The Double-Edged Sword: Perception and Uncertainty in Inverse Problems

Regev Cohen **Ehud Rivlin** **Daniel Freedman**

Verily AI, Israel
regevcohen@google.com

Abstract

Inverse problems pose significant challenges due to their inherent ambiguity in mapping observed data back to its original state. While recent advances have yielded impressive results in restoring degraded data, attaining high perceptual quality comes at the cost of increased hallucinations. This paper investigates this phenomenon to reveal a fundamental tradeoff between perception and uncertainty in solving inverse problems. Using error entropy as a measure of uncertainty, we demonstrate that higher perceptual quality in restoration algorithms is accompanied by a surge in uncertainty. Leveraging Rényi divergence as a perception metric, we derive bounds for this tradeoff, allowing for categorization of different inverse methods based on their performance. Additionally, we connect estimation distortion with uncertainty, offering novel insights into the traditional perception-distortion tradeoff. Our work provides a rigorous framework for analyzing uncertainty in the context of solving inverse problems, highlighting its interplay with perception and distortion, while underscoring the limitations of current approaches to achieving both high perceptual quality and low uncertainty simultaneously.

1 Introduction

Generative artificial intelligence (AI) has transformed inverse problems by delivering unprecedented performance in tasks like image denoising, super-resolution, and inpainting [8, 7]. These models have far-reaching applications in fields like medical imaging, computer vision, and signal processing, offering the ability to recover missing or corrupted data with astonishing realism. However, generative models are susceptible to "hallucinations," generating realistic but inaccurate content due to the ill-posed nature of restoration problems. Interestingly, the severity of hallucinations seems to correlate with the model's perceptual quality.

One of the seminal works investigating the perception of restoration models is the perception-distortion tradeoff [4, 3, 6]. The authors established a fundamental tradeoff between perceptual quality and distortion in image restoration, applicable to any distortion measure and distribution. When using mean squared error (MSE), they showed that achieving perfect perceptual quality could incur a maximum penalty of 3dB in peak signal-to-noise ratio. The work in [11] extended this concept by providing closed-form expressions for the tradeoff when considering MSE distortion and Wasserstein-2 distance as a perception measure.

In this paper, we complement the above studies by analyzing the relation of uncertainty to perception and distortion, establishing a fundamental tradeoff between uncertainty and perception. The main contributions are as follows: (i) The introduction of the uncertainty-perception function, a novel framework based on information-theoretic principles, proving the existence of a tradeoff between uncertainty and perception, regardless of the underlying data distribution. (ii) Derivation of lower and upper bounds for the UP function using Rényi divergence as a measure of perception, construct-

Mathematics of Modern Machine Learning Workshop at NeurIPS 2023.

Table 1: Information-Theory Measures.

Name	Definition
Differential Entropy	$h(X) \triangleq -\int p_X(x) \log p_X(x) dx$.
Entropy Power	$N(X) \triangleq \frac{1}{2\pi e} e^{\frac{2}{d} h(X)}$ $(X \in \mathbb{R}^d)$.
Rényi Entropy	For order $r \geq 0$: $h_r(X) \triangleq \frac{1}{1-r} \ln \int p_X^r(x) dx$.
Rényi Divergence	For order $r \geq 0$: $D_r(X,Y) \triangleq \frac{1}{r-1} \ln \int p_X^r(x) p_Y^{1-r}(x) dx$.
Conditional Entropy	$h(X\|Y) \triangleq \mathbb{E}_{y \sim p_Y} [h(X\|Y=y)]$.
Conditional Divergence	$D_v(X, Z \big\| Y) \triangleq \mathbb{E}_{y \sim p_Y} [D_v(X\|Y=y, Z\|Y=y)]$.

ing an uncertainty-perception plane that categorizes estimators into distinct performance domains. (iii) Extension of a well-known relationship between uncertainty and distortion to demonstrate that the uncertainty-perception tradeoff implies the seminal distortion-perception tradeoff. Theoretical findings are supported by numerical experiments using established super-resolution algorithms. Throughout the paper, we rely on the information-theoretic measures defined in Table 1, where we assume all mentioned quantities, which involve integrals, are well-defined and finite.

