



Национальный
исследовательский

**Томский
государственный
университет**



**РЕГИОНАЛЬНЫЙ
НАУЧНО-
ОБРАЗОВАТЕЛЬНЫЙ
МАТЕМАТИЧЕСКИЙ
ЦЕНТР**



Симметричные гиперболические вложения данных в машинном обучении

“Алгебраическая топология, гиперболическая геометрия и компьютерный анализ данных” – Томск, 05-09 декабря 2023

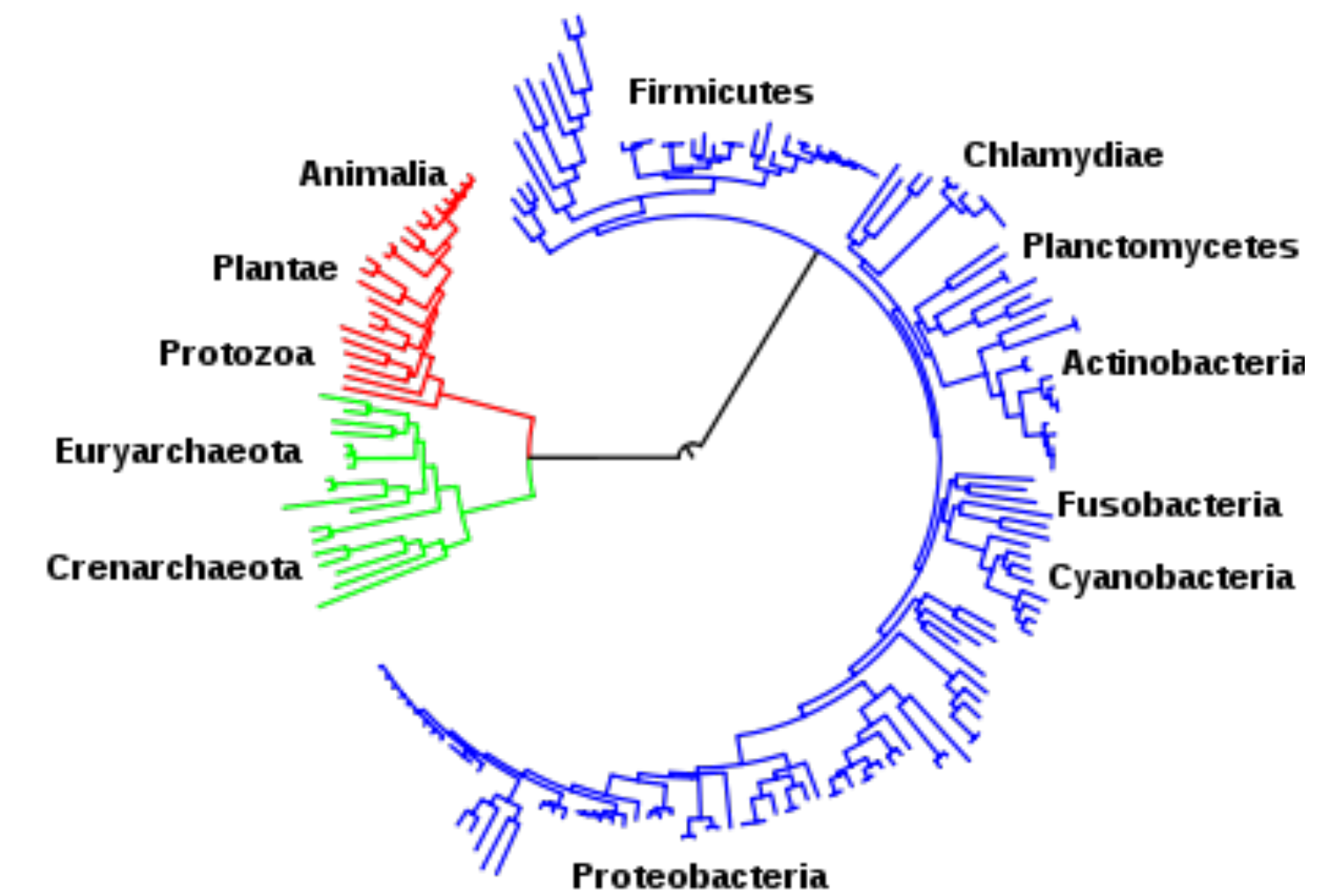
конференция совместного проекта ВШЭ и ТГУ “**Зеркальные Лаборатории**”

**Максим Бекетов, стажер-исследователь и аспирант,
Международная Лаборатория Алгебраической Топологии и её Приложений ФКН ВШЭ**

Мотивация

Во многих данных есть **иерархические связи**:

