An Efficient Quasi-Newton Method for Nonlinear Inverse Problems via Learned Singular Values

Danny Smyl, Tyler N. Tallman, Dong Liu, *Senior Member, IEEE*, and Andreas Hauptmann, *Member, IEEE*

***Abstract*—Solving complex optimization problems in engineering and the physical sciences requires repetitive computation of multi-dimensional function derivatives, which commonly require computationally-demanding numerical differentiation such as perturbation techniques. In particular, Gauss-Newton methods are used for nonlinear inverse problems that require iterative updates to be computed from the Jacobian and allow for flexible incorporation of prior knowledge. Computationally more efficient alternatives are Quasi-Newton methods, where the repeated computation of the Jacobian is replaced by an approximate update, but unfortunately are often too restrictive for highly ill-posed problems. To overcome this limitation, we present a highly efficient data-driven Quasi-Newton method applicable to nonlinear inverse problems, by using the singular value decomposition and learning a mapping from model outputs to the singular values to compute the updated Jacobian. Enabling time-critical applications and allowing for flexible incorporation of prior knowledge necessary to solve ill-posed problems. We present results for the highly non-linear inverse problem of electrical impedance tomography with experimental data.**

***Index Terms*—Quasi-Newton method, nonlinear inverse problems, neural networks, electrical impedance tomography.**

I. Introduction

HISTORICALLY, derivative-based optimization is at the heart of, and required for solving, complex systems in engineering, applied sciences, and applied mathematics [1], [2]. The computation of suitable update equations can be particularly problematic, when the system equations are computationally expensive to evaluate. This is especially a problem, when solving (large) nonlinear inverse problems of the form

$$g = \mathcal{A}(f) + \delta g, \tag{1}$$

with a nonlinear model $\mathcal{A}$, typically given by a finite element solver. Given data g with possible noise δg, we can then compute a solution f minimizing a suitable penalty functional

$$\mathcal{L}(f) = \|g - \mathcal{A}(f)\|_2^2 + R(f). \tag{2}$$

The first data-fidelty term ensures closeness of solutions to the measurements and the second term incorporates necessary regularization to counteract the ill-posedness and stabilize the reconstruction procedure as well as incorporate prior information. For nonlinear problems, second-order methods are typically employed to compute minimizers of eq. (2) due to fast (quadratic) convergence. However, given potentially expensive evaluations of the forward model $\mathcal{A}$ the computation of second order derivatives is undesirable. Thus, approximations such as the Gauss-Newton method are used, where only first order derivatives are required in each iteration, collected in the Jacobian $\mathbf{J} = \frac{\partial \mathcal{A}(f)}{\partial f}$. We can then form an approximate Hessian $\mathbf{H} = \mathbf{J}^T\mathbf{J}$, characterizing the GN approach. Yet, if no efficient/accurate analytical or semi-analytical algorithms are available, computing $\mathbf{J}$ is often problematic, requiring either (a) numerical differentiation or (b) updating approaches.

Both, (a) and (b), have advantages and disadvantages. Numerical differentiation approaches, namely the perturbation method, can indeed yield reliable derivative approximations. However, perturbation methods require repetitive function evaluations which can be extremely demanding. Therefore, when semi-analytical gradients are not available, conventional numerical differentiation approaches may be infeasible in solving high-dimensional nonlinear inverse problems - if not intractable on accessible computing resources [3], [4]. To address this challenge, a number of researchers have proposed methods for updating gradient terms, rather than, e.g., computing finite differences at each iteration; these approaches are referred to as Quasi-Newton (QN) methods. Possibly the most classical QN derivative approximator utilizes Broyden's method proposed in 1965 [5], taking advantage of the secant method. Since then, a number of successful updating methods have been proposed including the ever popular Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [6]–[9], the limited memory BFGS algorithm [10], and a suite of other QN-family derivative approximators [11]. Nevertheless, incorporation of prior information to regularize and stabilize the reconstruction for ill-posed problems is often either limited or not possible for QN methods [4], [12], thus effectively limiting applicability in practice. This is exacerbated in the case of nonlinear problems, where prior information is essential to stabilize reconstructions.

