Optimal Low-Rank Inverse Preconditioner

Matthias Chung and Julianne Chung

Department of Mathematics, Virginia Tech, {mcchung,jmchung}@vt.edu

October 2015 @ SIAM ALA 15

Optimal regularized inverse matrix (ORIM)

- Assume
  - ξ and δ independent
  - $\mathbb{E}(\delta) = \mathbf{0}$ $\qquad \mathbb{E}(\delta\delta^\top) = \eta^2 \mathbf{I}_m, \ (\eta \geq 0)$ $\qquad$ (w.l.o.g.)
  - $\mathbb{E}(\xi) = \mu$ $\qquad \mathbb{E}\left((\xi - \mu)(\xi - \mu)^\top\right) = \Gamma$ $\qquad$ (s.p.d.)
  - $\mathbf{M}\mathbf{M}^\top = \Gamma$ any symmetric factorization
- Using identity of quadratic form $\mathbb{E}\left(\epsilon^\top \Lambda \epsilon\right) = \mathrm{tr}\left(\Lambda \Sigma_\epsilon\right) + \mu_\epsilon^\top \Lambda \mu_\epsilon$

$$
\begin{aligned}
f(\mathbf{Z}) &= \mathbb{E}_{\xi,\delta} \left\|(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\xi + \mathbf{Z}\delta\right\|_2^2 \\
&= \mathbb{E}_{\xi}\left(\|(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\xi\|_2^2\right) + \mathbb{E}_{\delta}\left(\|\mathbf{Z}\delta\|_2^2\right) \\
&= \mu^\top(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)^\top(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\mu \\
&\qquad\qquad + \mathrm{tr}\left((\mathbf{Z}\mathbf{A} - \mathbf{I}_n)^\top(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\mathbf{M}\mathbf{M}^\top\right) + \eta^2 \mathrm{tr}\left(\mathbf{Z}^\top\mathbf{Z}\right) \\
&= \|(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\mu\|_2^2 + \|(\mathbf{Z}\mathbf{A} - \mathbf{I}_n)\mathbf{M}\|_\mathrm{F}^2 + \eta^2 \|\mathbf{Z}\|_\mathrm{F}^2
\end{aligned}
$$

Optimal regularized inverse matrix ($\mu = 0$)

Theorem

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank $k \le n \le m$ and an invertible matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, let their generalized singular value decomposition be $\mathbf{A} = \mathbf{U\Sigma G}^{-1}$, $\mathbf{M} = \mathbf{GS}^{-1}\mathbf{V}^\top$. Let η be a given parameter, nonzero if $k < m$. Let $r \le k$ be a given positive integer. Define $\mathbf{D} = \mathbf{\Sigma S}^{-2}\mathbf{\Sigma}^\top + \eta^2\mathbf{I}_m$. Let the symmetric matrix $\mathbf{H} = \mathbf{GS}^{-4}\mathbf{\Sigma}^\top\mathbf{D}^{-1}\mathbf{\Sigma G}^\top$ have eigenvalue decomposition $\mathbf{H} = \widehat{\mathbf{V}}\mathbf{\Lambda}\widehat{\mathbf{V}}^\top$ with eigenvalues ordered so that $\lambda_j \ge \lambda_i$ for $j < i \le n$. Then a global minimizer $\mathbf{Z}_r \in \mathbb{R}^{n\times m}$ of the problem

$$\mathbf{Z}_r = \underset{\mathrm{rank}(\mathbf{Z})\le r}{\arg\min} \left\|(\mathbf{ZA} - \mathbf{I}_n)\mathbf{M}\right\|_\mathrm{F}^2 + \eta^2 \left\|\mathbf{Z}\right\|_\mathrm{F}^2 \tag{1}$$

is

$$\mathbf{Z}_r = \widehat{\mathbf{V}}_r\widehat{\mathbf{V}}_r^\top\mathbf{GS}^{-2}\mathbf{\Sigma}^\top\mathbf{D}^{-1}\mathbf{U}^\top,$$

where $\widehat{\mathbf{V}}_r$ contains the first r columns of $\widehat{\mathbf{V}}$. Moreover this $\widehat{\mathbf{Z}}$ is the *unique* global minimizer of (1) if and only if $\lambda_r > \lambda_{r+1}$.

