Stochastic Multiresolution Image Sketching for Inverse Imaging Problems

Alessandro Perelli

Symposium on AI and Reconstruction for Biomedical Imaging

with Matthias J. Ehrhardt

09 March 2026

Outline

Summary of Content

- Inverse Problem and Model-based Optimization
- Dimensionality Reduction
- Image-domain Low-Resolution Sketching
- `ImaSk`: Randomized Algorithm for Saddle-Point Problems
- Numerical simulations for CT problem
- Variant `ImaSk-Seq`

Model-based Optimization

Noisy Linear model

$$\mathbf{b} = \mathbf{K}\mathbf{x} + \mathbf{n}$$

- $\mathbf{x} \in \mathbb{R}^d$: spatial-domain image, sometimes non-negative
- $\mathbf{b} \in \mathbb{R}^m$: noisy measurements
- $\mathbf{K} \in \mathbb{R}^{m \times d}$: physical forward model (application dependent)
- $\mathbf{n} \sim \mathcal{N}(0, \sigma^2) \in \mathbb{R}^m$ additive Gaussian noise

Applications: Imaging (CT, MRI, Radar); different noise models

Model-based Maximum a Posteriori (MAP) optimization problem

Estimate the unknown high-dimensional vector $\mathbf{x}$ from the measurements $\mathbf{b}$

$$\min_{\mathbf{x} \in \mathbb{R}^d_+} \frac{1}{2} \|\mathbf{K}\mathbf{x} - \mathbf{b}\|^2 + g(\mathbf{x})$$

- $g(\cdot)$: regularization term which enforce a particular structure on $\mathbf{x}$

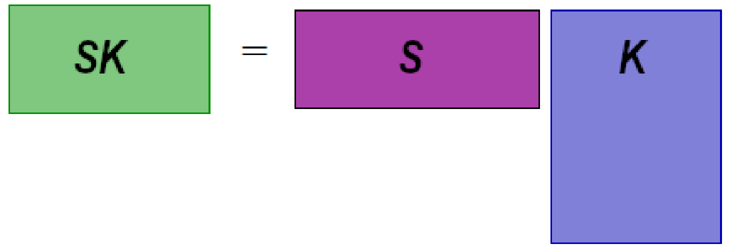
Dimensionality Reduction

Computational bottleneck: first-order iterative solvers require matrix-vector multiplications with $\mathbf{K}$ and its adjoint $\mathbf{K}^*$s.

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{Kx} - \mathbf{b}\|^2 \Rightarrow \nabla f(\mathbf{x}) = \mathbf{K}^*(\mathbf{Kx} - \mathbf{b}) = \mathbf{K}^*\mathbf{Kx} - \mathbf{K}^*\mathbf{b}$$

Subspace embedding[1]

$\mathbf{S} \in \mathbb{R}^{s\times m}$ is an $(1+\epsilon)$ l_p subspace embedding for some fixed matrix $\mathbf{K} \in \mathbb{R}^{m\times d}$ if, for all $\mathbf{x} \in \mathbb{R}^d$, it holds that $(1-\epsilon)\|\mathbf{Kx}\|_2 \leq \|\mathbf{SKx}\|_2 \leq (1+\epsilon)\|\mathbf{Kx}\|_2$



[1] Wang, S., Gittens, A. and Mahoney, M.W., Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging, ICML 2017.

Proposed idea for High-dimensional Inverse Problems

Develop iterative solvers with low computational cost

- **Image domain sketches:** the forward model is approximated by a family of deterministic multiresolution operators.
- **Solver**: algorithm uses at each iteration one of these multiresolution operators at random.

