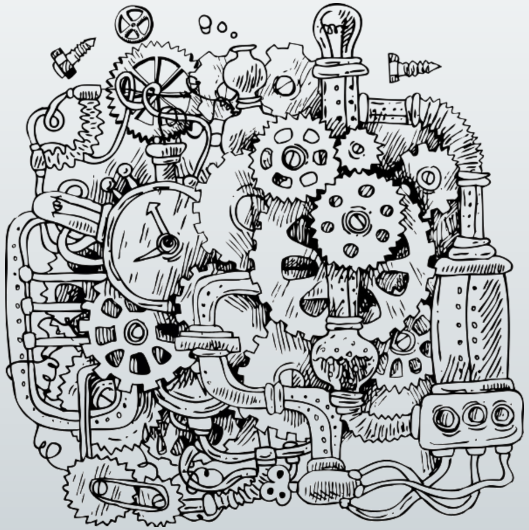
Generalization bounds for uniformly stable algorithms

Vitaly Feldman
Google Brain

with Jan Vondrak
Stanford University

Generalization bounds

Dataset $S = (z_1, \ldots, z_n) \sim P^n$



Learning algorithm A

Model w

Data distribution P

Loss function $\ell(w, z)$

Generalization error/gap for $w = A(S)$:

$$\Delta_S\big(\ell(w)\big) = \mathbf{E}_{z\sim P}[\ell(w,z)] - \frac{1}{n}\sum_{i=1}^{n} \ell(w, z_i)$$

Uniform stability [Bousquet,Elisseeff '02]

A has uniform stability γ w.r.t. loss ℓ if

for all S, S' that differ in a single element and $z \in Z$

$$|\ell(A(S), z) - \ell(A(S'), z)| \leq \gamma$$

Uniform stability [Bousquet,Elisseeff '02]

A has uniform stability γ w.r.t. loss ℓ if
for all S, S' that differ in a single element and $z \in Z$

$$|\ell(A(S), z) - \ell(A(S'), z)| \leq \gamma$$

Examples:

- Strongly convex ERM [BE '02; Shalev-Shwartz,Shamir,Srebro,Sridharan '09]
- Gradient descent on convex smooth losses [Hardt,Recht,Singer '16]

Typical $\gamma = 1/\sqrt{n}$

From stability to generalization

For ℓ with range $[0,1]$ and A with uniform stability $\gamma \in \left[\frac{1}{n}, 1\right]$

[Rogers,Wagner '78]

$$\left| \mathop{\mathbf{E}}_{S \sim P^n} [\Delta_S(\ell(A))] \right| \leq \gamma$$

From stability to generalization

For ℓ with range $[0,1]$ and A with uniform stability $\gamma \in \left[\frac{1}{n}, 1\right]$

[Rogers,Wagner '78]

$$\left| \mathop{\mathbf{E}}_{S \sim P^n} [\Delta_S(\ell(A))] \right| \leq \gamma$$

[Bousquet,Elisseeff '02]

$$\mathop{\mathbf{Pr}}_{S \sim P^n} \left[\Delta_S(\ell(A)) \geq \gamma \sqrt{\boldsymbol{n} \log(1/\delta)} \right] \leq \delta$$

From stability to generalization

For ℓ with range $[0,1]$ and A with uniform stability $\gamma \in \left[\frac{1}{n}, 1\right]$

[Rogers,Wagner '78]

$$\left| \mathop{\mathbf{E}}_{S \sim P^n} [\Delta_S(\ell(A))] \right| \leq \gamma$$

[Bousquet,Elisseeff '02]

