Score based learning in scientific computing

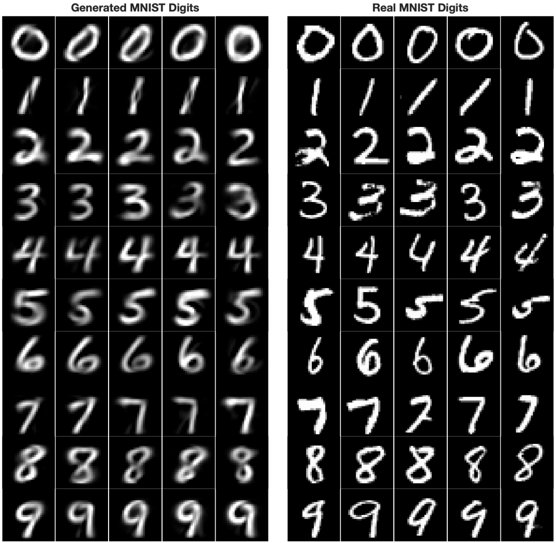
Yuehaw Khoo, Yifan Peng, **Mathias Oster**

May 20th 2025
Seminar of the Mathematical Modelling Group at OsloMet

igpm | RWTH AACHEN UNIVERSITY

SFB 1481
Sparsity and
Singular Structures

Generating new handwritten digits



Generating new samples from given one

Given: samples $x_i \sim \rho_0$ for some density ρ_0

Aim: Create new samples y_i such that $y_i \sim \rho_0$

Problem: ρ_0 is unknown

Idea:

- deform the samples into a new, simpler distribution
- create new samples from easier distribution
- reverse the deformation for the new samples

Deforming densities

Consider the SDE

$$dX_t = -V(X_t)dt + dW_t, \quad X_0 \sim \rho_0$$

with some

potential $V(x)$

and

Brownian motion W_t

Then

$$\rho_t = \mathsf{Law}(X_t)$$

fulfills

$$\partial_t \rho_t + \Delta \rho_t + \operatorname{div}(\rho_t \nabla V) = 0$$

and

$$\rho_\infty(x) \eqsim e^{-V(x)}$$

Examples for the potential

Examples:

- $V(x) = \alpha\|x\|^2 \rightsquigarrow$ Ornstein-Uhlenbeck process with ρ_∞ Gaussian
- $V(x) = 0$ and periodic boundary $\rightsquigarrow$ periodic Brownian motion with ρ_∞ uniform
- $V(x) = \sum_i f_i(x_i) \rightsquigarrow$ overdamped Langevin with ρ_∞ seperable Gibbs

The score and reverse time dynamics

One important quantity is the score

$$s(t,x) = \nabla \log \rho_t(x)$$

Reverse time dynamics

$$dY_{T-t} = [\nabla V(Y_{T-t}) + s(T-t, Y_{T-t})]\, dt + dW_{T-t}$$

with

$$Y_T \sim \exp(-V)$$

Observation: $\mathrm{Law}(Y_0) \approx \rho_0$ (in dependence of T)

Learning the score

Recover the score:

$$\min_{s(t,x)} \int_0^T \int \frac{1}{2}\|s(t,x) - \nabla \log \rho_t(x)\|_2^2 \rho_t(x) dx dt$$

Problem: one needs pointwise information of $\nabla \log \rho_t$

Idea: expand and partial integration

Expansion:

$$\begin{aligned}&\int_0^T \int \frac{1}{2}\|s(t,x) - \nabla \log \rho_t(x)\|_2^2 \rho_t(x) dx dt \\ =&\frac{1}{2}\int_0^T \int \|s(t,x)\|_2^2 \rho_t(x) dx dt - 2\int_0^T \int \langle s(t,x), \nabla \log \rho_t(x)\rangle \rho_t(x) dx dt \\ &+ \int_0^T \int \|\nabla \log \rho_t(x)\|_2^2 \rho_t(x) dx dt\end{aligned}$$

