Continuum Limits in Semi-Supervised Learning
Geometric and Topological Approaches to Data Analysis

Matthew Thorpe

Cantab Capital Institute for the Mathematics of Information
Department of Applied Mathematics and Theoretical Physics
University of Cambridge

13th June 2019
**Problem:** Given data $X_n = \{x_i\}_{i=1}^{n} \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$. 
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$. 
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ function $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.

Example by Jeff Calder.
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.

Example by Jeff Calder.
**Problem:** Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

**‘Classical’ approach:** minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$. 

Example by Jeff Calder.
1 **Problem:** Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

2 **‘Classical’ approach:** minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.

Example by Jeff Calder.
Let $X_n = \{x_i\}_{i=1}^n$, where $x_i \sim \mu \in \mathcal{P}(X)$, $X \subset \mathbb{R}^d$, be a point cloud that forms the nodes of a graph.
Let $X_n = \{x_i\}_{i=1}^n$, where $x_i \sim \mu \in \mathcal{P}(X), X \subset \mathbb{R}^d$, be a point cloud that forms the nodes of a graph. Place edges between all vertices with weights $w_{ij} = \eta_{\varepsilon}(|x_i - x_j|)$, where $\eta_{\varepsilon} = \frac{1}{\varepsilon d} \eta(\cdot/\varepsilon)$, e.g. $\eta(|x|) = \mathbb{I}_{|x| \leq 1}$. A lower bound is needed on $\varepsilon = \varepsilon_n$ in order for the graph to be connected. Theorem (Penrose, 03) If $R(n)$ is the connectivity radius of the graph then $R(n) \sim d \sum \log n$. 
Let $X_n = \{x_i\}_{i=1}^n$, where $x_i \overset{iid}{\sim} \mu \in \mathcal{P}(X)$, $X \subset \mathbb{R}^d$, be a point cloud that forms the nodes of a graph.

Place edges between all vertices with weights $w_{ij} = \eta_\varepsilon(|x_i - x_j|)$, where $\eta_\varepsilon = \frac{1}{\varepsilon^d} \eta(\cdot/\varepsilon)$, e.g. $\eta(|x|) = \mathbb{I}_{|x|\leq 1}$.

Keep edges with weights $w_{ij} > 0$. 

A lower bound is needed on $\varepsilon = \varepsilon_n$ in order for the graph to be connected.

Theorem (Penrose, 03) If $R_n$ is the connectivity radius of the graph then $R_n \sim \frac{d}{n} \log n$. 

1. Let $X_n = \{x_i\}_{i=1}^n$, where $x_i \sim \mu \in \mathcal{P}(X)$, $X \subset \mathbb{R}^d$, be a point cloud that forms the nodes of a graph.

2. Place edges between all vertices with weights $w_{ij} = \eta(\varepsilon d |x_i - x_j|)$, where $\eta(\varepsilon) = \frac{1}{\varepsilon^d} \eta(\cdot / \varepsilon)$, e.g. $\eta(|x|) = \mathbb{I}|x| \leq 1$.

3. Keep edges with weights $w_{ij} > 0$.

4. A lower bound is needed on $\varepsilon = \varepsilon_n$ in order for the graph to be connected.
Let $X_n = \{x_i\}_{i=1}^n$, where $x_i \overset{\text{iid}}{\sim} \mu \in \mathcal{P}(X)$, $X \subset \mathbb{R}^d$, be a point cloud that forms the nodes of a graph.

Place edges between all vertices with weights $w_{ij} = \eta(\varepsilon(|x_i - x_j|))$, where $\eta(\varepsilon \cdot / \varepsilon) = 1$.

Keep edges with weights $w_{ij} > 0$.

A lower bound is needed on $\varepsilon = \varepsilon_n$ in order for the graph to be connected.

**Theorem (Penrose, 03)**

If $R(n)$ is the connectivity radius of the graph then $R(n) \sim \sqrt{\frac{d \log n}{n}}$. 
Sensitivity to $\varepsilon$

(a) Data.
Sensitivity to $\varepsilon$

(a) Data.

(b) Large $\varepsilon$. 
Sensitivity to $\varepsilon$

(a) Data.  
(b) Large $\varepsilon$.  
(c) Small $\varepsilon$.  

4/18
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_2^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$. 

$p$-Laplacian Regularisation
**Problem:** Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \to \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

**‘Classical’ approach:** minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.

**$p$-Laplacian regularisation:** (Zhu, Ghahramani and Lafferty 03, and Zhou and Schölkopf 05)

\[
\mathcal{E}_n^{(p)}(f_n) = \frac{1}{\varepsilon_n^p n^2} \sum_{i,j=1}^n w_{ij} |f_n(x_i) - f_n(x_j)|^p
\]
Problem: Given data $X_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and a subset of labels $\{y_i\}_{i \in Z_n} \subset \mathbb{R}$, where $Z_n \subseteq \{1, \ldots, n\}$, find the ‘best’ $f_n : X_n \rightarrow \mathbb{R}$ such that $f_n(x_i) = y_i$ for all $i \in Z_n$.

