Bounded Independence Fools Halfspaces

Ragesh Jaiswal
(Columbia University)

joint work with:
- Ilias Diakonikolas (Columbia University)
- Parikshit Gopalan (Microsoft Research)
- Rocco Servedio (Columbia University)
- Emanuele Viola (Northeastern University)

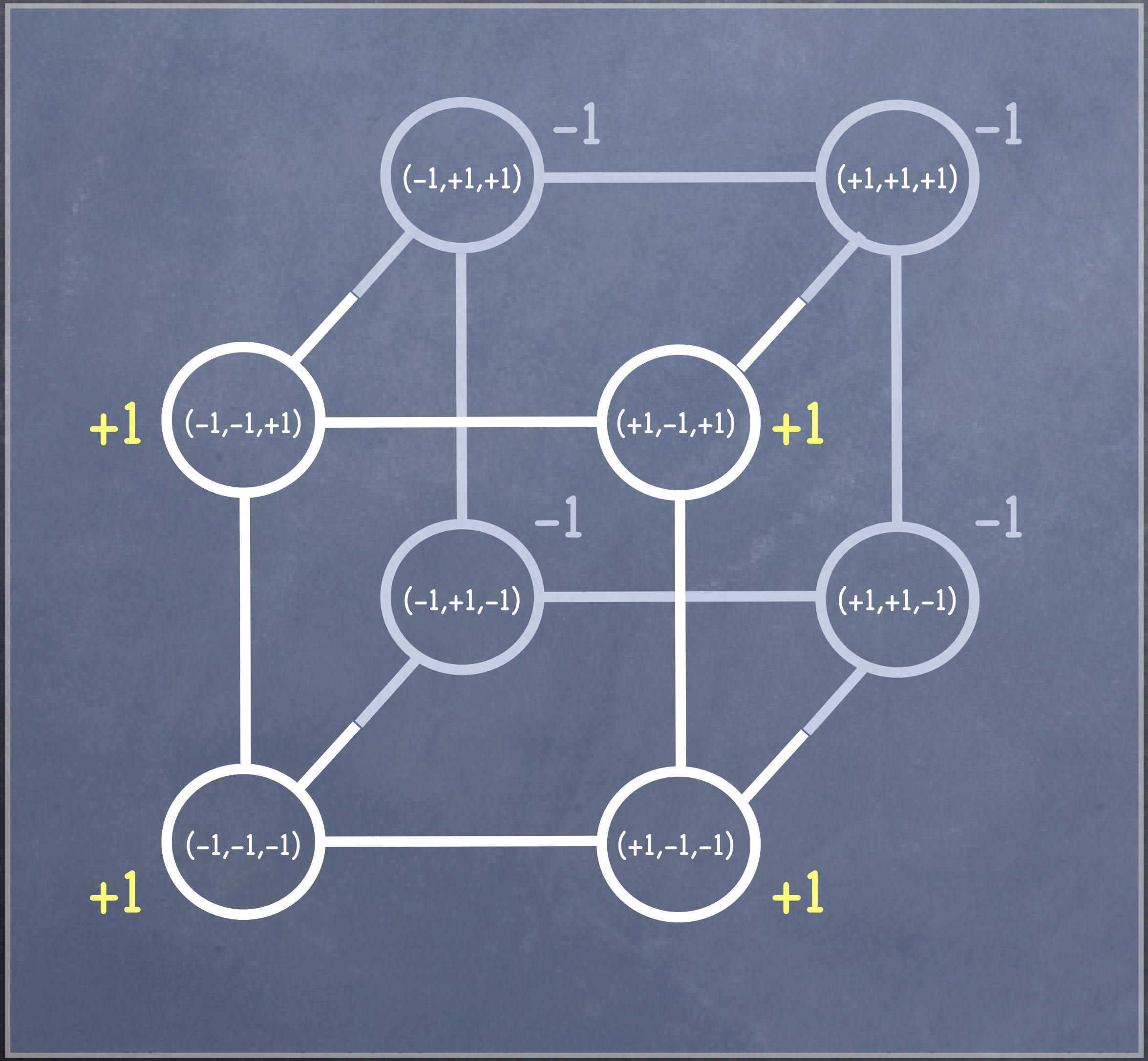
Motivation: Randomness in Computation

- "Does randomness helps in making computation efficient?"
- Universal derandomization:
  - Deterministic Turing machine for any randomized one.
  - Disadvantage: Conditioned on the existence of problems which are very hard to compute on average. Involves Direct Product Theorems.
- Consider restricted models of computation/ classes of functions.

Halfspaces

(also known as Linear Threshold Functions)

- A halfspace is a function $h:\{+1,-1\}^n \to \{+1,-1\}$
  - $h(x) = \text{sign}(w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots + w_n \cdot x_n - \Theta)$, where $w_1,\ldots,w_n,\Theta \in R$



- Studied in:
  - Machine Learning: winnow, perceptron
  - Complexity Theory: NP ⊂? halfspaces of halfspaces
  - Social Choice Theory: Weighted voting

Halfspaces: Examples

- MAJORITY(x) = sign($x_1 + x_2 + \ldots + x_n$)
  - $w_1=w_2=\ldots=w_n=1$, $\Theta=0$
- AND(x) = sign($x_1 + x_2 + \ldots + x_n - n + 1/2$)
  - $w_1=w_2=\ldots=w_n=1$, $\Theta = (n-1/2)$
- MAX-OR-BIT(x) := If $x_1=+1$ then $+1$; else if $x_2=-1$ then -1; else if $x_3=+1$ then $+1$; else...
  - MAX-OR-BIT(x) = sign($2^n x_1 - 2^{n-1} x_2 + 2^{n-2} x_3 - \ldots$)
  - Any representation with integer weights needs weights of exponential magnitude.

Bounded Independence

- A distribution D over $\{-1,+1\}^n$ is called k-wise independent if its projection on <u>any</u> k indices is uniform over $\{-1,+1\}^k$

-1,	-1,	-1,	+1
-1,	-1,	+1,	-1
-1,	+1,	-1,	+1
-1,	+1,	+1,	-1
+1,	-1,	-1,	-1
+1,	-1,	+1,	+1
+1,	+1,	-1,	-1
+1,	+1,	+1,	+1

- <u>Example</u>: This distribution over $\{-1,+1\}^4$ is 1,2-wise independent (but not 3,4-wise independent).