Manuscript received January 12, 2021; revised February 26, 2021; accepted February 26, 2021. Date of publication March 3, 2021; date of current version April 28, 2021. This work was supported in part by the Academy of Finland under Grant 336796 and Grant 334817, and in part by the CMIC-EPSRC platform under Grant EP/M020533/1. The work of DL was supported by the National Natural Science Foundation of China under Grant 61871356. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shiwen He. *(Corresponding author: Andreas Hauptmann.)*

Danny Smyl is with the Department of Civil and Structural Engineering, University of Sheffield, Sheffield, U.K. (e-mail: d.smyl@sheffield.ac.uk).

Tyler N. Tallman is with the School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN USA (e-mail: ttallman@purdue.edu).

Dong Liu is with the Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China (e-mail: dong2016@ustc.edu.cn).

Andreas Hauptmann is with the Research Unit of Mathematical Sciences, University of Oulu, 90570 Oulu, Finland, and also with the Department of Computer Science, University College London, London WC1E 6BT, U.K. (e-mail: andreas.hauptmann@oulu.fi).

Digital Object Identifier 10.1109/LSP.2021.3063622

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Motivated by recent advances of data-driven methods for nonlinear inverse problems [13]–[24] we propose a new QN approach for updating $\mathbf{J}$ based on the singular value decomposition (SVD). First we learn a representative decomposition of the Jacobian from a randomly generated set of training data. We then train a neural network to predict the associated singular values from the model outputs, making it possible to compute an approximate Jacobian without the need for numerical differentiation, resulting in the proposed neural network augmented Quasi-Newton method (NN-QN). In related studies, researchers have proposed learned SVD frameworks in applications to regularized inverse problems [25]–[27] We will first detail the mathematics and framework for the NN-QN approach in Section II. We then present an application to the highly nonlinear inverse problem of electrical impedance tomography in Section III. In the following, the approach will be tested with experimental data in Section IV. Lastly, conclusions will be provided.

II. A Neural Network Augmented Quasi Newton Method

A. Implementing an Approximate Jacobian

Computing minimizers of $\mathcal{L}(f)$ in eq. (2) requires an iterative scheme

$$f_{k+1} = f_k + \lambda_k \Delta f_k, \tag{3}$$

where the search direction Δf_k needs to be computed as well as a suitable step-length $\lambda_k \in (0, 2]$ [28].

Let us first discuss the unconstrained case without a regularization term, i.e. we are solving only the least squares problem $\|g - \mathcal{A}(f)|_2^2$. When using the GN method, we are required to compute $\mathbf{J}$ at each iteration in order to attain an unregularized update $\Delta f_k = \mathbf{H}^{-1}\mathbf{J}^T(g - \mathcal{A}(f_k))$. It is worth mentioning that a defining characteristic of the GN method is that higher order terms are ignored when computing the Hessian $\mathbf{H} = \mathbf{J}^T\mathbf{J}$. On the other hand, QN methods generally aim to either approximate/update (i) $\mathbf{H}$ or $\mathbf{H}^{-1}$ using gradient information or (ii) $\mathbf{J}$ and use the approximation of the type $\mathbf{H} = \mathbf{J}^T\mathbf{J}$ [29], [30]. Moreover, when statistical noise information is available, we may include noise weighting in the search direction via a weight matrix $\mathbf{W}$.

In this work, we will utilize option (ii) and consider the regularized problem in eq. (2) which is required for ill-posed problems. Given this choice, we obtain the following weighted description for the update

$$\Delta f_k = (\mathbf{J}^T\mathbf{W}\mathbf{J} + \Gamma_R)^{-1}(\mathbf{J}^T\mathbf{W}(g - \mathcal{A}(f_k)) - \partial R) \tag{4}$$

where Γ_R and ∂R correspond to the regularization and gradient of the regularization term [31].

