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

\[ G^T(i, j, \alpha) = \begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & \cos(\alpha) & \cdots & -\sin(\alpha) & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \sin(\alpha) & \cdots & \cos(\alpha) & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix} \]
Givens Factorization of Orthogonal Matrices

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

Exact Givens Factorization

\[ U = G_1 \cdots G_N \quad N = \frac{d(d - 1)}{2} \]
**Approximate Givens Factorization**

\[ U \approx G_1 \ldots G_N \quad N \ll \frac{d(d - 1)}{2} \]

*computationally hard problem*
Approximate Givens Factorization

\[
U \approx G_1 \ldots G_N \quad N \ll \frac{d(d - 1)}{2}
\]

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

\[
U \approx G_1 \ldots G_N \quad N \ll \frac{d(d-1)}{2}
\]

This is a 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)
<table>
<thead>
<tr>
<th>Advantageous Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once computed, applied many times</td>
</tr>
</tbody>
</table>
Motivation: Unitary Basis Transform / FFT

<table>
<thead>
<tr>
<th>Advantageous Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once computed, applied many times</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unitary Basis Transform</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT: $O(d^2) \rightarrow O(d \log(d))$</td>
</tr>
</tbody>
</table>
# Motivation: Unitary Basis Transform / FFT

## Advantageous Setting
- Once computed, applied many times

## Unitary Basis Transform
- **FFT**: $O(d^2) \rightarrow O(d \log(d))$
  
  **Application**: Graph Fourier Transform
**Theorem**

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

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

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

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

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

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 $SO(d)$**

<table>
<thead>
<tr>
<th>Sample $U = G_1 \ldots G_K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>• choose subspace $(i_k, j_k)$ uniformly with replacement</td>
</tr>
<tr>
<td>• choose rotation angle $\alpha_k \in [0, 2\pi)$ uniformly</td>
</tr>
</tbody>
</table>

$K$-planted matrices quickly become dense.
**K-planted Distribution over \( SO(d) \)**

**Sample** \( U = G_1 \ldots G_K \)

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

\[
\|U\|_0 / d^2
\]

\( K \)-planted matrices quickly become dense
Minimizing Sparsity-Inducing Norms over \( O(d) \)

\[ G_N^T \ldots G_2^T U \approx I \quad \hat{U} = G_1 \ldots G_N \]
Minimizing Sparsity-Inducing Norms over $O(d)$

$G_N^T \ldots G_N^T U \approx I \quad \hat{U} = G_1 \ldots G_N$

**Approximation criterion**

$$\| U - \hat{U} \|_{F,\text{sym}} := \min_{P \in \mathcal{P}_d} \| U - \hat{U}P \|_F$$
Minimizing Sparsity-Inducing Norms over $O(d)$

$$G_N^T \ldots G_1^T U \approx I \quad \hat{U} = G_1 \ldots G_N$$

**Approximation criterion**

$$\| U - \hat{U} \|_{F,\text{sym}} := \min_{P \in P_d} \| U - \hat{U}P \|_F$$

**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 $O(d)$

\[ G_N^T \ldots G_N^T U \approx I \quad \hat{U} = G_1 \ldots G_N \]

**Approximation criterion**

\[
\| U - \hat{U} \|_{F,\text{sym}} := \min_{P \in P_d} \| U - \hat{U}P \|_F
\]

**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 $O(d^2)$ amortized time complexity
Thank you

Poster #164

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