‘Classical’ approach: minimise $\|\nabla f_n\|_{L^2}^2$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$.

$p$-Laplacian regularisation: (Zhu, Ghahramani and Lafferty 03, and Zhou and Schölkopf 05)

$$E_n^{(p)}(f_n) = \frac{1}{\varepsilon_n^p n^2} \sum_{i,j=1}^n w_{ij} |f_n(x_i) - f_n(x_j)|^p$$

Variational Problem: Minimise $E_n^{(p)}$ over $f_n : X_n \rightarrow \mathbb{R}$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$. 
The following formal calculation gives intuition as to what we should expect. We assume \( x_i \overset{iid}{\sim} \mu \) and \( \mu \) has density \( \rho \).
The following formal calculation gives intuition as to what we should expect. We assume \( x_i \overset{\text{iid}}{\sim} \mu \) and \( \mu \) has density \( \rho \).

\[
E_n^{(p)}(f) = \frac{1}{n^2 \varepsilon_n^{p+d}} \sum_{i,j=1}^{n} \eta \left( \frac{|x_i - x_j|}{\varepsilon_n} \right) |f(x_i) - f(x_j)|^p
\]
The following formal calculation gives intuition as to what we should expect. We assume $x_i \overset{iid}{\sim} \mu$ and $\mu$ has density $\rho$.

$$E_n^{(p)}(f) = \frac{1}{n^2 \varepsilon_n^{p+d}} \sum_{i,j=1}^{n} \eta \left( \frac{|x_i - x_j|}{\varepsilon_n} \right) |f(x_i) - f(x_j)|^p$$

$$\approx \frac{1}{\varepsilon_n^{p+d}} \int \int \eta \left( \frac{|x - y|}{\varepsilon_n} \right) |f(x) - f(y)|^p \rho(x) \rho(y) \, dx \, dy$$
The following formal calculation gives intuition as to what we should expect. We assume $x_i \overset{iid}{\sim} \mu$ and $\mu$ has density $\rho$.

$$E_n^{(p)}(f) = \frac{1}{n^2 \varepsilon_n^{p+d}} \sum_{i,j=1}^{n} \eta \left( \frac{|x_i - x_j|}{\varepsilon_n} \right) |f(x_i) - f(x_j)|^p$$

$$\approx \frac{1}{\varepsilon_n^{p+d}} \int \int \eta \left( \frac{|x - y|}{\varepsilon_n} \right) |f(x) - f(y)|^p \rho(x)\rho(y) \, dx \, dy$$

$$= \frac{1}{\varepsilon_n^{p}} \int \int \eta(|z|)|f(y + \varepsilon_n z) - f(y)|^p \rho(y + \varepsilon_n z)\rho(y) \, dy \, dz$$
The following formal calculation gives intuition as to what we should expect. We assume $x_i \overset{iid}{\sim} \mu$ and $\mu$ has density $\rho$.

$$E_n^{(p)}(f) = \frac{1}{n^2 \varepsilon_n^{p+d}} \sum_{i,j=1}^{n} \eta \left( \frac{|x_i - x_j|}{\varepsilon_n} \right) |f(x_i) - f(x_j)|^p$$

$$\approx \frac{1}{\varepsilon_n^{p+d}} \int \int \eta \left( \frac{|x - y|}{\varepsilon_n} \right) |f(x) - f(y)|^p \rho(x)\rho(y) \, dx \, dy$$

$$= \frac{1}{\varepsilon_n^{p}} \int \int \eta(|z|)|f(y + \varepsilon_n z) - f(y)|^p \rho(y + \varepsilon_n z) \rho(y) \, dy \, dz$$

$$\approx \int \int \eta(|z|)|\nabla f(y) \cdot z|^p \rho^2(y) \, dy \, dz$$
The following formal calculation gives intuition as to what we should expect. We assume $x_i \sim \mu$ and $\mu$ has density $\rho$.

$$
\mathcal{E}_n^{(p)}(f) = \frac{1}{n^2 \varepsilon_n^{p+d}} \sum_{i,j=1}^n \eta \left( \frac{|x_i - x_j|}{\varepsilon_n} \right) |f(x_i) - f(x_j)|^p
$$

$$
\approx \frac{1}{\varepsilon_n^{p+d}} \int \int \eta \left( \frac{|x - y|}{\varepsilon_n} \right) |f(x) - f(y)|^p \rho(x)\rho(y) \, dx \, dy
$$

$$
= \frac{1}{\varepsilon_n^p} \int \int \eta(|z|) |f(y + \varepsilon_n z) - f(y)|^p \rho(y + \varepsilon_n z) \rho(y) \, dy \, dz
$$

$$
\approx \int \int \eta(|z|) |\nabla f(y) \cdot z|^p \rho^2(y) \, dy \, dz
$$

$$
= \sigma \eta \int |\nabla f(y)|^p \rho^2(y) \, dy =: \mathcal{E}_\infty^{(p)}(f)
$$

where

$$
\sigma \eta = \int_{\mathbb{R}^d} \eta(|z|) |z_1|^p \, dz.
$$
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to be respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No!
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No! Why not?
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No! Why not?

