Large-data limit of the MBO scheme for data clustering

Tim Laux

Hausdorff Center for Mathematics & Institute for Applied Mathematics
University of Bonn

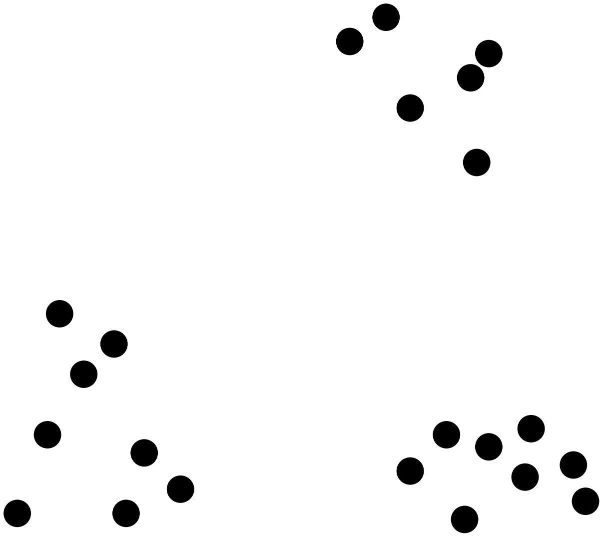
Mathematical aspects of interfaces and free boundaries

Preliminary meeting for the 81st Fujihara seminar

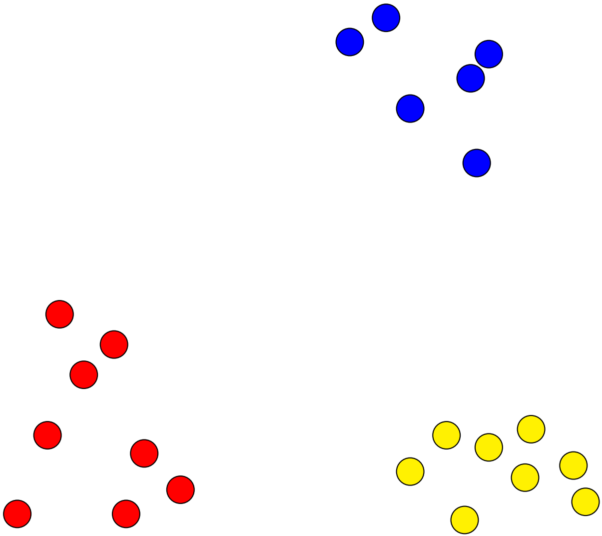
June 6, 2023

Joint work with Jona Lelmi (U Bonn)

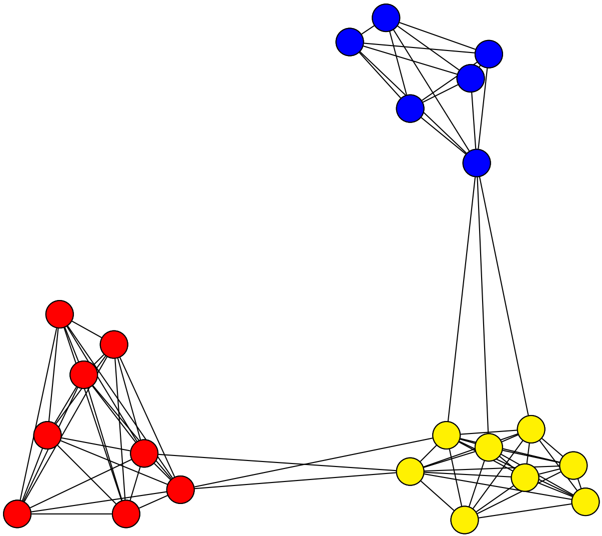
Data clustering



Data clustering



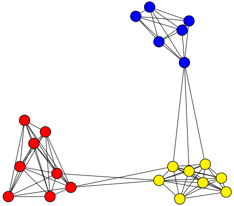
Data clustering



Similarity graph

Given data points $x_1, \dots, x_n \in \mathbb{R}^d$ w/ pairwise distances $|x_i - x_j|$.
Goal: Understand "structure", e.g. find clusters.
Define similarity graph

