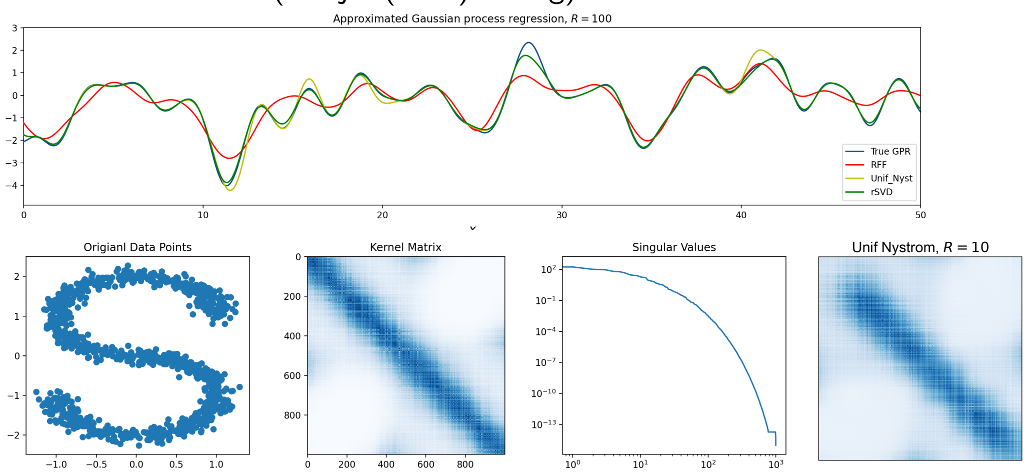
# APPM 5650
# "Randomized Algorithms"
# Prof. Becker, fall 2021
# **Student projects**

**Student backgrounds:**

- Applied Math (MS, PhD)
- Math (PhD)
- Computer Science (MS, PhD)
- Aerospace Engineering (PhD)

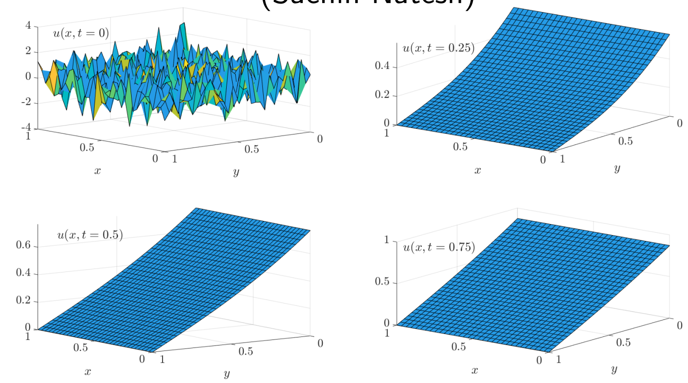
## Randomized Algorithms for Approximated Kernel Machine

(Nuojin (Noki) Cheng)




## Accelerated Local Reduced Order Basis Interpolation Applied to the Parabolic Diffusion Equation with a Random Coefficient Field

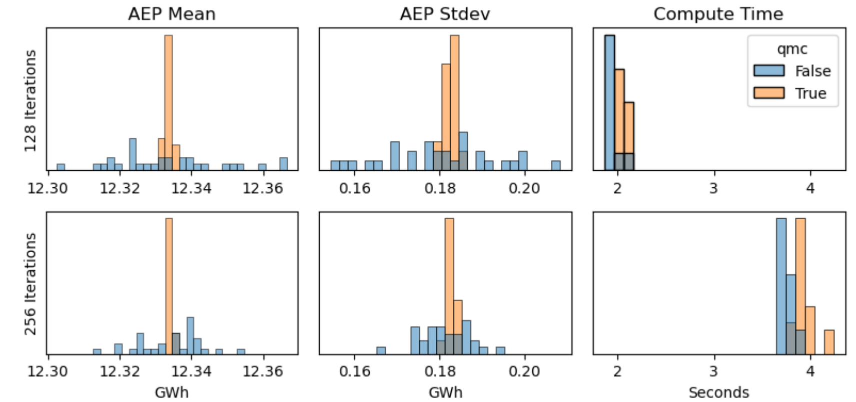
(Sachin Natesh)

