

A Mini-Batch Quasi-Newton Proximal Method for Constrained Total-Variation Nonlinear Image Reconstruction

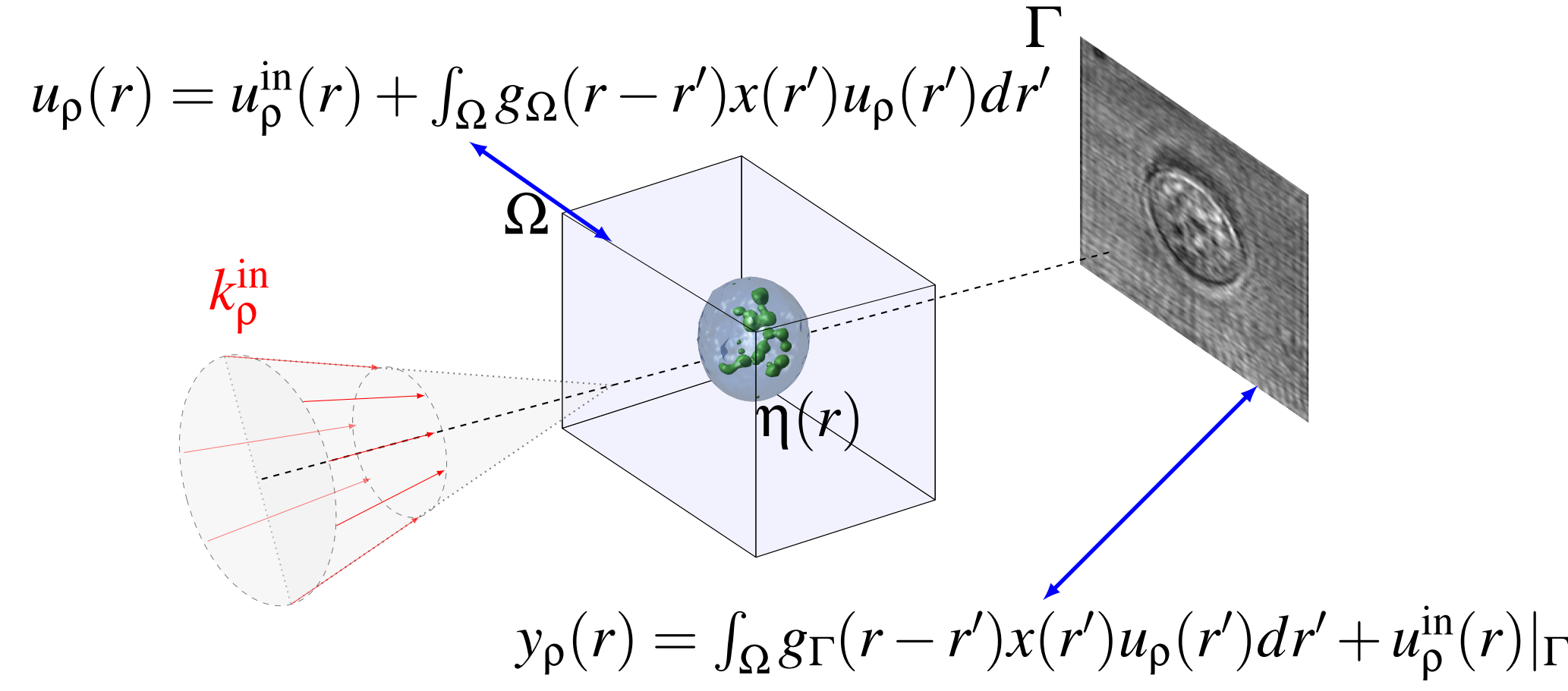


Tao Hong*, Thanh-an Pham**, Irad Yavneh#, and Michael Unserz###

*Functional MRI Lab, University of Michigan, Ann Arbor, **MIT USA, #Technion Israel, ###EPFL Switzerland

Problem formulation

Principle of optical diffraction tomography



- g Green's function of $\nabla^2 + k_0^2 \eta_0^2 I$ (Helmholtz equation)
- $x(r) = k_0^2(\eta(r)^2 - \eta_0^2)$ with $k_0 = \omega/c$
- **Goal:** recover $\eta(r)$ through $\{y_p(r)\}_p$

Composite minimization problem:

$$\min_{x \in \mathcal{C}} \Phi(x) \equiv \left(\frac{1}{L} \sum_{p=1}^L \underbrace{\frac{1}{2} \|\mathcal{H}_p(\mathbf{x}) - y_p\|_2^2}_{f_p} + \lambda \text{TV}(x) \right)$$