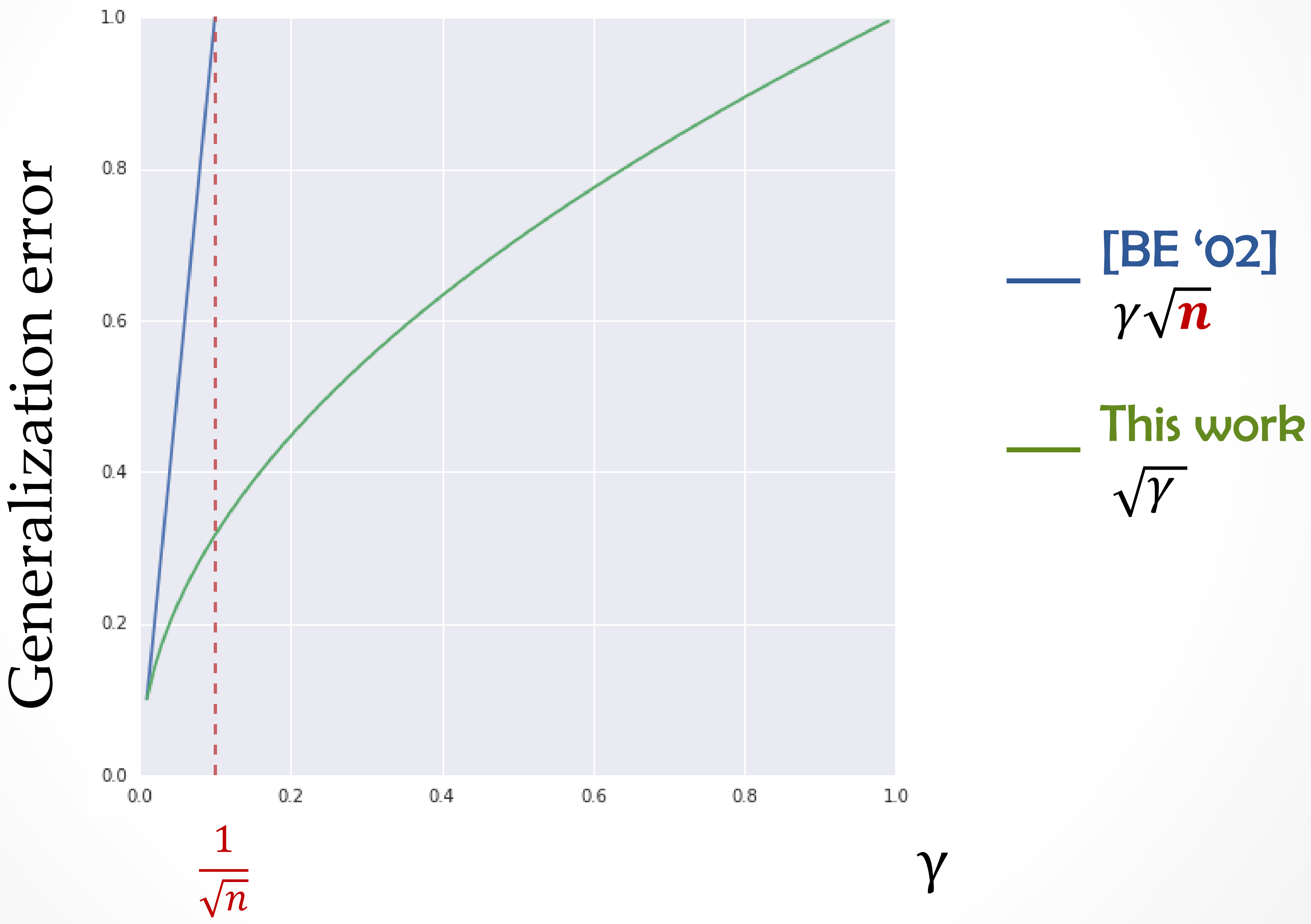
$$\Pr_{S \sim P^n} \left[\Delta_S(\ell(A)) \geq \gamma \sqrt{\boldsymbol{n} \log(1/\delta)} \right] \leq \delta$$

NEW!

$$\Pr_{S \sim P^n} \left[\Delta_S(\ell(A)) \geq \sqrt{\gamma \log(1/\delta)} \right] \leq \delta$$