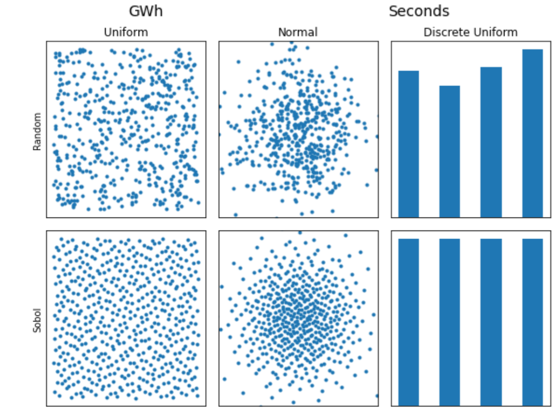



## Empirical Results Suggest Quasi-Monte Carlo Sampling Increases Accuracy in OpenOA AEP

(Jordan Perr-Sauer)

| Parameter | Distribution |
|---|---|
| Meter | $N(1, 0.005)$ |
| Loss | $N(1, 0.05)$ |
| Windiness | $\lfloor \text{Unif}(10, 21) \rfloor$ |
| Loss Threshold | $\lfloor \text{Unif}(10, 21) \rfloor$ |
| Reanalysis Product | $\lfloor \text{Unif}(0, 1) \rfloor$ |




## Fast and Randomized Principle Component Analysis

(Ayoub Ghriss)

**Theorem 1.** Define the $d \times d$ matrix $C$ as $\frac{1}{n}X^\top X = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$, and let $V_k$ denote the $d \times k$ matrix composed of the eigenvectors corresponding to the largest $k$ eigenvalues. Suppose that

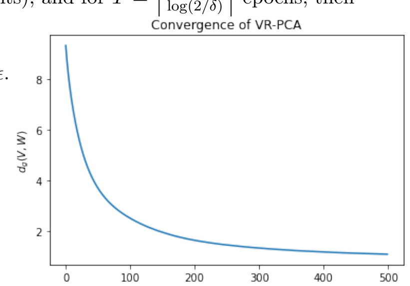
- $\max_i \|\mathbf{x}_i\|^2 \le r$ for some $r > 0$.
- $C$ has eigenvalues $s_1 > s_2 \ge \ldots \ge s_d$, where $s_k - s_{k+1} = \lambda$ for some $\lambda > 0$.
- $k - \|V_k^\top \tilde{W}_0\|_F^2 \le \frac{1}{2}$.

Let $\delta, \epsilon \in (0, 1)$ be fixed. If we run the algorithm with any epoch length parameter $m$ and step size $\eta$, such that

$$\eta \le \frac{c\delta^2}{r^2}\lambda \quad , \quad m \ge \frac{c' \log(2/\delta)}{\eta\lambda} \quad , \quad km\eta^2r^2 + rk\sqrt{m\eta^2\log(2/\delta)} \le c'' \qquad (9)$$

(where $c, c', c''$ designate certain positive numerical constants), and for $T = \left\lceil \frac{\log(1/\epsilon)}{\log(2/\delta)} \right\rceil$ epochs, then with probability at least $1 - \lceil \log_2(1/\epsilon) \rceil \delta$, it holds that

$$k - \|V_k^\top \tilde{W}_T\|_F^2 \le \epsilon.$$




## On Convergence of Stochastic Gradient Descent with Adaptive Step Sizes, from Li and Orabona '19

(Spencer Shortt)

