Stochastic Variance Reduction Optimisation Algorithms Applied to Iterative PET Reconstruction

Robert Twyman

Email: robert.twyman.18@ucl.ac.uk

Motivation

- Standard subset iterative PET reconstruction algorithms suffer from limit cycle behaviour and non-convergence

- This behaviour can lead to significant variations between sequential image updates

- Stochastic variance reduction algorithms can reduce the impact of these variations by computing a pseudo-full gradient at each update

Background

Objective Function

$$\Phi(x) = L(x;y) - \beta R(x)$$

x – Image estimate
y – Measured data
β – Penalty strength

Log-Likelihood

$$L(x;y) = y\log(\bar{y}) - \bar{y}$$
$$\bar{y} = Ax + b$$

$\bar{y}$ – Expected data

A – System matrix

b – Background

Relative Difference Prior[1]

$$R(x) = \sum_{i}^{n_v} \sum_{l \in \mathcal{N}_i,\, l>i} \hat{\kappa}_i \ \hat{\kappa}_l \ \frac{(x_i - x_l)^2}{x_i + x_l + \gamma|x_i - x_l|}$$

$\hat{\kappa}_i$ $\hat{\kappa}_l$ - spatially variant penalty strengths[2]
γ – edge preservation hyperparameter (γ=2)

[1] Nuyts et. al (2002), [2] Tsai et al (2020)

Update Equation

$$x_{k+1} = P_{+}\left(x_k + \alpha_k D_m(x_k)\nabla\Phi_m(x_k)\right)$$

m – subset number
k – iteration number
$P_{+}(\cdot)$ – non-negativity constraint

Subset Gradient

$$\nabla\Phi_m(x) = A_m^{\mathrm{T}}\left(\frac{y_m}{A_m x + b_m} - 1\right) - \frac{\beta}{M}\nabla R(x)$$

Example

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

BSREM[1,2]

$\mathrm{diag}(\cdot)$ – an operator to construct a diagonal matrix

[1]*De Pierro et. al. (2001),* [2]*Ahn et. al. (2003)*

Stochastic Variance Reduction Algorithms

Algorithm Properties

The gradient approximation $\widetilde{\nabla}_m(x_k)$ is computed using:

- The subset gradient $\nabla\Phi_m(x_k)$, and
- Previously computed subset gradients
  - One historical subset gradient for each subset
  - Requires the storage of *M* gradients

Each update has the approx. same computation cost as one subset gradient computation

Define an epoch as equivalent computation to a full forward and backward projection of the data

[1] *Chambolle, et al (2018)*

Three Stochastic Algorithms

SAG (*Stochastic Average Gradient*)

Roux, N. Le, et.al. (2012), Schmidt, et. al. (2017)

- A **low variance** estimate of the gradient
- After **each update**, replace stored gradient for m^{th} subset with $\nabla\Phi_m(x_k)$

SAGA

Defazio, A., el. al. (2014)

- An **unbiased** gradient estimate
- After **each update**, replace stored gradient for m^{th} subset with $\nabla\Phi_m(x_k)$

SVRG (*Stochastic Variance Reduction Gradient*)

Johnson, R., & Zhang, T. (2013)

- An **unbiased** gradient estimate
- **Periodically*** resets subset gradient history by **recomputing** every subset gradient

UCL

Adaption to PET

Preconditioner

$$D(x) = \mathrm{diag}\left(\frac{x+\delta}{A^T 1}\right)$$

δ – is small constant

Non-negativity constraint

Any voxel values < 0 after the update are projected to 0

$$x_{k+1} = P_+\left(x_k + \alpha_k D(x_k)\tilde{\nabla}_m(x_k)\right)$$

Stochastic Algorithms

SAG, SAGA, and SVRG

Subsets

Using the OS methodology[1], but randomly selecting subset number at each update

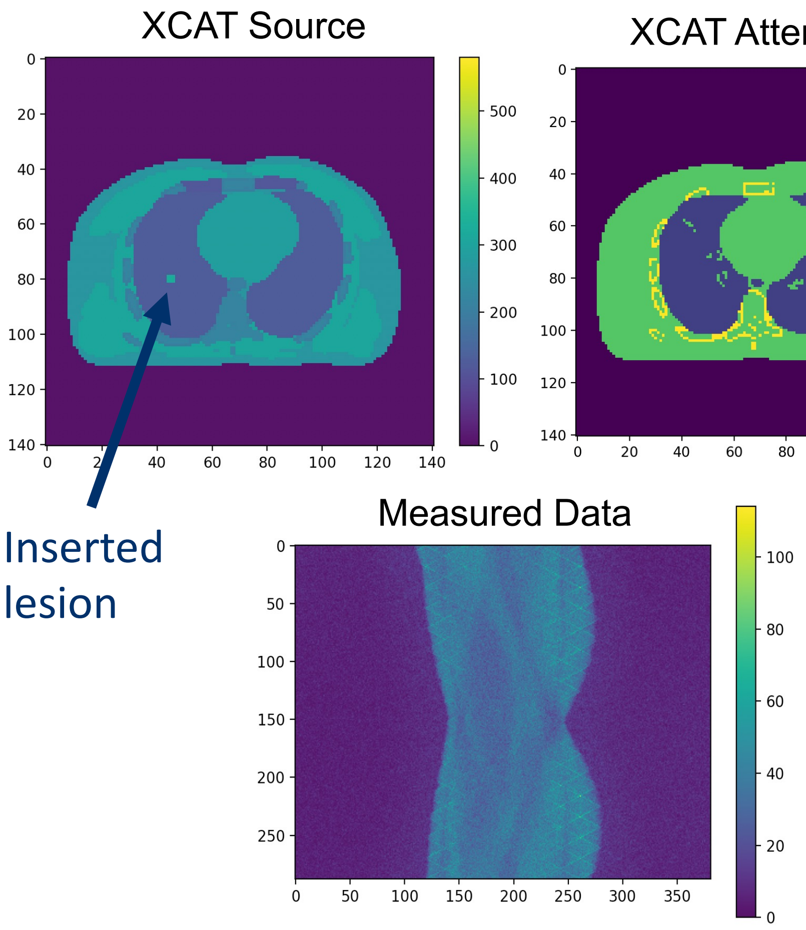
Step Size

Constant α_k=1

[1]*Hudson et al. (1994)*

Experiments

Simulated Phantom Data



*Open*GATE[1] Monte Carlo simulations
Back-to-Back emission without radioactive decay

