Accelerating Joint PET/SPECT/CT Reconstruction with Synergy-Aware Preconditioning and Variance Reduction

Sam Porter [1,2,3], Daniel Deidda [3],
Simon Arridge [2,4], Kris Thielemans [1,2]

Symposium on AI and Reconstruction for Biomedical Imaging, 10th February 2026

[1] The Institute of Nuclear Medicine, UCL
[2] The Hawkes Institute, UCL
[3] Medical, Marine and Nuclear, NPL
[4] The Centre for Inverse Problems, UCL

UCL

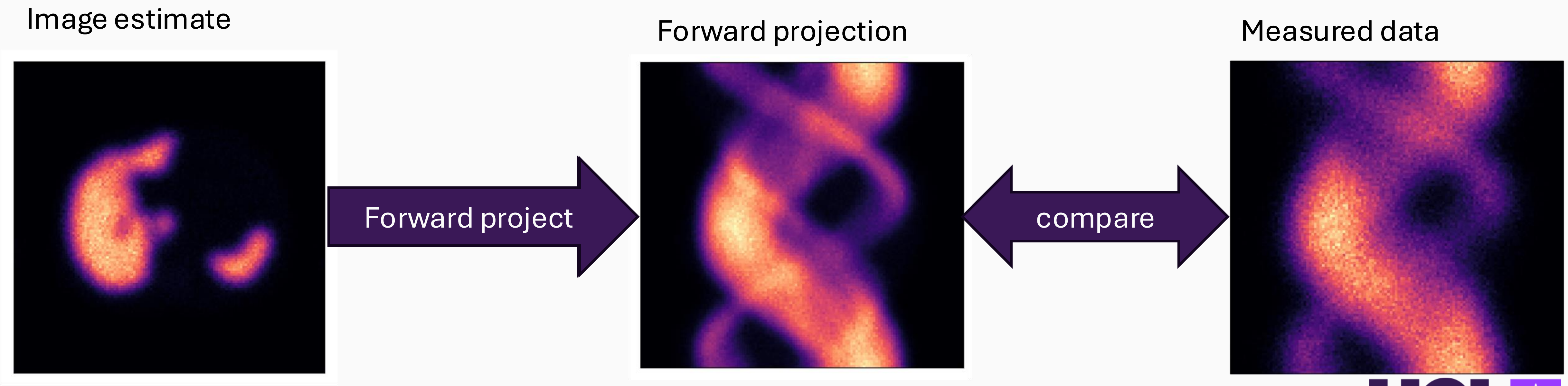
PET and SPECT

$$\bar{y}_i = \sum_{j=1}^{J} a_{ij} x_j + b_i, \qquad i \in \{\text{LoRs}\},\ j \in \{\text{voxels}\}$$

$$\mathcal{D}(x) = \sum_i \left(\bar{y}_i - y_i \log(\bar{y}_i) + \mathcal{C}\right.$$

$$\bar{y} = Ax + b, \qquad y_i \sim \text{Poisson}(\bar{y}_i)$$

$$\mathcal{O}(x) = \mathcal{D}(x) + \lambda \mathcal{R}(x)$$



Synergistic Reconstruction

We want to find the set of images that minimise a joint objective function,

$$(\hat{x}_1, \hat{x}_2) = \underset{x_1, x_2 \geq 0}{\arg\min} \mathcal{D}_1(x_1) + \mathcal{D}_2(x_2) + \lambda \mathcal{R}_{\mathrm{wdTNV}}(x_1, x_2)$$

With a joint prior, promoting shared anatomical boundaries. We use directional total nuclear variation (wdTNV)

$$Y_j(x_1, x_2) = \left[b_1 (D_1 \nabla x_1)_j \,\middle|\, b_2 (D_2 \nabla x_2)_j \right] \in \mathbb{R}^{d \times 2}, \qquad \mathcal{R}_{\mathrm{wdTNV},\delta}(x_1, x_2) = \sum_{j=1}^{J} \sum_{\ell=1}^{2} \sqrt{\sigma_\ell(Y_j)^2 + \delta^2}$$

$$\xi_j = \frac{(\nabla \mu)_j}{\|(\nabla \mu)_j\|_2 + \eta}, \qquad (D_m u)_j = (I - \gamma_m \, \xi_j \xi_j^\top) \, u_j, \qquad m \in \{1, 2\}, \quad \gamma_m = 1$$

Preconditioning

$$\mathbf{x}^{(k+1)} = \left[\mathbf{x}^{(k)} - \alpha_k \, P^{(k)} \, \nabla\mathcal{O}\left(\mathbf{x}^{(k)}\right)\right]_+$$

Preconditioned projected gradient descent

$$P_{\mathcal{D},m} = M_{\mathcal{D},m}^{-1} = \operatorname{diag}\left(\frac{x_m}{A^T\mathbf{1}}\right)$$

EM-type preconditioner: majorises Expected Hessian of PLL

$$\nabla^2\mathcal{R}(\mathbf{x}) \preceq M_{\mathcal{R}}^{(k)} \qquad \Longrightarrow \qquad P_{\mathcal{R}}^{(k)} = \left(M_{\mathcal{R}}^{(k)}\right)^{-1}$$