- Problems occur when spikes have low energy.
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

**No!** Why not?

- Problems occur when spikes have low energy.
- Consider the function $f_n(x_1) = 1$ and $f_n(x_i) = 0$ for all $i \geq 2$. 
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No! Why not?

- Problems occur when spikes have low energy.
- Consider the function $f_n(x_1) = 1$ and $f_n(x_i) = 0$ for all $i \geq 2$.
- $E_n^{(p)}(f_n) = \frac{2}{\varepsilon_n^{p+d} n^2} \sum_{j=2}^{n} \eta \left( \frac{|x_1 - x_j|}{\varepsilon_n} \right)$
Passing Constraints to the Limit

If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No! Why not?

- Problems occur when spikes have low energy.
- Consider the function $f_n(x_1) = 1$ and $f_n(x_i) = 0$ for all $i \geq 2$.
- $\mathcal{E}_n^{(p)}(f_n) = \frac{2}{\varepsilon_n^{p+d}n^2} \sum_{j=2}^{n} \eta \left( \frac{|x_1 - x_j|}{\varepsilon_n} \right)$

  $= \left( \frac{2}{\varepsilon_n^{p+d}n} \right) \times \left( \frac{1}{n\varepsilon_n^d} \# \{X_n \cap B(x_1, \varepsilon_n) \} \right)$ if $\eta(|x|) = \mathbb{I}_{|x| \leq 1}$. 

If we fix the number of training data points, say \(|Z_n| = N\), we see that \(p > d\) is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

**No!** Why not?

- Problems occur when spikes have low energy.
- Consider the function \(f_n(x_1) = 1\) and \(f_n(x_i) = 0\) for all \(i \geq 2\).
- \[\mathcal{E}_n^{(p)}(f_n) = \frac{2}{\varepsilon_n^{p+d}n^2} \sum_{j=2}^{n} \eta \left( \frac{|x_1-x_j|}{\varepsilon_n} \right) = \left( \frac{2}{\varepsilon_n^{p}n} \right) \times \left( \frac{1}{n\varepsilon_n^d} \# \{X_n \cap B(x_1, \varepsilon_n) \} \right) \text{ if } \eta(|x|) = \mathbb{I}_{|x| \leq 1}.\]
- If \(\varepsilon_n^{p}n \to \infty\) then \(\mathcal{E}_n^{(p)}(f_n) \to 0\) and the spike pays no cost in the limit!
If we fix the number of training data points, say $|Z_n| = N$, we see that $p > d$ is necessary for constraints to respected (i.e. constraints for the finite data problem pass to the limit), is it sufficient?

No! Why not?

- Problems occur when spikes have low energy.
- Consider the function $f_n(x_1) = 1$ and $f_n(x_i) = 0$ for all $i \geq 2$.
- $\mathcal{E}^{(p)} \left( f_n \right) = \frac{2}{\varepsilon_n^{p+d}} n^2 \sum_{j=2}^{n} \eta \left( \frac{|x_1 - x_j|}{\varepsilon_n} \right) = \left( \frac{2}{\varepsilon_n^p n} \right) \times \left( \frac{1}{n \varepsilon_n^d} #\{X_n \cap B(x_1, \varepsilon_n)\} \right)$ if $\eta(|x|) = \mathbb{I}_{|x| \leq 1}$.
- If $\varepsilon_n^p n \to \infty$ then $\mathcal{E}^{(p)} \left( f_n \right) \to 0$ and the spike pays no cost in the limit!

This elementary example turns out to be sharp: $\varepsilon_n^p n \to \infty$ implies ill-posedness and $\varepsilon_n^p n \to 0$ implies well-posedness.
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.

The idea is to define a space which contains both discrete and continuum functions; we do this by treating the coupling $(f_n, \mu_n)$ and $(f, \mu)$ where $f_n \in L^p(\mu_n)$ and $f \in L^p(\mu)$. 

Proposition (García Trillos and Slepčev 16) If $\mu$ is absolutely continuous, then $(f_n, \mu_n) \to (f, \mu)$ in $TL^p$ if and only if $\mu_n \rightharpoonup \mu$ and there exists a sequence of maps $T_n : X \to X$ such that $(T_n)_{\#} \mu = \mu_n$, $T_n \to \text{Id}$ in $L^p$ and $\|f_n \circ T_n - f\|_{L^p(\mu)} \to 0$. 
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.

The idea is to define a space which contains both discrete and continuum functions; we do this by treating the coupling $(f_n, \mu_n)$ and $(f, \mu)$ where $f_n \in L^p(\mu_n)$ and $f \in L^p(\mu)$.

Let

$$TL^p := \{(f, \mu) : f \in L^p(\mu), \mu \in \mathcal{P}(X)\}.$$
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.

