A. Upper Bound Proofs

A.1. Proof of Lemma 3.1

Lemma 3.1. *For $c \in [0,1]$, let $H := (1-c)H_0 + cH_1$ be a mixture of two absolutely continuous distributions H_0, H_1 admitting densities h_0, h_1. Let y be a sample from the distribution H, such that $y|z^* \sim H_{z^*}$ where $z^* \sim Bernoulli(c)$.*

Define $\widehat{c}_y = \frac{ch_1(y)}{(1-c)h_0(y)+ch_1(y)}$, and let $\widehat{z}|y \sim Bernoulli(\widehat{c}_y)$ be the posterior sampling of z^* given y. Then we have

$$\Pr_{z^*,y,\widehat{z}}[z^* = 0, \widehat{z} = 1] \leq 1 - TV(H_0, H_1).$$

Proof. We have

$$\Pr_{z^*,y,\widehat{z}}[z^* = 0, \widehat{z} = 1] = \Pr[z^* = 0] \mathop{\mathbb{E}}_{y \sim h_0, \widehat{z}|y}[1\{\widehat{z} = 1\}], \tag{7}$$

$$= (1-c) \int h_0(y) \Pr[\widehat{z} = 1|y] dy. \tag{8}$$

By definition, we have

$$\Pr[\widehat{z} = 1|y] = \frac{ch_1(y)}{(1-c)h_0(y) + ch_1(y)}.$$

Substituting, we have

$$\begin{aligned}
\Pr_{z^*,y,\widehat{z}}[z^* = 0, \widehat{z} = 1] &= \int \frac{(1-c)h_0(y)ch_1(y)}{(1-c)h_0(y) + ch_1(y)} dy \\
&\leq \int \frac{(1-c)h_0(y) \cdot ch_1(y)}{\max\{(1-c)h_0(y), ch_1(y)\}} dy \\
&= \int \min\{(1-c)h_0(y), ch_1(y)\} dy \\
&\leq \int \min\{h_0(y), h_1(y)\} dy \\
&= (1 - TV(H_0, H_1)).
\end{aligned}$$

□

A.2. Proof of Lemma 3.2

Lemma 3.2. *Let y be generated from x^* by a Gaussian measurement process with noise level σ. For a fixed $\tilde{x} \in \mathbb{R}^n$, and parameters $\eta > 0, c \geq 4e^2$, let P_{out} be a distribution supported on the set*

$$S_{\tilde{x},out} := \{x \in \mathbb{R}^n : \|x - \tilde{x}\| \geq c(\eta + \sigma)\}.$$

Let $P_{\tilde{x}}$ be a distribution which is supported within an $\eta-$radius ball centered at $\tilde{x}$.

For a fixed A, let $H_{\tilde{x}}$ denote the distribution of y when $x^* \sim P_{\tilde{x}}$. Let H_{out} denote the corresponding distribution of y when $x^* \sim P_{out}$. Then we have:

$$\mathop{\mathbb{E}}_{A}[TV(H_{\tilde{x}}, H_{out})] \geq 1 - 4e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

Proof. In order to prove the lemma, it suffices to show that on the set

$$B := \{y \in \mathbb{R}^m : \|y - A\tilde{x}\| \leq \sqrt{c}\,(\eta + \sigma)\},$$

we have

$$\mathbb{E}_A[H_{out}(B)] \leq 2e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}, \tag{9}$$

$$\mathbb{E}_A[H_{\tilde{x}}(B)] \geq 1 - 2e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}. \tag{10}$$

Using the above bounds, we can conclude that

$$\mathbb{E}_A\left[TV(H_{out}, H_{\tilde{x}})\right] \geq \mathbb{E}_A[H_{\tilde{x}}(B)] - \mathbb{E}_A[H_{out}(B)] \geq 1 - 4e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

First we prove Equation (9).

Consider the joint distribution of y, A. We have

$$\mathbb{E}_A\left[H_{out}(B)\right] = \mathbb{E}_A\left[\mathbb{E}_{x\sim P_{out}}\left[\mathcal{N}\left(Ax, \frac{\sigma^2}{m}I_m\right)(B)\right]\right], \tag{11}$$

$$= \mathbb{E}_{x\sim P_{out}}\left[\mathbb{E}_A\left[\mathcal{N}(Ax, \sigma^2/m)(B)\right]\right], \tag{12}$$

where the first line follows from the definition of H_{out} and the fact that x, A are independent. The last line follows by switching the order of integrating A, x. Here $\mathcal{N}(Ax, \sigma^2/m)(B)$ refers to the mass $\mathcal{N}(Ax, \sigma^2/m)$ places on B.

Consider a fixed $x \in S_{\tilde{x},out}$, that is, x lies in the support of P_{out} and satisfies $\|x - \tilde{x}\| \geq c(\eta + \sigma\sqrt{m})$. We split the above expectation into two conditions over the matrix A.

- Case 1: $\|Ax - A\tilde{x}\| \leq 2\sqrt{c}\,(\eta + \sigma)$. Since A is i.i.d. Gaussian, $A\,(x - \tilde{x})$ is distributed as $\mathcal{N}\left(0, \frac{\|x-\tilde{x}\|^2}{m}I_m\right)$. This gives

$$\Pr_A\left[\|Ax - A\tilde{x}\| < 2\sqrt{c}\,(\eta+\sigma)\right] \leq \Pr_A\left[\|Ax - A\tilde{x}\| \leq \frac{2}{\sqrt{c}}\|x - \tilde{x}\|\right],$$
$$\leq \frac{2}{\sqrt{m\pi}}\left(\frac{2e}{\sqrt{c}}\right)^m,$$
$$= \frac{2}{\sqrt{m\pi}}e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)},$$
$$\leq e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)} \quad \text{if } m > 1.$$

  This implies

$$\mathbb{E}_{x\sim P_{out}}\left[\mathbb{E}_A\left[\mathcal{N}(Ax, \sigma^2/m)(B)1_{\|Ax-A\tilde{x}\|<2\sqrt{c}(\eta+\sigma)}\right]\right] \leq \mathbb{E}_{x\sim P_{out}}\left[\mathbb{E}_A\left[1_{\|Ax-A\tilde{x}\|<2\sqrt{c}(\eta+\sigma)}\right]\right],$$
$$= \mathbb{E}_{x\sim P_{out}}\left[\Pr_A\left[\|Ax - A\tilde{x}\| \leq 2\sqrt{c}\,(\eta+\sigma)\right]\right],$$
$$\leq e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

- Case 2: $\|Ax - A\tilde{x}\| > 2\sqrt{c}\,(\eta + \sigma)$.

  Recall the definition of $B := \{y \in \mathbb{R}^m : \|y - A\tilde{x}\| \leq \sqrt{c}\,(\eta+\sigma)\}$. For any $y \in B$, x in the support of P_{out} and for A such that $\|Ax - A\tilde{x}\| > 2\sqrt{c}\,(\eta+\sigma)$, we have

$$\|y - Ax\| \geq \|Ax - A\tilde{x}\| - \|y - A\tilde{x}\| \geq 2\sqrt{c}\,(\eta+\sigma) - \sqrt{c}\,(\eta+\sigma) = \sqrt{c}\,(\eta+\sigma)\,.$$

  For each x in the support of P_{out}, define the set $B_x := \{y \in \mathbb{R}^m : \|y - Ax\| \geq \sqrt{c}\,(\eta+\sigma)\}$. The above inequality gives $B \subseteq B_x$ for each x in the support of P_{out}. This gives

$$\mathcal{N}(Ax, \sigma^2)(B) \leq \mathcal{N}(Ax, \sigma^2)(B_x) \leq e^{-2\left(\sqrt{c}-1\right)^2 m} \leq e^{-\frac{mc}{2}}.$$

  where the last inequality follows by the definition of B_x and Gaussian concentration of $\mathcal{N}(Ax, \sigma^2)$ on the set B_x, and since $2\left(\sqrt{c}-1\right)^2 > \frac{c}{2}$ if $c \geq 4$.

Substituting the inequalities from Case 1 and Case 2 in Eqn (12), we have

$$\begin{aligned}\mathop{\mathbb{E}}_{A}\left[H_{out}(B)\right] &= \mathop{\mathbb{E}}_{x\sim P_{out}}\left[\mathop{\mathbb{E}}_{A}\left[\mathcal{N}(Ax,\sigma^2/m)(B)\right]\right],\\ &\leq e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)} + e^{-\frac{cm}{2}},\\ &\leq 2e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)} \quad \text{if } c \geq 4e^2.\end{aligned}$$

This proves Eqn (9).

A similar proof can be used to show that

$$\mathop{\mathbb{E}}_{A}\left[H_{\tilde{x}}(B^c)\right] \leq 2e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

This proves Eqn (10).