In our proposed QN method we utilize the SVD for computing the Jacobian,

$$\mathbf{J} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{5}$$

with $\mathbf{U}$ and $\mathbf{V}^T$ containing the left and the right singular vectors, respectively, and $\mathbf{S}$ is a diagonal matrix containing the singular values [32]. The crucial assumption for our QN method is that we herein assume that the matrices $\mathbf{U}$ and $\mathbf{V}^T$ can be reasonably approximated, given a suitable initialization. That is, given an initial f_0 we compute the SVD of the corresponding Jacobian i.e. $\mathbf{J}(f_0) = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^T$. We then fix $\mathbf{V} = \mathbf{V}_0$ and $\mathbf{U} = \mathbf{U}_0$ for any approximation of $\mathbf{J}$ during the iterative process of eq. (3). Since $\mathbf{S}$ contains the non-negative singular values, it centrally influences the estimation of $\mathbf{J}$ and we aim to estimate the updated Jacobian given the fixed singular vectors as

$$\mathbf{J} \approx \mathbf{U}_0\mathbf{S}\mathbf{V}_0^T. \tag{6}$$

The challenge is now to obtain a good estimate of the singular values in $\mathbf{S}$ in order to compute the Jacobian (the target quantity) for the QN update. To do this, we propose the use of a neural network Λ to find an approximate set of singular values $\mathbf{S}_\Lambda \approx \mathbf{S}$ and obtain the learned estimate for $\mathbf{J}$ as

$$\mathbf{J}_\Lambda = \mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T. \tag{7}$$

The obtained approximation to the Jacobian can now be readily employed to compute the search direction in eq. (3) and forms the basis for the learned QN method proposed here. In the next subsection, we will present the neural network (NN) used for predicting and approximating $\mathbf{S}_\Lambda$.

B. Learned Prediction of the Singular Values

The keystone to the viability of the NN augmented Jacobian approximation, as proposed here, is an accurate and reliable prediction of the singular values in $\mathbf{S}_\Lambda$. We note, that $\mathbf{S}_\Lambda$ is a diagonal matrix and hence we are only left with estimating the vector of singular values $\mathrm{diag}(\mathbf{S}_\Lambda)$, simplifying the implementation of a learned estimator. What remains unclear at present is the selection of the input to the network Λ to predict $\mathrm{diag}(\mathbf{S}_A)$. An obvious choice is the model output itself $\mathcal{A}(f)$, since $\mathbf{J}$ is a derivative of $\mathcal{A}(f)$. While other choices may be viable, for simplicity we choose this mapping and aim to predict

$$\Lambda(\mathcal{A}(f)) = \mathrm{diag}(\mathbf{S}_\Lambda) \tag{8}$$

where Λ denotes the functional mapping between NN inputs and predictions. We can then readily form the estimated singular value matrix S_Λ. The proposed neural network augmented Quasi-Newton method (NN-QN) can then be obtained by simply substituting eq. (7) into the update in eq. (4):

$$\Delta f_k = ((\mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T)^T\mathbf{W}(\mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T) + \Gamma_R)^{-1} \\ ((\mathbf{U}_0\mathbf{S}_A\mathbf{V}_0^T)^T\mathbf{W}(g - \mathcal{A}(f_k)) - \partial R). \tag{9}$$

The major obvious advantage to using the NN-QN method is computational speed, since $\mathcal{A}(f)$ need not be repetitively computed to form the Jacobian after initialization. We also note that Jacobian updates are not susceptible to accumulating round-off errors present in other QN gradient estimators [33], the NN-QN method largely depends on the quality of U_0 and V_0 estimates. Additionally, it can be expected that $\mathbf{S}_\Lambda$ predictions are primarily valid within the space of the training data; but, as we will show in the following, highly general random fields can be successfully used for network training. Lastly, we present a summarizing pseudocode of the workflow in Algorithm 1 as a reference for solving nonlinear inverse problems using the proposed NN-QN method.

