NADA-p algorithm inherits the same guarantees of convergence from NADAM and AMSGrad under the assumptions that they both converge. The case where NADAM or AMSGrad or both fail to converge is not examined in this work.

Algorithm 1: Pseudocode of NADAM-AMSGrad Diversified Algorithm with ratio p

(NADA-p)

Require: $p, \mu_a^0, \alpha_{Nadam}, \alpha_{AMSGrad}, \rho_1, \rho_2$, and ε
Ensure: μ_a^n
while J not converged **do**
 $b \leftarrow bernoulli(p)$
 $k \leftarrow k + 1$
 $g_k \leftarrow \nabla J_{\mu_a}(\mu_a^{k-1})$
 $m_k \leftarrow \rho_1 \cdot m_{k-1} + (1 - \rho_2) \cdot g_k$
 $v_k \leftarrow \rho_1 \cdot v_{k-1} + (1 - \rho_2) \cdot g_k^2$
 $\hat{m}_k \leftarrow \frac{m_k}{(1 - \rho_1^k)}$
 $\hat{v}_k \leftarrow \frac{v_k}{(1 - \rho_2^k)}$
 $\hat{x}_k \leftarrow \max(v_k, \hat{x}_{k-1})$
if $b == 0$ **then**
 $\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\alpha_{NADAM}}{\sqrt{\hat{v}_k} + \varepsilon} (\rho_1 \hat{m}_k + \frac{1 - \rho_1}{1 - \rho_1^k} g_k)$
if $b == 1$ **then**
 $\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\alpha_{AMSGrad}}{\sqrt{\hat{x}_k} + \varepsilon} m_k$
end while

In the following outlined implementation of the NADA-p algorithm, p is a decision parameter taken in the range $[0, 1]$, and b is the value of the random variable that follow Bernoulli distribution with parameter p at each iteration. α_{Nadam} and $\alpha_{AMSGrad}$ are the learning rate parameters for Nadam and AMSGard algorithms, respectively, ρ_1 and ρ_2 are the exponential decay rates for the moment estimates. We denote by ε the stabilization parameter, and μ_a^0 is the initial guess estimation.

4. PERFORMANCE EVALUATION

To evaluate the reconstructed image quality of the proposed algorithm for the problem of DOT, peak noise ratio (PSNR) and Structural similarity (SSIM) [21,22], were calculated as expressed below

$$PSNR = 10\log_{10}\left(\frac{\max^2(\mu_a^{true})}{\frac{1}{N}\sum_{i=1}^N(\mu_a^{recon}(i) - \mu_a^{true}(i))^2}\right),$$

$$\begin{aligned} SSIM(\mu_a^{true}, \mu_a^{recon}) &= [l(\mu_a^{true}, \mu_a^{recon})]^x \\ &\quad + [c(\mu_a^{true}, \mu_a^{recon})]^y + [s(\mu_a^{true}, \mu_a^{recon})]^z, \end{aligned}$$

$$\begin{cases} l(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\bar{m}_{\mu_a^{true}}\bar{m}_{\mu_a^{recon}} + C_1)}{(\bar{m}_{\mu_a^{true}}^2 + \bar{m}_{\mu_a^{recon}}^2 + C_1)} \\ c(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_2)}{(\sigma_{\mu_a^{true}}^2 + \sigma_{\mu_a^{recon}}^2 + C_2)} \\ s(\mu_a^{true}, \mu_a^{recon}) = \frac{(\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_3)}{(\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_3)} \end{cases}$$

where $l(\mu_a^{true}, \mu_a^{recon})$, $c(\mu_a^{true}, \mu_a^{recon})$, and $s(\mu_a^{true}, \mu_a^{recon})$ are the luminance, contrast and structure variations between the true image μ_a^{true} and reconstructed image μ_a^{recon} , respectively, and $x > 0$, $y > 0$, and $z > 0$ are three parameters used to adjust relative importance of the three components of the similarity measure. $\bar{m}_{\mu_a^{true}}$ and $\bar{m}_{\mu_a^{recon}}$ are the means of pixel values of μ_a^{true} and μ_a^{recon} , respectively. We denote by $\sigma_{\mu_a^{true}}$, $\sigma_{\mu_a^{recon}}$, and $\sigma_{\mu_a^{true}\mu_a^{recon}}$ the standard deviation of μ_a^{true} and μ_a^{recon} , and the covariance of image μ_a^{true} and μ_a^{recon} , respectively. C_1 , C_2 , and C_3 are constants.

5. SIMULATION AND DATA PROCESSING

we executed a two stages plan for our simulation experiments. First, we run simulations for free noise data to compare the performance of the NADA-p with NADAM and AMSGrad, where we sampled the probability ration p uniformly randomly in the range $[0, 1]$. We run a similar simulation with the measured data contaminated with 5% white Gaussian noise in the second stage.

To generate a set of data, we use the Toast++ software [23], which solves the forward problem described above. We assume that the medium is highly scattering such that, $\mu_a \ll \mu_s$. 16 sources and 16 detectors are located on the boundary of the domain with equal distance. The location, size and number of anomalies in μ_a are chosen randomly with a background $\mu_a^{bkg} = 0.05mm^{-1}$

and $\mu_s'^{bkg} = 4mm^{-1}$. We solve the forward problem by using the finite element method as described in [24]. We use a circular mesh of radius 20 mm, containing different inclusion sizes and shapes, with 30 504 nodes and 82 332 tetrahedral elements for the forward problem. To avoid inverse crime [25], we use different meshes for the inverse problem with 18 635 nodes and 41 723 tetrahedral elements. The number of anomalies is taken randomly in the range 1 to 3. The regularization parameter λ is set to be equal to 10^{-10} . We solve the inverse problem of DOT by using the NADA-p algorithm presented in the pseudo-code below. For comparison purposes, we use Nadam and AMSGrad algorithms. For more information and details, we refer the reader to [20].