Alternative cost function

Integration by parts

$$\begin{aligned}
&\int \langle s(t,x), \nabla \log \rho_t(x) \rangle \rho_t(x) dx \\
\underbrace{=}_{\substack{\rho \nabla \log \rho \\ = \rho \frac{\nabla \rho}{\rho}}} &\int \langle s(t,x), \nabla \rho_t(x) \rangle dx \\
= - &\int \operatorname{div}(s(t,x)) \rho_t dx
\end{aligned}$$

Computable cost functional:

$$\min_{s(t,x)} \int_0^T \int \frac{1}{2} \|s(t,x)\|_2^2 \rho_t(x) + \operatorname{div}(s(t,x)) \rho_t(x) dx dt$$

Advantage: only samples from ρ_t needed

"Classical" score learning – I

Tasks:

- Deform original samples by solving the SDE
- Collect the deformed samples to recover the score
- Reverse the dynamics to create new samples

Solving SDE: use **Euler-Mayurana** for **Ornstein-Uhlenbeck** to create time series $x_i(t_j)$ from $x_i \sim \rho_0$

"Classical" score learning – II

Recover the score: Minimize

$$\sum_j \sum_i \mathrm{NN}^2(t_j, x_i(t_j)) + \mathrm{div}(\mathrm{NN}(t_j, x_i(t_j)))$$

over the set of neural networks NN by e.g. stochastic gradient descent

Spectral approach – I

Associated to an SDE is a generator

$$\mathcal{L}f = \Delta f - \nabla V(x) \cdot \nabla f$$

with

$$\partial_t \rho_t - \mathcal{L}^\dagger \rho_t = 0$$

Consider the eigenbasis φ_i of $\mathcal{L}$ with eigenvalues λ_i.

Then it holds

$$\begin{aligned}\int \varphi_i(x)\rho_t(x)dx &= \int \varphi_i(x) \exp(t\mathcal{L}^\dagger)\rho_0(x)dx \\ &= \int \rho_0(x) \exp(t\mathcal{L})\varphi_i(x)dx \\ &= e^{\lambda_i t} \int \varphi_i(x)\rho_0(x)dx\end{aligned}$$

$\rightsquigarrow$ no SDE needed anymore

Spectral approach – II

Examples:

- Ornstein-Uhlenbeck process: $\rightsquigarrow \varphi_i$ are Hermite polynomials
- periodic Brownian motion: $\rightsquigarrow \varphi_i$ are Fourier bases

Spectral approach – III

Tasks:

- Choose potential V such that one knows (or can approximate) the associated eigenfunctions
- calculate the "moments"

$$\int \varphi_i(x)\rho_t(x)dx = e^{\lambda_i t}\int \varphi_i(x)\rho_0(x)dx$$

- expand $s(t,x) = \sum_{i=1}^N c_i(t)\varphi_i(x)$
- calculate

$$\varphi_i(x)\varphi_j(x) = \sum_k \mu_{ij}^k \varphi_k(x) \text{ as well as } \frac{d}{dx}\varphi_i = \sum_k \nu_i^k \varphi_k(x)$$

- given a time discretization, minimize

$$\frac{1}{2}\sum_{i,j} c_i(t_k)c_j(t_k)\int \langle \varphi_i, \varphi_j\rangle_2^2 \rho_{t_k}(x)dx + \sum_i c_i(t_k)\int \operatorname{div}(\varphi_i(x))\rho_{t_k}(x)dx$$

Computational tricks

- exploit fast decay of higher order eigenfunctions over time to expand product and derivative in "few" basis functions
- for high-dimensional case use a band-limited cluster ansatz, i.e.:
  - $$\mathcal{B}_{j,d} = \Big\{ \phi^{(k_1)}_{n_{k_1}}(x_{k_1}) \cdots \phi^{(k_j)}_{n_{k_j}}(x_{k_j}) \Big| \; 1 \le k_1 < \cdots < k_j \le d, \; 1 \le n_{k_1}, \ldots, n_{k_j} \le n \Big\}$$
  - $$\mathcal{B}^{\text{local}}_{2,d} = \Big\{ \phi^{(k)}_{n_k}(x_k)\, \phi^{(k')}_{n_{k'}}(x_{k'}) \Big| \; 1 \le k < k' \le d, \; k' \le k + d_{\text{b}}, \; 1 \le n_k, n_{k'} \le n \Big\}$$

