SUPPLEMENTARY MATERIALS: A Mini-Batch Quasi-Newton Proximal Method for Constrained Total Variation Nonlinear Image Reconstruction *

Tao Hong†, Thanh-an Pham‡, Irad Yavneh§, and Michael Unser¶

S.I. Accelerated Stochastic Proximal Method. For completeness, we summarize the accelerated stochastic proximal methods (ASPM) in Algorithm S.I.1. The TV proximal operator (S.I.1) is solved by the algorithm proposed in [SM1].

S.II. Lippmann-Schwinger Forward Model. For completeness, we described the details of the nonliner forward model $\mathcal{H}_l(\cdot)$ in this section in a continuous domain. A comprehensive description and proper discretization can be found in [SM3]. The notations used in this section is self-contained. Fig. S.1 presents the principle of optical diffraction tomograhy.

Denote by η_b and $\eta(\mathbf{r})$ the refractive indices of the background and the sample. Then the scattering potential is defined as $f(\mathbf{r}) = k_b^2(\eta(\mathbf{r})^2/\eta_b^2 - 1)$ where $k_b = 2\pi\eta_b/\lambda$ is the wavenumber and λ is the wavelength. Then, for the l-th incident plane wave $u_{\text{in}}^l(\mathbf{r})$, the total field $u(\mathbf{r})$ in Ω can be obtained by solving the Lippmann-Schwinger equation:

$$u^l(\mathbf{r}) = u_{\text{in}}^l(\mathbf{r}) + \int_\Omega g(\mathbf{r}-\mathbf{r}')f(\mathbf{r}')u(\mathbf{r}')^l d\mathbf{r}', \quad \mathbf{r}\in\Omega \tag{S.II.1}$$

where $g(\mathbf{r}) = \frac{e^{jk_b\|\mathbf{r}\|}}{4\pi\|\mathbf{r}\|}$ is the free-space Green function on $\mathbf{r}\in\Omega$. After obtaining $u^l(\mathbf{r})$ on $\mathbf{r}\in\Omega$, we model the wave propagation from Ω to the measurements plane Γ, *i.e.* the l-th measurements $y_l(\mathbf{r})$, with

$$y_l(\mathbf{r}) = u_{\text{in}}^l(\mathbf{r}) + \int_\Omega g(\mathbf{r}-\mathbf{r}')f(\mathbf{r}')u(\mathbf{r}')^l d\mathbf{r}', \quad \mathbf{r}\in\Gamma. \tag{S.II.2}$$

In practice, we applied the Bi-CGSTAB method [SM4] to solve (S.II.2).

S.III. Born Forward Model. The Born model is a linearized version of (S.II.2)

$$y_l(\mathbf{r}) = u_{\text{in}}^l(\mathbf{r}) + \int_\Omega g(\mathbf{r}-\mathbf{r}')f(\mathbf{r}')u_{\text{in}}^l(\mathbf{r}')^l d\mathbf{r}', \quad \mathbf{r}\in\Gamma, \tag{S.III.1}$$

where the total field in the integral is replaced by the incident field. This approximation assumes that the sample is weakly-scattering.

*Submitted to the editors DATE.

†Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA (tao.hong@austin.utexas.edu).

‡Biomedical Imaging Group, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland (thanh-an.pham@proton.me).

§Department of Computer Science, Technion-Israel Institute of Technology, Haifa, 3200003 Israel (irad@cs.technion.ac.il).

¶Biomedical Imaging Group, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland (michael.unser@epfl.ch).

Algorithm S.I.1 Accelerated stochastic proximal method.

Initialization: Initial guess $\mathbf{x}_0 \in \mathbb{R}^N$; tradeoff parameter λ; pre-defined sets $\{\mathscr{S}_s\}_{s=1}^S$ or choose the partial measurements randomly at each iteration; Lipschitz constant κ; maximal number of iterations Max_Iter

