Progress Report

Objective of Score matching on non-negative data

Kamiya Takuto

July 14, 2025

Outline

- 1. Review of Score Matching
- 2. Score Matching on Non-negative Data
  - (1) Weighted Score Matching
  - (2) Proposed Method: Boundary Term Approximation

Score Function

Score Function

$$\nabla_x \log p(x) = \left(\frac{\partial}{\partial x_1} \log p(x), \ldots, \frac{\partial}{\partial x_m} \log p(x) \right)^\top \quad (x \in \mathbb{R}^m)$$

Key property: The score function is *invariant under normalization*.

Let $p(x) = \frac{\tilde{p}(x)}{Z}$, where $Z = \int \tilde{p}(x)dx$ is the partition function. Then:

$$\nabla_x \log p(x) = \nabla_x \log \tilde{p}(x) \quad \text{since } \log p(x) = \log \tilde{p}(x) - \log Z$$

The normalization constant Z disappears under differentiation.

Objective of Score Matching

Score Matching Loss

$$J(p) = \int_{\mathbb{R}^m} \|\nabla_x \log p(x) - \nabla_x \log q(x)\|^2 q(x) dx$$

This loss measures the discrepancy between the score functions of the model p and the data distribution q, using an expectation under $q(x)$.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, J(p_\theta, q) = \underset{\theta}{\operatorname{argmin}} \int_{\mathbb{R}^m} \|\nabla_x \log p_\theta(x) - \nabla_x \log q(x)\|^2 q(x) dx$$

We estimate the model parameter θ by minimizing the score-matching objective.

Approach to Reformulating the Objective

We begin by expressing the score matching loss as:

$$J = \int_{\mathbb{R}^m} F(x)\, q(x)\, dx \quad \Longleftrightarrow \quad J = \mathbb{E}_{x\sim q}[F(x)]$$

This suggests we can approximate J using the empirical average:

$$\mathbb{E}_{x\sim q}[F(x)] \approx \frac{1}{n}\sum_{i=1}^{n} F(x_i)$$

However, in its current form, $F(x)$ still depends on $q(x)$, which is unknown.
Goal: Reformulate the objective so that

$$J = \int_{\mathbb{R}^m} F'(x)\, q(x)\, dx$$

where $F'(x)$ no longer involves $q(x)$. This enables unbiased empirical estimation.

Fisher Divergence as a Decomposed Divergence

- Expanding the squared norm in $J(p)$ gives:

$$\int \|\nabla_x \log p(x)\|^2 \, q(x) \, dx - 2 \int \nabla_x \log p(x)^\top \nabla_x \log q(x) \, q(x) \, dx + \int \|\nabla_x \log q(x)\|^2 \, q(x) \, dx$$

- This can be written in the form:

$$J(p) = g(q) + d(p, q)$$

where

$$g(q) = \int \|\nabla_x \log q(x)\|^2 \, q(x) \, dx,$$

$$d(p, q) = \int \|\nabla_x \log p(x)\|^2 \, q(x) \, dx - 2 \int \nabla_x \log p(x)^\top \nabla_x \log q(x) \, q(x) \, dx$$

Structure of $J(p) = d(p, q)$

Since $g(q)$ does not depend on the model p, minimizing $J(p)$ is equivalent to minimizing $d(p, q)$. **Thus, we redefine $J(p) := d(p, q)$ for the remainder of this presentation.**:

$$J(p) = \int \|\nabla_x \log p(x)\|^2 \, q(x) \, dx - 2 \int \nabla_x \log p(x)^\top \nabla_x \log q(x) \, q(x) \, dx$$

This objective consists of two terms:

- The first term depends only on the model p, via its score function.
- The second term couples the model score $\nabla_x \log p(x)$ with the data score $\nabla_x \log q(x)$, and is problematic because $\nabla_x \log q(x)$ is unknown.

Goal: Eliminate the dependence on the unknown $\nabla_x \log q(x)$ using integration by parts.

Step 1: Substituting the Score Function

We begin with the problematic term:

$$-2 \int \nabla_x \log p(x)^\top \nabla_x \log q(x)\, q(x)\, dx.$$

Using the identity:

$$\nabla_x \log p(x) = \frac{\nabla_x q(x)}{q(x)} \quad \Rightarrow \quad \nabla_x \log q(x) \cdot q(x) = \nabla_x q(x),$$

we rewrite:

$$-2 \int \nabla_x \log p(x)^\top \nabla_x \log q(x)\, q(x)\, dx = -2 \int \nabla_x \log p(x)^\top \nabla_x q(x)\, dx.$$

This step assumes:

- $q(x) > 0$ almost everywhere,
- $q \in C^1(\mathbb{R}^d)$,
- The integral is well-defined and finite.

