Machine Learning Written Assignment 03

313652008 黃睿帆

September 21, 2025

Question 01

Focus on and explain the statements and ideas behind of Lemma 3.1 and Lemma 3.2 on the following paper **On the approximation of functions by tanh neural networks**.

Chapter 3: Uniform approximation of polynomials

本章節首先爲雙曲正切函數（tanh）神經網路的逼近誤差推導出界限。所以文章針對多項式的均勻逼近誤差，且在 Sobolev 範數下給出界限。固定大小的淺層神經網路，可以在 supremum norm（最大範數）下以任意精度逼近單項式（monomials），(By Pinkus,1999)。後來 Gühring 和 Raslan (2021) 又將這個結果推廣到 Sobolev 範數。文章繼續提出了一個新的推廣，能對「某一最大次數以下的所有多項式」得到顯式的誤差估計。對於有效率地逼近解析函數（analytic functions）非常關鍵。

Univariate Polynomials

用 $tanh$ 神經網路逼近任意次數的單變數多項式, 我們引入 p 階中心有限差分算子: The p-th order central finite difference operator δ_h^p for any $f \in C^{p+2}([a, b])$ and $p \in \mathbb{N}$, defined as

$$\delta_h^p[f](x) = \sum_{i=0}^{p} (-1)^i \binom{p}{i} f\left(x + \left(\tfrac{p}{2} - i\right) h\right),$$

where $\binom{p}{i}$ denotes the binomial coefficient (二項式係數). Next, for any $p \in \mathbb{N}$, $q \in 2\mathbb{N} - 1$ (odd), and $M > 0$, we define the following: The monomial

$$f_p : [-M, M] \to \mathbb{R}, \quad f_p(y) = y^p,$$

and the neural network function

$$\hat{f}_{(q,h)}(y) = \frac{\delta_{hy}^q[\sigma](0)}{\sigma^{(q)}(0)\, h^q}.$$

Example: the case $p = 1$

Before the lemma: First we introduce two notations that it will appear several times in the report: (以下介紹 L^p 空間，Sobolev 空間的定義)

1. $L^p(\Omega) = \{f \mid \int_\Omega |f|^p dx < \infty\}$
2. $L^\infty(\Omega) = \{f \mid sup_\Omega |f| < \infty\}$
3. Sobolev space is a functional space, denoting by $W^{k,p}(\Omega) = \{f \in L^p(\Omega) \mid D^\alpha f \in L^p(\Omega), \forall \alpha \in N_0^d, |\alpha| \leq k\}$ (即微分 α 次也在 L^p 空間)
4. $W^{k,\infty}(\Omega) = \{f \in L^\infty(\Omega) \mid D^\alpha f \in L^\infty(\Omega), \forall \alpha \in N_0^d, |\alpha| \leq k\}$
5. **Big O notation,** we say $f(\varepsilon) = O(\varepsilon)$ for small ε if $\lim_{\varepsilon \to 0} \frac{f(\varepsilon)}{\varepsilon} = c$ (即速度相同, 相除極限趨近於某個 constant.)

Lemma 3.1

Let $k \in \mathbb{N}_0$ and $s \in 2\mathbb{N}-1$. Then, for every $\varepsilon > 0$, there exists a shallow tanh neural network

$$\Psi_{s,\varepsilon} : [-M, M] \to \mathbb{R}^{(s+1)/2}$$

of width $(s+1)/2$ such that

$$\max_{\substack{p \le s \\ p \text{ odd}}} \|f_p - (\Psi_{s,\varepsilon})_{(p+1)/2}\|_{W^{k,\infty}} \le \varepsilon.$$

In other words, this neural network simultaneously approximates all odd-degree monomials up to degree s in the $W^{k,\infty}$ Sobolev norm with accuracy ε.

Moreover, the weights of $\Psi_{s,\varepsilon}$ scale as

$$O\Big(\varepsilon^{-s/2}\,[\,2(s+2)(2M)^{1/2}\,]^{\,s(s+3)}\Big),$$

for small ε and large s.