$$G_n = (V_n, W_n),$$

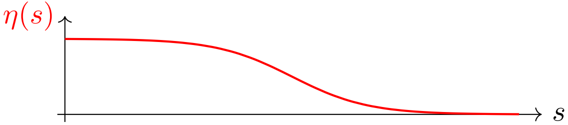


with vertices $V_n = \{x_1, \dots, x_n\}$
and edge weights $W_n = (w_{i,j})_{i,j}$

$$w_{i,j} := \eta\big(|x_i - x_j|\big), \quad i \neq j,$$
$$w_{i,i} := 0.$$

Profile η non-increasing. Typical choices are

$$\eta(s) = \exp(-s), \quad \eta(s) = \exp(-s^2), \quad \eta(s) = \mathbf{1}_{[0,1]}(s).$$



Graph Laplacian

For a weighted graph $G_n = (V, W)$, define the graph Laplacian:

$$\Delta_n = I - D^{-1}W.$$

Here, $D = \mathrm{diag}(d_1, \dots, d_n)$, where

$$d_i = \sum_{j=1}^{n} w_{i,j}$$

denotes the degree of a vertex x_i.
The Laplacian Δ_n is a non-negative self-adjoint operator w.r.t. the scalar product on $\mathcal{V} = \{u \colon V \to \mathbb{R}\}$:

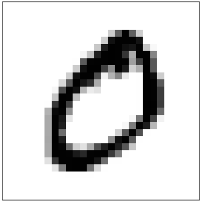
$$\langle u, v \rangle_{\mathcal{V}} = \sum_{i=1}^{n} d_i u_i v_i.$$

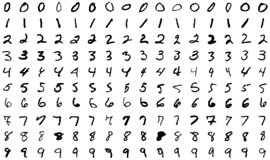
Example. On lattice $\mathbb{Z}^d$: difference quotients for Laplacian.

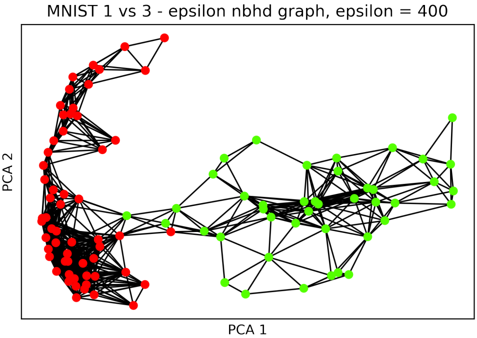
Example: MNIST data set

$\sim 50,000$ images of hand-written digits $0, 1, \ldots, 9$.

- Each image x is a (28×28)-pixels gray-scale image.
- $x \in \mathbb{R}^d$ with $d = 28^2 \approx 1000$.

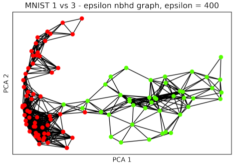






MBO scheme for data clustering

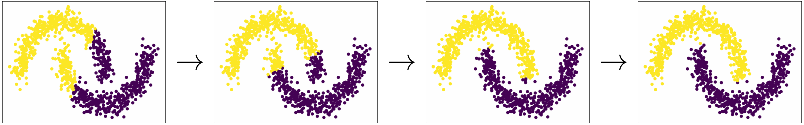
Given similarity graph $G_n = (V_n, W_n)$ with (positive) graph Laplacian Δ_n.
Goal: Find clusters $(\Omega, V_n \setminus \Omega)$.



Algorithm (Merkurjev-Kostić-Bertozzi '13; Merriman-Bence-Osher '92)

Given time step size $h > 0$ and an old cluster $\Omega \subset V_n$, obtain new cluster Ω' by:

1. Diffusion step: $\phi := e^{-h\Delta_n}\mathbf{1}_\Omega$.
2. Thresholding step: $\Omega' := \left\{\phi > \frac{1}{2}\right\}$.