Output: $\mathbf{x}^*$

1: $k \leftarrow 1$
2: **for all** $k \leq$ Max_Iter **do**
3: $s \leftarrow \mod(k, S)$
4: Compute partial gradient $\mathbf{g}_k \leftarrow \nabla F_s(\mathbf{x}_{k-1})$
5: Solve the following problem to get $\mathbf{x}_k$

$$\mathbf{x}_k \leftarrow \arg\min_{\mathbf{x}\in\mathcal{C}} \langle \mathbf{g}_k, \mathbf{x} - \mathbf{x}_{k-1}\rangle + \frac{\kappa}{2}\|\mathbf{x} - \mathbf{x}_{k-1}\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}) \tag{S.I.1}$$

6: $k \leftarrow k + 1$
7: **end for**
8: **return** $\mathbf{x}^* \leftarrow \mathbf{x}_{\text{Max\_Iter}}$

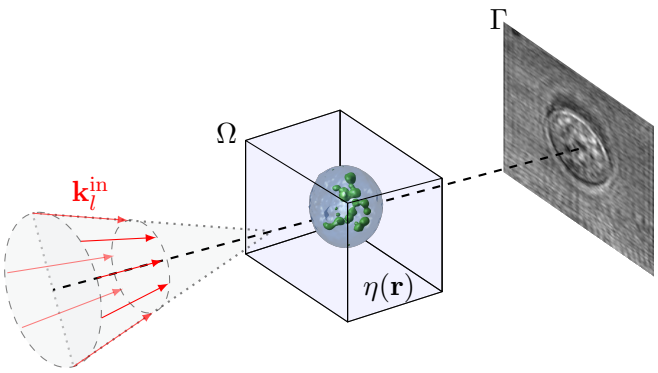

Figure S.1. *Principle of optical diffraction tomography. The arrows represent the wave vectors* $\{\mathbf{k}_l^{\text{in}} \in \mathbb{R}^3\}_{l=1}^L$ *of the* L *incident plane waves* $\{u_{\text{in}}^l\}_{l=1}^L$. *The angles of illumination are limited to a cone around the optical axis. The refractive-index map of the sample* $\eta(\mathbf{r})$ *is embedded in the domain* $\Omega \subset \mathbb{R}^3$, *and the recorded domain is denoted by* Γ. *The parameter* $\boldsymbol{r}$ *denotes the three-dimensional spatial coordinate.*

S.III.0.1. Linear Model–Weakly Scattering Sample. The tradeoff parameter $\lambda = 10/64^3$ was optimized by grid search and the stepsize in the ASPM was set to 0.1. Note that the stepsize was chosen to be the largest possible while still ensuring convergence. BQNPM and FBQNPM with the parameters $a_k = 1$ and $\gamma = 0.8$ performed well for our experiments. We set a total of $S = 4$ subsets ($\{\mathscr{S}_s\}_{s=1}^4$ with fifteen measurements each). Note that the fifteen illumination angles were equally spaced.[1] 100 iterations were performed to recover the RI maps. The Rytov approximation [SM2] was used as the initial guess for all competing methods.

[1]We also tried to choose the illumination angles randomly for ASPM but found that selecting them in equally spaced yielded slightly better performance.

