Variational methods for geometric statistical inference problems

Florian Theil

Collaborators: Neil Cade (Selex-ES), Adam Johansen (Warwick), Matt Thorpe (CMU, Pittsburgh)

Cambridge, December 2015

Objectives

- Variational tools for classification (labelling, partitioning) of data
- Rigorous characterization large data limit (classification only mildly affected by arrival of more data)
- Finite dimensional data (n data points in $\mathbb{R}^d$), infinite dimensional inference

Tracking problems

Examples of infinite dimensional inference problems.

- Tracking multiple objects
- Image segmentation

Typical structure:

- Data points $\xi_i \in X$, $i = 1 \ldots n$ (X is finite dimensional, n is large),
- Generating structure $\zeta \in Y^k$ (Banach space),
- Unknown classification $\mu(\xi) \in \{1, 2, \ldots, k\}$

Tracking of multiple targets

Random variational setting

Standard statistical approach: Maximum likelihood estimation

Mathematical task: Optimize random fitness functional

$$f_n \,:\, Y^k \mapsto \mathbb{R} \,:\, \zeta \to f_n(\zeta \mid \xi).$$

Interpretation: $p_\beta(\zeta) = \frac{1}{Z_\beta} e^{-\beta f_n}$ is a probability distribution, $\beta \gg 1$.
Mathematical challenges:

1. If $\dim(Y) = \infty$ (curves, surfaces etc) existence of minimizers $\zeta \in Y^k$ not immediate
2. As $n \to \infty$ convergence of extremizers ζ_n not clear (rich data limit).

Applications: Inverse problems, data assimilation, tracking, image analysis, ...

Definition

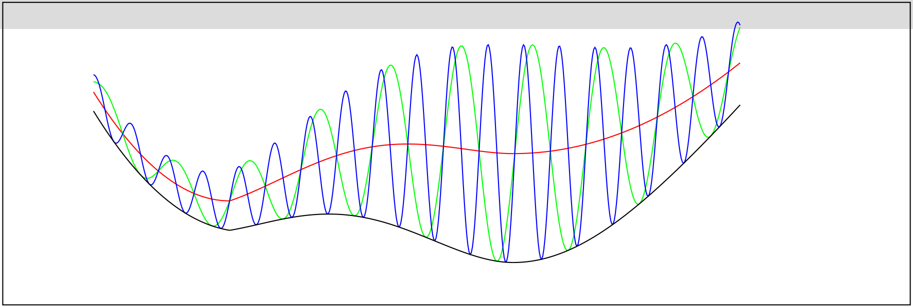
Let (Y, d_Y) be a metric space, $f, f_n : Y \to \mathbb{R} \cup \{-\infty, \infty\}$.
f is the Γ-limit of f_n if

1. (lim inf inequality) For every sequence ζ^n converging to ζ

$$f_\infty(\zeta) \leq \liminf_{n\to\infty} f_n(\zeta^n),$$

2. (Recovery sequence) There exists (ζ^n) converging to ζ such that

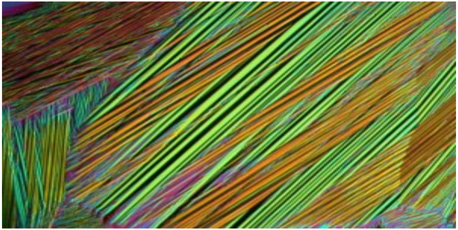
$$f_\infty(\zeta) \geq \limsup_{n\to\infty} f_n(\zeta^n).$$



Features of Γ-convergence

- Existence of f_∞ is typically not hard to show
- f_∞ is lower semi-continuous, i.e. minimizers exist even if constraints are applied
- Minimizers of f_n converge to minimizers of f_∞.
- Can deal with change of kinematics, e.g. $Y_n \subset Y$ can be finite dimensional
- Is useful in pure mathematics (tool for oscillations) and mathematical materials science (characterization of material instabilities and microstructure)

