Asymptotics of the Sketched Pseudoinverse

Daniel LeJeune[†‡], Pratik Patil[†§], Hamid Javadi[¶], Richard G. Baraniuk[‡], Ryan J. Tibshirani[‖]

[†] Equal contribution
[‡] Department of Electrical and Computer Engineering, Rice University
[§] Department of Statistics, Machine Learning Department, Carnegie Mellon University
[¶] Google
[‖] Department of Statistics, University of California, Berkeley

Abstract

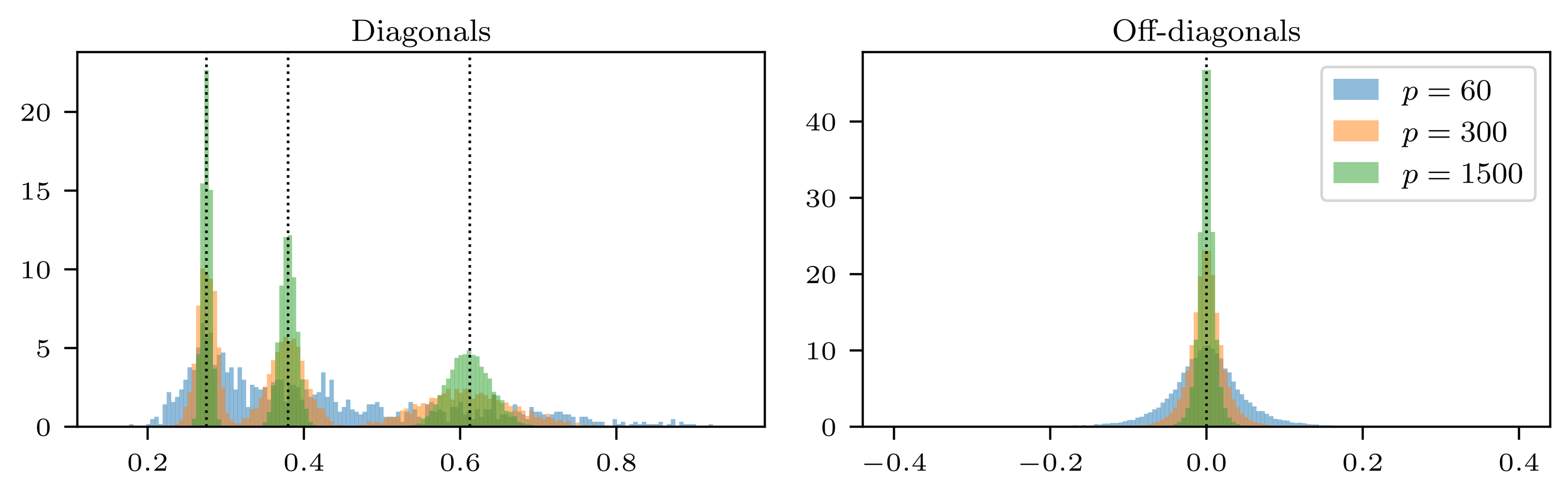
- **Problem:** Random sketching of regularized linear systems leads to a **different** solution than the original system.
- **Contribution:** We precisely characterize the pseudoinverse under sketching in the asymptotic regime, obtaining **first-** and **second-order equivalences** in terms of the ridge resolvent. We provide a conjecture that the same result holds for **general free sketching** supported by experiments.

