# 6. Appendix

## Symbols

$N(\delta, \Theta, ||\cdot||_q)$ $\delta$-cover of $\Theta$ w.r.t. $||\cdot||_q$ norm

$M(\delta, \Theta, ||\cdot||_q)$ $\delta$-packing of $\Theta$ w.r.t. $||\cdot||_q$ norm

$B_q^k(r)$ $k$-dimensional ball of radius $r$ w.r.t. $||\cdot||_q$ norm

$S_1 \oplus S_2$ Minkowski sum of the sets $S_1, S_2$, i.e. the set $\{x+y | x \in S_1, y \in S_2\}$

$\mathcal{G}(T)$ Gaussian complexity of set $T$

$[M]$ Set $\{1, ..., M\}$

### 6.1. Proofs

#### 6.1.1. Metric entropy for the $l_1$ ball

**Definition 2** (Covering number (Wainwright, 2019)). *A $\delta$-covering of a set $\mathbb{T}$ with respect to a metric $\rho$ is a set $\{\theta^1, ..., \theta^M\} \subset \mathbb{T}$ such that for each $\theta \in \mathbb{T}$, there exists some $i \in [N]$ such that $\rho(\theta, \theta^i) \leq \delta$. The $\delta$-covering number $N(\delta, \mathbb{T}, \rho)$ is the cardinality of the smallest $\delta$-cover.*

**Definition 3** (Packing number (Wainwright, 2019)). *A $\delta$-packing of a set $\mathbb{T}$ with respect to a metric $\rho$ is a set $\{\theta^1, ..., \theta^M\} \subset \mathbb{T}$ such that $\rho(\theta^i, \theta^j) > \delta$ for all distinct $i, j \in [M]$. The $\delta$-packing number $M(\delta, \mathbb{T}, \rho)$ is the cardinality of the largest $\delta$-packing.*

**Lemma 2** (Wainwright (2019)). *For all $\delta > 0$, the packing and covering numbers are related as follows:*

$$M(2\delta, \mathbb{T}, \rho) \leq N(\delta, T, \rho) \leq M(\delta, T, \rho). \tag{7}$$

**Theorem 2** (Maurey's Empirical Method (Pisier, 1986)). *Let $B_1^d(r) = \{x \in \mathbb{R}^d \mid ||x||_1 \leq r\}$. Then,*

$$\log N(\delta, B_1^d(r), ||\cdot||_2) \leq \frac{r^2}{\delta^2} \log(2d+1). \tag{8}$$

A short proof of this result follows.

*Proof.* Fix $x \in \mathbb{R}^d$. Let $Z$ be the following RV:

$$Z = \begin{cases} \text{sgn}(x_i) r e_i, & \text{w.p. } \frac{|x_i|}{r} \\ 0, & \text{w.p.} 1 - \frac{||x||_1}{r} \end{cases} \tag{9}$$

Observe that: $E[Z_i] = \text{sgn}(x_i) r \cdot \frac{|x_i|}{r} = x_i$ and $V[Z_i] = r^2 \cdot \frac{|x_i|}{r} = r|x_i|$.

Let

$$\bar{Z} = \frac{1}{t} \sum_{i=1}^{t} Z_i \tag{10}$$

where $Z_i$ are independent copies of $Z$.

We have that:

$$E[||\bar{Z} - x||^2] = E\left[\sum_{j=1}^{d} (\bar{Z}_j - x_j)^2\right] \tag{11}$$

$$= \sum_{j=1}^{d} E\left[(\bar{Z}_j - x_j)^2\right] = \sum_{j=1}^{d} V(\bar{Z}_j) \tag{12}$$

$$= \sum_{j=1}^{d} V\left(\frac{1}{t} \sum_{i=1}^{t} (Z_i)_j\right) = \tag{13}$$

$$\frac{1}{t^2} t \sum_{j=1}^{d} V(Z_j) = \frac{1}{t} \sum_{j=1}^{d} r|x_i| = \frac{r||x||_1}{t} \leq \frac{r^2}{t}. \tag{14}$$

If we choose $t$ such that: $\frac{r^2}{t} \leq \delta^2$, then, we have that $E[||\bar{Z} - x||^2] \leq \delta^2$. Hence, for $t \geq \frac{r^2}{\delta^2}$, by the Pigeonhole Principle, we have that there is a $\bar{Z}$ such that: $||\bar{Z} - x|| \leq \delta$. In other words, the set of all possible $\bar{Z}$ form an $\delta-$net for $B_1^d(r)$ for $t \geq \frac{r^2}{\delta^2}$. Set $t = \frac{r^2}{\delta^2}$. We will now count how many $\bar{Z}$ there are. For each $\bar{Z}$, we have $t$ choices, each one of which can take one value among $2d+1$ values. Hence, there are $(2d+1)^t$ different $\bar{Z}$. Therefore, we can create an $\delta$-net of $B_1^d(r)$ that has $(2d+1)^{\frac{r^2}{\delta^2}}$ elements, i.e.

$$\log N(\delta, B_1^d(r), ||\cdot||_2) \leq \frac{r^2}{\delta^2} \log(2d+1).$$

□

The same result (up to constants) for the size of the $\epsilon$-net for an $l_1$ ball follows from Sudakov's minoration inequality.