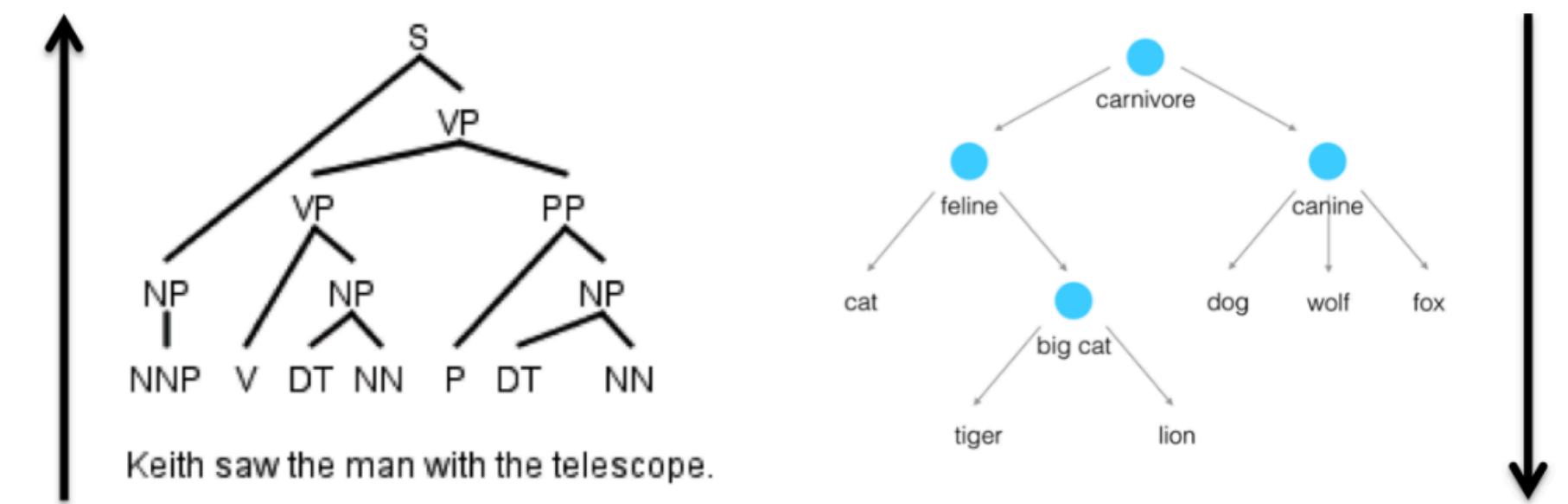
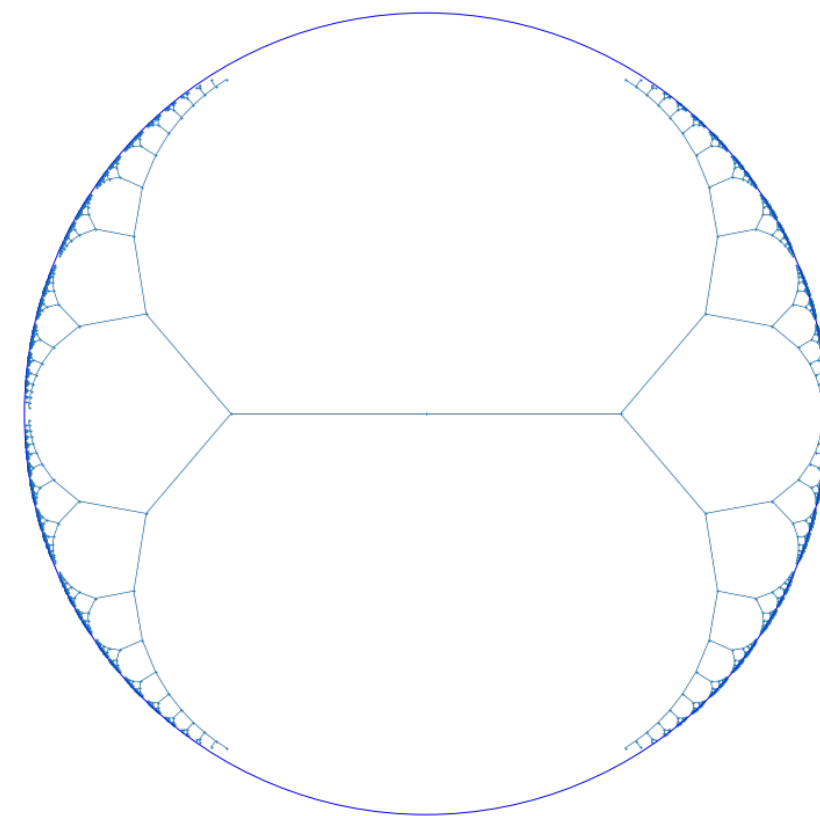
- социальные сети (родства / рабочих отношений)
- филогенетические деревья (родства биологических видов)
- слова в предложениях естественного языка
- экспертно-определенные dissimilarity measures*



филогенетическое дерево

а деревья хорошо вкладываются
в гиперболическое пр-во:

т.к. $Vol(B_r) \sim \exp(r)$



синтаксические деревья

*Duin R. P. W., Pełalska E. Non-euclidean dissimilarities: Causes and informativeness //Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, SSPR&SPR 2010, Cesme, Izmir, Turkey, August 18-20, 2010. Proceedings. – Springer Berlin Heidelberg, 2010. – С. 324-333.

Эмбединги: word2vec

Объекты отображаются в вектора.

Например, word2vec*

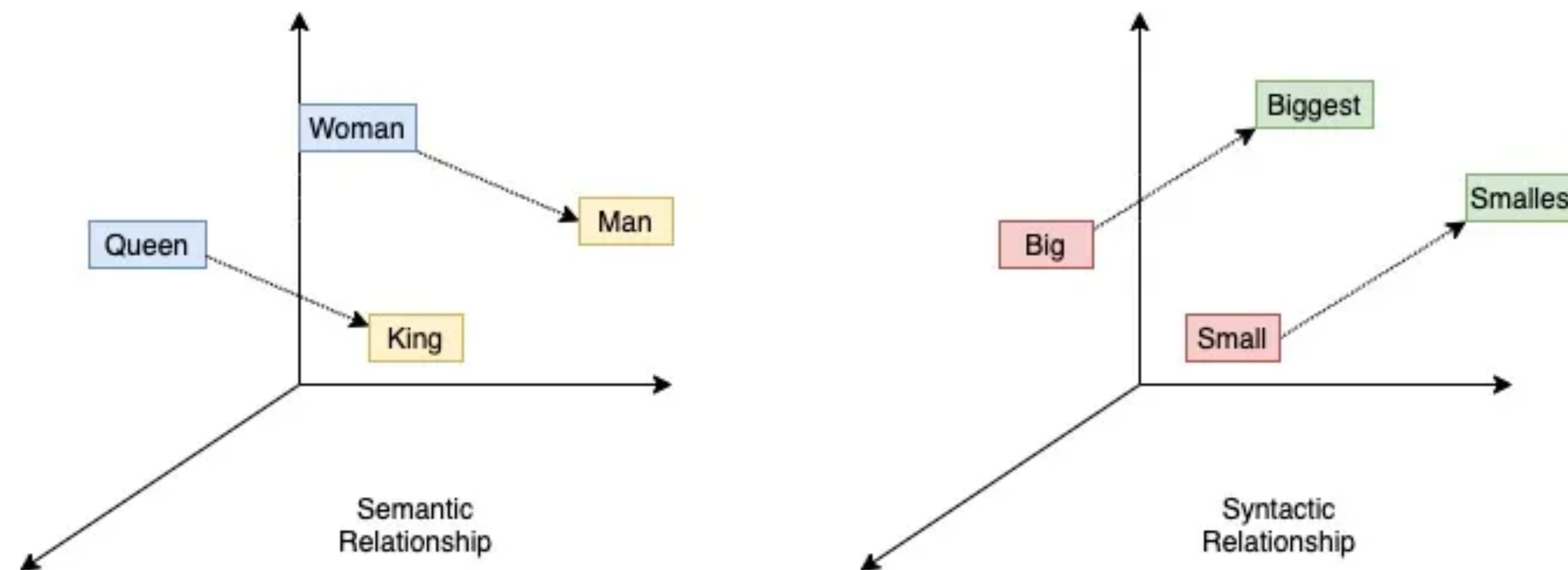
Слово $u \in \mathcal{V}$ из словаря \mathcal{V} – дается слоем нейросети с набором параметров $\alpha \in \mathbb{R}^{d \times |\mathcal{V}|}$, на выходе которой $\beta \in \mathbb{R}^{d \times |\mathcal{V}|}$

Пусть $w_0 \in \mathcal{V}$ – слово **контекста**, и $\{w_1, \dots, w_k\}$ – шумные примеры слов контекста.

