Improving Numerical Stability of Normalized Mutual Information Estimator on High Dimensions

Scan for the paper

Marko Tuononen and Ville Hautamäki

UNIVERSITY OF EASTERN FINLAND

NOKIA

Key Contribution: A log-domain reformulation that prevents numerical overflow in high-dimensional KSG-based NMI estimation.

Motivation and Problem

Mutual information measures statistical dependence but is intractable to compute exactly for continuous variables, requiring estimation in practice. Since it is unbounded and scale-dependent, its values are not directly comparable across datasets. **Normalized variants provide a bounded and comparable dependency measure**, useful in applications such as molecular dynamics [2] and interpretable machine learning (IML) [4].

Problem: KSG[1]-based normalized mutual information estimator [2] suffers from numerical overflow when computing the scaling-invariant k-NN radii (Eq. 3) in high dimensions.

Background

Mutual Information (MI) can be expressed via Shannon entropy:

$$I(X;Y) = H(X) + H(Y) - H(X;Y), \quad I(X;Y) \geq 0. \tag{1}$$

A normalized version is **Normalized Mutual Information (NMI)**:

$$I_{\text{NMI}}(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}, \quad 0 \leq I_{\text{NMI}}(X;Y) \leq 1. \tag{2}$$

KSG[1]-based NMI estimation [2]. For continuous variables, relative entropy with invariant measures is used. Scaling-invariant k-Nearest-Neighbors (k-NN) radii are used to estimate local data densities:

$$\tilde{\varepsilon} = \frac{\varepsilon}{\langle \varepsilon^{d_X+d_Y} \rangle^{1/(d_X+d_Y)}} = \frac{\varepsilon}{V}. \tag{3}$$

where ε denotes the raw k-NN radii, d_X and d_Y are the dimensionalities of the marginal spaces, and V is the normalization factor. The marginal and joint relative entropy estimators are:

$$\hat{H}_r(X) = -\langle \psi(n_x+1) \rangle + \psi(N) + d_X \langle \ln \tilde{\varepsilon} \rangle, \tag{4}$$

$$\hat{H}_r(X;Y) = -\psi(k) + \psi(N) + (d_X+d_Y) \langle \ln \tilde{\varepsilon} \rangle, \tag{5}$$

where $\psi(\cdot)$ is the digamma function and $\langle \cdot \rangle$ denotes the sample mean. The NMI estimate is computed by substituting $\hat{H}_r(X)$, $\hat{H}_r(Y)$, and $\hat{H}_r(X;Y)$ into Equations (1) and (2).

[1] Kraskov-Stögbauer-Grassberger (KSG) estimator [1]

Proposed Fix [3] to the Problem

Take the logarithm of the normalization factor V in Eq. 3:

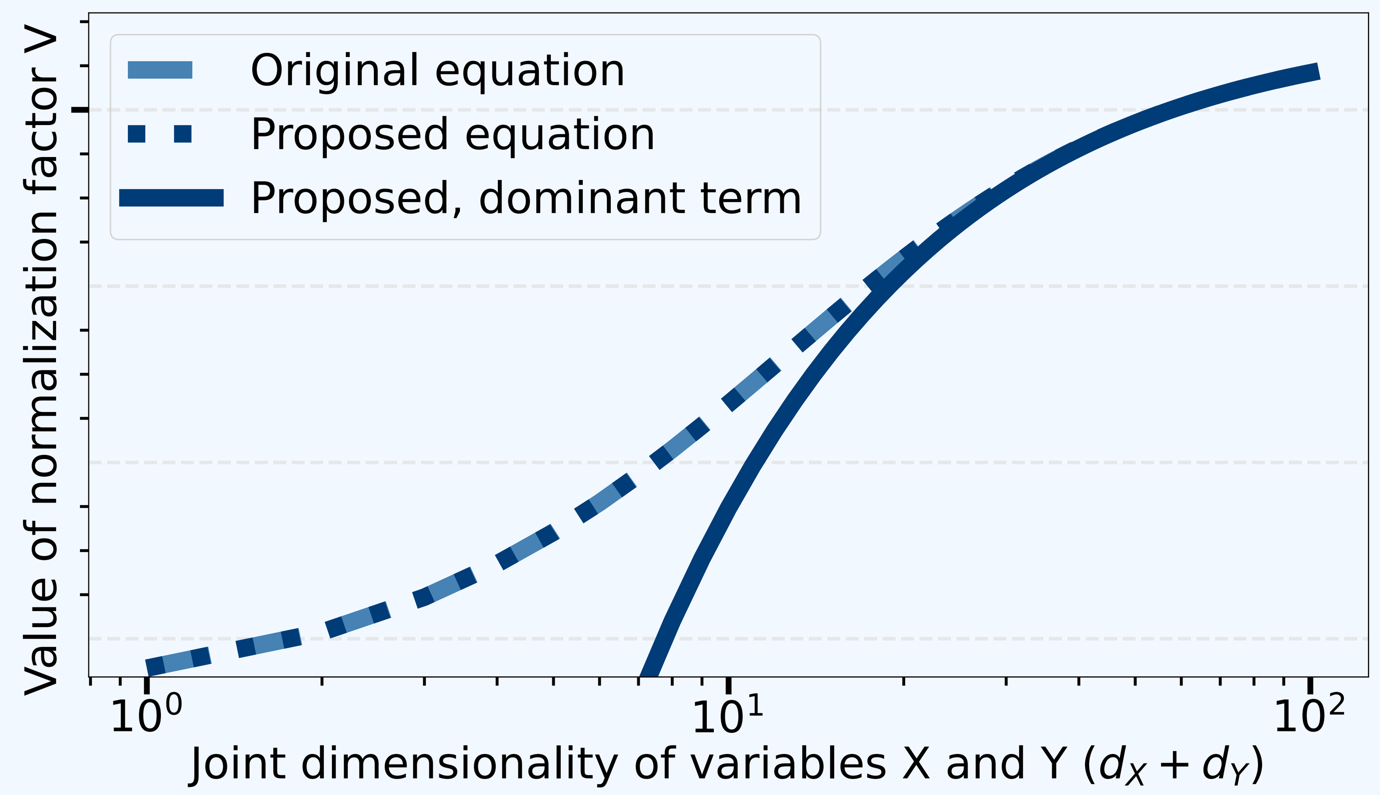
$$\ln V = \frac{1}{d_X+d_Y} \left[\ln \frac{1}{N} + \ln \sum_{i=1}^{N} \varepsilon_i^{d_X+d_Y} \right]. \tag{6}$$

Let ε_{max} be the largest radius and apply the log-sum-exp trick:

$$\ln V = \ln \varepsilon_{\text{max}} + \frac{1}{d_X+d_Y} \ln \left(\frac{\sum_{i=1}^{N} (\varepsilon_i/\varepsilon_{\text{max}})^{d_X+d_Y}}{N} \right). \tag{7}$$

All terms $\varepsilon_i/\varepsilon_{\text{max}}$ are now ≤ 1, preventing numerical overflow when Eq. 7 is substituted back into Eq. 3.

In high dimensions, the normalization factor V becomes dominated by the largest radius. Thus, as $d_X + d_Y \to \infty$, $V \to \varepsilon_{\text{max}}$ as illustrated in the figure below.