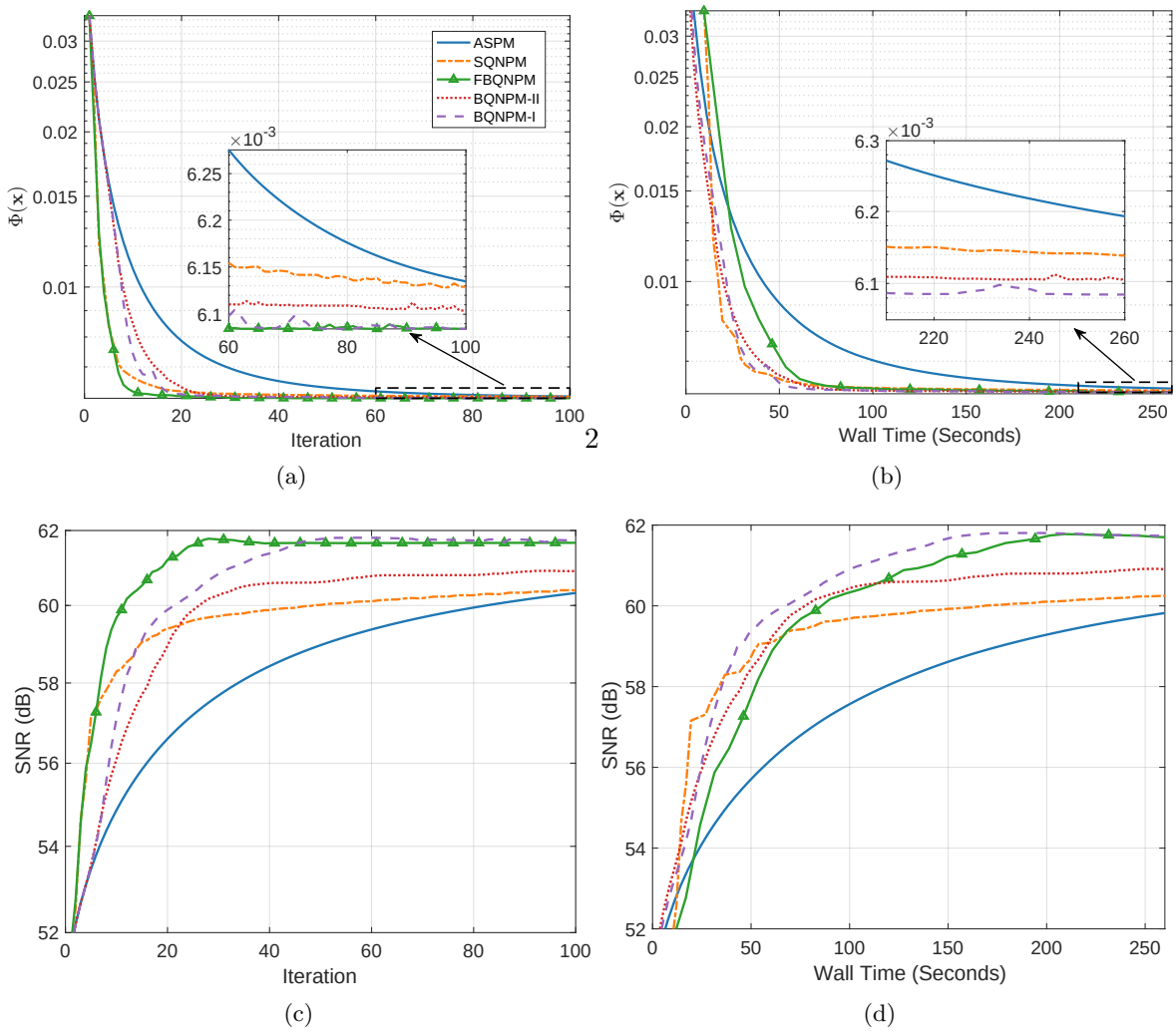

Figure S.2. *Performance of ASPM, SQNPM [SM5], BQNPM with strategies I and II, and FBQNPM algorithms on the weakly scattering simulated sample using the first-order Born approximation. From top to bottom rows: Full cost and SNR values versus iterations and wall time, respectively.*

The first row of Fig. S.2 presents the full cost with respect to the number of iterations and wall time for all methods. It is evident that BQNPM-I/II converged faster than ASPM in terms of iterations and wall time. The computational cost per iteration for BQNPM-I/II was similar to that of ASPM, which indicates that the computational overhead of WPM is negligible. At the beginning of iterations, SQNPM reached the lowest full cost among ASPM and BQNPM-I/II. However, BQNPM-I/II outpaced SQNPM at later iterations. Fig. SM2(b) shows that BQNPM-I/II converged as well as SQNPM in terms of wall time, even in the first iterations. This is because SQNPM requires computing the full gradient every S iterations, whereas BQNPM-I/II never compute the full gradient. Clearly, FBQNPM was the fastest algorithm in terms of iterations but it converged slower than BQNPM-I/II in terms of wall