Putting the two above inequalities together, we have

$$\mathop{\mathbb{E}}_{A} TV(H_{out}, H_{\tilde{x}}) \geq \mathop{\mathbb{E}}_{A}[H_{\tilde{x}}(B)] - \mathop{\mathbb{E}}_{A}[H_{out(B)}] \geq 1 - 4e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

This concludes the proof. □

A.3. Proof of Lemma A.1

Lemma A.1. *Let R, P be arbitrary distributions on $\mathbb{R}^n$. Let $p \geq 1$ and $\eta, \rho, \delta > 0$, be parameters.*

If $\mathcal{W}_p(R, P) \leq \rho$ and $\min\{\log \mathrm{Cov}_{\eta,\delta}(P), \log \mathrm{Cov}_{\eta,\delta}(R)\} \leq k$, then there exist distributions R', R'', P', P'', and a finite discrete distirbution Q with $|\operatorname{supp}(Q)| \leq e^k$ satisfying:

1. $\min\{\mathcal{W}_\infty(P', Q), \mathcal{W}_\infty(R', Q)\} \leq \eta$,
2. $\mathcal{W}_\infty(R', P') \leq \frac{\rho}{\delta^{1/p}}$,
3. $P = (1-2\delta)P' + (2\delta)P''$ *and* $R = (1-2\delta)R' + (2\delta)R''$

Proof. Since the statement of the lemma is symmetric with respect to P and R, WLOG let $\log \mathrm{Cov}_{\eta,\delta}(P) \leq k$. Then there is an $S \subset \mathbb{R}^n$ such that $|S| \leq e^k$ and

$$\Pr_{x\sim P}[x \in \cup_{u\in S} B(u,\eta)] = 1 - c_P \geq 1 - \delta,$$

We define the function $f : \mathbb{R}^n \to \mathbb{R}_+$ as

$$f(x) = \begin{cases} \frac{1}{|\{u\in S | x \in B(u,\eta)\}|} & \text{if } \exists u \in S\ s.t.\ x \in B(u,\eta), \\ 0 & \text{otherwise.}\end{cases}$$

By construction, f is a piecewise constant function that is inversely proportional to the number of $\eta-$radius balls centered around points in S cover a point x.

For each $u \in S$, we define the measure Q'' as

$$Q''(u) := \int_{B(u,\eta)} f\, dP.$$

Observe that

$$\sum_{u\in S} Q''(u) = \sum_{u\in S}\int_{B(u,\eta)} f dP,$$

$$= \int_{\cup_{u \in S} B(u,\eta)} dP = 1 - c_P$$

Notice that Q'' is not a probability distribution, since it only has mass $1 - c_P$. However we can create a distribution Q' from Q'' by putting an additional c_P mass on some arbitrary point in $\mathbb{R}^n$ (say, 0). By construction, there exists a coupling Π of P and Q' where the coupling distributes the mass at each point in $\mathbb{R}^n$ to points η close to it in S, such that

$$c_P = \Pr_{(x_1,x_2)\sim\Pi} [\|x_1 - x_2\| \geq \eta] \leq \delta. \tag{13}$$

Additionally, since $W_p(R, P) \leq \rho$, there exists a coupling Γ such that.

$$c_R = \Pr_{(x_1,x_2)\sim\Gamma} \left[\|x_1 - x_2\| \geq \frac{\rho}{\delta^{1/p}}\right] \leq \frac{\mathbb{E}\left[\|x_1 - x_2\|^p\right]}{\frac{\rho^p}{\delta}} \leq \delta. \tag{14}$$

where c_P is defined by the first equality. We can hence define a couple between P, Q', R whose distribution is given by the following – for any borel measurable sets B_1, B_2, B_3 we have $\Omega(B_1, B_2, B_3) = P(B_1)\Pi(B_2 \mid B_1)\Gamma(B_3 \mid B_1)$. To verify that this is indeed a coupling of the kind we want, we observe that the marginals of Ω are P, Q and R respectively.

1. $\Omega(B_1, \mathbb{R}^n, \mathbb{R}^n) = P(B_1)\Pi(\mathbb{R}^n \mid B_1)\Gamma(\mathbb{R}^n \mid B_1) = P(B_1)$.
2. $\Omega(\mathbb{R}^n, B_2, \mathbb{R}^n) = P(\mathbb{R}^n)\Pi(B_2 \mid \mathbb{R}^n)\Gamma(\mathbb{R}^n \mid \mathbb{R}^n) = 1 \cdot \frac{\Pi(B_2,\mathbb{R}^n)}{P(\mathbb{R}^n)} \cdot 1 = Q'(B_2)$.
3. $\Omega(\mathbb{R}^n, \mathbb{R}^n, B_3) = P(\mathbb{R}^n)\Pi(\mathbb{R}^n \mid \mathbb{R}^n)\Gamma(B_3 \mid \mathbb{R}^n) = R(B_3)$.

To define P', Q, R', we look at Ω conditioned on the event $E := \{(x, y, z) \mid \|x - z\| \leq \rho/\delta^{1/p} \text{ and } \|x - y\| \leq \eta\}$. To estimate the probability of E, we define $E_1 := \{(x, y, z) \mid z \in \mathbb{R}^n \text{ and } \|x - y\| > \eta\}$ and $E_2 := \{(x, y, z) \mid \|x - z\| > \rho/\delta^{1/p} \text{ and } y \in \mathbb{R}^n\}$. Then, $\overline{E} = E_1 \vee E_2$.

We now show that $\Omega(E_1) \leq \delta$. Let $(E_1)_I$ denote E_1 restricted to the coordinates in I.

$$\Omega(E_1) := P((E_1)_1)\Pi((E_1)_{1,2} \mid (E_1)_1)\Gamma((E_1)_{1,3} \mid (E_1)_1) \leq \Pi((E_1)_{1,2}) \leq \delta,$$

where the first inequality is because $\Gamma((E_1)_{1,3} \mid (E_1)_1) \leq 1$ and $\Pi((E_1)_{1,2} \mid (E_1)_1) = \Pi((E_1)_{1,2})/P((E_1)_1)$ and the final inequaity follows from equation (13). The bound for E_2 follows similarly. A union bound shows that $\Omega(E) \geq 1 - 2\delta$. We can restrict the event E further to have mass $1 - 2\delta$.

We look at the marginals of the conditional couple $\Omega(\cdot \mid E)$ to get distributions P', Q, R' as follows. We define $P'(\cdot) := \Omega(\cdot, \mathbb{R}^n, \mathbb{R}^n \mid E)$, $Q(\cdot) := \Omega(\mathbb{R}^n, \cdot, \mathbb{R}^n \mid E)$ and $R'(\cdot) := \Omega(\mathbb{R}^n, \mathbb{R}^n, \cdot \mid E)$. P'' and R'' are defined similarly via conditioning on $\overline{E}$. Hence, $P(\cdot) = \Omega(\cdot, \mathbb{R}^n, \mathbb{R}^n) = \Omega(E)\Omega(\cdot, \mathbb{R}^n, \mathbb{R}^n \mid E) + \Omega(\overline{E})\Omega(\cdot, \mathbb{R}^n, \mathbb{R}^n \mid \overline{E}) = (1 - 2\delta)P'(\cdot) + (2\delta)P''(\cdot)$. The statement for R follows similarly.

This finally gives distributions P', R', Q, such that:

1. $\mathcal{W}_\infty(P', Q) \leq \eta$
2. $\mathcal{W}_\infty(R', P') \leq \rho/\delta^{1/p}$
3. $P = (1 - 2\delta)P' + (2\delta)P''$ and $R = (1 - 2\delta)R' + (2\delta)R''$.

The first two statements follow because of the event we condition over.

Note that this restriction does not change the fact that $\text{supp}(Q) < e^k$, and hence we have our result.

□

A.4. Proof of Lemma 3.3

Lemma 3.3. *Let R, P, denote arbitrary distributions over $\mathbb{R}^n$ such that $\mathcal{W}_\infty(R, P) \leq \varepsilon$.*

Let $x^* \sim R$ and $z^* \sim P$ and let y and u be generated from x^* and z^* via a Gaussian measurement process with m measurements and noise level σ. Let $\widehat{x} \sim P(\cdot|y, A)$ and $\widehat{z} \sim P(\cdot|u, A)$. For any $d > 0$, we have

$$\Pr_{x^*, A, \xi, \widehat{x}} [\|x^* - \widehat{x}\| \geq d + \varepsilon] \leq$$
$$e^{-\Omega(m)} + e^{\left(\frac{4\varepsilon(\varepsilon+2\sigma)m}{2\sigma^2}\right)} \Pr_{z^*, A, \xi, \widehat{z}} [\|z^* - \widehat{z}\| \geq d] .$$

Proof. Let B_1 denote the event

$$B_1 = \{\|x^* - \widehat{x}\| \geq d + \varepsilon\} .$$

Similarly, let B_2 denote the event

$$B_2 = \{\|z^* - \widehat{x}\| \geq d\} .$$

We have

$$\Pr_{x^* \sim R, A, \xi, \widehat{x} \sim P(\cdot|A, y)} [B_1] = \mathop{\mathbb{E}}_{x^* \sim R} \mathop{\mathbb{E}}_{A} \left[\mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_1}] \right] \right] .$$

We can write the integral over R as an integral over the coupling Π between R, P. This gives

$$\Pr_{x^*, A, \xi, \widehat{x} \sim P(\cdot|A, y)} [B_1] = \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[\mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_1}] \right] \right] .$$

Since x^*, z^* are coupled and $W_\infty(R, P) \leq \varepsilon$, we have $\|x^* - z^*\| \leq \varepsilon$ almost surely. This gives $B_1 \subseteq B_2$ if x^*, z^* are distributed according to Π. Hence,

$$\Pr_{x^*, A, \xi, \widehat{x} \sim P(\cdot|A, y)} [B_1] \leq \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[\mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_2}] \right] \right] .$$

We can split the above integral into two parts: one where the matrix A satsifies $\|Ax^* - Az^*\| \leq 2\varepsilon$, and another case where $\|Ax^* - Az^*\| > 2\varepsilon$. This gives

$$\Pr_{x^*, A, \xi, \widehat{x} \sim P(\cdot|A, y)} [B_1] \leq \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[1_{\|Ax^* - Az^*\| > 2\varepsilon} \mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_2}] \right] \right] (*) \tag{15}$$

$$+ \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[1_{\|Ax^* - Az^*\| \leq 2\varepsilon} \mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_2}] \right] \right] .(**) \tag{16}$$

Consider the term$(*)$ in line (15). We have

$$\mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[1_{\|Ax^* - Az^*\| > 2\varepsilon} \mathop{\mathbb{E}}_{y|A, x^*} \left[\mathop{\mathbb{E}}_{\widehat{x} \sim P(\cdot|y, A)} [1_{B_2}] \right] \right] \leq \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[1_{\|Ax^* - Az^*\| > 2\varepsilon} \right] , \tag{17}$$

$$\leq \mathop{\mathbb{E}}_{x^*, z^*} \left[e^{-\Omega(m)} \right] \leq e^{-\Omega(m)}, \tag{18}$$

where the last inequality follows from the Johnson-Lindenstrauss lemma for a fixed x^*, z^*, and hence is true on average over x^*, z^* drawn independent of A.

