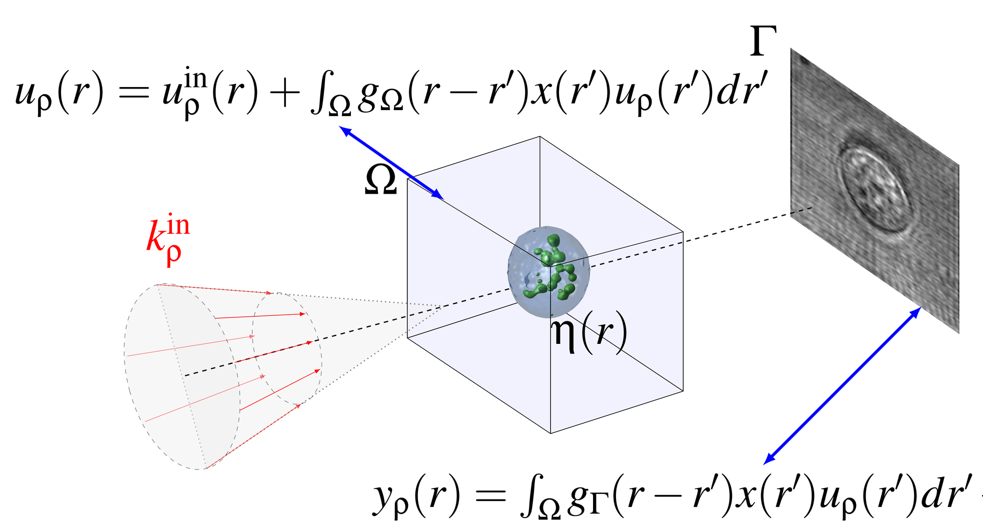
A Mini-Batch Quasi-Newton Proximal Method for Constrained Total-Variation Nonlinear Image Reconstruction

Tao Hong[1], Thanh-an Pham[2], Irad Yavneh[3], and Michael Unserz[4]

Functional MRI Lab, University of Michigan[1], Ann Arbor, MIT USA[2], Technion Israel[3], EPFL Switzerland[4]

Problem formulation

Principle of optical diffraction tomgraphy



- 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 $\{y_\rho(r)\}_\rho$

Composite minimization problem:

$$\min_{x\in\mathcal{C}} \Phi(x) \equiv \left(\frac{1}{L}\sum_{\rho=1}^{L} \underbrace{\frac{1}{2}\|\mathcal{H}_\rho(\mathbf{x}) - y_\rho\|_2^2}_{f_\rho} + \lambda\mathrm{TV}(x) \right)$$

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

Classical solver: accelerated stochastic proximal method (ASPM)

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

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

Proposed method

Our mini-batch quasi-Newton proximal Method:

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

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

2.

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

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

(1) is equivalent to

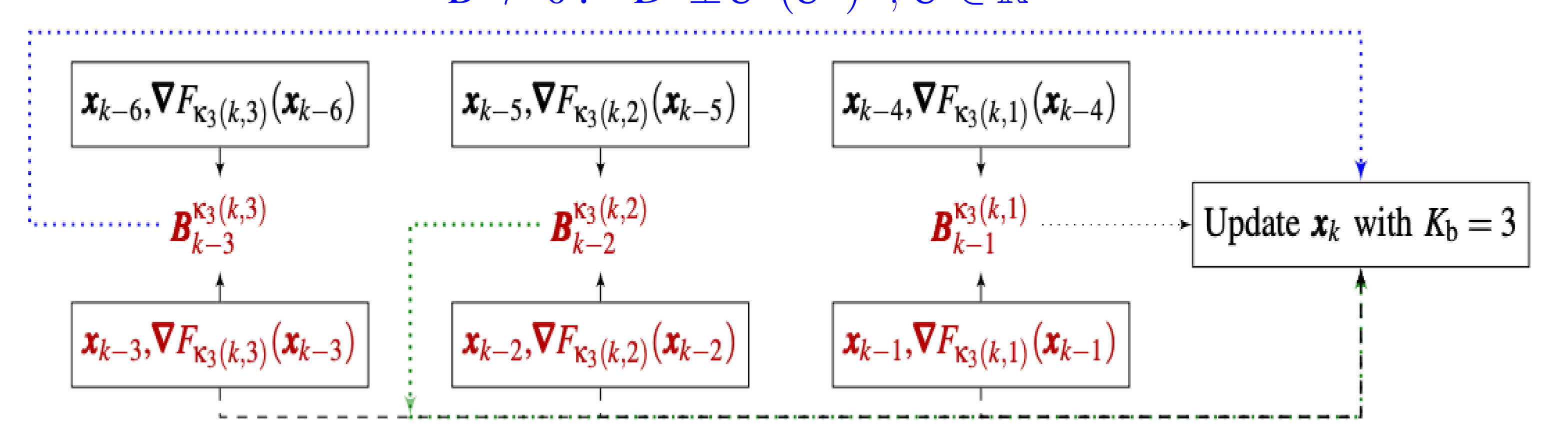
$$x_k = \arg\min_{x\in\mathcal{C}} \left(\frac{1}{2}\|x - v_k\|_{B^k}^2 + a_k K_b\lambda\mathrm{TV}(x) \right), \quad (2)$$