For guaranteed descent with unit step size , we want a preconditioner using a hessian approximation that majorises the prior curvature

$$P^{(k)} = \left(\left(P_{\mathcal{D}}^{(k)}\right)^{-1} + \lambda\left(P_{\mathcal{R}}^{(k)}\right)^{-1}\right)^{-1} = \left(M_{\mathcal{D}}^{(k)} + \lambda M_{\mathcal{R}}^{(k)}\right)^{-1}$$

We can sum these majorising curvature approximations and use these to form preconditioner

An iteratively reweighted least squares upper bound

1. Redefine prior in terms of Gram matrix

$$G_j(x_1, x_2) := Y_j(x_1, x_2)^\top Y_j(x_1, x_2) \in \mathbb{R}^{2\times 2} \qquad \mathcal{R}_{\mathrm{wdTNV}}(x_1, x_2) = \sum_j \mathrm{tr}\Big((G_j(x_1, x_2) + \delta I_2)^{1/2}\Big),$$

2. Construct quadratic surrogate to estimate upper bound for curvature

$$W_j^{(k)} := \Big(G_j(x_1^{(k)}, x_2^{(k)} + \delta I_2)\Big)^{-1/2}, \qquad \mathcal{R}_{\mathrm{wdTNV}}(x_1, x_2) \leq C^{(k)} + \frac{1}{2}\sum_j \mathrm{tr}\left(W_j^{(k)} Y_j^\top Y_j\right)$$

3. Pull this curvature estimate back to image space

$$H^{(k)} = \sum_j J_j^\top \left(W_j^{(k)} \otimes I_d\right) J_j, \qquad (J_j x) := \mathrm{vec}\Big(\big[b_1 D_1(\nabla x_1)_j,\ b_2 D_2(\nabla x_2)_j\big]\Big)$$

4. Ignore off-diagonal terms

$$\left(P_j^{(k)}\right)^{(-1)} := \left[H^{(k)}\right]_{jj} \approx \sum_{\ell \in \mathcal{N}(j)} \begin{pmatrix} w_{\ell,11}^{(k)}\, b_1^2\, c_{\ell,j}^\top D_1^\top D_1 c_{\ell,j} & w_{\ell,12}^{(k)}\, b_1 b_2\, c_{\ell,j}^\top D_1^\top D_2 c_{\ell,j} \\ w_{\ell,21}^{(k)}\, b_1 b_2\, c_{\ell,j}^\top D_2^\top D_1 c_{\ell,j} & w_{\ell,22}^{(k)}\, b_2^2\, c_{\ell,j}^\top D_2^\top D_2 c_{\ell,j} \end{pmatrix}$$

$$c_{\ell,j}[k] := \frac{1}{h_k}\Big(\mathbf{1}_{\{j=\ell+s_k\}} - \mathbf{1}_{\{j=\ell\}}\Big)\mathbf{1}_{\{\ell+s_k\in\Omega\}}, \qquad k = 1, \ldots, d,$$

Where can we go from here?

1. We use the full pulled-back block Jacobi IRLS surrogate:

$$P_{\mathrm{blk},j} := \begin{bmatrix} p_{11,j} & p_{12,j} \\ p_{12,j} & p_{22,j} \end{bmatrix}$$

2. We construct an approximately majorising diagonal:

$$P_{\mathrm{maj},j} := \mathrm{diag}(p_{11,j} + |p_{12,j}|,\ p_{22,j} + |p_{12,j}|)\,.$$

3. How much can cross-modality coupling matter?

$$P_{\mathrm{diag},j} := \mathrm{diag}(p_{11,j},\ p_{22,j})\,.$$

The Experiment

We test 4 preconditioning strategies:

$$\left(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1}\right)^{-1}$$
$$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$$
$$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$$
$$P_{\mathcal{D}}$$

- 10 epochs SVRG
- Initialise with OSEM + Gaussian smooth
- Initial step sizes, $\alpha_0 \in [1, 2]$
- Step size relaxation, $\alpha_k = \dfrac{\alpha_0}{1 + 0.005k}$
- 5 repeated runs
- Preconditioner updated once per epoch

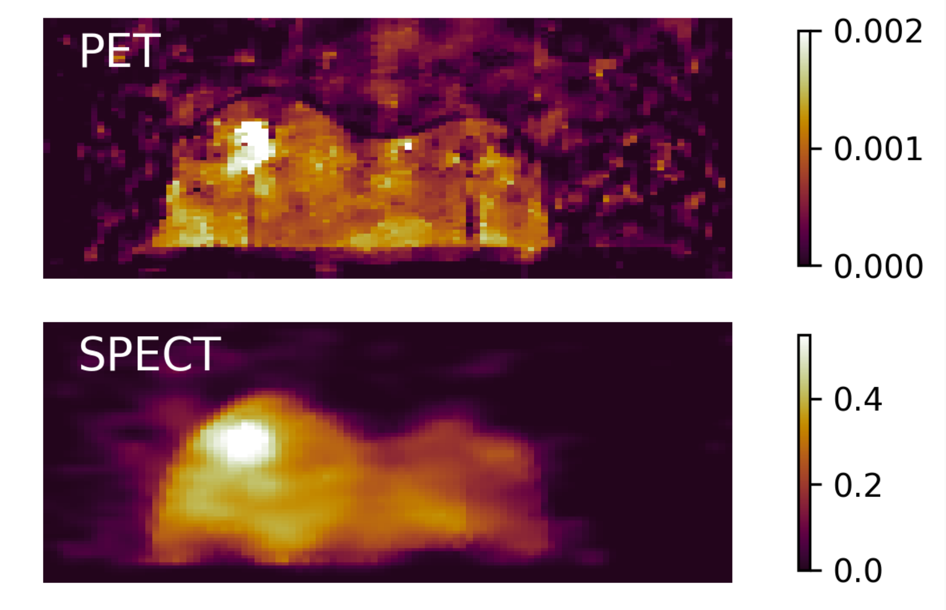
Versus one "converged reconstruction”

