Title: LECTURE: Generative Modelling

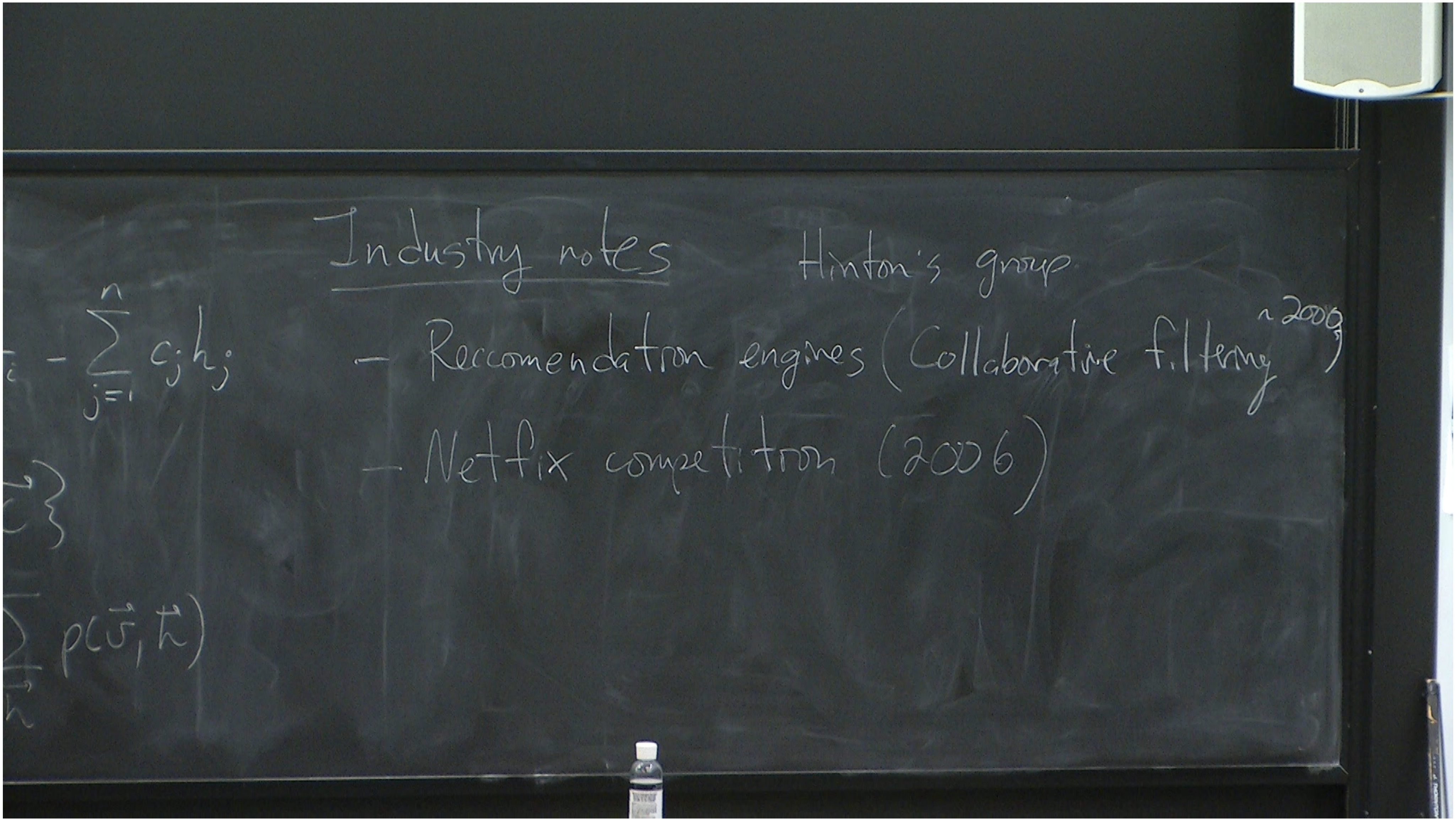
Speakers: Roger Melko

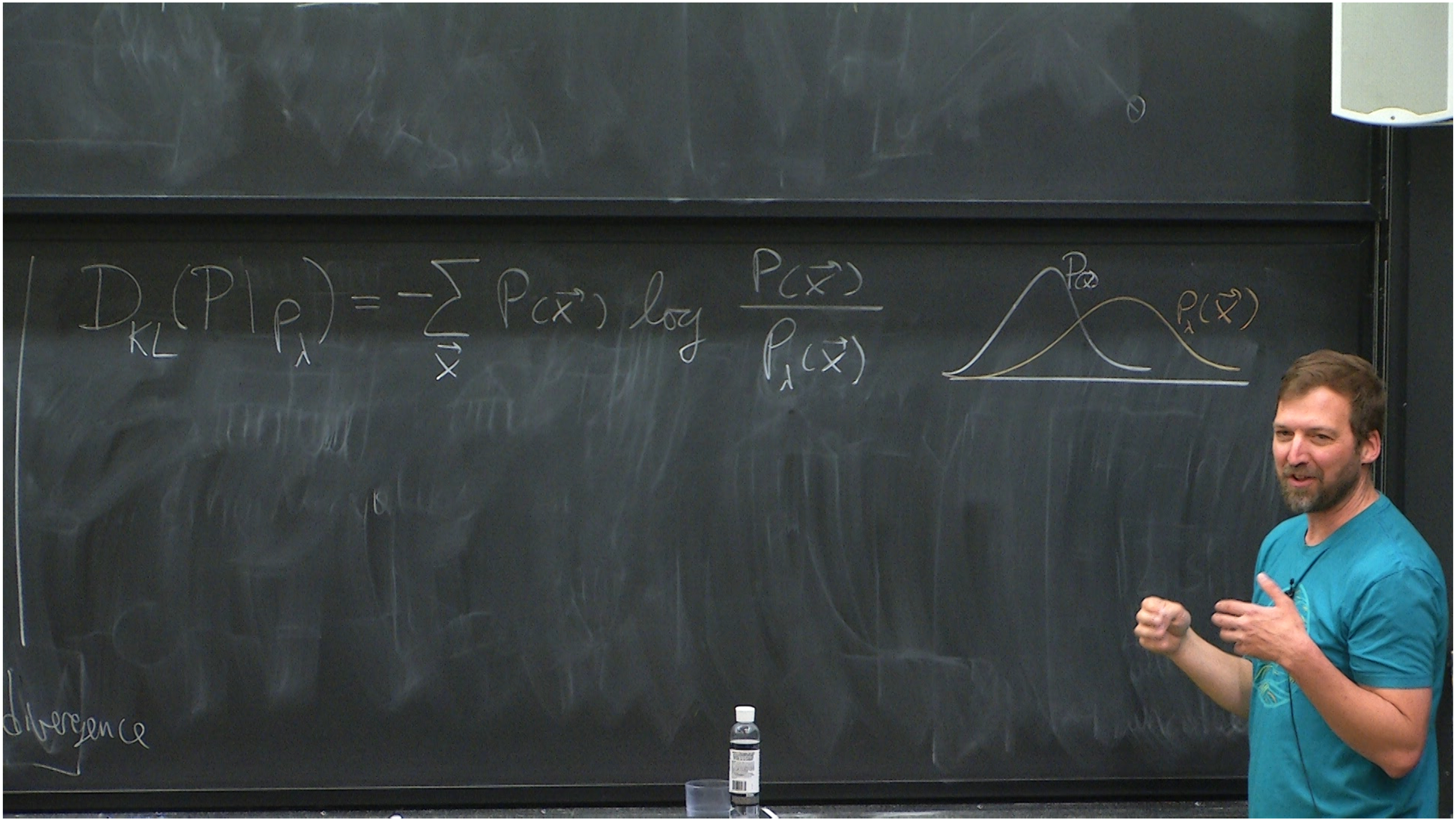
Collection: Quantum and AI Career Trajectories Mini-Course: Computational Methods and their Applications