Precise setting

We will work under the manifold assumption:
Assume that data is distributed on a submanifold (M, g) of $\mathbb{R}^d$.
Precisely:

- Let (M, g) be a smooth closed k-dim submanifold of $\mathbb{R}^d$.
- Let $\nu = \rho \, \mathrm{Vol}_M \in \mathcal{P}(M)$ be a prob. meas. on M with $\rho > 0$.
- Let the data points $\{x_i\}_{i \in \mathbb{N}}$ be iid random variables $\sim \nu$.

Then we consider the scaled graphs

$$G_{n,\varepsilon} = (V_n, W_{n,\varepsilon}),$$

with vertices $V_n = \{x_1, \ldots, x_n\}$, and edge weights

$$\begin{aligned} w_{i,j} &:= \frac{1}{\varepsilon^k} \eta\Big(\frac{|x_i - x_j|}{\varepsilon}\Big), \quad i \neq j, \\ w_{i,i} &:= 0. \end{aligned}$$

The graph Laplacian then is $\Delta_{n,\varepsilon} = \frac{1}{\varepsilon^2}\big(I_n - \frac{1}{n} D_{n,\varepsilon}^{-1} W_{n,\varepsilon}\big)$.

Gradient-flow structure

[Esedoğlu-Otto '15], [v.Gennip-Guillen-Osting-Bertozzi '14]

Each step decreases the graph heat content energy

$$E_h^{n,\varepsilon}(\chi) = \frac{1}{\sqrt{h}} \langle 1 - \chi, e^{-h\Delta_{n,\varepsilon}} \chi \rangle_{\mathcal{V}_n}.$$

More precisely, each step of MBO is equivalent to

$$\chi^\ell \in \arg\min_{\chi} \Big\{ E_h^{n,\varepsilon}(\chi) - E_h^{n,\varepsilon}(\chi - \chi^{\ell-1}) \Big\}.$$

Question: What is the effective behavior of (local) minimizers of the energy $E_h^{n,\varepsilon}$ when

$$n \gg 1, \quad \varepsilon \ll 1, \quad h \ll 1\ ?$$

$$\begin{tikzcd}
E_h^{n,\varepsilon_n} \arrow[r, "n \to \infty"] \arrow[dr, "?"'] & E_h \arrow[d, "h \downarrow 0"] \\
& E
\end{tikzcd}$$

Known: ε and h have to be sufficiently large!
Otherwise, scheme gets pinned/frozen [Bertozzi et al. '14].

1st main result: large data limit $n \to \infty$, $\varepsilon \to 0$

Theorem 1 (L., Lelmi '21)

In the regime $\log(n^{1/k})\left(\frac{1}{n^{1/k}}\right)^{\frac{k}{k+2}} \ll \varepsilon_n \ll 1$ *it holds almost surely*

$$E_h^{n,\varepsilon_n} \overset{\Gamma(\textit{weak-}TL^2)}{\longrightarrow} E_h,$$

where

$$E_h(u) = \frac{1}{\sqrt{h}} \int_M (1-u) e^{-\kappa h \Delta_{M,\rho^2}} u \, \rho^2 d\,\mathrm{Vol}_M$$

for $u \colon M \to [0,1]$; *and* $E_h(u) = +\infty$ *otherwise.*

Here, $\kappa = \kappa(\eta)$ and the Laplacian Δ_{M,ρ^2} on (M, g, ρ^2) is given by

$$\Delta_{M,\rho^2} u = -\frac{1}{\rho^2} \mathrm{div} \rho^2 \nabla u.$$

2nd main result: sharp interface limit $h \to 0$

Theorem 2 (L., Lelmi '21)
As $h \to 0$ it holds

$$E_h \overset{\Gamma(\textit{strong-}L^1)}{\longrightarrow} E,$$

where

$$E(\chi) = \frac{1}{\sqrt{\pi}} |D\chi|_{\rho^2}(M)$$

for $\chi \colon M \to \{0,1\}$; and $E(\chi) = +\infty$ otherwise.