Fooling a Function

- A distribution D over $\{-1,+1\}^n$ is said to fool a function $h:\{-1,+1\}^n \to \{-1,+1\}$ with error ε if

$$|E_{x<-D}[h(x)] - E_{x<-U}[h(x)]| \leq \varepsilon,$$

where U is the uniform distribution over $\{-1,+1\}^n$.

- Example:

D (n is odd) perfectly fools MAJORITY= $\mathrm{sign}(x_1+\ldots+x_n)$

$-1,-1,-1,\ldots,-1$

$+1,+1,+1,\ldots,+1$

Bounded Independence Fools Halfspaces

- Main Theorem(This work): Any k-wise independent distribution fools any halfspace with error ε, provided $k \geq (C/\varepsilon^2)\cdot\log^2(1/\varepsilon)$, where $C>1$ is some fixed constant.

- Observations:
  - The result is interesting only when $\varepsilon \geq 1/\sqrt{n}$.
  - The result is tight up to $\log(1/\varepsilon)$ factors.
    - There exists a halfspace and a k-wise independent distribution D with $k<1/\varepsilon^2$ such that D does not fool this halfspace with error ε.

Interesting Implications

Pseudorandom Generators for Halfspaces

- A pseudorandom generator for halfspaces is an efficiently computable function $G:\{-1,+1\}^s \to \{-1,+1\}^n$ such that $s \ll n$ and for any halfspace h, $E_{x \leftarrow \{-1,+1\}^s}[h(G(x))] \approx E_{x \leftarrow \{-1,+1\}^n}[h(x)]$.
  - G fools h with error ε if the above expectations differ by at most ε.
- [Alon-Babai-Itai'86; Chor-Goldreich'85]: There is an explicit generator $G:\{-1,+1\}^{k \cdot \log(n)} \to \{-1,+1\}^n$ such that the output of G is a k-wise independent distribution.

Interesting Implications

Pseudorandom Generators for Halfspaces

- Corollary of our main theorem: There is a pseudorandom generator $G:\{-1,+1\}^s \to \{-1,+1\}^n$ such that G fools any halfspace with error ε and $s = O((1/\varepsilon^2) \cdot \log^2(1/\varepsilon) \cdot \log(n))$.

- Observations:
  - First such generator for general halfspaces.
  - [Nisan'92]: Implies a generator ($s=\log^2(n)$) for a subset of halfspaces where $\forall i$, $|w_i|=\mathrm{poly}(n)$.
  - [Bazzi'07, Razborov'08, Braverman'09]: Generators for constant depth circuits.

Interesting Implications

Derandomization of Berry-Esséen Central Limit Theorem

- Theorem[Berry-Esséen]: Let $Y = \Sigma w_i \cdot x_i$, where $x_i \in \{-1,+1\}$. If $\Sigma w_i^2 = 1$ and $\forall i, |w_i| \leq \varepsilon$, then

$$\forall t \in R,\ |\Pr_{x \leftarrow U}[Y \leq t] - \Phi(t)| \leq \varepsilon,$$

where U is the uniform distribution over $\{-1,+1\}^n$ and Φ is the cumulative distribution function of the standard normal $N(0,1)$.

- Theorem[Our result]: Let $Y = \Sigma w_i \cdot x_i$, where $x_i \in \{-1,+1\}$. If $\Sigma w_i^2 = 1$ and $\forall i, |w_i| \leq \varepsilon$, then

$$\forall t \in R,\ |\Pr_{x \leftarrow D}[Y \leq t] - \Phi(t)| \leq \varepsilon,$$

where D is any k-wise independent distribution over $\{-1,+1\}^n$ and Φ is the cumulative distribution function of the standard normal $N(0,1)$ provided $k \geq (C/\varepsilon^2) \cdot \log^2(1/\varepsilon)$.

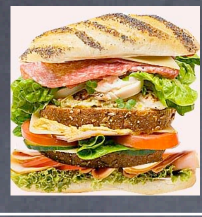
Proof Sketch

Proof Sketch

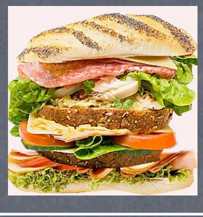
- Main Theorem: Any k-wise independent distribution fools any halfspace with error ε, provided $k \geq (C/\varepsilon^2)\log^2(1/\varepsilon)$, where $C>1$ is some fixed constant.

- For halfspace $h(x) := \mathrm{sign}(w_1x_1+\ldots+w_nx_n-\Theta)$
  - WLOG assume that $\sum w_i^2 = 1$.
  - For simplicity of discussion we assume $\Theta = 0$.
- Proof by case analysis:
  - case (a): $\forall i, |w_i| \leq \varepsilon$. Any h with this property is called ε-regular.
  - case (b): $\exists i, |w_i| > \varepsilon$.

Proof Sketch



Sandwiching Polynomials



- Fact: Any k-wise independent distribution fools a function $f:\{-1,+1\}^n \to \{-1,+1\}$ with error ε if and only if there are two multivariate polynomials q_u and q_l such that:

1. $\text{degree}(q_u), \text{degree}(q_l) \leq k$,
2. $\forall x \in \{-1,+1\}^n,\ q_l(x) \leq f(x) \leq q_u(x)$,
3. $E_{x \leftarrow U}[q_u(x) - f(x)] \leq \varepsilon$, and
   $E_{x \leftarrow U}[f(x) - q_l(x)] \leq \varepsilon$.