2 The Uncertainty-Perception Tradeoff

Problem Formulation We study the problem of estimating a random vector $X \in \mathbb{R}^d$ from its observations $Y \in \mathbb{R}^{d'}$. This translates to devise an estimator $\hat{X}(Y)$ which induces a conditional distribution measure $p_{\hat{X}|Y}$ on $\mathbb{R}^d$. We rely on the following mild assumptions:

Assumption 1 (Markovian Process)**.** *The estimation process is a Markov chain $X \to Y \to \hat{X}$, such that $\hat{X}$ is independent of X given Y.*

Assumption 2 (Loss of Information)**.** *The problem is ill-posed so X cannot be perfectly recovered from Y. Namely, $p_{X|Y}(\cdot|y)$ is not a delta function for almost every y.*

Assumption 3 (Unbiasedness)**.** *$\hat{X}$ is an unbiased estimator of X, implying $\mathbb{E}(\hat{X}) = \mathbb{E}(X)$.*

Our aim is to establish the relationship between two performance criteria of an estimator $\hat{X}(Y)$:

- Estimation uncertainty $U_{nc}(\hat{X}|Y)$ of estimator $\hat{X}$ given the information available in Y.
- Perception quality $D_v(X, \hat{X} \big| Y)$ of estimator $\hat{X}$ in capturing the true distribution of X.

While there are diverse approaches to define and quantify uncertainty [12, 1], we rely on an information-theoretic approach and adopt the concept of entropy, a measure of statistical dispersion. Specifically, we employ error entropy power as our uncertainty measure to formulate the following uncertianty-perception (UP) function:

$$U(P) \triangleq \min_{p_{\hat{X}|Y}} \Big\{ N(\hat{X} - X|Y) \ : \ D_v(X, \hat{X} \big| Y) \leq P \Big\}. \tag{1}$$

In words, $U(P)$ represents the minimum achievable uncertainty for an estimator with a perception quality of at least P, considering the information provided by the measurements Y. The above objective focuses on the information content of the error signals rather than their energy (second-order statistics), forcing errors to be concentrated and thus leading to robust predictions.

The Uncertainty-Perception Plane Following the definition of the uncertainty-perception function, we deduce its properties to establish the uncertainty-perception tradeoff. The exact form of the tradeoff generally governed by the underlying distributions of X and Y, as well as the chosen divergence measure $D_v(\cdot, \cdot)$. However, the following theorem identifies fundamental properties of the uncertainty-perception function, U(P), that hold regardless of these specifics.

Theorem 1. *The uncertainty-perception function $U(P)$ displays the following properties*

1. *Quasi-linearity (monotonically non-increasing and continuous):*

$$\min\Big(U(P_1), U(P_2)\Big) \leq U\Big(\lambda P_1 + (1-\lambda) P_2\Big) \leq \max\Big(U(P_1), U(P_2)\Big), \ \forall \lambda \in [0,1].$$

2. *Boundedness:* $N(X|Y) \leq U(P) \leq 2N(X|Y)e^{\frac{2}{d}D_{KL}(X,X_G|Y)}$.
3. *Assume* $D_v(X, \hat{X}\big|Y)$ *is convex in its second argument. Then, for any* $P \geq 0$*, the minimum is attained on the boundary where* $D_v(X, \hat{X}\big|Y) = P$*.*

Here $X_G \sim \mathcal{N}(0, \Sigma_x)$ *is a Gaussian random variable, independent of* X*, with the identical covariance matrix* $\Sigma_{X_G} = \Sigma_X$*, and* $D_{KL}(X, X_G|Y)$ *is the conditional Kullback–Leibler divergence.*

Irrespective of the divergence measure or underlying distributions, the theorem establishes a quasi-linear tradeoff between perceptual quality and uncertainty. This tradeoff is intrinsically linked to $N(X|Y)$, representing the inherent uncertainty rooted in information loss. Remarkably, for sufficiently large dimension d such that $e^{\frac{2}{d}D_{KL}(X,X_G|Y)} \approx 1$, perfect perceptual quality can be attained at a expense of at most twice the inherent uncertainty.

To gain deeper insights, we adopt a specific perception measure based on Rényi divergence with $r = 1/2$, leading to the following form:

$$U(P) = \min_{p_{\hat{X}|Y}} \left\{ N(\hat{X} - X|Y) \ : \ D_{1/2}(X, \hat{X}\big|Y) \leq P \right\}. \tag{2}$$

Rényi divergence, with its connection to Rényi entropy, generalizes various entropy measures like Hartley, Shannon, collision, and min-entropy. Thus, analyzing the specific formulation in (2) may provide insights applicable to a broader range of divergence measures. The following theorem shows the tradeoff, defined by (2), is confined between two convex functions.