First-order Equivalence

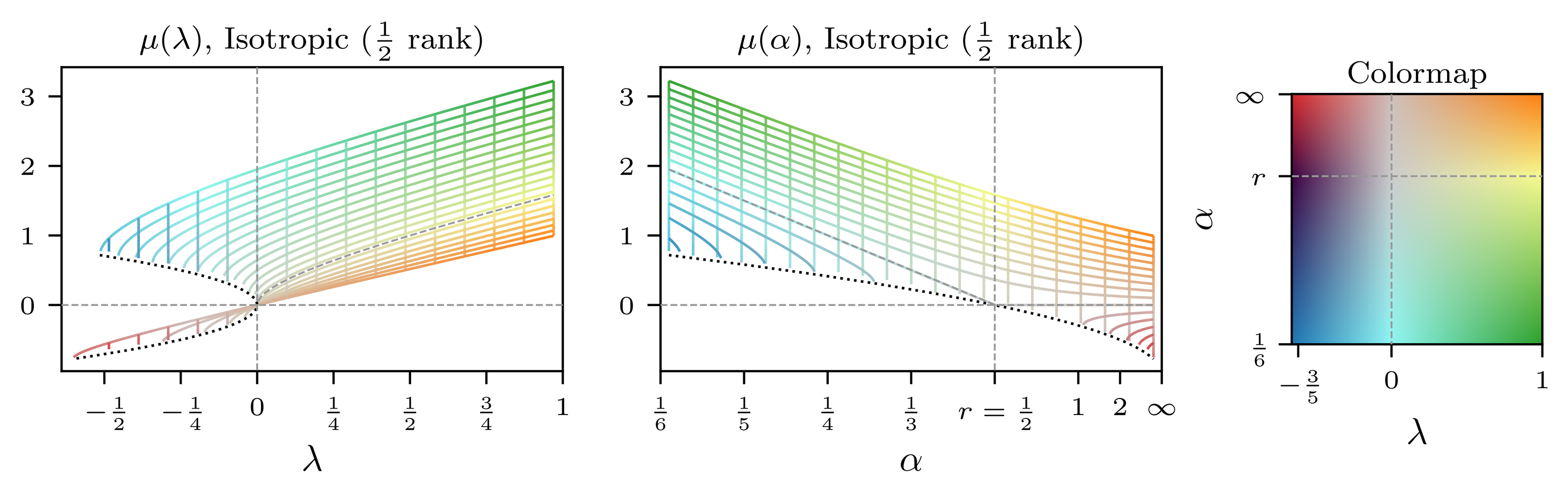
- Sketching yields a **first-order** equivalence to **ridge regularization** under inversion.
- **Setup:**
  - $\mathbf{A} \in \mathbb{C}^{p \times p}$ is positive semidefinite with uniformly upper bounded $\|\mathbf{A}\|_{\mathrm{op}}$ and lower bounded $\lambda_{\min}^{+}(\mathbf{A})$.
  - $\sqrt{q}\mathbf{S} \in \mathbb{C}^{p \times q}$ is a random matrix consisting of i.i.d. random variables that have mean 0, variance 1, and finite $8+\delta$ moment for some $\delta > 0$.

> ***Theorem 1.*** *For any $\lambda > \limsup \lambda_{\min}^{+}(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S})$, as $q, p \to \infty$ such that $0 < \liminf \frac{q}{p} \leq \limsup \frac{q}{p} < \infty$,*
> $$\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} + \lambda \mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}} \simeq \left(\mathbf{A} + \mu \mathbf{I}_p\right)^{-1},$$
> *where μ is the most positive solution to*
> $$\lambda = \mu\left(1 - \tfrac{1}{q}\mathrm{tr}\left[\mathbf{A}\left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1}\right]\right).$$

- That is, for any $\boldsymbol{\Theta}$ with uniformly bounded trace norm,
$$\mathrm{tr}\left[\boldsymbol{\Theta}\left(\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} + \lambda\mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}} - \left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1}\right)\right] \xrightarrow{\text{a.s.}} 0.$$
- Implies, e.g., **elementwise convergence**.
- **Example:** $\mathbf{A} = \mathrm{diag}\left(0\ldots, 1\ldots, 2\ldots\right)$, $q = \lfloor 0.8p \rfloor$, $\lambda = 1$.



- We provide a **detailed characterization** of the map from λ and sketch size $\alpha = \frac{q}{p}$ to μ.
  - $\lambda \mapsto \mu$ is **increasing**,
  - For $\lambda \geq 0$, $\alpha \mapsto \mu$ is **decreasing**, and $\mu \geq \lambda$
  - Distinct behavior for $\alpha < r(\mathbf{A})$ and $\alpha > r(\mathbf{A})$ (rank)
  - Precise limits of **negative regularization** $\lambda < 0$ and $\mu < 0$.
- **Example:** $\mathbf{A} = \begin{bmatrix} \mathbf{I}_{\lfloor rp \rfloor} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $r = \frac{1}{2}$.