Proof Sketch

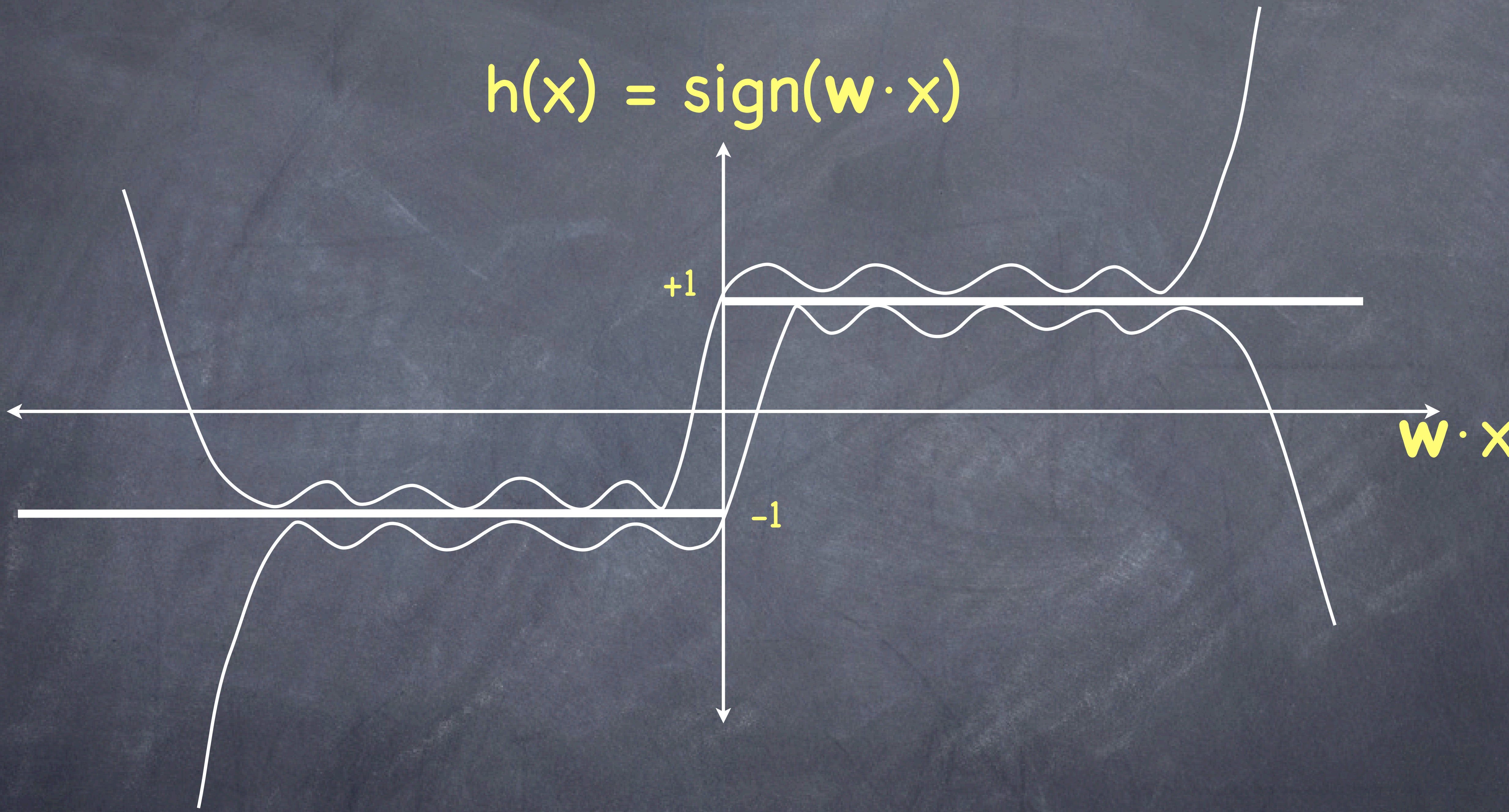
- Case(a): h is ε-regular ($\forall i$, $|w_i| \leq \varepsilon$)

Show that the sandwiching polynomials exist.

- Case (b): h is not ε-regular ($\exists i$, $|w_i| \geq \varepsilon$)
  - Reduce it to case (a).

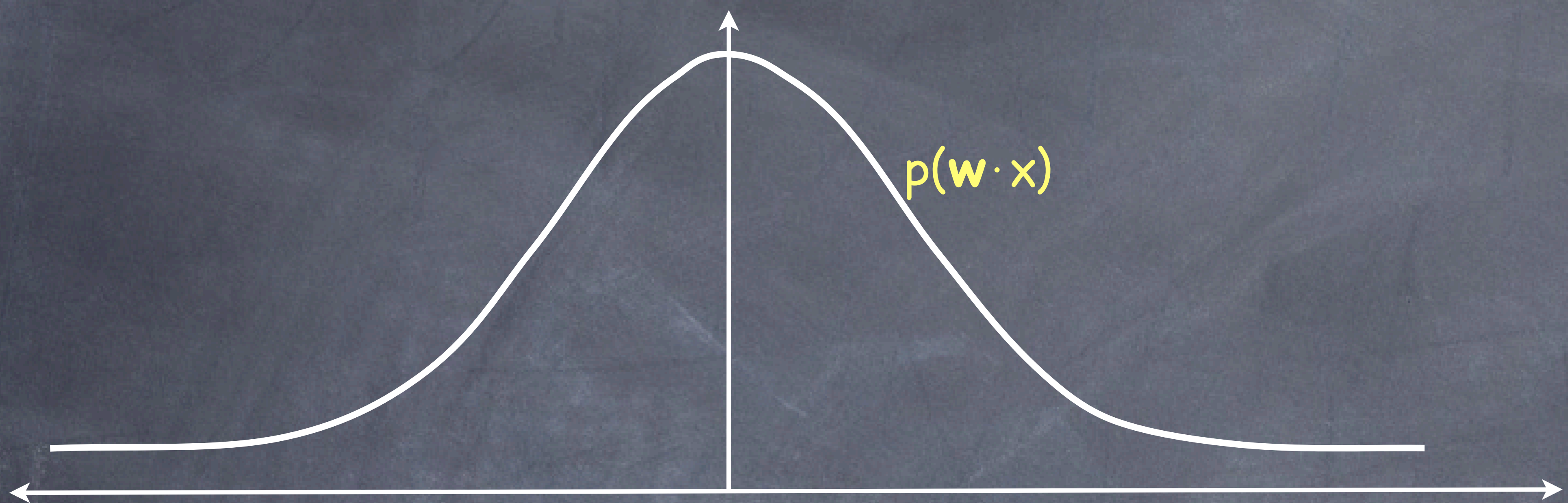
Proof Sketch: (Case (a): h is ε-regular)