GE PET/CT Discovery 690 Scanner[2]

Source and attenuation map of a 3D XCAT[3] torso phantom with, inserted lung lesion

Coincidence events recorded in list mode (~1.2B events), unlisted into non-TOF sinograms

STIR[4]: Scatter correction, Randoms correction (from delayed coincidence events), and Normalization

Using the STIR-GATE-Connection:
https://github.com/UCL/STIR-GATE-Connection

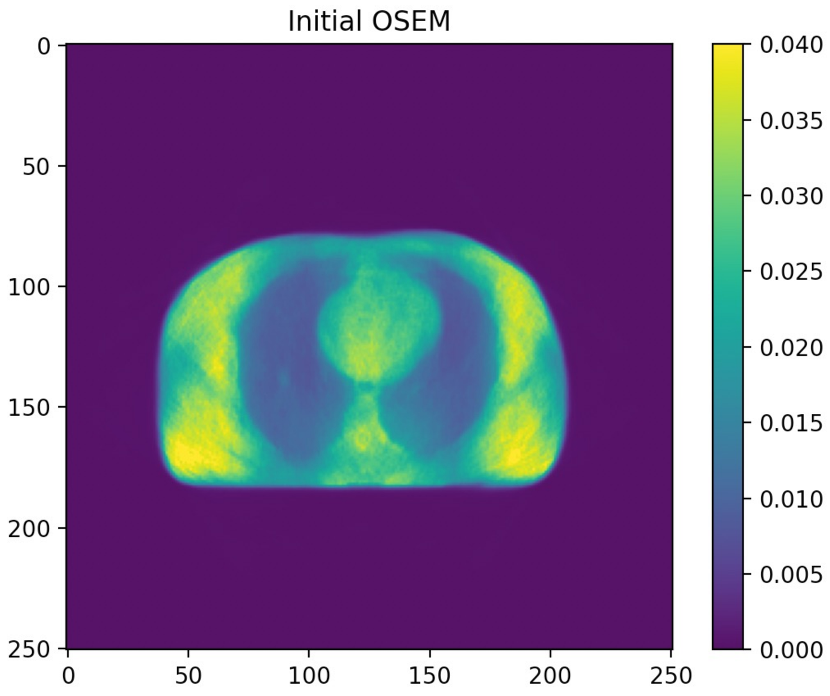
[1]Jan et al (2004), [2]Bettinardi et al (2011), [3]Segars et. al (2010), [4]Thielemans et al (2012)

Warm Starting

x_{init} is computed by OSEM 24 subsets for 1 epoch (24 updates)

Two reasons for warm starting the stochastic algorithms:

- **Stochastic algorithms** are sensitive to initial conditions
- **Spatially variant penalty strength** $\hat{\kappa}$ can be used for free[1]



[1] *Tsai et al (2020)*

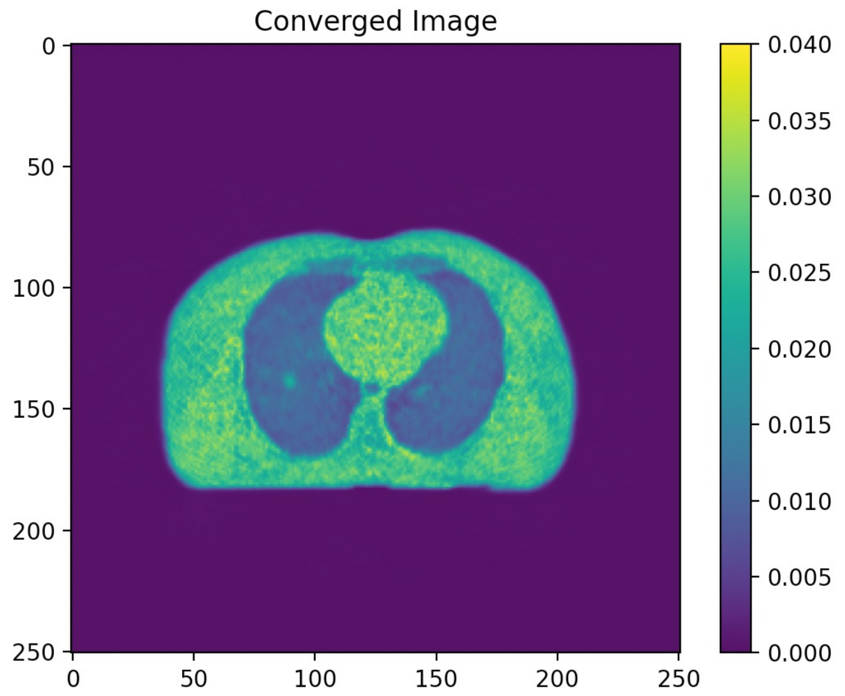
Evaluation strategy

- Objective function concavity a unique solution $\hat{x}$ exists

- Each image update is compared to $\hat{x}$
  Computed with:
  - 1000 epochs of **SAGA** reconstruction
  - Followed by a line search reconstruction

Evaluations:

