Exponentiated Symmetric Matrix Distributions with Application to Linear Inverse Problems

Piyush Agarwal, Sean Cottrell, Yi Luo

Simon Fraser University, Michigan State University, Fudan University

Polymath Jr. Program

Advised by Aimee Maurais (MIT) and David Gomez (Georgia Tech)

Symmetric Positive Definite Matrices

$A \in \mathbb{R}^{d \times d}$ is *symmetric positive definite* (SPD) if

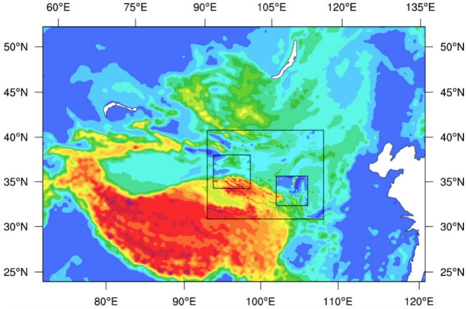
- $A = A^T$
- $\boldsymbol{y}^\top A \boldsymbol{y} > 0 \; \forall \boldsymbol{y} \neq 0 \Leftrightarrow \lambda_\mathrm{i}(\mathrm{A}) > 0, \; \mathrm{i} \in \{1, \ldots, \mathrm{d}\}$

For a random vector $\mathrm{X} \in \mathbb{R}^d$, the *covariance matrix* of X is defined as

- $\mathrm{cov}(\mathrm{X}, \mathrm{X}) = \mathbb{E}\left[(\mathrm{X} - \mathbb{E}[\mathrm{X}])(\mathrm{X} - \mathbb{E}[\mathrm{X}])^\top\right]$
- Always **positive semidefinite**.

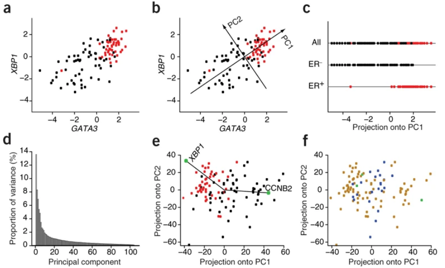
Why are Symmetric Positive Definite Matrices Useful?

Inverse problems. Source: TSA

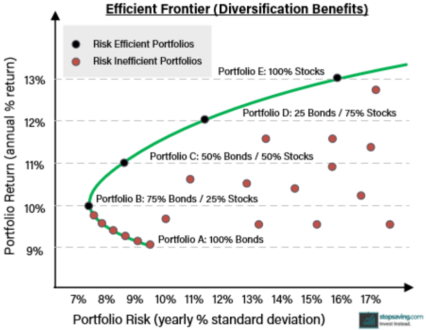


Sequential inference. Source: NCAR ensemble wind forecast

Why are Symmetric Positive Definite Matrices Useful?



Dimension reduction/PCA. Source: Ringnér 2008



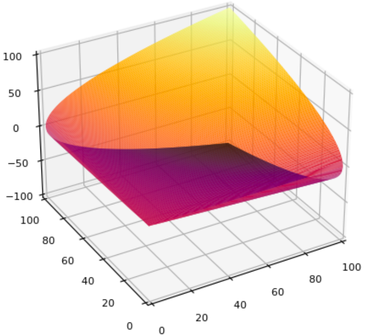
Finance, e.g., portfolio theory. Source: StopSaving.com

The Manifold of SPD Matrices

The set

$$\mathbb{P}_d = \{S \in \mathbb{R}^{d\times d} : S = S^\top,\ \lambda_i(S) > 0,\ i = 1, \ldots, d\}$$

is a **Riemannian manifold** with non-Euclidean geometry.



Exponentiation can be used to map symmetric matrices onto the manifold: given $\Sigma \in \mathbb{P}_d$, define

$$\exp_\Sigma(B) = \Sigma^{\frac{1}{2}} \exp(\Sigma^{-\frac{1}{2}} B \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}},$$

for $B \in \mathbb{H}_d = \{A \in \mathbb{R}^{d\times d} : A = A^\top\}$.