$$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$$

- 5000 epochs SVRG
- Initialise with OSEM + Gaussian smooth
- 4 restarts (once per 1000 epochs)
- Initial step size, $\alpha_0 = 0.8$
- Step size relaxation, $\alpha_k = \dfrac{\alpha_0}{1 + 0.001k}$
- Preconditioner updated once per epoch

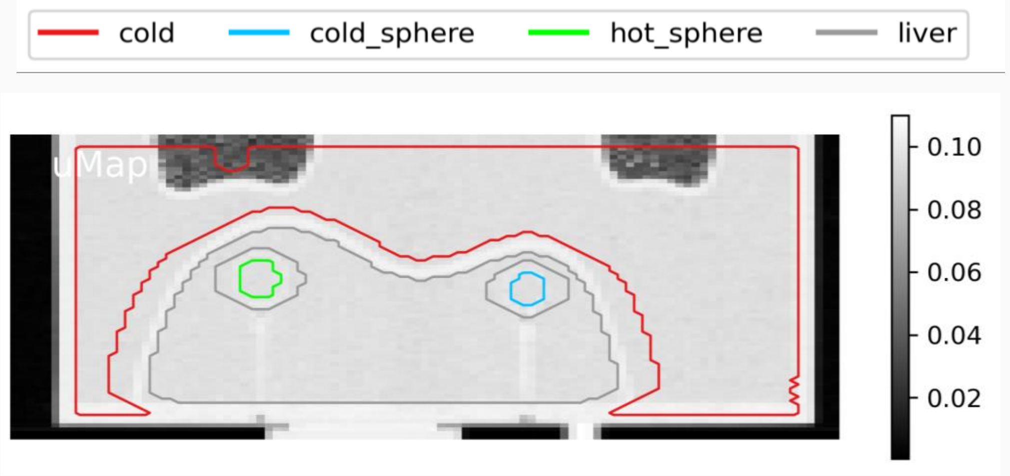
The Data

- Yttrium-90 anthropomorphic phantom data
- Warm liver, cold background, hot sphere, cold sphere
- $b_1 = 0.028, \quad b_2 = 9.06$



- Compared NRMSE for each VOI $\mathcal{V}$

$$\mathrm{NRMSE}_{\mathcal{V}}(\hat{x}, x_{\mathrm{true}}) = \frac{\sqrt{\frac{1}{|\mathcal{V}|}\sum_{j\in\mathcal{V}}\left(\hat{x}_j - x_{\mathrm{true},j}\right)^2}}{\sqrt{\frac{1}{|\mathcal{V}|}\sum_{j\in\mathcal{V}} x_{\mathrm{true},j}^2}}$$



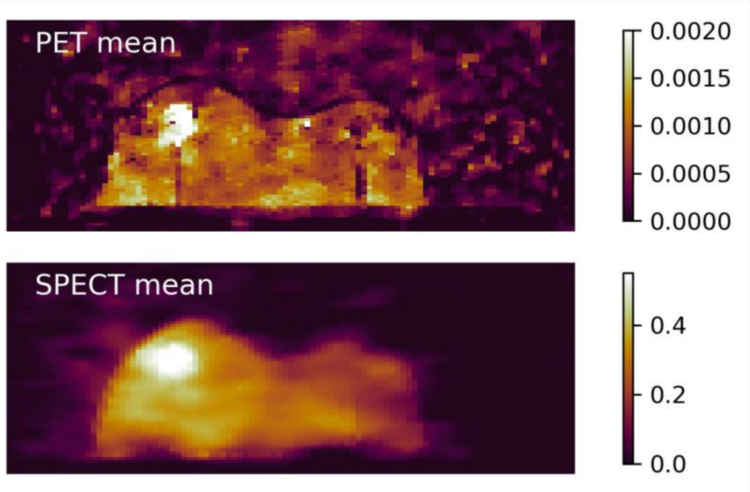
Results – speed (10 repeats)

Preconditioner	Time taken (ave.)	Standard Deviation
$\left(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1}\right)^{-1}$	1.502	0.031
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	0.844	0.021
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	0.854	0.026
$P_{\mathcal{D}}$	0.257	0.024