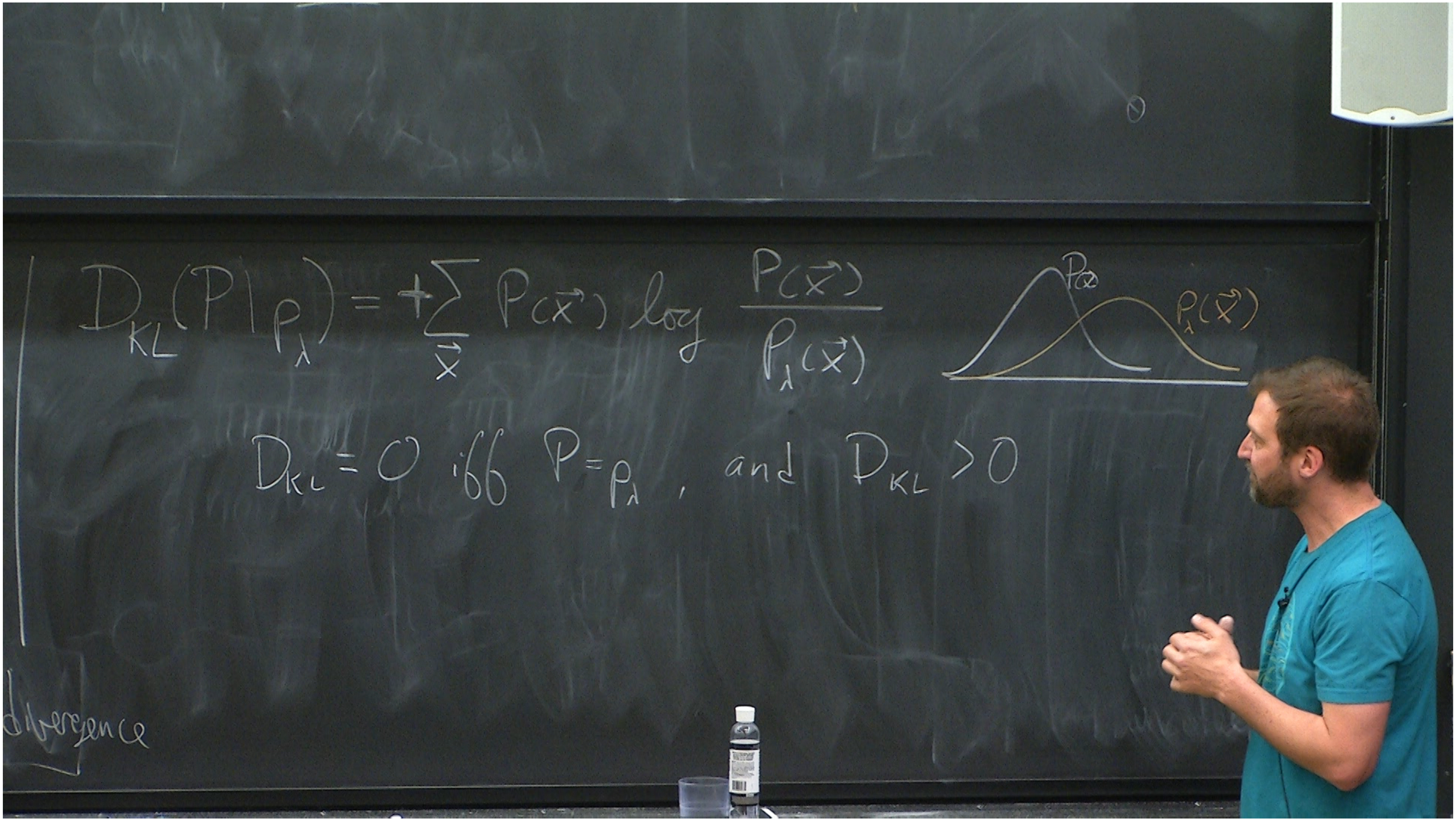
Date: May 09, 2023 - 11:00 AM

URL: https://pirsa.org/23050097

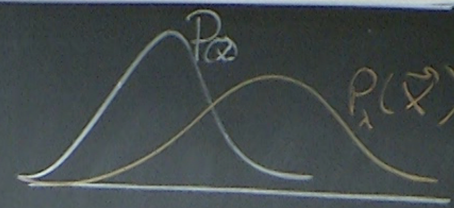
Abstract: TBC


Industry notes
Hinton's group
– Reccomendation engines (Collaborative filtering ~2000s)
– Netflix competition (2006)
$-\sum_{j=1}^{n} c_j h_j$
$p(\vec{v}, \vec{h})$


$D_{KL}(P \| P_\lambda) = -\sum_{\vec{x}} P(\vec{x}) \log \frac{P(\vec{x})}{P_\lambda(\vec{x})}$
$P(\vec{x})$
$P_\lambda(\vec{x})$
divergence


$D_{KL}(P|\rho_\lambda) = +\sum_{\vec{x}} P(\vec{x}) \log \frac{P(\vec{x})}{\rho_\lambda(\vec{x})}$
$P(x)$
$\rho_\lambda(\vec{x})$
$D_{KL} = 0$ iff $P = \rho_\lambda$, and $D_{KL} > 0$
divergence

$$D_{KL}(P|p_\lambda) = +\sum_{\vec{x}} P(\vec{x}) \log \frac{P(\vec{x})}{p_\lambda(\vec{x})}$$



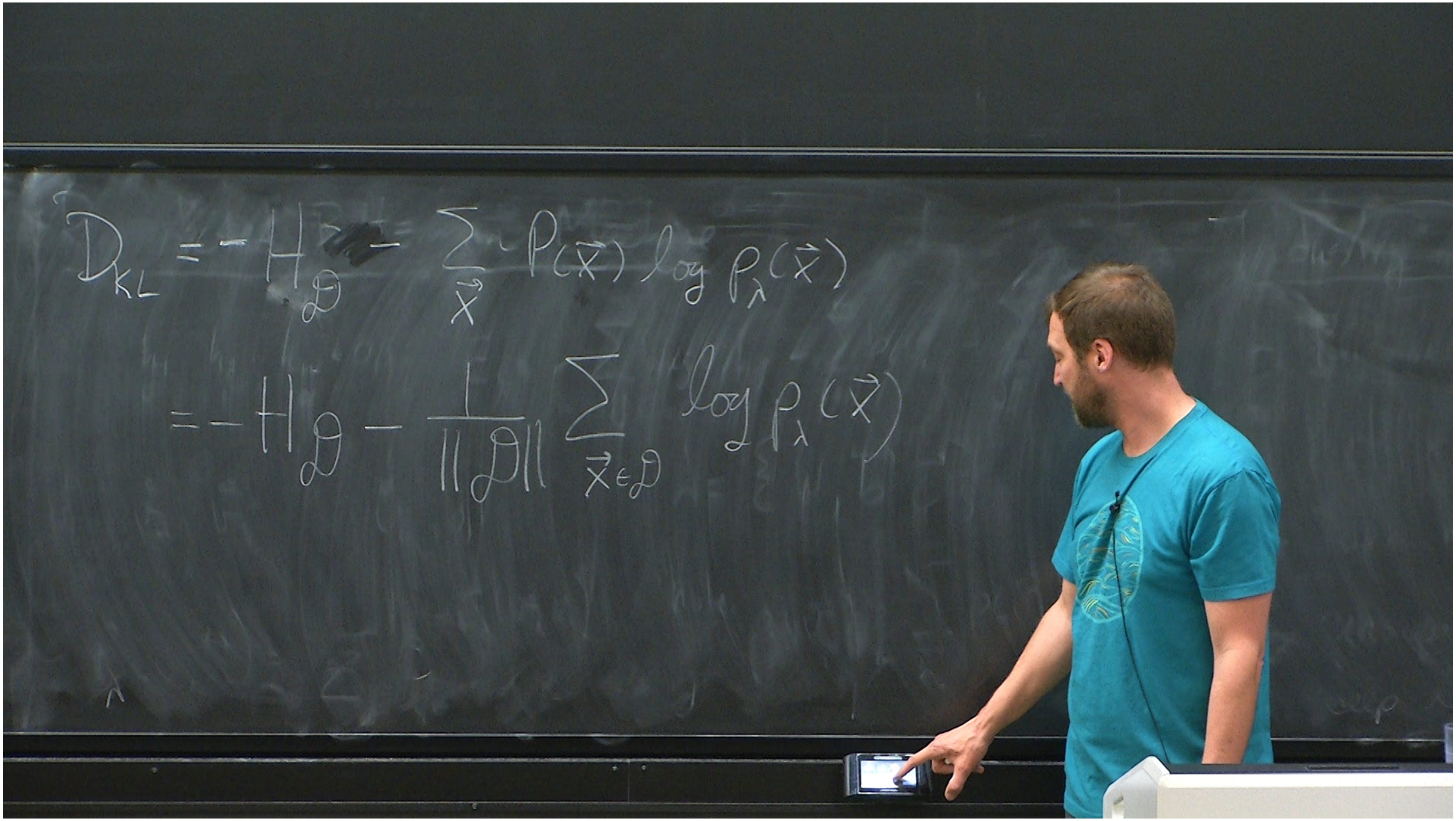
$D_{KL} = 0$ iff $P = p_\lambda$, and $D_{KL} > 0$

$$D_{KL} = \sum_{\vec{x}} P(\vec{x}) \log P(\vec{x}) - \sum_{\vec{x}} P(\vec{x}) \log p_\lambda(\vec{x})$$

divergence

$$\langle Q \rangle_{p_\lambda(\vec{v},\vec{h})} = \cancel{\frac{1}{Z}} \sum_{\vec{v}} \sum_{\vec{h}} Q\, p_\lambda(\vec{v},\vec{h})$$ e.g. $Q = E$ (of the RBM)

$Q = Q_v$ ... $\langle Q_v \rangle = \cancel{\frac{1}{Z}} \sum Q_v \sum p(\vec{v},\vec{h}) = \cancel{\frac{1}{Z}} \sum Q_v\, p(\vec{v})$


$D_{KL} = -H_{\mathcal{D}} - \sum_{\vec{x}} P(\vec{x}) \log p_{\lambda}(\vec{x})$
$= -H_{\mathcal{D}} - \frac{1}{\|\mathcal{D}\|} \sum_{\vec{x} \in \mathcal{D}} \log p_{\lambda}(\vec{x})$

...ata-driven setting } Kullback-Leibler divergence

X

$$D_{KL} = -H_{\mathcal{D}} - \sum_{\vec{x}} P(\vec{x}) \log p_\lambda(\vec{x})$$

$$= -H_{\mathcal{D}} - \underbrace{\frac{1}{\|\mathcal{D}\|} \sum_{\vec{x} \in \mathcal{D}} \log p_\lambda(\vec{x})}_{\langle \log p_\lambda \rangle_{\mathcal{D}, P(x)} = \mathbb{E}_{P(\vec{x})}(\log p_\lambda)}$$

from

$$P(\vec{x}) \simeq \frac{1}{\|\mathcal{D}\|} \sum_{x_k \in \mathcal{D}} \delta_{\vec{x}, \vec{x}_k}$$


$L_\lambda = -\langle \log p_\lambda(\vec{x}) \rangle_{P(\vec{x})}$

$$L_\lambda = -\langle \log p_\lambda(\vec{x}) \rangle_{P(\vec{x})}$$

"loss" function.

can include H_θ in verif.

