Hybrid Projection Methods with Recycling for Large Inverse Problems

Jiahua Jiang* Julianne Chung* Eric de Sturler*

*Department of Mathematics, Virginia Tech

Outline

Outline

1. Background on Hybrid Methods
2. Hybrid Projection Methods with Recycling
3. Numerical Results
4. Conclusions and Future Directions

Linear Inverse Problem

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{true} + \boldsymbol{\epsilon}$$

where

$\mathbf{b} \in \mathbb{R}^M$ - observed data

$\mathbf{x}_{true} \in \mathbb{R}^N$ - desired solution

$\mathbf{A} \in \mathbb{R}^{M \times N} (M \geq N)$ - models the forward processes

$\boldsymbol{\epsilon} \in \mathbb{R}^M$ - noise, statistical properties may be known

Goal: Given $\mathbf{b}$ and $\mathbf{A}$, compute approximation of $\mathbf{x}_{true}$

An Ill-Posed Inverse Problem

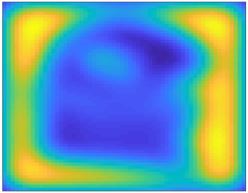
Challenges:

- Solution may not exist
- Solution may not be unique
- Solution may not depend continuously on the data

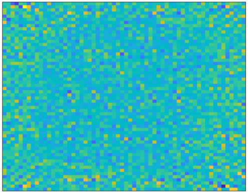
Image deblurring: $\mathbf{x}_{LS} = \mathbf{A}^{-1}\mathbf{b}$

Hokie Bird

Hokie Blurred



Hokie Absurd



Example courtesy of Prof. Mark Embree(VT)

Tikhonov Regularization

Regularization

$$\min_{\mathbf{x}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^2 R(x) \right\}$$

Many choice of $R(x)$

- Tikhonov reguarization(/ridge regression/weight decay)
- Total Variation regularization
- ℓ^1 regularization
- Sparsity regularization

Tikhonov solution

$$\begin{aligned} \mathbf{x}_\lambda &= \arg\min_{\mathbf{x}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^2 \|x\|_2^2 \right\} \\ &= (\mathbf{A}^\top \mathbf{A} + \lambda^2 \mathbf{I}) \mathbf{A}^\top \mathbf{b} \\ &= \mathbf{A}_\lambda^\dagger \mathbf{b} \end{aligned}$$

Choosing Regularization Parameter λ

- Discrepancy Principle (DP): $\|(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{b}\|_2 < \nu_{\text{DP}}\sqrt{N}\sigma$
- Unbiased Predictive Risk Estimator (UPRE) - Mallow (1973)

$$\min_\lambda \frac{1}{N}\|(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{b}\|_2^2 + \frac{2\sigma^2}{N}\text{trace}(\mathbf{A}\mathbf{A}_\lambda^\dagger) - \sigma^2$$

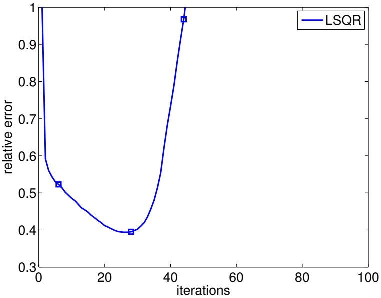
- Generalized Cross Validation (GCV) - Golub, Heath and Wahba (1979)

$$\min_\lambda \frac{N\|(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{b}\|_2^2}{\left[\text{trace}(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\right]^2}$$

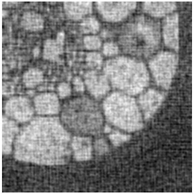
- Weighted Generalized Cross Validation (WGCV) - Golub, Heath and Wahba (1979)

$$\min_\lambda \frac{N\|(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{b}\|_2^2}{\left[\text{trace}(\mathbf{I} - \omega\mathbf{A}\mathbf{A}_\lambda^\dagger)\right]^2}$$

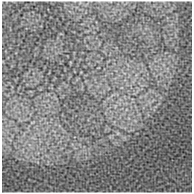
- Apply standard iterative method to least squares problem, $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2$, and terminate early
- Relative error: $\frac{\|\mathbf{x}_m - \mathbf{x}_{true}\|_2}{\|\mathbf{x}_{true}\|_2}$



