# Approximating Orthogonal Matrices with Effective Givens Factorization

**Thomas Frerix**

Technical University of Munich

*joint work with*

**Joan Bruna**

(NYU)

*Poster #164*

# Givens Factorization of Orthogonal Matrices

$$G^T(i,j,\alpha) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\alpha) & \cdots & -\sin(\alpha) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin(\alpha) & \cdots & \cos(\alpha) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

# Givens Factorization of Orthogonal Matrices

$$G^T(i,j,\alpha) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\alpha) & \cdots & -\sin(\alpha) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin(\alpha) & \cdots & \cos(\alpha) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

## Exact Givens Factorization

$$U = G_1 \dots G_N \qquad N = \frac{d(d-1)}{2}$$

# Approximate Givens Factorization

## Approximate Givens Factorization

$$U \approx G_1 \dots G_N \qquad N \ll \frac{d(d-1)}{2}$$

computationally hard problem

# Approximate Givens Factorization

## Approximate Givens Factorization

$$U \approx G_1 \dots G_N \qquad N \ll \frac{d(d-1)}{2}$$

computationally hard problem

## Our Questions in this Context

1. Which orthogonal matrices can be effectively approximated?
   (not all of them)

# Approximate Givens Factorization

## Approximate Givens Factorization

$$U \approx G_1 \dots G_N \qquad N \ll \frac{d(d-1)}{2}$$

computationally hard problem

## Our Questions in this Context

1. Which orthogonal matrices can be effectively approximated?
   (not all of them)
2. Which principles are behind effective approximation algorithms?
   (sparsity-inducing algorithms)

# Motivation: Unitary Basis Transform / FFT

**Advantageous Setting**

Once computed, applied many times

# Motivation: Unitary Basis Transform / FFT

### Advantageous Setting

Once computed, applied many times

### Unitary Basis Transform

FFT: $\mathcal{O}\left(d^2\right) \rightarrow \mathcal{O}\left(d \log(d)\right)$

# Motivation: Unitary Basis Transform / FFT

## Advantageous Setting

Once computed, applied many times

## Unitary Basis Transform

FFT: $\mathcal{O}\left(d^2\right) \rightarrow \mathcal{O}\left(d \log(d)\right)$

Application: Graph Fourier Transform

# Which Matrices can be Effectively Approximated?

**Theorem**

*Let* $\epsilon > 0$*. If* $N = o\left(d^2 / \log(d)\right)$*, then as* $d \to \infty$*,*

$$\mu\left(\left\{U \in U(d) \middle| \inf_{G_1 \ldots G_N} \|U - \prod_n G_n\|_2 \leq \epsilon\right\}\right) \to 0 \ ,$$

*where* $\mu$ *is the Haar measure over* $U(d)$*.*

# Which Matrices can be Effectively Approximated?

**Theorem**

*Let* $\epsilon > 0$. *If* $N = o\big(d^2/\log(d)\big)$, *then as* $d \to \infty$,

$$\mu\left(\left\{U \in U(d) \middle| \inf_{G_1 \dots G_N} \|U - \prod_n G_n\|_2 \leq \epsilon\right\}\right) \to 0 \ ,$$

*where* $\mu$ *is the Haar measure over* $U(d)$.

- proof is based on an $\epsilon$-covering argument
- suggests computational-to-statistical gap together with experimental results (details at poster)

## $K$-planted Distribution over $SO(d)$

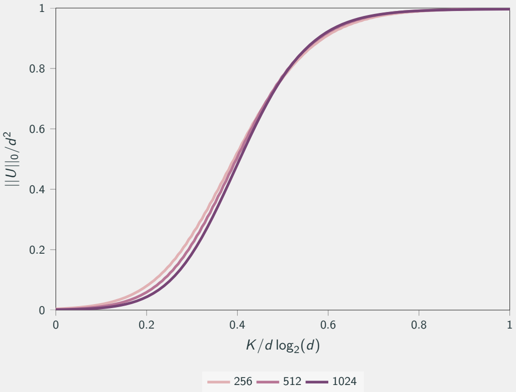
**Sample** $U = G_1 \dots G_K$

- choose subspace $(i_k, j_k)$ uniformly with replacement
- choose rotation angle $\alpha_k \in [0, 2\pi)$ uniformly

# $K$-planted Distribution over $SO(d)$

**Sample $U = G_1 \dots G_K$**

- choose subspace $(i_k, j_k)$ uniformly with replacement
- choose rotation angle $\alpha_k \in [0, 2\pi)$ uniformly




$K$-planted matrices quickly become dense

# Minimizing Sparsity-Inducing Norms over $O(d)$

$$G_N^T \dots G_N^T U \approx I \qquad \hat{U} = G_1 \dots G_N$$

# Minimizing Sparsity-Inducing Norms over $O(d)$

$$G_N^T \dots G_N^T U \approx I \qquad \hat{U} = G_1 \dots G_N$$

## Approximation criterion

$$\left\| U - \hat{U} \right\|_{F,\mathrm{sym}} := \min_{P \in \mathcal{P}_d} \left\| U - \hat{U} P \right\|_F$$

# Minimizing Sparsity-Inducing Norms over $O(d)$

$$G_N^T \dots G_N^T U \approx I \qquad \hat{U} = G_1 \dots G_N$$

## Approximation criterion

$$\left\| U - \hat{U} \right\|_{F,\mathrm{sym}} := \min_{P \in \mathcal{P}_d} \left\| U - \hat{U} P \right\|_F$$

## Better functions to be minimized greedily?

$$f(U) := d^{-1} \|U\|_1 = d^{-1} \sum_{i,j=1}^{d} |U_{ij}|$$

# Minimizing Sparsity-Inducing Norms over $O(d)$

$$G_N^T \dots G_N^T U \approx I \qquad \hat{U} = G_1 \dots G_N$$

## Approximation criterion

$$\left\| U - \hat{U} \right\|_{F,\mathrm{sym}} := \min_{P \in \mathcal{P}_d} \left\| U - \hat{U} P \right\|_F$$

## Better functions to be minimized greedily?

$$f(U) := d^{-1} \|U\|_1 = d^{-1} \sum_{i,j=1}^{d} |U_{ij}|$$

- *Non-convex* greedy step
- global optimum in $\mathcal{O}(d^2)$ amortized time complexity

# Thank you

*Poster #164*

`https://github.com/tfrerix/givens-factorization`