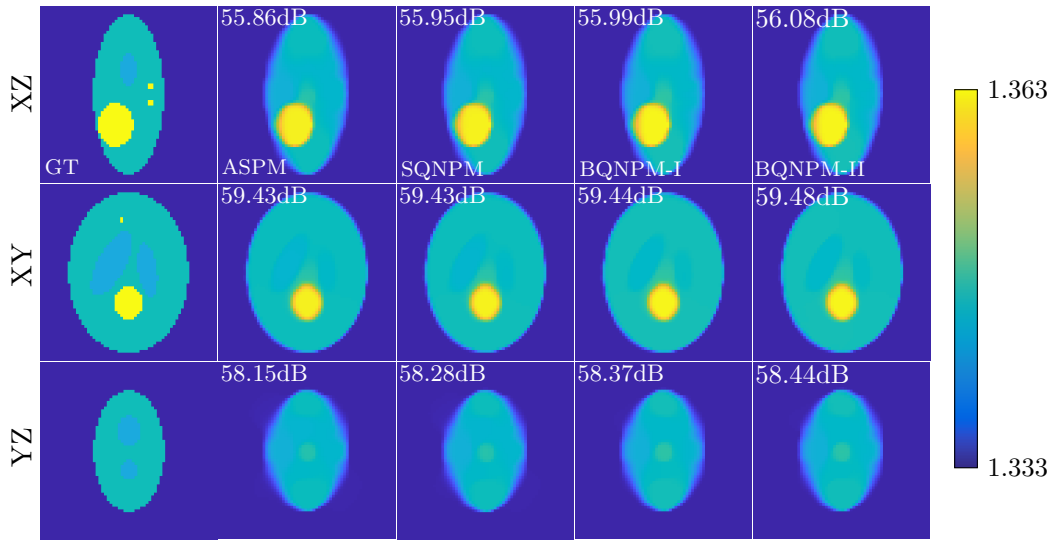

Figure S.3. *Orthoviews of the 3D refractive index maps obtained by ASPM (iter. $k = 100$), SQNPM [SM5] (iter. $k = 100$), BQNPM-I (iter. $k = 26$), and BQNPM-II (iter. $k = 35$) algorithms on the weakly scattering simulated sample using the first-order Born approximation. The SNR for each slice is displayed in the top-left corner of each image.*

time since FBQNPM requires computing the full gradient at each iteration.

Fig. S.2 shows that BQNPM-I converged faster than BQNPM-II in terms of iterations and wall time. Although our theoretical analysis does not demonstrate the superiority of BQNPM-I, we observed that BQNPM-I converged faster than BQNPM-II in the following experiments. This may be due to the fact that BQNPM-I processes all subsets over each S iterations, allowing it to utilize the most recent information, whereas BQNPM-II uniformly samples subsets at each iteration, potentially skipping some subsets during S iterations.

The second row of Fig. S.2 demonstrates that BQNPM-I/II required less wall time than ASPM and SQNPM to achieve the highest SNR, further corroborating our previous observations. Moreover, ASPM and SQNPM needed 100 iterations to reach its highest SNR while BQNPM-I/II required much less number of iterations to get a comparable SNR. Fig. S.3 displays the orthoviews of the RI maps obtained with ASPM, SQNPM, and BQNPM-I/II at the 100th, 100th, and 26th/35th iterations, respectively. Here, we observed that the first-order Born approximation was accurate for the weakly scattering sample.

S.IV. Linear Model–Strongly Scattering Sample. In this part, we studied the performance of ASPM, SQNPM [SM5], BQNPM-I/II, and FBQNPM on a strongly-scattering sample. In this experiment, the measurements were simulated with the Lippmann-Schwinger forward model, but the reconstructions were obtained by using the linear (Born) model. The tradeoff parameter $\lambda = 10/64^3$ was optimized by grid search and the stepsize was set at 0.1. We ran 100 iterations to recover the RI maps in this setting. Note that the shape of the