Remark: In case no regularization is required, the NN-QN method also holds for unregularized weighted solutions by using $\Delta f_k = ((\mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T)^T\mathbf{W}(\mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T))^{-1}(\mathbf{U}_0\mathbf{S}_\Lambda\mathbf{V}_0^T)^T\mathbf{W}(g - \mathcal{A}(f_k))$. If no weighting is used and assuming $\mathbf{J}_\Lambda^T\mathbf{J}_\Lambda$ is invertable, we can represent the update in the reduced form $\Delta f_k = (\mathbf{V}_0\mathbf{S}_\Lambda^{-1}\mathbf{U}_0^T)(g - \mathcal{A}(f_k))$.

Algorithm 1: NN Augmented QN Workflow.

Given trained estimator $\Lambda(\mathcal{A}(f)) = \mathrm{diag}(\mathbf{S}_\Lambda)$
$f_0 \leftarrow$ Initial estimate
Compute $\mathbf{J}(f_0) = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^T$
function NN-QN$(f_0, \mathbf{U}_0, \mathbf{V_0}, \mathbf{\Lambda}, \mathbf{k} = \mathbf{0})$
  while Stopping criterion $>$ tolerance **do**
    Compute Γ_R, ∂R, $\mathcal{A}(f_k)$
    $\mathrm{diag}(\mathbf{S}_\Lambda) \leftarrow \Lambda(\mathcal{A}(f_k))$
    $\Delta f_k \leftarrow$ Solve eq. (9)
    Use linesearch to compute λ_k
      $f_{k+1} \leftarrow f_k + \lambda_k \Delta f_k$
      $k \leftarrow k + 1$
  end while
end function (**return** $f_{\text{rec}} = f_k$)

III. Neural Network Augmented Quasi-Newton Method for Nonlinear Inverse Problems

A. Electrical Impedance Tomography

Herein we demonstrate the proposed NN-QN method using Electrical Impedance Tomography (EIT) [34], [35], a nonlinear inverse problem aiming to estimate the conductivity σ provided voltage data g. Due to notational convention we write σ instead of f for the unknown. The resultant observation model for EIT is then written as $g = \mathcal{A}(\sigma) + \delta g$ where δg is a noise term and $\mathcal{A}(\sigma)$ is the nonlinear forward model characterized by the Complete Electrode Model [36], [37] and solved using finite elements. In addition to the nonlinear relationship between σ and g, EIT is ill-posed and susceptible to measurement noise. As such, regularization is essential to attain suitable reconstructions; to do this, we aim to minimize the following functional characterizing absolute EIT imaging

$$\mathcal{L}_{\text{EIT}}(\sigma) = ||\mathbf{L}_e(g - \mathcal{A}(\sigma))||_2^2 + R(\sigma) \tag{10}$$

where $\mathbf{L}_e^T\mathbf{L}_e = \mathbf{W}^{-1}$ and $R(\sigma)$ is the regularization term. Lastly, we require a linearization point from which to compute the left and right singular vector matrices $\mathbf{U}_0$ and $\mathbf{V}_0^T$. For this, we simply utilize $\sigma_0 = \sigma_{\text{exp}}$ where σ_{exp} is a constant homogeneous estimate based on the expected conductivity, providing a good estimate and general starting point.

In the experimental EIT part, data from water tank (non-conductive inclusion localization) and carbon black-modified glass fiber/epoxy laminate (damage localization) tests are considered. In both cases, 256 measurements are collected using adjacent excitation and measurement following [38], [39]. Geometries for the water tank and laminate tests are circular (radius = 14 cm) and square (width = 9.54 cm), respectively. These circular and square geometries have 16 electrodes and are discretized using quadratic triangular finite elements meshes consisting of $N_{\text{el}} = 2034$ and $N_{\text{el}} = 2528$, respectively. To attain reliable electrode contact impedances for these finite element models, the best homogeneous estimate was used. Lastly, in order to demonstrate EIT reconstruction generalizability using the proposed NN-QN method, two different regularization forms are used. To accomplish this, we use total variation regularization and Laplacian regularization for the water tank and laminate tests, respectively.