Consider a single training example $\vec{v}$

$$\log p_\lambda(\vec{v}) = \log \frac{1}{Z} \sum_{\vec{h}} e^{-E_\lambda(\vec{v}, \vec{h})}$$

$$= \log \sum_{\vec{h}} e^{-E} - \log \sum_{\vec{v}\vec{h}} e^{-E}$$

The gradient:

$$\frac{\partial \log p_\lambda(\vec{v})}{\partial \lambda} = -\frac{1}{\sum_h e^{-E}} \sum_h e^{-E} \cdot \frac{\partial E}{\partial \lambda}$$

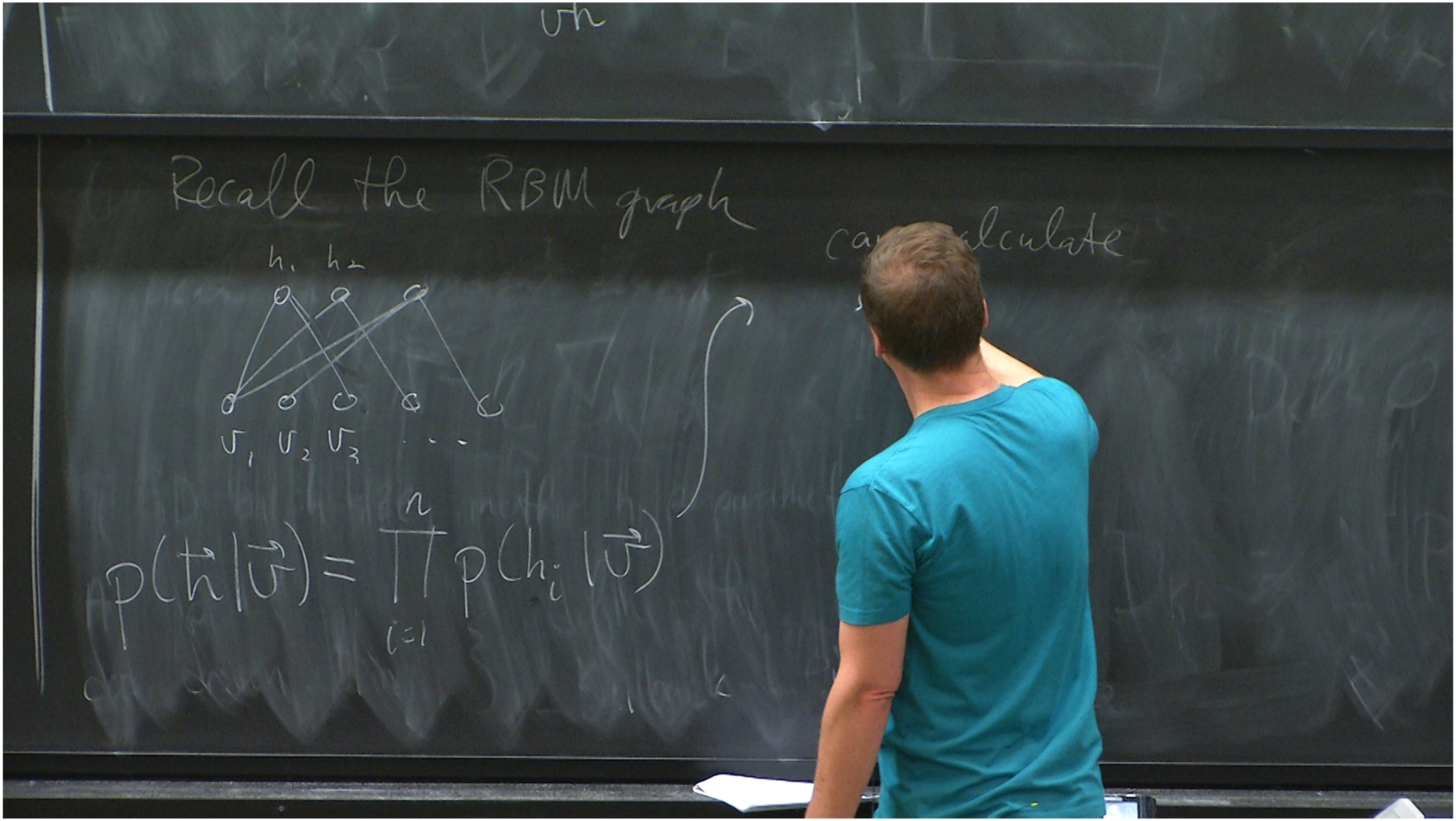
$$+ \frac{1}{\sum_{vh} e^{-E}} \sum_{vh} e^{-E} \cdot \frac{\partial E}{\partial \lambda}$$

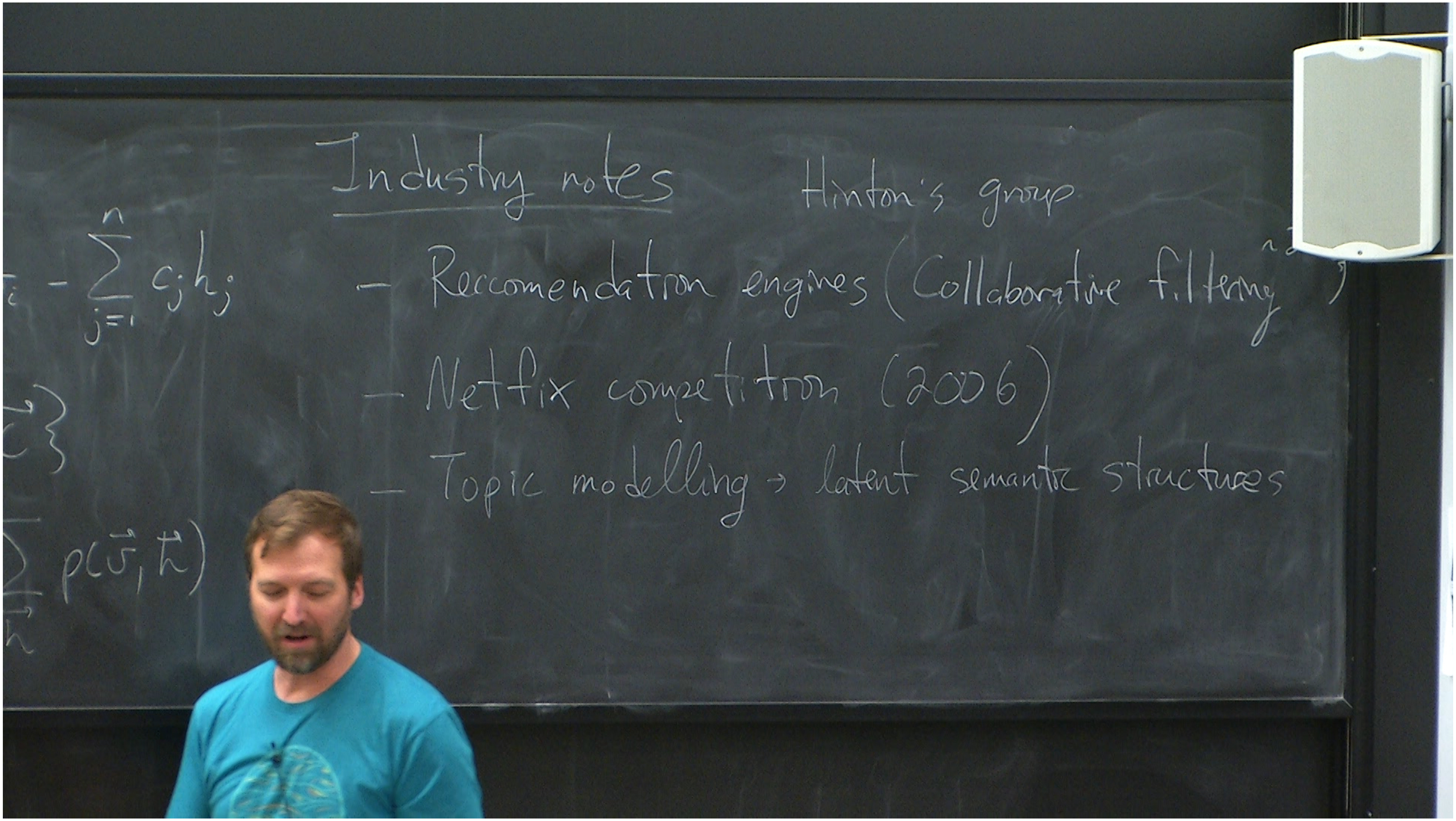
use $p(h|v) = \frac{p(v,h)}{p(v)} = e^{-E} \Big/ \sum_h e^{-E}$

$$\boxed{\frac{\partial \log p_\lambda}{\partial \lambda} = -\sum_h p(h|v)\frac{\partial E}{\partial \lambda} + \sum_{vh} p(v,h)\frac{\partial E}{\partial \lambda}}$$