The idea is to define a space which contains both discrete and continuum functions; we do this by treating the coupling $(f_n, \mu_n)$ and $(f, \mu)$ where $f_n \in L^p(\mu_n)$ and $f \in L^p(\mu)$.

Let

$$TL^p := \{(f, \mu) : f \in L^p(\mu), \mu \in \mathcal{P}(X)\}.$$

The $TL^p$ distance is the defined for $(f, \mu), (g, \nu) \in TL^p$ by

$$d_{TL^p}((f, \mu), (g, \nu)) = \inf_{\{T : T \# \mu = \nu\}} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p \, d\mu(x)$$
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.

The idea is to define a space which contains both discrete and continuum functions; we do this by treating the coupling $(f_n, \mu_n)$ and $(f, \mu)$ where $f_n \in L^p(\mu_n)$ and $f \in L^p(\mu)$.

Let

$$TL^p := \{(f, \mu) : f \in L^p(\mu), \mu \in \mathcal{P}(X)\}.$$  

The $TL^p$ distance is the defined for $(f, \mu), (g, \nu) \in TL^p$ by

$$d_{TL^p}^p((f, \mu), (g, \nu)) = \inf_{\{T : T_{\#} \mu = \nu\}} \int_X |x - T(x)|^p + |f(x) - g(T(x))|^p \, d\mu(x).$$

Compare this to the Wasserstein distance

$$d_W^p(\mu, \nu) = \inf_{\{T : T_{\#} \mu = \nu\}} \int_X |x - T(x)|^p \, d\mu(x).$$
Before we can say more we need a topology in which we can say $f_n \to f$ where $f_n : X_n \to \mathbb{R}$ and $f : X \to \mathbb{R}$.

The idea is to define a space which contains both discrete and continuum functions; we do this by treating the coupling

Proposition (García Trillos and Slepčev 16)

If $\mu$ is absolutely continuous, then $(f_n, \mu_n) \to (f, \mu)$ in $TL^p$ if and only if $\mu_n \rightharpoonup^* \mu$ and there exists a sequence of maps $T_n : X \to X$ such that $(T_n)^# \mu = \mu_n$, $T_n \to \text{Id}$ in $L^p$ and

$$\|f_n \circ T_n - f\|_{L^p(\mu)} \to 0.$$ 

Compare this to the Wasserstein distance

$$d^p_W(\mu, \nu) = \inf_{\{T : T^# \mu = \nu\}} \int_X |x - T(x)|^p \, d\mu(x).$$
\[ \Gamma \text{-Convergence} \]

Definition

We say \( F^\infty = \Gamma \lim_n F^n \), if for all \( f \) we have

\[(i) \quad \forall f_n \to f, \quad F^\infty(f) \leq \liminf_{n \to \infty} F^n(f_n) ;\]

\[(ii) \quad \exists f_n \to f, \quad F^\infty(f) \geq \limsup_{n \to \infty} F^n(f_n).\]

Theorem

Let \( f_n \) be a sequence of almost minimizers of \( F^n \). If \( f_n \) are precompact and \( F^\infty = \Gamma \lim_n F^n \) where \( F^\infty \) is not identically \(+\infty\) then

\[ \min F^\infty = \lim_{n \to \infty} \inf F^n. \]

Furthermore any cluster point of \( f_n \) minimizes \( F^\infty \).

Green - \( F_n \), Blue - \( F_m \) for \( m > n \)
$\Gamma$-Convergence

Green - $\mathcal{F}_n$, Blue - $\mathcal{F}_m$ for $m > n$, Red - weak limit
$\Gamma$-Convergence

<table>
<thead>
<tr>
<th>Color</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>$\mathcal{F}_n$</td>
</tr>
<tr>
<td>Blue</td>
<td>$\mathcal{F}_m$ for $m &gt; n$</td>
</tr>
<tr>
<td>Red</td>
<td>Weak limit</td>
</tr>
<tr>
<td>Black</td>
<td>$\Gamma$-limit</td>
</tr>
</tbody>
</table>

**Definition**

We say $F^\infty = \Gamma\lim_n F_n$, if for all $f$ we have:

1. $\forall f_n \to f$, $F^\infty(f) \leq \lim inf_n F_n(f_n)$;
2. $\exists f_n \to f$, $F^\infty(f) \geq \lim sup_n F_n(f_n)$.

**Theorem**

Let $f_n$ be a sequence of almost minimizers of $F_n$. If $f_n$ are precompact and $F^\infty = \Gamma\lim_n F_n$ where $F^\infty$ is not identically $+\infty$ then $\min F^\infty = \lim_{n \to \infty} \inf F_n$.

Furthermore any cluster point of $f_n$ minimizes $F^\infty$. 