Challenges

- how to design multi-resolution operators
- reformulate the original problem as a primal-dual optimization problem

Intuition: Image-domain Sketching

Aim to exploit sketching in the image domain

- define a family of operators $\{\mathbf{S}_i\}$ which significantly make the evaluation of $\mathbf{Kx}$ computationally cheaper, i.e., $\mathbf{KS}_i\mathbf{x}$ is cheaper than $\mathbf{Kx}$,
- new operators do not introduce bias in the solution, i.e., on average they cancel another,

$$\mathbb{E}_i \mathbf{S}_i = \sum_{i=1}^{r} p_i \mathbf{S}_i = \mathbf{I},$$

- The operators $\mathbf{K}_i = \mathbf{KS}_i$ are similar to $\mathbf{K}$, i.e., $\mathbf{Kx} \approx \mathbf{KS}_i\mathbf{x}$.

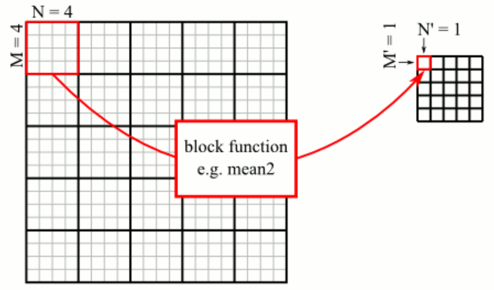
Challenges

- how to design image domain sketches
- solving the optimization problem with the new structure

Construction Low-Resolution Mappings

Decimators - Interpolators

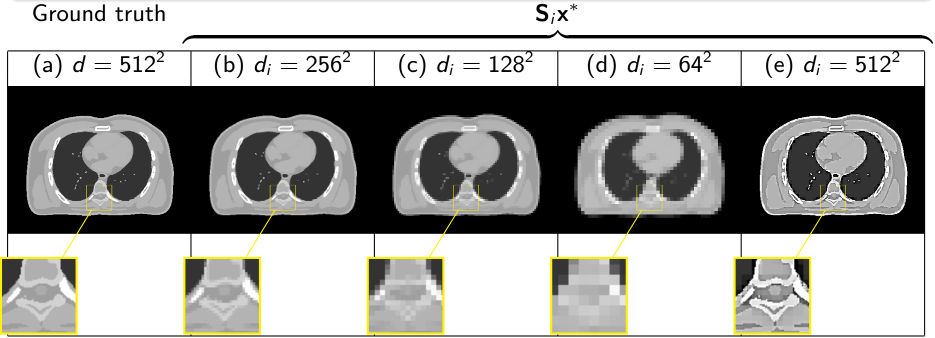
- linear mappings from full to low resolution grids in the image domain.
- decimator operators $\mathbf{T}_i : \mathbb{R}^d \to \mathbb{R}^{d_i}$, $i = 1, \ldots, r-1$ from the full-resolution r, images $\sqrt{d} \times \sqrt{d}$, to low-resolution $\mathbf{T}_i^* \mathbf{x}$ images.
- performing downsampling of the factor $2^i \times 2^i$ by block averaging.
- $\mathbf{S}_i = \mathbf{T}_i^* \mathbf{T}_i$ is proportional to the projection onto the subspace of piece-wise constant images.
- $\mathbf{T}_i$ a family of low resolution mappings with $i \in [r-1]$ which are constructed as doing down-sampling by 2^i-block averaging of the input image.



Visualization Low-Resolution Mappings

Illustration of multiresolution sketches

The sketches ($r = 4$) are averaging over small neighbourhoods. (a) Ground truth image $\mathbf{x}^*$ at full resolution $d = 512^2$, (b, c, d) increasingly lower resolution mappings $\mathbf{S}_i\mathbf{x}^*$ and (e) resolution compensation.



The original image on the left is approximated by lower and lower resolutions. This is compensated by the image on the right which is a sharpened version of the ground truth.