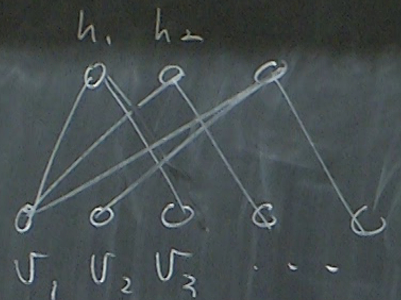
Tractability? first $\frac{\partial E}{\partial \lambda}$ e.g. $\lambda = W_{ij}$, $\frac{\partial E}{\partial W_{ij}} = -v_i h_j$

$\min \| M - \tilde{M} \|$ → obtained from


Recall the RBM graph
h_1 h_2
v_1 v_2 v_3 . . .
$p(\vec{h}|\vec{v}) = \prod_{i=1}^{n} p(h_i|\vec{v})$


Industry notes
Hinton's group
– Reccomendation engines (Collaborative filtering)
– Netfix competition (2006)
– Topic modelling → latent semantic structures

Recall the RBM graph



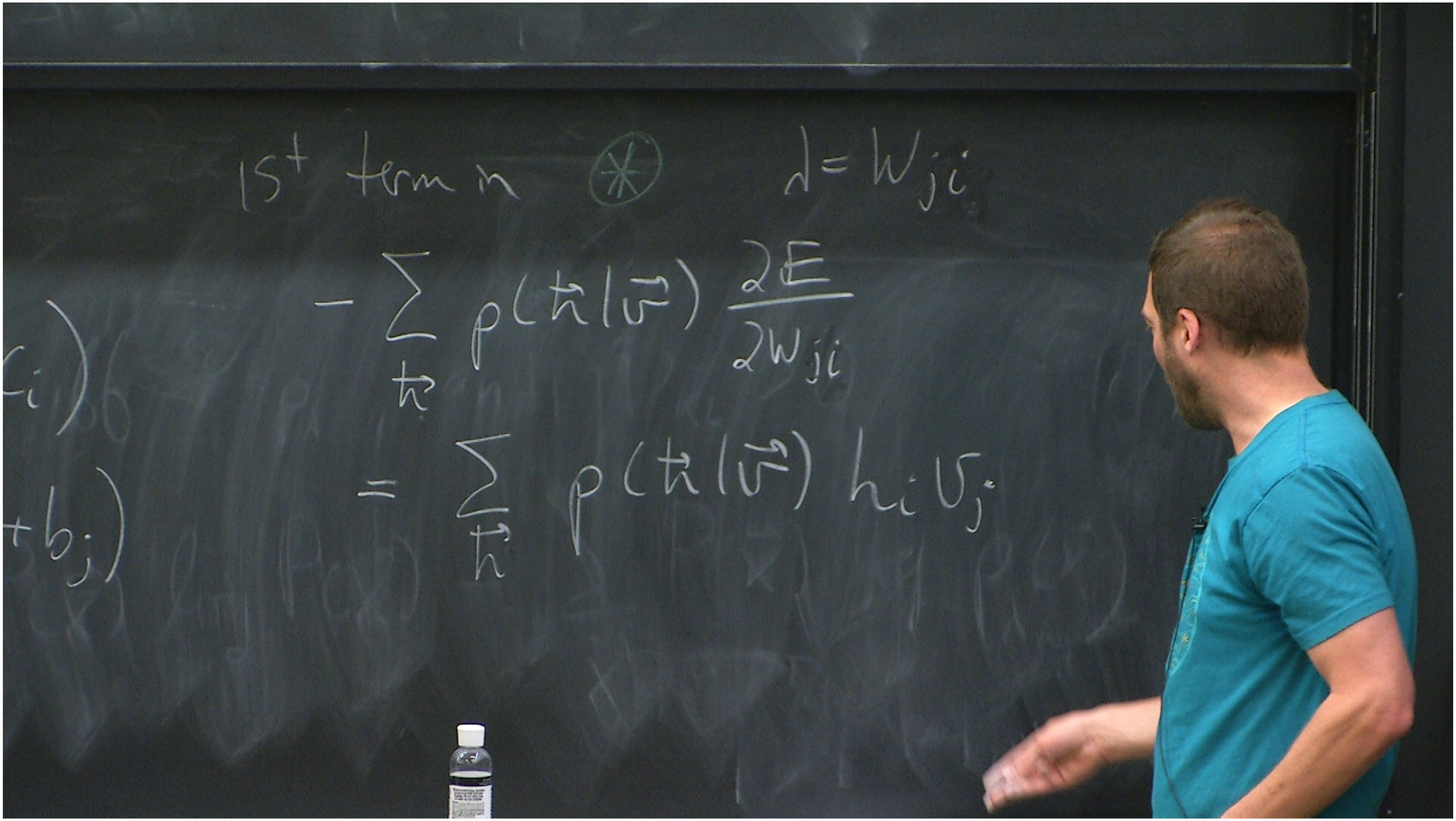
can calculate exactly (exercise)

$$p(h_i = 1 | \vec{v}) = \sigma\left(\sum_{i=1}^{m} W_{ij} v_i + c_i \right)$$

$$p(v_i = 1 | \vec{h}) = \sigma\left(\sum_{j=1}^{n} W_{ij} h_j + b_j \right)$$

$$p(\vec{h}|\vec{v}) = \prod_{i=1}^{n} p(h_i | \vec{v})$$

where $\sigma(x) = \dfrac{1}{1+e^{-x}}$



1^{st} term in (*)
$\lambda = W_{ji}$
$- \sum_{\vec{h}} p(\vec{h}|\vec{v}) \frac{\partial E}{\partial W_{ji}}$
$= \sum_{\vec{h}} p(\vec{h}|\vec{v}) h_i v_j$

1^{st} term in (*) $\qquad \lambda = W_{ji}$

$c_i)$

$+b_j)$

$$-\sum_{\vec{h}} p(\vec{h}|\vec{v}) \frac{\partial E}{\partial W_{ji}}$$

$$= \sum_{\vec{h}} \underbrace{p(\vec{h}|\vec{v})}_{p(h_i|\vec{v})\, p(\vec{h}_{-i}|\vec{v})} h_i v_j$$

$$= \sum_{h_i} p(h_i|\vec{v})\, h_i v_j \sum_{\vec{h}_{-i}} p(\vec{h}_{-i}|\vec{v})$$

recall $\sum_{\vec{h}_{-i}}$

recall $\sum_{\vec{h}_{-i}} = \underbrace{\sum_{h_1=\{0,1\}}}_{p(h_1=0|\vec{v}) + p(h_1=1|\vec{v}) = 1} \sum_{h_2=\{0,1\}} \dots$ etc $\leftarrow (1)^{(n-1)}$ here

$$= \sum_{h_i} p(h_i|\vec{v}) h_i v_j \sum_{\vec{h}_{-i}} p(\vec{h}_{-i}|\vec{v})$$

recall $\sum_{\vec{h}_{-i}}$

$$= \sum_{h_i} p(h_i|\vec{v}) h_i v_j$$

$$= p(h_i = 0|\vec{v}) \cdot 0 + p(h_i = 1|\vec{v}) \cdot v_j$$

$$= p(h_i = 1|\vec{v}) v_j = \sigma\left(\sum_k W_{ik} v_k + c_k\right) v_j$$

$$= \sum_{h_i} p(h_i|\vec{v})\, h_i v_j \sum_{\vec{h}_{-i}} p(\vec{h}_{-i}|\vec{v})$$