Martensitic microstructures:



Example 1: Infinite dimensional k-means method

Standard clustering: Given $\xi \in X^k$ (data) find cluster centers $\zeta \in X^k$ which minimize

$$f_n(\zeta, \mu \mid \xi) = \frac{1}{n} \sum_{i=1}^{n} |\xi_i - \zeta_{\mu(i)}|^p.$$

Generalization: $\zeta \in Y^k$

$$f_n(\zeta, \mu \mid \xi) = \frac{1}{n} \sum_{i=1}^{n} d(\xi_i, \zeta_{\mu(i)}).$$

where $d : X \times Y \to \mathbb{R}$.

Motivation: ξ could be observation, ζ is a path.

$$\begin{aligned} \xi_i &= (t_i, x_i) \in [0, T] \times \mathbb{R}^d, \quad \zeta \in H^1([0, T]), \\ d(\xi, \zeta) &= |x - \zeta(t)|. \end{aligned}$$

Observation:
(a) Regularization required, (b) Existence of $\lim_{n\to\infty} \zeta^n$ unclear
(c) Problem involves inference (ζ) and classification ($\mu(\xi) \in \{1, \dots, k\}$).

Tracking example

Regularization

We consider

$$f_n^{(\lambda)}(\zeta,\mu) = f_n(\zeta,\mu \mid \xi) + \lambda\, r(\zeta) \text{ with } r(\zeta) = \|\partial^2\zeta\|_{L^2}^2.$$

Theorem (Cade,Johansen,T.,Thorpe SIAP 2014)

- *$d(x,\cdot)$ is lower semicontinuous and ρ-integrable,*
- *ξ_i is iid with law ρ, $\mu \in \{1,\ldots,k\}$ a.e. x,*
- $f_\infty^{(\lambda)}(\zeta,\mu) = \int d(x,\zeta_\mu(t))\,\rho(\mathrm{d}x,\mathrm{d}t) + \lambda\, r(\zeta).$

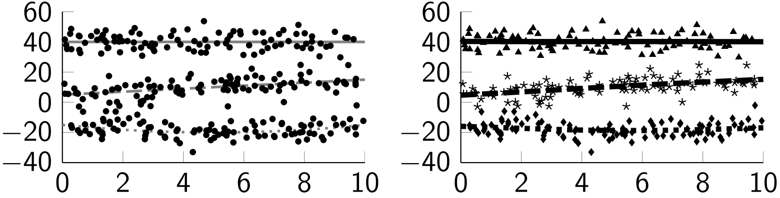
Then almost surely $f_\infty^{(\lambda)}$ is the Γ-limit of $f_n^{(\lambda)}$ in $H^1([0,T]) \times L^\infty([0,T])$.

Key results:

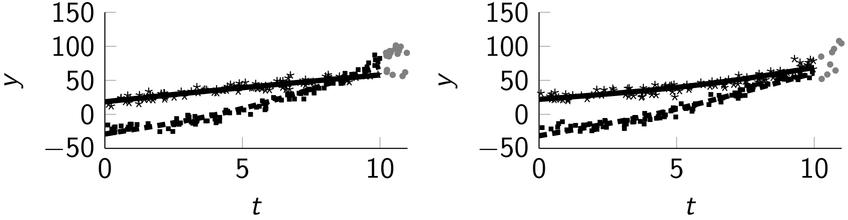
- Minimizers of $f_n^{(\lambda)}$ converge in $H^1([0,T])$ to minimizers of f_∞,
- f_n, $f_n^{(\lambda)}$ random but f_∞ deterministic.

Tracking of multiple targets

Left: Generating process, Right: Partitioned data & estimated paths



Crossing tracks



Remarks

Minimizers depend on regularization parameter λ.