Design Image Sketched Linear Operator

Given a full resolution image $\mathbf{x} \in \mathbb{R}^d$ and bounded linear operator $\mathbf{K}$, define a family of low resolution linear sketches $\mathbf{K}_i : \mathbb{R}^d \to \mathbb{R}^m, i = 1, \dots, n-1$

$$\mathbf{K}_i\mathbf{x} = \begin{cases} \mathbf{K}\mathbf{S}_i\mathbf{x} = \underbrace{\mathbf{K}\mathbf{T}_i^*}_{\mathbf{R}_i}\mathbf{T}_i\mathbf{x} = \mathbf{R}_i\mathbf{T}_i\mathbf{x} & , \quad i = 1, \dots, r-1 \\ \mathbf{K}\mathbf{S}_r\mathbf{x} & , \quad i = r \end{cases}$$

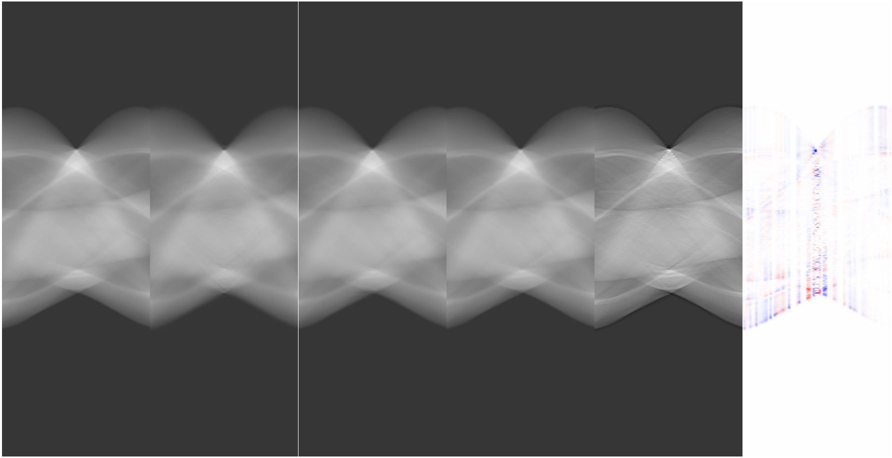
- **low resolution** $i = 1, \dots, n-1$**:** $\mathbf{K}\mathbf{T}_i^*$ can be approximated with $\mathbf{R}_i$ which is the forward model from the low resolution $\sqrt{d_i} \times \sqrt{d_i}$ input grid to the m measurements; sketch $\mathbf{S}_i = \mathbf{T}_i^*\mathbf{T}_i$
- the up-sampling $\mathbf{T}_i^*$ is never explicitly computed.

- **full resolution** $i = r$**:** linear mapping $\mathbf{T}_r^*\mathbf{T}_r$ is derived from the constraint

$$\sum_{i=1}^{r-1} p_i\mathbf{T}_i^*\mathbf{T}_i + p_r\mathbf{S}_r = \mathbf{I} \quad \Rightarrow \quad \mathbf{S}_r = p_r^{-1}\left(\mathbf{I} - \sum_{i=1}^{r-1} p_{L_i}\mathbf{T}_i^*\mathbf{T}_i\right)$$

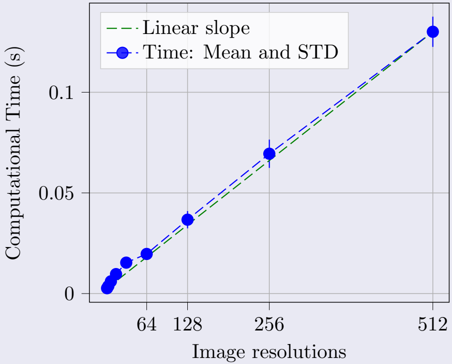
- p_i probability of selecting low resolution sketches
- $p_r = 1 - \sum_{i=1}^{r-1} p_i$ probability full resolution operator

Visualization CT Data Decomposition $\mathbf{K}_i\mathbf{x}$ CT



Computational Analysis CT operators in ASTRA

- The operator **K** scales linearly in the image dimension $\sqrt{d}$.
- Monte–Carlo simulations: computational time CT operator $\mathbf{R}_i\mathbf{x}$ in ASTRA for different input resolutions $\sqrt{d_i} = 2^i, \; i = 1, \ldots, 10$ and linear interpolation.