Step 2: Integration by Parts

We now handle the term:

$$-2 \int \nabla_x \log p(x)^\top \nabla_x q(x) \, dx.$$

Component-wise:

$$= -2 \sum_{i=1}^{m} \int \frac{\partial}{\partial x_i} \log p(x) \cdot \frac{\partial q(x)}{\partial x_i} \, dx.$$

By integration by parts:

$$= 2 \sum_{i=1}^{m} \int \frac{\partial^2}{\partial x_i^2} \log p(x) \cdot q(x) \, dx = 2 \int \Delta_x \log p(x) \cdot p(x) \, dx.$$

Assumption: boundary term vanishes,

$$\lim_{\|x\| \to \infty} \frac{\partial}{\partial x_i} \log p(x) \cdot q(x) = 0.$$

Final Form of $J(p)$

Combining the two terms, we have:

$$J(p) = \int \|\nabla_x \log p(x)\|^2 \, q(x) \, dx + 2 \int \Delta_x \log p(x) \, q(x) \, dx.$$

Key features:

- The expression depends only on the model p,
- The expectation is taken under $q(x)$, which can be approximated from data.

This forms the basis of the score matching objective.

Empirical Score Matching Objective

From the previous analysis, we obtained the following quantity to minimize:

$$J(p) = \int \left(\|\nabla_x \log p(x)\|^2 + 2\Delta_x \log p(x)\right) q(x)\, dx.$$

This is an expectation over the data distribution $p(x)$, which is unknown.

However, given i.i.d. samples $x^{(1)}, \ldots, x^{(n)} \sim p(x)$, we approximate it as:

$$\hat{J}(p) = \frac{1}{n} \sum_{i=1}^{n} \left(\|\nabla_x \log p(x^{(i)})\|^2 + 2\Delta_x \log p(x^{(i)})\right).$$

Outline

- 1. Review of Score Matching
- 2. Score Matching on Non-negative Data
  - (1) Weighted Score Matching
  - (2) Proposed Method: Boundary Term Approximation

Score Matching for Non-negative Data

Weighted Score Matching

Let $h_1, \dots, h_m : \mathbb{R}_+ \to \mathbb{R}_+$ be a.s. positive functions that are absolutely continuous on every bounded subinterval of $\mathbb{R}_+$, and set $h(x) = [h_1(x_1), \dots, h_m(x_m)]^\top$, which is absolutely continuous on $\mathbb{R}_+^m$.

Then, the weighted score matching objective is defined as

$$J_h(p) = \int_{\mathbb{R}_+^m} \left\| h(x)^{1/2} \odot \nabla_x \log p(x) - h(x)^{1/2} \odot \nabla_x \log q(x) \right\|^2 q(x)\, dx,$$

where $h(x)^{1/2} = \left[h_1(x_1)^{1/2}, \dots, h_m(x_m)^{1/2}\right]^\top$, and $\odot$ is denote the element-wise product

$$\left(\mathbf{y} \odot \mathbf{z} = [y_1 \cdot z_1, \cdots, y_m \cdot z_m] \quad \mathbf{y}, \mathbf{z} \in \mathbb{R}^m \right)$$

Rewriting Weighted Score Matching Objective

We expand the weighted score matching loss:

$$J_h(p) = \int_{\mathbb{R}_+^m} \left\| h(x)^{1/2} \odot (\nabla_x \log p(x) - \nabla_x \log q(x)) \right\|^2 q(x)\, dx - C(q) =$$

$$\int_{\mathbb{R}_+^m} \left\| h(x)^{1/2} \odot \nabla_x \log p(x) \right\|^2 q(x)\, dx - 2 \int_{\mathbb{R}_+^m} (h(x) \odot \nabla_x \log p(x))^\top (h(x) \odot \nabla_x \log q(x))\, q(x)\, dx$$

where $C(q) = \int \left\| h(x)^{1/2} \odot \nabla_x \log q(x) \right\|^2 q(x) dx$ is constant in p.

Goal: Eliminate dependence on $\nabla_x \log q(x)$ via integration by parts.