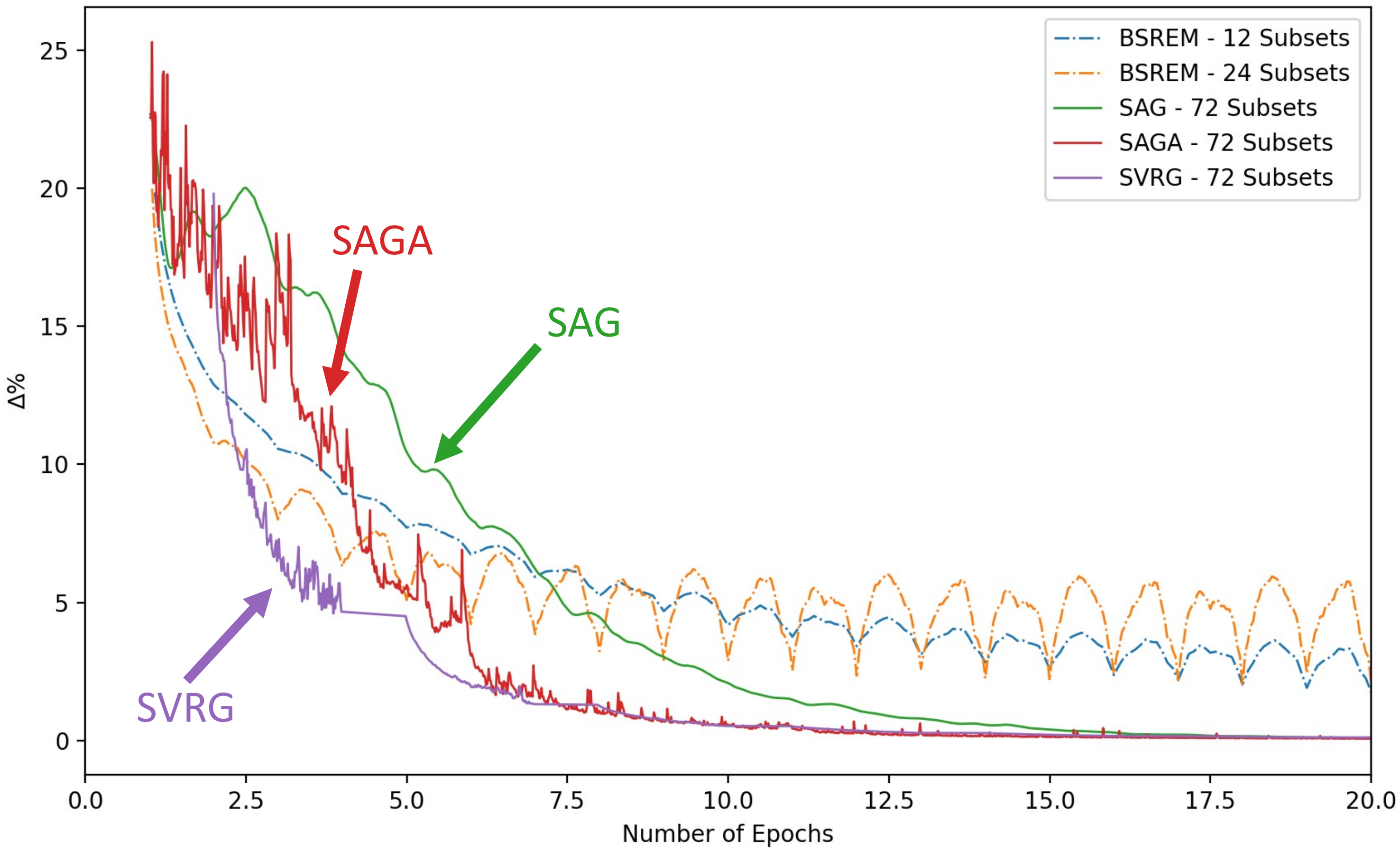
- Visual Assessment
- Distance from convergence $\Delta\% = \frac{|x_k - \hat{x}|_2}{|\hat{x}|_2} \times 100\%$
- Lesion ROI values



Results

Results

Global Performance (distance from converged image)



Δ is a global image performance assessment

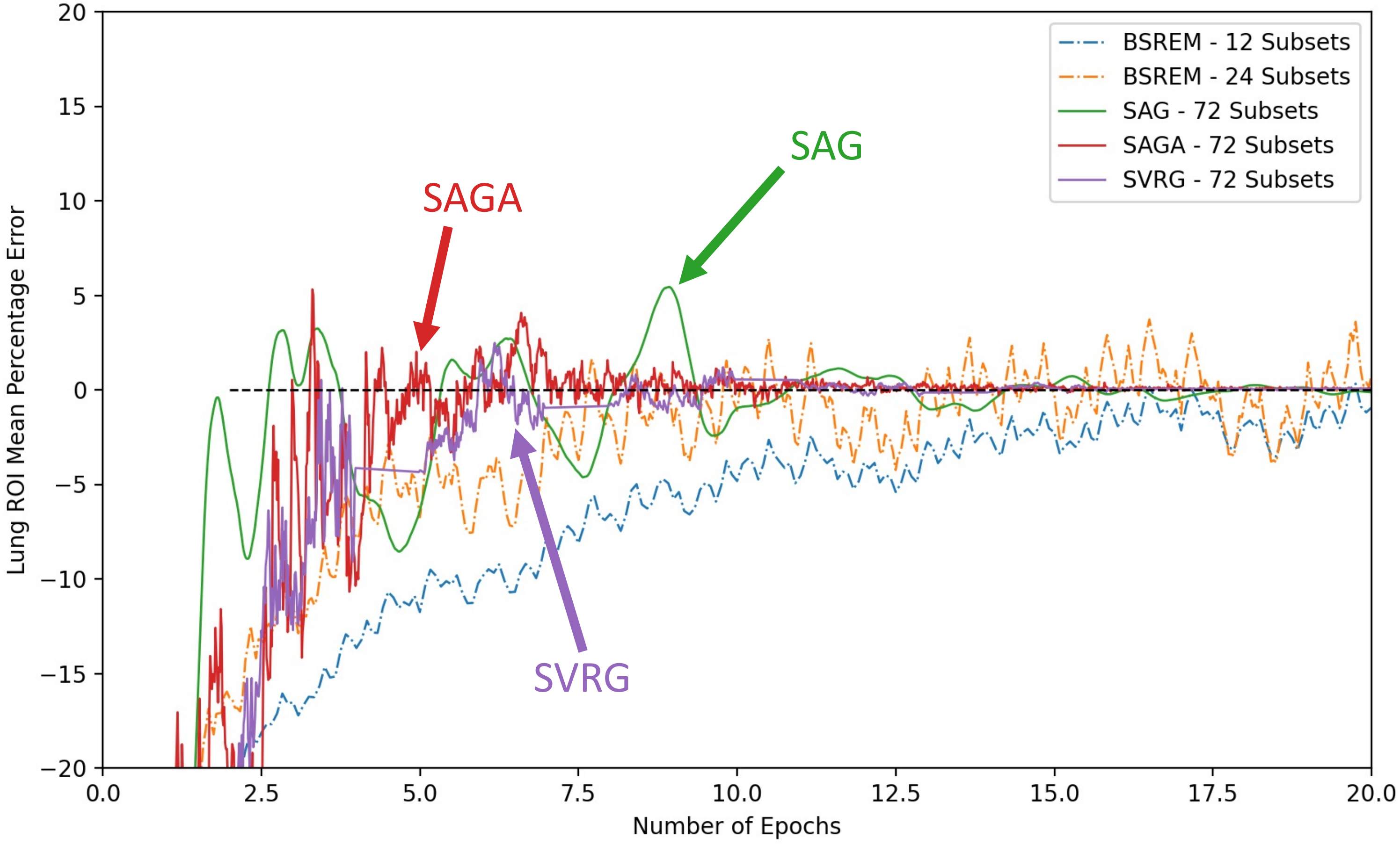
$$\Delta\% = \frac{|x_k - \hat{x}|_2}{|\hat{x}|_2} \times 100\%$$