recall $\sum_{\vec{h}_{-i}}$

Tractable

$$= \sum_{h_i} p(h_i|\vec{v})\, h_i v_j$$

$$= p(h_i=0|\vec{v})\cdot 0 + p(h_i=1|\vec{v})\cdot v_j$$

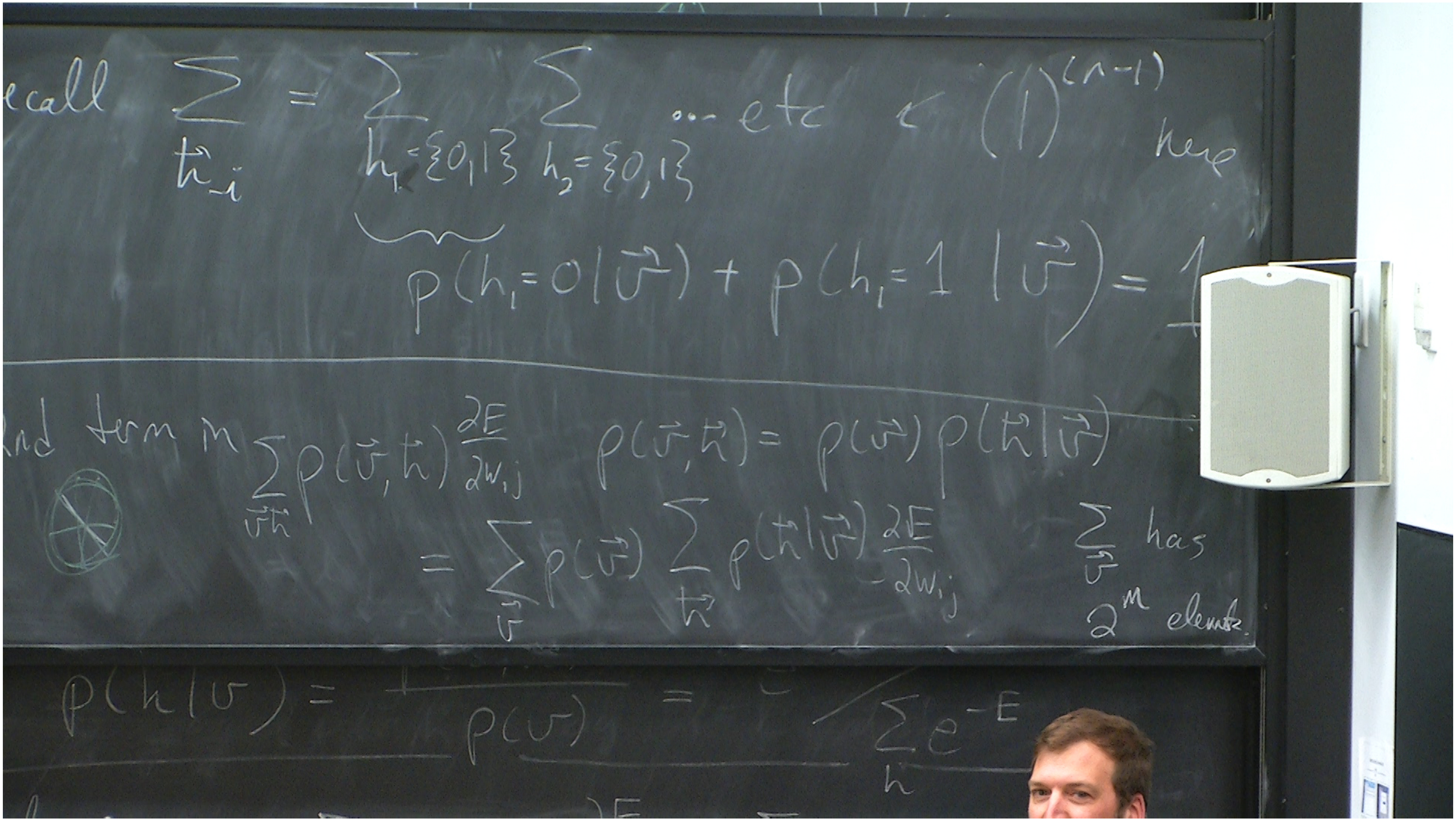
$$= p(h_i=1|\vec{v})\, v_j = \sigma\left(\sum_k w_{ik} v_k + c_k\right) v_j$$

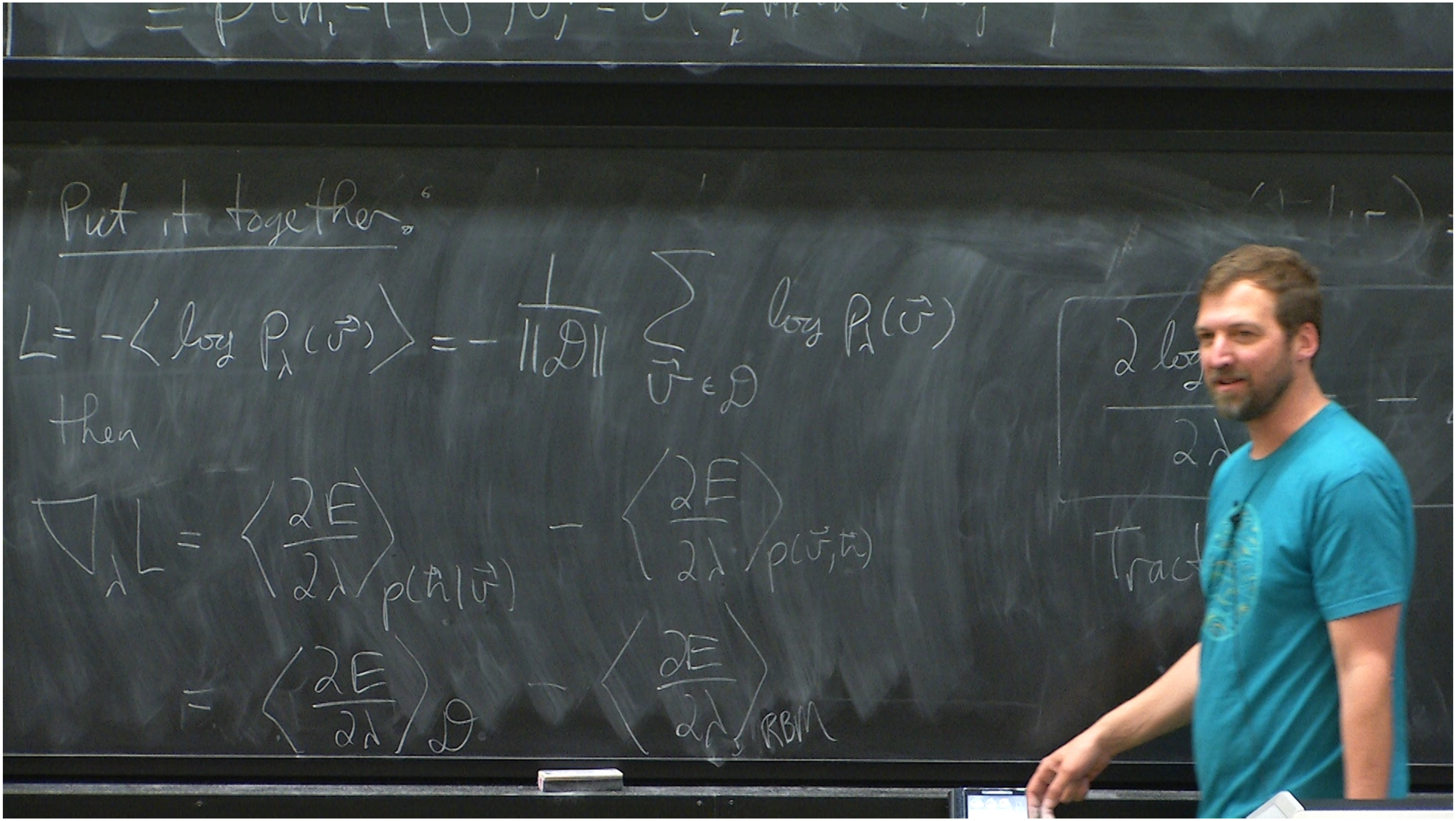
table

recall $\sum_{\vec{h}_{-i}} = \underbrace{\sum_{h_1=\{0,1\}}}_{p(h_1=0|\vec{v}) + p(h_1=1|\vec{v}) = 1} \sum_{h_2=\{0,1\}} \dots$ etc $\leftarrow (1)^{(n-1)}$ here

2nd term in $\sum_{\vec{v}\vec{h}} p(\vec{v},\vec{h}) \frac{\partial E}{\partial w_{ij}}$ $\quad p(\vec{v},\vec{h}) = p(\vec{v})\, p(\vec{h}|\vec{v})$

$$= \sum_{\vec{v}} p(\vec{v}) \sum_{\vec{h}} p(\vec{h}|\vec{v}) \frac{\partial E}{\partial w_{ij}}$$


ecall $\sum_{\vec{h}_i} = \sum_{h_1=\{0,1\}} \sum_{h_2=\{0,1\}} \dots$ etc $\leftarrow (1)^{(n-1)}$ here
$p(h_1=0|\vec{v}) + p(h_1=1|\vec{v}) = 1$
2nd term in $\sum_{\vec{v}\vec{h}} p(\vec{v},\vec{h}) \frac{\partial E}{\partial w_{ij}}$
$p(\vec{v},\vec{h}) = p(\vec{v})\, p(\vec{h}|\vec{v})$
$= \sum_{\vec{v}} p(\vec{v}) \sum_{\vec{h}} p(\vec{h}|\vec{v}) \frac{\partial E}{\partial w_{ij}}$
$\sum_{\vec{v}}$ has 2^M elemts.
$p(h|v) = $
$p(v)$
$\sum_h e^{-E}$