Обучение = максимизация $\mathcal{L}_{u, w_0}(\alpha, \beta) = \prod_{i=0}^k P(\delta_{i0} | w_i, u) = \prod_{i=0}^k \sigma((-1)^{1-\delta_{i0}} \langle \alpha_u, \beta_{w_i} \rangle_{\mathbb{R}^d})$

(часто встречающиеся вместе слова – отображаются в близкие вектора)

после этого эмбединги слов можно складывать/вычитать (на них есть структура \mathbb{R}^n)



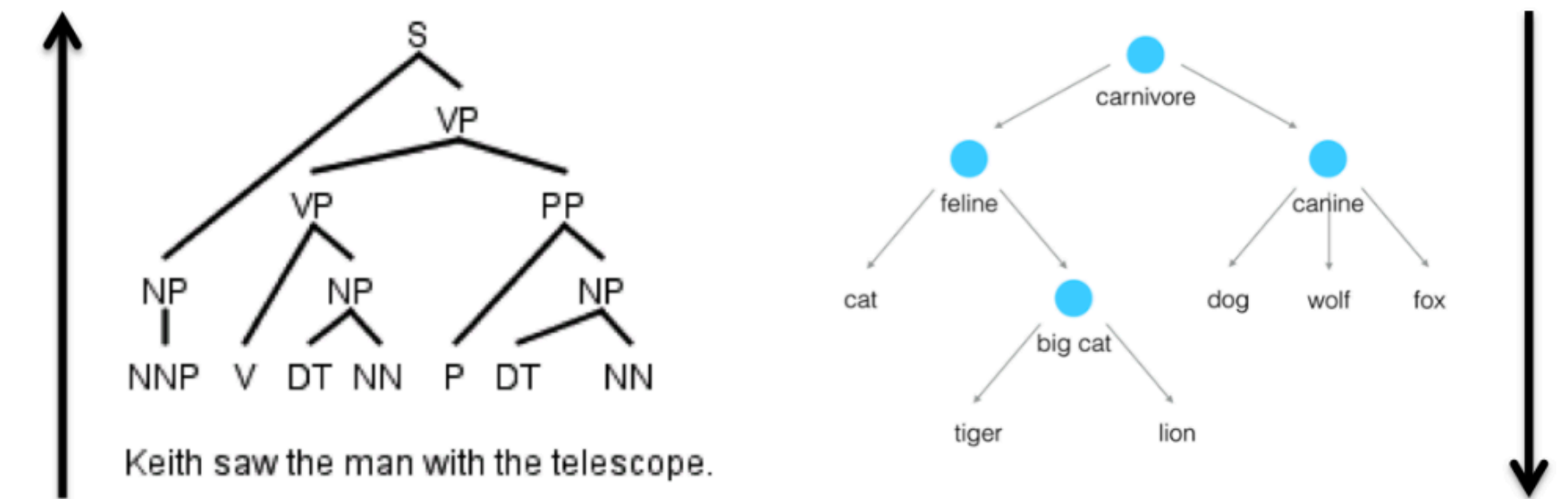
* Mikolov T. et al. Efficient estimation of word representations in vector space //arXiv preprint arXiv:1301.3781. – 2013. (~38k citations)

Идея: гиперболические эмбединги

Раз на словах есть иерархические связи, скалярное произведение в \mathbb{R}^n – плохая мера “близости”.

Идея: заменим $\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$ на $\langle p, q \rangle_{\mathbb{H}^d} = -\cosh(d_{\mathbb{H}^d}(p, q))$

(в целевой функции $\mathcal{L}_{u, w_0}(\alpha, \beta)$)