The choice of the learning rate parameters of both NADAM and AMSGrad is based on the results from our previous work [17]. For Nadam, we choose a learning rate that gives the best possible reconstructed image quality without paying attention to the resulting low convergence speed. On the other hand, for AMSGrad we take a learning rate value that gives the best convergence rate despite the resulting low quality of reconstructed images. Based on the results in [17], α_{Nadam} and $\alpha_{AMSGrad}$ are set to be equal to 0.01 and 0.2, respectively. We fixed all the other hyperparameters for all optimizers to the recommended values from the corresponding literature with $\rho_1 = 0.9$ and $\rho_2 = 0.999$. The algorithm stops until the stopping criterion is satisfied. The complete data analysis was carried out in R [26].

6. RESULTS AND DISCUSSION

Our study is divided into two parts. For the first section, we discuss the performance of the proposed algorithm in the case of free noise measurement data. In the second section, we will investigate our algorithm's performance and robustness when white Gaussian noise is added to measurement data. For both cases, we study the behavior of speed of convergence and the quality of the reconstructed images with respect to the output of the three algorithms, namely NADAM, AMSGrad, and NADA-p. To statistically compare different groups of resulted simulation samples, Kruskal-Wallis paired test [27] will be carried out. The resulting p-value of the test is at the top of each graph.

6.1. Simulation with free of noise Data measurements. We discuss in this first part the results from the simulation with free noise measurement data. We present the simulation results of 200 instances. For this simulation, the maximal number of iterations has been set to be equal to 120 iterations.

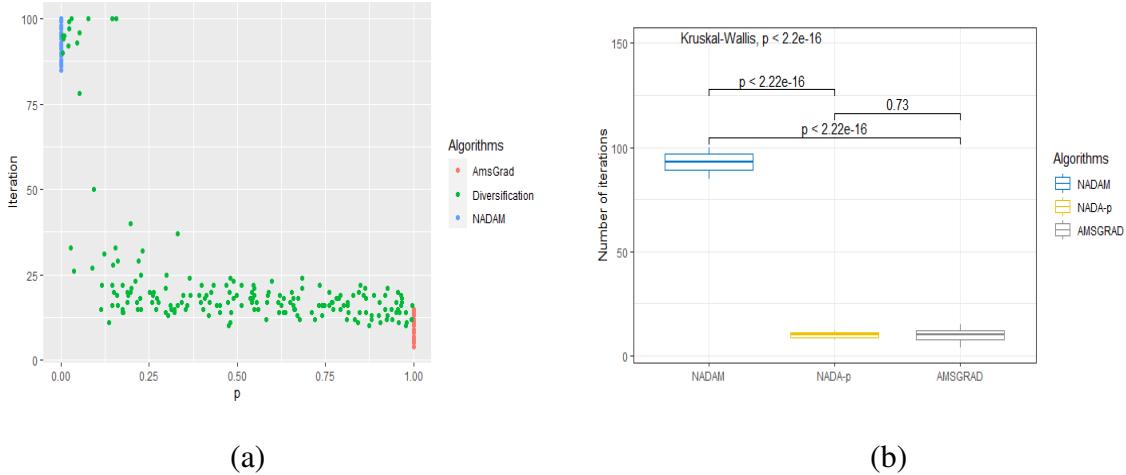


FIGURE 1. (a) Relationship between number of iterations and the choice of the ratio p for Nada-p, data from Nadam ($p=0$) and AMSGrad ($p=1$) is add for visual comparison. (b) Kruskal–Wallis test results for the number of iterations between different optimizers.

Figure 1(a) presents the impact of the ratio p on the number of iterations. We take $p = 0$ for the case of Nadam (which is equivalent to NADA-0) and $p = 1$ for the case of AMSGrad (which is equivalent to NADA-1) algorithm. From Figure 1(a), we note that our proposed algorithm converges faster than the NADAM algorithm when p is greater than 0.25. However, it takes more iteration number when p ranges in $[0, 0.25]$. To gain more credible evidence about this remark, we compare the mean of the number of iteration for each algorithm as shown in Figure 1(b), restricting our attention to the NADA-p where p is in the range $[0.2, 0.99]$. Box plot in Figure 1(b) show that there is a significant difference ($p - value < 0.05$) for the speed of convergence between NADAM on the one hand and NADA-p and AMSGrad on the other hand. At the same time, there is no significant difference between NADA-p and AMSGrad in terms of speed of convergence.