J Chung, M Chung, and D P O'Leary. **Optimal regularized low rank inverse approximation.** In: *Linear Algebra and its Applications* 468 (2015), pp. 260–269.

Truncated Tikhonov ($\mu = 0$ and $\mathbf{M} = \mathbf{I}_n$)

Theorem

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ of rank $k \leq n \leq m$, let its singular value decomposition be $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ with the singular values arranged in non-increasing order. Let η be a given parameter, nonzero if $k < m$. Let $r \leq k$ be a given positive integer. Then a global minimizer $\mathbf{Z}_r \in \mathbb{R}^{n \times m}$ of the problem

$$\mathbf{Z}_r = \underset{\text{rank}(\mathbf{Z}) \leq r}{\arg\min} \; \|\mathbf{Z}\mathbf{A} - \mathbf{I}_n\|_{\text{F}}^2 + \eta^2 \|\mathbf{Z}\|_{\text{F}}^2 \tag{2}$$

is

$$\mathbf{Z}_r = \mathbf{V}_r \boldsymbol{\Psi}_r \mathbf{U}_r^\top,$$

where $\mathbf{V}_r$ contains the first r columns of $\mathbf{V}$, $\mathbf{U}_r$ contains the first r columns of $\mathbf{U}$, and $\boldsymbol{\Psi}_r = \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\eta^2}, \ldots, \frac{\sigma_r}{\sigma_r^2+\eta^2}\right)$. Moreover, this $\widehat{\mathbf{Z}}$ is the *unique* global minimizer of (2) if and only if $\sigma_r > \sigma_{r+1}$.

Chung, Chung, and O'Leary 2015.

Using $\mathbf{Z}_r$ as a preconditioner

- Assume $\boldsymbol{\mu} = \mathbf{0}$ and $m = n$ then by Theorem

$$\mathbf{Z}_r = \widehat{\mathbf{V}}_r \widehat{\mathbf{V}}_r^\top \mathbf{G} \boldsymbol{\Phi} \boldsymbol{\Sigma}^{-1} \mathbf{U}^\top,$$

where $\boldsymbol{\Phi} = \mathbf{S}^{-2}\mathbf{D}\boldsymbol{\Sigma}^2$ and

$$\mathbf{Z}_r \mathbf{A} = \widehat{\mathbf{V}}_r \widehat{\mathbf{V}}_r^\top \mathbf{G} \boldsymbol{\Phi} \mathbf{G}^{-1}.$$

- Least squares problem

$$\min_{\boldsymbol{\xi}} \left\| \widehat{\mathbf{V}}_r \widehat{\mathbf{V}}_r^\top \mathbf{G} \boldsymbol{\Phi} \mathbf{G}^{-1} \boldsymbol{\xi} - \widehat{\mathbf{V}}_r \widehat{\mathbf{V}}_r^\top \mathbf{G} \boldsymbol{\Phi} \boldsymbol{\Sigma}^{-1} \mathbf{U}^\top \mathbf{b} \right\|_2^2$$

with solution

$$\boldsymbol{\xi}_{\mathrm{LS}} = \mathbf{G} \boldsymbol{\Phi}^{-1} \mathbf{G}^{-1} \widehat{\mathbf{V}}_r \widehat{\mathbf{V}}_r^\top \mathbf{G} \boldsymbol{\Phi} \boldsymbol{\Sigma}^{-1} \mathbf{U}^\top \mathbf{b}$$

- Hence LSQR will converge in at most r iterations
- For $\mathbf{M} = \mathbf{I}_n$, we have

$$\boldsymbol{\xi}_{\mathrm{LS}} = \mathbf{V}_r \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_r^\top \mathbf{b}$$

the rank-r TSVD solution (independent of η^2)

Efficient computation of $\mathbf{Z}_r$...