**Theorem 3** (Sudakov minoration (Sudakov, 1969; Wainwright, 2019)). *Let $\{X_\theta, \theta \in T\}$ be a zero-mean Gaussian process on $T \subset R^d$. Then,*

$$\mathcal{G}(T) \geq \frac{\delta}{2} \sqrt{\log M(\delta/2, T, \rho_X)}, \tag{15}$$

*with $\rho_X(\theta_1, \theta_2) = \sqrt{E[(X_{\theta_1} - X_{\theta_2})^2]}$.*

**Corollary 1.**

$$\log N(\delta, B_1^d(r), ||\cdot||_2) \leq \frac{16 r^2}{\delta^2} \log d. \tag{16}$$

*Proof.* Observe that:

$$\mathcal{G}(B_1^d(r)) = E_w\left[\sup_{||u||_1 \leq r} u^T w\right] \tag{17}$$

$$\leq r E_w\left[||w||_\infty\right] \tag{18}$$

$$\leq 2r\sqrt{\log d}. \tag{19}$$

By Inequality (15),

$$\mathcal{G}(T) \geq \frac{\delta}{2}\sqrt{\log M(\delta/2, T, \rho_X)} \Rightarrow \tag{20}$$

$$\log M(\delta/2, B_1^d(r), ||\cdot||_2) \leq 16\frac{r_1^2}{\delta^2}\log d. \tag{21}$$

It follows from the definition of the covering number that:

$$M(\delta, T, ||\cdot||_2) \leq M(\delta/2, T, ||\cdot||_2).$$

By Inequality (7), we also have:

$$N(\delta, T, ||\cdot||_2) \leq M(\delta, T, ||\cdot||_2). \tag{22}$$

Hence,

$$\log N(\delta, B_1^d(r), ||\cdot||_2) \leq \frac{16r^2}{\delta^2}\log d. \tag{23}$$

□

**Theorem 4** (Volume rations and metric entropy (Wainwright, 2019))**.** *Let $\Theta$ be an arbitrary set. Then,*

$$\frac{\text{vol}(\Theta)}{\text{vol}(\delta B_q^d(1))} \leq N(\delta, \Theta, ||\cdot||_q) \leq \frac{\text{vol}\left(\frac{2}{\delta}\Theta \oplus B_q^d(1)\right)}{\text{vol}(B_q^d(1))} \tag{24}$$

**Corollary 2.**

$$\log N(\delta, B_1^d(r), ||\cdot||_2) \leq d\log\frac{4r}{\delta} \tag{25}$$

*Proof.* By Theorem 4, we have that:

$$N(\delta, B_1^d(r), ||\cdot||_2) \leq \frac{\text{vol}\left(\frac{2}{\delta}B_1^d(r) \oplus B_2^d(1)\right)}{\text{vol}(B_2^d(1))} \tag{26}$$

$$\frac{\text{vol}\left(\frac{2}{\delta}B_2^d(r) \oplus B_2^d(1)\right)}{\text{vol}(B_2^d(1))} \leq \left(\frac{2r}{\delta}+1\right)^d \tag{27}$$

$$\leq \left(\frac{4r}{\delta}\right)^d. \tag{28}$$

□

**Remark 5.** *Observe that by Theorem 2 and Corollary 2, we get two different upper bounds regarding the covering of the $l_1$-ball. With Maurey's method, the covering number depends logarithmically in the dimension but polynomially on $\frac{1}{\epsilon}$. On the other hand, the volumetric argument gives polynomial dependence on the dimension and logarithmic dependence on $\frac{1}{\epsilon}$. The Maurey's bound is tighter when $\epsilon = \Omega\left(\frac{r}{\sqrt{d}}\right)$.*

### 6.1.2. S-REC

**Lemma 3** (S-REC for nested $l_1$-ball)**.** *Let $G = G_2 \circ G_1$ with $G_1 : \mathbb{R}^k \to \mathbb{R}^p$ be an $L_1$-Lipschitz function and $G_2 : \mathbb{R}^p \to \mathbb{R}^n$ be an $L_2$-Lipschitz function. Let $A \in R^{m\times n}$ be a random matrix with $A_{ij} \sim \mathcal{N}(0, 1/m)$ i.i.d. entries.*

*Then, if*

$$m = \frac{1}{(1-\gamma)^2}\Omega\left(k\log\frac{L_1L_2r_1}{\delta} + K^2\log p\right) \tag{29}$$

$$r_2 = \frac{K\cdot\delta}{L_2}, \quad 1 < K < \sqrt{p} \tag{30}$$

*w.p. $1 - e^{-\Omega((1-\gamma)^2 m)}$, we have that $A$ satisfies S-REC$(G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2)), \gamma, \log(4K)\cdot\frac{\sqrt{p}}{K}\cdot\log\frac{\sqrt{p}}{K})$.*

*Proof.* Using Theorem 4, we get that:

$$N\left(\frac{\delta}{L_1\cdot L_2}, B_2^k(r_1), ||\cdot||_2\right) \leq \left(\frac{2L_1L_2r_1}{\delta}+1\right)^k \leq \left(\frac{4L_1L_2r_1}{\delta}\right)^k \tag{31}$$

Using the fact that $G_2 \circ G_1$ is $L_1L_2$ Lipschitz, we get that:

$$N(\delta, G(B_2^k(r_1)), ||\cdot||_2) \leq \left(\frac{4L_1L_2r_1}{\delta}\right)^k \tag{32}$$

Using Maurey's Empirical Method (see Theorem 2), we get that:

$$\log N\left(\frac{\delta}{L_2}, B_1^p(r_2), ||\cdot||_2\right) \leq \frac{r_2^2L_2^2}{\delta^2}\log(2p+1). \tag{33}$$

Setting $r_2 = \frac{K\cdot\delta}{L_2}$ and using the fact that $G_2$ is $L_2$-Lipschitz, we get:

$$\log N\left(\delta, G_2(B_1^p(r_2)), ||\cdot||_2\right) \leq K^2\log 3p \tag{34}$$

By (32), (34), we get that:

$$\log N\left(\delta, G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2)), ||\cdot||_2\right) \leq k\log\frac{4L_1L_2r_1}{\delta} + K^2\log 3p. \tag{35}$$

By JL lemma, if $m = \frac{1}{a^2}\Omega\left(k\log\frac{4L_1L_2r_1}{\delta} + K^2\log 3p\right)$, then w.p. $1 - e^{-\Omega(a^2m)}$, we have that:

$$||AG_2(\hat{z}_2^p) - AG_2(\hat{z}_1^p)|| \geq (1-a)||G_2(\hat{z}_2^p) - G_2(\hat{z}_1^p)||, \ \forall \hat{z}_1^p, \hat{z}_2^p \in S \tag{36}$$

where $S$ is a minimal $\delta$-net of $G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2))$.