Now consider the term $(**)$ in line (16). Notice that since the noise in the measurements is Gaussian, we have

$$y|x^*, A \sim \mathcal{N}(Ax^*, \sigma^2/m).$$

We break the integral over y in $(**)$ into two cases:

1. Case 1: $\|y - Ax^*\| > 2\sigma$. Since $p(y|A, x^*)$ is distributed as $\mathcal{N}\left(Ax^*, \frac{\sigma^2}{m} I_m\right)$, by standard Gaussian concentration, we have

$$\int_{y:\|y-Ax^*\|>2\sigma} p(y|A, x^*) dy \leq e^{-\Omega(m)}.$$

2. Case 2: $\|y - Ax^*\| \leq 2\sigma$. This gives

$$\begin{aligned}\|Ax^* - y\|^2 &= \|Ax^* - y\|^2 - \|y - Az^*\|^2 + \|y - Az^*\|^2, \\ &= \|Ax^* - y\|^2 - \|y - Ax^* + Ax^* - Az^*\|^2 + \|y - Az^*\|^2, \\ &= -\|Ax^* - Az^*\|^2 - 2\langle y - Ax^*, Ax^* - Az^*\rangle + \|y - Az^*\|^2.\end{aligned}$$

Observe that in $(**)$, we have

$$\|Ax^* - Az^*\| \leq 2\varepsilon \Rightarrow \|Ax^* - Az^*\|^2 \leq 4\varepsilon^2.$$

By the Cauchy-Schwartz inequality and the assumption that $\|y - Ax^*\| \leq 2\sigma$, we have

$$2\langle y - Ax^*, Ax^* - Az^*\rangle \leq 8\sigma\varepsilon.$$

Substituting the above two inequalities, we have

$$\|Ax^* - y\|^2 \geq -4\varepsilon^2 - 8\sigma\varepsilon + \|y - Az^*\|^2, \tag{19}$$

$$\Rightarrow \exp\left(-\frac{\|Ax^* - y\|^2}{2\sigma^2/m}\right) \leq \exp\left(\frac{4\varepsilon\,(\varepsilon + 2\sigma)\,m}{2\sigma^2}\right) \exp\left(-\frac{\|Az^* - y\|^2}{2\sigma^2/m}\right), \tag{20}$$

(21)

Observe that the LHS has the density of measurements from x^*, while the RHS has the density of measurements from z^* with an exponential scaling. From the above inequality, we can replace the expectation over $y|A, x^*$ in $(**)$ with $u|A, z^*$ with an exponential factor.

Similarly, since posterior sampling now uses u in place of y, we can replace $\widehat{x}$ in $(**)$ with $\widehat{z}$.

Combining Case 1 and 2 gives

$$\begin{aligned}(**) &\leq e^{-\Omega(m)} + e^{\left(\frac{4\varepsilon(\varepsilon+2\sigma)m}{2\sigma^2}\right)} \mathop{\mathbb{E}}_{x^*, z^*} \mathop{\mathbb{E}}_{A} \left[\mathop{\mathbb{E}}_{u|A,z^*} \left[\mathop{\mathbb{E}}_{\widehat{z}\sim P(\cdot|u,A)} [1_{B_2}]\right]\right], \\ &= e^{-\Omega(m)} + e^{\left(\frac{4\varepsilon(\varepsilon+2\sigma)m}{2\sigma^2}\right)} \mathop{\mathbb{E}}_{z^*\sim P} \mathop{\mathbb{E}}_{A} \left[\mathop{\mathbb{E}}_{u|A,z^*} \left[\mathop{\mathbb{E}}_{\widehat{z}\sim P(\cdot|u,A)} [1_{B_2}]\right]\right].\end{aligned}$$

From the above inequality and eqn. (18), we have

$$\Pr_{x^*\sim R,\xi,A,\widehat{x}\sim P(\cdot|A,y)} [\|x^* - \widehat{x}\| \geq d + \varepsilon] \leq e^{-\Omega(m)} + e^{\left(\frac{4\varepsilon(\varepsilon+2\sigma)m}{2\sigma^2}\right)} \Pr_{z^*\sim P,\xi,A,\widehat{z}\sim P(\cdot|u,A)} [\|z^* - \widehat{z}\| \geq d].$$

□

A.5. Proof of Theorem 3.4

Theorem 3.4. *Let $\delta \in [0, 1/4)$, $p \geq 1$, and $\varepsilon, \eta > 0$ be parameters. Let R, P be arbitrary distributions over $\mathbb{R}^n$ satisfying $\mathcal{W}_p(R, P) \leq \varepsilon$.*

Let $x^* \sim R$ and suppose y is generated by a Gaussian measurement process from x^* with noise level $\sigma \gtrsim \varepsilon/\delta^{1/p}$ and $m \geq O(\min(\log \mathrm{Cov}_{\eta,\delta}(R), \log \mathrm{Cov}_{\eta,\delta}(P)))$ measurements. Given y and the fixed matrix A, let $\widehat{x}$ output of posterior sampling with respect to P.

Then there exists a universal constant $c > 0$ such that with probability at least $1 - e^{-\Omega(m)}$ over A, ξ,

$$\Pr_{x^* \sim R, \widehat{x} \sim P(\cdot|y)} [\|x^* - \widehat{x}\| \geq c\eta + c\sigma] \leq 2\delta + 2e^{-\Omega(m)}.$$

Proof. We know from Lemma A.1 that there exist R', P', R'', P'' and a finite distribution Q supported on the set S such that

1. $\mathcal{W}_\infty(R', P') \leq \frac{\varepsilon}{\delta^{1/p}}$,
2. $\min\{\mathcal{W}_\infty(P', Q), \mathcal{W}_\infty(R', Q)\} \leq \eta$,
3. $R = (1-2\delta)R' + 2\delta R''$ and $P = (1-2\delta)P' + 2\delta P''$,
4. $|S| \leq e^k$.

Suppose $\mathcal{W}_\infty(P', Q) \leq \eta$. If not, then $\mathcal{W}_\infty(R', Q) \leq \eta$, and by (1), we see that $\mathcal{W}_\infty(P', Q) \leq \eta + \frac{\varepsilon}{\delta^{1/p}}$, and we will use this in the proof instead. This gives us

$$\begin{aligned}\Pr_{x^* \sim R, \widehat{x} \sim P(\cdot|y)} [\|x^* - \widehat{x}\| \geq (c+1)\eta + (c+1)\sigma] &\leq \Pr_{x^* \sim R, \widehat{x} \sim P(\cdot|y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\eta + c\sigma + (\varepsilon/\delta^{1/p})\right] \\ &\leq 2\delta + (1-2\delta) \Pr_{x^* \sim R', \widehat{x} \sim P(\cdot|y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\eta + c\sigma + (\varepsilon/\delta^{1/p})\right], \end{aligned} \tag{22}$$

where the first line follows since $\sigma \geq \varepsilon/\delta^{\frac{1}{p}}$, and the second line follows by decomposing $R = (1-2\delta)R' + 2\delta R''$.

We now bound the second term on the right hand side of the above equation. For this term, consider the joint distribution over $x^*, A, \xi, \widehat{x}$. By Lemma 3.3, we can replace $x^* \sim R'$ with $z^* \sim P'$, replace $y = Ax^* + \xi$ with $u = Az^* + \xi$, and replace $\widehat{x} \sim P(\cdot|A, y)$ with $\widehat{z} \sim P(\cdot|A, u)$ to get the following bound

$$\begin{aligned}&\Pr_{x^* \sim R', A, \xi, \widehat{x} \sim P(\cdot|A,y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\,\eta + c\sigma + (\varepsilon/\delta^{1/p})\right] \leq \\ &e^{-\Omega(m)} + e^{\left(\frac{2(\varepsilon/\delta^{1/p})\left((\varepsilon/\delta^{1/p}) + 2\sigma\right)m}{\sigma^2}\right)} \Pr_{z^* \sim P', A, \xi, \widehat{z} \sim P(\cdot|u,A)} [\|z^* - \widehat{z}\| \geq (c+1)\eta + c\sigma]\,. \end{aligned} \tag{23}$$

We now bound the second term in the right hand side of the above inequality. Let Γ denote an optimal $\mathcal{W}_\infty-$coupling between P' and Q.