Rank update approach

Corollary

Assume all conditions of Theorem ✪ are fulfilled. Let $\mathbf{Z}_r$ be a global minimizer of (1) of maximal rank r and $\mathbf{Z}_{r+\ell}$ be a global minimizer of (1) for maximal rank $r+\ell$. Then $\widehat{\mathbf{Z}}_\ell = \mathbf{Z}_{r+\ell} - \mathbf{Z}_r$ is of maximal rank ℓ and the global minimizer of

$$\min_{\operatorname{rank}(\mathbf{Z})\leq\ell} \|((\mathbf{Z}_r + \mathbf{Z})\mathbf{A} - \mathbf{I})\mathbf{M}\|_{\mathrm{F}}^2 + \eta^2 \|\mathbf{Z}_r + \mathbf{Z}\|_{\mathrm{F}}^2 . \qquad (3)$$

Furthermore, $\widehat{\mathbf{Z}}_\ell$ is the unique global minimizer of (3) if and only if $\lambda_r > \lambda_{r+1}$ and $\lambda_{r+\ell} > \lambda_{r+\ell+1}$ with $r+\ell+1 < \operatorname{rank}(\mathbf{A})$.

J. Chung and M. Chung. **An efficient approach for computing optimal low-rank regularized inverse matrices.** In: *Inverse Problems* 30 (2014). Inverse Problems, pp. 1–19.

Rank-1 update

- $\mathbf{X}_{r-1}\mathbf{Y}_{r-1}^\top = \mathbf{Z}_{r-1}$ with

$$\mathbf{X}_{r-1} = [\mathbf{x}_1, \ldots, \mathbf{x}_{r-1}] \in \mathbb{R}^{n \times r-1}$$
$$\mathbf{Y}_{r-1} = [\mathbf{y}_1, \ldots, \mathbf{y}_{r-1}] \in \mathbb{R}^{m \times r-1}$$

- Optimization problem

$$(\mathbf{x}_r, \mathbf{y}_r) = \arg\min_{\mathbf{x},\mathbf{y}} f_r(\mathbf{x}\mathbf{y}^\top)$$

subject to $\|\mathbf{x}\|_2 = 1$ and $\operatorname{sign}(x_1) = 1$

- $\widehat{\mathbf{Z}} = \mathbf{x}_r\mathbf{y}_r^\top$

- Use constrained optimization to solve problem?

Alternating directions

- Use alternating directions:

$$\widehat{\mathbf{x}}(\mathbf{y}) = \arg\min_{\mathbf{x}} f_r(\mathbf{x}\mathbf{y}^\top) \quad \text{given } \mathbf{y}$$

then

$$\widehat{\mathbf{y}}(\mathbf{x}) = \arg\min_{\mathbf{y}} f_r(\mathbf{x}\mathbf{y}^\top) \quad \text{given } \mathbf{x}$$

- Closed form solution for $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{y}}$ with $\mathbf{M}_{\boldsymbol{\mu}} = [\boldsymbol{\mu}, \mathbf{M}]$

$$\widehat{\mathbf{x}}(\mathbf{y}) = \frac{\mathbf{M}_{\boldsymbol{\mu}}\mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{A}^\top\mathbf{y} - \mathbf{Z}_{r-1}\mathbf{A}\mathbf{M}_{\boldsymbol{\mu}}\mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{A}^\top\mathbf{y} - \eta^2\mathbf{Z}_{r-1}^\top\mathbf{y}}{\mathbf{y}^\top\mathbf{A}\mathbf{M}_{\boldsymbol{\mu}}\mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{A}^\top\mathbf{y} + \eta^2\mathbf{y}^\top\mathbf{y}}$$

$$\widehat{\mathbf{y}}(\mathbf{x}) = \left(\mathbf{A}\mathbf{M}_{\boldsymbol{\mu}}\mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{A}^\top + \eta^2\mathbf{I}_m\right)^{-1}\mathbf{A}\mathbf{M}_{\boldsymbol{\mu}}\mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{x}$$

equivalently solve

$$\begin{bmatrix} \mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{A}^\top \\ \eta\mathbf{I}_m \end{bmatrix} \widehat{\mathbf{y}}(\mathbf{x}) = \begin{bmatrix} \mathbf{M}_{\boldsymbol{\mu}}^\top\mathbf{x} \\ \mathbf{0} \end{bmatrix}$$