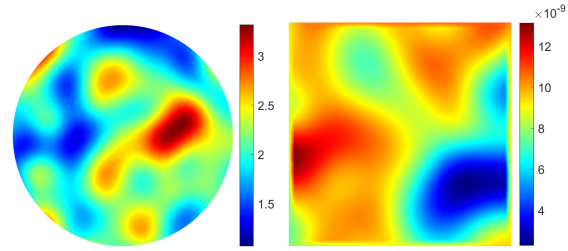

Fig. 1. Conductivity samples used in generating training data: left, random sample for the water tank geometry (mS/cm); right, random sample used for the laminate geometry (S/m).

B. Neural Network Implementation

The NN architectures used herein employ the same structure, since data sizes for both example problems are identical. Firstly, we note that the input and output data vectors are of size $\mathbb{R}^{256}$, because the number of singular values is equal to the number of model outputs for the test examples. Given this, we select a fully connected regression network consisting of three hidden layers composed of 300 neurons each. The first two layers utilize LeakyReLU activation functions with a negative region slope of $\alpha = 0.1$ while the last hidden layer uses ReLU as activation to ensure non-negative singular values.

To train the networks, $N = 5000$ randomized training and $N_t = 1000$ validation data sets are used. The randomized conductivity training samples σ_{rand} were generated using a spatially-weighted Gaussian kernel $\mathbf{G}$ perturbed with a random vector $\mathbf{r}$, i.e. $\sigma_{\text{rand}} = \mathbf{G}^{-1}\mathbf{r} + \sigma_{\text{exp}}$. We note that this is highly general training data, where we ensure the conductivity mean is centered at σ_{exp} while the range of the random samples varies within two standard deviations of the mean. To illustrate this, Fig. 1 shows samples for both EIT geometries. In optimizing the network using adaptive learning rates (with initial 10^{-5}), the following loss function is minimized

$$L = \frac{1}{N}\sum_{n=1}^{N}(\mathrm{diag}(\mathbf{S}^n) - \mathrm{diag}(\mathbf{S}_\Lambda^n))^2 + \kappa\|w\|_2^2, \tag{11}$$

where $\mathrm{diag}(\mathbf{S}^n)$ and $\mathrm{diag}(\mathbf{S}_\Lambda^n)$ represent the nth vector of singular values in the training set and output from the NN, w are the network weights, and κ is the regularization hyperparameter. For regularizing the network during training, $\kappa = 0.001$ and a dropout rate between layers of 20% are used. To prevent overfitting, training is terminated when the change in the validation loss was less than 10^{-2} between epochs. Summarily, learning singular values requires (i) generating model output and singular value samples, (ii) prescribing NN architecture/parameters, and finally (iii) minimizing eq. (11).

IV. Experimental Results

In this section we compare absolute EIT reconstructions from experimental water tank and laminate data as shown in Fig. 2. These reconstructions are computed using (i) the GN method using an accurate perturbed $\mathbf{J}$, (ii) the proposed NN-QN method for computing $\mathbf{J}$, and (iii) Broyden's classical method for computing $\mathbf{J}$. We note that Broyden's method was selected as it readily

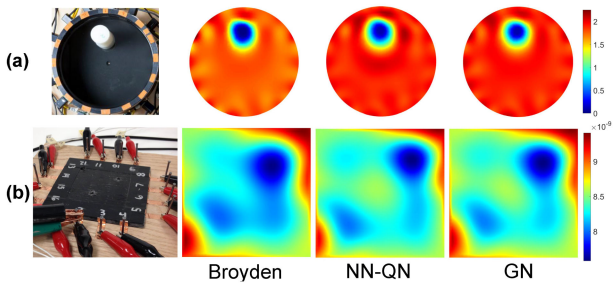

Fig. 2. EIT reconstructions using (a) water tank data and (b) laminate data. Images in row 1 are experiment photographs while rows 2-4 are computed using the labeled optimization approaches. Colorbar units are mS/cm and S/m for watertank and laminate reconstructions, respectively.