Comparison

$n = 100$



Second moment

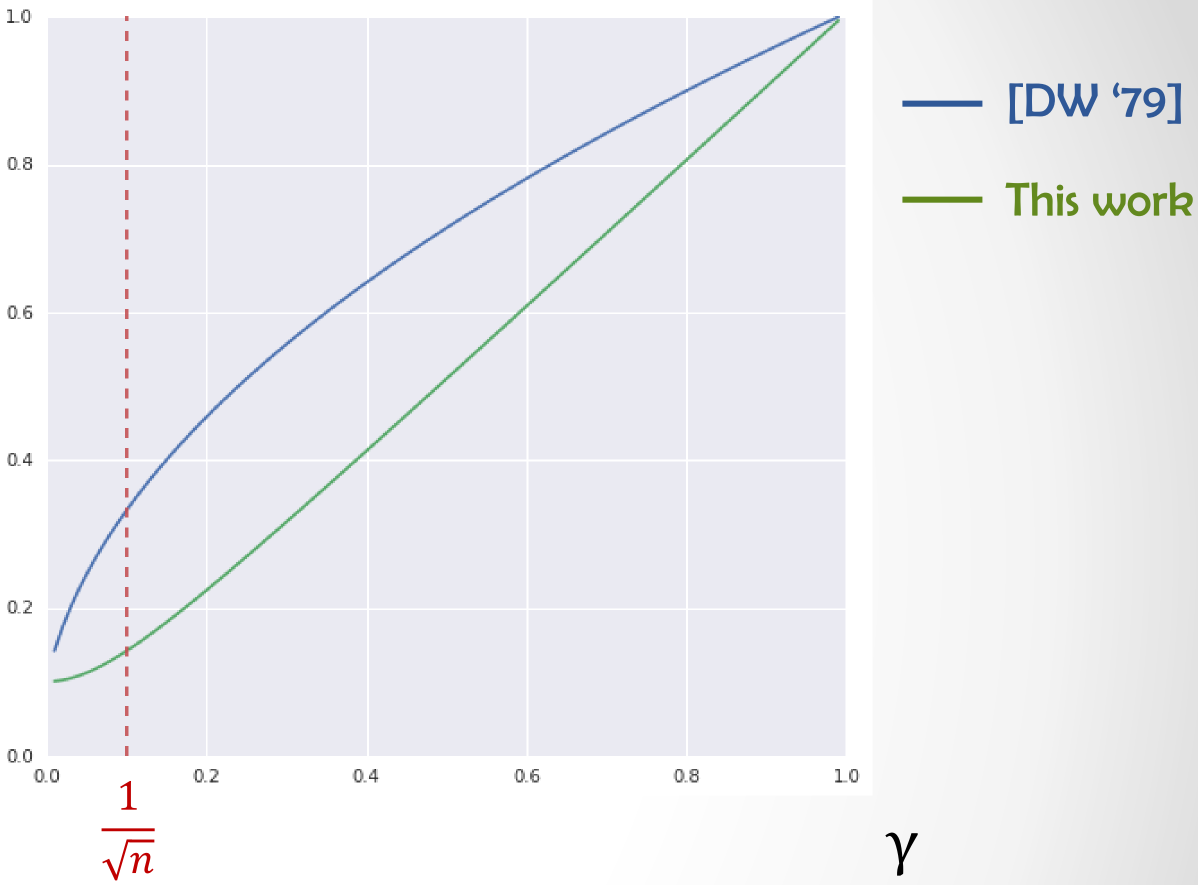
[Devroye,Wagner '79; BE '02]

$$\sqrt{\mathop{\mathbf{E}}_{S\sim P^n}\left[\Delta_S\left(\ell(A)\right)^2\right]} \leq \sqrt{\gamma}$$

Second moment

[Devroye,Wagner '79; BE '02]

$$\sqrt{\mathop{\mathbf{E}}_{S\sim P^n}\left[\Delta_S(\ell(A))^2\right]} \leq \sqrt{\gamma}$$



NEW!

$$\sqrt{\mathop{\mathbf{E}}_{S\sim P^n}\left[\Delta_S(\ell(A))^2\right]} \leq \gamma + \frac{1}{\sqrt{n}}$$

TIGHT!

There is more ….

- New proof technique
- Applications to stochastic convex optimization
- Connections to learning with differential privacy