Alternating directions

Solve rank-1 problem

Require: $\mathbf{A}, \mu, \mathbf{M}, \eta, \mathbf{X}_{r-1}, \mathbf{Y}_{r-1}, r$

initial guess $\widehat{\mathbf{y}} = \mathbf{1}$

while stopping criteria not reached **do**

  compute $\widehat{\mathbf{x}}(\widehat{\mathbf{y}})$

  orthogonalize by $\widehat{\mathbf{x}} = \widehat{\mathbf{x}} - \mathbf{X}_{r-1}\mathbf{X}_{r-1}^\top \widehat{\mathbf{x}}$

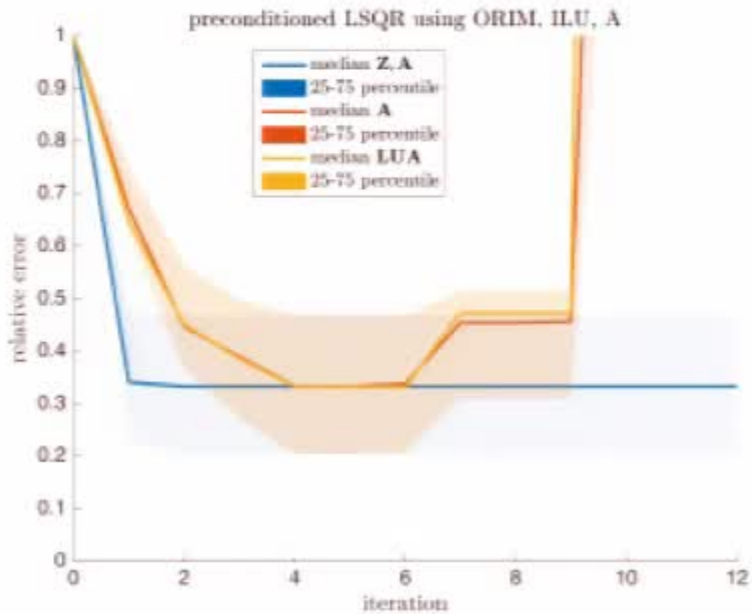
  normalize $\widehat{\mathbf{x}} = \widehat{\mathbf{x}} / \left\|\widehat{\mathbf{x}}\right\|_2$

  compute $\widehat{\mathbf{y}}(\widehat{\mathbf{x}})$

end while

Ensure: optimal $\mathbf{x}_r = \widehat{\mathbf{x}}$ and $\mathbf{y}_r = \widehat{\mathbf{y}}$