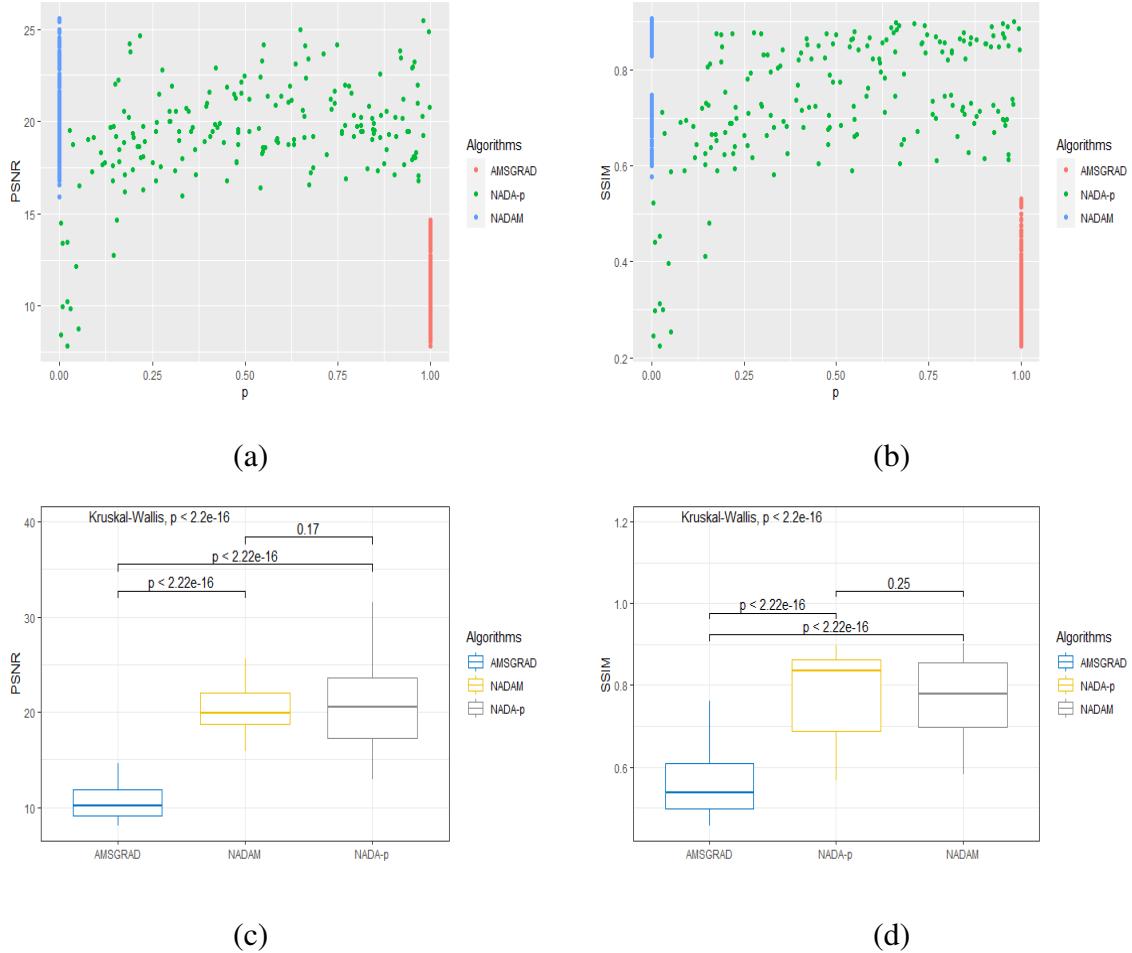


FIGURE 2. (a) Impact of ratio p on PSNR. (b) Impact of ratio p on SSIM. (c) Kruskal–Wallis test results for the PSNR difference between optimizers. (d) Kruskal–Wallis test for the SSIM difference between different optimizers.

Concerning the quality of the reconstructed images, the PSNR, and the SSIM values have been calculated as described in Section 4, for each simulated instance. The SSIM values range between 0 and 1. A result with a high value of SSIM is considered a good reconstructed image. Figure 2 (a) shows the relationship between the PSNR and the probability ratio p parameter choice for the NADA-p algorithm. From Figure 2 (a), we note that the PSNR of our proposed algorithm is the same as NADAM algorithm when p is in the range $[0.25, 1]$. The same observation can be concluded about the SSIM from Figure 2(b). Figure 2 (c) and Figure 2 (d) show the Kruskal-Wallis test results for PSNR and SSIM, respectively, between different pairs

of algorithms. From Figure 2(c) and Figure 2(d), we can conclude that there is significant statistical differences ($p - \text{value} < 0.05$) between AMSGrad and the other two algorithms, and no significant statistical difference between NADAM and NADA-p ($p - \text{value} > 0.05$) in terms of PSNR and SSIM.

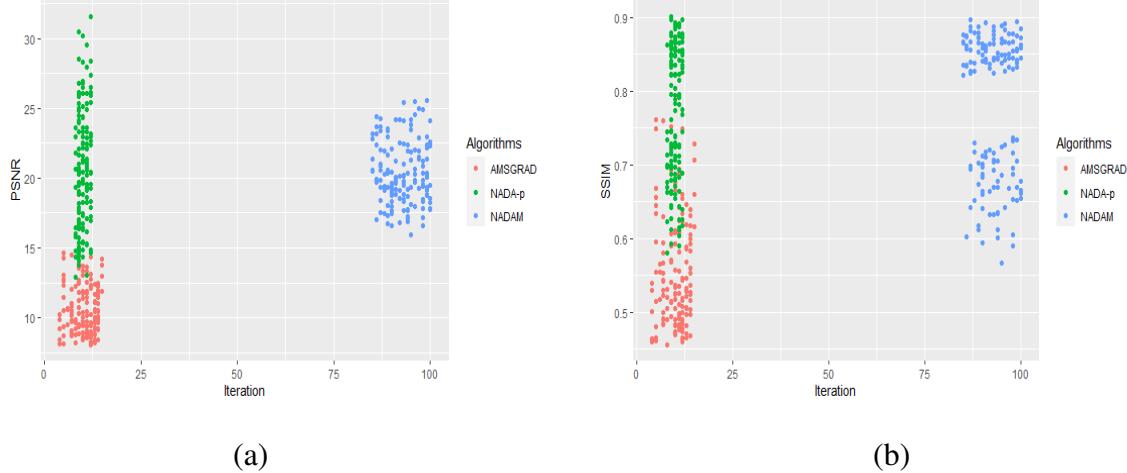


FIGURE 3. (a) Relationship between PSNR and the number of iterations. (b) Relationship between SSIM and the number of iterations.