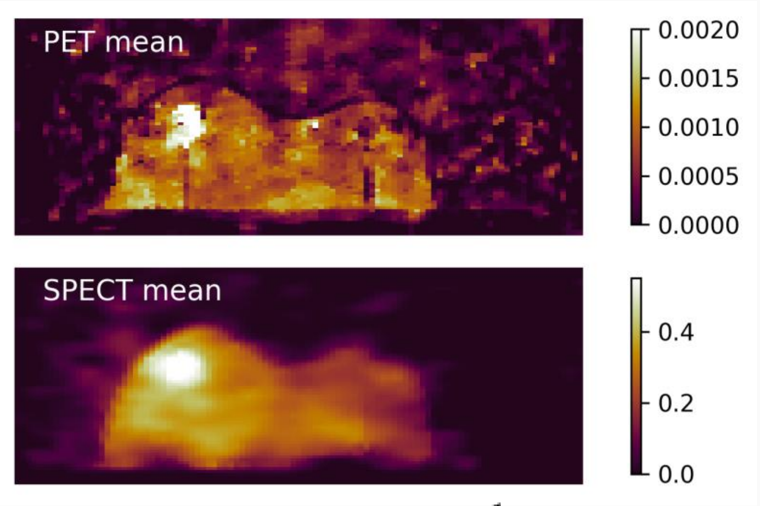
*But only once an epoch – insignificant wrt. projections

Convergence

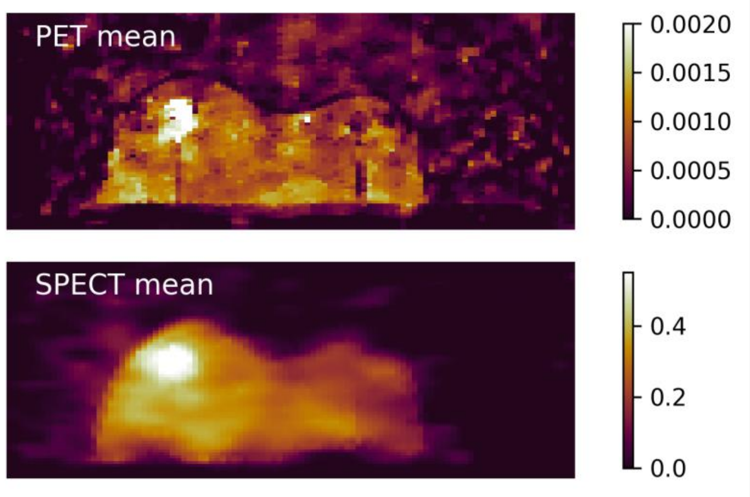
- Visually very similar reconstructions after 10 epochs
- Good news for any potential clinical use!
- But does beg the question: was this all necessary?



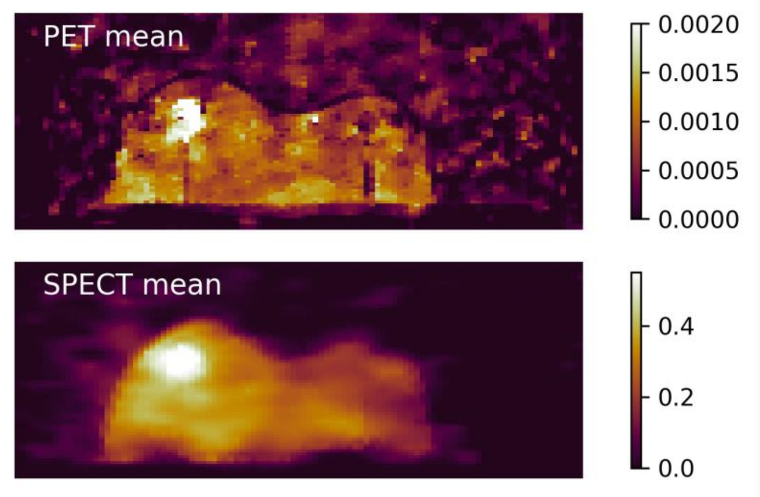
$P_{\mathcal{D}}$



$\left(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1}\right)^{-1}$



$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$



$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$

Convergence – NRMSE after 10 iterations

PET

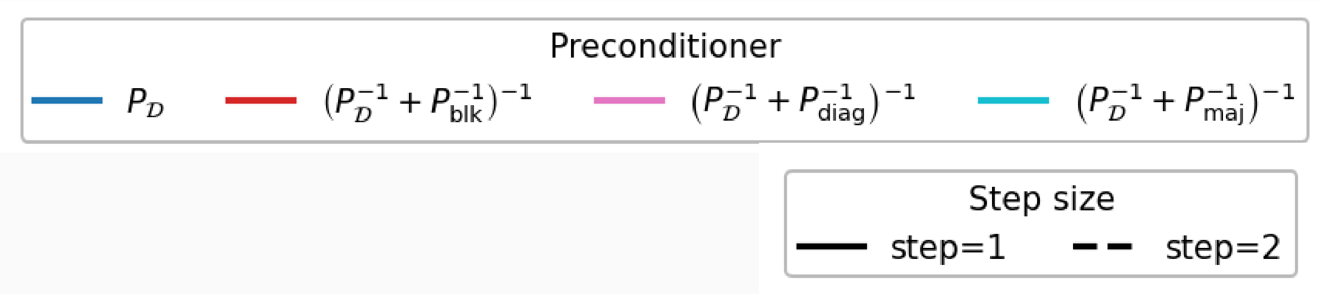
Preconditioner	Step size	Hot_sphere	Cold_sphere	Liver	Cold background
$P_{\mathcal{D}}$	1	0.003509	0.004234	0.017391	0.122952
$P_{\mathcal{D}}$	2	0.009736	0.003244	0.082271	0.082471
$P_{\mathcal{D}}$	3	0.043283	0.026009	0.161189	0.145153
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	1	0.032112	0.069766	0.029725	0.119204
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	2	0.005733	0.015707	0.012188	0.042838
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	3	0.011918	0.023016	0.015334	0.087762
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	1	0.033237	0.071119	0.029964	0.142970
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	2	0.005502	0.018060	0.012526	0.058083
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	3	0.039508	0.057516	0.037004	0.135318
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	1	0.042722	0.072326	0.030352	0.144030
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	2	0.008082	0.018050	0.012528	0.057777
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	3	0.026250	0.029120	0.026902	0.139474