Example II: Fredholm integral, first kind

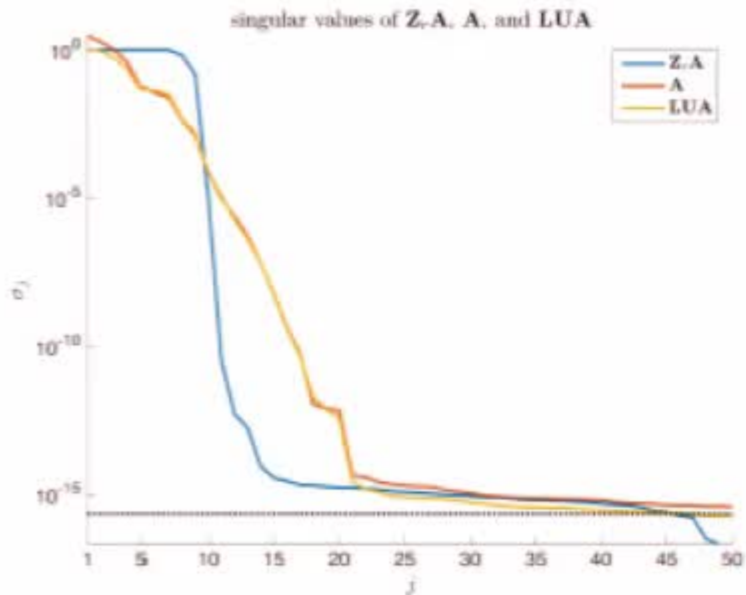


operator	# iter	timing [sec]	rel error
$\mathbf{A}$	11 ± 2	0.79 ± 0.097	6.31 ± 3.89
$\mathbf{U}^{-1}\mathbf{L}^{-1}\mathbf{A}$	11 ± 2	0.79 ± 0.096	6.62 ± 4.30
$\mathbf{Z}_r\mathbf{A}$	3.75 ± 0.5	1.11 ± 0.077	0.34 ± 0.15

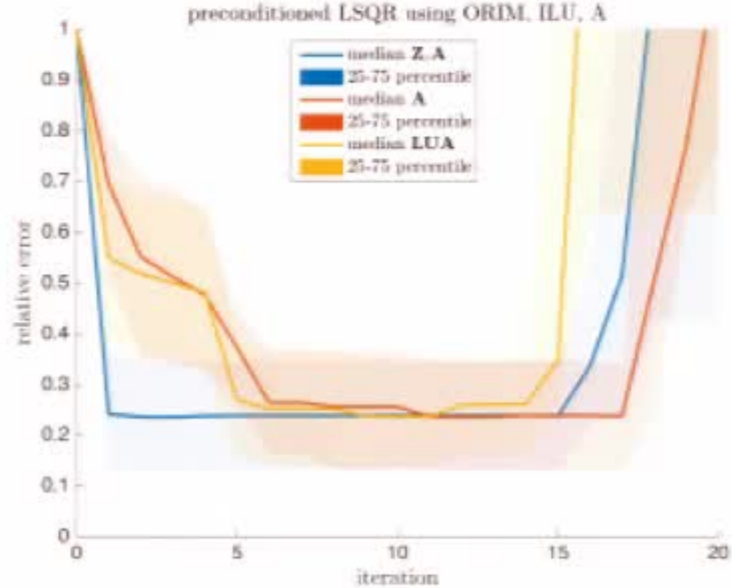
time to compute $\mathbf{Z}_r$, with $r = 4$ was 19.46 sec

Computations on *MacPro with 3.5 GHz, 6 core Intel Xeon E5; 32GB, 1866 MHz, DDR3 ECC; Yosemite; Matlab 2015b*

Example III: $\mathbf{Z}_r + \mathbf{D}$ Fredholm integral, first kind



singular values



LSQR

operator	# iter	timing [sec]	rel error
$\mathbf{A}$	$29{\pm}0$	$1.96{\pm}0.03$	$8.94{\pm}7.00$
$\mathbf{U}^{-1}\mathbf{L}^{-1}\mathbf{A}$	$29.5{\pm}7$	$1.99{\pm}0.45$	$23.62{\pm}38.76$
$(\mathbf{Z}_r + \mathbf{D})\mathbf{A}$	$7{\pm}0$	$2.52{\pm}0.45$	$0.23{\pm}0.13$

time to compute $\mathbf{Z}_r$, with $r = 9$ (plus $\mathbf{D}$) was 421.44 sec (62.52 sec without diagonal)