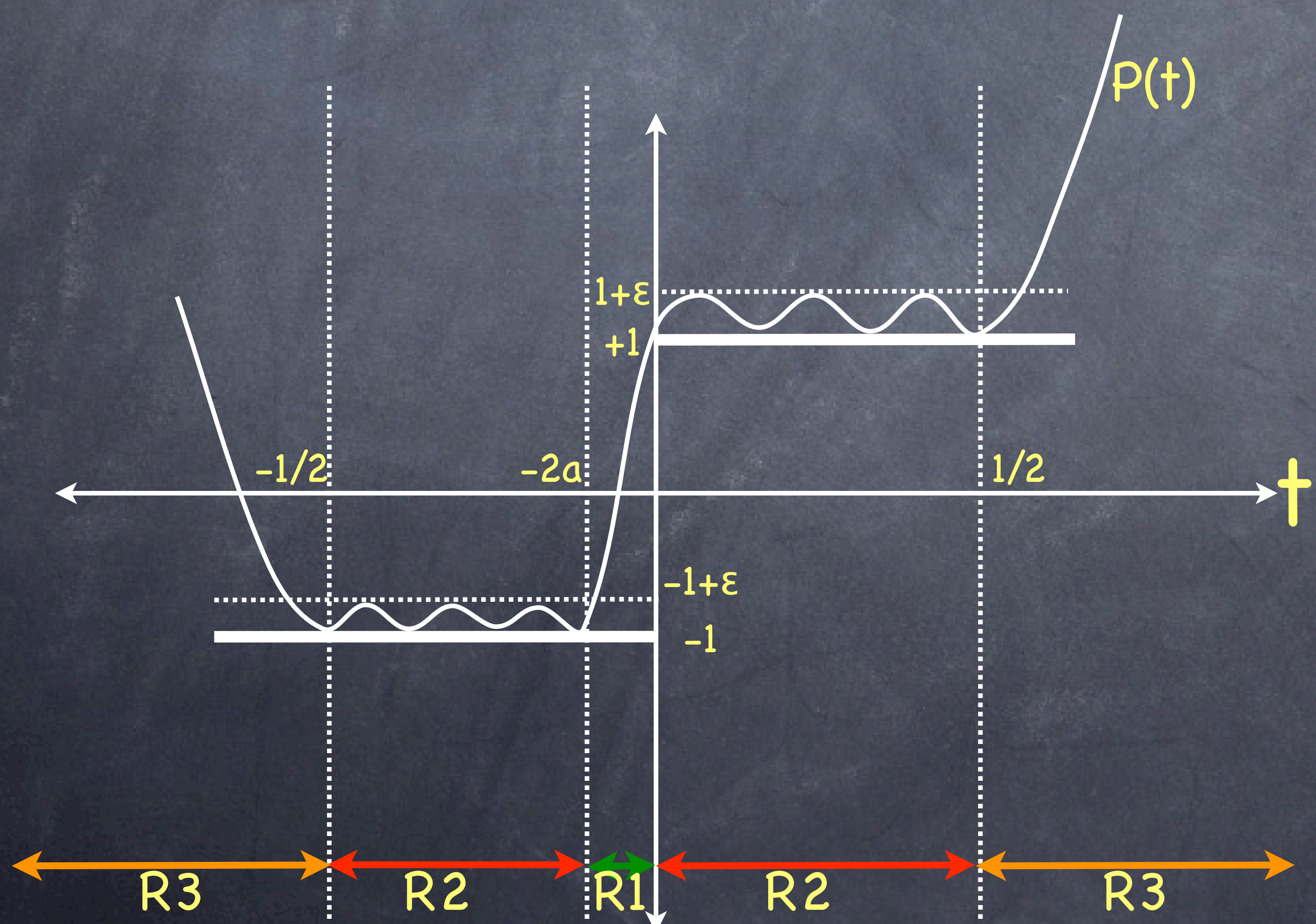
$h(x) = \text{sign}(\mathbf{w}\cdot x)$



- If there is a <u>univariate</u> polynomial P of bounded degree that approximates the sign function, then we can perhaps plug in $\mathbf{w}\cdot x$ into P to get our sandwiching polynomials.

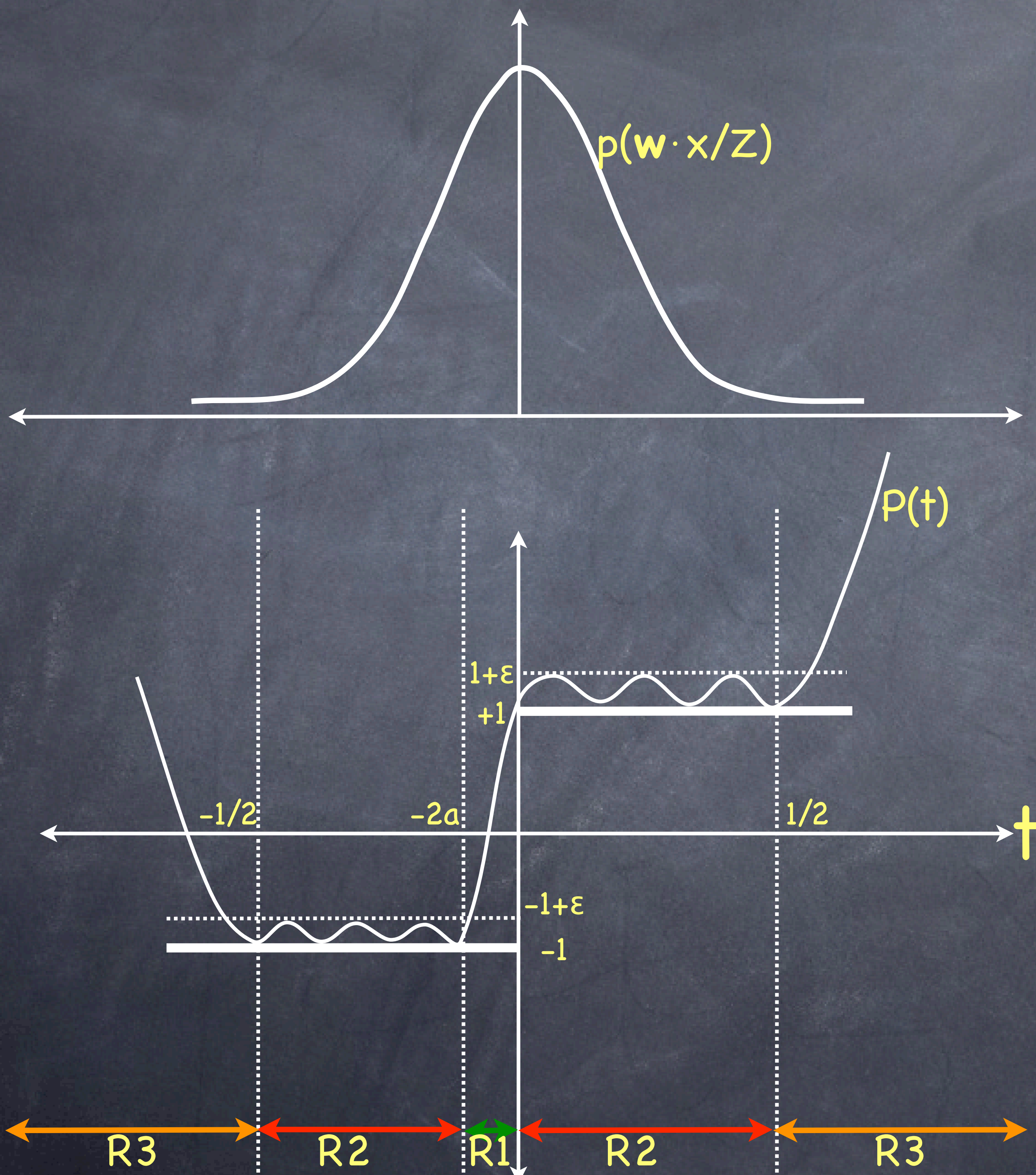
Proof Sketch: (Case (a): h is ε-regular)