[1] Van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K.J. and Sijbers, J., Fast and flexible X-ray tomography using the ASTRA toolbox, Optics express, 2016.

Image-domain Sketching Optimization Problem

- By design $\mathbf{K}\mathbf{x} = \sum_{i=1}^{r} \mathbf{K}_i\mathbf{x}$
- The computational cost of $\mathbf{K}_i\mathbf{x}$ compared to $\mathbf{K}\mathbf{x}$ is about $\sqrt{d}/\sqrt{d_i}$ cheaper.

Optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}_+^d} \frac{1}{2}\left\|\sum_{i=1}^{r} p_i\mathbf{K}_i\mathbf{x} - \mathbf{b}\right\|^2 + g(\mathbf{x})$$

- only evaluate one $\mathbf{K}_i$ per iteration
- $f(\mathbf{K}\mathbf{x}, \mathbf{b}) = \frac{1}{2}\|\mathbf{K}\mathbf{x} - \mathbf{b}\|^2$ and $g : \mathbb{R}^d \to \mathbb{R}$ are convex functions
- objective function is not linear in the sampling probability p_i
- prevents to use traditional SGD solvers which require the objective update iteratively by using "unbiased" samples of the iterates' gradients

ImaSk: Randomized Algorithm for Saddle-Point Problems

Reformulate the original minimization as a saddle-point problem: regime where linearity property is preserved

Fenchel convex conjugate function $f^* : \mathbb{R}^m \to \mathbb{R}$

Let f be a proper, convex function, then

$$f^*(\mathbf{y}) := \sup_{\mathbf{z} \in \mathrm{dom} f} \Big(\langle \mathbf{z}, \mathbf{y} \rangle - f(\mathbf{z}) \Big).$$

Equivalent saddle-point problem

Using f^*, solving original problem is equivalent to finding the primal $\mathbf{x} \in \mathbb{R}^d$ of a solution to the saddle point problem

$$\min_{\mathbf{x} \in \mathbb{R}^d_+} \max_{\mathbf{y} \in \mathbb{R}^m} \sum_{i=1}^{n} \langle \mathbf{y}, \mathbf{K}_i \mathbf{x} \rangle - f^*(\mathbf{y}) + g(\mathbf{x}).$$

different from Stochastic Primal-Dual Hybrid Gradient algorithm (SPDHG) where f and $\mathbf{y}$ are indexed by i.

ImaSk: SAGA solver for saddle-point optimization

Require: number of iterations K, probabilities $p = (p_i)_{i=1}^r \in (0,1]^r$, step size σ, strong convexity of regularizer μ

Input: initial iterate $\mathbf{x}^0 = \mathbf{0}$, $\mathbf{y}^0 = \mathbf{0}$, memory $\phi^0 = \mathbf{0}$, $\psi^0 = \mathbf{0}$

Output: $\mathbf{x}^K$

for $k = 0, \ldots, K-1$ **do**

Sample $i_k \in \{1, \ldots, r\}$ at random with probability p.

$$\phi_i^{k+1} = \begin{cases} \mathbf{K}_{i_k}^T \mathbf{y}^k & \text{if } i = i_k \\ \phi_i^k & \text{else} \end{cases}$$

$$\psi_i^{k+1} = \begin{cases} \mathbf{K}_{i_k} \mathbf{x}^k & \text{if } i = i_k \\ \psi_i^k & \text{else} \end{cases}$$

$$\boldsymbol{\xi}^k = \phi_{i_k}^{k+1} - \phi_{i_k}^k + \textstyle\sum_{i=1}^r p_i \phi_i^k$$

$$\boldsymbol{\zeta}^k = \psi_{i_k}^{k+1} - \psi_{i_k}^k + \textstyle\sum_{i=1}^r p_i \psi_i^k$$