Stochastic algorithms use 72 subsets

SAG and **SAGA** initial performance is worse than BSREM's

Performance after 5 epochs is significantly better than BSREM

Lung Lesion ROI: Overview



Comparison with the converged image

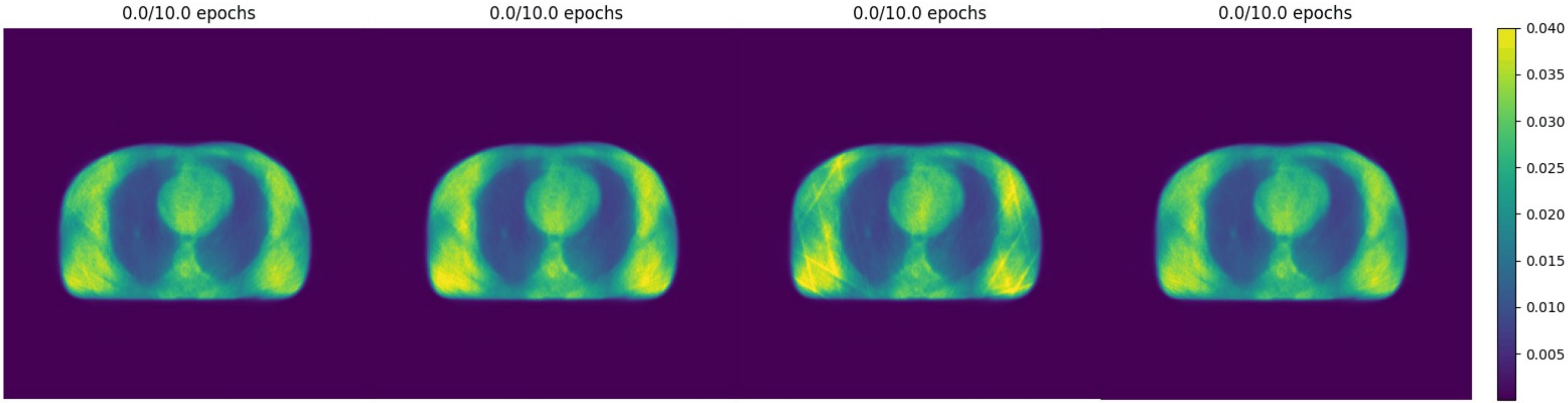
Stochastic algorithms use 72 subset

Some significant variations in the stochastic algorithms

The stochastic algorithm tend towards to 0% error before 20 epochs

Animation of the Reconstructions

BSREM 24 Subsets | SAG 72 Subsets | SAGA 72 Subsets | SVRG 72 Subsets



An epoch is an effective pass through all data

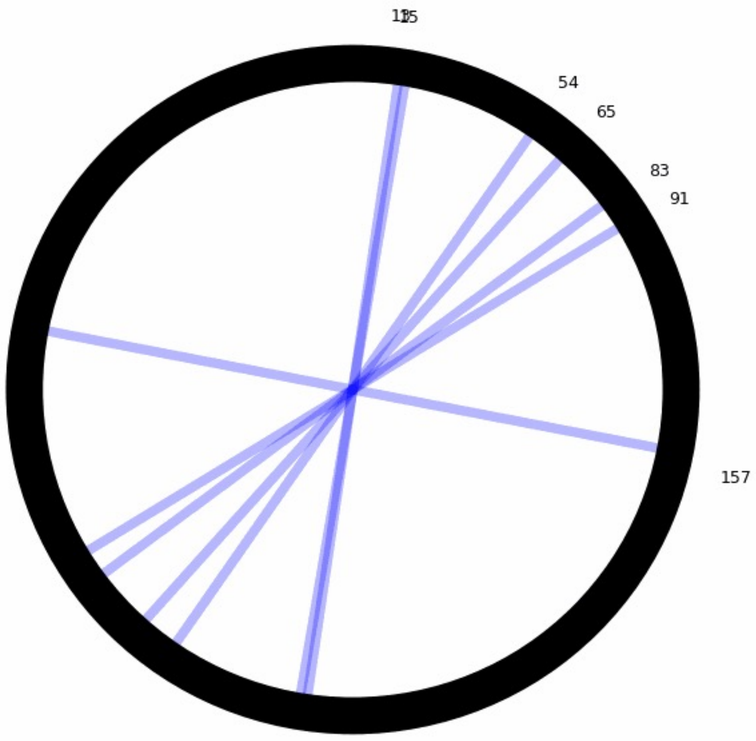
Investigating Subset Sampling Methodologies