Let $z_1^p, z_2^p \in G_1(B_2^k(r_1))) \oplus B_1^p(r_2)$ and $\hat{z}_1^p = \text{argmin}_{\tilde{z}_1^p \in S}||z_1^p - \tilde{z}_1^p||$, $\hat{z}_2^p = \text{argmin}_{\tilde{z}_2^p \in S}||z_2^p - \tilde{z}_2^p||$. Then,

$$||AG_2(z_2^p) - AG_2(z_1^p)|| \geq ||AG_2(\hat{z}_2^p) - AG_2(\hat{z}_1^p)|| \\ -||AG_2(z_2^p) - AG_2(\hat{z}_2^p)|| - ||AG_2(z_1^p) - AG_2(\hat{z}_1^p)|| \quad (37)$$

$$\geq (1-a)||G_2(\hat{z}_2^p) - G_2(\hat{z}_1^p)|| - ||AG_2(z_2^p) - AG_2(\hat{z}_2^p)|| \\ -||AG_2(z_1^p) - AG_2(\hat{z}_1^p)|| \quad (38)$$

$$\geq (1-a)||G_2(z_2^p) - G_2(z_1^p)|| - \\ (1-a)\left(||G_2(z_2^p) - G_2(\hat{z}_2^p)|| + ||G_2(z_1^p) - G_2(\hat{z}_1^p)||\right) \\ -||AG_2(z_2^p) - AG_2(\hat{z}_2^p)|| - ||AG_2(z_1^p) - AG_2(\hat{z}_1^p)|| \quad (39)$$

$$\geq (1-a)||G_2(z_2^p) - G_2(z_1^p)|| - 2\delta \\ -||AG_2(z_2^p) - AG_2(\hat{z}_2^p)|| - ||AG_2(z_1^p) - AG_2(\hat{z}_1^p)|| \quad (40)$$

By Lemma 4, we have that w.p. $1 - e^{-\Omega(m)}$, $||AG_2(z_2^p) - AG_2(\hat{z}_2^p)|| + ||AG_2(z_1^p) - AG_2(\hat{z}_1^p)|| = O\left(\log(4K) \cdot \frac{\sqrt{p}}{K} \cdot \log\frac{\sqrt{p}}{K}\right) \cdot \delta$. Let $a = 1 - \gamma$. Hence,

$$||AG_2(z_2^p) - AG_2(z_1^p)|| \geq \gamma||G_2(z_2^p) - G_2(z_1^p)|| \\ - \log(4K) \cdot \frac{\sqrt{p}}{K} \cdot \log\frac{\sqrt{p}}{K} \cdot \delta. \quad (41)$$

□

**Lemma 4.** *Let $G = G_2 \circ G_1$ with $G_1 : \mathbb{R}^k \to \mathbb{R}^p$ be an $L_1$-Lipschitz function and $G_2 : \mathbb{R}^p \to \mathbb{R}^n$ be an $L_2$-Lipschitz function. Let $A \in R^{m \times n}$ be a random matrix with $A_{ij} \sim \mathcal{N}(0, 1/m)$ i.i.d. entries. Let $M_0$ be a $\frac{\delta}{L_2}$ net of $G_1(B_2^k(r_1)) \oplus B_1^p(r_2)$ such that $\log|M_0| \leq k\log\left(\frac{4L_1L_2r_1}{\delta}\right) + K^2\log 3p$.*

*Then, if*

$$m = \Omega\left(k\log\left(\frac{4L_1L_2r_1}{\delta}\right) + K^2\log p\right), \\ r_2 = \frac{K\cdot\delta}{L_2}, \quad 1 < K < \sqrt{p}. \quad (42)$$

*then for any $x \in G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2))$, if $x' = \text{argmin}_{\hat{x} \in G_2(M_0)}||x - \hat{x}||$, w.p. $1 - e^{\Omega(m)}$, we have that:*

$$||A(x - x')|| = O\left(\log(4K) \cdot \frac{\sqrt{p}}{K} \cdot \log\frac{\sqrt{p}}{K}\right) \cdot \delta. \quad (43)$$

*Proof.* From Lemma 8.2 of (Bora et al., 2017), we have that if $\epsilon \geq 2 + \frac{4}{m}\log\frac{2}{f}$, then

$$P(||Ax|| \geq (1+\epsilon)||x||) \leq f. \quad (44)$$

Let $N_0 \subseteq N_1 \subseteq ... \subseteq N_l$ be a chain of minimal $\delta_i$-nets of $G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2))$.