For each $\tilde{z} \in S$, the conditional coupling can be defined as

$$\Gamma(\cdot|\tilde{z}) = \frac{\Gamma(\cdot, \tilde{z})}{Q(\tilde{z})}.$$

By the $\mathcal{W}_\infty$ condition, each $\Gamma(\cdot|\tilde{z})$ is supported on a ball of radius η around $\tilde{z}$.

Let $E = \{z^*, \widehat{z} \in \mathbb{R}^n : \|z^* - \widehat{z}\| \geq (c+1)\,\eta + c\sigma\}$ denote the event that $z^*, \widehat{z}$ are far apart. By the coupling, we can express P' as

$$P' = \sum_{\tilde{z} \in S} Q(\tilde{z})\Gamma(\cdot|\tilde{z}).$$

This gives

$$\Pr_{z^* \sim P', A, \xi, \widehat{z} \sim P(\cdot|A,u)} [E] = \sum_{\tilde{z}^* \in S} Q(\tilde{z}^*) \mathop{\mathbb{E}}_{z^* \sim \Gamma(\cdot|\tilde{z}^*), A, \xi, \widehat{z} \sim P(\cdot|A,u)} [1_E] .$$

For each $\tilde{z}^* \in S$, we now bound $Q(\tilde{z}^*) \, \mathbb{E}_{z^* \sim \Gamma(\cdot|\tilde{z}^*), A, \xi, \widehat{z} \sim P(\cdot|A,u)} [1_E]$.

For each $\tilde{z}^* \in S$, we can write P as $P = (1-2\delta) \, Q_{\tilde{z}^*} P_{\tilde{z}^*,0} + c_{\tilde{z}^*,1} P_{\tilde{z}^*,1} + c_{\tilde{z}^*,2} P_{\tilde{z}^*,2}$, where the components of the mixture are defined in the following way. The first component $P_{\tilde{z}^*,0}$ is $\Gamma(\cdot|\tilde{z}^*)$, the second component is supported within a $c(\eta+\sigma)$ radius of $\tilde{z}^*$, and the third component is supported outside a $c\,(\eta+\sigma)$ radius of $\tilde{z}^*$.

Formally, let $B_{\tilde{z}^*}$ denote the ball of radius $c(\eta+\sigma)$ centered at $\tilde{z}^*$, and let $B^c_{\tilde{z}^*}$ be its complement. The constants are defined via the following Lebesque integrals, and the mixture components for any Borel measurable B are defined as

$$c_{\tilde{z}^*,1} := \int_{B_{\tilde{z}^*}} dP - (1-2\delta) \, Q_{\tilde{z}^*} \int_{B_{\tilde{z}^*}} d\Gamma(\cdot|\tilde{z}^*),$$

$$c_{\tilde{z}^*,2} := \int_{B^c_{\tilde{z}^*}} dP - (1-2\delta) \, Q_{\tilde{z}^*} \int_{B^c_{\tilde{z}^*}} d\Gamma(\cdot|\tilde{z}^*),$$

$$P_{\tilde{z}^*,0}(B) := \Gamma(B \cap B_{\tilde{z}^*}|\tilde{z}^*) = \Gamma(B|\tilde{z}^*) \text{ since } \operatorname{supp}(\Gamma(\cdot|\tilde{z}^*)) \subset B_{\tilde{z}^*},$$

$$P_{\tilde{z}^*,1}(B) := \begin{cases} \frac{1}{c_{\tilde{z}^*,1}} P(B \cap B_{\tilde{z}^*}) - \frac{1-2\delta}{c_{\tilde{z}^*,1}} Q_{\tilde{z}^*} \Gamma(B \cap B_{\tilde{z}^*}|\tilde{z}^*) & \text{if } c_{\tilde{z}^*,1} > 0, \\ \text{do not care} & \text{otherwise.} \end{cases},$$

$$P_{\tilde{z}^*,2}(B) := \begin{cases} \frac{1}{c_{\tilde{z}^*,2}} P(B \cap B^c_{\tilde{z}^*}) - \frac{1-2\delta}{c_{\tilde{z}^*,2}} Q_{\tilde{z}^*} \Gamma(B \cap B^c_{\tilde{z}^*}|\tilde{z}^*) & \text{if } c_{\tilde{z}^*,2} > 0, \\ \text{do not care} & \text{otherwise.} \end{cases}.$$

Notice that if z^* is sampled from $\Gamma(\cdot|\tilde{z}^*)$, then by the W_∞ condition, we have $\|z^* - \tilde{z}^*\| \leq \eta$. Furthermore, if $\widehat{z}$ is $(c+1)\,\eta + c\sigma$ far from z^*, an application of the triangle inequality implies that it must be distributed according to $P_{\tilde{z}^*,2}$. That is,

$$\begin{aligned} Q(\tilde{z}^*) \mathop{\mathbb{E}}_{z^* \sim \Gamma(\cdot|\tilde{z}^*), A, \xi, \widehat{z} \sim P(\cdot|A,u)} [1_E] &\leq \mathop{\mathbb{E}}_{A,\xi,z^*} \Pr\left[z^* \sim P_{\tilde{z}^*,0}, \widehat{z} \sim P_{\tilde{z}^*,2}(\cdot|u)\right] \\ &\leq \frac{1}{1-2\delta} \mathop{\mathbb{E}}_{A} \left[1 - TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2})\right], \end{aligned}$$

where $H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2}$ are the push-forwards of $P_{\tilde{z}^*,0}, P_{\tilde{z}^*,2}$ for A fixed and the last inequality follows from Claim A.2.

Notice that if we sum over all $\tilde{z}^* \in S$, then the LHS of the above inequality is an expectation over $z^* \sim P'$. This gives:

$$\Pr_{z^* \sim P', A, \xi, \widehat{z} \sim P(\cdot|u,A)} [E] \leq \frac{1}{1-2\delta} \sum_{\tilde{z}^* \in S} \mathop{\mathbb{E}}_{A} \left[1 - TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2})\right].$$

Notice that $P_{\tilde{z}^*,0}$ is supported within an $\eta-$ball around $\tilde{z}^*$, and $P_{\tilde{z}^*,2}$ is supported outside a $c(\eta+\sigma)-$ball of $\tilde{z}^*$. By Lemma 3.2 we have

$$\mathop{\mathbb{E}}_{A}[TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2})] \geq 1 - 4e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)}.$$

This implies

$$\Pr_{z^* \sim P', A, \xi, \widehat{z} \sim P(\cdot|u,A)} \left[\|z^* - \widehat{z}\| \geq (c+1)\eta + c\sigma\right] \leq \frac{1}{1-2\delta} \sum_{\tilde{z}^* \in S} \mathop{\mathbb{E}}_{A} \left[(1 - TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2}))\right],$$

$$\leq \frac{1}{1-2\delta} 4|S| e^{-\frac{m}{2}\log\left(\frac{c}{4e^2}\right)},$$
$$\leq \frac{1}{1-2\delta} 4 e^{-\frac{m}{4}\log\left(\frac{c}{4e^2}\right)},$$

where the last inequality is satisfied if $m \geq 4\log(|S|)$.

Substituting in Eqn (23), if $c > 4\exp\left(2 + \frac{8(\varepsilon/\delta^{1/p})\left((\varepsilon/\delta^{1/p})+2\sigma\right)}{\sigma^2}\right)$, we have

$$\Pr_{x^* \sim R', A, \xi, \widehat{x} \sim P(\cdot|A,y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\,\eta + c\sigma + (\varepsilon/\delta^{1/p})\right] \leq e^{-\Omega(m)} + \frac{1}{1-2\delta} e^{-\Omega(m\log c)}.$$

This implies that there exists a set $S_{A,\xi}$ over A, ξ satisfying $\Pr_{A,\xi}[S_{A,\xi}] \geq 1 - e^{-\Omega(m)}$, such that for all $A, \xi \in S_{A,\xi}$, we have

$$\Pr_{x^* \sim R', \widehat{x} \sim P(\cdot|y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\eta + c\sigma + (\varepsilon/\delta^{1/p})\right] \leq \frac{1}{1-2\delta} e^{-\Omega(m)}.$$

Substituting in Eqn (22), we have

$$\Pr_{x^* \sim R, \widehat{x} \sim P(\cdot|y)} \left[\|x^* - \widehat{x}\| \geq (c+1)\eta + c\sigma + (\varepsilon/\delta^{1/p})\right] \leq 2\delta + \frac{1}{1-2\delta} e^{-\Omega(m)} \leq 2\delta + 2e^{-\Omega(m)}.$$

Rescaling c gives us our result.

At the beginning of the proof, we had assumed that $\mathcal{W}_\infty(P', Q) \leq \eta$. If instead $\mathcal{W}_\infty(R', Q) \leq \eta$, then we need to replace η in the above bound by $\eta + \frac{\varepsilon}{\delta^{1/p}}$. Rescaling c in the above bound gives us the Theorem statement.