Lemma 3.1 的目標是建立一個具體的淺層 tanh 神經網路，這個網路寬度大約是 $(s+1)/2$。它能夠在 Sobolev 範數下，也就是說它不是只逼近「一個」單項式，而是同時逼近所有次數 $\le s$ 的奇數次單項式 $f_p(y) = y^p$。同時誤差量度使用的是 $W^{k,\infty}$ norm（即函數及其導數到 k 階爲止的最大誤差）。此 Lemma 保證：對所有奇數次 $p \le s$，都可以把誤差壓到小於 ε。因此可以以任意精度 ε 逼近所有小於等於某個奇數次 s 的單項式，並且給出權重大小隨 s, ε 的成長率。(給出權重大小的量級估計, 且不只是函數值的誤差（sup-norm），而是包括導數的誤差都被控制。)

Proof. Let $p \le s$ be odd and let $0 < h < \frac{2}{pM}$. Fix $0 \le m \le \min\{k, p+1\}$.

(條件: 取一個奇數次多項式次數 $p \le s$, 步長 $0 < h < \frac{2}{pM}$, 並令導數階數 $m = min\{k, p+1\}$).

利用有限差分運算子 δ_h^p 和泰勒定理 (Taylor's theorem) 可得到: For each x there exists $\xi_{x,i}$ such that

$$\frac{d^m}{dx^m}\delta_h^p = \sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^m h^m\,\sigma^{(m)}\big((\tfrac{p}{2}-i)hx\big).$$

Expanding $\sigma^{(m)}$ in Taylor series at 0 gives

$$\begin{aligned}\frac{d^m}{dx^m}\delta_h^p &= \sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^m h^m\left[\sum_{l=m}^{p+1}\frac{\sigma^{(l)}(0)}{(l-m)!}\Big(\tfrac{p}{2}-i\Big)^{l-m}(hx)^{l-m}\right]\\ &\quad+\sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^m h^m\left[\frac{\sigma^{(p+2)}(\xi_{x,i})}{(p+2-m)!}\Big(\tfrac{p}{2}-i\Big)^{p+2-m}(hx)^{p+2-m}\right].\end{aligned}$$

(上述式子分兩項, 第一部分是「有限展開到 $p+1$ 階」的泰勒級數。第二部分是誤差項，取在某個中間點 $\xi_{x,i}$)

From Katsuura in 2009, we know that

$$\sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^l = \begin{cases} p!, & l = p, \\ 0, & l \ne p, \quad l = 0,\dots,p. \end{cases} \tag{1}$$

Moreover, the above equality remains valid also for $l = p+1$, since all summands change sign when i is replaced by $p-i$.

((1) 式的意思是, 有限差分結構會「過濾掉」除了 $l = p$ 的所有項，保留單一最高次項。而且 (1) 對於 $l = p+1$ 仍然成立)

Using this, we can rewrite the first term above as

$$\begin{aligned}&\sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^m h^m\left[\sum_{l=m}^{p+1}\frac{\sigma^{(l)}(0)}{(l-m)!}\Big(\tfrac{p}{2}-i\Big)^{l-m}(hx)^{l-m}\right]\\ &= h^m\left[\sum_{l=m}^{p+1}\frac{\sigma^{(l)}(0)}{(l-m)!}(hx)^{l-m}\right]\sum_{i=0}^{p}(-1)^i\binom{p}{i}\Big(\tfrac{p}{2}-i\Big)^l.\end{aligned}$$

By the above result, only the term $l = p$ survives, so for $0 \le m \le p$ we obtain

$$h^m\frac{\sigma^{(p)}(0)}{(p-m)!}(hx)^{p-m}\,p!, \text{and for } m = p+1, \text{ the expression vanishes.}$$

由於 (1) 式的結果，除了 $l = p$ 以外的項都會消失，因此，這代表著有限差分能精確抓出第 p 階導數項，這就是爲什麼它能近似多項式。同時，上式等價於下面式子：

$$\frac{d^m}{dx^m}\delta_h^p = h^p\,\sigma^{(p)}(0)\,f_p^{(m)}(x).$$

整理來自泰勒展開式的剩餘項後，可以得到近似誤差

$$\hat{f}_{p,h}^{(m)}(x) - f_p^{(m)}(x) = \sum_{i=0}^{p}(-1)^i\binom{p}{i}\frac{1}{(p+2-m)!}\frac{\sigma^{(p+2)}(\xi_{x,i})}{\sigma^{(p)}(0)}\left(\frac{p}{2}-i\right)^{p+2}h^2x^{p+2-m}.$$