Green - $\mathcal{F}_n$, Blue - $\mathcal{F}_m$ for $m > n$, Red - weak limit, Black - $\Gamma$-limit.
Γ-Convergence

Definition

We say $\mathcal{F}_\infty = \Gamma\text{-}\lim_n \mathcal{F}_n$, if for all $f$ we have

(i) $\forall f_n \to f,$
$\mathcal{F}_\infty(f) \leq \liminf_{n \to \infty} \mathcal{F}_n(f_n);$

(ii) $\exists f_n \to f,$
$\mathcal{F}_\infty(f) \geq \limsup_{n \to \infty} \mathcal{F}_n(f_n).$

Green - $\mathcal{F}_n$, Blue - $\mathcal{F}_m$ for $m > n$, Red - weak limit, Black - Γ-limit.
\section*{Definition}

We say $\mathcal{F}_\infty = \Gamma\text{-}\lim_n \mathcal{F}_n$, if for all $f$ we have

\begin{enumerate}[label=(\roman*)]
  \item $\forall f_n \to f$, \quad $\mathcal{F}_\infty(f) \leq \liminf_{n \to \infty} \mathcal{F}_n (f_n)$;
  \item $\exists f_n \to f$, \quad $\mathcal{F}_\infty(f) \geq \limsup_{n \to \infty} \mathcal{F}_n (f_n)$.
\end{enumerate}

\section*{Theorem}

Let $f_n$ be a sequence of almost minimizers of $\mathcal{F}_n$. If $f_n$ are precompact and $\mathcal{F}_\infty = \Gamma\text{-}\lim_n \mathcal{F}_n$ where $\mathcal{F}_\infty$ is not identically $+\infty$ then

$$\min \mathcal{F}_\infty = \lim_{n \to \infty} \inf \mathcal{F}_n.$$ 

Furthermore any cluster point of $f_n$ minimizes $\mathcal{F}_\infty$. 

---

Green - $\mathcal{F}_n$, Blue - $\mathcal{F}_m$ for $m > n$, Red - weak limit, Black - $\Gamma$-limit.
Theorem (Slepčev and T. 17)

Let $p > 1$ and assume $\varepsilon_n$ satisfies a lower bound. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints where $Z_n = \{1, \ldots, N\}$ for $N$ fixed. Then, almost surely, $f_n$ converges in $TL^p$ along subsequences to some $f \in W^{1,p}(X)$. Furthermore,
Theorem (Slepčev and T. 17)

Let $p > 1$ and assume $\varepsilon_n$ satisfies a lower bound. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints where $Z_n = \{1, \ldots, N\}$ for $N$ fixed. Then, almost surely, $f_n$ converges in $T L^p$ along subsequences to some $f \in W^{1,p}(X)$. Furthermore,

(i) (well-posed regime) if $n \varepsilon_n^p \to 0$ as $n \to \infty$ then
Theorem (Slepčev and T. 17)

Let $p > 1$ and assume $\varepsilon_n$ satisfies a lower bound. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints where $Z_n = \{1, \ldots, N\}$ for $N$ fixed. Then, almost surely, $f_n$ converges in $TL^p$ along subsequences to some $f \in W^{1,p}(X)$. Furthermore,

(i) (well-posed regime) if $n\varepsilon_n^p \to 0$ as $n \to \infty$ then

(a) the whole sequence $f_n$ converges to $f$ locally uniformly, meaning that for any $X' \subset X$

$$\lim_{n \to \infty} \max_{\{k \leq n : x_k \in X'\}} |f(x_k) - f_n(x_k)| = 0,$$
Theorem (Slepčev and T. 17)

Let $p > 1$ and assume $\varepsilon_n$ satisfies a lower bound. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints where $Z_n = \{1, \ldots, N\}$ for $N$ fixed. Then, almost surely, $f_n$ converges in $TL^p$ along subsequences to some $f \in W^{1,p}(X)$. Furthermore,

(i) **(well-posed regime)** if $n\varepsilon_n^p \to 0$ as $n \to \infty$ then

(a) the whole sequence $f_n$ converges to $f$ locally uniformly, meaning that for any $X'$ with $\overline{X'} \subset X$

$$\lim_{n \to \infty} \max_{\{k \leq n : x_k \in X'\}} |f(x_k) - f_n(x_k)| = 0,$$

(b) $f$ is a minimizer of $\mathcal{E}_\infty^{(p)}$ subject to the constraints;
Theorem (Slepčev and T. 17)

Let $p > 1$ and assume $\varepsilon_n$ satisfies a lower bound. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints where $Z_n = \{1, \ldots, N\}$ for $N$ fixed. Then, almost surely, $f_n$ converges in $TL^p$ along subsequences to some $f \in W^{1,p}(X)$. Furthermore,

(i) (well-posed regime) if $n\varepsilon_n^p \to 0$ as $n \to \infty$ then

(a) the whole sequence $f_n$ converges to $f$ locally uniformly, meaning that for any $X'$ with $X' \subset X$

$$\lim_{n \to \infty} \max_{\{k \leq n : x_k \in X'\}} |f(x_k) - f_n(x_k)| = 0,$$

(b) $f$ is a minimizer of $\mathcal{E}_\infty^{(p)}$ subject to the constraints;

(ii) (ill-posed regime) if $n\varepsilon_n^p \to \infty$ as $n \to \infty$ then $f$ is constant.
Numerical Comparisons

(a) $p = 4$ continuum limit minimiser.