Subset Methodologies

Three Subset methods for subset construction and selection:

- Two are stochastic
- One is deterministic

1. Randomised Batches
   - At each iteration, a new subset (size *J/M*) is constructed as randomly selected projection angles
   - No usage of structure



Investigated Subset Methods

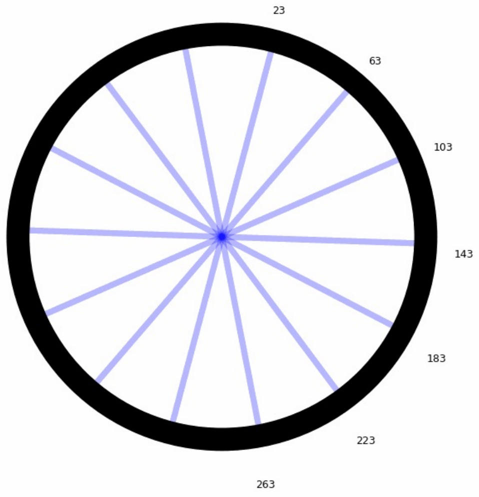
Structured methods

- Algorithm initialisation, construct *M* equally sized subsets
- Each subset are composed of equidistant projection angles
- Sequential subsets are construct from projection angles, with phase *m*
- Projection angles in a subset are as geometrically incoherent as possible from one another

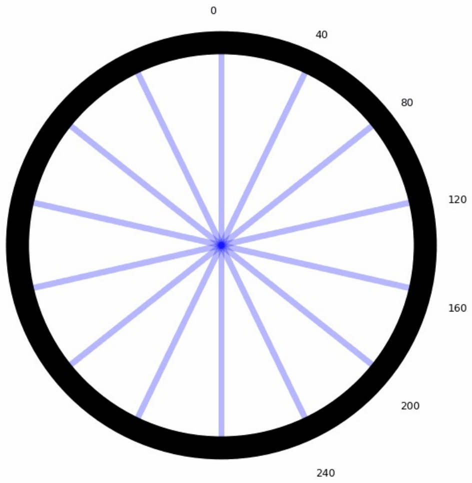
2. Stochastic Subsets
    - At each iteration, randomly select a subset index *m*
    - Some regard for projection angle coherence

3. Ordered Subsets *(Herman & Meyer, 1993/ Hudson & Larkin, 1994)*
    - Create a cyclical deterministic subset sequence to apply a subsets are as orthogonal as possible to the space generated by recently used subsets
    - At each iteration, increment through the cyclical sequence
    - Attempts to apply subsets that are as incoherent as possible to previously applied