To further illustrate these observations, we present the relationship between PSNR/SSIM and the number of iterations to localize the NADA-p algorithm's performance relative to both others. From Figure 3(a) and Figure 3(b), we outline that NADA-p algorithm can achieve high values of PSNR and SSIM with a few numbers of iterations contrarily to NADAM that performs the same in terms of quality reconstruction but with a much higher number of iterations.

The most remarkable feature of the NADA-p algorithm that we note from Figure 3 is that, even though NADA-p is in a way an "average" algorithm from NADAM and AMSGrad, it does not yield an "average" response with respect to the rate of convergence and quality of reconstructed images. Still, it inherits the good features from both NADAM and AMSGrad, which is a very interesting behavior that allows us to construct a better optimizer from combination of two lesser good ones. This fact is significantly exhibited when we simulate with a specific value of ratio p around the middle (like $p = 0.6$) as shown in Figure 4.

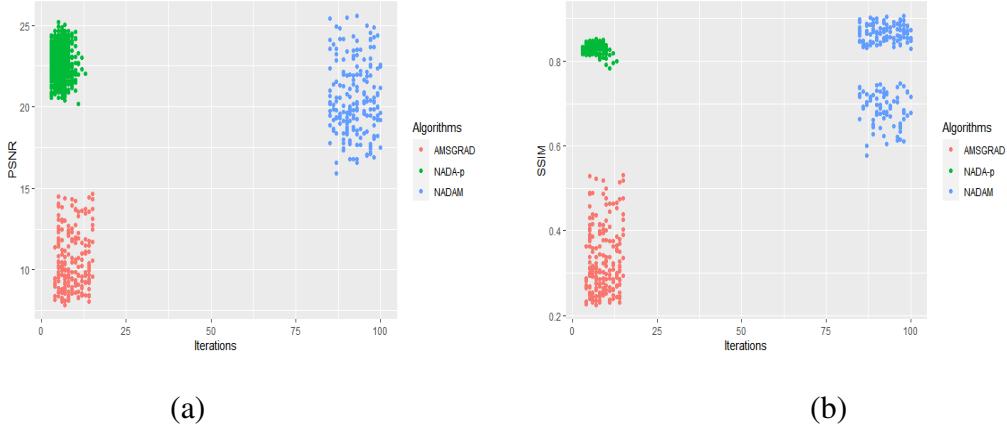


FIGURE 4. (a) Relationship between the PSNR and the number of iterations for NADAM, AMSGrad and NADA-0.6. (b) Relationship between the SSIM and the number of iterations for NADAM, AMSGrad and NADA-0.6.

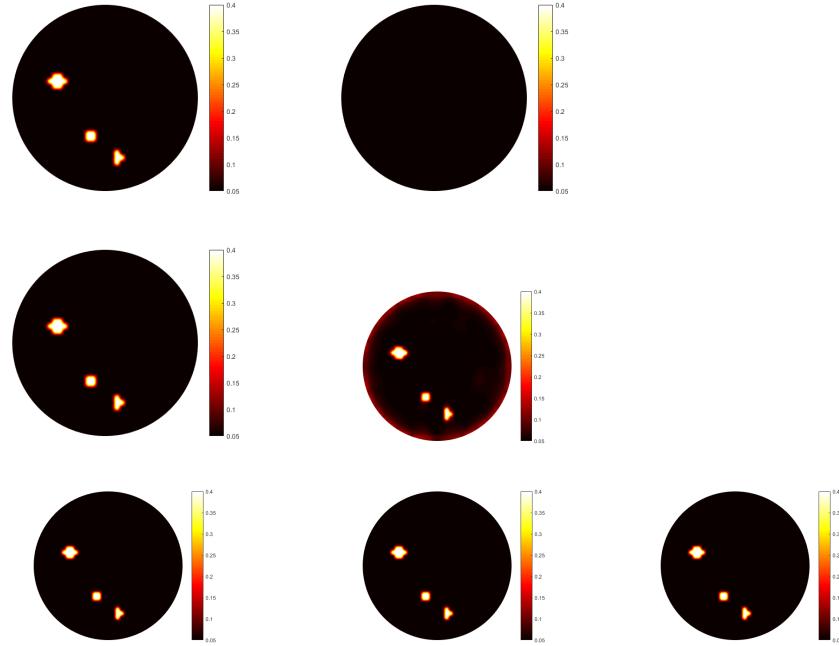


FIGURE 5. Reconstruction results of the absorption coefficient μ_a with three inclusions with different algorithms in the case of free noise measurement data. The first row presents the target image (left) and initial guess image (right). The second row, presents the reconstruction images using NADA-0 (left), and NADA-1 (right). The third row, presents the reconstruction images using NADA-0.3, NADA-0.5, and NADA-0.6 from left to right, respectively.