Random Matrix Theory

Random SPD matrix ensembles can be generated by exponentiating random symmetric matrices,

$$S = \exp_{\Sigma} \mathcal{E}, \quad \mathcal{E} \in \mathbb{H}_d,\ \mathbb{E}[\mathcal{E}] = 0$$

Our goal is to characterize these SPD matrix ensembles

- Joint element densities
- Spectral properties

Three Subprojects

Motivating Question

Given a parametric model for $\mathcal{E} \in \mathbb{H}_d$ with $\mathbb{E}[\mathcal{E}] = 0$, what can we say about the <u>transformed</u> random matrix $S = \exp_\Sigma \mathcal{E} \in \mathbb{P}_d$?

I. Characterizing symmetric matrix ensembles

II. Exponentiating symmetric matrix ensembles, $S = \exp_\Sigma \mathcal{E}$

III. Application: SPD Matrix Ensembles in Linear Inverse Problems

Characterizing Symmetric Matrix Ensembles

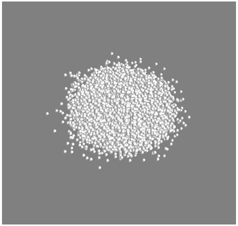
- Symmetric Matrix Ensembles
  - Elements are Independent
    - Gaussian Orthogonal, Uniform Ensemble
      - $A = (B + B^T)/2$ where B has ind. entries
        - $f(A) = \prod_{i \leq j} f(a_{ij})$
  - Elements are correlated
    - Colored Gaussian Symmetric Ensemble
      - $A = (B + B^T)/2$ where $B = \text{mat}(X)$, $X \sim \mathcal{N}(0, \Phi)$
        - $f(A)$ lacks product structure

Gaussian Orthogonal Ensemble

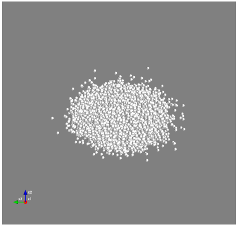
Let $B = [b_{ij}]$ where $b_{ij} \sim \mathcal{N}(0,1)$ and $A = \dfrac{B + B^T}{2}$

Then we have $A = [a_{ij}]$ with $a_{ij} \sim \mathcal{N}\left(0, \frac{1}{2}\right)$ for $i \neq j$ and $\mathcal{N}(0,1)$ for $i = j$

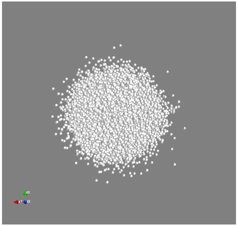
For $d = 2$, Let $A = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix}$. We generate 10^4 realizations of A and plot in a_1-a_2-a_3 space



(a) Isometric view



(b) a_3-a_2 plane



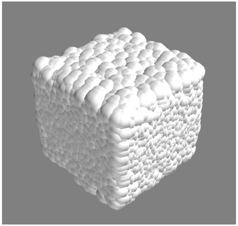
(c) a_1-a_3 plane

Uniform Ensemble

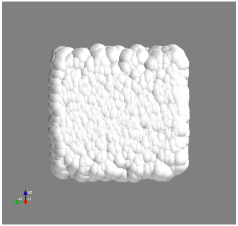
Let $B = [b_{ij}]$ where $b_{ij} \sim U(0,1)$ and $A = \dfrac{B + B^T}{2}$

Then we have $A = [a_{ij}]$ with $a_{ij} \sim Tri\left(0, 1, \frac{1}{2}\right)$ for $i \neq j$ & $U(0,1)$ for $i = j$

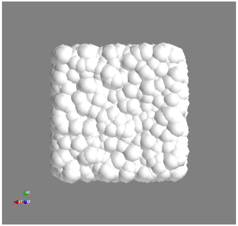
Visualizing A as previously,



(a) Isometric View



(b) a_3-a_2 plane



(c) a_1-a_3 plane

Colored Gaussian Symmetric Ensemble

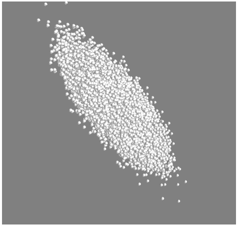
Illustration for $d = 2$

Let Φ be a 4×4 covariance matrix and $X = [x_1 \; x_2 \; x_3 \; x_4]^\top \sim \mathcal{N}(0, \Phi)$.