Stochastic Subsets

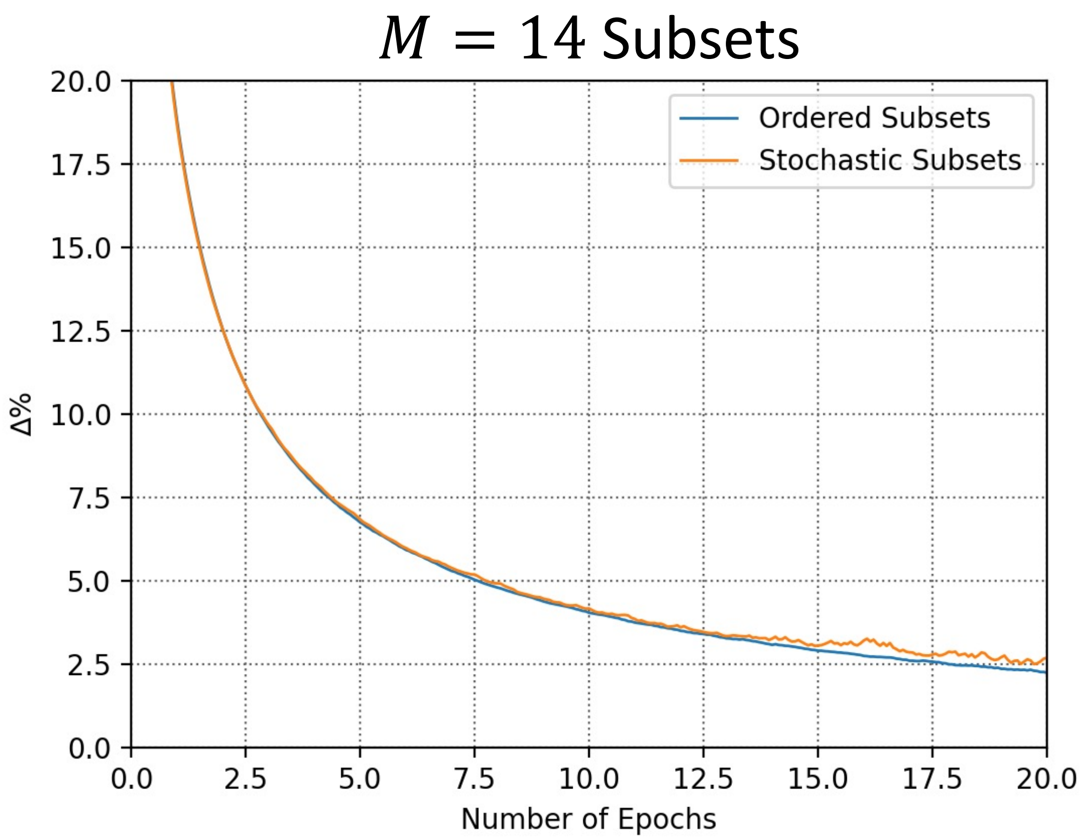


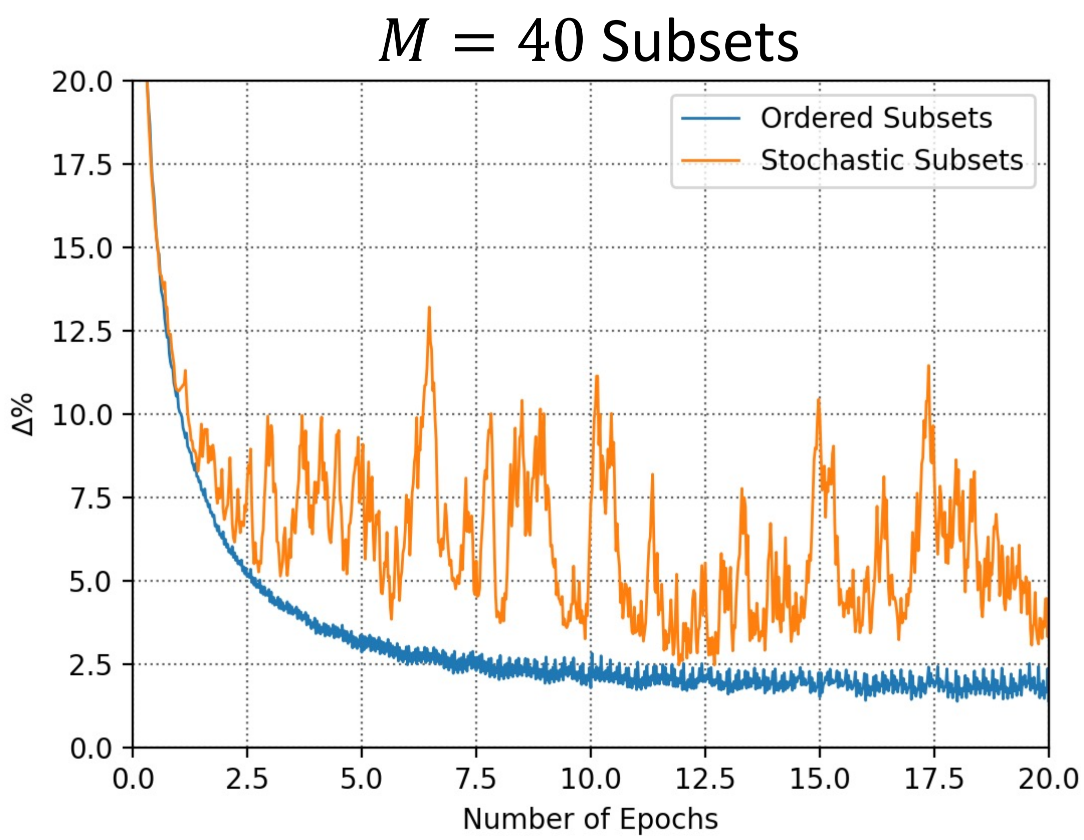
Ordered Subsets



Application to BSREM

An epoch is an effective pass through the data set





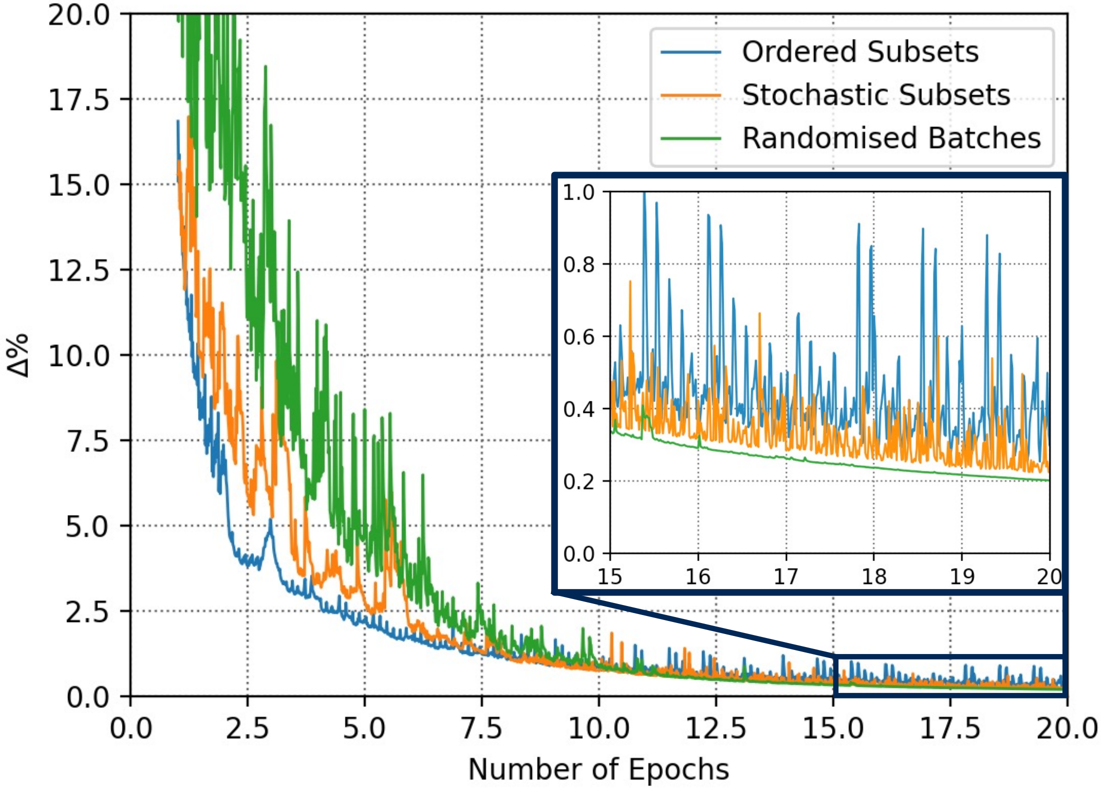
Random Batches is not plotted as due to poor performance