Theorem 2. *The uncertainty-perception function is confined to the following region*

$$\eta(P) \cdot N(X|Y) \ \leq \ U(P) \ \leq \ \eta(P) \cdot N(X_G|Y)$$

where $\eta(P) = \left(2e^{\frac{2P}{d}} - \sqrt{(2e^{\frac{2P}{d}} - 1)^2 - 1}\right) \in [1, 2]$ *is a convex function w.r.t the perception index.*

Theorem 2 defines the uncertainty-perception plane, shown in Fig. 1, which segmentes the space into three distinct regions: (i) The impossible region, inaccessible to any estimator. (ii) The tradeoff feasible region, containing all optimal estimators according to (2). (iii) The overestimation region, occupied by suboptimal estimators that overestimate uncertainty for a given perceptual quality.
Thus, the uncertainty-perception plane is a valuable tool for evaluating and optimizing estimator performance, suggesting that estimators in the overestimation region can be optimized to achieve lower uncertainty without compromising perceptual quality. Next, we investigate the influence of dimension on the tradeoff by analyzing $\eta(d; P)$, a variation of $\eta(P)$ with dimension for a fixed perceptual quality. Fig. 2 reveals a severe tradeoff in high-dimensional settings, indicating that marginal perceptual gains entail substantial uncertainty increases.

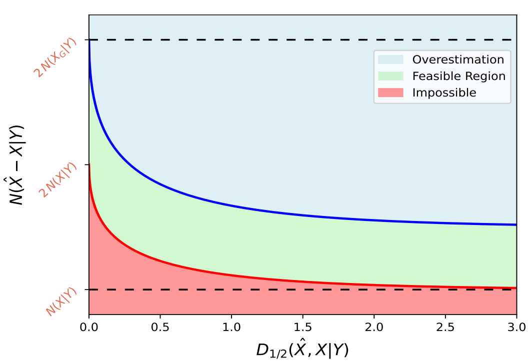

Figure 1: The uncertainty-perception plane.

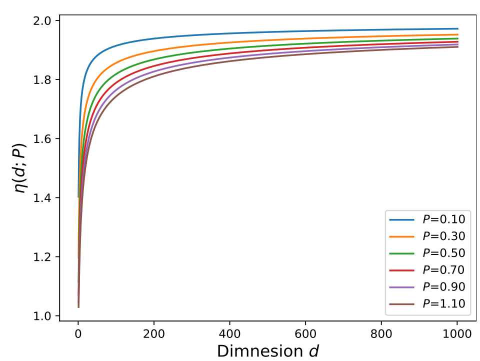

Figure 2: Curse of dimensionality.

The Distortion-Perception Tradeoff Building upon the established uncertainty-perception tradeoff, we extend our analysis to encompass estimation distortion. The following theorem establishes a connection between uncertainty and mean-squared error distortion.

Theorem 3. *For any random variable* X*, observation* Y *and unbiased estimator* $\hat{X}$*, it holds that*

$$\frac{1}{d}\mathbb{E}\left[||\hat{X} - X||^2\right] \geq N\left(\hat{X} - X|Y\right).$$

By considering the estimator $\hat{X}$ as a function of the perception index P, we derive the next corollary.

Corollary 1. *Define the distortion-perception function as*

$$D(P) \triangleq \min_{p_{\hat{X}|Y}} \left\{ \frac{1}{d}\mathbb{E}\left[||\hat{X} - X||^2 \right] :\ D_v(X, \hat{X}\big|Y) \leq P \right\}.$$

Then, for any perceptual index P, we have $D(P) \geq U(P)$.

By employing MSE as a distortion measure, the uncertainty-perception tradeoff gives rise to a distortion-perception tradeoff [5, 11], providing a fresh perspective on the latter and highlighting the fundamental interplay between uncertainty, distortion, and perception.

3 Numeric Evaluation

To empirically validate the uncertainty-perception tradeoff and its connection to MSE distortion, we follow previous works [11, 5] and evaluate super-resolution (SR) algorithms on the BSD100 dataset [18]. The algorithms include EDSR [16], ESRGAN [22], SinGAN-Z [20], SANGAN [13], DIP [21], SRResNet and SRGAN variants [15], EnhanceNet [19], and bicubic recovery. Low-resolution observations are generated by downsampling ground-truth images by a factor of 4 using a bicubic kernel. We treat each image as a stationary random source and extract 9x9 patches for statistical measures. R'enyi divergence is computed via kernel density estimation and empirical expectations. Differential entropy is estimated using the Kozachenko-Leonenko nearest neighbor method [14, 10, 2].