Features: ∇f_p expensive, f_p nonconvex, $\text{TV}(\cdot)$ nonsmooth, constrained convex set \mathcal{C}

Classical solver: accelerated stochastic proximal method (ASPM)

$$\begin{cases} x_k = \text{prox}_{a_k \lambda \text{TV}}(v_{k-1} - a_k \sum_{p \in \mathcal{S}_k} \nabla f_p(v_{k-1})) & \text{Dual} \\ v_k = x_k + c_k(x_k - x_{k-1}) \end{cases}$$

where $\text{prox}_{a_k \lambda \text{TV}}(x) = \arg \min_u \frac{1}{2} \|u - x\|_2^2 + a_k \lambda \text{TV}(u)$ and \mathcal{S}_k defines the chosen indices at k th iteration.

Proposed method

Our mini-batch quasi-Newton proximal Method:

1. Split the index set $\{1, 2, \dots, L\}$ into K_b subsets $\{\mathcal{S}_t\}_{t=1}^{K_b}$ and then we have

$$\min_{x \in \mathcal{C}} \left(\frac{1}{K_b} \sum_{t=1}^{K_b} F_t(x) + \lambda \text{TV}(x) \right)$$

- 2.

$$x_k = \arg \min_{x \in \mathcal{C}} \left(\sum_t \left(\left\langle \nabla F_{\kappa_{K_b}(k,t)}(x_{k-t}), x - x_{k-t} \right\rangle + \frac{1}{2a_k} (x - x_{k-t})^T B_{k-t}^{\kappa_{K_b}(k,t)} (x - x_{k-t}) \right) + K_b \lambda \text{TV}(x) \right), \quad (1)$$

where $\kappa_{K_b}(k, t) = \text{mod}(k - 1 - t, K_b) + 1$ and $B_{k-t}^{\kappa_{K_b}(k,t)}$ is the approximate Hessian (symmetric rank-1) of F_t at k th iteration.

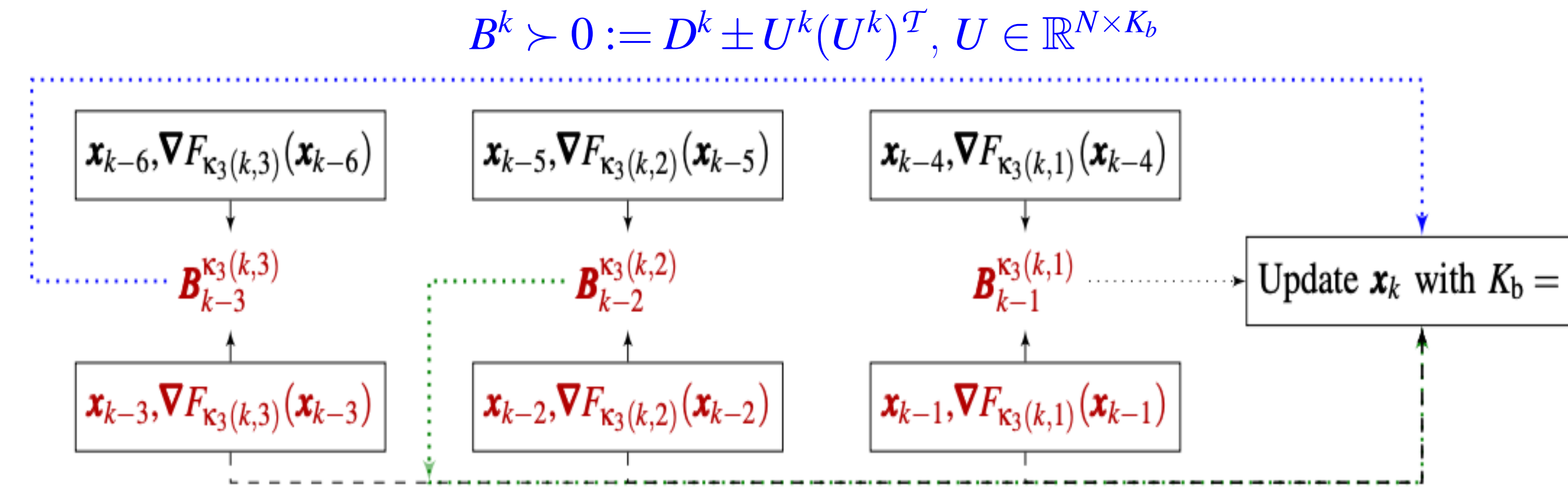
(1) is equivalent to

$$x_k = \arg \min_{x \in \mathcal{C}} \left(\frac{1}{2} \|x - v_k\|_{B^k}^2 + a_k K_b \lambda \text{TV}(x) \right), \quad (2)$$