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 $\mathrm{prox}(\cdot)$.

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

$B^k \succ 0 := D^k \pm U^k(U^k)^{\mathcal{T}},\ U \in \mathbb{R}^{N\times K_b}$



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

Isotropic TV:

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

while the anisotropic version is

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

Equivalent formulation:

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

and

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

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

Compute (2) efficiently from the dual formulation

Dual problem of (2)

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

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

$$x_k = \mathrm{prox}_{\delta_{\mathcal{C}}}^{B^k}(w_k(P^*)).$$

Gradient of (3)

$$-2a_k K_b\lambda\mathbf{d}\left(\mathrm{prox}_{\delta_{\mathcal{C}}}^{B^k}(w_k(P))\right),$$

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

Theorem

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

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

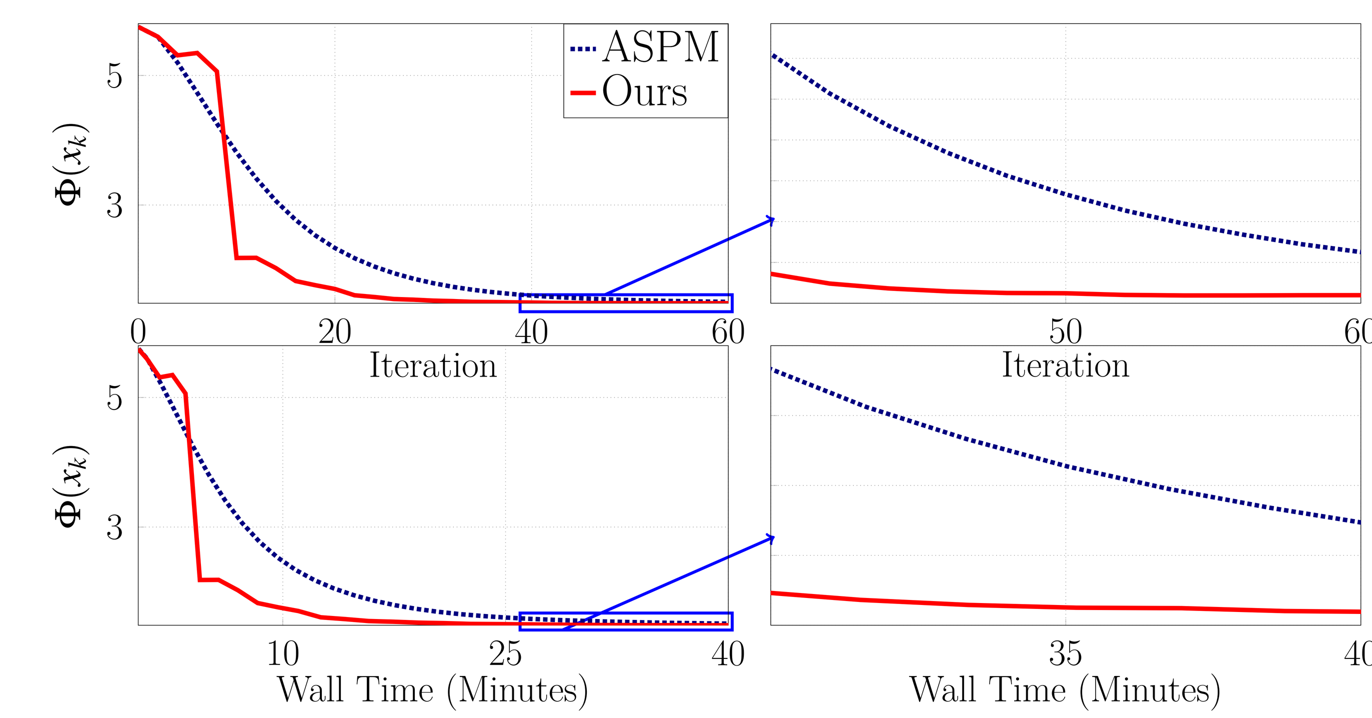
where $\beta^* \in \mathbb{R}^{\bar{r}}$ is the unique solution of the nonlinear system of equation