Iteration 6



Iteration 28



Iteration 44

Given $\mathbf{A}$ and $\mathbf{b}$, initialize $\beta_1 = \|\mathbf{b}\|_2, \mathbf{u}_1 = \mathbf{b}/\beta_1, \alpha_1\mathbf{v}_1 = \mathbf{A}^\top\mathbf{u}_1$. At each iteration,

$$\beta_{k+1}\mathbf{u}_{k+1} = \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{u}_k$$
$$\alpha_{k+1}\mathbf{v}_{k+1} = \mathbf{A}^\top\mathbf{u}_{k+1} - \beta_{k+1}\mathbf{v}_k.$$

At iteration m,

$$\mathbf{A}\mathbf{V}_m = \mathbf{U}_{m+1}\mathbf{B}_m$$

where $\mathbf{U}_{m+1} = \begin{bmatrix} \mathbf{u}_1, & \cdots & , \mathbf{u}_{m+1} \end{bmatrix}$ and $\mathbf{V}_m = \begin{bmatrix} \mathbf{v}_1, & \cdots & , \mathbf{v}_m \end{bmatrix}$ have orthonormal columns, and $\mathbf{B}_m = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_m & \alpha_m \\ & & & \beta_{m+1} \end{bmatrix}$.

Hybrid Method

After m steps of GK bidiagonalization, solve the *projected* problem:

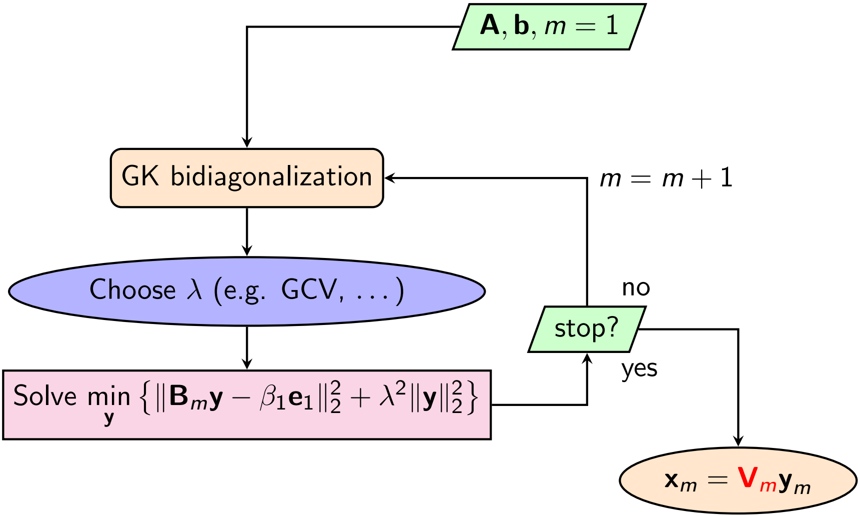
$$\min_{\mathbf{x}\in\mathcal{R}(\mathbf{V}_m)} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\lambda^2\|\mathbf{x}\|_2^2=\min_{\mathbf{y}} \|\mathbf{B}_m\mathbf{y}-\beta_1\mathbf{e}_1\|_2^2+\lambda^2\|\mathbf{y}\|_2^2$$

where $\mathbf{x}_m=\mathbf{V}_m\mathbf{y}$, and $\mathbf{V}_m \in N\times m$

Remarks:

- Ill-posed problem $\Rightarrow$ $\mathbf{B}_m$ may be very ill-conditioned.
- $\mathbf{B}_m$ is much smaller than $\mathbf{A}$
- Standard techniques (e.g. GCV) to find λ and stopping point

Hybrid Method(HyBR)



[O'Leary, Simmons 1981], [Björck, Grimme, Van Dooren 1988], [Larsen 1998], [Kilmer, O'Leary 2001], [Golub,Von Matt,1991], [Bazan, Borges, 2010], [Renaut, Hnětynková, Mead, 2010]

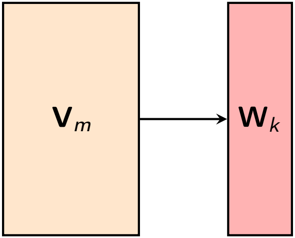
Storage challenge

Storage cost of $\mathbf{V}_m \in \mathbb{R}^{N\times m}$ is large

- Slow convergence: image deblurring problem with small noise level
- Large linear system: dynamic problems, streaming data, limited angle problems, modified angle problems
- Large m causes large projected problem

Outline

Compression Strategies



After m iterations of GK bidiagonalization,

$$\mathbf{A}\mathbf{V}_m = \mathbf{U}_{m+1}\mathbf{B}_m,$$

where $\mathbf{V}_m \in \mathbb{R}^{N \times m}$.