Convergence – NRMSE after 10 iterations

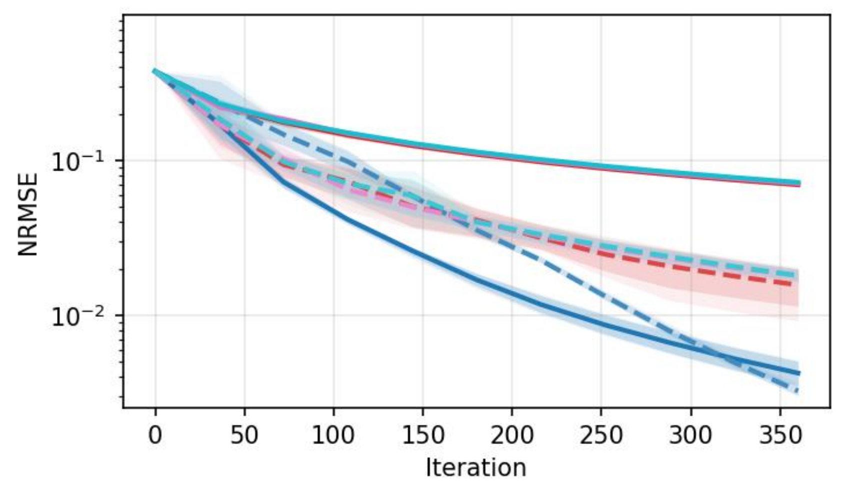
SPECT

Preconditioner	Step size	Hot_sphere	Cold_sphere	Liver	Cold background
$P_{\mathcal{D}}$	1	0.006053	0.007483	0.020221	0.199658
$P_{\mathcal{D}}$	2	0.004162	0.003000	0.013491	0.106508
$P_{\mathcal{D}}$	3	0.010964	0.003838	0.018268	0.070286
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	1	0.006894	0.016440	0.035976	0.210475
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	2	0.005720	0.008208	0.020695	0.140382
$(P_{\mathcal{D}}^{-1} + P_{\text{blk}}^{-1})^{-1}$	3	0.005286	0.006048	0.015079	0.116656
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	1	0.003591	0.028608	0.041008	0.215486
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	2	0.006362	0.010533	0.018831	0.110105
$\left(P_{\mathcal{D}}^{-1} + P_{\text{diag}}^{-1}\right)^{-1}$	3	0.004935	0.005018	0.009987	0.067594
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	1	0.003545	0.028281	0.040955	0.215470
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	2	0.006112	0.010463	0.018901	0.110269
$\left(P_{\mathcal{D}}^{-1} + P_{\text{maj}}^{-1}\right)^{-1}$	3	0.003868	0.004468	0.010004	0.068204