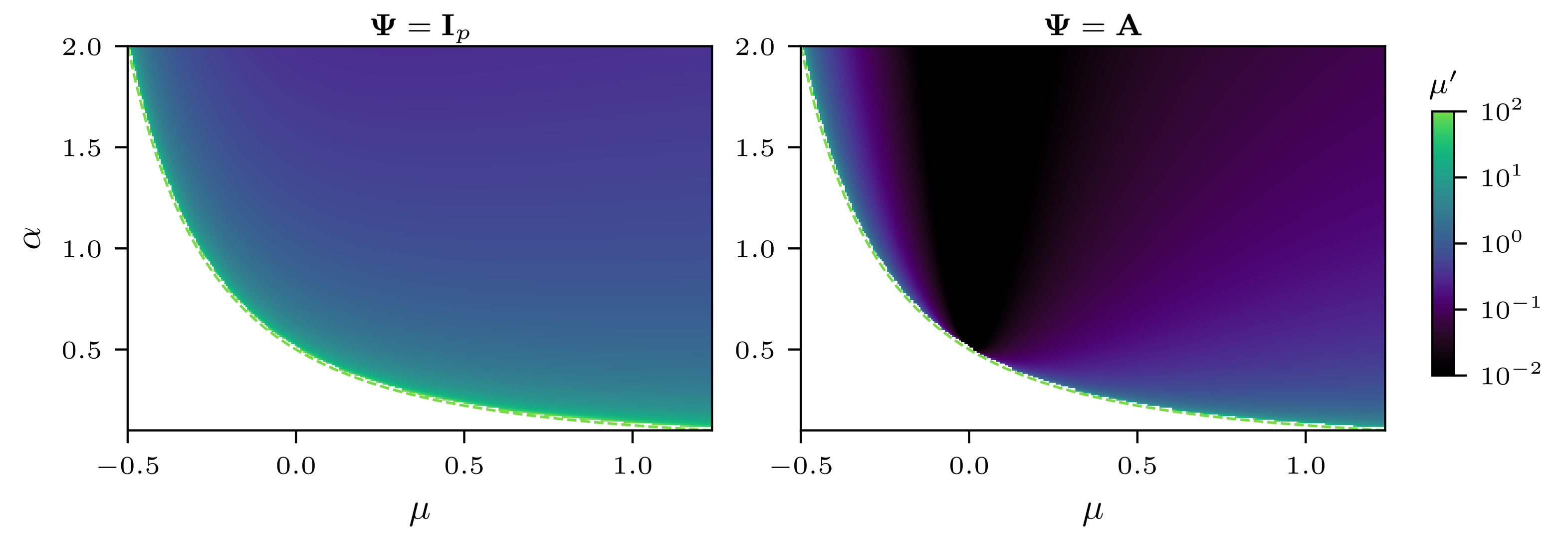


Second-order Equivalence

- **Setup:**
  - $\mathbf{A}$, $\mathbf{S}$, λ, μ, $\frac{q}{p}$ as in **Theorem 1.**
  - $\boldsymbol{\Psi} \in \mathbb{C}^{p \times p}$ is a positive semidefinite matrix independent of $\mathbf{S}$ with uniformly upper bounded $\|\boldsymbol{\Psi}\|_{\mathrm{op}}$.

> ***Theorem 2.*** *As $q, p \to \infty$,*
> $$\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} + \lambda\mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}}\boldsymbol{\Psi}\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} + \lambda\mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}} \simeq \left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1}\left(\boldsymbol{\Psi} + \mu'\mathbf{I}_p\right)\left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1},$$
> where
> $$\mu' = \frac{\frac{1}{q}\mathrm{tr}\left[\mu^3\left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1}\boldsymbol{\Psi}\left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-1}\right]}{\lambda + \frac{1}{q}\mathrm{tr}\left[\mu^2\mathbf{A}\left(\mathbf{A} + \mu\mathbf{I}_p\right)^{-2}\right]} \geq 0.$$