- Decompose $\mathbf{B}_m \in \mathbb{R}^{(m+1)\times m}$ using a low-rank approximation e.g. TSVD, Reduced Basis Decomposition- Y. Chen(2015)
- Use $\mathbf{y}_m$ to identify the important columns of $\mathbf{V}_m$: $\mathbf{x}_m = \mathbf{V}_m \mathbf{y}_m$ e.g. solution-oriented or sparsity enforcing compression

GK Bidiagonalization with Recycling

Given $\mathbf{A}, \mathbf{b}$ and $\mathbf{W}_k$, for $\ell = 1, 2, \ldots,$ compute

- $\mathbf{U}_{\ell+1} = \left[\ \mathbf{u}_1, \quad \cdots \quad, \mathbf{u}_{\ell+1} \ \right]$
- $\mathbf{V}_{\ell} = \left[\ \mathbf{v}_1, \quad \cdots \quad, \mathbf{v}_{\ell} \ \right]$
- $\mathbf{B}_\ell = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_\ell & \alpha_\ell \\ & & & \beta_{\ell+1} \end{bmatrix}$
- QR factorization: $\mathbf{A}\mathbf{W}_k = \mathbf{Y}_k \mathbf{R}_k$

where $\mathbf{U}_{\ell+1}, \mathbf{V}_\ell$ and $\mathbf{Y}_k$ have orthonormal columns, and

$$\mathbf{A} \begin{bmatrix} \mathbf{W}_k & \mathbf{V}_\ell \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_k & \mathbf{U}_{\ell+1} \end{bmatrix} \begin{bmatrix} \mathbf{R}_k & \mathbf{Y}_k^\top \mathbf{A} \mathbf{V}_\ell \\ \mathbf{0} & \mathbf{B}_\ell \end{bmatrix}$$

Solve the Regularized Projected Problem

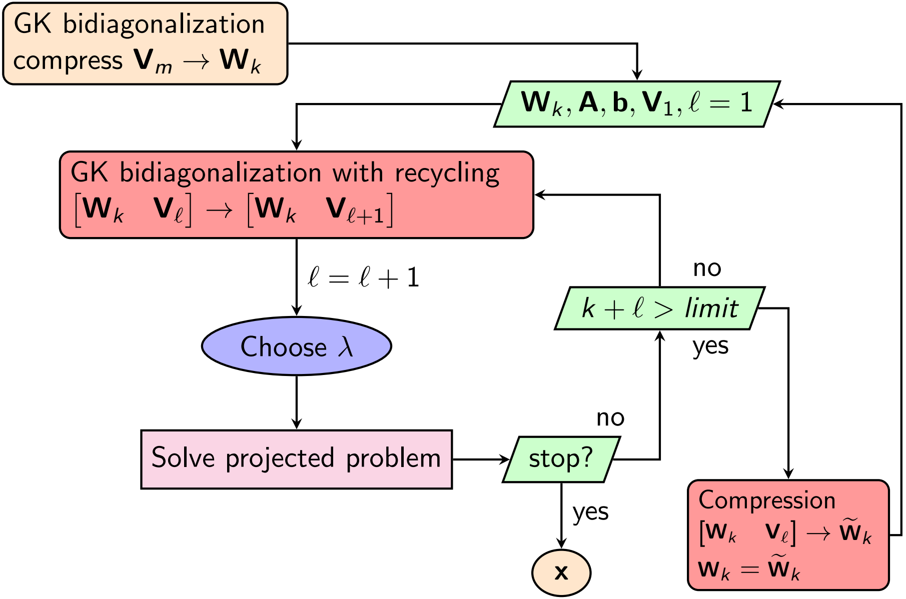
$$\mathbf{x}_{m+\ell} = \underset{\mathcal{R}\left([\mathbf{W}_k \quad \mathbf{V}_\ell]\right)}{\arg\min} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2$$

$$= \begin{bmatrix} \mathbf{W}_k & \mathbf{V}_\ell \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$$