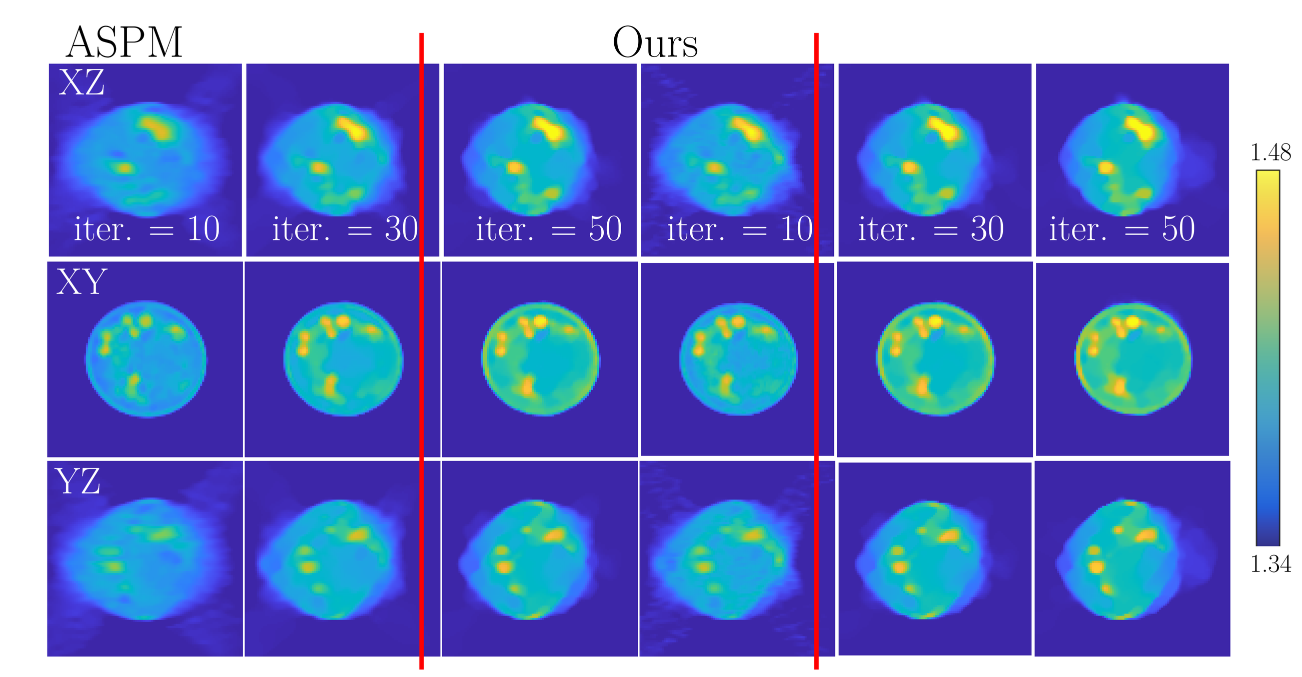
$$\underbrace{U^{\mathcal{T}}\left(x - \mathrm{prox}_g^\Sigma(x \mp \Sigma^{-1}U\beta)\right) + \beta}_{\varphi(\beta)} = 0.$$

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., $y_\rho \in \mathbb{C}^{150\times 150}$, $K_b = 5$ and run on a GPU





Open problems

- Convergence (rate) and recovery guarantee?
- Nonsmooth mini-batch quasi-Newton proximal? – only know $\{|y_\rho|\}_\rho$
- 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.