Let also:

$$T_i = \{x_{i+1} - x_i | x_{i+1} \in N_{i+1}, x_i \in N_i\}. \quad (45)$$

By union bound,

$$P(||At|| \leq (1+\epsilon_i)||t||, \quad \forall i \in [0, ..., l-1], \quad \forall t \in T_i) \geq \\ 1 - \sum_{i=0}^{l-1} |T_i| f_i, \quad (46)$$

where $\epsilon_i = 2 + \frac{4}{m}\log\frac{2}{f_i}$. We want to choose $f_i$ such that $\sum_{i=0}^{l-1}|T_i|f_i$ decays exponentially with $m$.

First notice that:

$$\log|T_i| \leq \log|N_{i+1}| + \log|N_i| \quad (47)$$

To develop bounds for $\log|N_i|, \log|N_{i+1}|$ we first need to decide how $\delta_i$ decays and then whether we are going to use Maurey's method or the volumetric argument.

We choose $\delta_i = \frac{\delta}{2^i}$. Now assume $m = K^2\log(3p) + k\log\left(\frac{L_1L_2r_1}{\delta}\right)$.

For $0 \leq i < \log\frac{\sqrt{p}}{K}$ we will use Maurey's method.

$$\log|T_i| \leq 2\log|N_{i+1}| \quad (48)$$

$$\leq 2\left(\left(\frac{K\delta}{\delta_{i+1}}\right)^2\log(3p) + k\log\left(\frac{L_1L_2r_1}{\delta_i}\right)\right) \quad (49)$$

$$\leq 2\cdot\left(4^{i+1}K^2\log(3p) + k\log\left(\frac{L_1L_2r_1}{\delta}\right) + k(i+1)\right) \quad (50)$$

$$\leq 2\cdot\left(4^{i+1}K^2\log(3p) + 2k\log\left(\frac{L_1L_2r_1}{\delta}\right)(i+1)\right) \quad (51)$$

$$\leq 2\cdot 4^{i+1}m. \quad (52)$$

To get probability that decays exponentially with $m$, we choose:

$$\log f_i = -3\cdot 4^{i+1}m \quad (53)$$

$$\epsilon_i = O(1) + 3\cdot 4^{i+1}. \quad (54)$$

For $\log\frac{\sqrt{p}}{K} \leq i \leq l-1$, we will use the volumetric argument.

$$\log|T_i| \leq p\log\left(4K\frac{\delta}{\delta_i}\right) + p\log\left(4K\frac{\delta}{\delta_{i+1}}\right) + \\ k\log\left(\frac{L_1L_2r_1}{\delta_i}\right) + k\log\left(\frac{L_1L_2r_1}{\delta_{i+1}}\right) \quad (55)$$

$$\leq 2p\log(4K) + p(2i+1) + \\ 2k\log\left(\frac{L_1L_2r_1}{\delta}\right) + k(2i+1) \quad (56)$$

$$\leq 2p\log(4K) + 3pi + 2k\log\left(\frac{L_1L_2r_1}{\delta}\right) + 3ki \quad (57)$$

$$\leq 5ip\log(4K) + 5ik\log\left(\frac{L_1L_2r_1}{\delta}\right) \quad (58)$$

We choose:

$$\log f_i = -6ip\log(4K) - 6ik\log\left(\frac{L_1L_2r_1}{\delta}\right) \quad (59)$$

$$\epsilon_i \leq O(1) + \log(4K)\frac{ip}{m} + i. \quad (60)$$

Notice that:

$$\log|T_i|f_i \leq -ip\log(4K) - ik\log\left(\frac{L_1L_2r_1}{\delta}\right) \leq -im. \quad (61)$$

For that choice of parameters, observe that:

$$P(||At|| \leq (1+\epsilon_i)||t||, \quad \forall i \in [0,...,l-1], \quad \forall t \in T_i) \quad (62)$$

$$= 1 - e^{-\Omega(m)}. \quad (63)$$

Let $x$ be the image we want to recostruct and $x_i$ be the closest point of that image to the $\delta_i$ net. Then,

$$x - x_0 = \sum_{i=0}^{l-1}(x_{i+1} - x_i) + x - x_l \Rightarrow \quad (64)$$

$$||Ax - Ax_0|| \leq \sum_{i=0}^{l-1}||Ax_{i+1} - Ax_i|| + ||Ax - Ax_l||. \quad (65)$$

Now w.h.p. $||Ax_{i+1} - Ax_i|| \leq (1+\epsilon_i)||x_{i+1} - x_i||$. Therefore, w.h.p.:

$$||Ax - Ax_0|| \leq \sum_{i=0}^{l-1}(1+\epsilon_i)||x_{i+1} - x_i|| + ||Ax - Ax_l|| \quad (66)$$

$$\leq \sum_{i=0}^{l-1}(1+\epsilon_i)\delta_i + ||Ax - Ax_l|| \quad (67)$$

$$\leq \sum_{i=0}^{\log\frac{\sqrt{p}}{K}-1}\left(O(1) + 3\cdot 4^{i+1}\right)\frac{\delta}{2^i} + \sum_{i=\log\frac{\sqrt{p}}{K}}^{l-1}\left(O(1) + \log(4K)\frac{ip}{K^2\log 3p} + i\right)\frac{\delta}{2^i} + ||Ax - Ax_l|| \quad (68)$$

$$\leq O\left(\log(4K)\cdot\frac{\sqrt{p}}{K}\cdot\log\frac{\sqrt{p}}{K}\right)\cdot\delta + ||Ax - x_l||. \quad (69)$$

Observe that:

$$||Ax - Ax_l|| \leq ||A||\cdot||x - x_l|| \quad (70)$$

$$\leq 2\sqrt{n}||x - x_l|| \quad (71)$$

$$\leq 2\frac{\sqrt{n}}{2^l}\delta. \quad (72)$$

For $l = \log n$, we have that $||Ax - Ax_l|| \leq \delta$. Hence,

$$||Ax - Ax_0|| \leq O\left(\log(4K)\cdot\frac{\sqrt{p}}{K}\cdot\log\frac{\sqrt{p}}{K}\right)\cdot\delta. \quad (73)$$