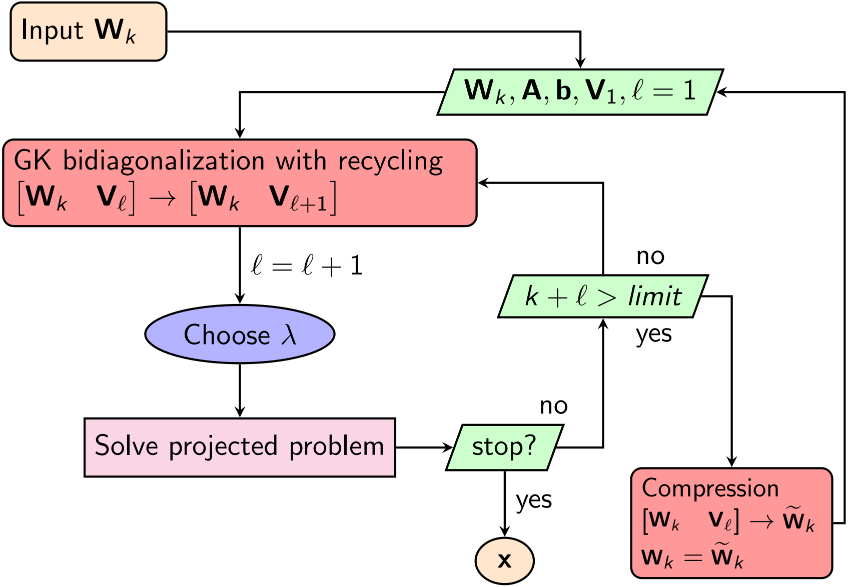
where $\begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$ solves

$$\min_{\mathbf{c},\mathbf{d}} \left\| \begin{bmatrix} \zeta + \mathbf{R}_k \mathbf{e}_k \\ \beta_1 \mathbf{e}_1 \end{bmatrix} - \begin{bmatrix} \mathbf{R}_k & \mathbf{Y}_k^\top \mathbf{A} \mathbf{V}_\ell \\ \mathbf{O} & \mathbf{B}_\ell \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} \right\|_2^2 + \lambda^2 \left\| \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} \right\|_2^2$$

where $\zeta = \mathbf{Y}_k^\top \mathbf{r}^{(1)}$ and $\mathbf{r}^{(1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(1)}$


GK bidiagonalization
compress $\mathbf{V}_m \to \mathbf{W}_k$
$\mathbf{W}_k, \mathbf{A}, \mathbf{b}, \mathbf{V}_1, \ell = 1$
GK bidiagonalization with recycling
$\begin{bmatrix} \mathbf{W}_k & \mathbf{V}_\ell \end{bmatrix} \to \begin{bmatrix} \mathbf{W}_k & \mathbf{V}_{\ell+1} \end{bmatrix}$
$\ell = \ell + 1$
Choose λ
Solve projected problem
stop?
no
yes
$\mathbf{x}$
$k + \ell > limit$
no
yes
Compression
$[\mathbf{W}_k \quad \mathbf{V}_\ell] \to \widetilde{\mathbf{W}}_k$
$\mathbf{W}_k = \widetilde{\mathbf{W}}_k$

Hybrid Projection Methods with Recycling(HyBR-recycle)



Interlacing Property

Define

- $\widehat{\mathbf{B}}_{m,k,\ell} = \begin{bmatrix} \mathbf{R}_k & \mathbf{Y}_k^\top \mathbf{A} \mathbf{V}_\ell \\ \mathbf{O} & \mathbf{B}_\ell \end{bmatrix}$ (TSVD compression)
- $\mathbf{B}_{m+\ell} = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_{m+\ell} & \alpha_{m+\ell} \\ & & & \beta_{m+\ell+1} \end{pmatrix}$

Theorem

Let σ_j denotes the j-th largest singular values of the matrix. Then

$$\sigma_{m-k+j}(\mathbf{B}_{m+\ell}) \leq \sigma_j(\widehat{\mathbf{B}}_{m,k,\ell}) \leq \sigma_j(\mathbf{B}_{m+\ell})$$

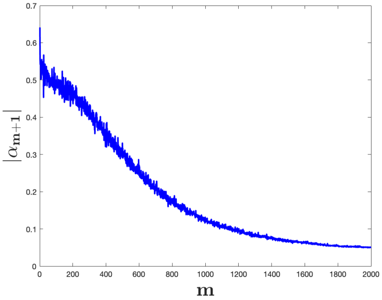
with $j = 1, \ldots, k+\ell$.