Fig. 3 showcases the tradeoff between uncertainty and perception in super-resolution methods. Algorithms with superior perceptual quality exhibit higher uncertainty values, and vice versa. This trend aligns with our theoretical findings. Operating in a high-dimensional space ($d = 243$), we observe that even minor improvements in perceptual quality result in substantial increases in uncertainty. Fig. 3 further confirms the relationship between uncertainty and distortion. A clear correlation is observed, where increasing uncertainty leads to a significant rise in distortion. This reinforces the findings from the previous section, highlighting the impact of uncertainty on distortion.

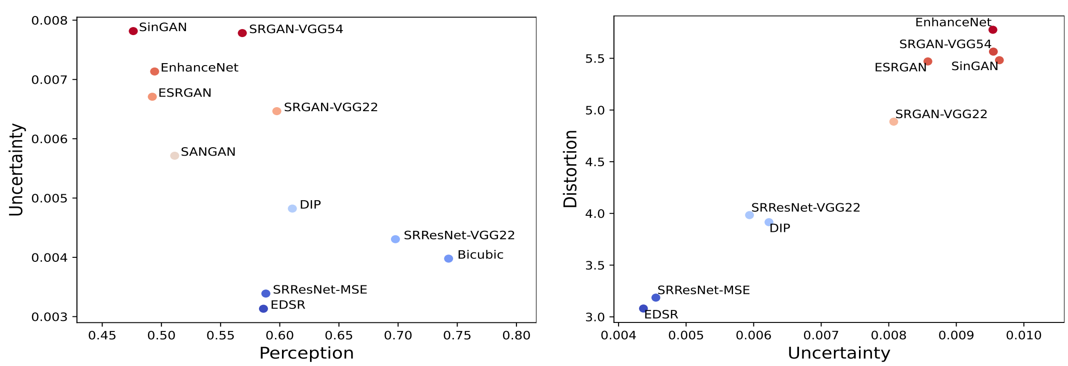

Figure 3: Evaluation on SR algorithms on (left) the uncertainty-distortion plane and (right) on the uncertainty-distortion plane.

4 Conclusion

This study unveils the uncertainty-perception tradeoff in restoration tasks, demonstrating that achieving superior perceptual quality leads to higher uncertainty levels. The tradeoff is characterized for Rényi divergence, revealing its quasi-linear nature and its curse of dimensionality. We derived the uncertainty-perception plane which emerges as a tool for evaluating estimator performance and identifying potential improvements. By bridging the gap between uncertainty and MSE distortion, a novel interpretation of the uncertainty-distortion tradeoff is presented. This work highlights the intrinsic relationship between uncertainty, distortion, and perception in restoration tasks.

References

[1] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information fusion **76**, 243–297 (2021)

[2] Beirlant, J., Dudewicz, E.J., Györfi, L., Van der Meulen, E.C., et al.: Nonparametric entropy estimation: An overview. International Journal of Mathematical and Statistical Sciences **6**(1), 17–39 (1997)

[3] Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 pirm challenge on perceptual image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018)

[4] Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE (jun 2018). https://doi.org/10.1109/cvpr.2018.00652, `https://doi.org/10.1109%2Fcvpr.2018.00652`

[5] Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6228–6237 (2018)

[6] Blau, Y., Michaeli, T.: Rethinking lossy compression: The rate-distortion-perception tradeoff. In: International Conference on Machine Learning. pp. 675–685. PMLR (2019)

[7] Cohen, R., Blau, Y., Freedman, D., Rivlin, E.: It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. Advances in Neural Information Processing Systems **34**, 18152–18164 (2021)

[8] Cohen, R., Elad, M., Milanfar, P.: Regularization by denoising via fixed-point projection (RED-PRO). SIAM Journal on Imaging Sciences **14**(3), 1374–1406 (2021)

[9] Cover, T.M.: Elements of information theory. John Wiley & Sons (1999)

[10] Delattre, S., Fournier, N.: On the Kozachenko–Leonenko entropy estimator. Journal of Statistical Planning and Inference **185**, 69–93 (2017)

[11] Freirich, D., Michaeli, T., Meir, R.: A theory of the distortion-perception tradeoff in Wasserstein space. Advances in Neural Information Processing Systems **34**, 25661–25672 (2021)

[12] Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al.: A survey of uncertainty in deep neural networks. Artificial Intelligence Review pp. 1–77 (2023)

[13] Kligvasser, I., Michaeli, T.: Sparsity aware normalization for GANs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8181–8190 (2021)

[14] Kozachenko, L.F., Leonenko, N.N.: Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsii **23**(2), 9–16 (1987)

[15] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)

[16] Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)

[17] Madiman, M., Melbourne, J., Xu, P.: Forward and reverse entropy power inequalities in convex geometry. In: Convexity and concentration, pp. 427–485. Springer (2017)

[18] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 416–423. IEEE (2001)

[19] Sajjadi, M.S., Scholkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: Proceedings of the IEEE international conference on computer vision. pp. 4491–4500 (2017)

[20] Shaham, T.R., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4570–4580 (2019)

[21] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018)

[22] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops. pp. 0–0 (2018)

A Information-Theory Preliminaries

Here, we provide a brief overview of essential definitions and fundamental results that stand in the center of our study. Let X, Y and Z be continuous random variables with probability density functions $p_X(x)$, $p_Y(y)$ and $p_Z(z)$ respectively, defined over a space Ω.