□

### 6.1.3. Proof of main theorem

*Proof of Theorem 1.* Let $\delta_{l_1} = \left(\log(4K)\cdot\frac{\sqrt{p}}{K}\log\frac{\sqrt{p}}{K}\right)\delta$. Then,

$$||G_2(\tilde{z}^p) - G_2(\tilde{z}^p)|| \leq \quad (74)$$

$$\frac{||AG_2(\tilde{z}^p) - AG_2(\bar{z}^p)|| + \delta_{l_1}}{\gamma} \quad (75)$$

$$\leq \frac{||Ax - AG_2(\bar{z}^p)|| + ||Ax - AG_2(\hat{z}^p)|| + \delta_{l_1}}{\gamma} \quad (76)$$

$$\leq \frac{2||Ax - AG_2(\bar{z}^p)|| + \delta_{l_1}}{\gamma} \quad (77)$$

$$\leq \frac{4||G_2(\bar{z}^p) - x|| + \delta_{l_1}}{\gamma}. \quad (78)$$

Finally, observe that:

$$||G_2(\tilde{z}^p) - x|| \leq ||G_2(\bar{z}^p) - x|| + ||G_2(\bar{z}^p) - G_2(\tilde{z}^p)|| \quad (79)$$

$$\leq \left(1 + \frac{4}{\gamma}\right)||x - G_2(\bar{z}^p)|| + \frac{\delta_{l_1}}{\gamma}. \quad (80)$$

□

**Remark 6.** *Similar to the analysis of the CSGM paper (see Lemma 4.3), $\gamma$ is a constant that we control and we may set it to $\gamma = \frac{4}{5}$ to get the same scaling term with CSGM.*

### 6.1.4. Proof of Lemma 1

**Lemma 5.** *Consider the setting of Theorem 1. Let $g = [g_1, \cdots, g_n]$ be a vector with i.i.d. Gaussian entries of variance $1/m$, let $F \in \mathbb{R}^{m\times n}$ be a partical circulant matrix that has $g$ in its first row, and let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix with uniform $\pm 1$ entries along its diagonal. Then for $m = \Omega\left(\frac{1}{(1-\gamma)^2}(k\log\frac{L_1L_2r_1}{\delta} + K^2\log p)\log^4(n)\right)$, $FD$ satisfies S-REC$(G_2(G_1(B_2^k(r_1)) \oplus B_1^p(r_2)), 1-\gamma, \delta\cdot\frac{\log(4K)}{\gamma}\cdot\frac{\sqrt{p}}{K}\log\frac{\sqrt{p}}{K})$ with probability $1 - e^{-\Omega(m)}$.*

*Proof.* The proof follows from the proof of Lemma 3 above and Theorem 3.1 in (Krahmer & Ward, 2011). The proof of Lemma 3 requires the Johnson-Lindenstrauss guarantee for a set of size $2^{O(m)}$, and invoking Theorem 3.1 in (Krahmer & Ward, 2011), this is guaranteed to hold for the matrix $FD$.

The proof of Lemma 3 also requires $\|FD\|_{op} \leq \sqrt{n}$. This is also guaranteed by noting that

$$\|FD\|_{op} \leq \|FD\|_F \leq \sqrt{n},\ w.p. 1 - e^{-\Omega(m)}.$$

□

### 6.2. Code

We plan to release the code we used for all our experiments. The whole package is included in the Supplementary Material submission and we are working towards a public release soon.

Our code is implemented in PyTorch (Paszke et al., 2019). Our code is based on the following open-source implementations of StyleGAN-2: https://github.com/rosinality/stylegan2-pytorch, https://github.com/NVlabs/stylegan2. We also draw inspiration from the open-source implementation of PULSE: https://github.com/tg-bomze/Face-Depixelizer. A Tensorflow (Abadi et al., 2016) implementation is in the works.

Our current repository includes:

- Detailed instructions on how to setup the environment and download the dependencies.
- Code for image pre-processing, such as random inpainting, interactive masks, noise addition, automatic face alignment, etc.
- Examples on how to run inpainting, denoising, super-resolution and compressed-sensing with circulant matrices for custom images.
- Code for out-of-distribution generation on 1000 ImageNet (Deng et al., 2009) classes using a robust classifier. We use a robust classifier from the `robustness` (Engstrom et al., 2019) library.
- Code for evaluating the performance of ILO and previous methods on Celeba-HQ (Liu et al., 2018; Lee et al., 2020).
- Tools to visualize performance and track experiments.
- Code for generating GIF files by collecting frames during the optimization.

Our code is GPU/CPU compatible.

### 6.3. Experimental details

We performed all our experiments on a single GPU. As mentioned in the paper, obtaining a solution for a single inverse problem requires less than a minute on a single 1080Ti. All the experiments can be reproduced in less than a day on a single GPU.

Unless mentioned otherwise, we use Adam (Kingma & Ba, 2014) optimizer with an initial learning rate of $0.1$ for each layer. During a single layer optimization, learning rate ramps up linearly and is ramped down using a cosine scheduler, as proposed by (Karras et al., 2020).

Loss functions are changed for each task as explained in the paper. For all tasks, we use a geodesic loss with coefficient $0.01$. For random inpainting, we use both MSE and LPIPS when we have more than $20\%$ observed pixels, otherwise we only use MSE. When both MSE and LPIPS are used, we search co-efficients in the set $\{0.5, 1, 2, 5\}$ for each of the terms. For inpainting with continuous black boxes, we used both MSE and LPIPS. For the experiments of Figure 1 of the main paper, we used the same co-efficient for both MSE and LPIPS.