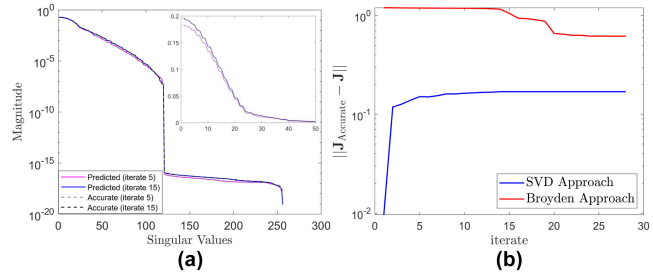

Fig. 3. Representative verification using EIT water-tank data: (a) Comparison of accurate and predicted singular values at iterates 5 and 15; (b) iterative accuracy of the obtained Jacobians compared with Broyden's approach.

TABLE I
COMPUTATIONAL DEMAND (AFTER INITIALIZATION)

Method	Mean computing time (s)	Mean Iterations (K)
NN-QN	5.22	30
Broyden	4.93	29.5
GN	1858	10.5

supports the integration of $\mathbf{W}$ in the inversion. Additionally, the perturbation method was selected for GN optimization to illustrate computing demands when a semi-analytical algorithm for calculating $\mathbf{J}$ is not available (as is often the case).

Water tank reconstructions in Fig. 2(a) all show the localization of the non-conductive inclusion well. In fact, by visual inspection, NN-QN produces the fewest artifacts near the electrodes but slightly more near the target. We may conclude that NN-QN is comparable to GN, even though GN is based on the more accurate $\mathbf{J}$ and $\mathbf{H}$. Meanwhile, it is also apparent that the background water conductivity estimated using Broyden's method has significant fluctuations near the electrodes and is significantly lower than the accurate GN estimate. This observation is likely owed to roundoff errors accumulating in successive iterations, consistent with findings in [4]. On the other hand, such roundoff errors do not occur in the proposed NN-QN method, resulting in a more accurate estimation of the background conductivity.

Images reported in Fig. 2(b) endeavor to quantitatively localize two through-holes in the composite laminate. These absolute images show that the holes are visible in all cases, although the damage is clearly blurred using Broyden's method. Due to the small size of the holes relative to the imaged domain and the diffusive nature of EIT, it is clear that the presence of errors/noise in $\mathbf{J}$ and $\mathbf{H}$ resulting from the use of Broyden's method significantly impacts the ability of EIT to detect highly localized changes in conductivity. Comparing the proposed QN-NN and GN reconstructions, the drop in conductivity observed near the holes is visually comparable – although the NN-QN reconstruction is somewhat smoother relative to the GN reconstruction. Note that the conductivity of carbon filler-modified polymers is known to vary appreciably even for a set filler concentration [40]. Hence, conductivity fluctuations away from the holes are expected in Fig. 2(b).

A table is provided (Table I), reporting key variables associated with the computational demand of reconstructions presented. Owing to the fact that the GN method used forward differencing to compute $\mathbf{J}$, GN reconstructions took approximately 100-200x longer ($\approx$ 15 minutes) to reach an equal stopping criteria ($||f_{k+1} - f_k||_2^2/||f_k||_2^2 \leq 10^{-2}$) than the Broyden- or NN-based approaches ($\approx$ 5 seconds). We point out that the computation times are after initialization. Overall, NN-QN and Broyden's method exhibited similar minimization behavior, requiring less time to reach a minimum but requiring more iterations than GN.

We illustrate the effectiveness of singular value predictions, relative to those obtained from an accurate Jacobian $\mathbf{J}_{\text{Accurate}}$, in Fig. 3(a) which compares closely. More importantly, the accurately predicted singular values contribute to a significantly improved approximation of $\mathbf{J}$ relative to the Broyden approach, as measured by the differences in matrix norms $||\mathbf{J}_{\text{Accurate}} - \mathbf{J}||_2$ shown in Fig. 3(b), which explains the improved reconstruction quality. Specifically, the higher accuracy in early iterates may be beneficial for the final reconstruction quality.