We may also write

$$E(\chi) = \frac{1}{\sqrt{\pi}} \int_{\partial^* \{\chi = 1\}} \rho^2 d\mathcal{H}^{k-1}.$$

This means, (local) minimizers of E are minimal surfaces in the weighted manifold (M, g, ρ^2).

Recap of main results

Theorem 1 (L., Lelmi '21)
Large-data limit, or,
discrete to non-local.

Theorem 2 (L., Lelmi '21)
Sharp-interface limit, or,
non-local to local.

$$\begin{tikzcd} E_h^{n,\varepsilon_n} \arrow[r, "\text{Thm 1}"] \arrow[dr] & E_h \arrow[d, "\text{Thm 2}"] \\ & E \end{tikzcd}$$

Generality of results:

- arbitrary number of clusters (for simplicity in this talk only 2)
- different "surface tension" between different labels
- external forcing/drift f:
  - efficient algorithm: change threshold value from $\frac{1}{2}$ to $\frac{1}{2} - \frac{\sqrt{h}}{\sqrt{\pi}} f$
  - simple analysis: additional term in energy $-\frac{1}{\sqrt{\pi}} \langle f, \chi \rangle_{\nu_n}$
- Other variants of graph Laplacians

A word on TL^2

As $u_n \colon V_n \subset M \to \mathbb{R}$ and $u \colon M \to \mathbb{R}$ live in different spaces, we cannot immediately compare u_n to u.

Helpful analytical tool: optimal transport.
Let $T_n \colon M \to V_n$ be "nice" optimal transport maps.
[García Trillos, Slepčev '16]: In this case we have

$$u_n \to u \quad \text{in } TL^2 \quad \Leftrightarrow \quad u_n \circ T_n \to u \quad \text{in } L^2(M).$$

For our case, we introduce weak TL^2 convergence. We have

$$u_n \rightharpoonup u \text{ weakly in } TL^2 \quad \Leftrightarrow \quad u_n \circ T_n \rightharpoonup u \text{ weakly in } L^2.$$

This is the natural space, because our sequence u_n will be only pre-compact in the weak topology of TL^2.

Elements of proof: large-data limit

In fact we prove a stronger statement:

Proposition

In the setting of Theorem 1, let $u_n \in \mathcal{V}_n$ be a sequence of functions converging weakly to $u \in L^2(M)$ in TL^2, then for every $t > 0$ we have

$$\lim_{n \to +\infty} e^{-t\Delta_{n,\varepsilon_n}} u_n = e^{-t\Delta_{M,\rho^2}} u \text{ strongly in } TL^2.$$

Idea [De Giorgi; Sandier–Serfaty; AGS; ...]: solutions to diffusion/heat equation are characterized by optimal energy dissipation relation

$$\tfrac{1}{2}\|\nabla_n v_n(t)\|^2_{\mathcal{V}_n} + \tfrac{1}{2}\int_0^t \left(\|\Delta_n v_n(s)\|^2_{\mathcal{V}_n} + \|\tfrac{d}{ds} v_n(s)\|^2_{\mathcal{V}_n}\right) ds \leq \tfrac{1}{2}\|\nabla_n u_n\|^2_{\mathcal{V}_n},$$

in which we can pass to the limit by lower semi-continuity, which identifies the limit and gives regularity.

Elements of proof: large-data limit (cont'd)

As a direct consequence, we have the following sharper version of the Γ-convergence:

Corollary

In the setting of the proposition, $\lim_{n\to\infty} E_h^{n,\varepsilon_n}(u_n) = E_h(u)$.

We also give a positive answer to a question of [Bertozzi et al. '14]:

Corollary

If additionally $\bar{u}_n \rightharpoonup \bar{u}$*, then the min. mov. functional converges:*

$$\lim_{n\to\infty} \left(E_h^{n,\varepsilon_n}(u_n) - E_h^{n,\varepsilon_n}(u_n - \bar{u}_n) \right) = E_h(u) - E_h(u - \bar{u}).$$

In particular,

MBO on $(V_n, W_{n,\varepsilon_n})$ *converges to MBO on* (M, g, ρ^2)*.*

Background on sharp-interface limit $h \downarrow 0$

To study the asymptotic behavior of

$$E_h(u) = \frac{1}{\sqrt{h}} \int_M (1-u) e^{-h\Delta_{M,\rho^2}} u \, \rho^2 d\,\mathrm{Vol}_M$$

is a challenging mathematical problem, see e.g.