Definition 1 (Entropy). *The differential entropy of X, whose support is a set S_x, is defined by*

$$h(X) \triangleq -\int_{S_X} p_X(x) \log p_X(x) dx.$$

Definition 2 (Rényi Entropy). *The Rényi entropy of order $r \geq 0$ of X is defined by*

$$h_r(X) \triangleq \frac{1}{1-r} \ln \int p_X^r(x) dx.$$

The above quantity generalizes various notions of entropy, including Hartley entropy, collision entropy, and min-entropy. In particular, for $r = 1$ we have

$$h_1(X) \triangleq \lim_{r \to 1} h_r(X) = h(X).$$

Definition 3 (Entropy Power). *Let be $h(X)$ be the differential entropy of $X \in \mathbb{R}^d$. Then, the entropy Power of X is given by*

$$N(X) \triangleq \frac{1}{2\pi e} e^{\frac{2}{d} h(X)}.$$

Definition 4 (Divergence). *A statistical divergence is any function $D_v : \Omega \times \Omega \to \mathbb{R}^+$ which satisfies the following conditions for all $p, q \in \Omega$:*

1. $D_v(p, q) \geq 0$. 2. $D_v(p, q) = 0$ *iff* $p = q$ *almost everywhere.*

Definition 5 (Rényi Divergence). *The Rényi divergence of order $r \geq 0$ between p_X and p_Y is*

$$D_r(X, Y) \triangleq \frac{1}{r-1} \ln \int p_X^r(x) p_Y^{1-r}(x) dx.$$

The above establishes a spectrum of divergence measures, generalising the Kullback–Leibler divergence as $D_1(X, Y) = D_{KL}(X, Y)$.

Definition 6 (Conditioning). *Consider the joint probability p_{XY} and the conditional probabilities $p_{X|Y}(x|y)$ and $p_{Z|Y}(z|y)$. The conditional differential entropy of $X \in \mathbb{R}^d$ given Y is defined as*

$$h(X|Y) \triangleq -\int_{S_{XY}} p_{XY}(x, y) \log p_{X|Y}(x|y) dx dy = \mathbb{E}_{y \sim p_Y} \left[h(X|Y = y) \right]$$

where S_{XY} is the support set of p_{XY}. Then, the conditional entropy power of X given Y is

$$N(X|Y) = \frac{1}{2\pi e} e^{\frac{2}{d} h(X|Y)}.$$

Similarly, the conditional divergence between X and Z given Y is defined as

$$D_v(X, Z \big| Y) \triangleq \mathbb{E}_{y \sim p_Y} \left[D_v(X|Y = y, Z|Y = y) \right].$$

For example, the conditional Rényi divergence is given by

$$D_r(X, Z \big| Y) \triangleq \int \left(\frac{1}{r-1} \ln \int p_{X|Y}^r(x|y) p_{Z|Y}^{1-r}(x|y) dx \right) p_Y dy.$$

Table 2: Formulas for Multivariate Gaussian Distribution

Distribution	Quantity	Closed-Form Expression
$X \sim \mathcal{N}(\mu_x, \Sigma_x)$	$h(X)$	$\frac{1}{2}\ln\{(2\pi e)^d \lvert\Sigma_x\rvert\}$.
$X \sim \mathcal{N}(\mu_x, \Sigma_x)$	$N(X)$	$\lvert\Sigma_x\rvert^{1/n}$.
$X \sim \mathcal{N}(\mu_x, \Sigma_x)$	$h_{\frac{1}{2}}(X)$	$\frac{1}{2}\ln\{(8\pi)^d \lvert\Sigma_x\rvert\}$.
$X \sim \mathcal{N}(\mu_x, \Sigma_x)$, $Y \sim \mathcal{N}(\mu_y, \Sigma_y)$	$D_{1/2}(X,Y)$	$\frac{1}{4}(\mu_x-\mu_y)^T\left(\frac{\Sigma_x+\Sigma_y}{2}\right)^{-1}(\mu_x-\mu_y)+\ln\left(\frac{\left\lvert\frac{\Sigma_x+\Sigma_y}{2}\right\rvert}{\sqrt{\lvert\Sigma_x\rvert\lvert\Sigma_y\rvert}}\right)$.