Figure 5 shows the reconstruction results of the absorption coefficient with three inclusions by different reconstruction algorithms. The first row presents the target image and the initial guess image from left to right. The second row shows the reconstruction images using NADAM after 100 iterations (left) and AMSGrad after 15 iterations (right). The third row presents the reconstruction images using NADA-0.3 , NADA-0.5 , and NADA-0.6 after 40, 25 and 10 iterations. Results demonstrated in Figure 5 show that our algorithm can localize the shape and the size of inclusions in a fewer number of iteration compared to NADAM, with high contrast compared to AMSGrad.

6.2. Simulation with white noise contaminated Data measurements. Since the measurement data is often contaminated with noise, and to further assess the performance and effectiveness of the proposed algorithm against noisy data, we add 5% Gaussian noise to the boundary measured data. We increase the maximum number limit of iterations to 200 due to the slow convergence rate of the algorithms in the presence of noise. We run the experiment for 200 simulation instances.

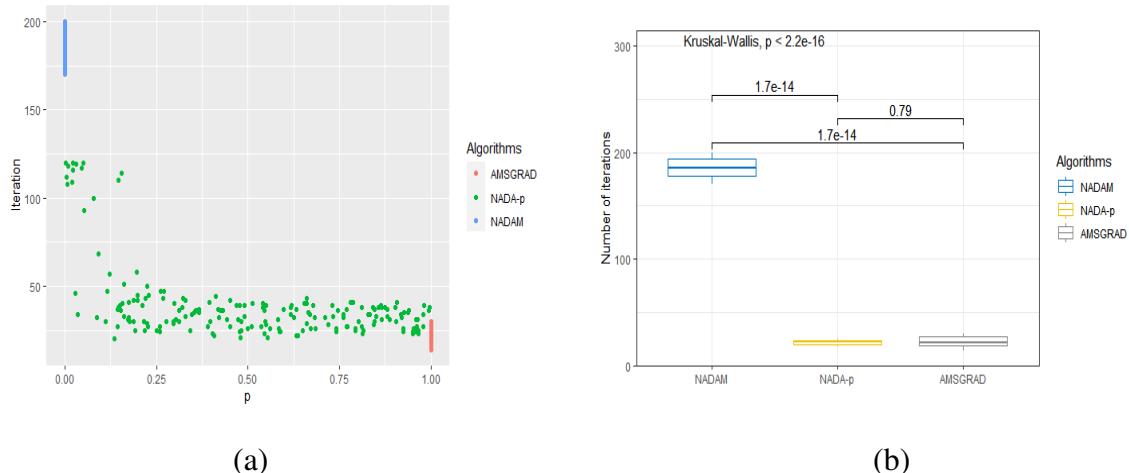


FIGURE 6. (a) Relationship between number of iterations and the choice of the ratio p for NADA-p. (b) Kruskal–Wallis test results for the number of iterations between different optimizers.

As shown in Figure 6(a), regarding the speed that the NADA-p algorithm takes to converge on average, it is much faster than the NADAM algorithm when the ratio p ranges in $[0.125, 1]$. Moreover, the speed of convergence of the proposed algorithm can be considered similar to

the speed of convergence of AMSGrad algorithm when p is greater than 0.125. To statistically confirm this observation, we run the Kruskal-Wallis test restricting our attention only to the instances with values of p in the range $[0.125, 0.99]$. Kruskal-Wallis test results are shown in Figure 6(b). From Figure 6(b), we note that there is a significant statistical difference between NADAM and the other two algorithms ($p-value < 0.05$) in terms of rate of convergence. However, we note that there is no significant difference between AMSGrad and NADA-p algorithms ($p-value > 0.05$).