Mean field approximation

the linear ansatz works well in a "perturbative regime", i.e.
$\rho_0(x) = \rho_\infty(x)\,(1 + \delta \mathrm{pert}(x))$

Idea: Calculate mean-field approximation to ρ_0

- Calculate (empirical) moments $m_{i,j} = \int x_i^j \rho_0(x) dx$
- solve maximal entropy program

$$\max_{\rho_i} \int \rho_i(x_i) \log \rho_i(x_i)\, dx_i$$

$$\text{subject to} \quad \int x_i^j \rho_i(x_i)\, dx_i = \mu_{i,j}, \quad \text{for } 0 \leq j \leq m.$$

- then $\rho_i(x_i) = \exp(-\sum_{j=0}^m \nu_j^{(i)} x_i^j - 1)$ where ν are the Lagrange multiplier
- given $V(x) = \sum_{j=0}^m \nu_j^{(i)} x_i^j + 1$, approximate φ_i associated to V

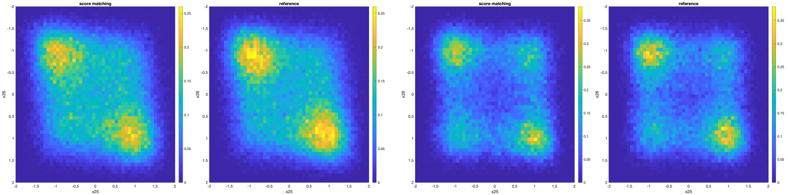
Example: Ginsburg-Landau

$$\rho_0 = \frac{1}{Z}\exp(-\beta_{\mathrm{GL}} V_{\mathrm{GL}}(x))$$

with

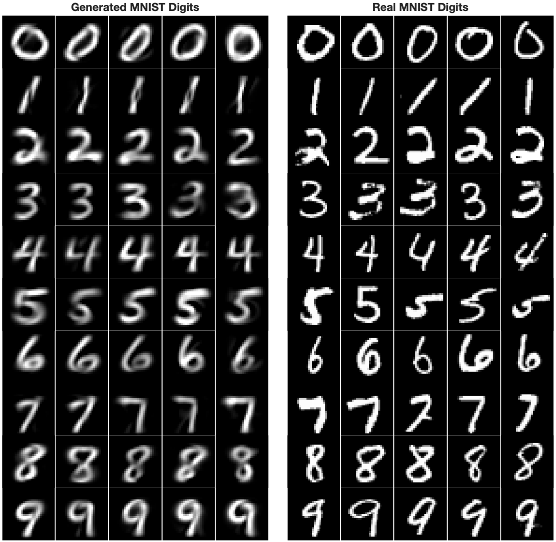
$$V_{\mathrm{GL}}(x) = \sum_{i=1}^{d} \left(\frac{\lambda_{\mathrm{GL}}}{2} \left(\frac{x_i - x_{i-1}}{h} \right)^2 + \frac{1}{4\lambda_{\mathrm{GL}}} (1 - x_i^2)^2 \right)$$

	Mean-field	Fourier basis	Hermite polynomial
strong interaction	0.0512	0.0575	0.0588
weak interaction	0.0657	0.0652	0.0685



Example: MNIST

Trick: PCA: 728 dim $\to$ 10 dim



Summary

- Choose (appropriate) potential V
- calculate the "moments"

$$\int \varphi_i(x)\rho_t(x)dx = e^{\lambda_i t} \int \varphi_i(x)\rho_0(x)dx$$

- expand $s(t,x) = \sum_{i=1}^{N} c_i(t)\varphi_i(x)$
- given a time discretization, minimize

$$\frac{1}{2}\sum_{i,j} c_i(t_k)c_j(t_k) \int \langle \varphi_i, \varphi_j \rangle_2^2 \rho_t(x)dx + \sum_i c_i(t_k) \int \operatorname{div}(\varphi_i(x))\rho_t(x)dx$$

Thank you for your attention