Equipped with the above, we present two important results that are used throughout our derivations.

Lemma 1 (**Maximum Entropy Principle** [9]). *Let $X \in \mathbb{R}^d$ be a continuous random variable with zero mean and covariance Σ_x. Define $X_G \sim \mathcal{N}(0, \Sigma_x)$ to be a Gaussian random variable, independent of X, with the identical covariance matrix $\Sigma_{x_G} = \Sigma_x$. Then,*

$$h(X) \leq h(X_G),$$
$$N(X) \leq N(X_G) = \lvert\Sigma_x\rvert^{1/n}.$$

Lemma 2 (**Entropy power inequality** [17]). *Let X and Y be independent continuous random variables. Then, the following inequality holds*

$$N(X) + N(Y) \leq N(X+Y),$$

where equality holds iff X and Y are multivariate Gaussian random variables with proportional covariance matrices. Equivalently, let X_g and Y_g be defined as independent, isotropic multivariate Gaussian random variables satisfying $h(X_g) = h(X)$ and $h(Y_g) = h(Y)$. Then,

$$h(X) + h(Y) = h(X_g) + h(Y_g) = h(X_g + Y_g) \leq h(X+Y).$$

B Proof of Theorem 1

We first prove the uncertainty-perception function is quasi-linear. The constraint $\mathcal{C}(P) \triangleq \{\hat{X} : D_v(X, \hat{X}\big|Y) \leq P\}$ defines a compact set which is continuous in P. Hence, by the Maximum Theorem [9], $U(P)$ is continuous. In addition, $U(P)$ is the minimal error entropy power obtained over a constraint set whose size does not decrease with P, thus, $U(P)$ is non-increasing in P. Any continuous non-increasing function is quasi-linear. Now, we drive the bounds of the uncertainty-perception function, starting with the lower bound. Consider the case where $P = \infty$, leading to the following unconstrained problem

$$U(\infty) \triangleq \min_{p_{\hat{X}|Y}} N(\hat{X} - X|Y). \tag{3}$$

For any $P \geq 0$ it holds that $U(\infty) \leq U(P)$. In addition, by Lemma 2 we have

$$N(X|Y) + \min_{p_{\hat{X}|Y}} N(\hat{X}|Y) \leq U(\infty), \tag{4}$$

where we use the property that $N(-X|Y) = N(X|Y)$. As $\min_{p_{\hat{X}|Y}} N(\hat{X}|Y) \geq 0$ we obtain

$$\forall P \geq 0: \quad N(X|Y) \leq U(P). \tag{5}$$

For the upper bound we have $U(P) \leq U(0) = N(\hat{X}_0 - X|Y)$ where $p_{\hat{X}_0|Y} = p_{X|Y}$. Define $V \triangleq \hat{X}_0 - X$, then, we obtain $\Sigma_{v|y} = \Sigma_{\hat{x}|y} + \Sigma_{x|y} = 2\Sigma_{x|y}$ where we use that X and $\hat{X}$ are independent given Y. Thus,

$$U(0) = N(V|Y) \leq N(V_G|Y) = \left\lvert\Sigma_{v|y}\right\rvert^{1/d} = \left\lvert 2\Sigma_{x|y}\right\rvert^{1/d} = 2\left\lvert\Sigma_{x|y}\right\rvert^{1/d} = 2N(X_G|Y), \tag{6}$$

where the inequality above is due to Lemma 1. For any $P \geq 0$ it holds that $U(P) \leq U(0)$ which implies $U(P) \leq 2N(X_G|Y)$.

Finally, we prove the third claim. Assuming $D_v(X, \hat{X}\big|Y)$ is convex in its second argument, the constraint represent a compact, convex set. Moreover, $h(\hat{X} - X|Y)$ is strictly-concave w.r.t $p_{\hat{X}|Y}$ as a composition of a linear function (convolution) with a strictly-concave function (entropy). Therefore, we minimize a log-concave function over a convex domain and thus the global minimum is attained on the set boundary where $D_v(X, \hat{X}\big|Y) = P$.