V. CONCLUSION

In this work, we presented a new data driven approach to the Quasi-Newton method intended for application to non-linear inverse problems. The method uses a neural network to map non-linear numerical model outputs to the singular values of the Jacobian matrix, which can be used in forming an approximation of the (otherwise costly) Jacobian. By doing this, the method falls within a subset of Quasi-Newton methods.

Preliminary, the efficacy of the proposed NN-QN method has been demonstrated in the context of an ill-posed non-linear inverse problem - EIT. It was shown that the use of different priors may be effectively incorporated into the approach. While these results showed that the NN-QN method may improve reconstructions relative to the use of other QN methods, specific further research is required to enable a more concrete evaluation. Lastly, it was found that application-specific prior information may be included in the data driven Jacobian estimate via the selection of training data. Whereas very general training data was used herein, more specific distributions may improve solutions to nonlinear inverse problems and is the source of future research.

REFERENCES

[1] A. D. Belegundu and T. R. Chandrupatla, *Optimization Concepts and Applications in Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[2] L. T. Biegler, Y.-D. Lang, and W. Lin, "Multi-scale optimization for process systems engineering," *Comput. Chem. Eng.*, vol. 60, pp. 17–30, 2014.

[3] A. K. Saibaba, T. Bakhos, and P. K. Kitanidis, "A flexible Krylov solver for shifted systems with application to oscillatory hydraulic tomography," *SIAM J. Sci. Comput.*, vol. 35, no. 6, pp. A3001–A3023, 2013.

[4] D. Smyl and D. Liu, "Less is often more: Applied inverse problems using HP-forward models," *J. Comput. Phys.*, vol. 399, 2019, Art no. 108949.

[5] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Math. Comput.*, vol. 19, no. 92, pp. 577–593, 1965.

[6] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA J. Appl. Math.*, vol. 6, no. 1, pp. 76–90, 1970.

[7] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, no. 3, pp. 317–322, 1970.

[8] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comput.*, vol. 24, no. 109, pp. 23–26, 1970.

[9] D. F. Shanno, "Conditioning of Quasi-Newton methods for function minimization," *Math. Comput.*, vol. 24, no. 111, pp. 647–656, 1970.

[10] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, nos. 1–3, pp. 503–528, 1989.

[11] A. S. Lewis and M. L. Overton, "Nonsmooth optimization via Quasi-Newton methods," *Math. Program.*, vol. 141, nos. 1/2, pp. 135–163, 2013.

[12] E. Haber, "Quasi-Newton methods for large-scale electromagnetic inverse problems," *Inverse Problems*, vol. 21, no. 1, 2004, Art. no. 305.

[13] S. Martin and C. T. M. Choi, "A post-processing method for three-dimensional electrical impedance tomography," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.

[14] J. S. Hamilton and A. Hauptmann, "Deep D-bar: Real-time electrical impedance tomography imaging with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 10, pp. 2367–2377, Oct. 2018.

[15] S. J. Hamilton, A. Hänninen, A. Hauptmann, and V. Kolehmainen, "Beltrami-Net: Domain-independent deep D-bar learning for absolute imaging with electrical impedance tomography (a-eit)," *Physiol. Meas.*, vol. 40, no. 7, 2019, Art no. 074002.

[16] D. Smyl, T. N. Tallman, J. A. Black, A. Hauptmann, and D. Liu, "Learning and correcting non-Gaussian model errors," *J. Comput. Phy.*, vol. 432, 2020, Art no. 110152.

[17] D. Smyl and D. Liu, "Optimizing electrode positions in 2D electrical impedance tomography using deep learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6030–6044, Sep. 2020.

[18] S. Arridge and A. Hauptmann, "Networks for nonlinear diffusion problems in imaging," *J. Math. Imag. Vis.*, vol. 62, pp. 471–487, 2020.

[19] Z. Lin *et al.*, "Neural network based supervised descent method for 2 D electrical impedance tomography," *Physiological Measurement*, vol. 41, no. 7 2020, Art. no. 074003.

[20] J. Yoo *et al.*, "Deep learning diffuse optical tomography," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 877–887, Apr. 2020.