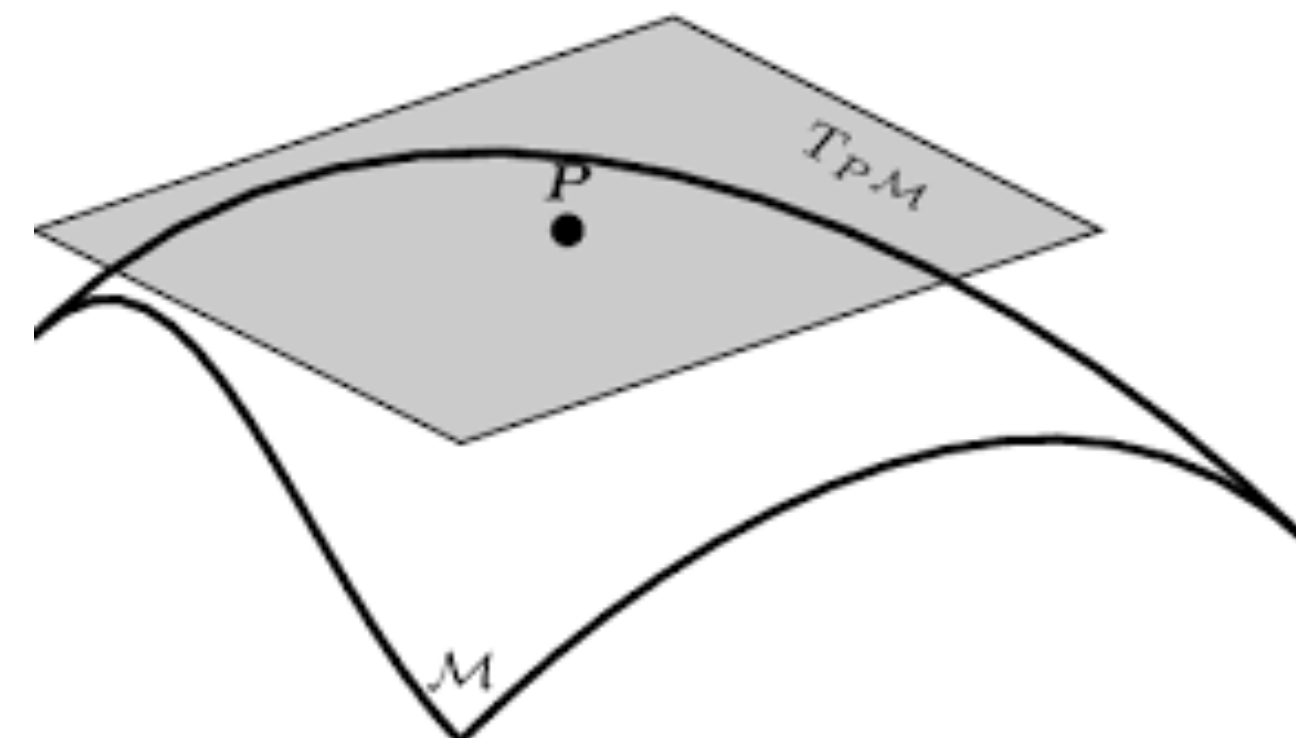
синтаксические деревья

Для обучения (максимизации \mathcal{L}) нужно уметь считать её производные по α, β

для этого градиенты $\nabla_{\alpha}, \nabla_{\beta} \log \mathcal{L}$ – надо просто **спроецировать** на $T_p \mathbb{H}^d$

$$\text{proj}_p(v) = v + \langle p, v \rangle_{\mathbb{H}^d} \cdot p$$

То есть структура векторного пр-ва на словах будет в касательном пр-ве



* Leimeister M., Wilson B. J. Skip-gram word embeddings in hyperbolic space //arXiv preprint arXiv:1809.01498. – 2018.

Идея: гиперболические эмбединги

В $T_p M$ можно работать с т.н. гировекторами (Мёбиуса) (в модели \mathbb{B})

их можно складывать $x \oplus y = \frac{(1 + 2\langle x, y \rangle + \|y\|^2)x + (1 - \|x\|^2)y}{1 + 2\langle x, y \rangle + \|x\|^2 \|y\|^2}$

вычитать $x \ominus y = x \oplus (-y)$

умножать на скаляр $r \otimes x = \text{Exp}_0(r \text{Log}_0(x))$

умножать на матрицу $M^{\otimes}(x) = \tanh\left(\frac{\|Mx\|}{\|x\|} \text{atanh}\|x\|\right) \frac{Mx}{\|Mx\|}$

Модели гиперболической геометрии

1) Лоренца

$$\mathbb{L}^n = \{x = (x_0, \dots, x_n) \in \mathbb{R}^{n+1} : \langle x, x \rangle_{\mathbb{L}} = -1, x_0 > 0\}$$

$$\text{где } \langle x, y \rangle_{\mathbb{L}} = x^T g^L y, \quad g^L = \text{diag}(1, \dots, 1, -1)$$

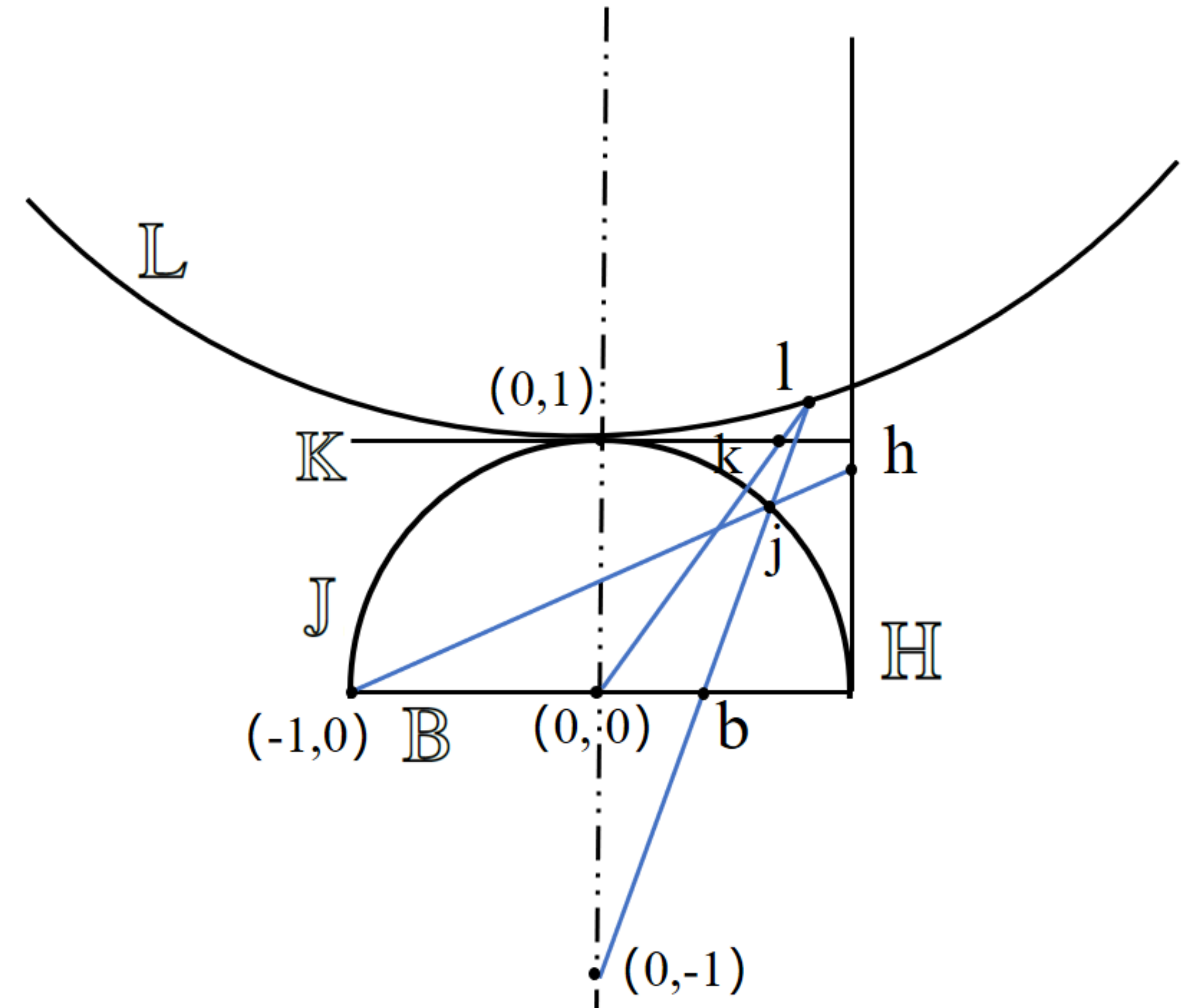
$$d(x, y) = \text{acosh}(-\langle x, y \rangle_{\mathbb{L}})$$