Application to Variance Reduction Methods

$M = 70$ subsets
$\alpha = 1$

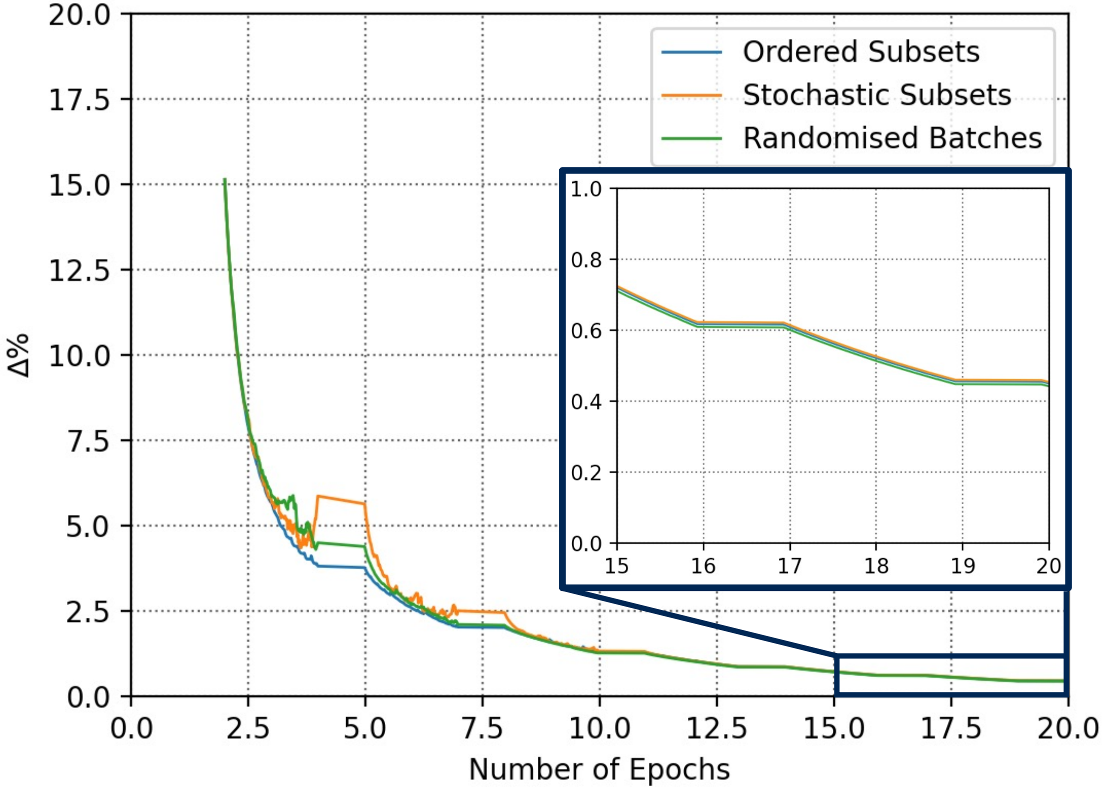
SAGA

$$\widetilde{\nabla}_{k,m}(x_k) = \nabla\Phi_m(x_k) - g_m + \eta$$



SVRG

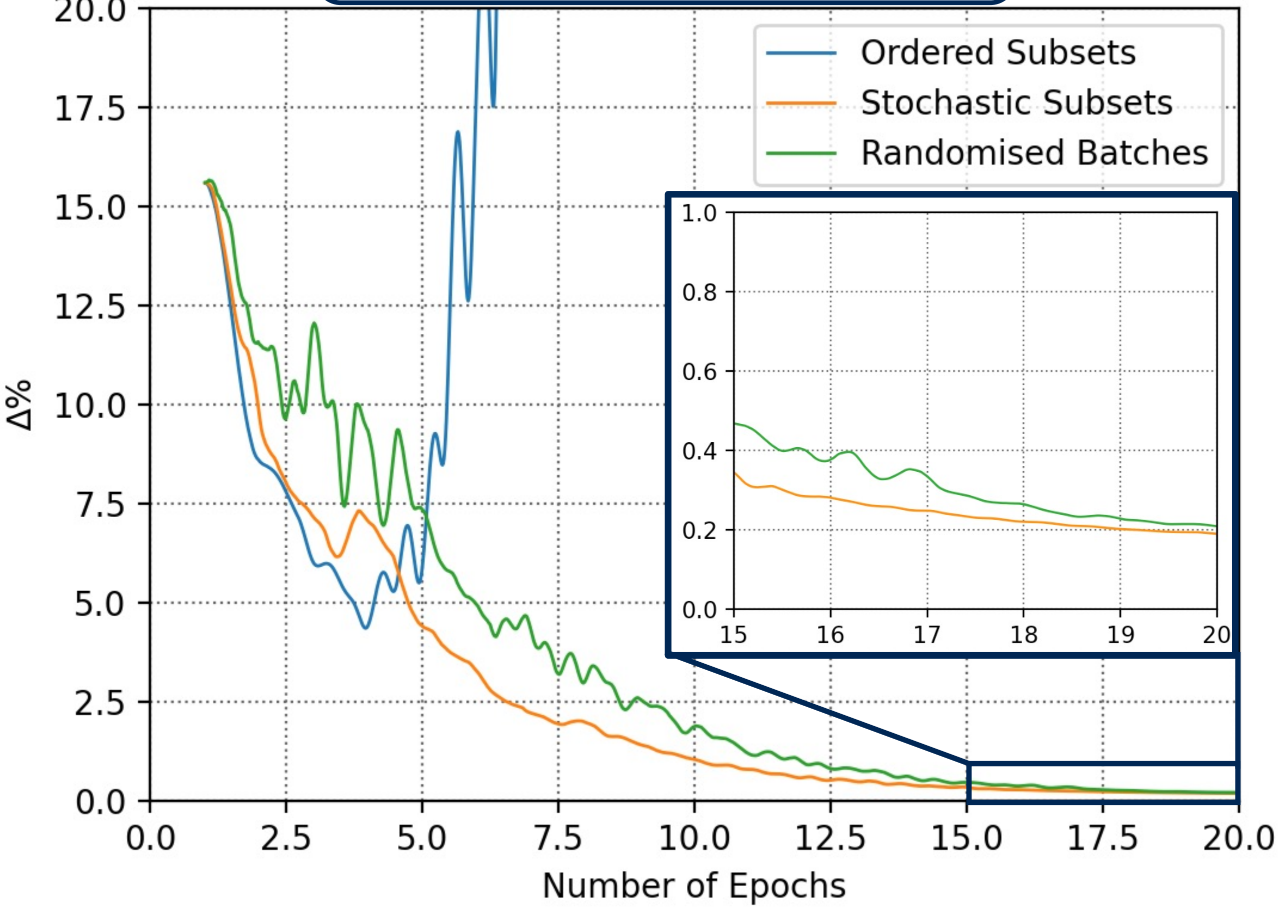
$$\widetilde{\nabla}_{k,m}(x_k) = \nabla\Phi_m(x_k) - g_m + \bar{\mu}$$