可以注意到誤差中的 h^2，所以有限差分對於 monomial 的近似是二階精確的。

文章中的 Lemmas A.1 and A.4 給出了 $\sigma = \tanh$ 的各階導數的上下界估計, 將其帶入上式，可得：For $m \le \min\{k, p+1\}$:

$$\|f_p - \hat{f}_{p,h}\|_{W^{m,\infty}} \le \sum_{i=0}^{p}\binom{p}{i}\frac{|\sigma^{(p+2)}(\xi_{x,i})|}{|\sigma^{(p)}(0)|}\left|\frac{p}{2}-i\right|^{p+2}h^2M^{p+2}.$$

Estimating the binomial sums gives

$$\|f_p - \hat{f}_{p,h}\|_{W^{m,\infty}} \le 2^p(2(p+2))^{p+3}\left(\frac{p}{2}\right)^{p+2}h^2M^{p+2} \le (2(p+2)pM)^{p+3}h^2.$$

Hence, if $k \le p+1$,

$$\|f_p - \hat{f}_{p,h}\|_{W^{k,\infty}} \le (2(p+2)pM)^{p+3}h^2.$$

Now suppose $k > p+1$, and consider $m \ge p+2$. In this case $f_p^{(m)} = 0$, so it suffices to bound $\hat{f}_{p,h}^{(m)}$. For $0 < h < 1$,

$$\begin{aligned}|\hat{f}_{p,h}^{(m)}(x)| &= \left|\frac{1}{h^p\sigma^{(p)}(0)}\sum_{i=0}^{p}(-1)^i\binom{p}{i}\left(\frac{p}{2}-i\right)^m h^m\sigma^{(m)}\big((\tfrac{p}{2}-i)hx\big)\right| \\ &\le 2\sum_{i=0}^{p}\binom{p}{i}\left|\frac{p}{2}-i\right|^m h^2(2m)^{m+1}.\end{aligned}$$

(如果 $k \le p+1$，直接得到這個估計; 但如果 $k > p+1$, 因爲 $f_p^{(m)} = 0$，我們只需界定近似函數的高階導數即可，最後仍可以得到誤差 $\sim h^2$)

Bounding the sum gives

$$|\hat{f}_{p,h}^{(m)}(x)| \le 2^{p+1}\left(\frac{p}{2}\right)^k(2k)^{k+1}h^2 \le (2pk)^{k+1}h^2.$$

We thus obtain, for arbitrary $k \in \mathbb{N}$,

$$\|f_p - \hat{f}_{p,h}\|_{W^{k,\infty}} \le \Big((2(p+2)pM)^{p+3} + (2pk)^{k+1}\Big)h^2 = \varepsilon.$$

(即對於任意 Sobolev 階數 k，眞實 monomial $f_p(x) = x^p$ 與近似函數 $\hat{f}_{p,h}$ 的差距，可以用一個與 h^2 成正比的誤差上界來控制，而常數則由 p, k, M 決定。

Furthermore, observe that the weights scale as

$$O\left(\max_i\binom{p}{i}h^{-p}\right).$$

For $\varepsilon \to 0$ and large p, it holds that

$$O(h^{-p}) = O\Big(\varepsilon^{-p/2}\left((p+2)(2M)^{1/2}\right)^p(p+3)\Big),$$

where the implied constant depends on k.

(權重的數値大小與二項式係數的最大值、以及 h^{-p} 成正比。即當我們要更精準近似 $\varepsilon = 0$ 時，權重會急遽放大，並且放大的速率大約是 $\varepsilon^{-p/2}$，還要乘上一個依賴於 p, M 的因子。)

Next, using Stirling’s approximation, we find that for $0 \le i \le p$ it holds that

$$\binom{p}{i} \le \binom{p}{\lfloor (p-1)/2 \rfloor} \le \frac{e\, p^{(p+1)/2}}{2\pi\, ((p-1)/2)^{p/2}\, ((p+1)/2)^{p/2+1}} = O\Big(\frac{2^p}{\sqrt{p}}\Big).$$