(b) $p = 4$ minimiser ($\varepsilon = 0.06$, $n = 1280$).
Development of Spikes ($p = 4$)

(a) $\varepsilon = 0.05$.  
(b) $\varepsilon = 0.1$.  
(c) $\varepsilon = 0.2$.  

12 / 18
Step 1: We show $\mathcal{E}_n^{(p)}(f_n) \approx \mathcal{E}_\infty^{(p)}(J_{\epsilon_n} \ast \tilde{f}_n)$ where $\tilde{f}_n = f_n \circ T_n$ and $J$ is a mollifier, $T_n : X \to X_n$. 
Step 1: We show $\mathcal{E}_n^{(p)}(f_n) \approx \mathcal{E}_\infty^{(p)}(J_{\varepsilon_n} \ast \tilde{f}_n)$ where $\tilde{f}_n = f_n \circ T_n$ and $J$ is a mollifier, $T_n : X \to X_n$.

Step 2: We show $\text{osc}^{(n)}_{\varepsilon}(f_n) \leq C \sqrt{\varepsilon_n^{p} \mathcal{E}_n^{(p)}(f_n)}$ where

$$\text{osc}^{(n)}_{\varepsilon}(f_n)(x_k) = \max_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z) - \min_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z).$$
Step 1: We show \( \mathcal{E}_n^{(p)}(f_n) \approx \mathcal{E}_\infty^{(p)}(J_{\varepsilon_n} \ast \tilde{f}_n) \) where \( \tilde{f}_n = f_n \circ T_n \) and \( J \) is a mollifier, \( T_n : X \to X_n \).

Step 2: We show \( \text{osc}_{\varepsilon_n}^{(n)}(f_n) \leq C \sqrt{n \varepsilon_n^{p} \mathcal{E}_n^{(p)}(f_n)} \) where

\[
\text{osc}_{\varepsilon}^{(n)}(f_n)(x_k) = \max_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z) - \min_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z).
\]

Step 3: Sobolev embedding of \( J_{\varepsilon_n} \ast \tilde{f}_n \) plus the control over oscillations is enough to infer uniform convergence:

\[
\lim_{n \to \infty} \max_{\{k \leq n : x_k \in \mathcal{X}'\}} |f(x_k) - f_n(x_k)| = 0.
\]
Intuition on the Proof for Finite Training Size

1. **Step 1:** We show $\mathcal{E}_n^{(p)}(f_n) \approx \mathcal{E}_\infty^{(p)}(J_{\varepsilon_n} \ast \tilde{f}_n)$ where $\tilde{f}_n = f_n \circ T_n$ and $J$ is a mollifier, $T_n : X \to X_n$.

2. **Step 2:** We show $\text{osc}_{\varepsilon_n}^{(p)}(f_n) \leq C \sqrt{\sum_{n}^{\infty} \mathcal{E}_n^{(p)}(f_n)}$ where

   \[
   \text{osc}_{\varepsilon}^{(p)}(f_n)(x_k) = \max_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z) - \min_{z \in B(x_k, \varepsilon) \cap X_n} f_n(z).
   \]

3. **Step 3:** Sobolev embedding of $J_{\varepsilon_n} \ast \tilde{f}_n$ plus the control over oscillations is enough to infer uniform convergence:

   \[
   \lim_{n \to \infty} \max_{\{k \leq n : x_k \in X'\}} |f(x_k) - f_n(x_k)| = 0.
   \]

4. **Step 4:** $\Gamma$-convergence of $\mathcal{E}_n^{(p)}$ to $\mathcal{E}_\infty^{(p)}$ plus a $TL^P$ compactness result is now enough to get convergence of constrained minimizers.
The reason we had asymptotic ill-posedness whenever $n^p \epsilon_n \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.
The reason we had asymptotic ill-posedness whenever $n^p \epsilon_n \rightarrow \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?
The reason we had asymptotic ill-posedness whenever $n^p_n \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).
The reason we had asymptotic ill-posedness whenever \( n \varepsilon^p_n \to \infty \) or \( p \leq d \) is because we worked in \( W^{1,p} \) with a fixed the number of constraints.

**Question:** How can we get a well-posed result when \( p = 2 \)?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take \( |Z_n| \to \infty \).
The reason we had asymptotic ill-posedness whenever $\epsilon_n^p \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take $|Z_n| \to \infty$.

**Model:** Recall $\{(x_i, y_i)\}_{i \in Z_n}$ is the training data.
The reason we had asymptotic ill-posedness whenever $n \in P \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take $|Z_n| \to \infty$.

**Model:** Recall $\{(x_i, y_i)\}_{i \in Z_n}$ is the training data.

1. Assume $\mathbb{P}(i \in Z_n) = \beta_n$. 
The reason we had asymptotic ill-posedness whenever $n \in \mathbb{P}_n \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take $|Z_n| \to \infty$.