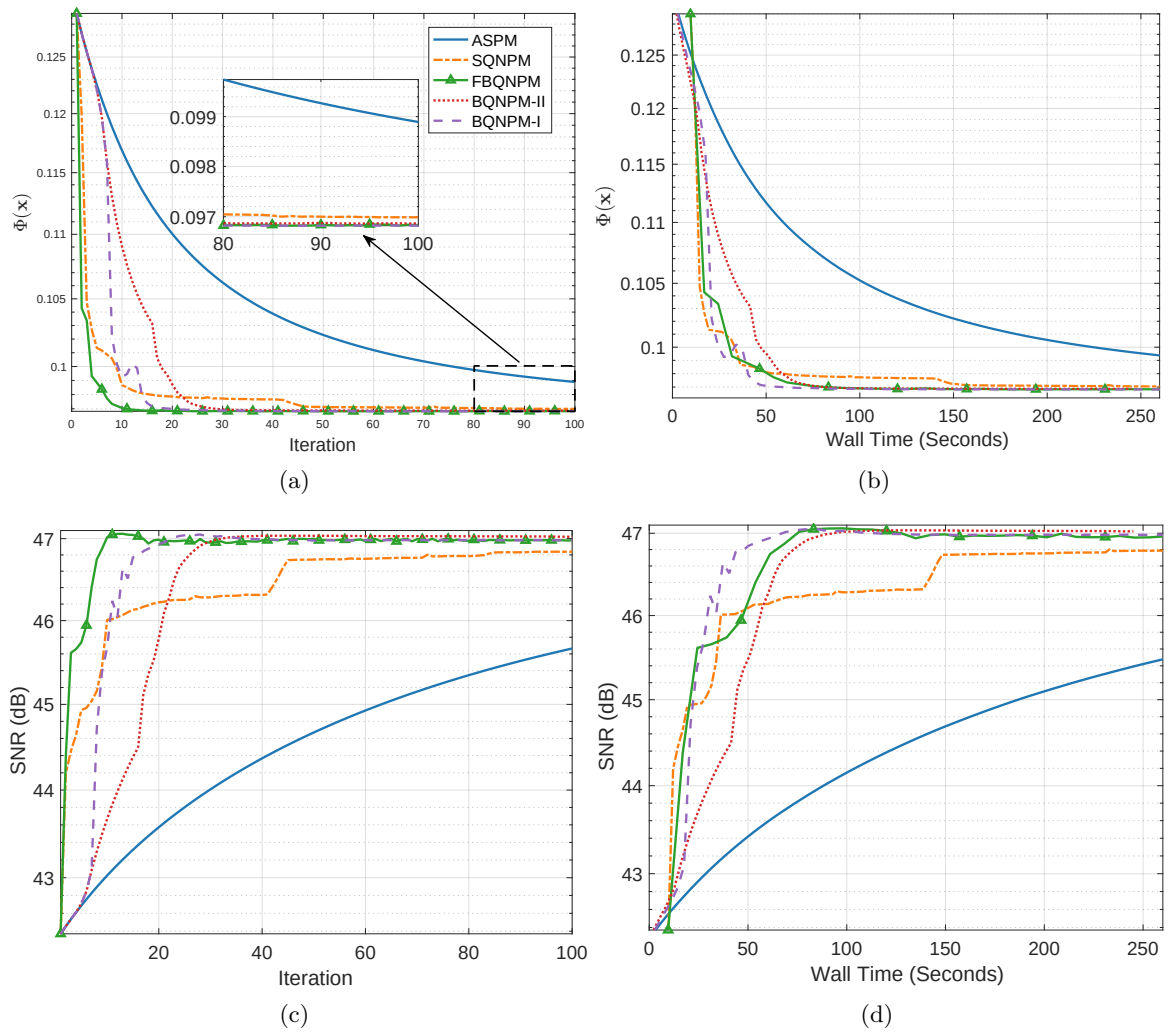

Figure S.4. *Performance of ASPM, SQNPM [SM5], and BQNPM algorithms on the strongly scattering simulated sample using the first-order Born approximation. From top to bottom rows: Full cost and SNR values versus iterations and wall time, respectively.*

phantom used here is identical to the one in Sec. V.A.1) of the main paper, but the maximal RI was set to a higher value (*i.e.*, 1.413). In this regime, the linear model was inaccurate in modeling light scattering.

