# Energy-Tweedie:
# Score meets Score, Energy meets Energy

**A general form of Tweedie's formula**

Andrej Leban
leban@umich.edu
University of Michigan · NeurIPS 2026

# Denoising & Tweedie's formula (1956)

- Data $X \sim p$ noised with $\epsilon \sim q$: $Y = X + \epsilon$.
- $m = p * q$ – the marginal.
- $s_m = \nabla_y \log m$ – the *score*.

Under *Gaussian* noise ($q = \mathcal{N}(0, \Sigma)$) the **mean** of the *denoising posterior* $P(X \mid Y = y)$ gives the score:

$$s_m(y) \;=\; \Sigma^{-1}(\mathbb{E}_X[X \mid Y = y] - y)$$

This is the identity behind diffusion models, and the posterior mean can be learned with the *squared loss* (MSE).

**What if the noise is not Gaussian?**

# Ingredients: Gibbs noise and scoring rules

- **Gibbs noise** (an *energy-based model*): $q_\eta(u) \propto e^{-c\,\mathcal{E}(u)}$, with potential $\mathcal{E}$, scale $c$, and parameters $\eta$. Examples: Gaussian, Laplace, generalized Gaussian, Student-$t$ ...
- **Scoring rules** evaluate a distributional forecast $P$ against an observation $y$. *Proper* rules are minimized in expectation by the true distribution, so they also serve as *distributional training losses*.
- A **kernel score** for a given *discrepancy* $\rho$:

$$S_\rho(P, y) \;=\; \mathbb{E}_{X\sim P}[\rho(X, y)] - \tfrac{1}{2}\,\mathbb{E}_{X,X'\overset{\text{iid}}{\sim}P}[\rho(X, X')]$$

An example is the *energy score*:

$$\mathrm{ES}_{I_d,\beta}(P, y) = \mathbb{E}_{X\sim P}\big[\|X - y\|^\beta\big] - \tfrac{1}{2}\mathbb{E}_{X,X'\overset{\text{iid}}{\sim}P}\big[\|X - X'\|^\beta\big],$$

i.e.: $\rho(x, y) = \|x - y\|^\beta$.

# The Energy-Tweedie identity

$$s_m(y) \;=\; -c\,\nabla_y^{\mathrm{PD}}\, S_\rho\big(P(X \mid Y = y),\, y\big)$$

where $\rho(x, y) = \mathcal{E}(y - x)$ is the discrepancy *induced* by the *noise potential* $\mathcal{E}$, and $\nabla_y^{\mathrm{PD}}$ is the *path derivative* (*stopgrad*) with respect to $y$.

| | GENERALIZED GAUSSIAN | GAUSSIAN |
|---|---|---|
| Noise | $\propto \exp\!\big(-\frac{\lambda}{\beta}\|u\|_{\Sigma^{-1}}^{\beta}\big)$ | $\propto \exp\!\big(-\frac{1}{2}u^\top \Sigma^{-1} u\big)$ |
| Objective | $\mathrm{ES}_{\Sigma^{-1},\beta}(P(X \mid Y = y), x)$ | $\|\mathbb{E}_X[X \mid Y = y] - x\|_{\Sigma^{-1}}^2$ |
| Posterior object | $P(X \mid Y = y)$ | $\mathbb{E}_X[X \mid Y = y]$ |
| Score | $-\frac{\lambda}{\beta}\nabla_y^{\mathrm{PD}}\mathrm{ES}_{\Sigma^{-1},\beta}(P(X \mid Y = y), y)$ | $\Sigma^{-1}(\mathbb{E}_X[X \mid Y = y] - y)$ |

# Score estimation

**Posterior-based:**

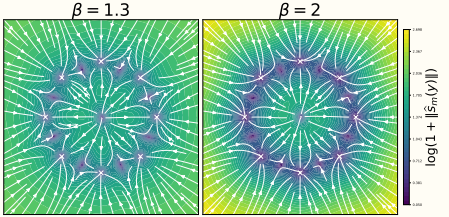
$$\widehat{s}_m(y) \;=\; -\frac{c}{N}\sum_{i=1}^{N} \nabla_y^{\mathrm{PD}}\, S_\rho(\delta_{X^{(i)}}, y), \quad X^{(i)} \overset{\text{iid}}{\sim} P_\theta(X \mid Y = y)$$