**Model:** Recall $\{(x_i, y_i)\}_{i \in Z_n}$ is the training data.

1. Assume $\mathbb{P}(i \in Z_n) = \beta_n$.
2. If $i \in Z_n$ we set $y_i = g^\dagger(x_i)$.
The reason we had asymptotic ill-posedness whenever $n^p \rightarrow \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take $|Z_n| \rightarrow \infty$.

**Model:** Recall $\{(x_i, y_i)\}_{i \in Z_n}$ is the training data.

1. Assume $\mathbb{P}(i \in Z_n) = \beta_n$.
2. If $i \in Z_n$ we set $y_i = g^\dagger(x_i)$.

**Well-posed case:** Minimisers of $\mathcal{E}_n^{(p)}$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$ converge to $g^\dagger$. 
The reason we had asymptotic ill-posedness whenever $n \in_{n}^{p} \to \infty$ or $p \leq d$ is because we worked in $W^{1,p}$ with a fixed the number of constraints.

**Question:** How can we get a well-posed result when $p = 2$?

**One approach:** Increase the weight around labelled data (Calder and Slepčev 18).

**This talks approach:** We take $|Z_n| \to \infty$.

**Model:** Recall $\{(x_i, y_i)\}_{i \in Z_n}$ is the training data.

1. Assume $P(i \in Z_n) = \beta_n$.
2. If $i \in Z_n$ we set $y_i = g^{\dagger}(x_i)$.

**Well-posed case:** Minimisers of $\mathcal{E}_n^{(p)}$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$ converge to $g^{\dagger}$.

**Ill-posed case:** Minimisers of $\mathcal{E}_n^{(p)}$ subject to $f_n(x_i) = y_i$ for all $i \in Z_n$ converge to constants.
Ill-Posed Regime. Let $p > 1$ and assume $\varepsilon_n$ satisfies the same lower bound as before. Let $f_n$ be a sequence of minimizers of $\mathcal{E}_n^{(p)}$ satisfying the constraints. Assume $\beta_n \ll \varepsilon_n^p$ and $\frac{n\varepsilon_n^p}{\log n} \gg 1$. Then, almost surely, $\{f_n\}_{n \in \mathbb{N}}$ is precompact and any convergent subsequence converges to a constant.
**Theorem (Calder, Slepčev and T., 2019 (unpublished))**

**Ill-Posed Regime.** Let $p > 1$ and assume $\varepsilon_n$ satisfies the same lower bound as before. Let $f_n$ be a sequence of minimizers of $E_n^{(p)}$ satisfying the constraints. Assume $\beta_n \ll \varepsilon_n^p$ and $\frac{n \varepsilon_n^p}{\log n} \gg 1$. Then, almost surely, $\{f_n\}_{n \in \mathbb{N}}$ is precompact and any convergent subsequence converges to a constant.

**Well-Posed Regime.** Let $p = 2$ and assume $\varepsilon_n$ satisfies the same lower bound as before. Let $f_n$ be a sequence of minimizers of $E_n^{(p)}$ satisfying the constraints. Assume $\beta_n \gg \varepsilon_n^p$ and $\beta_n \varepsilon_n^d \gg \frac{\log n}{n}$. Then, almost surely, $f_n$ converges to $g^\dagger$ uniformly, i.e.

$$\max_{i=1, \ldots, n} |f_n(x_i) - g^\dagger(x_i)| \to 0 \quad \text{as } n \to \infty.$$
Our proof for the well-posed regime when $p = 2$ makes explicit use of the random walk interpretation of minimisers.
Our proof for the well-posed regime when $p = 2$ makes explicit use of the random walk interpretation of minimisers.

1. Let $G_n = (X_n, W)$ be the graph with edge weights $W = (w_{ij})$. 

\[ Z_{x,t} \] be the random walk on $X_n$ starting from $Z_{x,0} = x \in X_n$ and transitioning with probability $P(Z_{x,t+1} = x_k | Z_{x,t} = x_\ell) = w_{k\ell}d_{\ell}$ where $d_{\ell} = \sum_{n=k}^{\ell} w_{k\ell}$.

Proposition

Define $f_n(x) = E[g^*(Z_{x,S(x)})]$. Then $f_n$ minimises $E[2^n]$ subject to the constraints.
Our proof for the well-posed regime when $p = 2$ makes explicit use of the random walk interpretation of minimisers.

1. Let $G_n = (X_n, W)$ be the graph with edge weights $W = (w_{ij})$.
2. Let $Z^x_t$ be the random walk on $X_n$ starting from $Z^x_0 = x \in X_n$ and transitioning with probability

$$\mathbb{P}(Z^x_{t+1} = x_k \mid Z^x_t = x_\ell) = \frac{w_{k\ell}}{d_\ell}$$

where $d_\ell = \sum_{k=1}^n w_{k\ell}$. 
Random Walks on Graphs

Our proof for the well-posed regime when $p = 2$ makes explicit use of the random walk interpretation of minimisers.