- [Alberti, Bellettini '98]
- [Miranda, Pallara, Paronetto, Preunkert '07a]
- [Esedoğlu, Otto '15]

Much easier question [MPPP '07b]: convergence of

$$F_h(u) = \int_M \left|\nabla e^{-h\Delta_{M,\rho^2}} u\right| \rho^2 \, d\,\mathrm{Vol}_M \, .$$

Also sparks interest in the context of

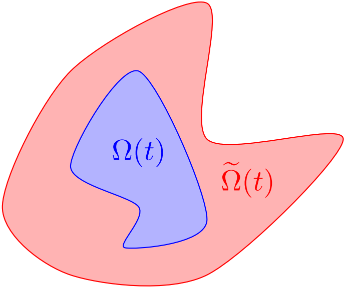
- Metric geometry [De Ponti, Mondino '20]
- Sub-Riemannian geometry [Agrachev, Rossi, Rizzi '20, '21, '22]

MBO & comparison principle

Comparison principle for MCF

Let $\partial\Omega(t)$, $\partial\widetilde{\Omega}(t) \subset (M, g, \rho^2)$ move by mean curvature flow. Then for any $t > 0$:

$$\Omega(0) \subset \widetilde{\Omega}(0) \Rightarrow \Omega(t) \subset \widetilde{\Omega}(t).$$



[Chen, Giga, Goto '91], [Evans, Spruck '91]

Comparison principle for MBO

Let Ω^ℓ, $\widetilde{\Omega}^\ell$ be obtained by the MBO scheme. Then for any $\ell \in \mathbb{N}$:

$$\Omega^\ell \subset \widetilde{\Omega}^\ell \Rightarrow \Omega^{\ell+1} \subset \widetilde{\Omega}^{\ell+1}.$$

Proof: Write $\chi = \chi_{\Omega^\ell}$, $\tilde{\chi} = \chi_{\tilde{\Omega}^\ell}$. Then

$$\chi \leq \tilde{\chi} \implies e^{-h\Delta}\chi \leq e^{-h\Delta}\tilde{\chi} \implies H(e^{-h\Delta}\chi) \leq H(e^{-h\Delta}\tilde{\chi}). \quad \square$$

Abstract convergence result

Theorem 3 (L., Lelmi '22)
Suppose that

(i) *For all* $u, v \in \mathcal{V}_n$

$$u \leq v \Rightarrow S_n(h_n)u \leq S_n(h_n)v + \|(u,v)\|_{V_n,\infty} O(h_n^{\frac{3}{2}}).$$

(ii) *For all* $f \in C^\infty(M)$

$$\|S_n(h_n)f - e^{-h_n \Delta_{\rho^2}} f\|_{\infty,V_n} = \|f\|_\infty o(h_n^{\frac{1}{2}}) + \|\nabla f\|_\infty O(h_n^{\frac{3}{2}}).$$

(iii) $\|S_n(h_n)1 - 1\|_{\infty,V_n} = O(h_n^{\frac{3}{2}})$.

Then the MBO scheme with $S_n(t)$ *instead of* $e^{-t\Delta_n}$ *converges to the unique viscosity solution of MCF on* (M, g, ρ^2).

Assumption (iii) is only needed to check initial conditions.
(Seems to have been overlooked in the past.)

Application to random geometric graphs

Theorem 4 (L., Lelmi '22)
Let G_n be a random geometric graph.
Assume that $q, \alpha, \beta > 0$ are suitably chosen and

(i) $h_n \gg (\log(n))^{-\alpha}$.