這裡使用 Stirling's 近似來估計二項式係數最大值，可知最大的二項式係數大約和 $\frac{2^p}{\sqrt{p}}$ 同階，因此，權重的最終尺度是：

$$O\Big(\varepsilon^{-p/2}\, \big(2(p+2)(2M)^{1/2}\big)^p (p+3)\Big).$$

Regarding the network architecture, note that the neurons needed for all $\hat{f}_{p,h}$ are already available in the network $\hat{f}_{s,h}$. This allows us to define the shallow tanh neural network $\Psi_{s,\varepsilon}$ by

$$(\Psi_{s,\varepsilon})_p = \hat{f}_{p,h},$$

such that it only has $(s+1)/2$ neurons in its hidden layer. The width follows directly from its definition and from the fact that σ is an odd function.

(要構造 $\hat{f}_{p,h}$ 所需的神經元，已經包含在 $\hat{f}_{s,h}$ 的網路裡。要同時近似所有次數 $\le s$ 的單項式，其實只需要 $\frac{s+1}{2}$ 個神經元。) □

We now extend the previous result to monomials with even degree. To this end, we rely on the observation that for $n \in \mathbb{N}$ and $\alpha > 0$, it holds that

$$y^{2n} = \frac{1}{2\alpha(2n+1)}\Bigg((y+\alpha)^{2n+1} - (y-\alpha)^{2n+1} - 2\sum_{k=0}^{n-1}\binom{2n+1}{2k}\alpha^{2(n-k)+1}y^{2k}\Bigg). \tag{2}$$

This formula allows us to construct recursively defined tanh neural network approximations of even powers of y. The following lemma quantifies the uniform approximation accuracy of these networks in the Sobolev norm.

(接著我們處理偶數次 monomial(形如 y^{2n})，其可以用「相鄰兩個奇次冪」以及更低階偶次冪的線性組合 (即 (2) 式) 來表示。這也提供了一個可行的遞迴構造方式：先近似奇次 monomial，再利用這個公式去近似偶次 monomial。)

Lemma 3.2

Let $k \in \mathbb{N}_0$, $s \in 2\mathbb{N}-1$ and $M > 0$. For every $\varepsilon > 0$, there exists a shallow tanh neural network

$$\psi_{s,\varepsilon} : [-M, M] \to \mathbb{R}^s$$

of width $\frac{3}{2}(s+1)$ such that

$$\max_{p \le s} \big\| f_p - (\psi_{s,\varepsilon})_p \big\|_{W^{k,\infty}} \le \varepsilon.$$

Furthermore, the weights scale as

$$O\Bigg(\varepsilon^{-s/2}\, \big(\sqrt{M}(s+2)\big)^{\frac{3s(s+3)}{2}}\Bigg)$$

for small ε and large s.

(給定 $k \in \mathbb{N}_0$（非負整數）、$s \in 2\mathbb{N}-1$（奇數）、以及 $M > 0$。對於每個 $\varepsilon > 0$，都存在一個淺層 tanh 神經網路，隱藏層寬度爲 $\frac{3}{2}(s+1)$，並且這個神經網路可以同時近似所有次方單項式 $f_p(y) = y^p$ ($p \le s$)，而且誤差在 Sobolev 範數下小於 ε。另外，這個網路的權重大小隨著上面給的參數縮放。

因此，Lemma 3.2 是 Lemma 3.1 的推廣，從「奇數次冪」擴展到「偶數次冪」，最後保證整個多項式空間（最高次數 s）都能用一個淺層 tanh 網路統一逼近。)

Proof. For $h, M > 0$ and $p \le s$, we define $\hat{f}_{p,h} : [-M-1, M+1] \to \mathbb{R}$ by

$$f_p(y) = y_p, \qquad \hat{f}_{q,h}(y) = \frac{\delta_h^q y}{\sigma^{(q)}(0)\, h^q}.$$