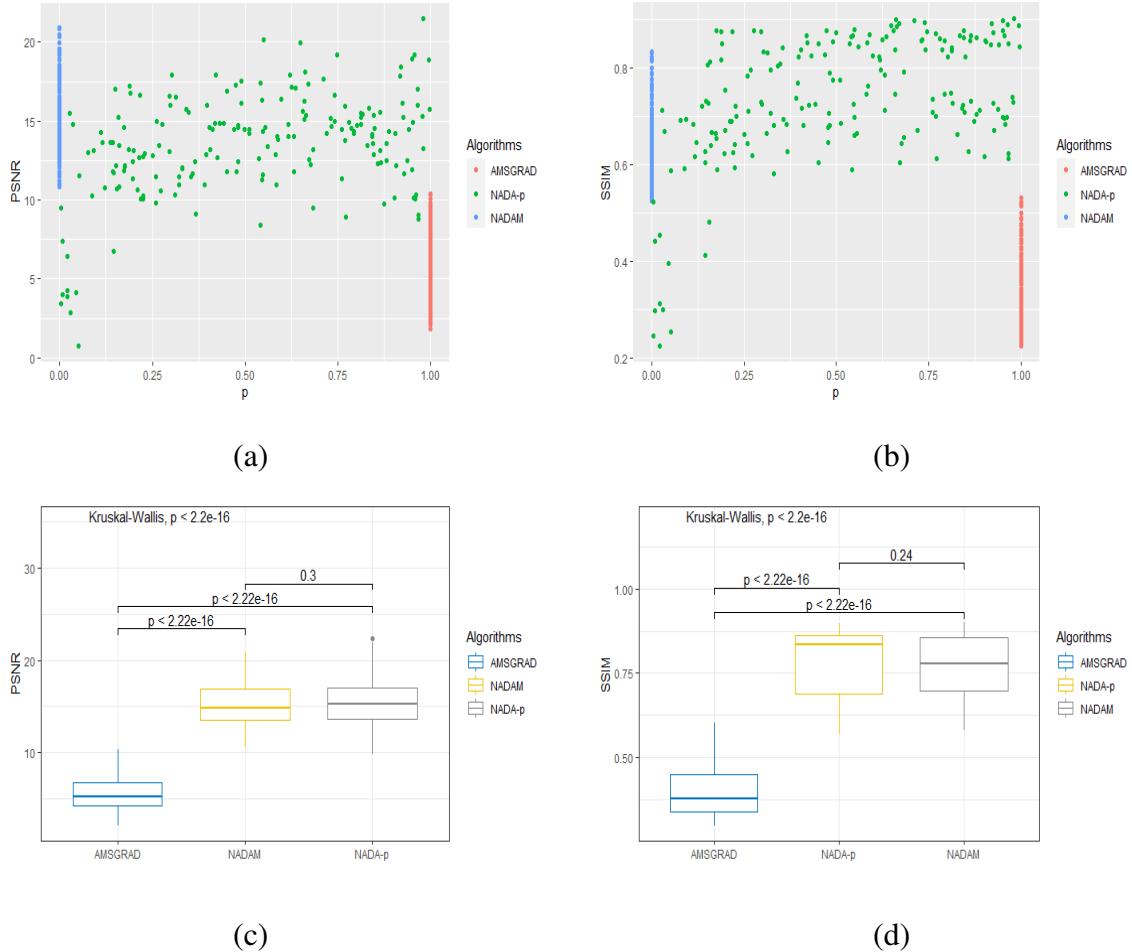


FIGURE 7. (a) Impact of ratio p on PSNR. (b) Impact of ratio p on SSIM. (c) Kruskal–Wallis test results for the PSNR difference between optimizers. (d) Kruskal–Wallis test for the SSIM difference between different optimizers.

Finishing our evaluating of the impact of noise on the performance of our algorithm, this time in terms of the quality of reconstructed images. We notice first from Figure 7(a) and Figure

7(b) that the NADA-p algorithm behaves as NADAM algorithm when p ranges in $[0.125, 1]$ while outperforming AMSGrad in terms of PSNR and SSIM in that same range. We run the Kruskal-Wallis test to confirm this observation as shown in Figure 7(c) and Figure 7(d). From Figure 7(a) and Figure 7(b), we conclude that there is no significant statistical differences ($p - value > 0.05$) between NADAM and NADA-p in term of PSNR and SSIM. However, we notice a significant statistical difference between AMSGrad and both other two algorithms regarding the quality of reconstructed images.

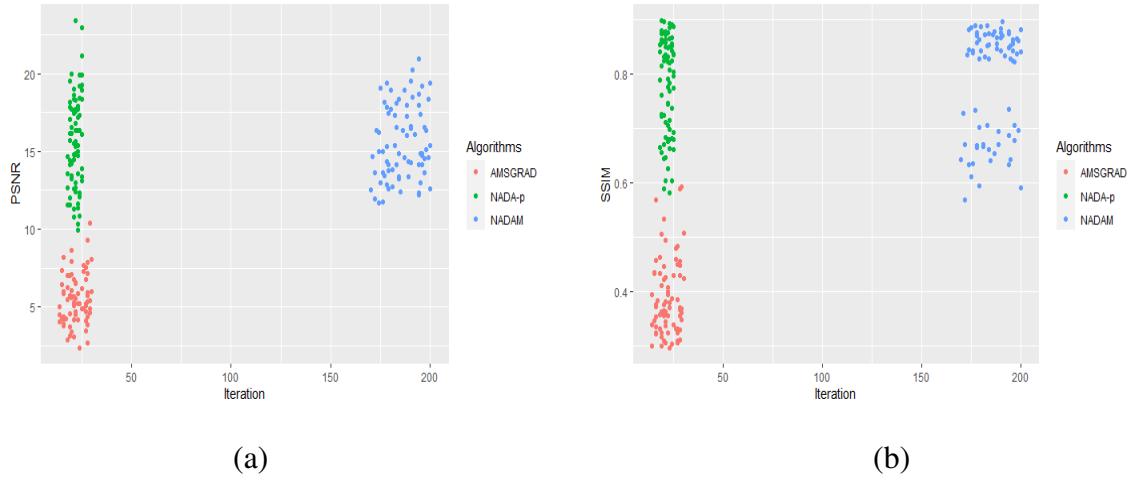


FIGURE 8. (a) Relationship between the PSNR and the number of iterations for each algorithm. (b) Relationship between the SSIM and the number of iterations for each algorithm.

Finally, to conclude this performance comparison as we did in the previous section about free noise data, Figure 8 illustrate clearly the relative performance of the three algorithms with respect to the rate of convergence and quality of reconstruction. From Figure 8, we notice that NADA-p and NADAM algorithm outperform AMSGrad algorithm in terms of quality of reconstructed images. Furthermore, the NADA-p algorithm can achieve the same quality of reconstructed images as the NADAM algorithm in fewer iterations. Thus, we can conclude that the NADA-p algorithm is more efficient and robust even in noisy measurement data and can inherit the advantages of NADAM and AMSGrad.

Here too, we outline the same remarkable feature of the NADA-p algorithm that we noted before in the case of the free noise data. From Figure 8 is that, even though NADA-p is in a way

an "average" algorithm from NADAM and AMSGrad, it does not yield an "average" response with respect to the rate of convergence and quality of reconstructed images. Still, it inherits the good features from both NADAM and AMSGrad, which is a very interesting behavior that allows us to construct a better optimizer from combination of two lesser good ones.