(ii) $\left(\frac{\log(n)}{n^{1/k}}\right)^{\frac{k}{k+4}} \lesssim \varepsilon_n \ll (\log(n))^{-\beta}$.

(iii) $K_n \geq (\log(n))^q$.

(iv) *The eigenvalues of* Δ_{ρ^2} *satisfy* $\inf_{i\in\mathbb{N}}(\lambda_i - \lambda_{i-1}) > 0$.

Then the operators $e^{-t\Delta_n}$ *and* $e^{-t\Delta_n} P_{\langle \psi_1^n, \ldots, \psi_{K_n}^n \rangle}$ *satisfy conditions (i), (ii), and (iii) in Theorem 3 (prev. slide) with probability greater than*

$$1 - C\varepsilon_n^{-6k} \exp\left(-\frac{n\varepsilon_n^{k+4}}{C}\right) - Cn \exp\left(-\frac{n}{CK_n^2}\right).$$

Convergence of MBO to MCF

Combining Theorems 3 and 4 from the previous slides:

Corollary

In the scaling regime of Theorem 4, the MBO scheme (with or without frequency cut-off) on random geometric graphs converges to the mean curvature flow on (M, g, ρ^2) in probability.

Remark

- Can apply Theorem 3 on $G_n = \varepsilon_n \mathbb{Z}^2$ to recover convergence result of [Misiats, Yip '16] + initial conditions.
- Mean curvature flow in (M, g, ρ^2) takes the form

$$V = -\frac{1}{\rho^2}\mathrm{div}(\rho^2 \nu) = -H - \nu \cdot \nabla \log \rho^2.$$

This PDE is driven by surface tension + density.

Ingredient: new heat kernel estimates

More precisely, the frequency cut-off at K can be expressed as

$$e^{-t\Delta_n} P_{\langle \psi_1^n, \ldots, \psi_K^n \rangle} u(x) = \sum_{y \in V_n} H_n^K(t,x,y) u(y),$$

where H_ε^K is the truncated heat kernel

$$H_n^K(t,x,y) = \sum_{i=1}^{K} e^{-t\lambda_i^n} \psi_i^n(x) \psi_i^n(y) \frac{d_n(y)}{n}.$$

Lemma

In the setting of Theorem 4, for $n \gg 1$*,*

$$\max_{x,y \in V_n} \left| H_n^{K_n}(h_n,x,y) - \frac{\rho(y)}{n} H(h_n,x,y) \right| = o\left(\frac{\sqrt{h_n}}{n} \right).$$

This improves the recent work [Dunson, Wu, Wu: ACHA '21] on heat kernel estimates.

Summary

- Intuitively, MBO finds local minimizers of the graph heat content energy $E_h^{n,\varepsilon}$.
- Our results show: local minimizers are close to local minimizers of the area functional E on (M, g, ρ^2).
- The energy landscape is highly non-convex, hence we expect to have many local minimizers.
- Our results show: MBO-dynamics converge to (viscosity solution of) mean curvature flow in (M, g, ρ^2).

- T.L., Jona Lelmi: Large data limit of the MBO scheme for data clustering: Γ-convergence of the thresholding energies arXiv:2112.06737 [math.AP]
- T.L., Jona Lelmi: Large data limit of the MBO scheme for data clustering: convergence of the dynamics arXiv:arxiv:2209.05837 [math.AP]

Summary

- Intuitively, MBO finds local minimizers of the graph heat content energy $E_h^{n,\varepsilon}$.
- Our results show: local minimizers are close to local minimizers of the area functional E on (M, g, ρ^2).
- The energy landscape is highly non-convex, hence we expect to have many local minimizers.
- Our results show: MBO-dynamics converge to (viscosity solution of) mean curvature flow in (M, g, ρ^2).

- T.L., Jona Lelmi: Large data limit of the MBO scheme for data clustering: Γ-convergence of the thresholding energies arXiv:2112.06737 [math.AP]
- T.L., Jona Lelmi: Large data limit of the MBO scheme for data clustering: convergence of the dynamics arXiv:arxiv:2209.05837 [math.AP]

Thank you for your attention!