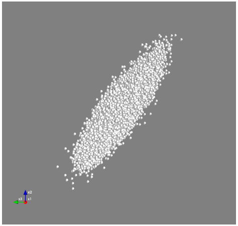
Set $B = \mathrm{mat}(X) = \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}$ and $A = \begin{bmatrix} x_1 & (x_2 + x_3)/2 \\ (x_2 + x_3)/2 & x_4 \end{bmatrix}$.

We say $A \sim \mathrm{CGSE}(2, \Phi)$.

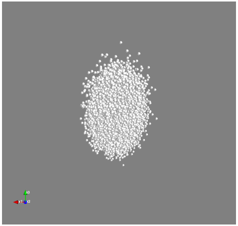
Visualizing A as previously,



(a) Isometric view



(b) a_3-a_2 plane



(c) a_1-a_3 plane

Eigenvalue Densities

Joint eigenvalue density for the Gaussian orthogonal ensemble:

$$f(\lambda_1, \ldots, \lambda_d) \propto \prod_{i<j} |\lambda_i - \lambda_j| \exp\left(-\sum_{i=1}^{d} \frac{\lambda_i^2}{2}\right)$$

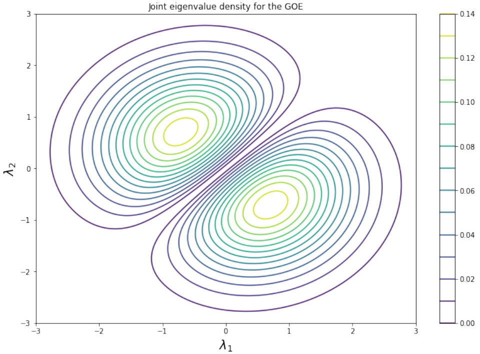
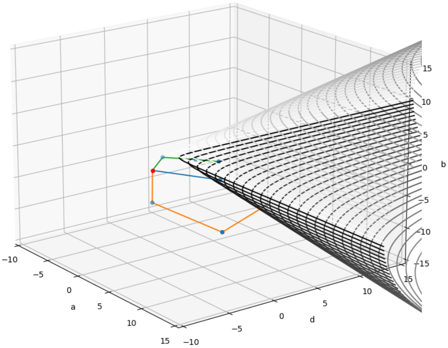


Figure 6: Contour plot of $f(\lambda_1, \lambda_2)$ in $d = 2$

Exponentiating Symmetric Matrix Ensembles



$$\mathcal{E} \in \mathbb{H}_d,\ \Sigma \in \mathbb{P}_d$$

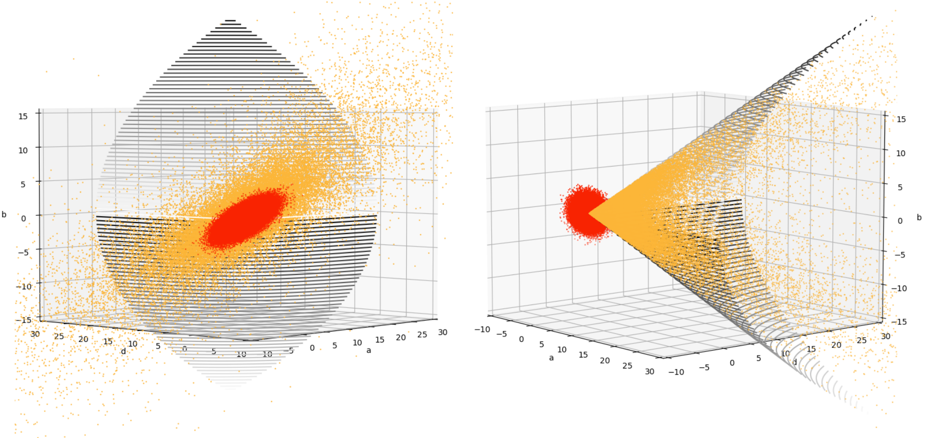
Matrix Exponential:

$$\begin{aligned}\exp(\mathcal{E}) &= \exp(Q^\top \Lambda Q) \\ &= Q^\top \exp(\Lambda) Q \\ &= Q^\top \mathrm{diag}\{e^{\lambda_1}, \cdots, e^{\lambda_n}\} Q\end{aligned}$$