For $\varepsilon > 0$, choose $h > 0$ small enough depending on ε, let $\alpha \le 1$, and let $y \in [-M, M]$. We then define

$$(\psi_{s,\varepsilon}(y))_p = \hat{f}_{p,h}(y), \quad \text{for } p \text{ odd},$$

and for $p = 2n$ even, we set $(\psi_{s,\varepsilon}(y))_{2n}$ recursively by

$$(\psi_{s,\varepsilon}(y))_0 = 1,$$

and

$$(\psi_{s,\varepsilon}(y))_{2n} = \frac{1}{2\alpha(2n+1)} \left(\hat{f}_{2n+1,h}(y+\alpha) - \hat{f}_{2n+1,h}(y-\alpha) - 2\sum_{k=0}^{n-1} \binom{2n+1}{2k} \alpha^{2(n-k)+1} (\psi_{s,\varepsilon}(y))_{2k} \right). \tag{3}$$

(對奇數次單項式: 我們直接用 $\hat{f}_{p,h}$ 就能逼近，Lemma 3.1 已經保證誤差 $\leq \varepsilon$。對於偶數次單項式, 我們使用上面的遞迴公式 (3) 來構造：將偶數次冪 y^{2n} 展開爲「高次冪差」減去「低次冪的線性組合」。遞迴的 idea 是：如果低次偶數次方已經能被近似，則更高次的偶數次方也能近似。)

We introduce the notation

$$E_p = \left\| f_p - (\psi_{s,\varepsilon})_p \right\|_{W^{k,\infty}}.$$

We claim that for all $\varepsilon > 0$, there exists $h > 0$ such that for all $p \leq s$,

$$E_p \leq E_p^* = \frac{2^{p/2}(1+\alpha)^{(p^2+p)/2}}{\alpha^{p/2}} \varepsilon. \tag{4}$$

Step 1: Odd p. Choosing h as in Lemma 3.1 implies

$$\max_{\substack{p \leq s \\ p \text{ odd}}} E_p \leq \varepsilon, \tag{5}$$

which proves (4) for odd p, since $(1+\alpha)/\alpha \geq 1$.

Step 2: Even p, base case. For $p = 2$, we find

$$E_2 \leq \frac{2\varepsilon}{6\alpha} \leq E_2^*,$$

which establishes the base step.

Step 3: Even p, induction step. Let $n \in \mathbb{N}$ with $2n+1 \leq s$ and $n > 1$. Assume by the induction hypothesis that

$$E_{2k} \leq E_{2k}^*, \quad \forall\, k < n.$$

From (3) it follows that

$$E_{2n} \leq \frac{1}{2\alpha(2n+1)} \left(E_{2n+1} + E_{2n+1} + 2\sum_{k=1}^{n-1} \binom{2n+1}{2k} \alpha^{2(n-k)+1} E_{2k} \right). \tag{6}$$

By the induction hypothesis and the monotonicity of E_{2k}^* in k, we have

$$E_{2k} \leq E_{2k}^* \leq E_{2(n-1)}^*.$$

Using also (5) and the fact that $\varepsilon \leq E_{2(n-1)}^*$, we can estimate (6) as

$$E_{2n} \leq \frac{1}{\alpha(2n+1)} \left(\max_{\substack{p \leq s \\ p \text{ odd}}} E_p + \sum_{k=1}^{n-1} \binom{2n+1}{2k} \alpha^{2(n-k)+1} E_{2(n-1)}^* \right).$$

This yields

$$E_{2n} \leq \frac{1}{\alpha}\left(E_{2(n-1)}^* + (1+\alpha)^{2n+1} E_{2(n-1)}^* \right) \leq \frac{2}{\alpha}(1+\alpha)^{2n+1} E_{2(n-1)}^*.$$

Recalling the definition of $E_{2(n-1)}^*$, we obtain

$$E_{2n} \leq \left(\tfrac{2}{\alpha}(1+\alpha)^{2n+1} \right)^n \varepsilon = E_{2n}^*.$$

(此處進行誤差分析 E_p, 對奇數情況：Lemma 3.1 直接給 $E_p \leq \varepsilon \leq E_p^*$. 而對偶數情況：使用數學歸納法。Base case（$p = 2$）：驗算成立。歸納步：假設所有 $2k < 2n$ 都成立，利用展開式與估計不等式，把

高次誤差 E_{2n} 控制在 E_{2n}^* 之內。綜合奇數與偶數情況，得到所有 $p \leq s$ 的誤差都被控制在 ε 以內，完成證明。)