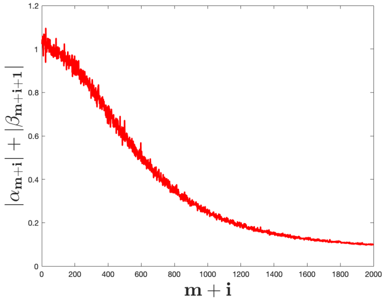
Norm Difference

Theorem

Let σ_k is the k-th largest singular values of $\mathbf{B}_m$. Then

$$\left| \|\mathbf{B}_{m+\ell}\|_F - \|\widehat{\mathbf{B}}_{m,k,\ell}\|_F \right| \leq 3\sqrt{\ell} \max_{1 \leq i \leq \ell} \{\sigma_k, (|\alpha_{m+i}| + |\beta_{m+i+1}|)\} + 2\sigma_k + 5\alpha_{m+1}$$





Outline

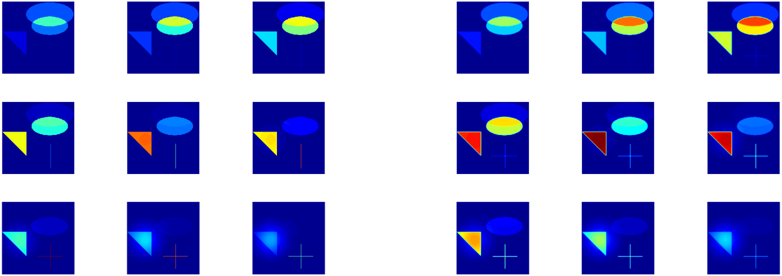
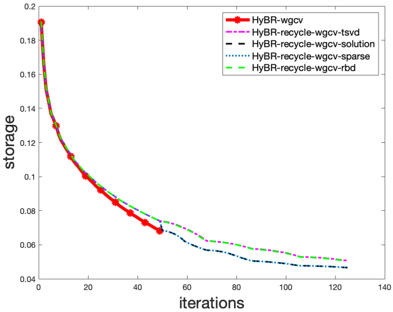
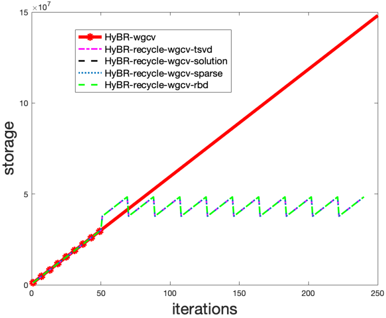


Figure 1: True image(left) and Observed image(right)

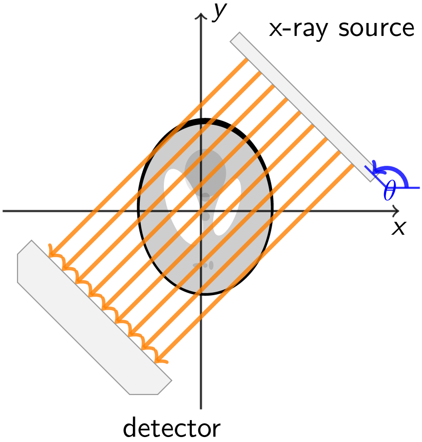
- $\mathbf{x}_{true} \in \mathbb{R}^{256^2 \times 9}$
- Noise level: 0.1%
- Relative error: $\frac{\|\mathbf{x}_m - \mathbf{x}_{true}\|_2}{\|\mathbf{x}_{true}\|_2}$
- Maximum storage for solution basis: 50
- Maximum storage for compression: $k = 30$

HyBR-recycle with WGCV

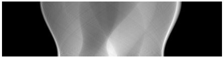




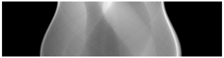
Tomography Reconstruction



True image



$0^o \sim 89^o$



$90^o \sim 179^o$

- **HyBR on all data:**

$$\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \right\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2,$$