- Properties of the poly P(t): (obtained using Jackson +Chebyshev+amplification)
  - degree(P) ~ $1/\varepsilon^2 \log^2(1/\varepsilon)$
  - t∈R1: P(t) ∈ [-1, 1+ε]
  - t∈R2: P(t)-sign(t) ≤ ε
  - t∈R3: P(t) does not grow too fast.

Proof Sketch: (Case (a): h is ε-regular)



- Let us try polynomials:
  - $q_u := P(\mathbf{w}\cdot x/Z)$
  - $q_l := -P(-\mathbf{w}\cdot x/Z)$
- Properties for sandwiching:

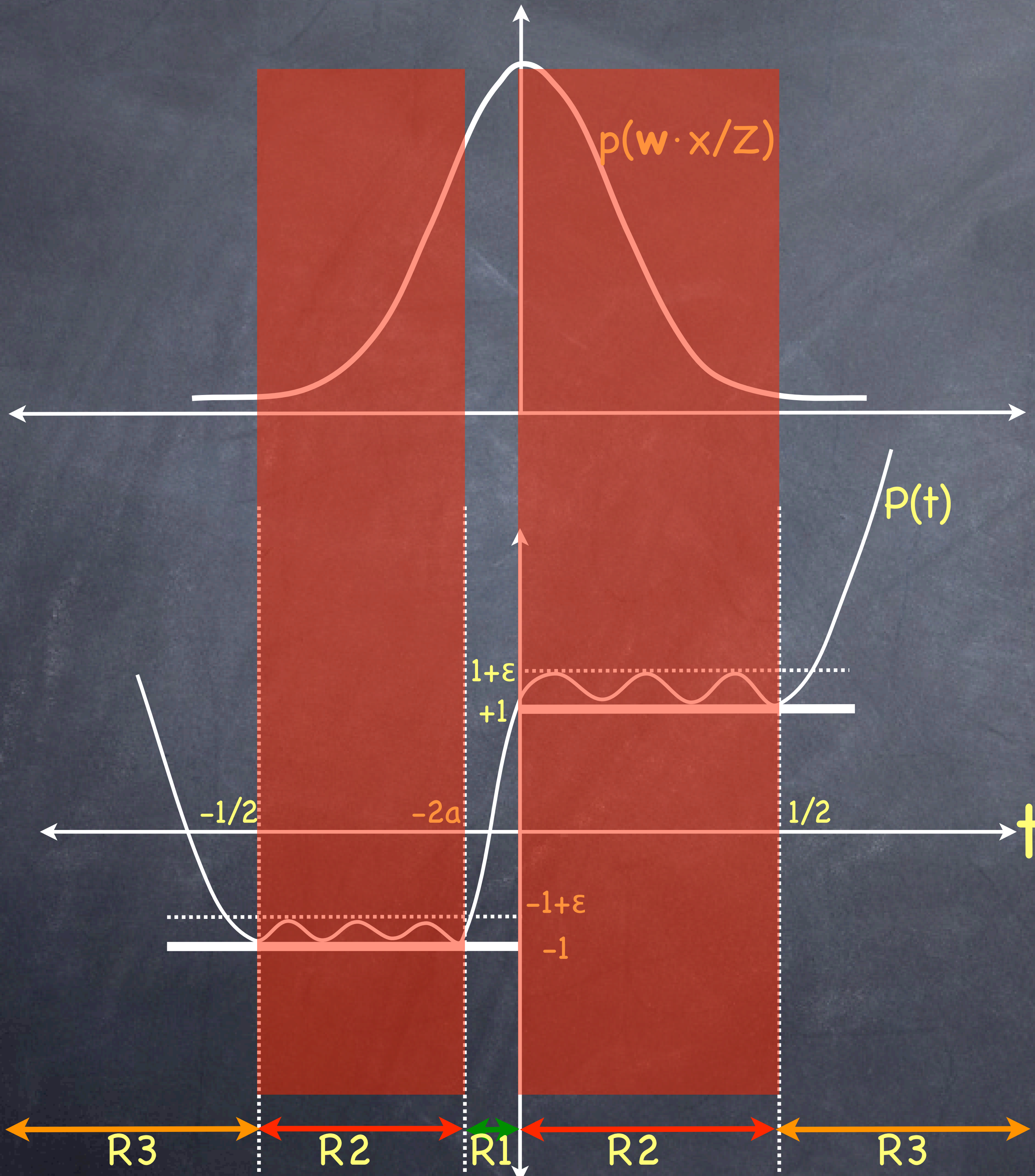
[✔] 1. $\deg(q_l), \deg(q_u) \sim 1/\varepsilon^2 \log^2(1/\varepsilon)$

[✔] 2. $\forall x, q_l(x) \le \mathrm{sign}(\mathbf{w}\cdot x) \le q_u(x)$

3. $E_x[q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)] \le \varepsilon$, $E_x[\mathrm{sign}(\mathbf{w}\cdot x) - q_l(x)] \le \varepsilon$