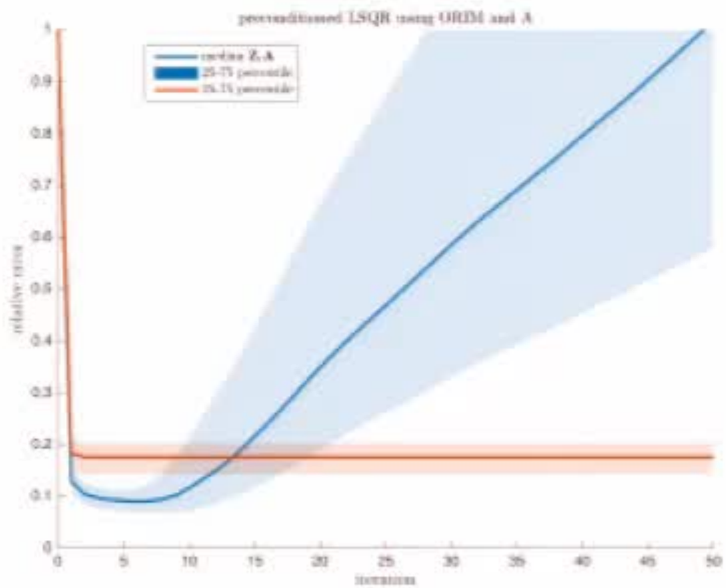
Computations on *MacPro with 3.5 GHz, 6 core Intel Xeon E5; 32GB, 1866 MHz, DDR3 ECC; Yosemite; Matlab 2015b*

Example IV: deblurring

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m = n = 1{,}048{,}576$

  $\mathbf{A} =$ `blur(n)`, from `regtools`
- Bayes assumptions: $\mathbf{M} = \mathbf{I}_n$, $\mu = \mathbf{1}$, $\eta^2 = 10^{-2}$, $r = 1{,}000$

operator	# iter	timing [sec]	rel error
$\mathbf{A}$	21.25 ± 0.957	14.77 ± 0.589	0.4234 ± 0.2788
$(\mathbf{Z}_r + \mathbf{D})\mathbf{A}$	2 ± 0	4.43 ± 0.022	0.18 ± 0.032



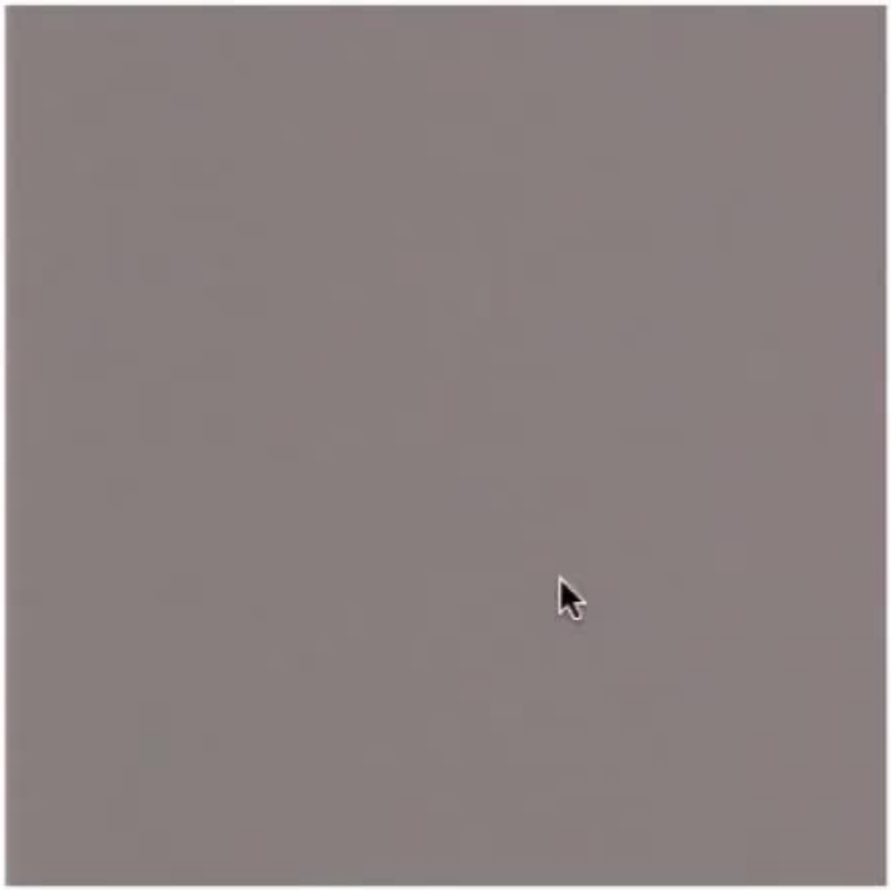
LSQR

Computations on MacPro with 2.6 GHz, 12 core Intel Xeon E5; 64 GB, 1866 MHz, DDR3 ECC; Yosemite; Matlab 2015b