□

Claim A.2. *Consider the setting of the previous theorem. We have*

$$\mathop{\mathbb{E}}_{A,\xi,z^*} \Pr\left[z^* \sim P_{\tilde{z}^*,0}, \widehat{z} \sim P_{\tilde{z}^*,2}(\cdot|u)\right] \leq \frac{1}{1-\delta_2} \mathop{\mathbb{E}}_{A}\left[1 - TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2})\right], \tag{24}$$

Proof. For a fixed A, let h_0, h_2 denote the corresponding densities of the push forward of $P_{\tilde{z}^*,0}, P_{\tilde{z}^*,2}$. Then we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{A,\xi,z^*} \Pr\left[z^* \sim P_{\tilde{z}^*,0}, \widehat{z} \sim P_{\tilde{z}^*,2}(\cdot|u)\right] &= \mathop{\mathbb{E}}_{A} \int \frac{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)}{(1-\delta_2)\, Q_{\tilde{z}^*,0} h_{\tilde{z}^*,0}(u) + c_{\tilde{z}^*,1} h_{\tilde{z}^*,1}(u) + c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)} du, &(25)\\
&\leq \mathop{\mathbb{E}}_{A} \int \frac{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)}{(1-\delta_2)\, Q_{\tilde{z}^*,0} h_{\tilde{z}^*,0}(u) + c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)} du, &(26)\\
&\leq \mathop{\mathbb{E}}_{A} \int \frac{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)}{(1-\delta_2)\, Q_{\tilde{z}^*,0} h_{\tilde{z}^*,0}(u) + (1-\delta_2) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)} du, &(27)\\
&\leq \mathop{\mathbb{E}}_{A} \frac{1}{1-\delta_2} \int \frac{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)}{Q_{\tilde{z}^*,0} h_{\tilde{z}^*,0}(u) + c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)} du, &(28)\\
&\leq \mathop{\mathbb{E}}_{A} \frac{1}{1-\delta_2} \int \frac{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u) c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)}{\max\{Q_{\tilde{z}^*,0} h_{\tilde{z}^*,0}(u)\,,\, c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)\}} du, &(29)\\
&= \mathop{\mathbb{E}}_{A} \frac{1}{1-\delta_2} \int \min\{Q_{\tilde{z}^*} h_{\tilde{z}^*,0}(u), c_{\tilde{z}^*,2} h_{\tilde{z}^*,2}(u)\} du, &(30)\\
&\leq \mathop{\mathbb{E}}_{A} \frac{1}{1-\delta_2} \int \min\{h_{\tilde{z}^*,0}(u), h_{\tilde{z}^*,2}(u)\} du, &(31)\\
&= \frac{1}{1-\delta_2} \mathop{\mathbb{E}}_{A}\left[1 - TV(H_{\tilde{z}^*,0}, H_{\tilde{z}^*,2})\right]. &(32)
\end{aligned}$$

□

B. Lower Bound Proofs

B.1. Proof of Lemma 4.2

Lemma 4.2. *Consider the setting of Theorem (4.1). If A is a deterministic matrix, we have*

$$I(y;x^*) \le \frac{m}{2}\log\left(1+\frac{mr^2\|A\|_\infty^2}{\sigma^2}\right).$$

If A is a Gaussian matrix, then $I(y;x^*|A) \le \frac{m}{2}\log\left(1+\frac{r^2}{\sigma^2}\right)$.

Proof. First, we consider the case where A is a deterministic matrix.

We have $y = Ax^* + \xi$. Let $z = Ax^*$, which gives $y = z + \xi$.

We have $z_i = a_i^T x^*$ where a_i is the i^{th} row of A, and $y_i = z_i + \xi_i$. Since x^* is supported within the sphere of radius r, we have $\mathbb{E}[z_i^2] = \mathbb{E}[\langle a_i, x\rangle^2] \le \|a_i\|^2 r^2$. Since the Gaussian noise ξ has variance σ^2/m in each coordinate, every coordinate of y_i is a Gaussian channel with power constaint $\|a_i\|^2 r^2$ and noise variance σ^2/m. Using Shannon's AWGN theorem (Cover & Thomas, 2012; Polyanskiy & Wu, 2014; Shannon, 1948), the mutual information between y_i, z_i, is bounded by

$$I(y_i;z_i) \le \frac{1}{2}\log\left(1+\frac{\|a_i\|^2 r^2 m}{\sigma^2}\right).$$

The chain rule of entropy and sub-addditivity of entropy implies,

$$\begin{aligned}
I(y;z) &= h(y) - h(y|z) = h(y) - h(y-z|z),\\
&= h(y) - h(\xi|z) = h(y) - \sum h(\xi_i|z,\xi_1,\cdots,\xi_{i-1}),\\
&= h(y) - \sum h(\xi_i),\\
&\le \sum h(y_i) - \sum h(\xi_i),\\
&= \sum h(y_i) - \sum h(y_i|z_i),\\
&= \sum I(y_i;z_i),\\
&\le \sum_{i=1}^{m} \frac{1}{2}\log\left(1+\frac{\|a_i\|^2 r^2 m}{\sigma^2}\right),\\
&\le \frac{m}{2}\log\left(1+\frac{mr^2\|A\|_\infty^2}{\sigma^2}\right).
\end{aligned}$$

Since $x^* \to z \to y$ is a Markov chain, we can conclude that

$$I(y;x^*) \le I(y;z) \le \frac{m}{2}\log\left(1+\frac{mr^2\|A\|_\infty^2}{\sigma^2}\right).$$

Now, if A is a Gaussian matrix with i.i.d. entries drawn from $\mathcal{N}(0,1/m)$, then the power constraint is $\mathbb{E}[\langle a_i, x\rangle^2] \le r^2/m$. This gives us

$$I(y;z) \le \frac{m}{2}\log\left(1+\frac{r^2}{\sigma^2}\right). \tag{33}$$

Now since A is a random matrix, we cannot directly apply the Data Processing Inequality of x^*, y, z as before, and need to prove that $I(x^*;y|A) \le I(y;z)$.

Consider the mutual information $I(x^*, A, z; y)$. By the chain rule of mutual information, we have

$$I(x^*, A, z; y) = I(A;y) + I(x^*;y|A) + I(z;y|x^*, A),$$

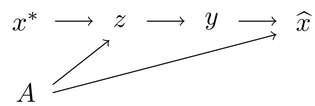

Figure 6: DAG relating $x^*, A, z, y, \widehat{x}$. The conditional independencies we use are $x^* \perp\!\!\!\perp y|z, A$ and $A \perp\!\!\!\perp y|z$.

$$
\begin{aligned}
&= I(A; y) + I(z; y|A) + I(x^*; y|z, A), \\
\Leftrightarrow I(x^*; y|A) + I(z; y|x^*, A) &= I(z; y|A) + I(x^*; y|z, A).
\end{aligned}
$$

From Figure 6, note that x^*, y, are conditionally independent given z, A. This gives $I(x^*; y|z, A) = 0$.

This gives

$$
\begin{aligned}
I(x^*; y|A) + I(z; y|x^*, A) &= I(z; y|A), && (34) \\
\Rightarrow I(x^*; y|A) &\leq I(z; y|A). && (35)
\end{aligned}
$$

We can bound $I(z; y|A)$ in the following way.

$$
\begin{aligned}
I(A, z; y) &= I(A; y) + I(z; y|A), && (36) \\
&= I(z; y) + I(A; y|z), && (37) \\
\Leftrightarrow I(A; y) + I(z; y|A) &= I(z; y) + I(A; y|z), && (38) \\
\Leftrightarrow I(A; y) + I(z; y|A) &= I(z; y), && (39) \\
\Rightarrow I(z; y|A) &\leq I(z; y), . && (40)
\end{aligned}
$$

where the second last line follows from $I(A; y|z) = 0$, and the last line follows from $I(A; y) \geq 0$.

From Eqn (33), (35), (40), we have

$$
I(x^*; y|A) \leq \frac{m}{2}\left(1 + \frac{mr^2\|A\|_\infty^2}{\sigma^2}\right).
$$

□

B.2. Proof of Lemma 4.3

Lemma 4.3. *Consider the setting of Theorem (4.1). If A is a deterministic matrix, we have $I(x^*; \widehat{x}) \leq I(y; x^*)$.*

If A is a random matrix, then $I(x^*; \widehat{x}) \leq I(y; x^*|A)$.

Proof. When A is a deterministic matrix, the proof follows directly from the Data Processing Inequality (Cover & Thomas, 2012). Since $x^* \to y \to \widehat{x}$ is a Markov chain, we get

$$
I(x^*; \widehat{x}) \leq I(y; x^*).
$$

Now when A is a random matrix, we need to show $I(x^*; \widehat{x}) \leq I(y; x^*|A)$. Consider the mutual information $I(x^*; y, A, \widehat{x})$. By the chain rule of mutual information, we can express it in two ways:

$$
\begin{aligned}
I(x^*; y, A, \widehat{x}) &= I(x^*; y, A) + I(x^*; \widehat{x}|y, A), && (41) \\
&= I(x^*; \widehat{x}) + I(x^*; y, A|\widehat{x}). && (42)
\end{aligned}
$$

As $\widehat{x}$ is a function of y, A, we have $I(x^*; \widehat{x}|y, A) = 0$. Also, $I(x^*; y, A|\widehat{x}) \geq 0$. Substituting in Eqn (41), (42), we have

$$
I(x^*; \widehat{x}) \leq I(x^*; y, A),
$$

$$= I(x^*; A) + I(x^*; y|A),$$
$$= I(x^*; y|A),$$

where the second line follows from the chain rule of mutual information, and the last line follows because x^*, A, are independent.

□

B.3. Proof of Fano variant Lemma 4.4

We will build up Lemma 4.4 in sequence. Before showing it in its full generality, we will show when $x, \widehat{x}$, are discrete random variables and x is uniform (Lemma B.1. We then lift the uniformity restriction on x (Lemma B.3) before extending to continuous distributions (Lemma 4.4).