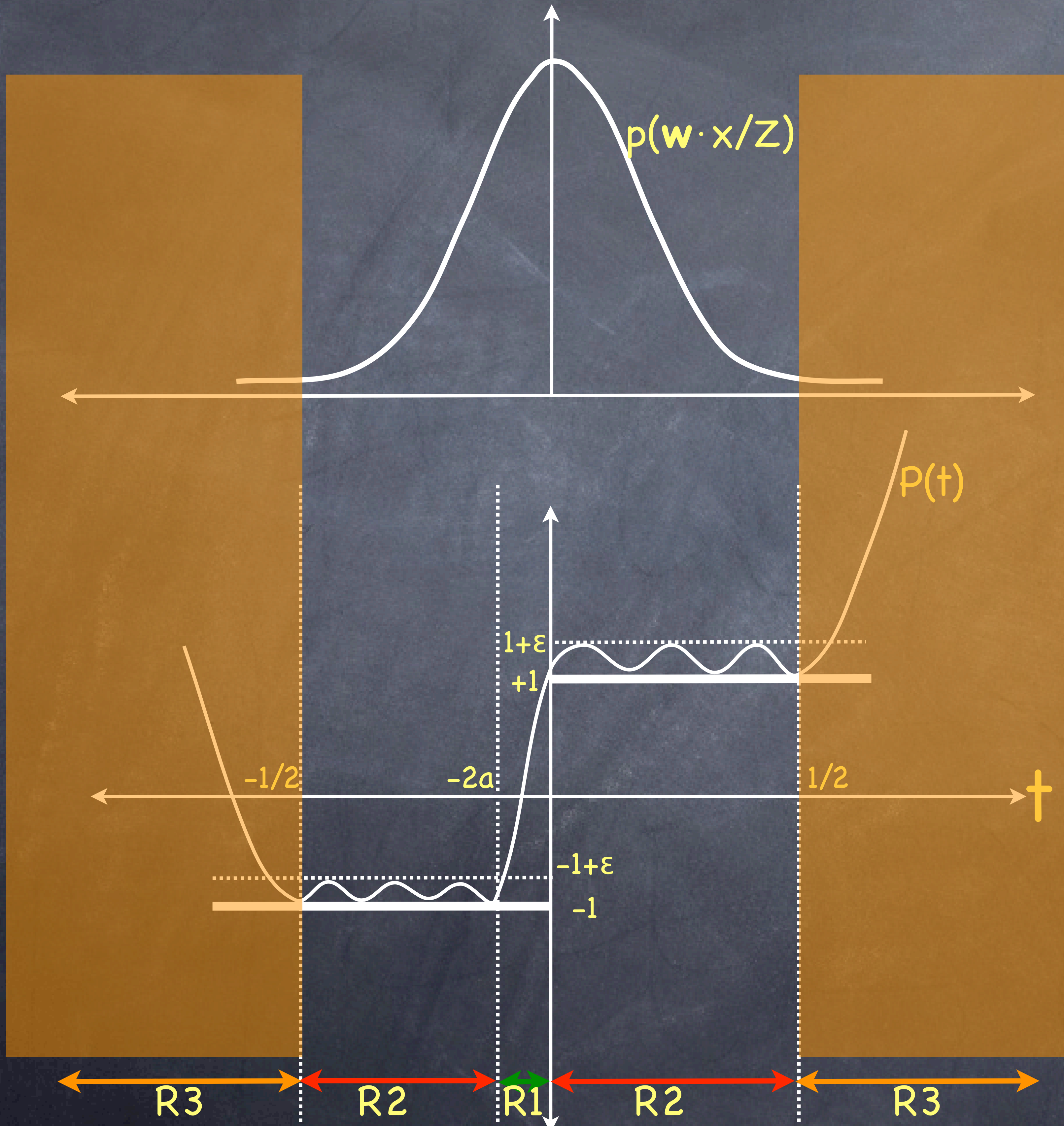
- We will show $E_x[q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)] \le \varepsilon$. $E_x[\mathrm{sign}(\mathbf{w}\cdot x) - q_l(x)] \le \varepsilon$ follows from symmetry.

Proof Sketch: (Case (a): h is ε-regular)



- Lemma: $E_x[q_u(x)-\mathrm{sign}(w\cdot x)] \leq \varepsilon$.
- Proof: case analysis based on the value of $y = w\cdot x/Z$
  - $y \in R2$:
    - $q_u(x)-\mathrm{sign}(w\cdot x) \leq \varepsilon$. So, the contribution to the expectation is $\leq \varepsilon$.

Proof Sketch: (Case (a): h is ε-regular)



- Lemma: $E_x[q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)] \leq \varepsilon$.
- Proof: case analysis based on the value of $y = \mathbf{w}\cdot x/Z$
  - $y \in R2$:
    – $q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x) \leq \varepsilon$.
    So, the contribution to the expectation is $\leq \varepsilon$.
  - $y \in R3$:
    $q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)$ grows as $|y|$ grows larger but $\Phi(y)$ diminishes (by Hoeffding). Hoeffding overshadows P's growth.

Proof Sketch

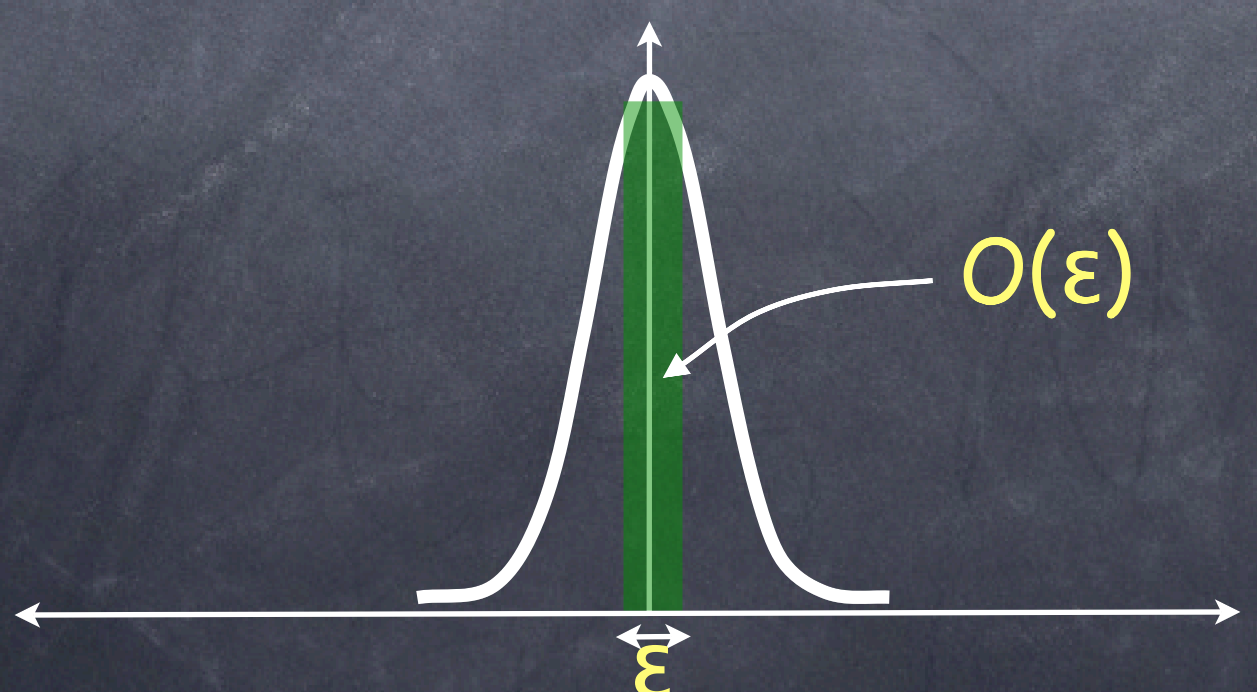
(Case (a): h is ε-regular)