Psychedelic orthogonal vectors $\mathbf{x}_j$

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m = n = 1{,}048{,}576$

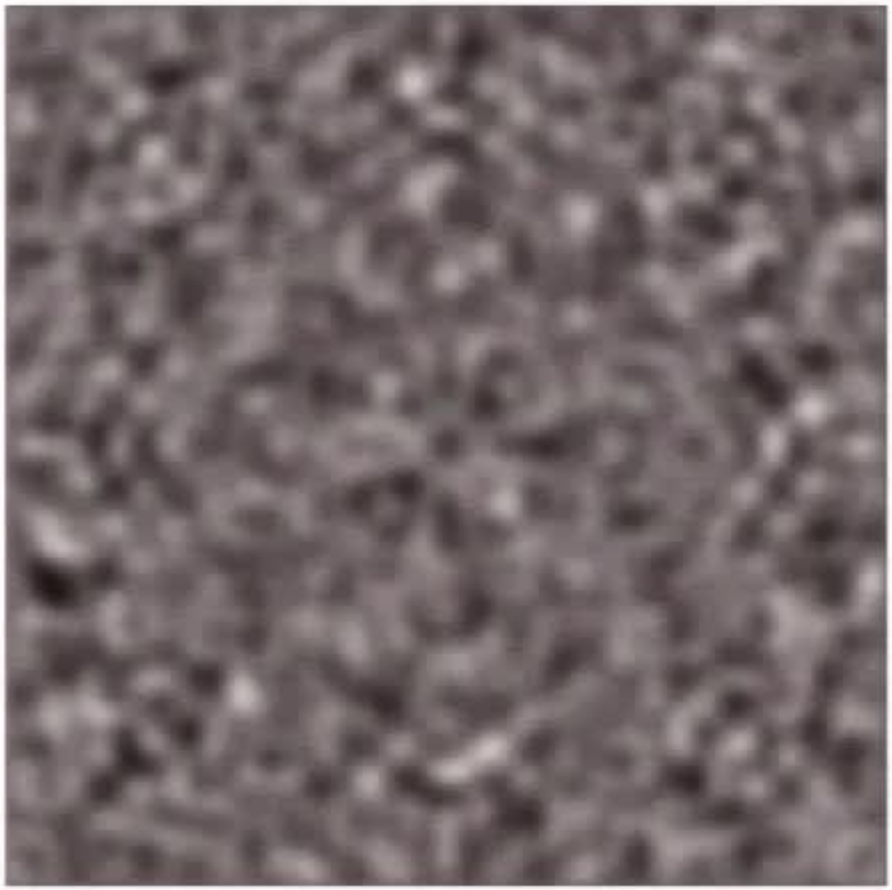
  $\mathbf{A} =$ `blur(n)`, from `regtools`
- Bayes assumptions: $\mathbf{M} = \mathbf{I}_n$, $\boldsymbol{\mu} = \mathbf{1}$, $\eta^2 = 10^{-2}$



first 400 orthogonal vectors $\mathbf{x}_j$

Psychedelic orthogonal vectors $\mathbf{x}_j$

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m = n = 1{,}048{,}576$
  $\mathbf{A} = \texttt{blur(n)}$, from $\texttt{regtools}$
- Bayes assumptions: $\mathbf{M} = \mathbf{I}_n$, $\boldsymbol{\mu} = \mathbf{1}$, $\eta^2 = 10^{-2}$



first 400 orthogonal vectors $\mathbf{x}_j$

Conclusion & Outlook

Presented work

- low-rank ORIM $\mathbf{Z}_r$ as a preconditioner
- convergence results, e.g., low-rank ORIM preconditioner can lead to TSVD solutions
- efficient rank-1 update approach
- fast convergence and may avoid semi-convergence

Future work

- constraints on ORIM e.g., sparsity or structure
- compare with other preconditioners and iterative solvers
- use ORIM to update an existing preconditioner