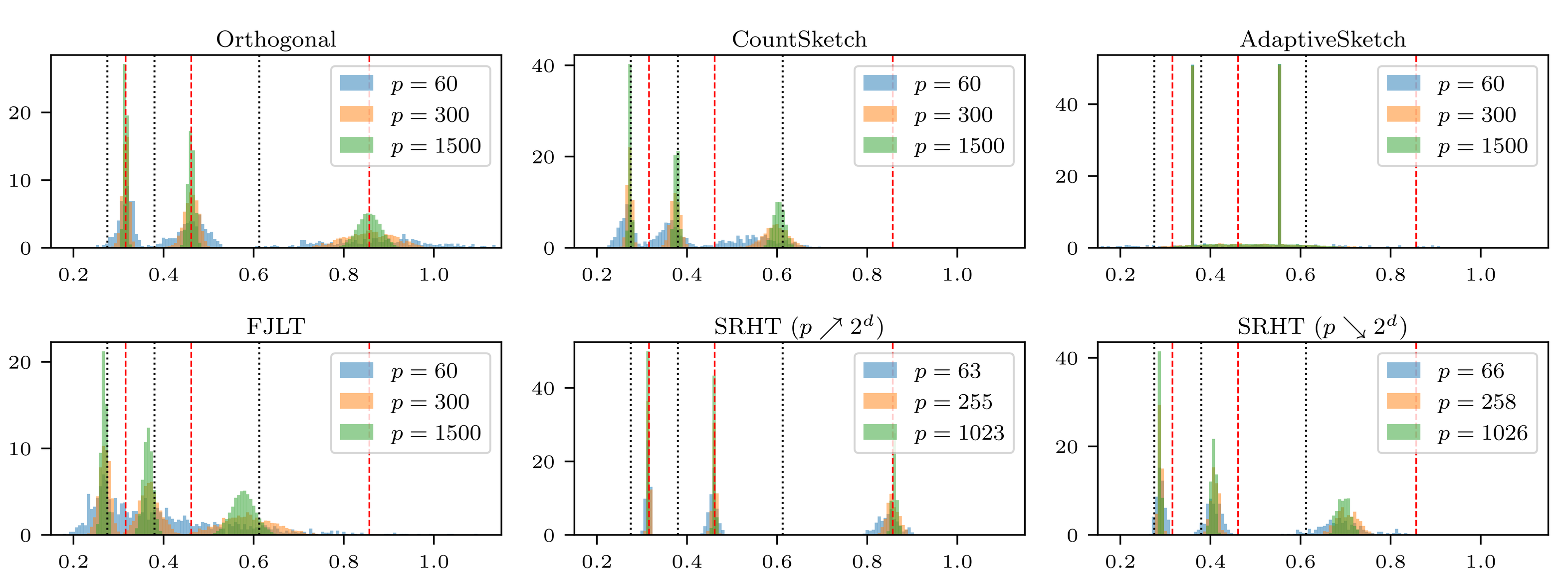
- **Inflation** μ' depends crucially on choice of $\boldsymbol{\Psi}$.
- When $\boldsymbol{\Psi} \in \mathrm{Range}(\mathbf{A})$, **no inflation** for $\mu = 0, \alpha > r(\mathbf{A})$ (rank).



General Free Sketching

> ***Conjecture 3.*** *An analogous result to **Theorem 1** holds for a general class of asymptotically free sketching matrices $\mathbf{S}$:*
> $$\mathbf{S}\left(\mathbf{S}^{\mathsf{H}}\mathbf{A}\mathbf{S} + \lambda\mathbf{I}_q\right)^{-1}\mathbf{S}^{\mathsf{H}} \simeq \left(\mathbf{A} + \gamma\mathbf{I}_p\right)^{-1},$$
> *and the map $\lambda \mapsto \gamma$ is increasing.*

- **Common sketches** used in practice **follow this conjecture**.
- We obtain the resulting γ for **orthogonal sketching**.



Conclusion

- **Interpretation:** sketching in linear systems maps regularization to effective regularization, generally **increasing regularization** and adding inflation due to randomness.
- **Impact:** Machine learning problems involving random projections induce **implicit regularization** in a precise way determined by our results.

Acknowledgments

This work was sponsored by Office of Naval Research MURI grant N00014-20-1-2787. DL, HJ, and RGB were also supported by NSF grants CCF-1911094, IIS-1838177, and IIS-1730574; ONR grants N00014-18- 12571 and N00014-20-1-2534; AFOSR grant FA9550-22-1-0060; and a Vannevar Bush Faculty Fellowship, ONR grant N00014-18-1-2047.