Matrix Exponential With Base Σ:

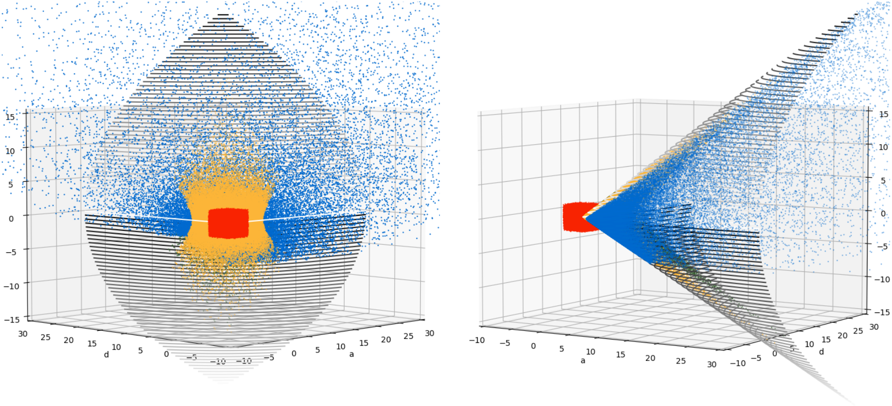
$$\begin{aligned}S &= \exp_\Sigma \mathcal{E} \\ &\equiv \Sigma^{1/2} \exp\left(\Sigma^{-1/2}\mathcal{E}\Sigma^{-1/2}\right)\Sigma^{1/2}\end{aligned}$$

Exponentiating Symmetric Matrix Ensembles



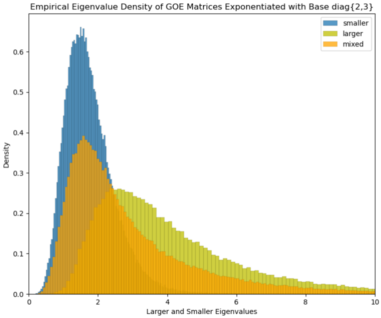
CGSE with $\Phi = \begin{bmatrix} 2 & 1 & 1 & -1 \\ 1 & 2 & 1 & -1 \\ 1 & 1 & 2 & -1 \\ -1 & -1 & -1 & 2 \end{bmatrix}$ exponentiated with base $\Sigma = I$.

Exponentiating Symmetric Matrix Ensembles

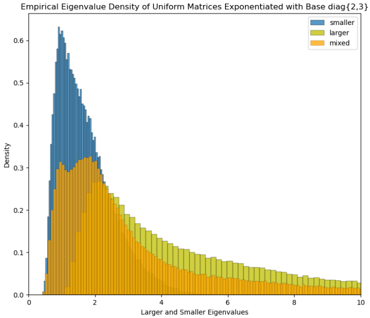


Uniform symmetric ensemble exponentiated with bases $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$ (yellow) and $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ (blue).

Empirical Eigenvalue Density



(a) Exponentiated GOE eigenvalue density



(b) Exponentiated uniform symmetric ensemble ($d = 2, r = 2$) eigenvalue density

Can we obtain the analytic expression for the eigenvalue density of exponentiated matrices?

Special Case: Diagonal Base in $d = 2$

What about the eigenvalue density of $\exp_\Sigma \mathcal{E}$ where $\mathcal{E} \sim \mathrm{GOE}(2)$ and $\Sigma \in \mathbb{P}_2$ diagonal?