- Theorem (follows from Berry-Esséen CLT):

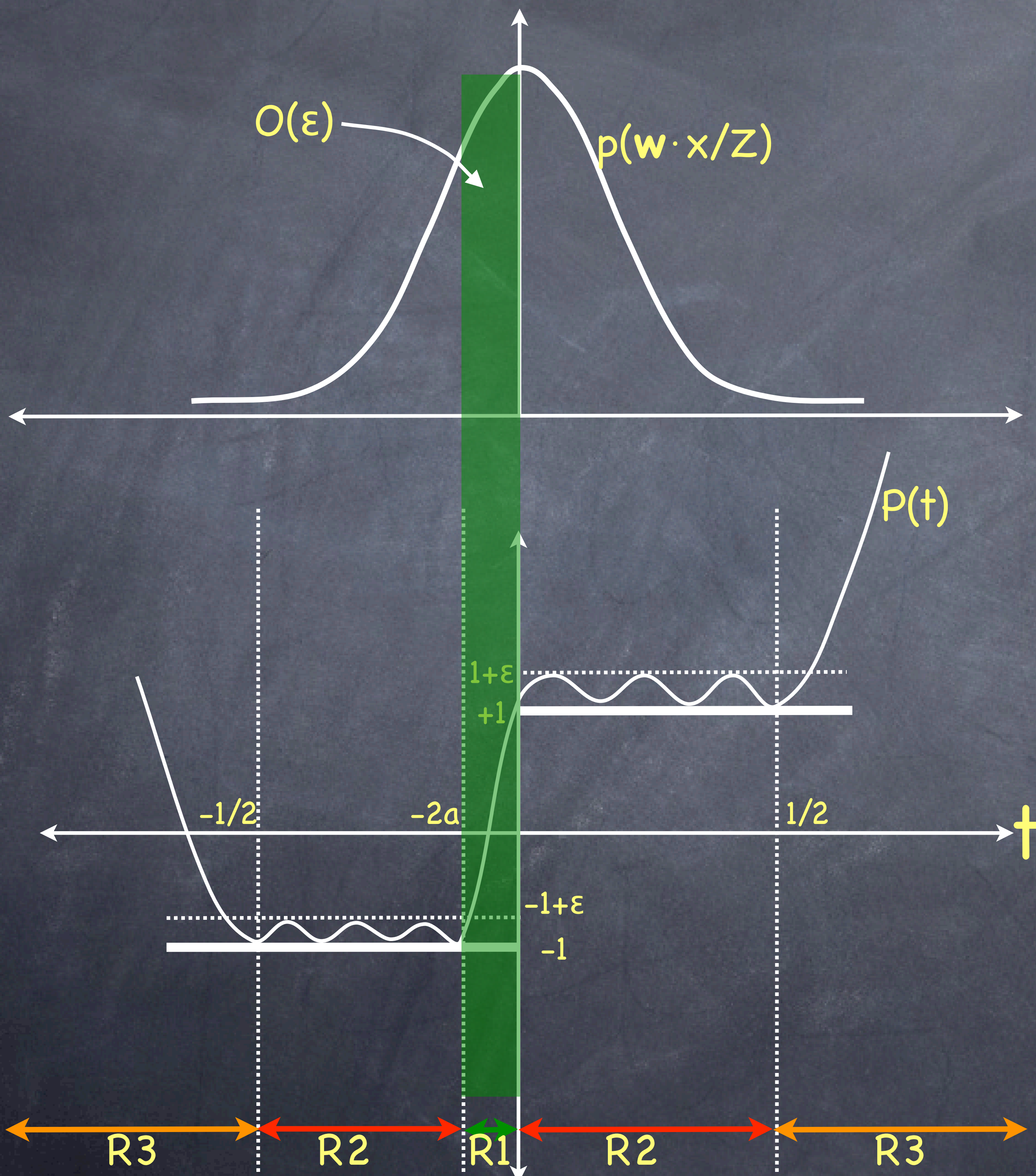
Let $\mathbf{w} = (w_1,\ldots,w_n)$ such that $\Sigma w_i^2 = 1$ and $\forall i$, $|w_i| \le \varepsilon$. Then

$$\forall t \in R, \ |\Pr_{x\text{<-}U}[\mathbf{w}\cdot x \le t] - \Phi(t)| \le \varepsilon,$$

where Φ is the cumulative distribution function of the standard normal $N(0, 1)$.



Proof Sketch: (Case (a): h is ε-regular)



- Lemma: $E_x[q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)] \le \varepsilon$.
- Proof: case analysis based on the value of $y = \mathbf{w}\cdot x/Z$
  - $y \in R2$:
    – $q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x) \le \varepsilon$.
    So, the contribution to the expectation is $\le \varepsilon$.
  - $y \in R3$:
    $q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x)$ grows as $|y|$ grows larger but $\Phi(y)$ diminishes (by Hoeffding). Hoeffding overshadows P's growth.
  - $y \in R1$:
    – $q_u(x) - \mathrm{sign}(\mathbf{w}\cdot x) \le (2+\varepsilon)$
    – $\Pr_x[y \in R1] = O(\varepsilon)$
    (from Berry-Esséen CLT)
    So, the contribution to the expectation is $O(\varepsilon)$.

Proof Sketch:

...where are we in the proof?