$$\mathbf{x}^{k+1} = \operatorname{prox}_{\sigma \mu_g^{-1} g} \left(\mathbf{x}^k - \sigma \mu_g^{-1} \boldsymbol{\xi}^k \right)$$

$$\mathbf{y}^{k+1} = \operatorname{prox}_{\sigma \mu_{f^*}^{-1} g} \left(\mathbf{y}^k + \sigma \mu_{f^*}^{-1} \boldsymbol{\zeta}^k \right)$$

end for

ImaSk Convergence

Convergence Theorem

Let $\mathbf{x}^t$ be defined in ImaSk and $\mathbf{x}^*$ the saddle point of the problem. Then after K iterations, for any $a, b > 0$ the iterates satisfy

$$\mathbb{E}\|\mathbf{x}^K - \mathbf{x}^*\|^2 \leq \theta^K C.$$

with the convergence factor

$$\theta := \max\left(\frac{1+\sigma^2(L^2+(1+a)\bar{L}_p^2)}{(1+\sigma)^2} + b\bar{L}^2, \frac{\left(1+a^{-1}\right)\sigma^2}{b(1+\sigma)^2} + 1 - \min_i p_i\right)$$

- $L = \|\mathbf{K}\|$: operator norm of the forward full-resolution matrix $\mathbf{K}$
- $\bar{L}_p = L \cdot N_T$ where $N_T = \left\|\left[\mathbf{T}_{i_1}^* \mathbf{T}_{i_1}, \ldots, \mathbf{T}_{i_{r-1}}^* \mathbf{T}_{i_{r-1}}, \mathbf{S}_r\right]\right\|$: operator norm of the stacked operator concatenating the different low resolution sketches
- a, b free scalar parameters and C a constant.

Mono-energetic X-ray CT Transmission Forward Model

Measurements at detector pixel i in the discrete domain

$$S_i \sim \mathrm{Poisson}\left\{ I_0 e^{-[\mathbf{Kx}]_i} \right\}, \quad i = 1, \ldots, m$$

- $[\mathbf{Kx}]_i$ represents the discrete linear projection of the i-th ray
- negligible scatter and other background noise.

Normal dose $\rightarrow$ quadratic approximation of the non-negative log-likelihood function

- Weighted Least Squares (WLS) approximation based on taking the logarithm of the data, $b_i = \log\left(\frac{I_0}{s_i}\right)$
- equivalent to linear measurements $\mathbf{Kx}$ corrupted with a data-dependent Gaussian noise, $\mathcal{N}(0, \mathbf{W}^{-1})$, with inverse covariance $\mathbf{W} = \mathrm{diag}(\mathbf{s})$, i.e.

$$\mathbf{b} = \mathbf{Kx} + \mathcal{N}(0, \mathbf{W}^{-1})$$

Numerical simulations

CT forward operator $\mathbf{K}$ and low resolution sketches $\mathbf{K}_i$ implemented with ASTRA Toolbox using a parallel beam geometry with line discretization.

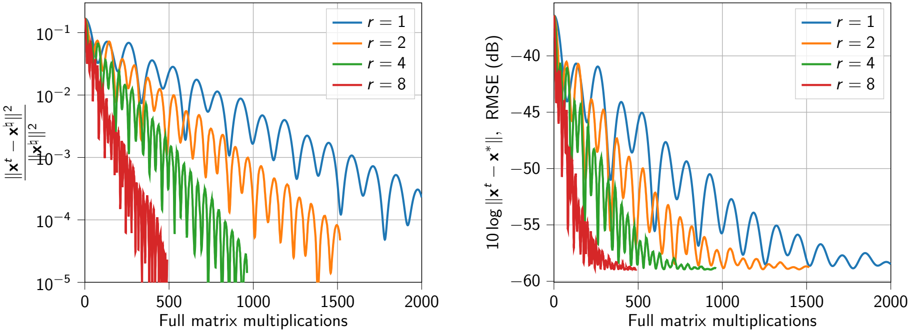
CT acquisition simulation

- $n_\theta = 100$ angular projections
- $n_d = \lceil \sqrt{2} d \rceil$ detectors
- dimension of the measurement vector $\mathbf{b}$ is $n = n_\theta \cdot n_d$
- Poisson noise to the projection data with $I_0 = 10^5$ X-ray photons
- input data: 1-mm pixel-width $512 \times 512\,(d = 512^2)$ torso axial slice images generated from the XCAT phantom and clinical AAPM images.