2) Пуанкаре $\mathbb{B}^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$

$$d(x, y) = \text{acosh}\left(1 + 2 \frac{\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)}\right)$$

$$g^{\mathbb{B}} = \lambda_x^2 g^{\mathbb{R}^n} \quad \text{где } \lambda_x = 2/(1 - \|x\|^2)$$

– конформна \mathbb{R}^n , что очень удобно



есть также модели Пуанкаре \mathbb{H}^n (полупр-ва), Клейна \mathbb{K}^n , полусферы \mathbb{J}^n , но Лоренц и Пуанкаре наиболее удобны

Проблема: стабильность гиперб. вложений

При численном построении гиперболических вложений мы работаем с фиксированной точностью (числом разрядов float-ов).

Из-за растяжения метрики вдали от центра, вложения далёких точек **нестабильны**

– вложение плохо* сохраняет метрику:

$$\text{Av. Distortion} = \frac{1}{\binom{n}{2}} \sum_{x,y \in X} \frac{1}{d(x,y)} \left| d_H(f(x), f(y)) - d(x,y) \right|$$

Казалось бы, можно запастись достаточной точностью; или сделать точность зависящей от этого удаления,

но есть и другая идея!

Mishne G. et al. The numerical stability of hyperbolic representation learning //International Conference on Machine Learning. – PMLR, 2023. – С. 24925-24949.

Sala F. et al. Representation tradeoffs for hyperbolic embeddings //International conference on machine learning. – PMLR, 2018. – С. 4460-4469.

* Linial N., London E., Rabinovich Y. The geometry of graphs and some of its algorithmic applications //Combinatorica. – 1995. – Т. 15. – С. 215-245.

Гиперболические вложения замощениями!