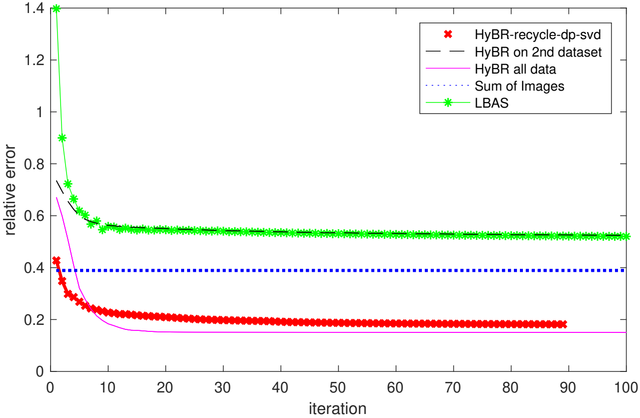
where $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{90\cdot 1448 \times 1024^2}$, $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^{90\cdot 1448}$ and $\mathbf{x} \in \mathbb{R}^{1024^2}$

- **HyBR on 1st dataset:** $\min_{\mathbf{x}} \|\mathbf{A}_1\mathbf{x} - \mathbf{b}_1\|_2^2 + \lambda^2\|\mathbf{x}\|_2^2$

- **HyBR on 2nd dataset:** $\min_{\mathbf{x}} \|\mathbf{A}_2\mathbf{x} - \mathbf{b}_2\|_2^2 + \lambda^2\|\mathbf{x}\|_2^2$

- **HyBR-recycle:** $\min_{\mathbf{x}} \|\mathbf{A}_2\mathbf{x} - \mathbf{b}_2\|_2^2 + \lambda^2\|\mathbf{x}\|_2^2$ with $\mathbf{W}_{10}$ from HyBR on 1st dataset

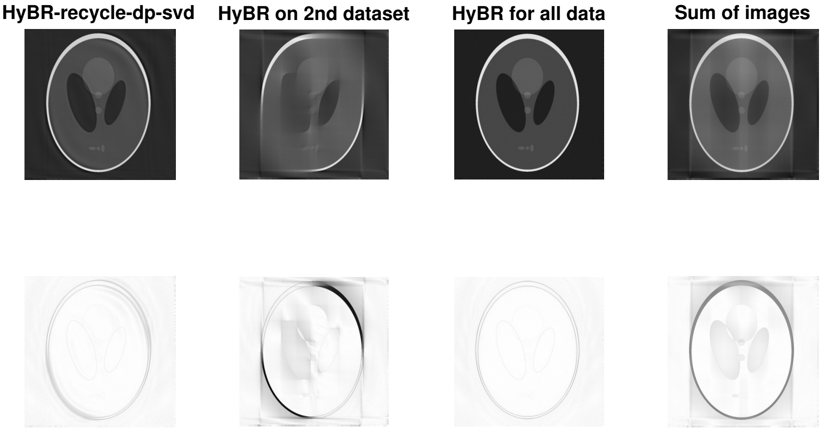
Relative Reconstruction Errors



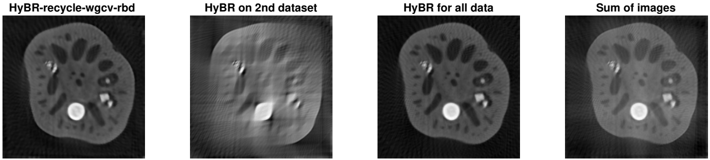
	HyBR-recycle	HyBR on the 2nd dataset	HyBR all data
CPU times:	**84.86s**	**83.20s**	**174.34s**

LBAS: P.C. Hansen, Y. Dong, and K. Abe. Hybrid enriched bidiagonalization for discrete ill-posed problems. Numerical Linear Algebra with Applications, 26(3):e2230, 2019.

Reconstructions and Error Images



Tomography reconstruction



- Image: lotus root (real data from Finish Inverse Problems Society)
- Limited angles: $1^o \sim 90^o$ and $91^o \sim 180^o$
- $$\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \right\|^2 + \lambda^2 \|\mathbf{x}\|^2,$$

where $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{R}^{60\cdot 328 \times 328^2}$, $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{R}^{60\cdot 328}$ and $\mathbf{x} \in \mathcal{R}^{256^2}$

Outline

Conclusions

- Use compression/ previous basis vectors to get $\mathbf{W}_k$
- Use recycling techniques with GK bidiagonalization to extend the solution space
- Enable large-scale inversion with limited storage of basis vectors

$$\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$$

- Automatic regularization parameter selection and stopping iteration

Thank you!!