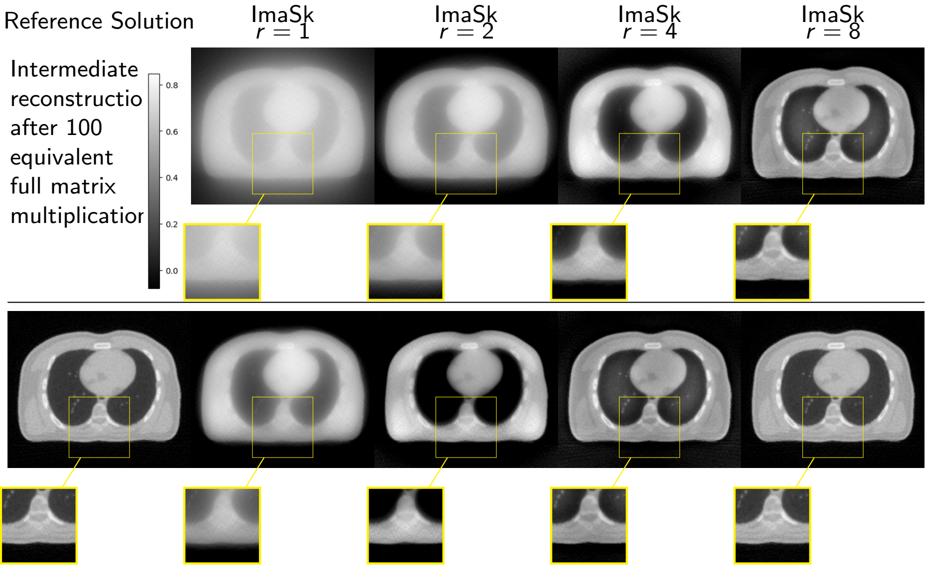
The `ImaSk` algorithm implemented in Python using the Operator Discretization Library (ODL)

Numerical results: L_2 regularization, Uniform probability

- distance of the optimizer $\frac{\|\mathbf{x}^t - \mathbf{x}^\natural\|^2}{\|\mathbf{x}^\natural\|^2}$ with `ImaSk` using uniform probabilities $p_i = \frac{1}{r}$, $i = 1, \ldots, r$ and distance to solution $\mathbf{x}^*$ RMSE

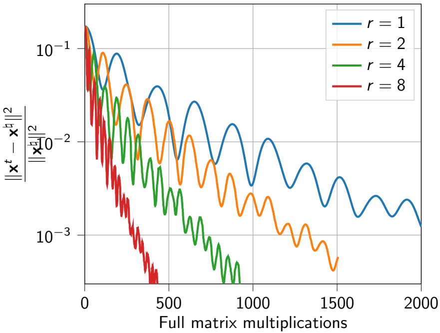


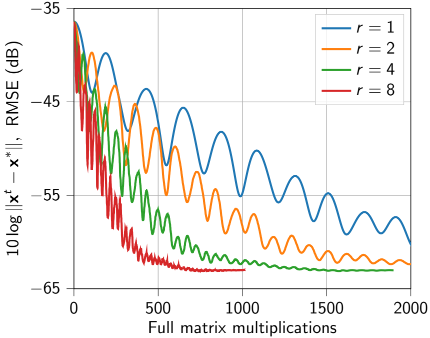
Comparison CT Reconstruction Results with XCAT Image