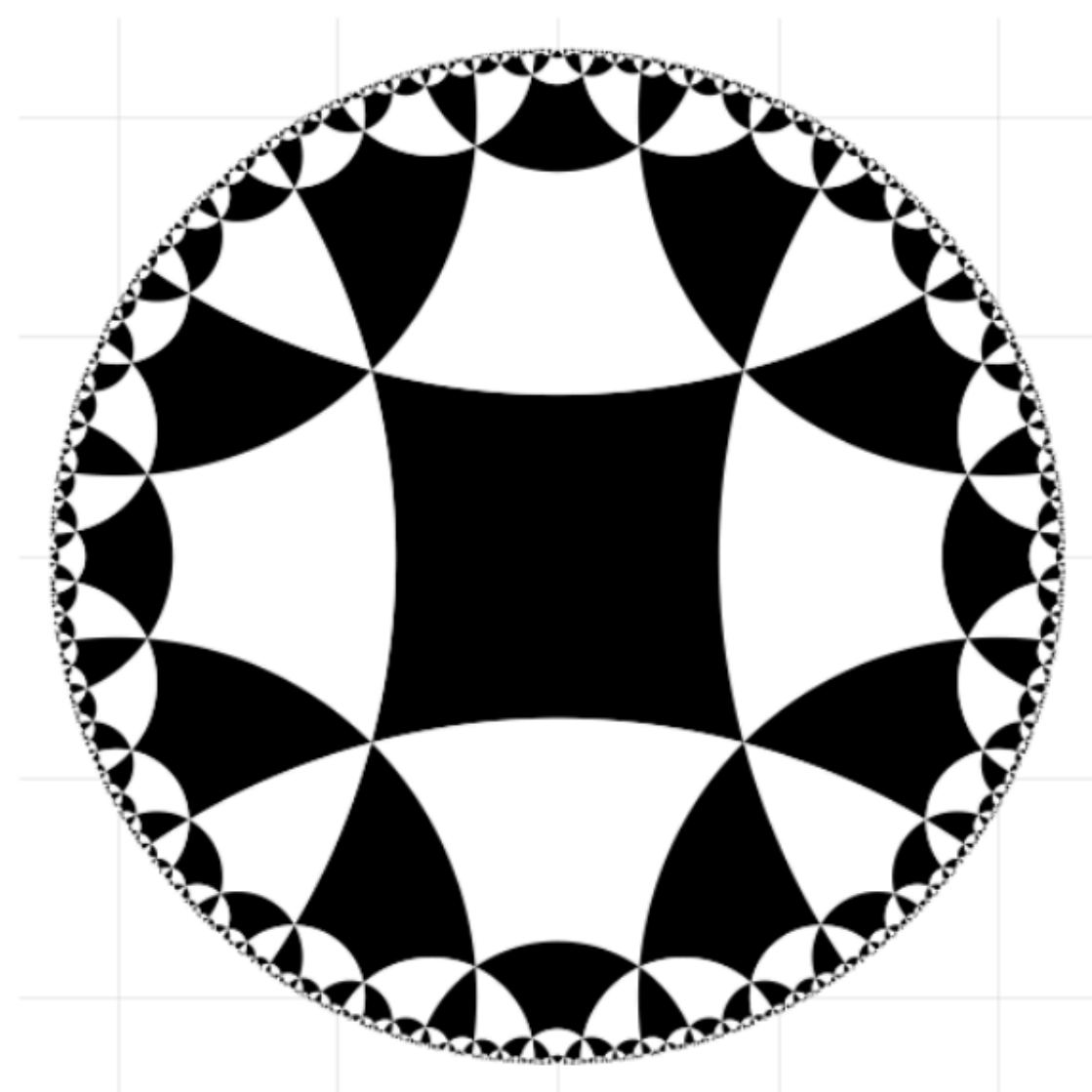
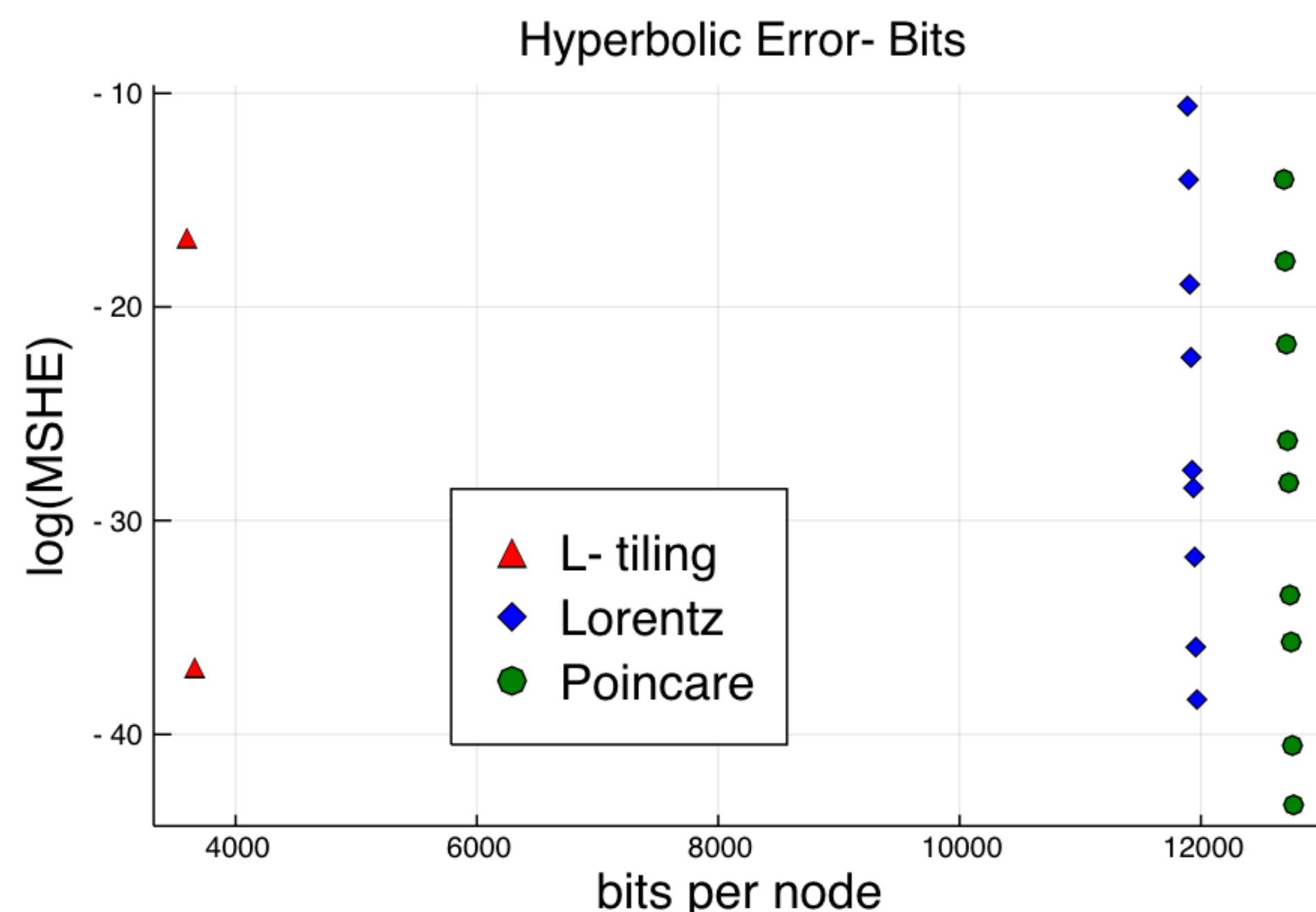


Figure 1: The regular quadrilateral tiling of hyperbolic space produced by the group G on the Poincaré disk.

Algorithm 1 Map Lorentz model to L -tiling model

Require: $x \in \mathcal{L}^2$
initialize $R \leftarrow I$
while $x \notin F$ **do**
 if $x_2 \leq -|x_3|$ **then** $S \leftarrow g_a^{-1}$
 else if $x_2 \geq |x_3|$ **then** $S \leftarrow g_b^{-1}$
 else if $x_3 < -|x_2|$ **then** $S \leftarrow g_b$
 else if $x_3 > |x_2|$ **then** $S \leftarrow g_a$
 $(R, x) \leftarrow (R \cdot S, L \cdot S^{-1} \cdot L^{-1} \cdot x)$
 $x_1 = \sqrt{x_2^2 + x_3^2 + 1}$ \triangleright renormalize x
end while
output (R, x)

Гиперболические вложения замощениями!



Models	size (MB)	bzip (MB)
Poincaré(16512B)	372	119
Poincaré(12688B)	287	81
Lorentz(11898B)	396	171
Matrix(VLQ)	600(286)	260(251)
Entries(VLQ)	132(63)	57(55)
Order	111	8.52
VBW	33.1	6.07
fpt-f32	6.2	1.96
fpt-f16	4.25	1.07

Figure 3: (Left) Hyperbolic error for WordNet Nouns; (Right) Compression statistics for WordNet under the same MSHE, first block contains the size of original poincare embedding, second block contains the size of compressed baseline models, third block contains the size of matrix part in the L -tiling model (size of compressed integers using VLQ is also reported), the last block contains size of float points (fpt, f32 or f16) in the fundamental domain of L -tiling model.