Lemma B.1. *Let Q be the uniform distribution over an arbitrary discrete finite set S. Let $(x, \widehat{x})$ be jointly distributed, where $x \sim Q$ and $\widehat{x}$ is distributed over an arbitrary countable set, satisfying*

$$\Pr\left[\|x - \widehat{x}\| \leq \varepsilon\right] \geq 1 - \delta.$$

Then for all $\tau \in (0, 1)$, we have

$$\tau(1 - \delta) \log \mathrm{Cov}_{2\varepsilon, \delta+\tau}(Q) \leq I(x; \widehat{x}) + 2.$$

Proof. Let $E = \mathbf{1}\{\|x - \widehat{x}\| \leq \varepsilon\}$ be the indicator random variable for x and $\widehat{x}$ being close.

Via claim B.2, we get

$$H(x|E = 1) \geq \log |S| - \frac{1}{1 - \delta}. \tag{43}$$

Recall,

$$I(x; \widehat{x}|E = 1) = H(x|E = 1) - H(x|\widehat{x}, E = 1)$$

By the Law of total probability, we have:

$$I(x; \widehat{x}|E = 1) = \sum_v \Pr[\widehat{x} = v|E = 1] \left(H(x|E = 1) - H(x|\widehat{x} = v, E = 1)\right).$$

We would like to apply a version of Markov's inequality to the above equation. However, the terms in the summation could be negative. However, from (43) we have that $H(x|E = 1) + \frac{1}{1-\delta} \geq \log |S|$. Furthermore, since x is supported on a discrete set of cardinality $|S|$, we have $H(x|\widehat{x} = v, E = 1) \leq \log |S|$. Adding and subtracting $\frac{1}{1-\delta}$, in the above equation, we have

$$\begin{aligned}
I(x; \widehat{x}|E = 1) &= \sum_v \Pr[\widehat{x} = v|E = 1] \left(H(x|E = 1) + \frac{1}{1 - \delta} - H(x|\widehat{x} = v, E = 1) - \frac{1}{1 - \delta}\right), \\
&= \sum_v \Pr[\widehat{x} = v|E = 1] \left(H(x|E = 1) + \frac{1}{1 - \delta} - H(x|\widehat{x} = v, E = 1)\right) - \frac{1}{1 - \delta}, \\
\Leftrightarrow I(x; \widehat{x}|E = 1) + \frac{1}{1 - \delta} &= \sum_v \Pr[\widehat{x} = v|E = 1] \left(H(x|E = 1) + \frac{1}{1 - \delta} - H(x|\widehat{x} = v, E = 1)\right)
\end{aligned}$$

Since the above summation has only non-negative terms that average to $I(x; \widehat{x}|E = 1) + \frac{1}{1-\delta}$, for all $\tau \in (0, 1)$, there exists $G_1 \subseteq \mathrm{supp}(\widehat{x})$ with $\Pr[G_1|E = 1] \geq 1 - \tau$, such that for all $v \in G_1$, we have

$$H(x|E = 1) + \frac{1}{1 - \delta} - H(x|\widehat{x} = v, E = 1) \leq \frac{I(x; \widehat{x}|E = 1) + \frac{1}{1-\delta}}{\tau}.$$

From (43), we have $H(x|E=1) + \frac{1}{1-\delta} \geq \log|S|$. Hence for all $v \in G_1$, we have

$$
\begin{aligned}
\log|S| - H(x|\widehat{x}=v, E=1) &\leq \frac{I(x;\widehat{x}|E=1) + \frac{1}{1-\delta}}{\tau}, \\
\Leftrightarrow H(x|\widehat{x}=v, E=1) &\geq \log|S| - \frac{I(x;\widehat{x}) + \frac{1}{1-\delta}}{\tau}, \\
\Rightarrow \log|\mathrm{supp}(x|\widehat{x}=v, E=1)| &\geq \log|S| - \left(\frac{I(x;\widehat{x}) + \frac{1}{1-\delta}}{\tau}\right), \\
\Rightarrow \log|S \cap B(v,\varepsilon)| &\geq \log|S| - \left(\frac{I(x;\widehat{x}) + \frac{1}{1-\delta}}{\tau}\right), \qquad (44)
\end{aligned}
$$

where the last inequality follows as conditioned on $E=1$, x must be supported on an ε-radius ball around $\widehat{x}$.

Now consider the set $G_2 = (S \times G_1) \wedge E_1$. That is, $G_2 \subseteq \mathrm{supp}(x, \widehat{x})$, such that $(u,v) \in G_2$ if and only if $\|u-v\| \leq \epsilon$ and $u \in S, v \in G_1$. Since $\Pr[E_1] \geq 1-\delta$ by the statement of the lemma, and $\Pr[G_1|E_1] \geq 1-\tau$ by construction, we have

$$\Pr[G_2] \geq (1-\delta)(1-\tau) \geq 1-\delta-\tau.$$

Now for all $(u,v) \in G_2$, we have

$$
\begin{aligned}
&\|u-v\| \leq \varepsilon, \\
&\log|S \cap B(v,\varepsilon)| \geq \log|S| - \left(\frac{I(x;\widehat{x}|E=1) + \frac{1}{1-\delta}}{\tau}\right). \qquad (45)
\end{aligned}
$$

Note that by the construction of G_2, the set $\bigcup_{v \in G_2} B(v,\varepsilon)$ covers a $1-\delta-\tau$ fraction of S. As each ball $B(v,\varepsilon)$ also has a large intersection with S, by the pigeon-hole principle, any 2ε-packing of this $1-\delta-\tau$ fraction of S must have size at most $2^{\left(I(x;\widehat{x}|E=1)+\frac{1}{1-\delta}\right)/\tau}$.

Hence, we can find a 2ε-cover of a $1-\delta-\tau$ fraction of S that has size at most $2^{\left(I(x;\widehat{x}|E=1)+\frac{1}{1-\delta}\right)/\tau}$.

This gives

$$\log \mathrm{Cov}_{2\varepsilon,\delta+\tau}(Q) \leq \frac{I(x;\widehat{x}|E=1) + \frac{1}{1-\delta}}{\tau}. \qquad (46)$$

We are almost done, since we now only need to relate $I(x;\widehat{x}|E=1)$ to $I(x;\widehat{x})$.

By the chain rule of mutual information, we have

$$
\begin{aligned}
I(x;\widehat{x},E) &= I(x;\widehat{x}) + I(x;E|\widehat{x}) = I(x;E) + I(x;\widehat{x}|E), \\
\Rightarrow I(x;\widehat{x}|E) &\leq I(x;\widehat{x}) + I(x;E|\widehat{x}), \\
&\leq I(x;\widehat{x}) + 1, \\
\Leftrightarrow I(x;\widehat{x}|E=0)\Pr[E=0] + I(x;\widehat{x}|E=1)\Pr[E=1] &\leq I(x;\widehat{x}) + 1, \\
\Rightarrow I(x;\widehat{x}|E=1) &\leq \frac{I(x;\widehat{x})+1}{1-\delta}.
\end{aligned}
$$

Substituting in Eqn (46), we have

$$
\begin{aligned}
\log \mathrm{Cov}_{2\varepsilon,\delta+\tau}(Q) &\leq \frac{I(x;\widehat{x})+2}{\tau(1-\delta)}, \\
\Rightarrow \tau(1-\delta)\log \mathrm{Cov}_{2\varepsilon,\delta+\tau}(Q) &\leq I(x;\widehat{x}) + 2.
\end{aligned}
$$

□

Claim B.2. *Let $x \sim Q$, where Q is the uniform distribution over an arbitrary discrete finite set S. Let E be a binary random variable such that $\Pr[E = 1] \geq 1 - \delta$.*

Then we have

$$H(x|E = 1) \geq \log |S| - \frac{1}{1 - \delta}.$$

Proof. Let $p = \Pr[E = 1]$. By the definition of conditional entropy, we have

$$\begin{aligned}
H(x|E) &= (1-p)H(x|E=0) + pH(x|E=1),\\
\Leftrightarrow H(x|E=1) &= \frac{1}{p}(H(x|E) - (1-p)H(x|E=0)),\\
&= \frac{1}{p}(H(x) - I(x;E) - (1-p)H(x|E=0)),\\
&= \frac{1}{p}(\log|S| - I(x;E) - (1-p)H(x|E=0)),\\
&\geq \frac{1}{p}(\log|S| - I(x;E) - (1-p)\log|S|),\\
&= \log|S| - \frac{I(x;E)}{p},\\
&\geq \log|S| - \frac{1}{1-\delta},
\end{aligned}$$

where the fourth line follows from $H(x) = \log |S|$ since x is uniform, the fifth line follows from $H(x|E = 0) \leq \log |S|$ since x is supported on a discrete set of size $|S|$, and the last line follows from $p \geq 1 - \delta$ and $I(x; E) \leq H(E) \leq 1$. □

The previous lemma handled the uniform distribution on x. Now we show that a similar result applies if x's distribution has quantized probability values.

Lemma B.3. *Let Q be a finite discrete distribution over $N \in \mathbb{N}$ points such that for each u in its support, $Q(u) = j\alpha$, where $j \in \mathbb{N}$ and $\alpha := \frac{1}{N_2}$ is a discretization level for $N_2 \in \mathbb{N}$ large enough.*

Let $(x, \widehat{x})$ be jointly distributed, where $x \sim Q$ and $\widehat{x}$ is distributed over a countable set, satisfying

$$\Pr[\|x - \widehat{x}\| \leq \varepsilon] \geq 1 - \delta.$$

Then we have

$$\tau(1 - \delta) \log \mathrm{Cov}_{2\varepsilon,\tau+\delta}(Q) \leq I(x; \widehat{x}) + 2\delta.$$

Proof. For each x in the support of Q, we know that its probability is an integral multiple of $\frac{1}{N_2}$. Hence we can define a new random variable $x' = (x, j), x \in \mathrm{supp}(Q), j \in [N_2]$ and a distribution Q' over x' in the following way:

$$Q'((x, j)) = \begin{cases} \alpha & \text{if } j\alpha \leq Q(x), \\ 0 & \text{otherwise}. \end{cases}$$

By definition, Q' is a uniform distribution, and its support is a discrete subset of $\mathbb{R}^n \times \mathbb{N}$.