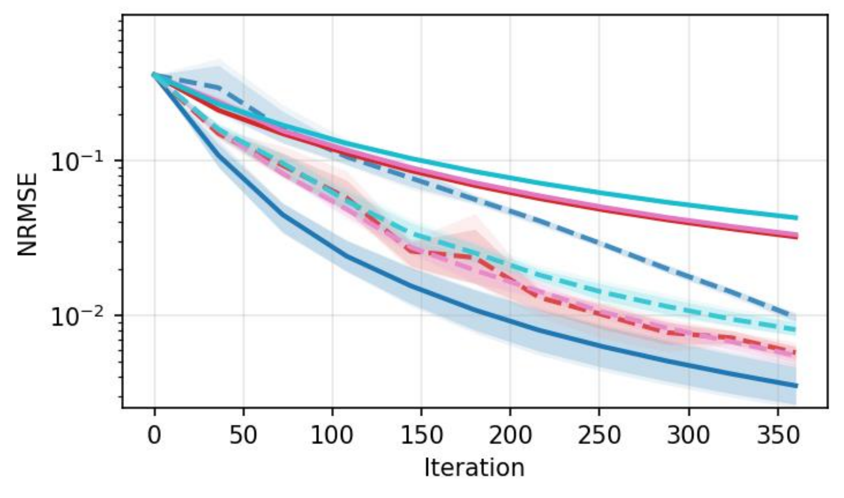
Convergence – PET



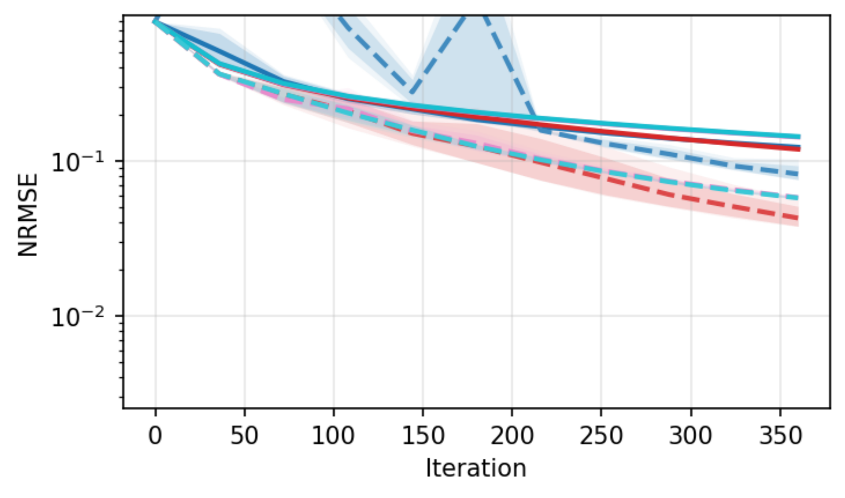
Cold Sphere



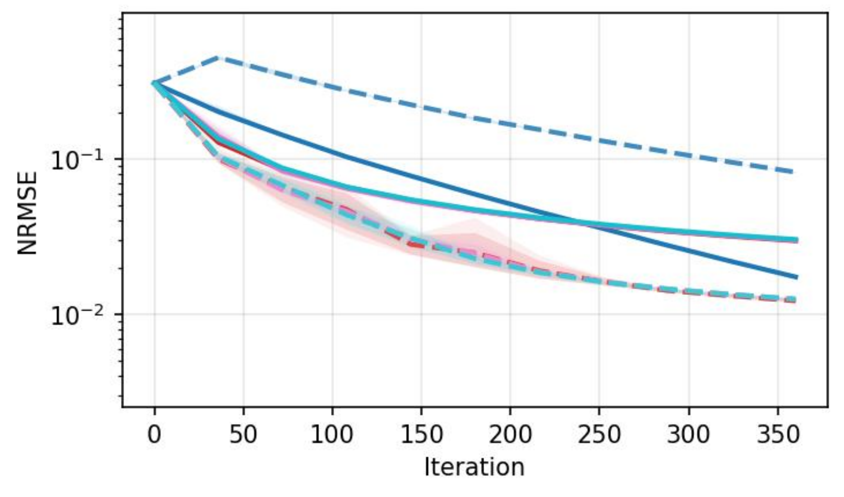
Hot Sphere



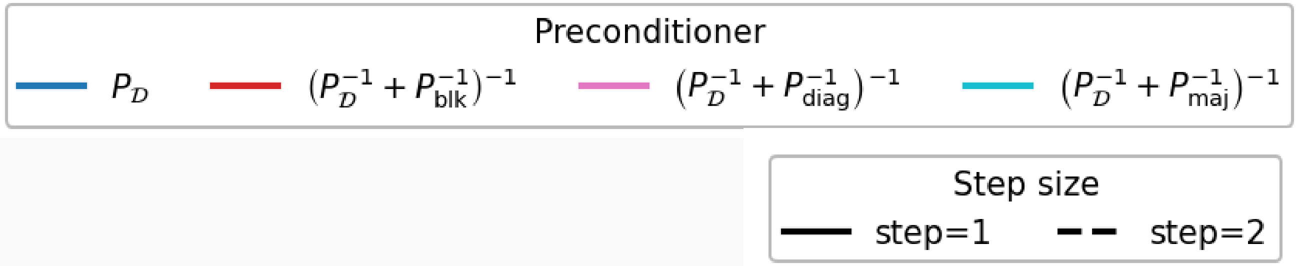
Cold Region



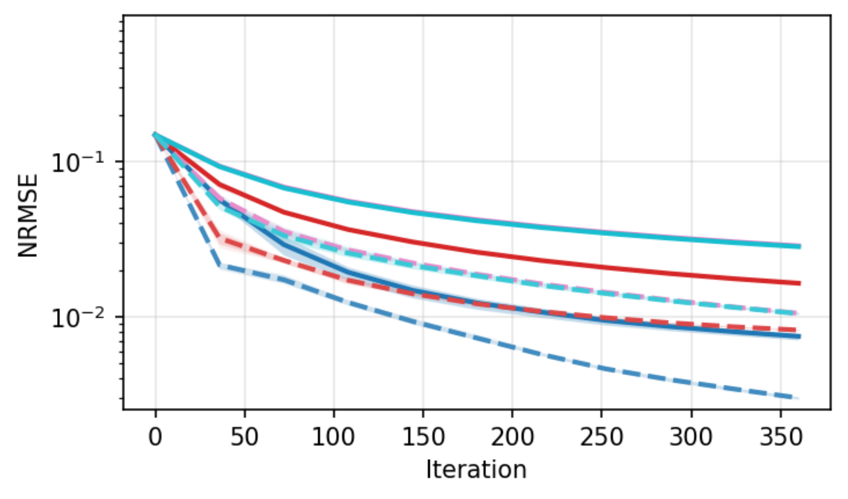
Liver



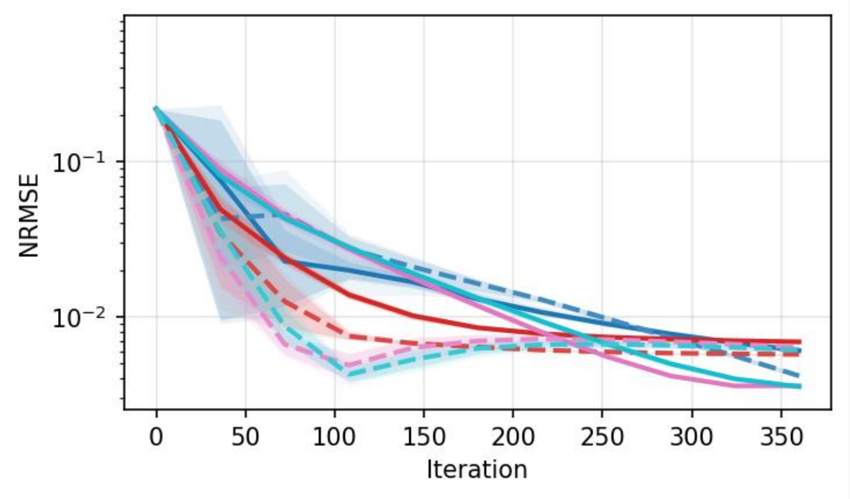
Convergence – SPECT



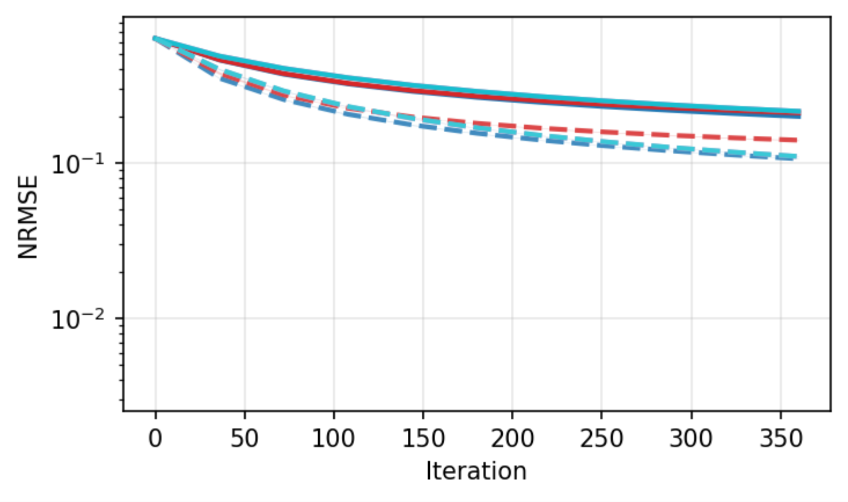
Cold Sphere



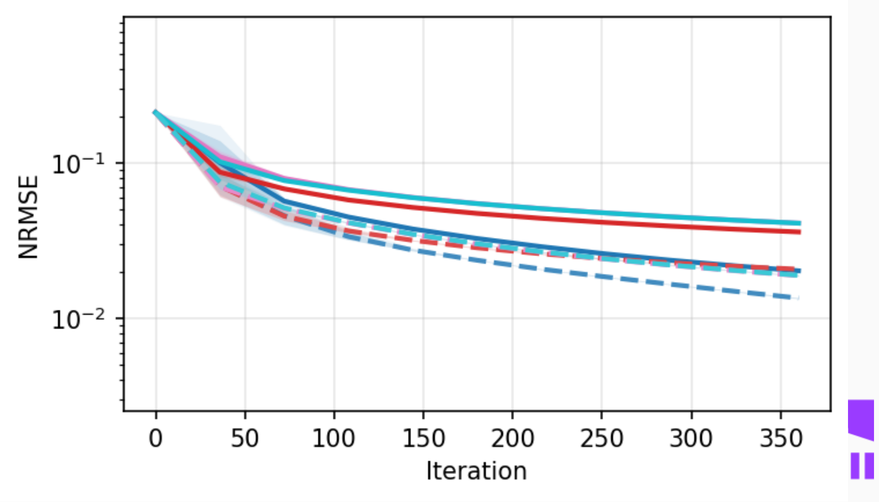
Hot Sphere



Cold Region



Liver



Conclusions

For this data set:

- EM-type precondtioner often out-performed the prior-aware preconditioners.

- Synergy-aware preconditioner did not significantly outperform diagonal preconditioner.

- Synergy-aware preconditioner costs 2x time for diagonal and 6x more than EM-type.

Discussion

Was this data set too simple?

- Regions of very uniform does
- Edge regions not included in analysis VOIs
- EM-type preconditioner fails with uniform step size for patient data

Do off diagonals matter?

- We ignore off diagonals in the Hessian (voxel-wise coupling)
- Ehrhardt et al. use an empirical weighting factor. Necessary?

Is the quadratic surrogate too conservative?

- Updated once per epoch – gets stale?
- Can we use a tighter alternative?

Future work:

- Same experiment on patient data!

Thanks for listening!

References

[1] K. Holt, “Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields,” IEEE Trans. Image Process., vol. 23, no. 9, pp. 3975–3989, Sep. 2014, doi: 10.1109/TIP.2014.2332397.

[2] D. S. Rigie and P. J. La Rivière, “Joint Reconstruction of Multi-channel, Spectral CT Data via Constrained Total Nuclear Variation Minimization,” Phys. Med. Biol., vol. 60, no. 5, pp. 1741–1762, Feb. 2015, doi: 10.1088/0031-9155/60/5/1741.

[3] M. J. Ehrhardt and M. M. Betcke, “Multicontrast MRI Reconstruction with Structure-Guided Total Variation,” SIAM J. Imaging Sci., vol. 9, no. 3, pp. 1084–1106, 2016, doi: 10.1137/15M1047325.