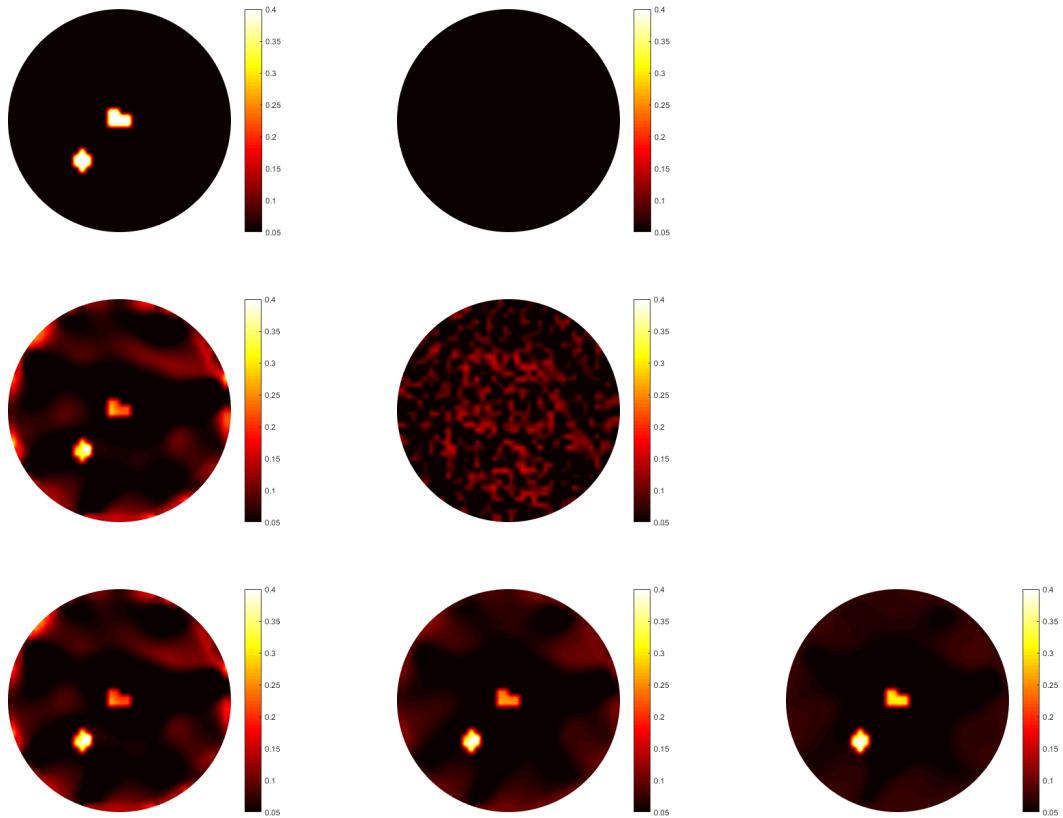


FIGURE 9. Reconstruction results of the absorption coefficient μ_a with two inclusions by different algorithms in the case of noisy measurement data. The first row presents the target image (left) and initial guess image (right). The second row, presents the reconstruction images using Nadam (left), and AMSGrad (right). The third row, presents the reconstruction images using NADA-0.1, NADA-0.4, and NADA-0.6 from left to right, respectively.

Figure 9 shows the reconstruction results of the absorption coefficient μ_a with two inclusions by different algorithms in the case when 5% of white Gaussian noise is added to boundary measurement data. The first row presents the actual image (left) and the initial guess image

(right). The second row shows the reconstruction images using NADAM with 200 iterations (left) and AMSGrad with 30 iterations (right). The third row presents the reconstruction images using NADA-0.1, NADA-0.4, and NADA-0.6 from left to right, with 100, 55, and 40 iterations. Reconstruction results show that AMSGrad fails to localize the inclusions in the case of noisy data. For the NADAM algorithm, we notice that the shape of inclusions matches those of the true image but with some artifacts. For the case of the reconstructed images obtained by NADA-p algorithm, we notice that, the shape and the size of inclusions match with those figuring in the actual image. Moreover, the image reconstructed by NADA-p algorithm has better contrast than those obtained by NADAM and AMSGrad.

7. CONCLUSION

In this paper, a new optimizer algorithm to solve the inverse problem of DOT has been presented and discussed. A large amount of random simulation has been carried out to evaluate the convergence behavior and quality of reconstructed images for both cases of free of noise and contaminated measurement data. Results and their analysis show that the proposed algorithm can achieve convergence significantly faster compared to NADAM algorithm. Also according to the quality of reconstructed images, results show that the proposed algorithm can achieve much better quality of reconstruction compared to AMSGrad algorithm.

By diversifying the two algorithms, NADAM and AMSGrad, in the way we used to design our NADA-p algorithm, the resulting optimizer exhibits a very interesting feature, by inheriting the speed of convergence from AMSGrad, alongside with the better quality of image reconstruction from NADAM. And even though being an algorithm that follows an "average" behavior from NADAM and AMSGrad, the resulting performance of NADA-p, as a hybrid mixed algorithm, is significantly better than the average performance of the mix components.