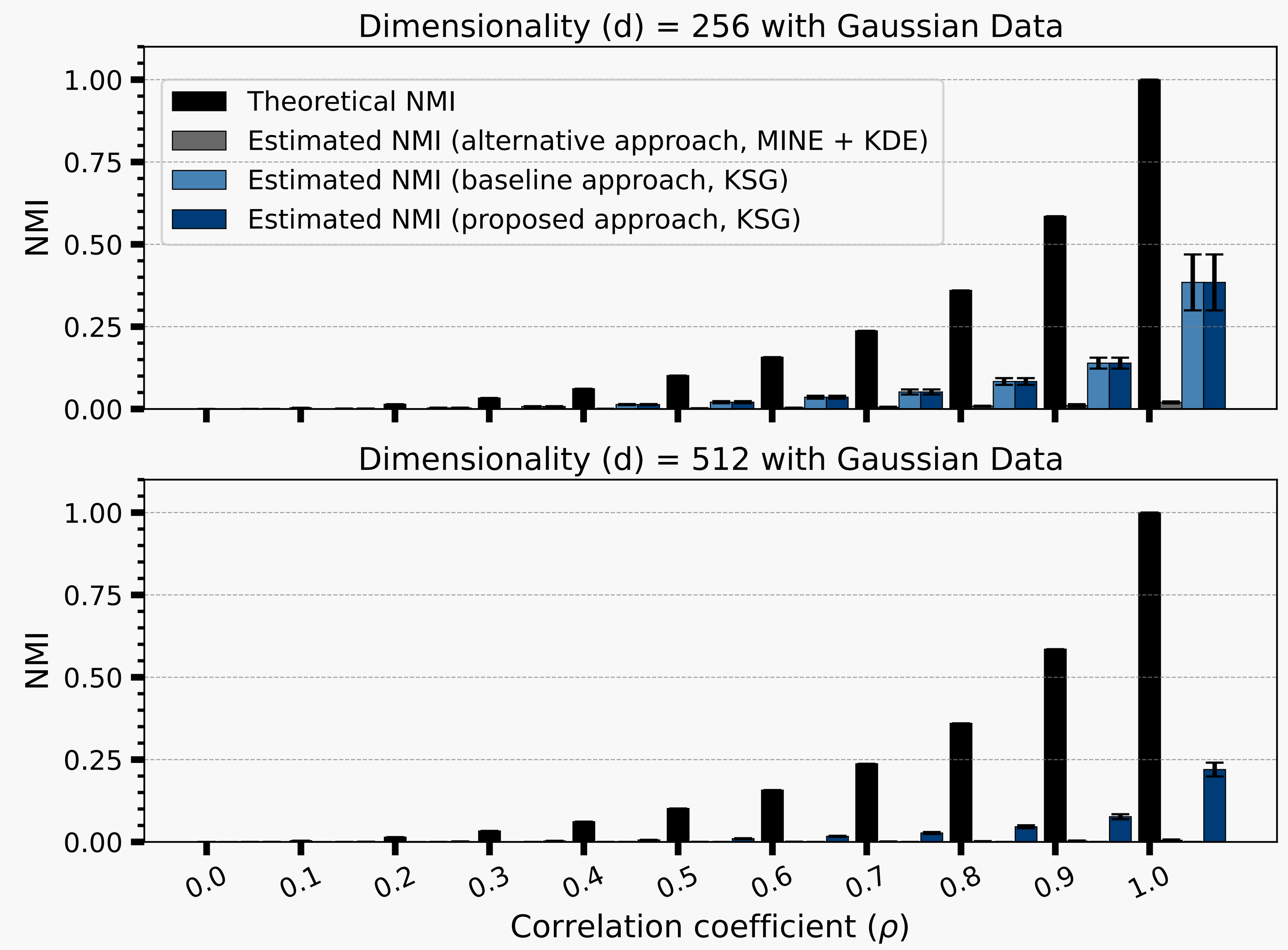
Experimental Setup

Data: Multivariate Gaussian with component-wise correlation ρ and Student's t (identity dispersion, ν); $10{,}000$ samples per case.

Gaussian $d \in \{1, 4, 16, 32, 64, 128, 256, 512\}$, $\rho \in [0, 1]$
Student's t $d \in \{1, 4, 16, 32\}$, $\nu \in [0.125, 10]$

Metrics: estimated NMI (mean and variance over 10 repetitions), theoretical NMI comparison, and numerical stability.

Experimental Results



The proposed fix prevents numerical overflow while preserving estimator accuracy and computational complexity, demonstrating stable operation up to 512 dimensions.

Future work: Experiments with real-world high-dimensional datasets and exploration of neural estimators (e.g., MINE) for direct NMI estimation, particularly for applications in IML.

References

[1] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 69(6):066138, 2004.

[2] Daniel Nagel, Georg Diez, and Gerhard Stock. Accurate estimation of the normalized mutual information of multidimensional data. *The Journal of Chemical Physics*, 161(5):054108, 2024.

[3] Marko Tuononen. Add 'volume_stable' invariant measure to improve numerical stability in NormalizedMI. `https://github.com/moldyn/NorMI/pull/14`, 2024. Pull request submitted on Oct 13, 2024, merged on Nov 21, 2024.

[4] Marko Tuononen, Dani Korpi, and Ville Hautamäki. Interpreting deep neural network-based receiver under varying signal-to-noise ratios. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2025.