We let

$$\Sigma = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \in \mathbb{P}_2,\ \mathcal{E} = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbb{H}_2$$

and consider the eigenvalues of

$$S = \exp_\Sigma \mathcal{E} = \Sigma^{\frac{1}{2}} \exp\left(\Sigma^{-\frac{1}{2}} \mathcal{E} \Sigma^{-\frac{1}{2}}\right) \Sigma^{\frac{1}{2}}$$

Special Case: Diagonal Base in $d = 2$

Set

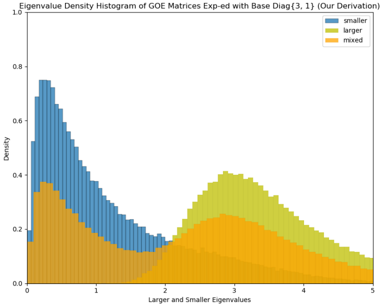
$$\Delta = a^2z^2 - 2abxz + b^2x^2 + 4aby^2, \quad u_1 = \frac{\sqrt{\Delta} + az - bx}{2\sqrt{\Delta}}, \quad u_2 = \frac{\sqrt{\Delta} - az + bx}{2\sqrt{\Delta}}$$

$$\lambda_1 = \frac{-\sqrt{\Delta} + az + bx}{2ab}, \quad \lambda_2 = \frac{\sqrt{\Delta} + az + bx}{2ab}$$

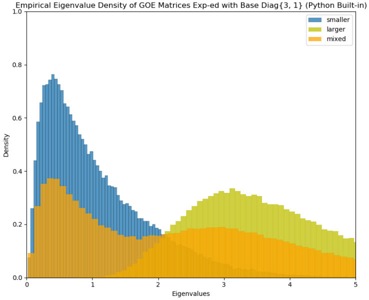
After much ~~weeping and gnashing of teeth~~ algebra, we have

$$\mu_{1,2} = \frac{1}{2}\left[\pm\sqrt{\begin{aligned}&\left(a(u_1e^{\lambda_1} + u_2e^{\lambda_2}) + b(u_2e^{\lambda_1} + u_1e^{\lambda_2})\right)^2 \\ &+ 4abe^{\lambda_1}e^{\lambda_2}(u_1 - u_2)^2\end{aligned}} \right. \\ \left. + a\left(u_1e^{\lambda_1} + u_2e^{\lambda_2}\right) + b\left(u_2e^{\lambda_1} + u_1e^{\lambda_2}\right)\right]$$

Special Case: Diagonal Base in $d = 2$



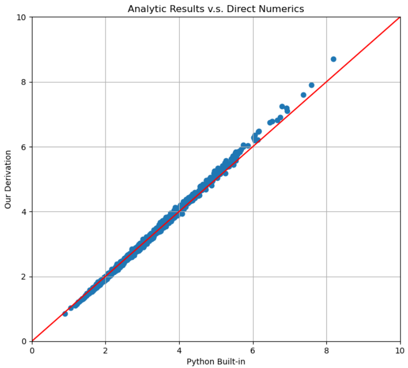
(a) Density rendered by our formula for eigenvalues



(b) Empirical density rendered by Python built-in functions (iterative method)

Special Case: Diagonal Base in $d = 2$

Reliability of The Derivation



Eigenvalue results computed from analytic formula compared with Python built-in functions for eigenvalue computation:

$$\mathcal{E} \sim \mathrm{GOE}(2)$$

$$\Sigma = \begin{bmatrix} 30 & 0 \\ 0 & 3 \end{bmatrix}$$

Linear Inverse Problems

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{d} \in \mathbb{R}^m$ is observed data
- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the known forward operator
- $\boldsymbol{s} \in \mathbb{R}^n$ is unknown information we want to recover
- $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is measurement error

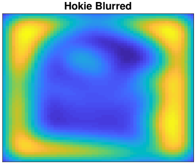
One may try to use least squares to solve for $\boldsymbol{s}$.