- We have shown the existence of sandwiching polynomials for halfspaces which are ε-regular.
- This implies that any k-wise independent distribution fools any ε-regular halfspace, provided $k \geq (C/\varepsilon^2)\cdot\log^2(1/\varepsilon)$.
- We need to show case(b), i.e., halfspaces that are not ε-regular.

Proof Sketch

(Case (b): h is not ε-regular)

- Based on structural properties of halfspaces studied in [Servedio'07].
- WLOG let $|w_1| \geq |w_2| \geq \ldots \geq |w_n|$.
- Definition (Critical Index): This is defined to be the smallest index I such that
  $|w_I| \leq \varepsilon \cdot \sigma_I, \quad \sigma_I = (w_I^2 + \ldots + w_n^2)$
  - For ε-regular halfspaces $I = 1$.

Proof Sketch

(Case (b): h is not ε-regular)

- Let head = $\{x_1,\dots,x_{I-1}\}$ and tail = $\{x_I,\dots,x_n\}$.
- Claim 1: For any fixing of head variables, the halfspace over the tail variables $h'(x_T)= (w_I\cdot x_I +\dots+w_n\cdot x_n + \Theta_H)$, $\Theta_H=(w_1\cdot x_1+\dots+w_{I-1}\cdot x_{I-1})$ is ε-regular.
  - Proof: From the definition of critical index

    $|w_n|\le|w_{n-1}|\le\dots\le|w_I| \le \varepsilon\cdot(w_I^2 + \dots+w_n^2)$.

    The claim follows from scaling.

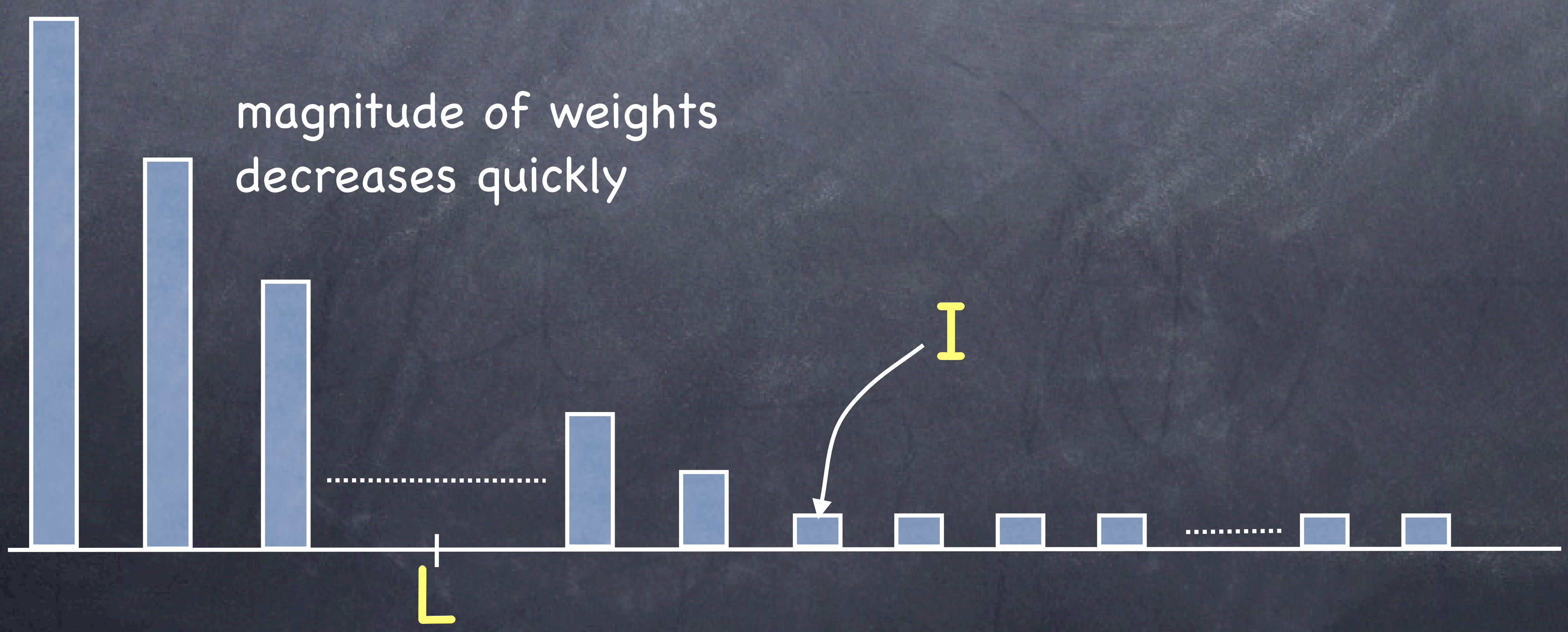
Proof Sketch

(Case (b): h is not ε-regular)

- Let $L = (8/\varepsilon^2) \cdot \log^2(10/\varepsilon)$.
- Claim 2: If the critical index $I \leq L$ for a halfspace h, then $(L + k)$-wise independence fools h with error ε, where $k \geq (C/\varepsilon^2) \cdot \log^2(1/\varepsilon)$.
  - Proof:
    (1) For any $(L+k)$-wise independent distribution D, conditioned on any fixed value of the head variables, the projection of D on the tail variables is at least k-wise independent.
    (2) For any fixing of the head variables the halfspace on tail variables is ε-regular.

Proof Sketch (Case (b): h is not ε-regular)

- Claim 3: If the critical index $I > L$ for a halfspace h, then $(L + 2)$-wise independence fools h with error ε.
  - Proof: For almost all fixings of the first L variables, the remaining variables hardly ever flips the value of the sign. (by Chernoff for uniform and by Chebychev for $(L+2)$-wise independence).



Proof Sketch: Summary

- Case(a): If h is ε-regular then any k-wise independent distribution fools h with error ε, provided $k \geq (C/\varepsilon^2)\cdot\log^2(1/\varepsilon)$.

- Case(b): If h is not ε-regular.
  Let $L = (8/\varepsilon^2)\cdot\log^2(10/\varepsilon)$

  - If $I \leq L$, then any $(L+k)$-wise independent distribution fools h with error ε.

  - If $I > L$, then any $(L+2)$-wise independent distribution fools h with error ε.

- So, any $(c/\varepsilon^2)\cdot\log^2(1/\varepsilon)$-wise independent distribution fools any halfspace with error ε.

Future Directions

- Polynomial threshold functions
  - Power of bounded independence [DKN'09]
  - Pseudorandom generators [MZ'09, LEY'09]

Questions?

Thank You