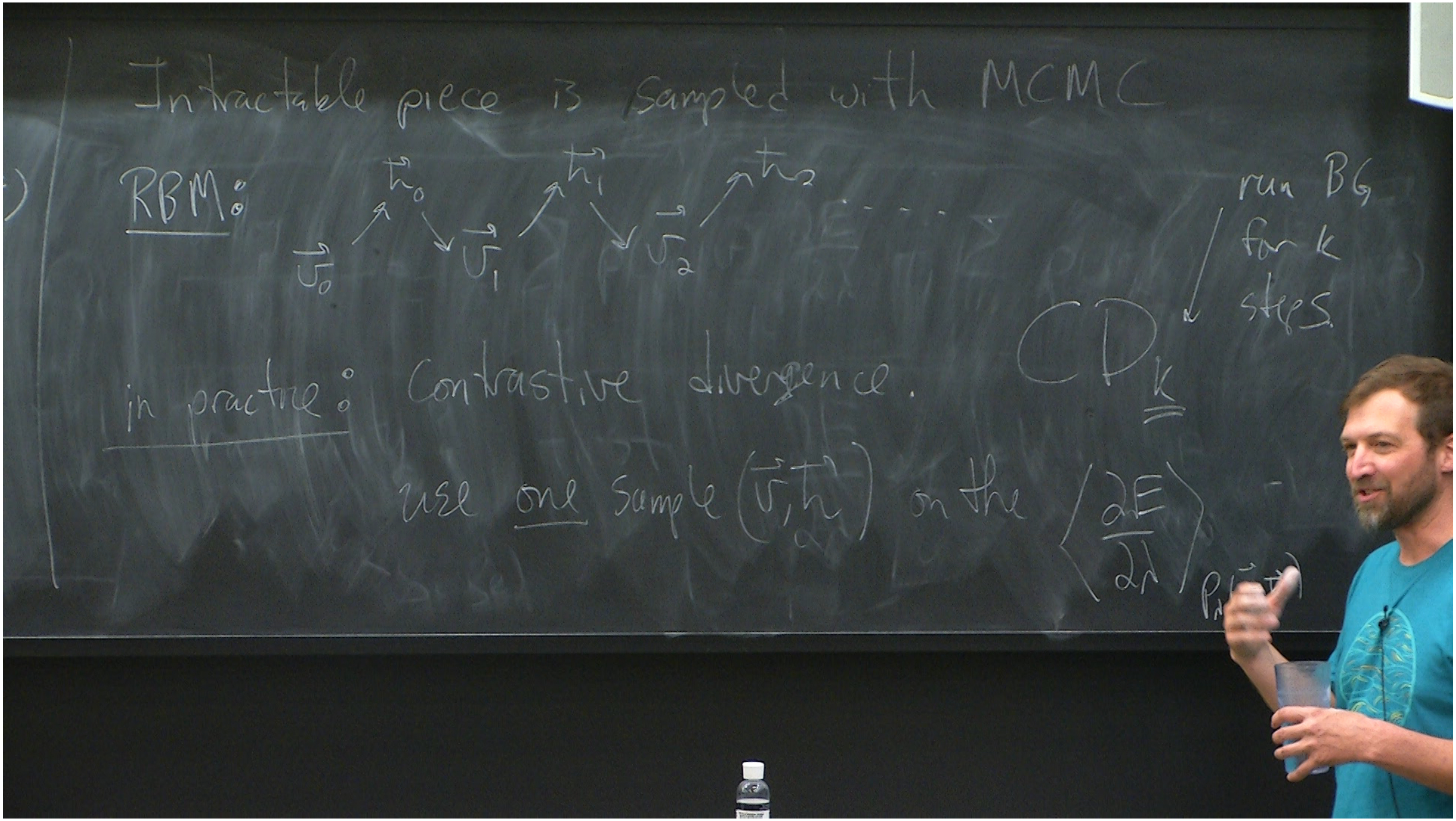

Put it together:
$L = -\langle \log P_\lambda(\vec{v}) \rangle = -\frac{1}{\|\mathcal{D}\|} \sum_{\vec{v} \in \mathcal{D}} \log P_\lambda(\vec{v})$
then
$\nabla_\lambda L = \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{p(\vec{h}|\vec{v})} - \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{p(\vec{v},\vec{h})}$
$= \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{\mathcal{D}} - \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{RBM}$