When Least Squares Fails

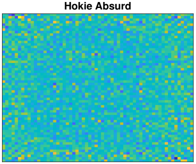
Least squares solution:

$$\boldsymbol{s}_{ls} = \underset{\boldsymbol{s}\in\mathbb{R}^n}{\arg\min}(||\boldsymbol{d} - \boldsymbol{A}\boldsymbol{s}||^2) \iff \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{s}_{ls} = \boldsymbol{A}^T\boldsymbol{d}$$

(a) Original Image



(b) Given Image($\boldsymbol{d}$)



(c) due Lst. squares ($\boldsymbol{s}_{ls}$)

We need **regularization** in order to obtain reasonable solutions[1]

[1] Example couresty of Mark Embree (Virginia Tech)

Bayesian Formulation

One way to regularize the solution is to adopt a *Bayesian* framework. Assume

$$\boldsymbol{s} \sim \mathcal{N}(\mu, \alpha^2 \Gamma_{\text{prior}}) \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \Gamma_{\text{noise}})$$

We use these assumptions and Bayes' Theorem to obtain a posterior probability distribution for $\boldsymbol{s}$ given $\boldsymbol{d}$, $\pi_{\text{post}}(\boldsymbol{s}|\boldsymbol{d})$

Bayesian Formulation

Bayes' Theorem: $\pi_{\text{post}}(\boldsymbol{s}|\boldsymbol{d}) \propto \pi_{\text{prior}}(\boldsymbol{s})\pi(\boldsymbol{d}|\boldsymbol{s})$.

Gaussian assumptions:

$$\pi_{\text{prior}}(\boldsymbol{s}) \propto \exp\left(-\frac{1}{2\alpha^2}\left((\boldsymbol{s}-\mu)^T\Gamma_{\text{prior}}^{-1}(\boldsymbol{s}-\mu)\right)\right)$$

$$\pi(\boldsymbol{d}|\boldsymbol{s}) \propto \exp\left(-\frac{1}{2\sigma^2}\left((\boldsymbol{d}-A\boldsymbol{s})^T\Gamma_{\text{noise}}^{-1}(\boldsymbol{d}-A\boldsymbol{s})\right)\right)$$

With $\Gamma_{\text{prior}}^{-1} = L^TL$ and $\Gamma_{\text{noise}}^{-1} = W^TW$ and setting $\lambda = \frac{\sigma}{\alpha}$ we have:

$$\pi_{\text{post}}(\boldsymbol{s}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2\sigma^2}\left(||W(\boldsymbol{d}-\boldsymbol{A}\boldsymbol{s})||^2 + \lambda^2||L(\boldsymbol{s}-\boldsymbol{\mu})||^2\right)\right)$$

Bayesian Formulation

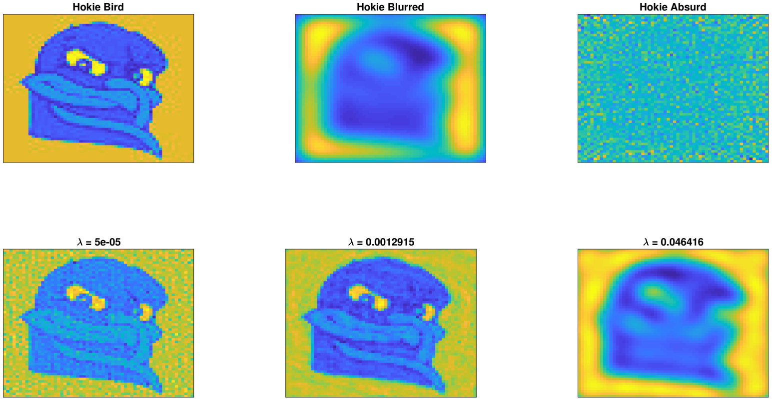
By maximizing $\pi_{\text{post}}(\boldsymbol{s}|\boldsymbol{d})$, we get the *maximum a posteriori* estimate $\boldsymbol{s}_{\text{map}}$ for $\boldsymbol{s}$,

$$\boldsymbol{s}_{\text{map}} = \arg\min_{\boldsymbol{s}}(||W(\boldsymbol{d} - \boldsymbol{A}\boldsymbol{s})||^2 + \lambda^2||L(\boldsymbol{s} - \boldsymbol{\mu})||^2)$$

$$\Updownarrow$$

$$\boldsymbol{s}_{\text{map}} = (\boldsymbol{A}^T\Gamma_{\text{noise}}^{-1}\boldsymbol{A} + \lambda^2\Gamma_{\text{prior}}^{-1})^{-1}(\boldsymbol{A}^T\Gamma_{\text{noise}}^{-1}\boldsymbol{d} + \lambda^2\Gamma_{\text{prior}}^{-1}\boldsymbol{\mu})$$