Our optimization algorithm is Projected Gradient Descent (Nocedal & Wright, 2006). First, we project each latent code to the unit sphere. Next, when optimizing over deeper layers, we use $l_1$ projection to stay close to the manifold induced by the previous layers. The projection in that case includes the solution of the previous layers, the latent codes (i.e. $w_i$) and the noises, (i.e. $u_i$). We tune seperately the $l_1$ radii for each one of the optimization variables and for each one of the layers. Empirically, we find that the following radii for the first four layers works decently for most of the tasks/images:

- Radius of noises: $300, 2000, 2000, 4000$.
- Radius of latent codes: $300, 500, 1000, 2000$.
- Radius of previous solutions: $500, 1000, 2000$.

Projection to the $l_1$ ball allows for optimization on deep layers of the generator (that is not possible without projections). By doing that we get better reconstruction that comes with the cost of increased number of optimization steps. Generally, tuning the radii for each layer is an especially difficult procedure. Even worse, these hyperparameters do not transfer across tasks. For the first four layers, we encourage the reader to use the parameters mentioned above.

To obtain the plots of Figure 2, we sampled (randomly) 5 images from Celeba-HQ and we reported the best score for each point on the horizontal axis over 5 different runs (25 runs in total for each method for each point in the plot) with different hyperparameters. The error bars are computed across the experiments for different images. For the plots of Figure 2, we searched over the following combinations of number of steps for each layer (starting from the first):
$\{300, 200, 200, 100\}, \{300, 200, 100\},$
$\{300, 200, 200, 100, 50\}, \{50, 50, 50, 50, 500\},$

$\{100, 100, 100, 100, 100\}$. Each reported point is the average (across images) of the minimums of those runs.

### 6.4. Additional Experiments

In this section, we list additional figures and experiments.

Figure 8 shows that by combining MSE loss to a reference image and the classification probability, we can morph a given person to an ImageNet (Deng et al., 2009) class. We observe experimentally that better results are obtained by only using MSE and LPIPS loss during the first ILO rounds and only using the additional classification term in the deeper layers of the generator. An extra benefit of this method is that we can interpolate intermediate frames to see how actually a human face can be transformed to an imagenet class since the generator first matches the phase and then uses the classifier to alter it.

Table 2 shows the effect of the radius of the $l_1$ deviations to the ground truth reconstruction error. For many measurements $m = n$, increasing the radius to infinity gives the smallest error since regularization is not needed. For lower measurements $m = 0.8n$, smaller radius is helping performance.

We also compare with the Projected Gradient Descent (PGD) algorithm, proposed by Shah & Hegde (2018b). This algorithm proceeds by first taking a step in the pixel space to match the measurements and then projecting the new image back to the generator range. Table 3 shows the results of the comparison. We evaluate on images with $10\%$ randomly observed pixels and additive noise of $\sigma = 12.5 \,|\, 25.0$. As shown, PGD improves upon CSGM but ILO is clearly superior in terms of ground truth MSE.

Finally, in order to show that our method can be successfully in other datasets as well, we perform inpainting experiments using a pre-trained StyleGAN-2 generators on cats. Results are shown in Figure 9.

| Task | Steps | $l_1$ radius | MSE | LPIPS |
|---|---|---|---|---|
| Inversion ($m = n$) | $100, 100, 100$ | 0 | 0.0315 | 0.1532 |
| | | 300 | 0.0259 | 0.1467 |
| | | $\infty$ | **0.0247** | **0.1402** |
| Inpainting ($m = 0.8n$) | $100, 100, 100$ | 0 | 0.0327 | 0.1576 |
| | | 300 | **0.0268** | **0.1495** |
| | | $\infty$ | 0.0285 | 0.1529 |

*Table 2.* Effect of radius of $l_1$ deviations from the range of an intermediate layer. For many measurements $m = n$, increasing the radius to infinity gives the smallest error since regularization is not needed. For lower measurements $m = 0.8n$, smaller radius is helping performance.

| Method | Observed Pixels | Noise std | MSE |
|---|---|---|---|
| PGD | 10% | 12.5 | 0.0370 |
| CSGM | | | 0.0611 |
| **ILO** | | | **0.0057** |
| PGD | 10% | 25.0 | 0.0418 |
| CSGM | | | 0.0643 |
| **ILO** | | | **0.0090** |

*Table 3.* Qualitative comparison with the Projected Gradient Descent (PGD) (Shah & Hegde, 2018b) algorithm. We evaluate on images with $10\%$ randomly observed pixels and additive noise of $\sigma = 12.5 \mid 25.0$. As shown, PGD improves upon CSGM but ILO is clearly superior in terms of ground truth MSE.

*Figure 8.* Morphing using a classifier for Bull Frog class, keeping also a loss term for distance to a well-known machine learning researcher.