[21] S. Siltanen and T. Ide, "Electrical impedance tomography, enclosure method and machine learning," in *Proc. IEEE 30th Int. Workshop Mach. Learn. Signal Process.*, 2020, pp. 1–6.

[22] T. Huuhtanen and A. Jung, "Anomaly location detection with electrical impedance tomography using multilayer perceptrons," in *Proc. IEEE 30th Int. Workshop Mach. Learn. Signal Process.*, 2020, pp. 1–6.

[23] J. P. Agnelli, A. Çöl, M. Lassas, R. Murthy, M. Santacesaria, and S. Siltanen, "Classification of stroke using neural networks in electrical impedance tomography," *Inverse Problems*, vol. 36, no. 11, 2020, Art. no. 115008.

[24] V. Candiani and M. Santacesaria, "Neural networks for classification of strokes in electrical impedance tomography on a 3D head model," 2020, *arXiv:2011.02852*.

[25] J. Schwab, S. Antholzer, and M. Haltmeier, "Big in Japan: Regularizing networks for solving inverse problems," *J. Math. Imag. Vis.*, vol. 62, pp. 1–11, 2019.

[26] J. Schwab, S. Antholzer, R. Nuster, G. Paltauf, and M. Haltmeier, "Deep learning of truncated singular values for limited view photoacoustic tomography," in *Proc. SPIE BiOS*, San Francisco, CA, USA, 2019, Art no. 1087836.

[27] Y. E. Boink and C. Brune, "Learned SVD: Solving inverse problems via hybrid autoencoding," 2019, Springer, New York, NY, USA, *arXiv:1912.10840*.

[28] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia, PA, USA: SIAM, 1996.

[29] R. M. Gower and P. Richtárik, "Randomized Quasi-Newton updates are linearly convergent matrix inversion algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 38, no. 4, pp. 1380–1409, 2017.

[30] M. J. Martınez, "Practical Quasi-Newton methods for solving nonlinear systems," *J. Comput. Appl. Math.*, vol. 124, nos. 1/2, pp. 97–121, 2000.

[31] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160. New York, NY, US: Springer, 2006.

[32] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Linear Algebra*. Berlin, Germany: Springer, 1971, pp. 134–151.

[33] Y. Ding, E. Lushi, and Q. Li, "Investigation of Quasi-Newton methods for unconstrained optimization," *Int. J. Comput. Appl.*, vol. 29, pp. 48–58, 2010.

[34] D. Liu, D. Smyl, D. Gu, and J. Du, "Shape-driven difference electrical impedance tomography," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3801–3812, Dec. 2020.

[35] D. Liu, D. Gu, D. Smyl, A. K. Khambampati, J. Deng, and J. Du, "Shape-driven EIT reconstruction using Fourier representations," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 481–490, Feb. 2021.

[36] E. Somersalo, M. Cheney, and D. Isaacson, "Existence and uniqueness for electrode models for electric current computed tomography," *SIAM J. Appl. Math.*, vol. 52, no. 4, pp. 1023–1040, 1992.

[37] P. J. Vauhkonen, M. Vauhkonen, T. Savolainen, and J. P. Kaipio, "Three-dimensional electrical impedance tomography based on the complete electrode model," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 9, pp. 1150–1160, Sep. 1999.

[38] A. Hauptmann, V. Kolehmainen, N. M. Mach, T. Savolainen, A. Seppänen, and S. Siltanen, "Open 2D electrical impedance tomography data archive," 2017, *arXiv:1704.01178*.

[39] T. N. Tallman, S. Gungor, K. W. Wang, and C. E. Bakis, "Damage detection via electrical impedance tomography in glass fiber/epoxy laminates with carbon black filler," *Struct. Health Monit.*, vol. 14, no. 1, pp. 100–109, 2015.

[40] S. Gungor and C. E. Bakis, "Anisotropic networking of carbon black in glass/epoxy composites using electric field," *J. Composite Mater.*, vol. 49, no. 5, pp. 535–544, 2015.