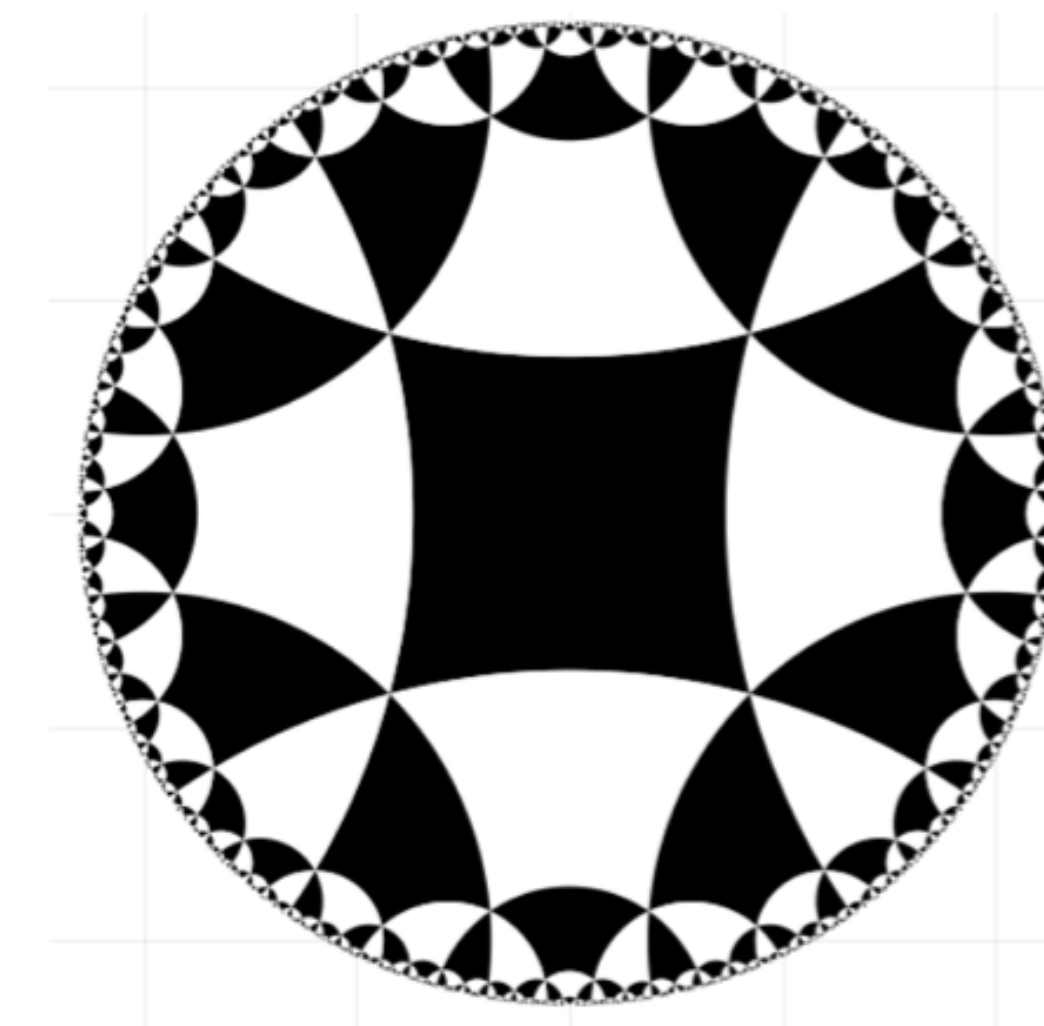
Гиперболические вложения замощениями!

Регулярное замощение плоскости Лобачевского

кодируется символом Шлефли $\{4,4\} \rightarrow$

другие: $\{3, 6\}$ – 6 треугольников смежны, $\{6,3\}$ – 3 6-угольника

Плоскость можно замостить любым $\{n,k\}$, если $1/n+1/k = 1/2$



В размерностях выше задача, казалось бы, проста:

у группы изометрий пр-ва $PSL_n(\mathbb{R})$ должна быть ко-компактная дискр. подгруппа (это называется Фуксовы группы)

К сожалению, классический результат* Кокстера гласит, что в $\dim \geq 6$ **такого нет**

* Coxeter H. S. M. Regular honeycombs in hyperbolic space //Proceedings of the International Congress of Mathematicians. – Amsterdam : North-Holland Publishing, 1954. – Т. 3. – С. 155-169.

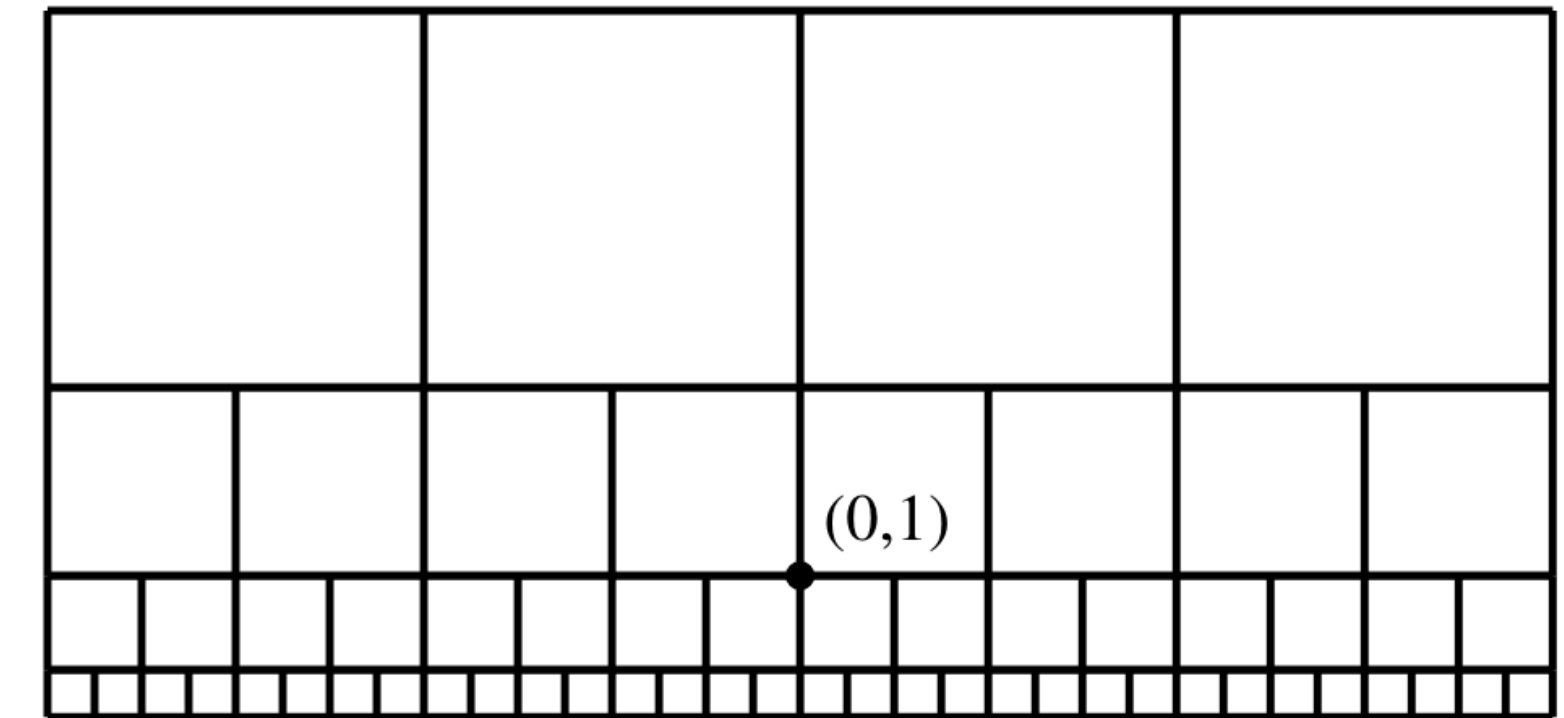
Гиперболические вложения замощениями!