Integration by Parts with Weight Function

We now handle the cross term:

$$\int_{\mathbb{R}_+^m} \left(h(x)^{1/2} \odot \nabla_x \log p(x)\right)^\top \left(h(x)^{1/2} \odot \nabla_x \log q(x)\right) q(x)\, dx$$

$$= \int_{\mathbb{R}_+^m} (h(x) \odot \nabla_x \log p(x))^\top \nabla_x \log q(x) \cdot q(x)\, dx = \int (h(x) \odot \nabla_x \log p(x))^\top \nabla_x q(x)\, dx$$

Then, by component-wise integration by parts:

$$= -\int \mathrm{div}(h(x) \odot \nabla_x \log p(x)) \cdot q(x)\, dx + \text{boundary term}$$

Assumption: boundary term vanishes or is handled separately.

Final Objective: Weighted Score Matching

Combining all terms, we get:

$$J_h(p) = \int_{\mathbb{R}_+^m} \left(\left\| h(x)^{1/2} \odot \nabla_x \log p(x) \right\|^2 + 2 \cdot \mathrm{div}(h(x) \odot \nabla_x \log p(x)) \right) q(x)\, dx$$

Empirical Approximation: Given samples $x^{(1)}, \ldots, x^{(n)} \sim q(x)$, we approximate $J_h(p)$ by:

$$\hat{J}_h(p) = \frac{1}{n} \sum_{i=1}^{n} \left(\left\| h(x^{(i)})^{1/2} \odot \nabla_x \log p(x^{(i)}) \right\|^2 + 2 \cdot \mathrm{div}(h(x^{(i)}) \odot \nabla_x \log p(x^{(i)})) \right)$$

Outline

- 1. Review of Score Matching
- 2. Score Matching on Non-negative Data
    - (1) Weighted Score Matching
    - (2) Proposed Method: Boundary Term Approximation

Proposed Method: Score Matching on Domain M

We propose a modified score matching objective over a domain $M \subset \mathbb{R}^m$ with boundary ∂M.

Proposed objective

$$J_{\text{prop}}(p) = \int_M \|\nabla \log p(x)\|^2 q(x)\, dx + 2 \int_M \Delta \log p(x) \cdot q(x)\, dx + 2B(p, q)$$

where $B(p, q)$ denotes the **boundary term** arising from applying Stokes' theorem to the cross term.

Key difference from conventional score matching: The boundary term $B(p, q)$ is preserved rather than discarded.

Boundary Term in the Proposed Objective

The boundary term $B(p, q)$ has the following explicit form:

$$B(p, q) = -\int_{\partial M} (\nabla \log p(x) \cdot \mathbf{n}(x))\, q(x)\, dS(x)$$

Explanation:

- $\mathbf{n}(x)$: outward unit normal vector on the boundary ∂M,
- $dS(x)$: surface measure on ∂M,

Interpretation: It penalizes mismatch between the model flow and the data density near the boundary.

For simplicity, let us assume:

$$M = \mathbb{R}_+^m = \{(x_1, \cdots, x_m) \mid x_1, \cdots, x_m \geq 0\}$$

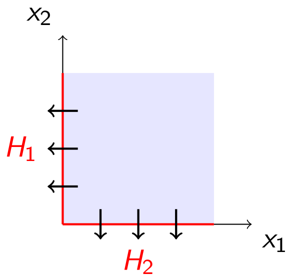
$$H_i = \{(x_1, \cdots, x_m) \in \mathbb{R}_+^m \mid x_i = 0\}$$

Then, the each component is

$$\partial M = \sum_{i=1}^{m} H_i$$

$$\mathbf{n}(x) = -e_i \quad (x \in H_i)$$

$$dS(x) = dx_{-i} = dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_m \quad (x \in H_i)$$



Transforming $B(p, q)$ and Preparing for Empirical Estimation

We start from the boundary term:

$$B(p, q) = -\int_{\partial M} (\nabla \log p(x) \cdot \mathbf{n}(x))\, q(x)\, dS(x)$$

In the case $M = \mathbb{R}_+^m$, this becomes:

$$= \sum_{i=1}^{m} \int_{H_i} \frac{\partial}{\partial x_i} \log p(x) \cdot q(x_i = 0, x_{-i})\, dx_{-i}$$

To make empirical estimation feasible, we rewrite the joint density using the identity:

$$q(x_i = 0, x_{-i}) = q(x_i = 0) \cdot q(x_{-i} \mid x_i = 0)$$

Hence:

$$= \sum_{i=1}^{m} q(x_i = 0) \int_{H_i} \frac{\partial}{\partial x_i} \log p(x) \cdot q(x_{-i} \mid x_i = 0)\, dx_{-i}$$