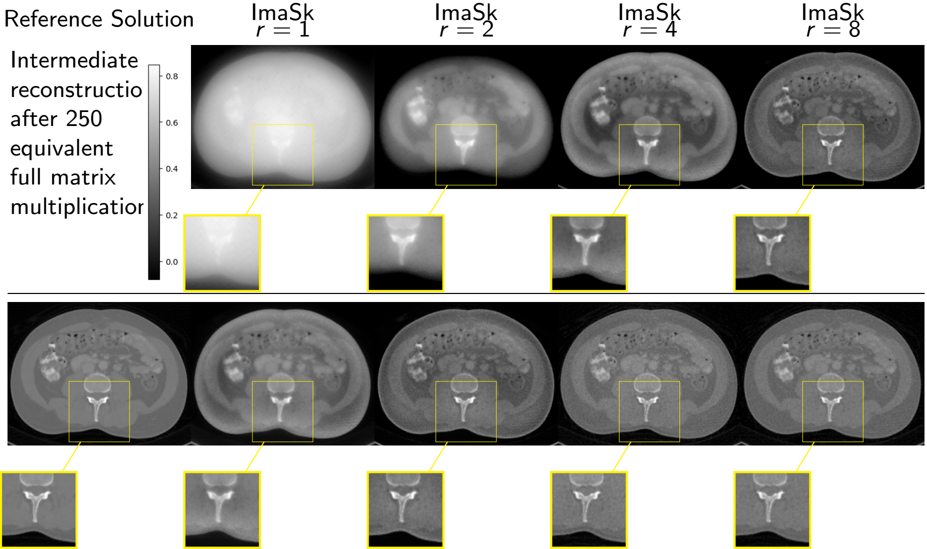
Numerical results: Non smooth regularization (TV)

- distance of the optimizer $\frac{\|\mathbf{x}^t - \mathbf{x}^\natural\|^2}{\|\mathbf{x}^\natural\|^2}$ with `ImaSk` using uniform probabilities $p_i = \frac{1}{r},\ i = 1, \ldots, r$ and distance to solution $\mathbf{x}^*$ RMSE





Comparison CT reconstruction results of the AAPM Images



Variant ImaSk-Seq for CT image reconstruction

Require: number of iterations K, probabilities $p = (p_i)_{i=1}^r \in (0,1]^r$, step size σ, strong convexity of regularizer μ

Input: initial iterate $\mathbf{x}^0 = \mathbf{0}$, $\bar{\mathbf{x}}^0 = \mathbf{0}$, $\mathbf{y}^0 = \mathbf{0}$, memory $\phi^0 = \mathbf{0}$, $\psi^0 = \mathbf{0}$

Output: $\mathbf{x}^K$

for $k = 0, \ldots, K-1$ **do**

Sample $i_k \in \{1, \ldots, r\}$ at random with probability p.

$$\psi_i^{k+1} = \begin{cases} \mathbf{K}_{i_k} \bar{\mathbf{x}}^k & \text{if } i = i_k \\ \psi_i^k & \text{else} \end{cases}$$

$\zeta^k = \psi_{i_k}^{k+1} - \psi_{i_k}^k + \sum_{i=1}^r p_i \psi_i^k$

$\mathbf{y}^{k+1} = (1+\sigma)^{-1}(\mathbf{y}^k + \sigma\zeta^k - \sigma\mathbf{b})$

$$\phi_i^{k+1} = \begin{cases} \mathbf{K}_{i_k}^T \mathbf{y}^{k+1} & \text{if } i = i_k \\ \phi_i^k & \text{else} \end{cases}$$

$\xi^k = \phi_{i_k}^{k+1} - \phi_{i_k}^k + \sum_{i=1}^r p_i \phi_i^k$

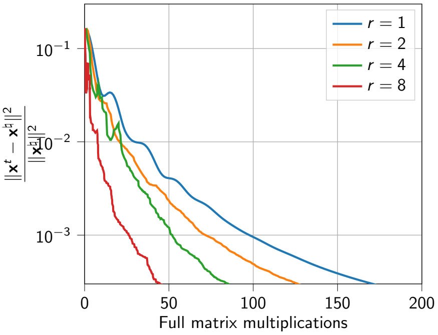
$\mathbf{x}^{k+1} = \operatorname{prox}_{\sigma\mu^{-1}R}\left(\mathbf{x}^k - \sigma\mu^{-1}\xi^k\right)$