In future work, we aim to apply the diversified algorithm to clinical data to evaluate its performance and accuracy in an actual data set.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

- [1] T.J. Huppert, M.A. Franceschini, D.A. Boas, Noninvasive imaging of cerebral activation with diffuse optical tomography. In: R.D. Frostig, editor. *In Vivo Optical Imaging of Brain Function*. 2nd edition. Boca Raton (FL): CRC Press/Taylor & Francis; 2009.
- [2] A. Gibson, H. Dehghani, Diffuse optical imaging, *Phil. Trans. R. Soc. A.* 367 (2009) 3055–3072.
- [3] C.P. Valdes, H.M. Varma, A.K. Kristoffersen, T. Dragojevic, J.P. Culver, T. Durduran, Speckle contrast optical spectroscopy, a non-invasive, diffuse optical method for measuring microvascular blood flow in tissue, *Biomed. Opt. Express.* 5 (2014) 2769-2784.
- [4] S. Zhu, M. Wang, Y. Bai, X. Cao, X. Chen, M. Xu, H. Wu, Feasibility study of silicon photomultiplier based frequency domain diffuse optical tomography, in: H. Bosmans, G.-H. Chen, T. Gilat Schmidt (Eds.), *Medical Imaging 2019: Physics of Medical Imaging*, SPIE, San Diego, United States, 2019: p. 140.
- [5] D.A. Boas, D.H. Brooks, E.L. Miller, C.A. DiMarzio, M. Kilmer, R.J. Gaudette, Quan Zhang, Imaging the body with diffuse optical tomography, *IEEE Signal Process. Mag.* 18 (2001) 57–75.
- [6] B.W. Pogue, K.D. Paulsen, C. Abele, H. Kaufman, Calibration of near-infrared frequency-domain tissue spectroscopy for absolute absorption coefficient quantitation in neonatal head-simulating phantoms. *J. Biomed. Optics*, 5(2) (2000), 185-193.
- [7] S.R. Arridge, Optical tomography in medical imaging, *Inverse Problems*. 15 (1999) R41–R93.
- [8] J.M. Cochran, D.R. Busch, L. Lin, D.L. Minkoff, M. Schweiger, S. Arridge, A.G. Yodh, Hybrid time-domain and continuous-wave diffuse optical tomography instrument with concurrent, clinical magnetic resonance imaging for breast cancer imaging. *J. Biomed. Optics*, 24(5) (2019), 051409.
- [9] X. Dai, T. Zhang, H. Yang, J. Tang, P.R. Carney, H. Jiang, Fast noninvasive functional diffuse optical tomography for brain imaging. *J. Biophotonics*, 11(3) (2018), e201600267.
- [10] S. Li, K. Huang, M. Zhang, K.S. Uddin, Q. Zhu, Effect and correction of optode coupling errors in breast imaging using diffuse optical tomography. *Biomed. Optics Express*, 12(2) (2021), 689-704.
- [11] H. Yang, H. Sun, X. Wei, T. Shi, H. Jiang, Impact of transmitted and reflected lights on medical imaging in a parallel-plane diffuse optical tomography system for breast cancer detection, in: *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, OSA, Hollywood, Florida, 2018: p. OTh4D.4.
- [12] A.G. Yodh, Diffuse optical monitoring of brain and breast, in: *CLEO Pacific Rim Conference*, OSA, Hong Kong, 2018: p. Tu2K.1. <https://doi.org/10.1364/CLEOPR.2018.Tu2K.1>.
- [13] S.R. Arridge, J.C. Schotland, Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12) (2009), 123010.
- [14] S. Wright, M. Schweiger, S.R. Arridge, Reconstruction in optical tomography using the P N approximations, *Meas. Sci. Technol.* 18 (2007), 79–86.

- [15] Y. Zhao, A. Raghuram, H.K. Kim, A.H. Hielscher, J.T. Robinson, A. Veeraraghavan, High resolution, deep imaging using confocal time-of-flight diffuse optical tomography, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (2021), 2206–2219.
- [16] M.B. Applegate, R.E. Istfan, S. Spink, A. Tank, D. Roblyer, Recent advances in high speed diffuse optical imaging in biomedicine, *APL Photonics.* 5 (2020) 040802.
- [17] N. Chakhim, M. Louzar, A. Lamnii, M. Alaoui, Image reconstruction in diffuse optical tomography using adaptive moment gradient based optimizers: a statistical study. *Appl. Sci.* 10(24) (2020), 9117.
- [18] T. Dozat, Incorporating nesterov momentum into adam. In: Proceedings of the ICLR 2016 Workshop Track. (2016).
- [19] P.T. Tran, On the convergence proof of amsgrad and a new version. *IEEE Access,* 7 (2019), 61706-61716.
- [20] S. Ruder, An overview of gradient descent optimization algorithms, ArXiv:1609.04747 [Cs]. (2017).
- [21] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, IEEE, Pacific Grove, CA, USA, 2003: pp. 1398–1402.
- [22] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, IEEE, Istanbul, Turkey, 2010: pp. 2366–2369.
- [23] M. Schweiger, S.R. Arridge, The Toast++ software suite for forward and inverse modeling in optical tomography. *J. Biomed. Optics,* 19(4) (2014), 040801.
- [24] S.R. Arridge, M. Schweiger, M. Hiraoka, D.T. Delpy, A finite element approach for modeling photon transport in tissue. *Med. Phys.* 20(2) (1993), 299-309.
- [25] D.L. Colton, R. Kress, Inverse acoustic and electromagnetic scattering theory, (Vol. 93, pp. xii+-334). (1998). Springer. Berlin.
- [26] J. Verzani, Getting started with RStudio: An integrated development environment for R. Sebastopol, CA: O'Reilly Media. (2011).
- [27] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Assoc.* 47(260) (1952), 583-621.