C Proof of Theorem 2

We begin with applying Lemma 1 and Lemma 2 to bound the objective function as follows

$$N(\hat{X}_g|Y) + N(X_g|Y) = N(\hat{X}_g - X_g|Y) \leq N(\hat{X} - X|Y) \leq N(\hat{X}_G - X_G|Y). \tag{7}$$

Note that the bounds are tight as the upper bound is attained when $\hat{X}|Y$ and $X|Y$ are multivariate Gaussian random variables, while the lower bound is attained if we further assume they are isotropic. Thus, we can bound the uncertainty-perception function as follows

$$U_g(P) \leq U(P) \leq U_G(P) \tag{8}$$

where we define

$$\begin{aligned} U_g(P) &\triangleq \min_{p_{\hat{X}_g|Y}} \Big\{ N(\hat{X}_g|Y) + N(X_g|Y) \ : \ D_{1/2}(X_g, \hat{X}_g\big|Y) \leq P \Big\}, \\ U_G(P) &\triangleq \min_{p_{\hat{X}_G|Y}} \Big\{ N(\hat{X}_G - X_G|Y) \ : \ D_{1/2}(X_G, \hat{X}_G\big|Y) \leq P \Big\}. \end{aligned} \tag{9}$$

The above quantities can be expressed in closed form. We start with minimization problem of the upper bound which can be written as

$$U_G(P) = \min_{p_{\hat{X}_G|Y}} \left\{ \frac{1}{2\pi e} e^{\frac{2}{d}\mathbb{E}[h(X_G - X_G|Y=y)]} \ : \ \mathbb{E}\left[D_{1/2}(X_G, \hat{X}_G\big|Y=y) \right] \leq P \right\}, \tag{10}$$

where the expectation is over $y \sim Y$. Substituting the expressions for $h(X_G - X_G|Y = y)$ and $D_{1/2}(X_G, \hat{X}_G\big|Y = y)$, we get

$$U_G(P) = \min_{\{\Sigma_{\hat{x}|y}\}} \left\{ \frac{1}{2\pi e} e^{\frac{2}{d}\mathbb{E}\left[\frac{1}{2}\log\left\{(2\pi e)^d \left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right|\right\}\right]} \ : \ \mathbb{E}\left[\log \frac{\left|\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)/2\right|}{\sqrt{\left|\Sigma_{\hat{x}|y}\right| \left|\Sigma_{x|y}\right|}} \right] \leq P \right\}. \tag{11}$$

Notice the optimization is with respect to the covariance matrices $\{\Sigma_{\hat{x}|y}\}$. Simplifying the above, we can equivalently solve the following minimization

$$\min_{\{\Sigma_{\hat{x}|y}\}} \mathbb{E}\left[\log\left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right|\right] \ \text{s.t.} \ \mathbb{E}\left[\log \frac{\left|\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)/2\right|}{\sqrt{\left|\Sigma_{\hat{x}|y}\right| \left|\Sigma_{x|y}\right|}} \right] \leq P. \tag{12}$$

The solution of a constrained optimization problem can be found by minizmiation the Lagrangian

$$L\left(\{\Sigma_{\hat{x}|y}\}, \lambda\right) \triangleq \mathbb{E}\left[\log\left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right|\right] + \lambda \left(\mathbb{E}\left[\log \frac{\left|\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)/2\right|}{\sqrt{\left|\Sigma_{\hat{x}|y}\right| \left|\Sigma_{x|y}\right|}} \right] - P \right). \tag{13}$$

Since expectation is a linear operation and using that $P = \mathbb{E}[P]$, we rewrite the above as

$$L\left(\{\Sigma_{\hat{x}|y}\}, \lambda\right) = \mathbb{E}\left[\log\left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right| + \lambda \left(\log \frac{\left|\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)/2\right|}{\sqrt{\left|\Sigma_{\hat{x}|y}\right| \left|\Sigma_{x|y}\right|}} - P \right) \right]. \tag{14}$$

The expression within the expectation can be written as

$$\log\left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right| + \lambda \left(\log\left|\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)/2\right| - \frac{1}{2}\log\left|\Sigma_{\hat{x}|y}\right| - \frac{1}{2}\log\left|\Sigma_{x|y}\right| - P \right). \tag{15}$$

Next, according to KKT conditions the solutions should satisfy $\frac{\partial L}{\partial \Sigma_{\hat{x}|y}} = 0$. Using the linearity of the expectation and differentiating (15) w.r.t $\Sigma_{\hat{x}|y}$ we obtain

$$\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)^{-1} + \lambda \left(\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)^{-1} - \frac{1}{2}\Sigma_{\hat{x}|y}^{-1} \right) = 0 \tag{16}$$