Next, we optimize inequality (4) by choosing the optimal value of α. Lemma A.2 shows that the optimal choice is $\alpha = 1/s$. We therefore conclude that, for any $\varepsilon > 0$, there exists a shallow tanh neural network

$$\psi_{s,\varepsilon},$$

with width independent of ε, such that

$$\max_{p \leq s} \|f_p - (\psi_{s,\varepsilon})_p\|_{W^{k,\infty}} \leq \sqrt{e}\,(2es)^{s/2}\,\varepsilon.$$

(接下來對式子 (4) 進行最佳化，方法是選擇一個最佳的參數 α。根據 Lemma A.2 的結果，最佳的 α 值就是 $\alpha = 1/s$。這裡 α 是一個控制平移間距的參數，最佳化 α 的目的是讓近似誤差最小。)

(因此，我們得到結論：對任何誤差精度 $\varepsilon > 0$，都可以構造一個淺層的 $tanh$ 神經網路，其寬度不依賴於 ε，滿足上面的誤差上界：這表示：我們能找到一個固定寬度的淺層網路，使得對所有 $p \leq s$，它對多項式的近似誤差在 Sobolev norm $W^{k,\infty}$ 意義下，被一個常數倍的 ε 控制住。)

Replacing $\varepsilon \mapsto \dfrac{\varepsilon}{\sqrt{e}(2es)^{s/2}}$ recovers the claimed error bound in the statement of the lemma.
(此處調整 ε 的尺度，來對齊 Lemma 3.2 開頭所要求的誤差形式。)

To quantify the size of the weights, we observe from equation (27) that the weight bound from Lemma 3.1 needs to be multiplied by the factor

$$\max_k \binom{s}{2k}\left(\frac{1}{s}\right)^{s-2k} \leq s\sum_{j=0}^{s}\binom{s}{j}\left(\frac{1}{s}\right)^{s-j} \leq s\left(1+\frac{1}{s}\right)^{s} = O(s),$$

where we used the binomial theorem.

(估計網路的權重大小，我們利用式子(3)。這顯示在 Lemma 3.1 的權重上限之外，我們需要再乘上一個因子：利用二項式展開，可以證明這個最大值被上面式子控制：並利用二項式展開，可以證明這個最大值權重增長到約 $O(s)$ 的等級，不會比指數更糟。)

Therefore, the final weight bound can be seen to satisfy

$$O\left(\varepsilon^{-s/2}\, s\left(2(2es)^{1/4}\,[2(M+1)]^{1/2}\,(s+2)\right)^{s(s+3)}\right) = O\left(\varepsilon^{-s/2}\left(\sqrt{M}(s+2)\right)^{\frac{3s(s+3)}{2}}\right),$$

for small ε and large s. This proves the weight bound as stated in the lemma.(證明了 Lemma 3.2 中所聲明的權重上界。)

Finally, we note that the constructed approximations indeed correspond to a shallow tanh neural network of the claimed size. In fact, by observing that σ is an odd function, and using

$$\delta_h^p[f](x) = \sum_{i=0}^{p}(-1)^i\binom{p}{i}f\left(x+\left(\tfrac{p}{2}-i\right)h\right),$$

together with Lemma 3.1 and equation (3), we conclude that a shallow tanh neural network suffices. Specifically, the values of the $\frac{3}{2}(s+1)$ neurons in the hidden layer are given by

$$\sigma\left(\left(\tfrac{s}{2}-i\right)h(y+\beta)\right), \qquad i = 0, 1, \ldots, \tfrac{s-1}{2}, \quad \beta \in \{-\alpha,\, 0,\, \alpha\}.\square$$

(由上面可以確認我們成功構造出一個隱藏層大小受控的淺層 $tanh$ 網路來近似多項式，並且誤差與權重都有清楚的上界。)

Question 02–Unanswered questions

1. In the paper we discussed in the class **(On the approximation of functions by** tanh **neural networks)**, we drew the neural network. If a network has more layer and more neuron, will it be better, that is, has better accuracy and less loss?

References

[1] On the approximation of functions by tanh neural networks, Tim DeRyck, Samuel Lanthaler, Siddhartha Mishra, Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland.