Efficient optimization of Poisson log likelihood and relative difference prior for PET using preconditioned SVRG

Matthias Ehrhardt[1], Zeljko Kereta[2], Georg Schramm[3]

[1]University of Bath, Department of Mathematical Sciences, Bath, United Kingdom

[2]University College London, Computer Science Department, London, United Kingdom

[3]KU Leuven, Department of Imaging and Pathology, Leuven, Belgium

Aim & General approach

find

$$x \in \underset{x \geq 0}{\operatorname{argmin}} \underbrace{-L(x) + R(x)}_{\Phi(x)} = \sum_{m=1}^{M} \underbrace{-L_m(x) + \frac{1}{M} R(x)}_{\Phi_m(x)}$$

as "quickly" as possible
(up to given tolerance)

$$-L(x) = -logL_{Poisson}(x) = \sum_{i=1}^{n_{data}} (Ax+s)_i - y_i \log(Ax+s)_i$$

$$R(x) = \frac{\beta}{2} \sum_{j=1}^{n_{voxel}} \sum_{k \in N_j} w_{jk} \kappa_j \kappa_k \frac{(x_j - x_k)^2}{x_j + x_k + \gamma |x_j - x_k| + \varepsilon}$$

Pre-conditioned Stochastic Gradient Descent

$$x^{(n+1)} = \left[x^{(n)} - \alpha \, D \, \tilde{\nabla}_m \Phi(x^{(n)}) \right]_+$$

scalar step size — α

(diagonal) preconditioner — D

stochastic gradient using data subset m — $\tilde{\nabla}_m \Phi(x^{(n)})$

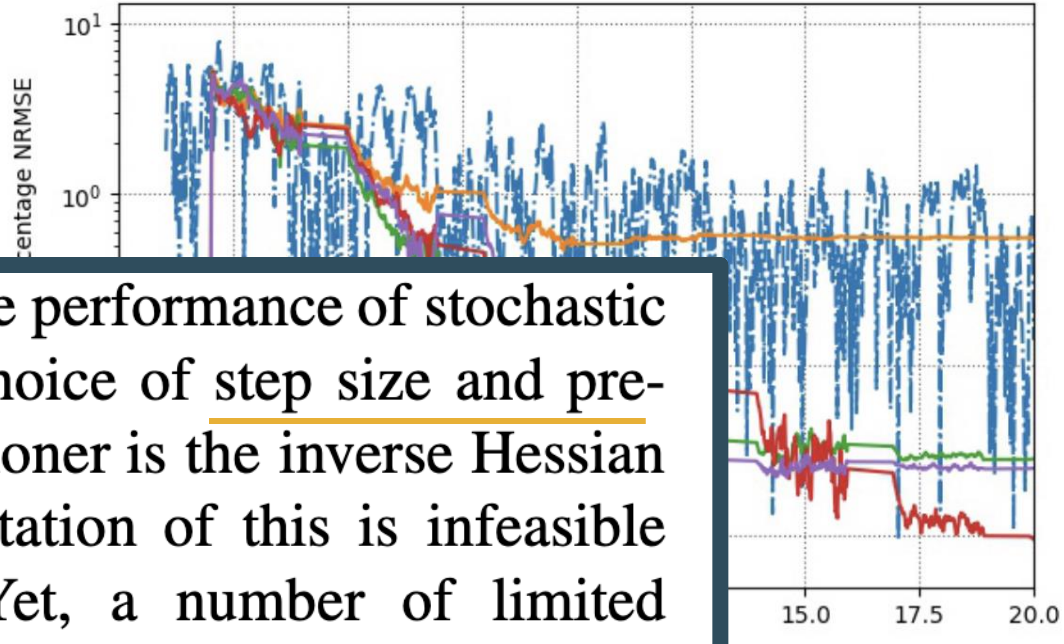
Nuyts et al. "A Concave Prior Penalizing Relative Differences for Maximum-a-Posteriori Reconstruction in Emission Tomography", IEEE TNS 2002