where $B^k = \sum_t B_{k-t}^{\kappa_{K_b}(k,t)}$ and $v_k = (B^k)^{-1} \sum_t \left(B_{k-t}^{\kappa_{K_b}(k,t)} x_{k-t} - a_k \nabla F_{\kappa_{K_b}(k,t)}(x_{k-t}) \right)$.

If $B_{k-t}^{\kappa_{K_b}(k,t)} = I, \forall t$, (2) is equivalent to $\text{prox}(\cdot)$.

Estimate $B^k \in \mathbb{R}^{N \times N}$: $K_b = 3$ example



Total variation preliminaries — $x \in \mathbb{R}^N$

Isotropic TV:

$$\text{TV}_{\text{iso}}(x) = \text{tr} \left(\sqrt{\sum_{n=1}^d (\mathbf{D}^n x) (\mathbf{D}^n x)^T} \right),$$

while the anisotropic version is

$$\text{TV}_{\ell_1}(x) = \text{tr} \left(\sum_{n=1}^d \sqrt{(\mathbf{D}^n x) (\mathbf{D}^n x)^T} \right).$$

Equivalent formulation:

$$\text{TV}_{\text{iso}}(x) = \max_{P \in \mathbb{R}^{d \times N}, \{\|\mathbf{c}_r\|_2 \leq 1\}_{r=1}^N} \mathbf{d}(P)^T x$$

and

$$\text{TV}_{\ell_1}(x) = \max_{P \in \mathbb{R}^{d \times N}, \{\|\mathbf{c}_r\|_{\infty} \leq 1\}_{r=1}^N} \mathbf{d}(P)^T x,$$

where $P = [\mathbf{c}_1 \cdots \mathbf{c}_N] = [\mathbf{r}_1 \cdots \mathbf{r}_d]^T$ and $\mathbf{d}(P) = \sum_{n=1}^d (\mathbf{D}^n)^T \mathbf{r}_n$.

Compute (2) efficiently from the dual formulation

Dual problem of (2)

$$P^* = \arg \min_{P \in \mathcal{P}} \left(-\|w_k(P) - \text{prox}_{\delta_C}^{B^k}(w_k(P))\|_{B^k}^2 + \|w_k(P)\|_{B^k}^2 \right), \quad (3)$$

where $w_k(P) = v_k - a_k K_b \lambda (B^k)^{-1} \mathbf{d}(P)$ and $\text{prox}_{\delta_C}^{B^k}(x) = \arg \min_{u \in \mathbb{R}^N} (\delta_C(u) + \frac{1}{2} \|u - x\|_{B^k}^2)$.

$$x_k = \text{prox}_{\delta_C}^{B^k}(w_k(P^*)).$$

Gradient of (3)

$$-2a_k K_b \lambda \mathbf{d} \left(\text{prox}_{\delta_C}^{B^k}(w_k(P)) \right),$$

with Lipschitz constant $16\omega_{\min} a_k^2 K_b^2 \lambda^2$ (or $24\omega_{\min} a_k^2 K_b^2 \lambda^2$) for 2D (or 3D), where ω_{\min} is the smallest eigenvalue of B^k .

Theorem

[2, Theorem 3.4] Let $W = \Sigma \pm U U^T$, $W \succ 0 \in \mathbb{R}^{N \times N}$, and $U \in \mathbb{R}^{N \times \tilde{r}}$. Then, it holds that