1. Let $G_n = (X_n, W)$ be the graph with edge weights $W = (w_{ij})$.
2. Let $Z_t^x$ be the random walk on $X_n$ starting from $Z_0^x = x \in X_n$ and transitioning with probability

$$
\mathbb{P}(Z_{t+1}^x = x_k \mid Z_t^x = x_\ell) = \frac{w_{k\ell}}{d_\ell}
$$

where $d_\ell = \sum_{k=1}^n w_{k\ell}$.
3. Define the stopping time

$$
S(x) = \min \{ t \in \mathbb{N} : Z_s^x \in \{x_i\}_{i \in Z_n} \}.
$$
Random Walks on Graphs

Our proof for the well-posed regime when $p = 2$ makes explicit use of the random walk interpretation of minimisers.

1. Let $G_n = (X_n, W)$ be the graph with edge weights $W = (w_{ij})$.
2. Let $Z_t^x$ be the random walk on $X_n$ starting from $Z_0^x = x \in X_n$ and transitioning with probability

$$
\mathbb{P}(Z_{t+1}^x = x_k \mid Z_t^x = x_\ell) = \frac{w_{k\ell}}{d_\ell}
$$

where $d_\ell = \sum_{k=1}^n w_{k\ell}$.
3. Define the stopping time

$$
S(x) = \min\{ t \in \mathbb{N} : Z_s^x \in \{x_i\}_{i \in Z_n} \}.
$$

Proposition

Define $f_n(x) = \mathbb{E}[g^\dagger(Z_s^x)]$. Then $f_n$ minimises $\mathcal{E}_n^{(2)}$ subject to the constraints.
Intuition on the Proof when \( p = 2 \)

1. **Step 1:** We show \( \mathbb{P}[S(x_i) > k|G_n] < e^{-Ck\beta_n} \) for \( i \notin Z_n \) with probability at least \( 1 - 4ne^{-\tilde{C}n\beta_n \varepsilon_n} \).
Intuition on the Proof when $p = 2$

1. **Step 1:** We show $\mathbb{P}[S(x_i) > k \mid G_n] < e^{-Ck\beta_n}$ for $i \notin Z_n$ with probability at least $1 - 4ne^{-\tilde{c}_n\beta_n\epsilon_n^d}$.

2. **Step 2:** We show, for $M > 0$ with high probability,
   \[
   \mathbb{P} \left[ |Z_k^{x_i} - x_i| \leq cM\epsilon_n \sqrt{k} \mid G_n \right] \geq 1 - 2e^{-CM^2}.
   \]
Intuition on the Proof when $p = 2$

1. **Step 1**: We show $\mathbb{P}[S(x_i) > k| G_n] < e^{-Ck\beta_n}$ for $i \not\in Z_n$ with probability at least $1 - 4ne^{-\tilde{c}n\beta_n\varepsilon_n^d}$.

2. **Step 2**: We show, for $M > 0$ with high probability,

$$\mathbb{P}\left[|Z_k^{x_i} - x_i| \leq cM\varepsilon_n\sqrt{k}| G_n\right] \geq 1 - 2e^{-CM^2}.$$

3. **Step 3**: Combining the two previous results we show, for $\alpha > 0$ with high probability,

$$\mathbb{P}\left[|Z_{S(x_i)}^x - x_i| \leq \alpha| G_n\right] \geq 1 - 3e^{-\frac{C\alpha\sqrt{\beta_n}}{\varepsilon_n}}.$$
Intuition on the Proof when $p = 2$

1. **Step 1:** We show $\mathbb{P}[S(x_i) > k | G_n] < e^{-Ck\beta_n}$ for $i \notin Z_n$ with probability at least $1 - 4ne^{-\tilde{C}n\beta_n\varepsilon^d_n}$.

2. **Step 2:** We show, for $M > 0$ with high probability,

$$\mathbb{P} \left[ |Z_k^{x_i} - x_i| \leq cM\varepsilon_n\sqrt{k} | G_n \right] \geq 1 - 2e^{-CM^2}.$$

3. **Step 3:** Combining the two previous results we show, for $\alpha > 0$ with high probability,

$$\mathbb{P} \left[ |Z_{S(x_i)}^{x_i} - x_i| \leq \alpha | G_n \right] \geq 1 - 3e^{-\frac{C\alpha\sqrt{\beta_n}}{\varepsilon_n}}.$$

4. **Step 4:** If $g^\dagger$ is Lipschitz then we can show, with high probability, for all $\xi > 0$ and by choosing $\alpha$ optimally

$$\|f_n - g^\dagger\|_{L\infty(X_n)} \leq \alpha + C_1e^{-\frac{C_2\alpha\sqrt{\beta_n}}{\varepsilon_n}} \approx O \left( \frac{\varepsilon_n}{\sqrt{\beta_n}} \right).$$
Thank you for listening!

In theory, there is no difference between theory and practice. But in practice, there is.

— Yogi Berra