Multiplying both sides by $\left(\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right)$, we have

$$\begin{aligned} & I + \lambda I - \frac{\lambda}{2} I - \frac{\lambda}{2}\Sigma_{x|y}\Sigma_{\hat{x}|y}^{-1} = 0 \\ \Rightarrow\ & (1 + \frac{\lambda}{2}) I = \frac{\lambda}{2}\Sigma_{x|y}\Sigma_{\hat{x}|y}^{-1} \\ \Rightarrow\ & (\lambda + 2)\Sigma_{\hat{x}|y} = \lambda \Sigma_{x|y} \\ \Rightarrow\ & \Sigma_{\hat{x}|y} = \frac{\lambda}{\lambda + 2}\Sigma_{x|y}. \end{aligned} \tag{17}$$

Define $\gamma = \frac{\lambda}{\lambda+2}$, so $\Sigma_{\hat{x}|y} = \gamma \Sigma_{x|y}$. Substituting the latter into the constraint we get

$$\begin{aligned} & \log\left|\left(\gamma\Sigma_{x|y} + \Sigma_{x|y}\right)/2\right| - \frac{1}{2}\log\left|\gamma\Sigma_{x|y}\right| - \frac{1}{2}\log\left|\Sigma_{x|y}\right| = P \\ \Rightarrow\ & n \log\frac{1+\gamma}{2} - \frac{n}{2}\log\gamma = P \\ \Rightarrow\ & \frac{(1+\gamma)^2}{4\gamma} = e^{\frac{2}{d}P} \\ \Rightarrow\ & \gamma^2 + 2\gamma + 1 = 4\gamma e^{\frac{2}{d}P} \\ \Rightarrow\ & \gamma(P) = 2e^{\frac{2}{d}P} - 1 - \sqrt{(2e^{\frac{2}{d}P} - 1)^2 - 1}. \end{aligned} \tag{18}$$

Thus, we obtain that

$$U_G(P) = \eta(P) \cdot N(X_G|Y) \tag{19}$$

where

$$\eta(P) = \gamma(P) + 1 = 2e^{\frac{2}{d}P} - \sqrt{(2e^{\frac{2}{d}P} - 1)^2 - 1}. \tag{20}$$

Notice that $\eta(0) = 2$, while $\lim_{P\to\infty}\ \eta(P) = 1$, so $1 \le \eta(P) \le 2$. Following similar steps where we replace $\Sigma_{\hat{x}|y}$ and $\Sigma_{x|y}$ with $N(\hat{X}|Y)$ and $N(X|Y)$ respectively, we derive

$$U_g(P) = \eta(P) \cdot N(X|Y). \tag{21}$$

D Proof of Theorem 3

Define $\mathcal{E} \triangleq \hat{X} - X$. Then,

$$\begin{aligned} \frac{1}{d}\mathbb{E}\left[||\hat{X} - X||^2\right] &\underset{(a)}{=} \mathbb{E}\left[\frac{1}{d}\mathbb{E}\left[||\hat{X} - X||^2 \middle| Y\right]\right] = \mathbb{E}\left[\frac{1}{d}\mathbb{E}\left[||\mathcal{E}||^2 \middle| Y\right]\right] = \mathbb{E}\left[\frac{1}{d}\mathbb{E}\left[\mathcal{E}^T\mathcal{E} \middle| Y\right]\right] \\ &= \mathbb{E}\left[\frac{1}{d}Tr\left(\mathbb{E}\left[\mathcal{E}\mathcal{E}^T \middle| Y\right]\right)\right] = \mathbb{E}\left[\frac{1}{d}Tr\left(\Sigma_{\varepsilon|y}\right)\right] \\ &\underset{(b)}{\ge} \mathbb{E}\left[\left|\Sigma_{\varepsilon|y}\right|^{1/d}\right] = \mathbb{E}\left[\left|\Sigma_{\hat{x}|y} + \Sigma_{x|y}\right|^{1/d}\right] \\ &\underset{(c)}{\ge} \mathbb{E}\left[\frac{1}{2\pi e}e^{\frac{2}{d}h(\hat{X} - X|Y=y)}\right] \\ &\underset{(d)}{\ge} \frac{1}{2\pi e}e^{\frac{2}{d}\mathbb{E}\left[h(\hat{X} - X|Y=y)\right]} = \frac{1}{2\pi e}e^{\frac{2}{d}h(\hat{X} - X|Y)} = N\left(\hat{X} - X\middle|Y\right), \end{aligned}$$

where (a) is by the law of total expectation, (b) is due to the inequality of arithmetic and geometric means, (c) follows Lemma 1, and (d) is according to Jensen's inequality.