Finer results:

- Convergence rate of minimizers ζ^n
- Vanishing regularization $\lambda = o(1)$ as $n \to \infty$

Convergence rate

Restriction of variational problem to non-crossing curves:

$$\Theta = \left\{ \zeta \ : \ \min_{t\in[0,T]} \min_{i\neq j} |\zeta_j(t) - \zeta_k(t)| > \theta \right\}.$$

Theorem (Johansen-Thorpe '15)

- *$\zeta^n \in \Theta$ minimizer of constrained problem,*

then there exist deterministic curves $\zeta^\infty \in \Theta$ such that

$$\lim_{n\to\infty} \zeta^n = \zeta^\infty \ \textit{a.s.},$$

and

$$\lim_{n\to\infty} n \operatorname{Var}(\|\zeta^n - \zeta^\infty\|_{L^2}) > 0 \ \textit{exists.}$$

Previous results: Pollard (Ann. Stat. '87) established a Central Limit Theorem when $\dim Y < \infty$.

Weak consistency

Truth is recoverable if $\lambda = o(1)$, $n \to \infty$.

Theorem (Johansen-Thorpe '15)

Assume data is generated by one curve $\zeta^\dagger$

$$\xi_i = \zeta^\dagger(t_i) + \varepsilon_i \text{ with } (j_i, t_i, \varepsilon_i) \text{ iid },$$

and $\lambda_n = n^{-p}$. *If*

- $p \in [0, \frac{1}{2}]$,

Then $\zeta^n(t) \to \zeta^\dagger(t)$, $n \to \infty$ *pointwise in probability.*

If $p > \frac{1}{2}$ then $\lim_{n\to\infty} \mathbb{E}\left(\|\partial^2 \zeta^n\|_{L^2}\right) = \infty$.

Alternative approach: Establish strong convergence in weaker space e.g. L^2. Requires additional spaces and destroys the Bayesian structure.

Data classification

Sometimes μ is not available (e.g. communities, medical data, etc) and the classification task μ needs additional structure.

Graphical methods

- Data $\Psi = \{\xi_1, \xi_2, \ldots\} \subset Y$.
- Classification $\mu : \Psi \to \{1, \ldots, k\}$.
- Low dimensional representation $\pi : Y \to \Omega \subset \mathbb{R}^d$
- Similarity measure $\eta : \mathbb{R}^d \to \mathbb{R}$.

Misfit of classification μ (Bertozzi,van-Gennip, Slepčev,...)

$$f_n(\mu \mid \xi) = \frac{1}{\varepsilon}\frac{1}{n^2} \sum_{\mu(\xi') \neq \mu(\xi')} \eta_\varepsilon(\pi(\xi) - \pi(\xi')),$$

with $\eta_\varepsilon = \varepsilon^{-d}\,\eta(\cdot/\varepsilon)$ and $\varepsilon \gg \left(\frac{(\log n)^2}{n}\right)^{\frac{1}{d}}$.

Soft anisotropic classification

Our generalization: $\mu \in \mathbb{R}$, η not radial:

$$
\begin{aligned}
f_n(\mu \mid \xi) \quad = \quad & \frac{1}{\varepsilon}\frac{1}{n}\sum_{\xi} V(\mathrm{dist}(\mu(\xi),\mathbb{Z})) \\
& + \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{\xi\neq\xi'} \eta_\varepsilon(\pi(\xi)-\pi(\xi')) \max\{|\mu(\xi)-\mu(\xi')|, 1\},
\end{aligned}
$$

where and $V(\mu) = 0$ if and only if $\mu = 0$.

Advantages:

- Suitable for gradient based optimization methods
- Allows unclassified data, $\mathrm{dist}(\mu^n, \mathbb{Z}) \to 0$ in measure as $n \to \infty$.