Define the following norm for x'. For $x'_1 = (x_1, j_1), x'_2 = (x_2, j_2)$, define

$$\|(x_1, j_1) - (x_2, j_2)\| := \|x_1 - x_2\|.$$

In order to apply Lemma B.1 on Q', it suffices to show that $I(x; \widehat{x}) = I(x'; \widehat{x})$.

By the chain rule of mutual information, we have

$$I(x'; \widehat{x}) = I((x, j); \widehat{x})$$

$$= I(x;\widehat{x}) + I(j;\widehat{x}|x).$$

Since $\widehat{x}$ is purely a function of x, we have $I(j;\widehat{x}|x) = 0$. This gives

$$I(x';\widehat{x}) = I(x;\widehat{x}).$$

Similarly construct a version $\widehat{x}' = (\widehat{x}, 0)$ of $\widehat{x}$, whose second coordinate is identically zero. Hence for $x' = (x, j) \sim Q'$, we have

$$\|x' - \widehat{x}'\| \le \varepsilon \text{ w.p. } 1-\delta,$$
$$I(x';\widehat{x}') = I(x;\widehat{x})$$

Applying Lemma B.1 on Q', we have

$$\tau(1-\delta)\log \mathrm{Cov}_{2\varepsilon,\tau+\delta}(Q') \le I(x;\widehat{x}) + 2.$$

Since the support of the first coordinate of Q' is the same as the support of Q, we have

$$\tau(1-\delta)\log \mathrm{Cov}_{2\varepsilon,\tau+\delta}(Q) \le I(x;\widehat{x}) + 2.$$

□

We now prove Lemma 4.4, which allows $(x,\widehat{x})$ to follow an arbitrary distribution.

Lemma 4.4 (Fano variant)**.** *Let $(x,\widehat{x})$ be jointly distributed over $\mathbb{R}^n \times \mathbb{R}^n$, where $x \sim R$ and $\widehat{x}$ satisfies*

$$\Pr[\|x - \widehat{x}\| \le \eta] \ge 1 - \delta.$$

Then for any $\tau \le 1 - 3\delta, \delta < 1/3$, we have

$$0.99\tau(1-2\delta)\log \mathrm{Cov}_{3\eta,\tau+3\delta}(R) \le I(x;\widehat{x}) + 1.98.$$

Proof. Let $\varepsilon = \eta$, which is the error in the statement of the lemma. Let $\gamma > 0$ be a small enough discretization level to be specified later. For every $x, \widehat{x} \in \mathbb{R}^n$, let $\bar{x}, \widehat{\bar{x}}$ denoted the rounding of $x, \widehat{x}$ to the nearest multiple of γ in each coordinate.

Let $\bar{R}$ be the discrete distribution induced by this discretization of x. We can create such a distribution by assigning the probability of each cell in the grid to its corresponding coordinate-wise floor. This discretization of the support changes the error between $x, \widehat{x}$ in the following way. If $\|x - \widehat{x}\| \le \varepsilon$ with probability $1-\delta$, an application of the triangle inequality gives

$$\|\bar{x} - \widehat{\bar{x}}\| \le \varepsilon + 2\gamma\sqrt{n} \text{ with probability } \ge 1 - \delta. \tag{47}$$

We also need to take into account the effect discretizing $x, \widehat{x}$ has on their mutual information. Note that since $\bar{x}$ is a function of x alone, and $\widehat{\bar{x}}$ is a function of $\widehat{x}$ alone, by the Data Processing Inequality, we have

$$I(\bar{x};\widehat{\bar{x}}) \le I(x;\widehat{x}). \tag{48}$$

Note that $\bar{R}$ is a distribution on a discrete but infinite set. However, for any $\beta \in (0,1]$, we can find a discrete and finite distribution Q such that $\bar{R} = (1-c_1)Q + c_1 D$, with $c_1 \le \beta$ and D is some other probability distribution. This is feasible because the probabilites of the infinite support of $\bar{R}$ must sum to 1, and hence we can find a finite subset that sums to atleast $1-\beta$ for any $\beta \in (0,1]$. Note that in this process, we only change the marginal of $\bar{x}$ without changing the conditional distribution of $\widehat{\bar{x}}|\bar{x}$. Let $I(\bar{x};\widehat{\bar{x}}), I_Q(\bar{x};\widehat{\bar{x}}), I_D(\bar{x};\widehat{\bar{x}})$ denote the mutual information between $\bar{x}, \widehat{\bar{x}}$ when the marginal of $\bar{x}$ is $\bar{R}, Q, D$, respectively. From Theorem 2.7.4 in (Cover & Thomas, 2012), mutual information is a concave function of the marginal distribution of $\bar{x}$ for a fixed conditional distribution of $\widehat{\bar{x}}|\bar{x}$. An application of Eqn (48) gives us,

$$I(x;\widehat{x}) \ge I(\bar{x};\widehat{\bar{x}}) \ge (1-c_1)I_Q(\bar{x};\widehat{\bar{x}}) + c_1 I_D(\bar{x};\widehat{\bar{x}}), \tag{49}$$

$$\geq (1-c_1) I_Q(\bar{x};\widehat{\bar{x}}), \tag{50}$$

$$\geq (1-\beta) I_Q(\bar{x};\widehat{\bar{x}}). \tag{51}$$

Now since the finite distribution Q has a TV distance of at most β to the countable distribution R, using Eqn (47), we have

$$\|\bar{x}-\widehat{\bar{x}}\| \leq \varepsilon + 2\gamma\sqrt{n} \text{ with probability } \geq 1-\beta-\delta \text{ if } \bar{x}\sim Q. \tag{52}$$

In order to apply Lemma B.3 on the distribution Q, we need its probability values to be multiples of some discretization level α. Let α be a small enough quantization level for the probability values. We will specify the value of α later. We can now express the distribution Q as a mixture of two distributions Q', Q''. The distribution Q' is obtained by flooring the probability values under Q and renormalizing to make them sum to 1. The distribution Q'' is the mass not contained in Q', normalized to sum to 1. Since each element in the support of Q loses at most α mass, the total mass in Q'' prior to normalization is at most αN_β, where N_β is the cardinaltiy of the support of Q. This gives

$$Q = (1-c_2)Q' + c_2 Q'',\ c_2 \leq \alpha N_\beta.$$

From Eqn (52), we have $\|\bar{x}-\widehat{\bar{x}}\| \leq \varepsilon + 2\gamma\sqrt{n}$ with probability $\geq 1-\beta-\delta$ when $\bar{x}\sim Q$. Since Q' has a TV distance of at most αN_β to Q, if $\bar{x}\sim Q'$, we have

$$\|\bar{x}-\widehat{\bar{x}}\| \leq \varepsilon + 2\gamma\sqrt{n} \text{ with probability } \geq 1-\beta-\delta-\alpha N_\beta \text{ if } \bar{x}\sim Q'. \tag{53}$$

Let $I_Q(\bar{x};\widehat{\bar{x}}), I_{Q'}(\bar{x};\widehat{\bar{x}}), I_{Q''}(\bar{x};\widehat{\bar{x}})$ denote the mutual information between $\bar{x},\widehat{\bar{x}}$ when the marginal of $\bar{x}$ is Q, Q', Q'' respectively. Mutual information is a concave function of the marginal distribution of $\bar{x}$ for a fixed conditional distribution of $\widehat{\bar{x}}|\bar{x}$. Hence using Eqn (51), we have

$$\frac{I(x;\widehat{x})}{1-\beta} \geq I_Q(\bar{x};\widehat{\bar{x}}) \geq (1-c_2) I_{Q'}(\bar{x};\widehat{\bar{x}}) + c_2 I_{Q''}(\bar{x};\widehat{\bar{x}}), \tag{54}$$

$$\geq (1-c_2) I_{Q'}(\bar{x};\widehat{\bar{x}}), \tag{55}$$

$$\geq (1-\alpha N_\beta) I_{Q'}(\bar{x};\widehat{\bar{x}}). \tag{56}$$

Hence if $\bar{x}\sim Q'$, we have $I(\bar{x};\widehat{\bar{x}}) \leq \frac{I(x;\widehat{x})}{(1-\alpha N_\beta)(1-\beta)}$. Applying Lemma B.3 on the distribution Q', for any $\tau > 0$, we have

$$\tau(1-\beta-\delta-\alpha N_\beta)\log \mathrm{Cov}_{2\varepsilon+4\gamma\sqrt{n},\tau+\beta+\delta+\alpha N_\beta}(Q') \leq \frac{I(x;\widehat{x})}{(1-\alpha N_\beta)(1-\beta)} + 2.$$