**ET-DSM – Energy-Tweedie Denoising Score Matching:**

$$\min_\theta \; \mathbb{E}_{X\sim p,\, Y\sim q_\eta(\cdot|X)} \left[ \left\| s_\theta(Y,\eta) + c\, \nabla_Y^{\mathrm{PD}}\, S_\rho(\delta_X, Y) \right\|_2^2 \right]$$

Examples from the *Eight Gaussians* dataset, noised with generalized Gaussian noise (over a range of parameters):

| SCORE ESTIMATOR | NORM. MSE ↓ | COSINE SIM. ↑ |
|---|---|---|
| ET-Posterior | 0.029–0.077 | 0.951–0.982 |
| ET-DSM | 0.014–0.019 | 0.990–0.992 |
| Hyvärinen SM | 0.022–0.027 | 0.987–0.989 |

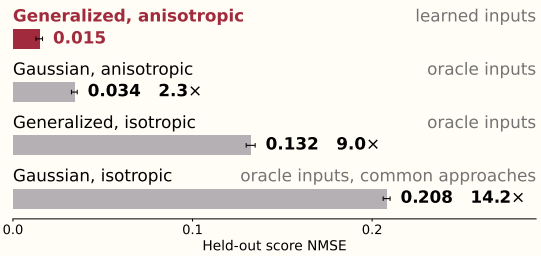


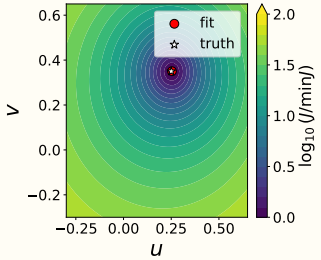
# Noise parameter estimation

Use *another score estimator* $\hat{s}_m$ and turn it into an *M-estimation* problem:

$$\min_{\eta} \; \mathbb{E}_{Y\sim m}\left[\left\|\hat{s}_m(Y) + c\,\nabla_Y^{\mathrm{PD}} S_\rho(P_\theta(X \mid Y), Y)\right\|_2^2\right]$$

Using generalized Gaussian noise with $\eta = (\beta, \lambda, u, v)$, $\Sigma = \sigma^2 \left(\begin{smallmatrix} 1+u & v \\ v & 1-u \end{smallmatrix}\right)$, we compare the full ET approach to those with restricted distributions *that we give oracle inputs*:



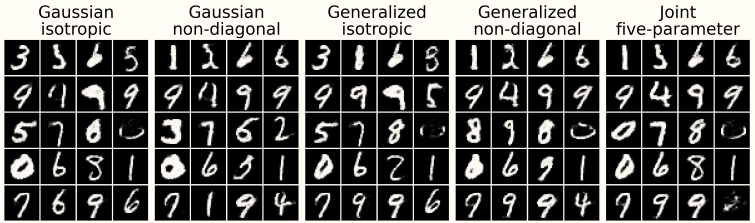



# Energy-Tweedie diffusion

Energy-Tweedie diffusion enables *paths* through the *parameters of the noise distribution* $t \mapsto \eta_t$ instead of a scalar *diffusion schedule*. This leads to a **path-induced training loss:**

$$\min_{\theta} \; \mathbb{E}_{t\sim\mathrm{Unif}([0,1]),\, X\sim p,\, Y\sim q_{\eta_t}(\cdot|X)} \Big[ S_{\rho_t}\big(P_\theta(X \mid Y, t),\, X\big) \Big],$$

and/or to **sampling from a freely chosen path**:




| | Gaussian isotropic | Gaussian non-diagonal | Generalized isotropic | Generalized non-diagonal | Joint five-parameter | |
|---|---|---|---|---|---|---|
| ED ↓ | 0.574 | 0.247 | 0.637 | 0.218 | 0.303 | (real: 0.015) |

# Energy-Tweedie

## A general form of Tweedie's formula

- Score estimation
- Noise parameter estimation
- Energy-Tweedie diffusion

PAPER

CODE

arXiv:2512.23818
github.com/andleb/et