SVRG – stochastic variance reduced gradient descent

Algorithm 2 SVRG Algorithm

Given $x_0 \in \mathbb{R}^{N_v}$, M, K, $\{\alpha_k > 0\}_{k=0}^{K}$ and $\gamma > 0$
for $k = 0, 1, \ldots, K$ **do**
  if $k \mod \gamma M \equiv 0$ **then**
    $\tilde{x} \leftarrow x_k$
    for $m \in \{1$
      Comput
      Store g
    end for
    $\tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(x_k$
  else
    Choose m_k
    Compute $\nabla\Phi_{m_k}(x_k)$
    $\tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(x_k) \leftarrow M\left(\nabla\Phi_{m_k}(x_k) - g_{m_k}\right) + \sum_{m=1}^{M} g_{m_k}$
  end if
  $x_{k+1} \leftarrow x_k + \alpha_k \tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(x_k)$ *# Update step*
end for

In this work, we observed that the performance of stochastic algorithms was impacted by the choice of step size and preconditioner. The optimal preconditioner is the inverse Hessian of the objective. However, computation of this is infeasible for the 3D PET problem [1]. Yet, a number of limited



(b) SVRG

Fig. 4. Lung lesion percentage NRMSE over multiple stochastic realisations with preconditioners anchored at different epochs. The BSREM NRMSE value was computed for a single deterministic realisation with (16).

Twynman et al. "An Investigation of Stochastic Variance Reduction Algorithms for Relative Difference Penalized 3D PET Image Reconstruction", IEEE TMI, 2023
Johnson and Zhang, "Accelerating stochastic gradient descent using predictive variance reduction", Proc. Adv. Neural Inf. Process. Syst, 2013

Proposed improved preconditioner

$$x \in \underset{x \geq 0}{\text{argmin}} \; -L(x) + R(x)$$

$$x^{(n+1)} = \left[x^{(n)} - \alpha \, D \, \tilde{\nabla} \Phi_m(x^{(n)}) \right]_+$$

$$D_{data} = diag\left(\frac{x}{A^T 1}\right)$$

$$D_{prior} = \left(H^R(x)\right)^{-1} \qquad H^R = \nabla^2 R$$

$$D_{prior} \approx diag\left(H^R(x)\right)^{-1} \qquad diag\left(H^R\right)_j = H^R_{jj} = \frac{\partial^2}{\partial x_j \partial x_j} R \qquad H^R_{jj} = 2\beta \sum_{k \in N_j} w_{jk} \kappa_j \kappa_k \frac{(2x_k + \varepsilon)^2}{((x_j + x_k) + \gamma |x_j - x_k| + \varepsilon)^3}$$

assuming $w_{jk} = w_{kj}$

$$D = \frac{D_{data} \, D_{prior}}{D_{data} + D_{prior}} = \frac{x}{A^T 1 + diag(H^R(x)) \, x}$$

Implemented diagonal preconditioner

$$D = \frac{x_{sm}}{A^T 1 + \boxed{1.5} \, diag(H^R(x_{sm})) \boxed{x_{sm}}}$$

fudge factor for effect of non-diagonal elements in H^R

smoothed version of current image

Nuyts et al. "A Concave Prior Penalizing Relative Differences for Maximum-a-Posteriori Reconstruction in Emission Tomography", IEEE TNS 2002

56 IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 49, NO. 1, FEBRUARY 2002

A Concave Prior Penalizing Relative Differences for *Maximum-a-Posteriori* Reconstruction in Emission Tomography