Now since Q' has at least $1-\alpha N_\beta$ of the mass under Q and Q has at least $1-\delta$ of the mass under $\bar{R}$, the mass $\tau+\beta+\delta+\alpha N_\beta$ not covered under Q' can be replaced with $\tau+\beta+2\delta+2\alpha N_\beta$ under $\bar{R}$. This gives

$$\tau(1-\beta-\delta-\alpha N_\beta)\log \mathrm{Cov}_{2\varepsilon+4\gamma\sqrt{n},\tau+\beta+2\delta+2\alpha N_\beta}(\bar{R}) \leq \frac{I(x;\widehat{x})}{(1-\alpha N_\beta)(1-\beta)} + 2.$$

Now since we can cover the whole distribution of R by extending each element in the support of $\bar{R}$ by γ in each coordinate, we can replace the radius $2\varepsilon+4\gamma\sqrt{n}$ for $\bar{R}$ by $2\varepsilon+6\gamma\sqrt{n}$ for R. This gives

$$\tau(1-\beta-\delta-\alpha N_\beta)\log \mathrm{Cov}_{2\varepsilon+6\gamma\sqrt{n},\tau+\beta+2\delta+2\alpha N_\beta}(R) \leq \frac{I(x;\widehat{x})}{(1-\alpha N_\beta)(1-\beta)} + 2.$$

For $\gamma = \frac{\varepsilon}{6\sqrt{n}}, \beta = \min\left\{\frac{\delta}{3}, 1-\sqrt{0.99}\right\}, \alpha N_\beta = \min\left\{\frac{\delta}{3}, 1-\sqrt{0.99}\right\}$, we have

$$0.99\tau(1-2\delta)\log \mathrm{Cov}_{3\varepsilon,\tau+3\delta}(R) \leq I(x;\widehat{x}) + 1.98.$$

□

B.4. Proof of Theorem 4.1

Theorem 4.1. *Let R be a distribution supported on a ball of radius r in $\mathbb{R}^n$, and $x^* \sim R$. Let $y = Ax^* + \xi$, where A is any matrix, and $\xi \sim \mathcal{N}(0, \frac{\sigma^2}{m} I_m)$. Assuming $\delta < 0.1$, if there exists a recovery scheme that uses y and A as inputs and guarantees*

$$\|\hat{x} - x^*\| \leq O(\eta),$$

with probability $\geq 1 - \delta$, then we have

$$m \geq \frac{0.15}{\log\left(1+\frac{mr^2\|A\|_\infty^2}{\sigma^2}\right)} \left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log 6\delta - O(1)\right).$$

If A is an i.i.d. Gaussian matrix where each element is drawn from $\mathcal{N}(0, 1/m)$, then the above bound can be improved to:

$$m \geq \frac{0.15}{\log\left(1 + \frac{r^2}{\sigma^2}\right)} \left(\log \mathrm{Cov}_{3\eta,4\delta}(R) + \log 6\delta - O(1)\right).$$

Proof. Throughout the proof, we use the notation $N(R, \delta)$ to denote a minimal set of 3η-radius balls that cover at least $1 - \delta$ mass under the distribution R.

Let B be the ball in $N(R, 10\delta)$ with smallest marginal probability. If we set $S \leftarrow N(R, 10\delta) \setminus B$, then S contains smaller than $1 - 10\delta$ of R.

Let $R = (1 - c)R' + cR''$, where the components R' and R'' are probability distributions restricted to S and its complement S^c respectively. By the construction of S, we have $c > 10\delta$. Note that since R'' contributes at least 10δ to R, any algorithm that succeeds with probability $\geq 1 - \delta$ over R must succeed with probability ≥ 0.9 over R''.

Now consider $x \sim R''$. By Lemma 4.3 and Lemma 4.2, we have

$$\begin{aligned} I(x; \widehat{x}) &\leq I(x; y|A), \\ &\leq \frac{m}{2} \log\left(1 + \frac{r^2}{\sigma^2}\right). \end{aligned}$$

Applying Lemma 4.4 on R'' with parameters $\tau = \delta = 0.1$, for the failure probability, we can conclude that

$$\begin{aligned} 0.99 \cdot 0.1 \cdot (1 - 0.2) \log |N(R'', 0.4)| \leq I(x; \widehat{x}) + 1.98 \leq \frac{m}{2} \log\left(1 + \frac{r^2}{\sigma^2}\right) + 1.98, \\ \Leftrightarrow m \geq \frac{0.1584\ \log |N(R'', 0.4)| - 3.96}{\log\left(1 + \frac{r^2}{\sigma^2}\right)}. \end{aligned} \tag{57}$$

We now need to express the covering number of R'' in terms of the covering number of R.

Note that as R'' contains at least 10δ mass under R, $N(R'', 0.4)$ contains at least 6δ mass under R, . Similarly, since $N(R, 10\delta)$ contains at least $1 - 10\delta$ mass under R, $N(R'', 0.4) \cup N(R, 10\delta)$ will contain at least 4δ mass under R. Hence, we get

$$|N(R'', 0.4)| + |N(R, 10\delta)| \geq |N(R, 4\delta)| \Leftrightarrow |N(R'', 0.4)| \geq |N(R, 4\delta)| - |N(R, 10\delta)|. \tag{58}$$

Now we need to relate $N(R, 4\delta)$ with $N(R, 10\delta)$. This can be accomplished via a simple counting argument. Assume that the balls in $N(R, 4\delta)$ are ordered in decreasing order of their marginal probability, then the last $\frac{10\delta}{1-4\delta}$-fraction of balls in $N(R, 4\delta)$ must contain at most 10δ mass. This implies that the first $\frac{1-10\delta}{1-4\delta}$-fraction of $N(R, 4\delta)$ must contain at least $1 - 10\delta$ mass. This gives:

$$\frac{1 - 10\delta}{1 - 4\delta} N(R, 4\delta) \geq N(R, 10\delta). \tag{59}$$

Combining Eqn (58), (59), we get

$$\begin{aligned} |N(R'', 0.4)| &\geq |N(R, 4\delta)| - \frac{1-10\delta}{1-4\delta}|N(R, 4\delta)|, \\ &= \frac{6\delta}{1-4\delta}|N(R, 4\delta)|, \\ &\geq 6\delta|N(R, 4\delta)|, \\ \Leftrightarrow \log|N(R'', 0.4)| &\geq \log|N(R, 4\delta)| + \log(6\delta). \end{aligned}$$

Substituting in Eqn (57), we get

$$m \geq \frac{0.1584\left(\log|N(R, 4\delta)| + \log(6\delta)\right) - 3.96}{\log\left(1 + \frac{r^2}{\sigma^2}\right)}.$$

Since $|N(R, 4\delta)| = \mathrm{Cov}_{3\eta, 4\delta}(R)$ by definition, this completes the proof. □

C. Experimental Setup

C.1. Datasets and Architecture

For the compressed sensing experiment in Fig 4a and the inpainting experiment in Figure 2 we used the 256×256 GLOW model (Kingma & Dhariwal, 2018) from the official repository. The test set for Fig 4a consists of the first 10 images used by (Asim et al., 2019) in their experiments.

For the compressed sensing experiment in Fig 1, 5a, 5b, we used the FFHQ NCSNv2 model (Song & Ermon, 2020) from the official repository. The test set for Fig 5a consists of the images 69000-69017 from the FFHQ dataset (this corresponds to the first 18 images in the last batch of FFHQ images).

In Fig 4a and Fig 5a, the measurements have noise satisfying $\sqrt{\mathbb{E}\,\|\xi\|^2} = 16$ and $\sqrt{\mathbb{E}\,\|\xi\|^2} = 4$ respectively.

C.2. Hyperparameter Selection

CelebA experiments For MAP, we used an Adam and Gradient Descent optimizer. Langevin dynamics only uses Gradient Descent. Each algorithm was run with learning rates varying over $\left[0.1, 0.01, 0.001, 5 \cdot 10^{-4}, 10^{-4}, 5 \cdot 10^{-5}, 10^{-5}, 5 \cdot 10^{-6}, 10^{-6}\right]$. For MAP and Modified-MAP, we also performed 2 random restarts for the initialization z_0.

The value of γ in Eqn (6) was varied over $[0, 0.1, 0.01, 0.001]$ for Modified-MAP. MAP uses the theoretically defined value of $\frac{\sigma^2}{m}$.

For Langevin dynamics, we vary the value of σ_i according to the schedule proposed by (Song & Ermon, 2019). We start with $\sigma_1 = 16.0$, and finish with $\sigma_{10} = 4.0$, such that σ_i decreases geometrically for $i \in [10]$. For each value of i, we do 200 steps of noisy gradient descent, with the learning rate schedule proposed by (Song & Ermon, 2019).

In order to select the optimal hyperparameters for each m, we chose the hyperparams that give maximum likelihood for Langevin and MAP. For Modified-MAP, we selected the hyperparameters based on reconstruction error on a holdout set of 5 images.

FFHQ experiments The NCSNv2 model is designed for Langevin dynamics. It can be adapted to MAP by simply not adding noise at each gradient step. We tune the initial and final values of σ used in (Song & Ermon, 2020), along with the initial learning rate.

Unfortunately, it is computationally difficult to obtain the likelihood associated with each reconstruction, since the NCSNv2 model only provides $\nabla \log p(x)$. Although one could, in theory, do numerical integration to find $p(x)$, we selected the optimal hyperparameters for each m based on reconstruction error on a holdout set of 5 images.

For the Deep-Decoder, we used the over-parameterized network described in (Asim et al., 2019), and tuned the learning rate over $[0.4, 0.004, 0.0004]$, and selected the hyper-parameters that optimized the reconstruction error on a holdout set of 5 images.

C.3. Computing Infrastructure

Experiments were run on an NVIDA Quadro P5000.