$$\text{prox}_g^W(x) = \text{prox}_g^{\Sigma}(x \mp \Sigma^{-1} U \beta^*),$$

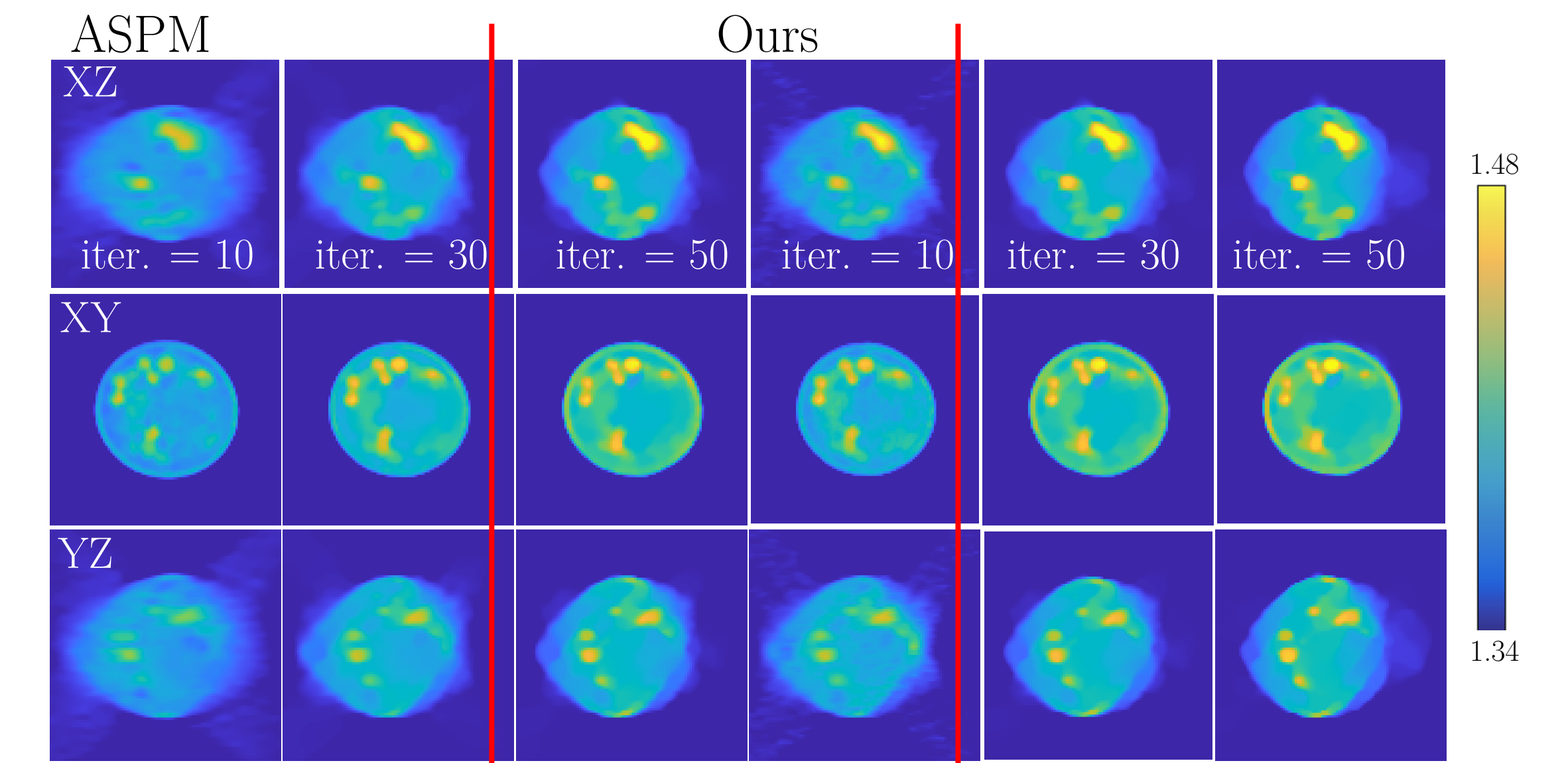
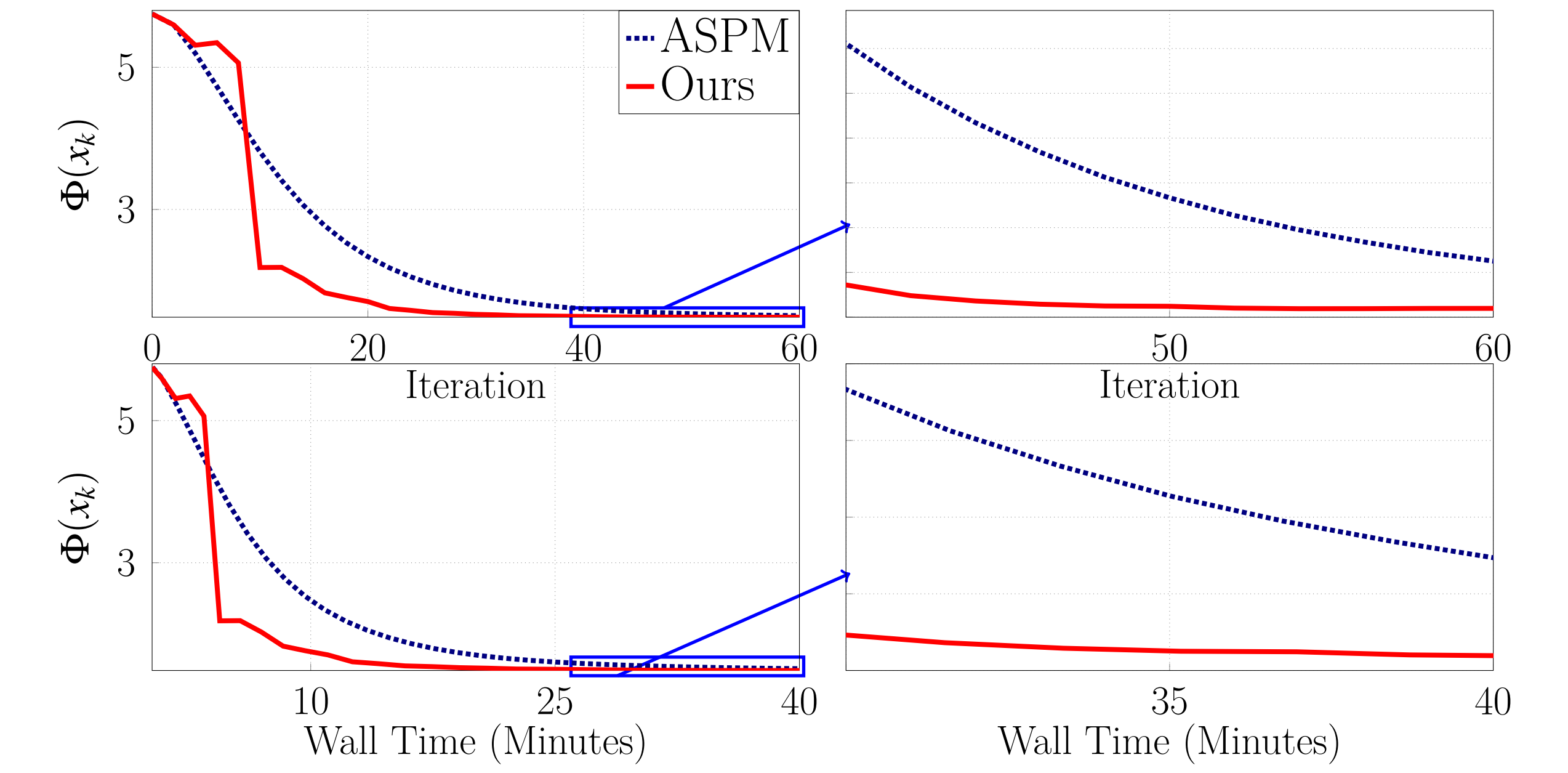
where $\beta^* \in \mathbb{R}^{\tilde{r}}$ is the unique solution of the nonlinear system of equation