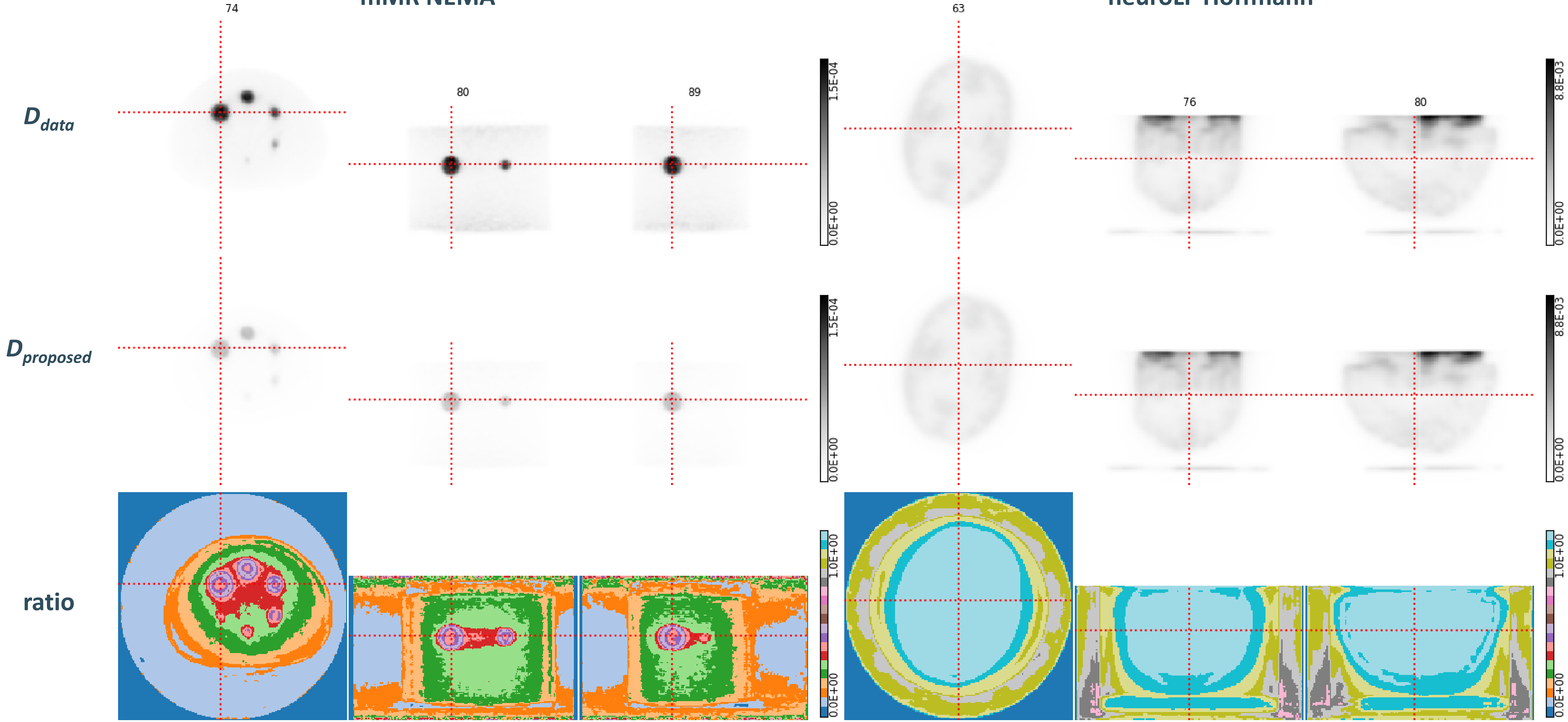
J. Nuyts, D. Bequé, P. Dupont, and L. Mortelmans

$$\lambda_j = \lambda_j^{\mathrm{old}} + \left(\frac{\partial L}{\partial \lambda_j} + \frac{\partial M}{\partial \lambda_j}\right) \Big/ \left(\frac{\sum_i c_{ij}}{\lambda_j^{\mathrm{old}}} - \frac{\partial^2 M}{\partial \lambda_j^2}\right)$$

Nuyts et al. “A Concave Prior Penalizing Relative Differences for Maximum-a-Posteriori Reconstruction in Emission Tomography”, IEEE TNS 2002

mMR NEMA

neuroLF Hoffmann



Algorithm hyper parameters

	ALG1
number of subsets	divisor of num views closest to 25
step size rule for α	“heuristic” 0 < update <= 10 → 3 10 < update <= 100 → 2 100 < update <= 200 → 1.5 200 < update <= 300 → 1.0 300 > update → 0.5
subset selection rule	“random”, every subset once in every epoch

Things we also tried but did not work well …

- using **elementwise minimum** of D_{data} and D_{prior} for D
- approximation of the inverse Hessian via the **Woodbury matrix identity**
- **subset sampling based on subset gradient norm** instead of uniform sampling (no real performance gain)
- (S)**PDHG** with **iterative** calculation of the **proximity operator** of $R(x)$ → numerical approximation too slow

https://www.xkcd.com/2451/

Finally - a big thanks to