Fast Tree-Field Integrators: From Low Displacement Rank to Topological Transformers

Google DeepMind

Krzysztof Choromanski*, Arijit Sehanobish*, Somnath Basu Roy Chowdhury*, Han Lin*, Avinava Dubey*, Tamas Sarlos, Snigdha Chaturvedi

Columbia University

Google Research

Problem: Tree-Field Integration

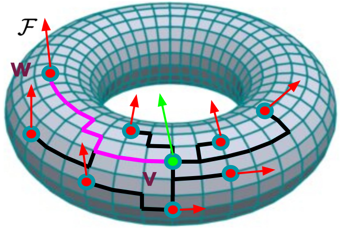
Compute efficiently (in **sub-quadratic** time in the number of nodes **N**) the following expressions **for every node v** of the given **tree T**, approximating graph G (e.g. *low-distortion-tree*):

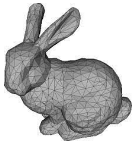
$$i(v) := \sum_{w \in V} K(w,v)\mathcal{F}(w)$$

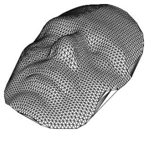
- $w \in V$: integration over all the nodes
- $K(w,v)$: similarity between two nodes (e.g. a function of the **shortest-path distance** between them)
- $\mathcal{F}(w)$: tensor field defined on the graph

$$K(w,v) = f_\theta(d_G(w,v))$$

- f_θ: learnable
- $d_G(w,v)$: shortest-path distance

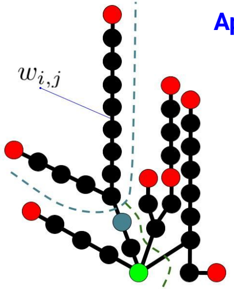






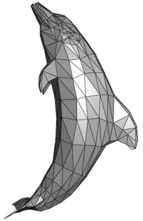
1. Popular trees used here: *minimum spanning trees* (MSTs); can be constructed in log-linear time for sparse graphs.

2. Brute-force integration with them still takes **quadratic** time.



Applications

- interpolation on manifolds,
- **topological masking mechanisms for Transformers** with structural inputs,
- physics simulations in curved spaces,
- Wasserstein barycenter, (Fused) Gromov Wasserstein

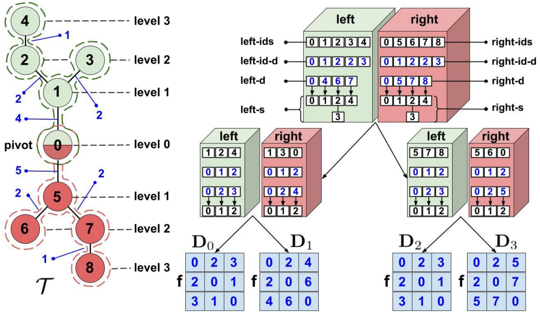


* equal contribution

Fast Tree-Field Integrators (FTFIs)

1. An **polylog-linear** field-integration algorithm for a large class of functions f, fusing *divide-and-conquer* and *FFT* methods.

2. New data structure to achieve it: **integration trees** (**ITs**).



Experiments: interpolation on meshes

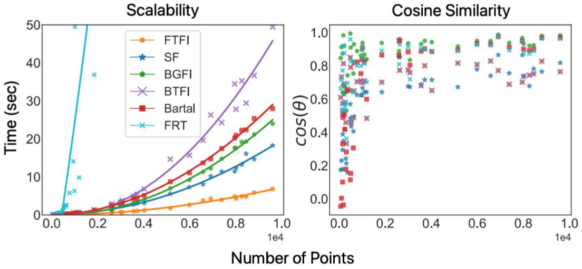


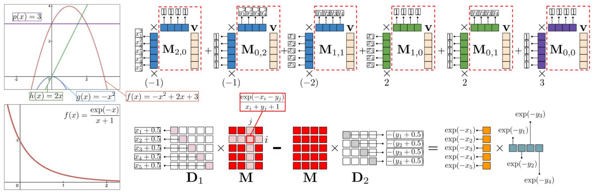
Fig.: Speed (pre-processing time) and accuracy (cosine similarity) comparison of the FTFI and other baselines for vertex normal prediction on meshes. Cosine similarity of BFFI and FTFI almost overlaps.

Cordial Functions f_θ

Definition [cordial functions] A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is *d-cordial* (or: *cordial* if d is not specified), if there exists $d \in \mathbb{N}$ such that matrix-vector multiplication with a matrix $\mathbf{M} = [f(x_i + y_j)]_{i=1,\ldots,a}^{j=1,\ldots,b}$ can be conducted in time $O((a+b)\log^d(a+b))$ for every $(x_i)_{i=1}^a$, $(y_j)_{j=1}^b$.

Lemma 3.3 [integration with cordial functions] *If f is d-cordial then f-integration for the general weighted tree of N vertices can be conducted in time $O(N \log^{d+1}(N))$.*

Examples: trigonometric (sin/cos), rational, exponentials and polynomials products



Experiments: Topological Transformers

	ImageNet															Place365			
ϕ :=RELU				$\phi := x \rightarrow x^2$				$\phi := x \rightarrow x^4$				$\phi :=$ exp				$\phi :=$ ReLU			
synced	**g**	**t**	**Acc. (%)**	**synced**	**g**	**t**	**Acc. (%)**	**synced**	**g**	**t**	**Acc. (%)**	**synced**	**g**	**t**	**Acc. (%)**	**synced**	**g**	**t**	**Acc. (%)**
NA	NA	NA	76.23	NA	NA	NA	75.03	NA	NA	NA	76.37	NA	NA	NA	76.76	NA	NA	NA	54.80
✓	exp	1	77.28	✓	exp	1	76.66	✓	exp	1	77.84	✗	exp	1	**78.79**	✗	exp	1	56.69
✓	exp	2	76.60	✓	exp	2	75.91	✓	exp	2	77.23	✗	exp	2	78.51	✗	$z \rightarrow z^{-1}$	1	56.44
✗	exp	1	**77.79**	✗	exp	1	**76.76**	✗	exp	1	77.94	✗	$z \rightarrow z^{-1}$	1	77.39	✗	$z \rightarrow z^{-1}$	5	56.32
✗	exp	2	77.43	✗	exp	2	76.27	✗	exp	2	**78.15**	✗	$z \rightarrow z^{-1}$	2	77.69	✗	$z \rightarrow z^{-1}$	10	56.51

Table: Performance of Topological Vision Transformers with tree-based masking. For each attention kernel, we present the results of the best variant in **bold** and Performer baselines in blue.

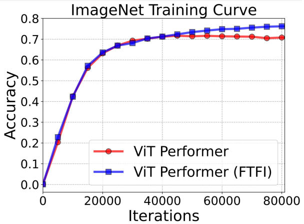


Fig.: Experiments with the RPE mechanism for ViT-L and on ImageNet. We observe that FTFI provides **7%** accuracy gain compared to the Performer variant.