Возможные решения (для увеличения размерности)

1) Рассматривать декартовы степени замощаемых

в статье* предлагается замощать $(\mathbb{H}^2)^n \rightarrow$

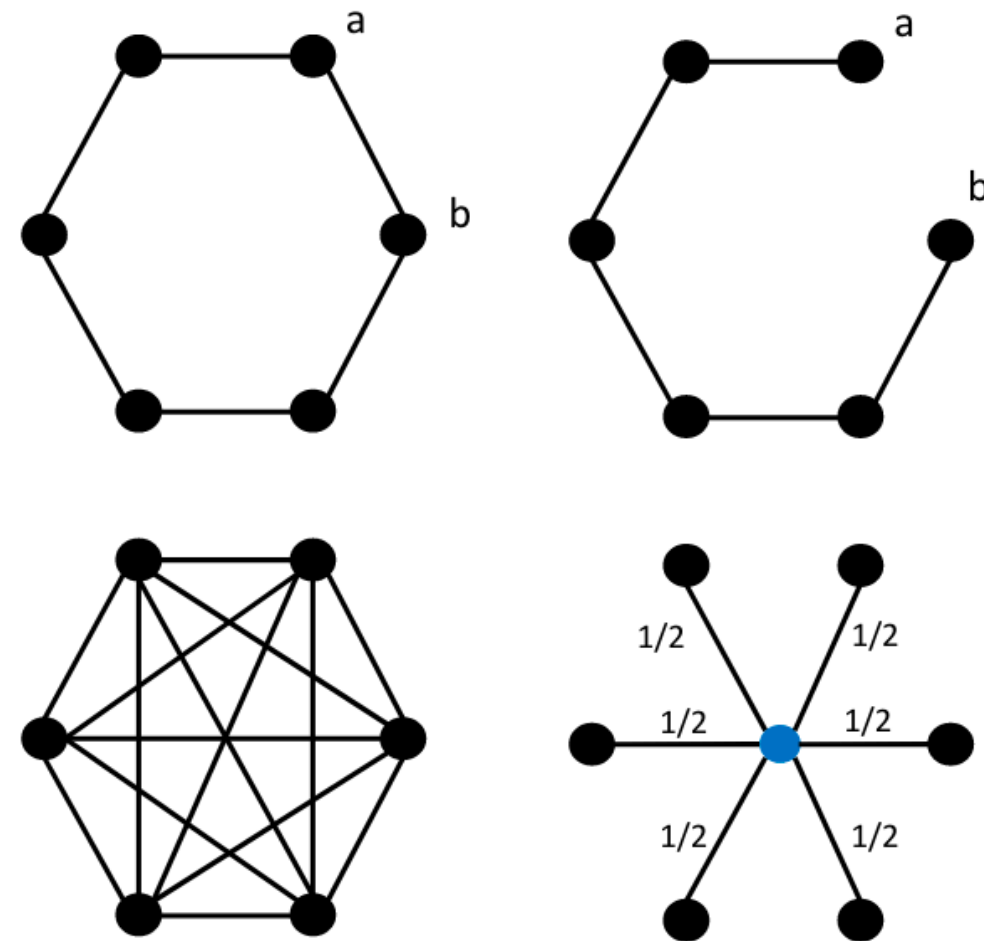
2) Строить замощения из изометрий, не образующих группу(?)



Гиперболические вложения замощениями!

Есть т.н. алгоритм Sarkar-а
вложения деревьев в \mathbb{H}_n

его идея: вершину-родителя
отправляем в центр, детей
ставим по кругу, потом
транспортируем всех обратно,
сшиваем



Algorithm 1 Sarkar's Construction

- 1: **Input:** Node a with parent b , children to place $c_1, c_2, \dots, c_{\deg(a)-1}$, partial embedding f containing an embedding for a and b , scaling factor τ
 - 2: $(0, z) \leftarrow \text{reflect}_{f(a) \rightarrow 0}(f(a), f(b))$
 - 3: $\theta \leftarrow \arg(z)$ {angle of z from x-axis in the plane}
 - 4: **for** $i \in \{1, \dots, \deg(a) - 1\}$ **do**
 - 5: $y_i \leftarrow \left(\frac{e^\tau - 1}{e^\tau + 1} \cdot \cos\left(\theta + \frac{2\pi i}{\deg(a)}\right), \frac{e^\tau - 1}{e^\tau + 1} \cdot \sin\left(\theta + \frac{2\pi i}{\deg(a)}\right) \right)$
 - 6: **end for**
 - 7: $(f(a), f(b), f(c_1), \dots, f(c_{\deg(a)-1})) \leftarrow \text{reflect}_{0 \rightarrow f(a)}(0, z, y_1, \dots, y_{\deg(a)-1})$
 - 8: **Output:** Embedded \mathbb{H}_2 vectors $f(c_1), f(c_2), \dots, f(c_{\deg(a)-1})$
-

несмотря на отсутствие глобальной симметрии у такого вложения, это довольно эффективный метод

в применении его к реальным данным бывают некоторые затруднения:

<- например, если в данных есть циклы, их приходится убирать, вводя новые вершины (дерево Штейнера)