Empirical Approximation of the Boundary Term

We consider the following form of the boundary term:

$$B(p,q) = \sum_{i=1}^{m} q(x_i = 0) \cdot \mathbb{E}_{x_{-i} \sim q(x_{-i}|x_i=0)} \left[\frac{\partial}{\partial x_i} \log p(x_i = 0, x_{-i}) \right]$$

Since exact sampling from $x_i = 0$ is infeasible, we approximate using small $\epsilon > 0$:

Density at the boundary:

$$q(x_i = 0) = \lim_{\epsilon \to 0} \frac{\mathbb{P}(0 \le x_i < \epsilon)}{\left| \{0 \le x_i < \epsilon\} \right|} \approx \frac{n_i^\epsilon}{n\epsilon} \quad \left(n_i^\epsilon = \sum_{j=1}^{n} \mathbf{1}_{\{x|0 \le x_i < \epsilon\}}(x^{(j)}) \right)$$

Conditional expectation:

$$\mathbb{E}_{x_{-i} \sim q(x_{-i}|x_i=0)} \left[\frac{\partial}{\partial x_i} \log p(x) \right] \approx \frac{1}{n_i^\epsilon} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} \log p(x^{(j)}) \mathbf{1}_{\{x|0 \le x_i < \epsilon\}}(x^{(j)})$$

Final Empirical Objective with Boundary Term

Final approximation of $B(p, q)$:

$$\hat{B}(p, q) = \frac{1}{n\epsilon} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} \log p(x^{(j)}) \cdot \mathbf{1}_{\{0 \le x_i^{(j)} < \epsilon\}}(x^{(j)})$$

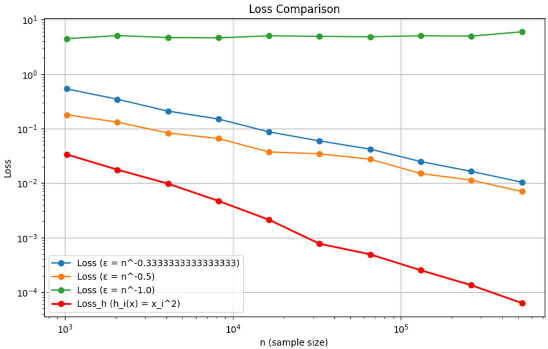
Thus, the full empirical objective becomes:

$$\hat{J}(p) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{m} \left(\frac{\partial}{\partial x_i} \log p(x^{(j)}) \right)^2 + 2 \cdot \frac{\partial^2}{\partial x_i^2} \log p(x^{(j)}) + \frac{2}{\epsilon} \cdot \frac{\partial}{\partial x_i} \log p(x^{(j)}) \cdot \mathbf{1}_{\{0 \le x_i^{(j)} < \epsilon\}}(x^{(j)})$$

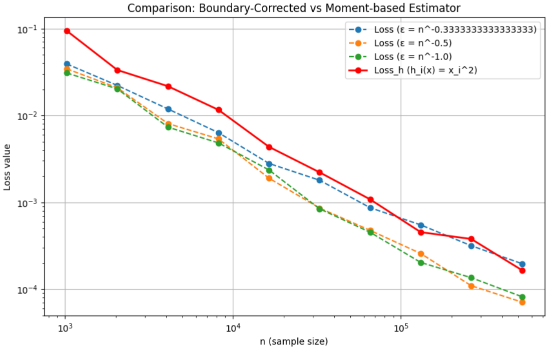
where $x^{(j)} \sim q(x)$, and $\epsilon > 0$ controls boundary proximity.

Experimental Comparison: Weighted vs. Proposed Score Matching

Exponential Distribution



Truncated Normal Distribution



Estimated Parameters:

- Weighted: $\hat{\theta}_w = 2\bar{x}/\overline{x^2}$
- Proposed: $\hat{\theta}_p = |H_i^\epsilon|/n\epsilon$

Estimated Parameters:

- Weighted: $\hat{\theta}_{iw} = \overline{x_i^4}/3\overline{x_i^2}$
- Proposed: $\hat{\theta}_p = n\epsilon\overline{x_i^2}/\left(\sum x_i^\epsilon + n\epsilon\right)$

References

Yu, S., Drton, M., & Shojaie, A. (2021).
Generalized Score Matching for General Domains.
Information and Inference: A Journal of the IMA, 11(2), 739–780.
https://doi.org/10.1093/imaiai/iaaa041