The first and second rows of Fig. S.4 contain the full cost and SNR versus iterations and wall time of ASPM, SQNPM, BQNPM-I/II, and FBQNPM. These algorithms behaved similarly to the case of the weakly scattering sample. SQNPM converged faster than BQNPM in the first iterations, but BQNPM-I/II became faster at later iteration. Moreover, BQNPM-I/II converged faster than SQNPM in terms of wall time since BQNPM-I/II required less computation than SQNPM. Similar to the weakly scattering experiment described in the main paper, we observed that BQNPM-I converged faster than BQNPM-II in terms of both

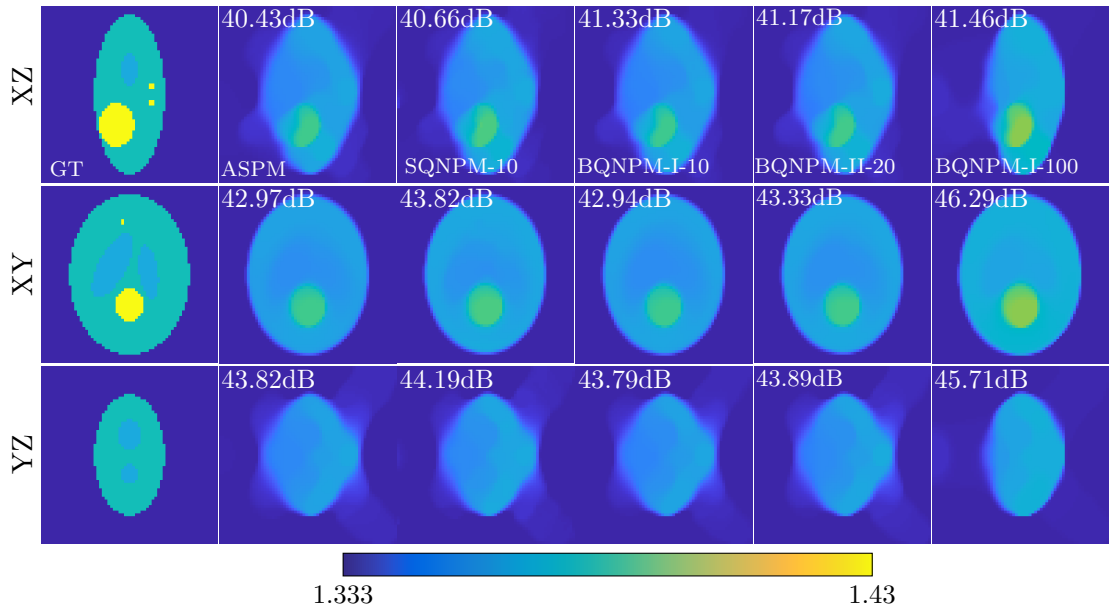

Figure S.5. *Orthoviews of the 3D refractive-index maps obtained by ASPM (iter. $k = 100$), SQNPM (iter. $k = 10$), BQNPM-I (iter. $k = 10$, and 100), and BQNPM-II (iter. $k = 20$) algorithms on the strongly scattering simulated sample using the first-order Born approximation. The SNR for each slice is displayed in the top-left corner of each image.*

iterations and wall time. Moreover, FBQNPM was the fastest algorithm in terms of iterations because it computed the full gradient at each iteration. Thus, FBQNPM lost its advantage in terms of wall time.

In this experiment, ASPM needed 100 iterations to reach its highest SNR, while SQNPM (BQNPM-I/II, respectively) only required 10 (10/20, respectively) iterations to get a comparable SNR. Fig. S.5 displays the orthoviews of the RI maps obtained using BQNPM-I/II, SQNPM, and ASPM at the 10th/20th, 10th, and 100th iterations, as well as the 100th iteration of BQNPM-I. These methods recovered similar images but fail to accurately reconstruct the RI map due to the inaccuracy of the first-order Born approximation.

REFERENCES

[SM1] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Transactions on Image Processing, 18 (2009), pp. 2419–2434.

[SM2] A. Devaney, *Inverse-scattering theory within the Rytov approximation*, Optics Letters, 6 (1981), pp. 374–376.

[SM3] T.-a. Pham, E. Soubies, A. Ayoub, J. Lim, D. Psaltis, and M. Unser, *Three-dimensional optical diffraction tomography with Lippmann-Schwinger model*, IEEE Transactions on Computational Imaging, 6 (2020), pp. 727–738.

[SM4] H. A. Van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing, 13 (1992), pp. 631–644.

[SM5] X. Wang, X. Wang, and Y.-X. Yuan, *Stochastic proximal quasi-Newton methods for non-convex composite optimization*, Optimization Methods and Software, 34 (2019), pp. 922–948.