$$\underbrace{U^T (x - \text{prox}_g^{\Sigma}(x \mp \Sigma^{-1} U \beta))}_{\varphi(\beta)} + \beta = 0.$$

Numerical experiment — real data

Experimental setting:

- A yeast cell immersed in water ($\eta_0 = 1.338$)
- $L = 60$ incident plane waves (wavelength: 532nm) embedded in a cone of illumination whose half-angle is 35°
- The discretized volume has 96^3 voxels of size 99^3nm^3
- 60×150^2 measurements, i.e., $y_p \in \mathbb{C}^{150 \times 150}$, $K_b = 5$ and run on a GPU



Open problems

- Convergence (rate) and recovery guarantee?
- Nonsmooth mini-batch quasi-Newton proximal? – only know $\{|y_p|\}_p$
- More accurate forward model? – Maxwell's equations?
- Nonlinear acceleration? – Anderson? (nonsmoothness & constraints)

References

- [1] Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Transactions on Image Processing, 18(11):2419–2434, 2009.
- [2] Stephen Becker, Jalal Fadili, and Peter Ochs. On quasi-Newton forward-backward splitting: Proximal calculus and convergence. SIAM Journal on Optimization, 29(4):2445–2481, 2019.
- [3] Aryan Mokhtari, Mark Eisen, and Alejandro Ribeiro. IQN: An incremental quasi-Newton method with local superlinear convergence rate. SIAM Journal on Optimization, 28(2):1670–1698, 2018.