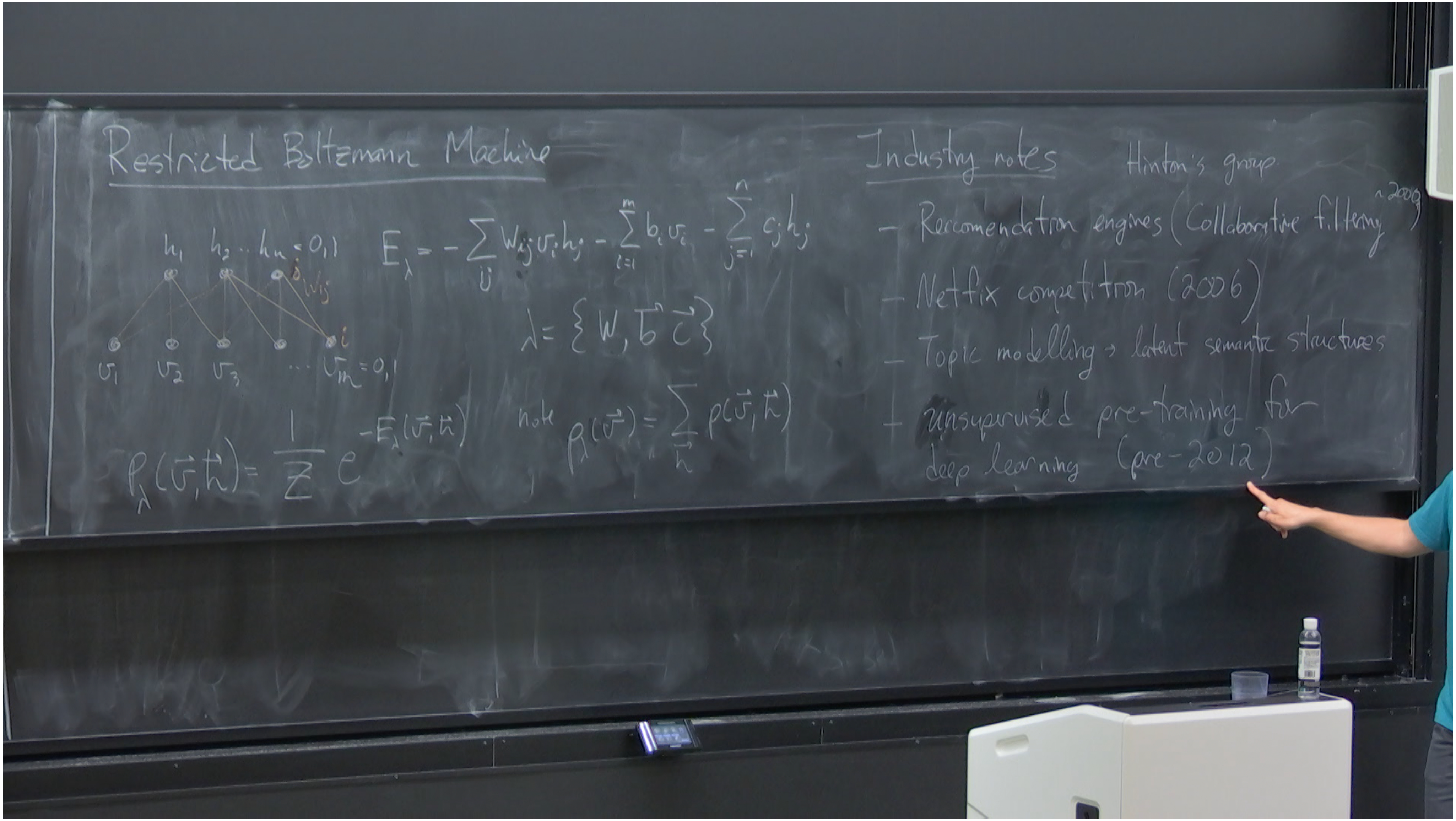
$$-\sum_{j=1}^{n} c_j h_j$$

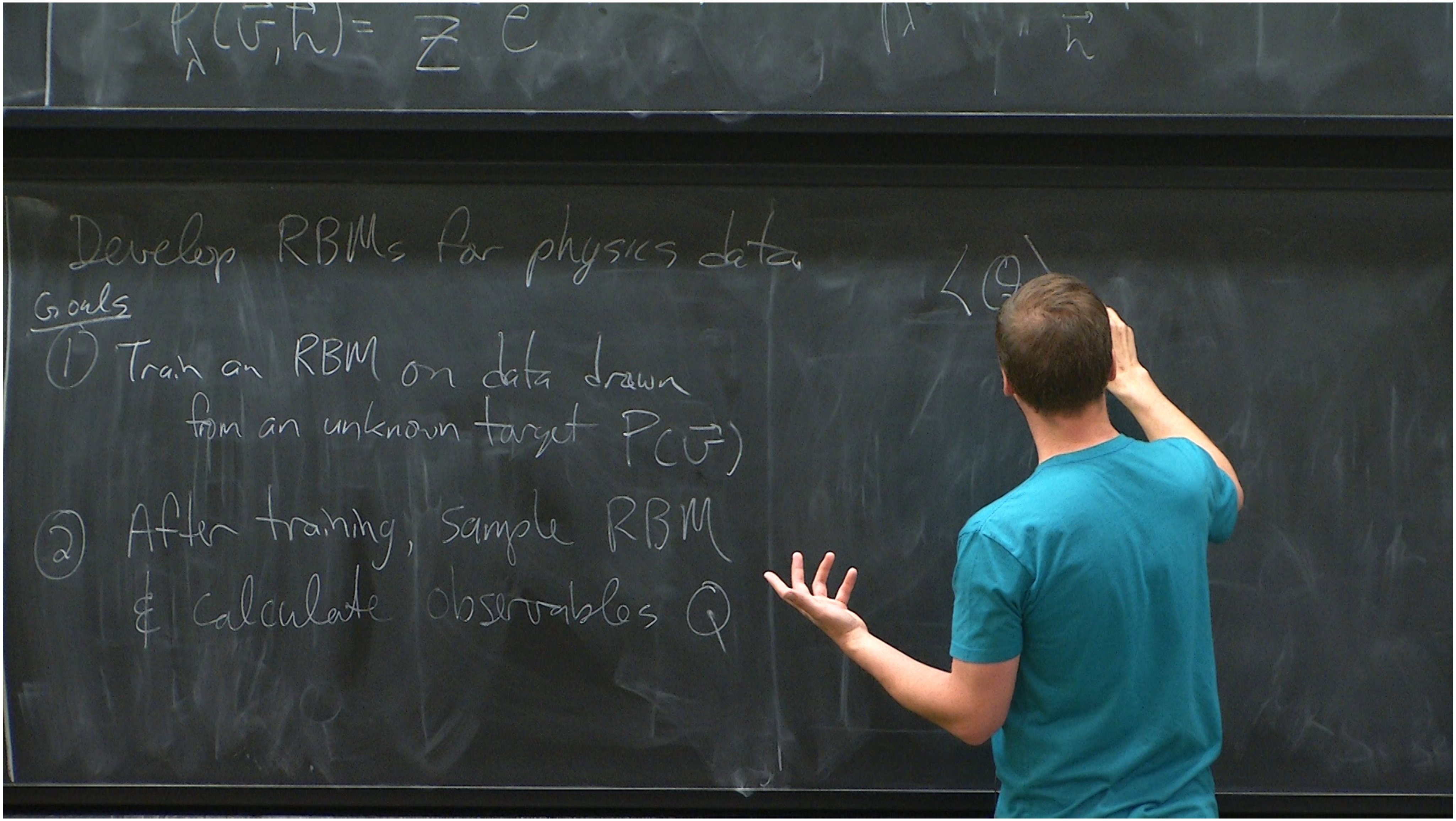
$$\sum_{\vec{h}} p(\vec{v},\vec{h})$$

Industry notes   Hinton's group

- Reccomendation engines (Collaborative filtering)
- Netflix competition (2006)
- Topic modelling → latent semantic structures
- unsupervised pre-training for


Intractable piece is sampled with MCMC
RBM:
run BG for k steps
CD_k
in practice: Contrastive divergence.
use one sample on the


Restricted Boltzmann Machine
Industry notes
Hinton's group
~2009
- Reccomendation engines (Collaborative filtering)
- Netfix competition (2006)
- Topic modelling → latent semantic structures
- unsupervised pre-training for deep learning (pre-2012)
h1 h2 .. hn = 0,1
v1 v2 v3 ... vm = 0,1
Wij
E_λ = - Σ_ij Wij vi hj - Σ_{i=1}^m bi vi - Σ_{j=1}^n cj hj
λ = {W, b, c}
P_λ(v, h) = 1/Z e^{-E_λ(v, h)}
note P_λ(v) = Σ_h p(v, h)


Develop RBMs for physics data
Goals
① Train an RBM on data drawn
from an unknown target P(v)
② After training, sample RBM
& calculate observables Q
⟨Q⟩

$P_\lambda(\vec{v},\vec{h}) = Z \; e \quad \ldots \quad \vec{h}$

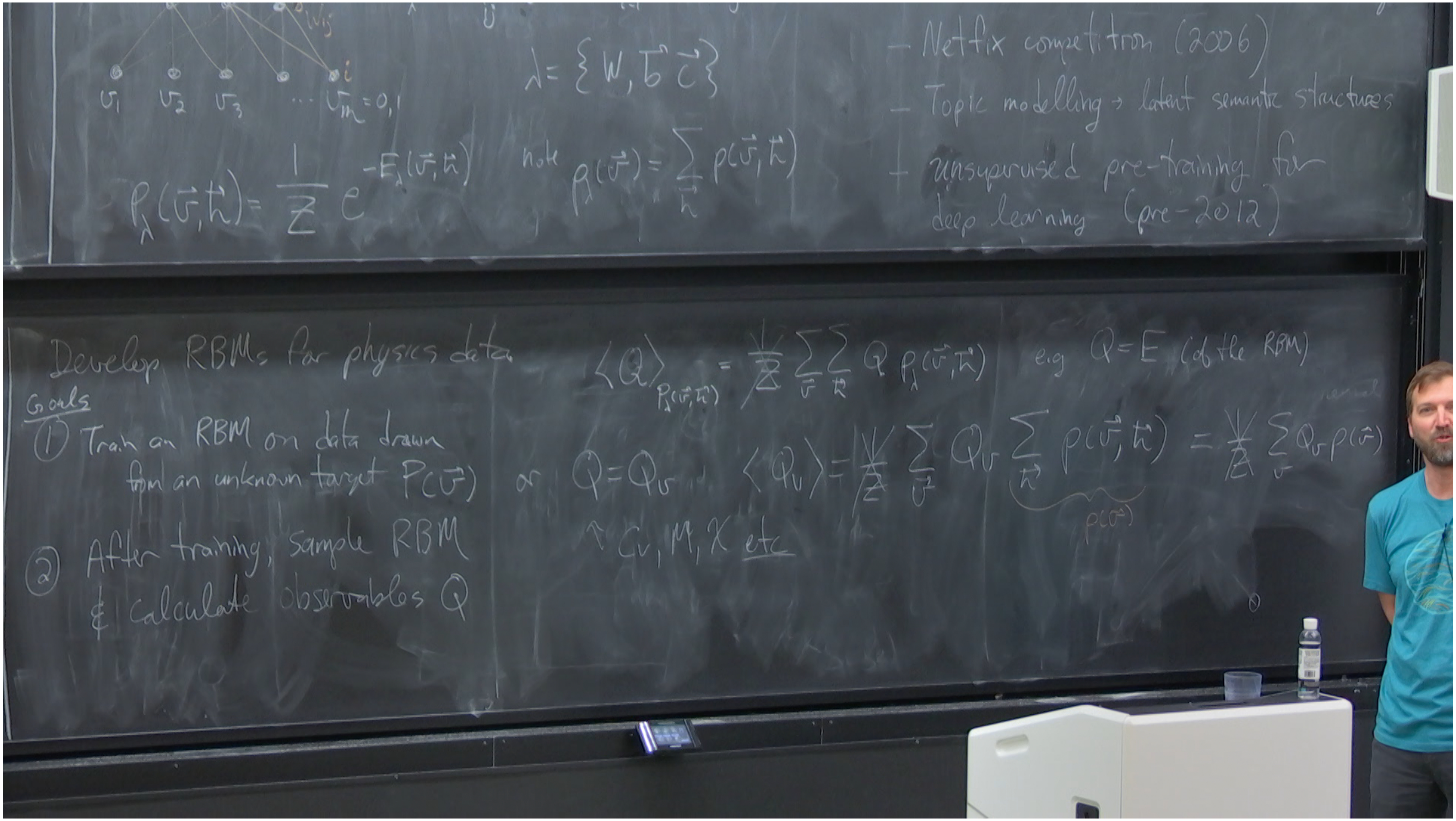
Develop RBMs for physics data

$\langle Q \rangle_{P_\lambda(\vec{v},\vec{h})}$

Goals

1. Train an RBM on data drawn from an unknown target $P(\vec{v})$

2. After training, sample RBM & calculate observables Q


v_1 v_2 v_3 ... $v_m = 0,1$
$\lambda = \{W, \vec{b}, \vec{c}\}$
$p_\lambda(\vec{v},\vec{h}) = \frac{1}{Z} e^{-E_\lambda(\vec{v},\vec{h})}$
note $p_\lambda(\vec{v}) = \sum_{\vec{h}} p(\vec{v},\vec{h})$
– Netflix competition (2006)
– Topic modelling → latent semantic structures
– unsupervised pre-training for deep learning (pre-2012)
Develop RBMs for physics data
Goals
① Train an RBM on data drawn from an unknown target $P(\vec{v})$
② After training, sample RBM & calculate observables Q
$\langle Q \rangle_{p_\lambda(\vec{v},\vec{h})} = \frac{1}{Z} \sum_{\vec{v}} \sum_{\vec{h}} Q \, p_\lambda(\vec{v},\vec{h})$ e.g. Q = E (of the RBM)
or $Q = Q_v$ $\langle Q_v \rangle = \frac{1}{Z} \sum_{\vec{v}} Q_v \underbrace{\sum_{\vec{h}} p(\vec{v},\vec{h})}_{p(\vec{v})} = \frac{1}{Z} \sum_{\vec{v}} Q_v \, p(\vec{v})$
C_v, M, χ etc