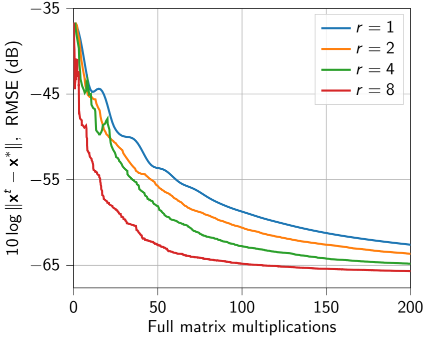
$\bar{\mathbf{x}}^{k+1} = \mathbf{x}^{k+1} + \theta(\mathbf{x}^{k+1} - \mathbf{x}^k)$, Extrapolation

end for

Numerical results: `ImaSk-Seq` (XCAT)

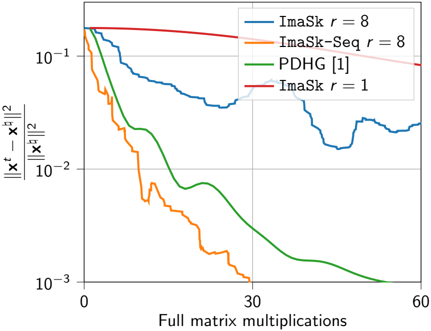
- distance of the optimizer $\frac{\|\mathbf{x}^t - \mathbf{x}^\natural\|^2}{\|\mathbf{x}^\natural\|^2}$ with `ImaSk` using uniform probabilities $p_i = \frac{1}{r}$, $i = 1, \ldots, r$ and distance to solution $\mathbf{x}^*$ RMSE

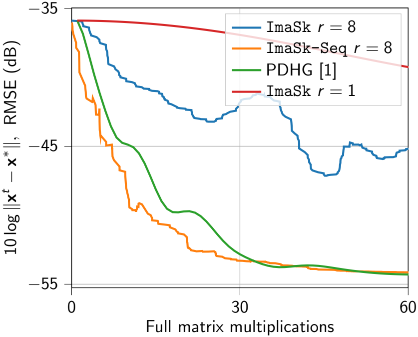




Numerical results: Comparison ImaSk and ImaSk-Seq

- distance of the optimizer $\frac{\|\mathbf{x}^t - \mathbf{x}^\natural\|^2}{\|\mathbf{x}^\natural\|^2}$ with ImaSk using uniform probabilities $p_i = \frac{1}{r}, \; i = 1, \dots, r$ and distance to solution $\mathbf{x}^*$ RMSE





[1] Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of mathematical imaging and vision, 40, pp.120-145, 2011

Conclusion and Future works

- `ImaSk` and `ImaSk-Seq`: iterative solvers based on image domain sketching leads to computation reduction with same solution accuracy
- linear convergence rate
- analysis for smooth regularization and numerical stable also for non-smooth penalty (Total Variation)

Future improvements

- analysis on different probability distributions and 3D data
- extend the framework to non-linear models

Paper available at `https://arxiv.org/abs/2412.10249`

Thanks for your attention

email: alessandro.perelli@glasgow.ac.uk

Step-size Analysis

step-size $\sigma = \eta/(1-\eta)$ parameterized over the new variable η is bounded by

$$\eta = \frac{2}{1 + L^2 + (1+a)\bar{L}^2} < \frac{2}{1 + L^2 + \bar{L}^2}$$

There is a unique optimal $\eta^* \in [0, c]$ that minimizes the convergence factor θ.

Optimal step-size σ (left) and convergence rate θ (right)