Global generalized AdaGrad: $\eta_t = \dfrac{\alpha}{\left(\beta + \sum_{i=1}^{t-1} ||\mathbf{g}(\mathbf{x}_i, \xi_i)||^2\right)^{1/2+\epsilon}}$

**Theorem 4**: Assume (H1, H3, H4'). Let $\eta_t$ be our global generalized AdaGrad stepsize from before, where $\alpha, \beta > 0$ and $\epsilon \in (0, 1/2)$, and $4\alpha M < \beta^{1/2+\epsilon}$. Then the iterates of SGD satisfy the following bound:

$$\mathbb{E}[\min_{1 \le t \le T} ||\nabla f(\mathbf{x}_t)||^{1-2\epsilon}] \le$$

$$\frac{1}{T^{1/2-\epsilon}} \max\left(2^{\frac{1/2+\epsilon}{1/2-\epsilon}}\gamma, 2^{1/2+\epsilon}(\beta + 2T\sigma^2)^{1/4-\epsilon^2}\gamma^{1/2-\epsilon}\right).$$

## A Nonlinear Extension to Kalman GD:

### Unsuccessfully Attempting to Combine Uncertainty Quantification with Variance Reduction

(Mike McCabe)

So why is filtering stochastic gradients a bad idea?

**Ideal Kalman SG**

Dynamics Model / Observation Model

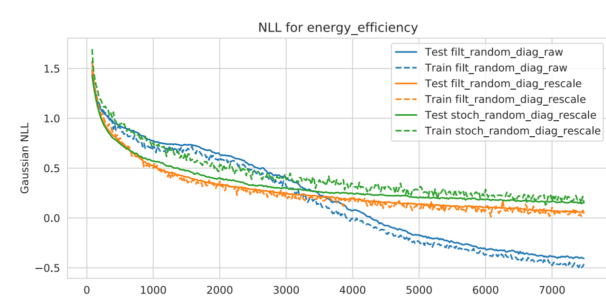
$$\begin{bmatrix}\dot{\theta} \\ \nabla \dot{f}(\theta)\end{bmatrix} = \begin{bmatrix}0 & -\alpha I \\ 0 & -\alpha\nabla^2 f(\theta)\end{bmatrix}\begin{bmatrix}\theta \\ \nabla f(\theta)\end{bmatrix} \quad y = Hx = [0 \quad I]x + \eta$$

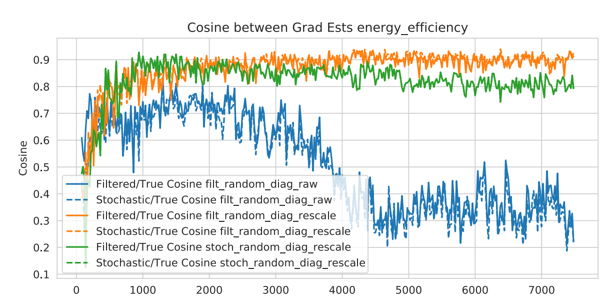
**Actual Kalman SG**

Dynamics Model / Observation Model

$$\begin{bmatrix}\Delta\tilde{\theta}_k \\ \Delta d_k\end{bmatrix} = \begin{bmatrix}0 & -\alpha I \\ 0 & -\alpha\nabla^2 f\left(\tilde{\theta}_{k-1}\right)\end{bmatrix}\begin{bmatrix}\tilde{\theta}_{k-1} \\ d_{k-1}\end{bmatrix} \quad y_k = Hx_k = [0 \quad I]x_k^f + \eta$$

AKA there is no true system being tracked, so pretty much all of the assumptions necessary for sequential filtering do not hold in practice.




## Accelerated Proper Orthogonal Decomposition for Turbulent Flows

(Aviral Prakash)

$$\mathbf{X} = \left[\mathbf{x}(t_1) \quad \mathbf{x}(t_2) \quad \mathbf{x}(t_3) \quad . \quad . \quad . \quad . \quad \mathbf{x}(t_n)\right]$$