Application to Variance Reduction Methods

SAG

$$\widetilde{\nabla}_{k,m}(x_k) = \frac{\nabla\Phi_m(x_k) - g_m}{M} + \eta$$



$M = 70$ subsets
$\alpha = 1$

SAG has a higher dependency on the $\eta = \sum_n^M g_n$ term

Ordered Subsets does not demonstrate this behaviour with a smaller α

Closing Remarks

Conclusion

- The **SAG**, **SAGA**, and **SVRG** algorithms are promising for PET image reconstructions
  - It appears that **SVRG** and **SAGA** perform better than **SAG** in most methods of assessment
  - During early reconstruction performance comparable to BSREM with 12/24 subsets
  - At later epochs (>5), the stochastic algorithms significantly outperform BSREM with no limit cycle behavior
- Future work will apply these algorithms to more datasets and investigating the impact of stochastic subset sampling on the reconstructions

Acknowledgements

A special thanks to:

Kris Thielemans
Simon Arridge
Brian Hutton
Bangti Jin
Ludovica Brusaferri

This research is supported by:

- **GE Healthcare**,
- **NIHR UCLH Biomedical Research Centre**, and
- **EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging** (EP/L016478/1)

Any Questions?

References

- Hudson, H. M., & Larkin, R. S. (1994). Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE Transactions on Medical Imaging*, *13*(4), 601–609. https://doi.org/10.1109/42.363108
- De Pierro, A. R., & Yamagishi, M. E. B. (2001). Fast EM-like methods for maximum “a posteriori” estimates in emission tomography. *IEEE Transactions on Medical Imaging*, *20*(4), 280–288. https://doi.org/10.1109/42.921477

- De Pierro, A. R., & Yamagishi, M. E. B. (2001). Fast EM-like methods for maximum “a posteriori” estimates in emission tomography. *IEEE Transactions on Medical Imaging*, *20*(4), 280–288. https://doi.org/10.1109/42.921477

- Ahn, S., & Fessler, J. A. (2003). Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Transactions on Medical Imaging*, *22*(5), 613–626. https://doi.org/10.1109/TMI.2003.812251
- Nuyts, J., Beque, D., Dupont, P., & Mortelmans, L. (2002). A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Transactions on Nuclear Science*, *49*(1), 56–60. https://doi.org/10.1109/TNS.2002.998681
- Tsai, Y.-J., Schramm, G., Ahn, S., Bousse, A., Arridge, S., Nuyts, J., Hutton, B. F., Stearns, C. W., & Thielemans, K. (2020). Benefits of Using a Spatially-Variant Penalty Strength With Anatomical Priors in PET Reconstruction. *IEEE Transactions on Medical Imaging*, *39*(1), 11–22. https://doi.org/10.1109/TMI.2019.2913889
- Herman, G. T., & Meyer, L. B. (1993). Algebraic Reconstruction Techniques Can Be Made Computationally Efficient. *IEEE Transactions on Medical Imaging*, *12*(3), 600–609. https://doi.org/10.1109/42.241889

References

- Chambolle, A., Ehrhardt, M. J., Richtárik, P., & Schönlieb, C.-B. (2018). Stochastic Primal-Dual Hybrid Gradient Algorithm with Arbitrary Sampling and Imaging Applications. *SIAM Journal on Optimization*, *28*(4), 2783–2808. https://doi.org/10.1137/17M1134834
- Roux, N. Le, Schmidt, M., & Bach, F. (2012). A Stochastic Gradient Method with an Exponential Convergence Rate for Finite Training Sets. *Nips*, *1*(2), 1–34. http://arxiv.org/abs/1202.6258
- Schmidt, M., Le Roux, N., & Bach, F. (2017). Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, *162*(1–2), 83–112. https://doi.org/10.1007/s10107-016-1030-6

- Schmidt, M., Le Roux, N., & Bach, F. (2017). Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, *162*(1–2), 83–112. https://doi.org/10.1007/s10107-016-1030-6

- Johnson, R., & Zhang, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. *Advances in Neural Information Processing Systems*.
- Jan, S., Benoit, D., Becheva, E., Lin, H., Chuang, K., Lin, Y., Pet, C., España, S., Herraiz, J. L., Vicente, E., Lazaro, D., Buvat, I., & Loudos, G. (2004). GATE : a simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*, *49*, 4543–4561.

- Jan, S., Benoit, D., Becheva, E., Lin, H., Chuang, K., Lin, Y., Pet, C., España, S., Herraiz, J. L., Vicente, E., Lazaro, D., Buvat, I., & Loudos, G. (2004). GATE : a simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*, *49*, 4543–4561.
- Jan, S., Benoit, D., Becheva, E., Lin, H., Chuang, K., Lin, Y., Pet, C., España, S., Herraiz, J. L., Vicente, E., Lazaro, D., Buvat, I., & Loudos, G. (2004). GATE : a simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*, *49*, 4543–4561.

- Thielemans, K., Tsoumpas, C., Mustafovic, S., Beisel, T., Aguiar, P., Dikaios, N., & Jacobson, M. W. (2012). STIR: software for tomographic image reconstruction release 2. *Physics in Medicine and Biology*, *57*(4), 867–883. https://doi.org/10.1088/0031-9155/57/4/867