Bayesian Linear Inverse Problems



Reconstruction quality is dependent on choice of **regularization parameter** λ

From Linear Inverse Problems to Random Matrices

For fixed data and forward operator, the solution to a linear inverse problem is *parametrized* by Γ_{prior} and Γ_{noise}

- Usually $\Gamma_{\text{noise}} = I$ for white noise
- Γ_{prior} is a modeling choice, and we want the final image to have high reconstruction accuracy

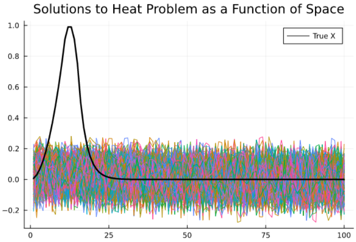
We model $\Gamma_{\text{prior}} = \exp_{\Sigma}(\mathcal{E})$ where $\mathcal{E}$ is a symmetric random matrix with $\mathbb{E}[\mathcal{E}] = 0$ and Σ is a chosen SPD matrix.

Given this distribution for Γ_{prior}, we are interested in the distribution of $\boldsymbol{s}_{\text{map}}$

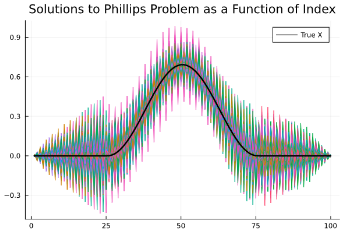
SPD Matrix Ensembles in Linear Inverse Problems

Matrix exponentiation tends to inflate condition number

- Well-studied phenomenon in differential equations (e.g., Loan 1977)
- When the prior covariance matrix is ill-conditioned, the optimal regularization parameter is $\approx 0 \Rightarrow$ **reversion to least squares**



(a) Heat Problem with $\lambda = 1$



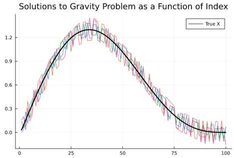
(b) Phillips Problem with $\lambda = 10^{-10}$

Controlling Conditioning

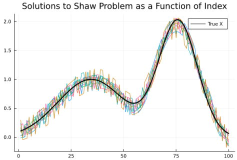
- Random Perturbation
  - For poorly conditioned A, the sum $M = A + N_n$ for random N_n is well-conditioned with high probability (Tao and Vu 2007)
  - If $A \in \mathbb{P}_d$, sample N_n from the Wishart Ensemble to ensure $M \in \mathbb{P}_d$
- Damping the Inversion
  - $M = A + cI$
  - Inflating the diagonal stabilizes eigenvalues
  - c is known as the Marquardt-Levenberg coefficient
  - Tradeoff between stabilizing matrix and retaining original information

Linear Inverse Problem Solutions

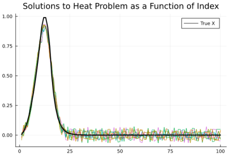
- Damping the inversion provides considerable improvement to input data reconstruction across a range of regularization parameters
  - $\mathcal{E} \sim \text{GOE}(100)$
  - $\Sigma \sim \text{Wishart Ensemble}$
- Inverse problems obtained via Matrix Depot package in Julia



(a) Gravity Surveying



(b) Image Restoration



(c) Heat Equation

In Summary

Contributions

- Intuitive visualization of random symmetric and SPD matrices in $d = 2$
- Generalization of exponentiated matrix eigenvalue distributions
- Application of exponentiated symmetric matrix ensembles to covariance modeling in linear inverse problems

Future Work

- More general eigenvalue densities
- Additional joint element densities
- Refined linear inverse problem results