Continuum analogy: Ginzburg-Landau functional

$$F_\varepsilon[\mu] = \varepsilon \int_X |\nabla\mu(x)|^2 \,\mathrm{d}x + \frac{1}{\varepsilon}\int_X (\mu^2 - 1)^2 \,\mathrm{d}x,$$

$$F_0[\mu] = \begin{cases} \sigma \,\mathrm{per}(\{\mu = 1\}) & \text{if } \mu \in \{0,1\} \text{ a.e. } x, \\ \infty & \text{else }, \end{cases}$$

There exists *surface energy* $\sigma > 0$ s.t. F_0 is the Γ-limit of F_ε as $\varepsilon \to 0$: Modica '87, Sternberg '88. Nonlocal generalizations: Alberti-Bellettini '98.

If $\sigma = \sigma(\nu)$, then

$$F_0[\mu] = \begin{cases} \displaystyle\int_{\partial\{\mu=1\}} \sigma(\nu) \,\mathrm{d}S & \text{if } \mu \in \{0,1\} \text{ a.e. } x, \\ \infty & \text{else }, \end{cases}$$

Γ-limit of f_n

Theorem (Thorpe-T '15)

Define for $\mu : \Omega \to \mathbb{R}$ *constant on* $T^{-1}(\xi)$

$$
\begin{aligned}
f_n(\mu \mid \xi) &= \frac{1}{\varepsilon}\frac{1}{n}\sum_{\xi} V(\mathrm{dist}(\mu(\pi(\xi)), \{0,1\})) \\
&\quad + \frac{1}{\varepsilon}\frac{1}{n^2}\sum_{\xi \neq \xi'} \eta_\varepsilon(\pi(\xi) - \pi(\xi')) \max\{|\mu(\pi(\xi)) - \mu(\pi(\xi'))|, 1\},
\end{aligned}
$$

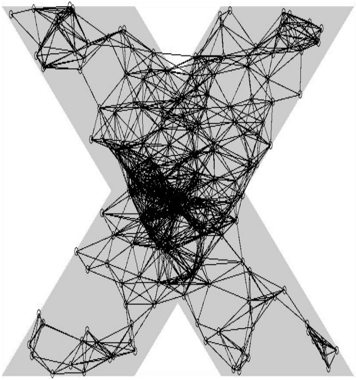
$$
\begin{aligned}
\sigma(\nu) &= \int \eta(x)\,|x \cdot \nu|\,\mathrm{d}x \text{ (convex 1-homogeneous function)}, \\
f_\infty(\mu) &= \begin{cases} \int_{\text{Jumpset of } \mu} \rho^2(x)\,\sigma(\nu)\,\mathrm{d}S & \text{if } \mu(x) \in \{0,1\} \text{ a.e. } x, \\ \infty & \text{else} \end{cases}
\end{aligned}
$$

If $\pi(\xi_i)$ *is iid with law* ρ*, then* f_∞ *is the* Γ*-limit of* f_n *a.s.*

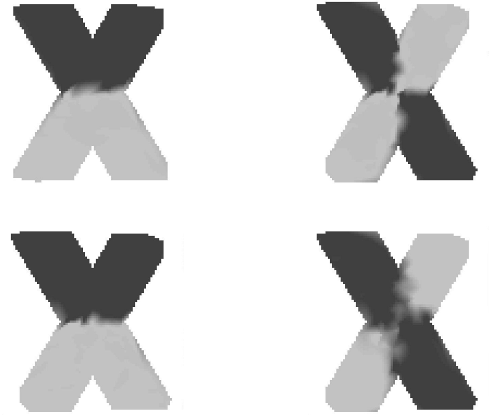
The minimization problem is unstable if σ is not convex.

Illustration of graph structure

If η has compact support draw an edge if $\eta(\pi(\xi) - \pi(\xi')) \neq 0$.



Constrained minimizers



Summary

- Γ-convergence is a useful tool to analyze statistical inference methods in the large data limit
- Can identify asymptotic regime
- Can characterize infinite dimensional objects such as paths or dividing surfaces
- Can deal with soft data classification/partition

Outlook

- Infinite dimensional data sets
- Low dimensional approximation, eg. PCA
- Posterior distribution