[4] M. J. Ehrhardt, Z. Kereta, and G. Schramm, “Fast PET reconstruction with variance reduction and prior-aware preconditioning,” *Front. Nucl. Med.*, vol. 5, Art. no. 1641215, Sep. 2025, doi: 10.3389/fnume.2025.1641215.

[5] A. R. De Pierro, “A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography,” IEEE Trans. Med. Imaging, vol. 14, no. 1, pp. 132–137, Mar. 1995, doi: 10.1109/42.370409.

[6] S. Ahn and J. A. Fessler, “Globally Convergent Image Reconstruction for Emission Tomography Using Relaxed Ordered Subsets Algorithms,” IEEE Trans. Med. Imaging, vol. 22, no. 5, pp. 613–626, May 2003, doi: 10.1109/TMI.2003.812251.

[7] F. R. K. Chung, Spectral Graph Theory, CBMS Regional Conf. Ser. Math., no. 92. Providence, RI, USA: American Mathematical Society, 1997.

[8] K. Lange, D. R. Hunter, and I. Yang, “Optimization Transfer Using Surrogate Objective Functions,” J. Comput. Graph. Stat., vol. 9, no. 1, pp. 1–20, 2000, doi: 10.1080/10618600.2000.10474858.

[9] A. S. Lewis, “Derivatives of Spectral Functions,” Math. Oper. Res., vol. 21, no. 3, pp. 576–588, Aug. 1996, doi: 10.1287/moor.21.3.576.

Variance Reduction

A quick foray into the stochastic variance reduced algorithm (SVRG) for stochastic reconstruction

$$\mathcal{O}(\mathbf{x}) = \sum_{s=1}^{S_1} \left(\mathcal{D}_{1,s}(x_1) + \frac{\lambda}{S}\mathcal{R}(\mathbf{x}) \right) + \sum_{s=S_1+1}^{S} \left(\mathcal{D}_{2,s}(x_2) + \frac{\lambda}{S}\mathcal{R}(\mathbf{x}) \right)$$

$$\mathcal{O}_s(\mathbf{x}) = \begin{cases} \mathcal{D}_{1,s}(x_1) + \frac{\lambda}{S}\mathcal{R}(\mathbf{x}) & \text{for } 1 \leq s \leq S_1 \\ \mathcal{D}_{2,s}(x_2) + \frac{\lambda}{S}\mathcal{R}(\mathbf{x}) & \text{for } S_1 < s \leq S \end{cases}$$

$$\tilde{\mu} = \sum_{s=1}^{S} \nabla \mathcal{O}_s(\tilde{\mathbf{x}}) \qquad v^{(k)} = \nabla \mathcal{O}_{s_k}(\mathbf{x}^{(k)}) - \nabla \mathcal{O}_{s_k}(\tilde{\mathbf{x}}) + \tilde{\mu}$$

$$\mathbf{x}^{(k+1)} = \left[\mathbf{x}^{(k)} - \alpha_k P^{(k)} v^{(k)} \right]_+$$

Preconditioning

$$
\begin{aligned}
\left(P^{(k)}\right)^{-1} &:= \mathrm{diag}_{\mathrm{vox}}\left(H^{(k)}\right) && \text{(block-Jacobi Hessian approximation; applied as a preconditioner)}\\
\left(P_j^{(k)}\right)^{-1} &:= \left[H^{(k)}\right]_{jj} && \text{(the } 2\times 2 \text{ block acting on } [x_1(j), x_2(j)]^\top\text{)}\\
\left(P_j^{(k)}\right)^{-1} &\approx \sum_{\ell\in\mathcal{N}(j)} \alpha_{\ell\to j}\, B\, W_\ell^{(k)}\, B && \text{(accumulate local contributions from anchors } \ell\text{)}\\
\alpha_{\ell\to j} &:= b_\ell^2\, c_{\ell,j}^\top D_\ell^\top D_\ell\, c_{\ell,j} = b_\ell^2\, c_{\ell,j}^\top D_\ell^2\, c_{\ell,j} && (D_\ell = I - \xi_\ell\xi_\ell^\top \text{ is symmetric})\\
c_{\ell,j}[k] &= \begin{cases} -1/h_k, & j = \ell,\\ +1/h_k, & j = \ell + s_k,\\ 0, & \text{otherwise,} \end{cases} && \text{(d-neighbour forward-difference stencil)}\\
D_\ell^2 &= (I - \xi_\ell\xi_\ell^\top)^2 = I - (2 - \|\xi_\ell\|_2^2)\,\xi_\ell\xi_\ell^\top && \text{(exact; since } (\xi\xi^\top)^2 = (\xi^\top\xi)\xi\xi^\top\text{)}\\
\alpha_{\ell\to j} &= b_\ell^2\left(\|c_{\ell,j}\|_2^2 - (2 - \|\xi_\ell\|_2^2)\,(\xi_\ell^\top c_{\ell,j})^2\right) && \text{(cheap: only } \|c\|^2,\ \xi^\top c, \text{ and } \|\xi\|^2\text{)}\\
B &:= \mathrm{diag}(b_1, b_2), \qquad W_\ell^{(k)} \in \mathbb{R}^{2\times 2} && \text{(per-modality weights } b_m, \text{ and IRLS coupling } W_\ell^{(k)}\text{).}
\end{aligned}
$$