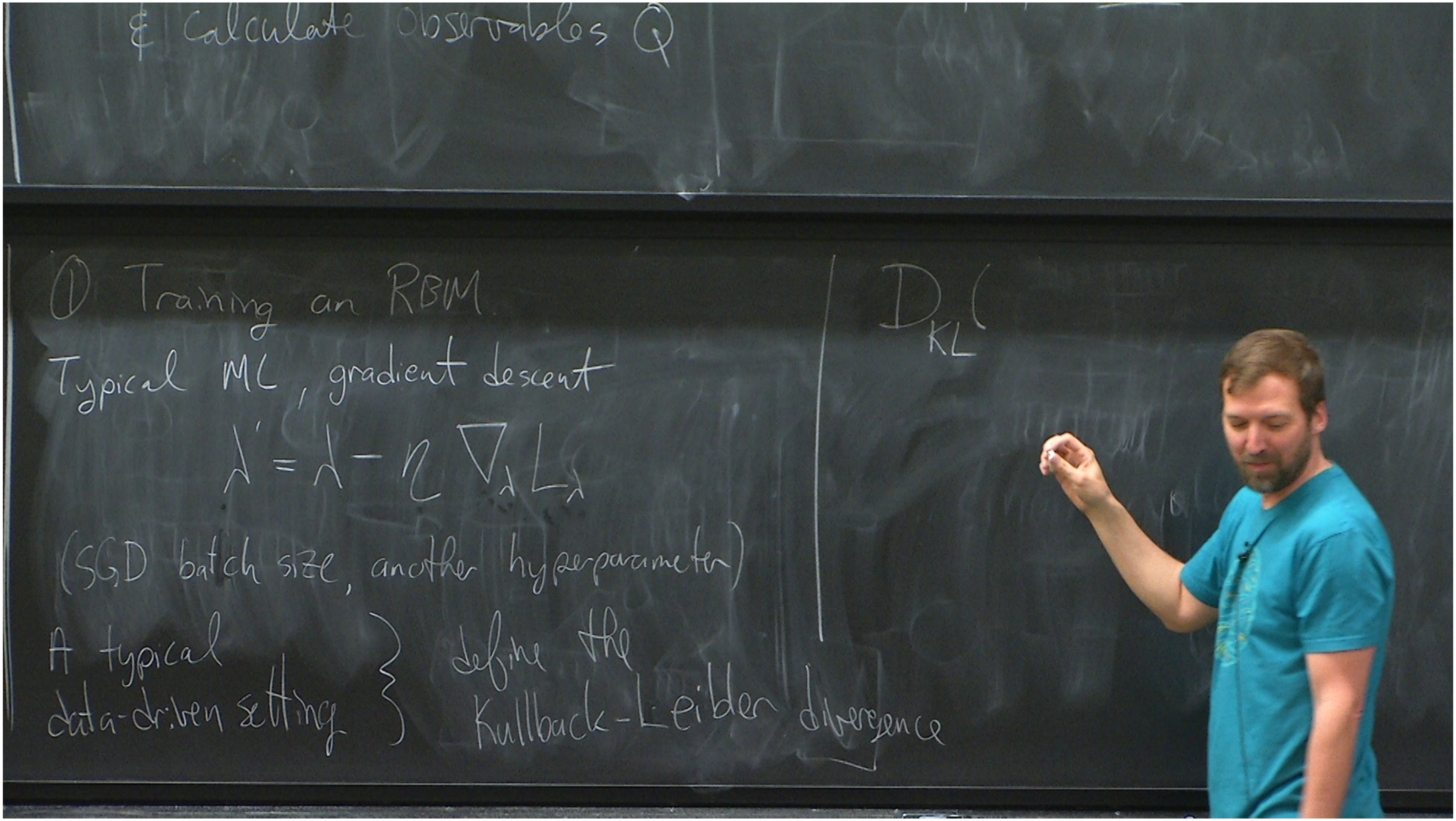
& Calculate observables Q

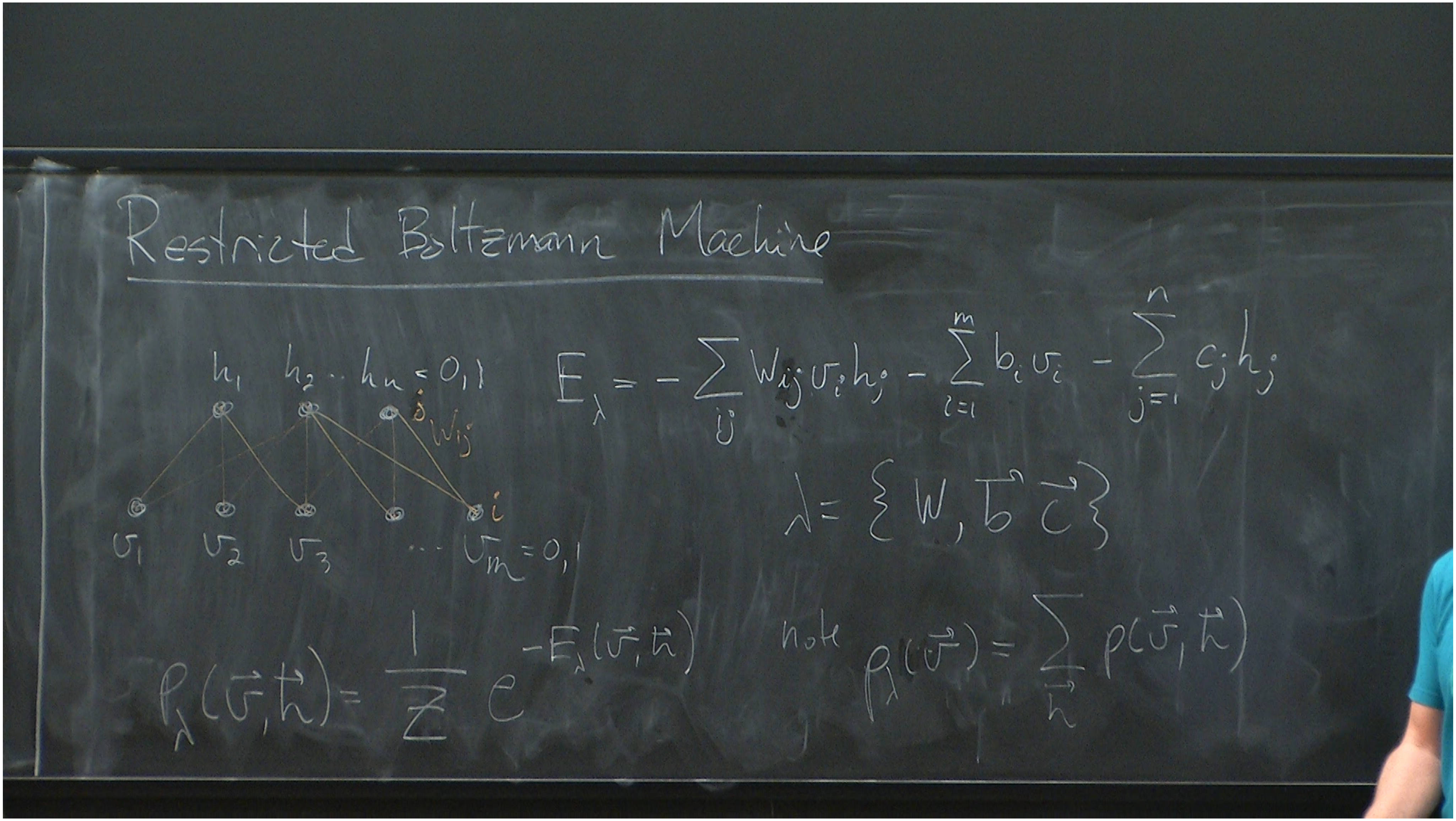
① Training an RBM

Typical ML, gradient descent

$$\lambda' = \lambda - \eta \nabla_\lambda L_\lambda$$

(SGD batch size, another hyperparameter)


& Calculate Observables Q
① Training an RBM
Typical ML, gradient descent
$\lambda' = \lambda - \eta \nabla_\lambda L_\lambda$
(SGD batch size, another hyperparameter)
A typical data-driven setting
define the Kullback-Leibler divergence
$D_{KL}($


Restricted Boltzmann Machine
$h_1 \quad h_2 \cdots h_n = 0,1$
j
W_{ij}
i
$v_1 \quad v_2 \quad v_3 \quad \cdots \quad v_m = 0,1$
$E_\lambda = -\sum_{ij} W_{ij} v_i h_j - \sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} c_j h_j$
$\lambda = \{ W, \vec{b}, \vec{c} \}$
$p_\lambda(\vec{v},\vec{h}) = \frac{1}{Z} e^{-E_\lambda(\vec{v},\vec{h})}$
note $p_\lambda(\vec{v}) = \sum_{\vec{h}} p(\vec{v},\vec{h})$