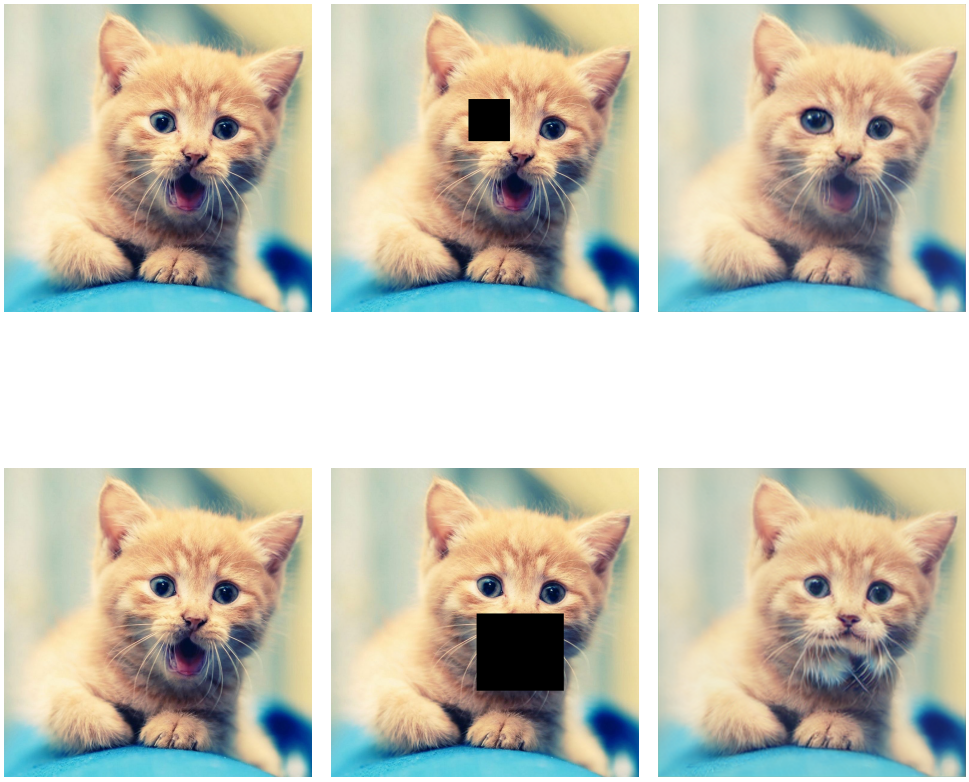

*Figure 9.* Inpainting using a StyleGAN trained to generate cat images. First column: Original image (never observed). Second column: Observed image. Third column: ILO reconstruction.

### 6.5. Ethical Considerations

Previous research has reported that StyleGAN leads to biased generations (Menon et al., 2020; Jain et al., 2020; Tan et al., 2020; Salminen et al., 2020). In practice, we observe that by extending the range of the generator we obtain more diverse generations. A similar finding has been reported by Abdal et al. (2019). Even though we observe less biased reconstructions, we encourage a lot more research on this topic.

Admittedly, our method makes the creation of DeepFakes (Korshunov & Marcel, 2018) easier, since it expands the range of the generator. Arguably, the technology behind DeepFakes is already very powerful so the negative effect of this work will be diminishing. An interesting topic of research is whether existing defences against DeepFakes (Matern et al., 2019; Güera & Delp, 2018; Nguyen et al., 2019; Yang et al., 2020) are robust to images that lie outside of the range of the GAN.

Experiments with a robust classifier combined with similarity to a reference image can be abused to generate images that are offensive in various ways. In this paper we are only exploring with what is possible, but future work should consider detecting and preventing such abuse.

### 6.6. Things that did not work

We share some negative results we encountered during the process of writing this paper. Our goal is to inform the research community about some methods that failed so that future research can avoid them, reformulate them or even contradict our findings. We also suggest ways to mitigate some of the issues we experienced.

First, we observed that joint optimization of all noise vectors leads to poor visual reconstructions. Even in cases where the MSE loss to the unobserved image was going down, joint-optimization of all noise vectors was giving blurry and/or unrealistic reconstructions. Thus we believe that expanding the generator space without sequential optimization and constraints fails since it makes it too powerful.

We also tried to establish a criterion on how many steps to run per layer. In practice, we observed that a working heuristic is to move to the next layer when the observed MSE error flattens. Even though this idea works well for the first layers, it can lead to unrealistic reconstructions when applied to deeper layers of the generator. To mitigate this issue, we choose very small radii when optimizing in deep layers. Tuning the hyperparameters (learning rates, number of steps and optimization radii) for each of the layers can be a particularly toilsome procedure. Sadly, we observed that these parameters do not generalize across different tasks (even though they mostly generalize across different images).

On the theoretical side, we tried (unsuccessfully) to obtain similar results for an $l_2$ dilation of the range of the first generator. The main bottleneck is the measurements bound; for the $l_2$ ball, we cannot avoid linear dependence on the intermediate dimension. It is not clear yet whether a similar result for the $l_2$ case could be proved. In practice, we did not observe significant difference between projecting in $l_1$ or $l_2$ balls. Moreover, it is known that $l_1$ projection encourages sparsity in some cases, e.g. see LASSO (Tibshirani, 1996). Establishing a connection between the $l_0$ and the $l_1$ solutions is left as future work.

## References


Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.

Abdal, R., Qin, Y., and Wonka, P. Image2stylegan: How to embed images into the stylegan latent space? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4432–4441, 2019.

Asim, M., Daniels, M., Leong, O., Ahmed, A., and Hand, P. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning*, pp. 399–409. PMLR, 2020.

Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3): 253–263, 2008.

Bau, D., Zhu, J.-Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., and Torralba, A. Seeing what a gan cannot generate, 2019.

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In *International Conference on Machine Learning*, pp. 537–546. PMLR, 2017.

Chen, G.-H., Tang, J., and Leng, S. Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. *Medical physics*, 35(2):660–663, 2008.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, pp. 606414. International Society for Optics and Photonics, 2006.

Daras, G., Odena, A., Zhang, H., and Dimakis, A. G. Your local gan: Designing two dimensional local attention mechanisms for generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

Dhar, M., Grover, A., and Ermon, S. Modeling sparse deviations for compressed sensing using generative models. In *International Conference on Machine Learning*, pp. 1214–1223. PMLR, 2018.

Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., and Baraniuk, R. G. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

Duchi, J., Shalev-Shwartz, S., Singer, Y., and Chandra, T. Efficient projections onto the l 1-ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pp. 272–279, 2008.

Engstrom, L., Ilyas, A., Salman, H., Santurkar, S., and Tsipras, D. Robustness (python library), 2019. URL https://github.com/MadryLab/robustness.

Güera, D. and Delp, E. J. Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6. IEEE, 2018.

Hand, P. and Voroninski, V. Global guarantees for enforcing deep generative priors by empirical risk. In *Conference On Learning Theory*, pp. 970–978. PMLR, 2018.

Hand, P., Leong, O., and Voroninski, V. Phase retrieval under a generative prior. *arXiv preprint arXiv:1807.04261*, 2018.

Hegde, C., Duarte, M. F., and Cevher, V. Compressive sensing recovery of spike trains using a structured sparsity model. In *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

Hinrichs, A. and Vybíral, J. Johnson-lindenstrauss lemma for circulant matrices. *Random Structures & Algorithms*, 39(3):391–398, 2011.

Jain, N., Olmo, A., Sengupta, S., Manikonda, L., and Kambhampati, S. Imperfect imaganation: Implications of gans exacerbating biases on facial data augmentation and snapchat selfie lenses. *arXiv preprint arXiv:2001.09528*, 2020.

Kabkab, M., Samangouei, P., and Chellappa, R. Task-aware compressed sensing with generative adversarial networks. In *AAAI*, 2018.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019. doi: 10.1109/cvpr.2019.00453. URL http://dx.doi.org/10.1109/CVPR.2019.00453.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. Analyzing and improving the image quality of stylegan. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020. doi: 10.1109/cvpr42600.2020.00813. URL http://dx.doi.org/10.1109/cvpr42600.2020.00813.

Keys, R. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Korshunov, P. and Marcel, S. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*, 2018.

Krahmer, F. and Ward, R. New and improved johnson–lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3): 1269–1281, 2011.

Lee, C.-H., Liu, Z., Wu, L., and Luo, P. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5549–5558, 2020.

Lei, Q., Jalal, A., Dhillon, I. S., and Dimakis, A. G. Inverting deep generative models, one layer at a time. In *NeurIPS*, 2019.

Liu, H., Jiang, B., Xiao, Y., and Yang, C. Coherent semantic attention for image inpainting. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019. doi: 10.1109/iccv.2019.00427. URL http://dx.doi.org/10.1109/ICCV.2019.00427.

Liu, Z. and Scarlett, J. Information-theoretic lower bounds for compressive sensing with generative models. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 292–303, May 2020. ISSN 2641-8770. doi: 10.1109/jsait.2020.2980676. URL http://dx.doi.org/10.1109/JSAIT.2020.2980676.

Liu, Z., Luo, P., Wang, X., and Tang, X. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August*, 15 (2018):11, 2018.

Lucas, A., Iliadis, M., Molina, R., and Katsaggelos, A. K. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018.

Lustig, M., Donoho, D., and Pauly, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008.

Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S. S., Zaharchuk, G., Xing, L., and Pauly, J. M. Deep generative adversarial neural networks for compressive sensing mri. *IEEE transactions on medical imaging*, 38(1):167–179, 2018.

Matern, F., Riess, C., and Stamminger, M. Exploiting visual artifacts to expose deepfakes and face manipulations. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 83–92. IEEE, 2019.

Menon, S., Damian, A., Hu, S., Ravi, N., and Rudin, C. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. doi: 10.1109/cvpr42600.2020.00251. URL http://dx.doi.org/10.1109/cvpr42600.2020.00251.

Mousavi, A., Dasarathy, G., and Baraniuk, R. G. A data-driven and distributed approach to sparse signal representation and recovery. In *International Conference on Learning Representations*, 2019.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., and Nahavandi, S. Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573*, 2019.

Nocedal, J. and Wright, S. *Numerical optimization*. Springer Science & Business Media, 2006.

Ongie, G., Jalal, A., Metzler, C. A., Baraniuk, R. G., Dimakis, A. G., and Willett, R. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

Pajot, A., de Bezenac, E., and Gallinari, P. Unsupervised adversarial image reconstruction. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=BJg4Z3RqF7.

Pandit, P., Sahraee, M., Rangan, S., and Fletcher, A. K. Asymptotics of map inference in deep networks. *2019 IEEE International Symposium on Information Theory (ISIT)*, Jul 2019. doi: 10.1109/isit.2019.8849316. URL http://dx.doi.org/10.1109/ISIT.2019.8849316.

Park, J. Y., Smedemark-Margulies, N., Daniels, M., Yu, R., van de Meent, J.-W., and HAnd, P. Generator surgery for compressed sensing. In *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020. URL https://openreview.net/forum?id=s2EucjZ6d2s.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2536–2544, 2016.

Pisier, G. Probabilistic methods in the geometry of banach spaces. In *Probability and analysis*, pp. 167–241. Springer, 1986.

Qaisar, S., Bilal, R. M., Iqbal, W., Naureen, M., and Lee, S. Compressive sensing: From theory to applications, a survey. *Journal of Communications and networks*, 15(5): 443–456, 2013.

Raj, A., Li, Y., and Bresler, Y. Gan-based projector for faster recovery with convergence guarantees in linear inverse problems. 2019.

Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., and Cohen-Or, D. Encoding in style: a stylegan encoder for image-to-image translation. *arXiv preprint arXiv:2008.00951*, 2020.

Salminen, J., Jung, S.-g., Chowdhury, S., and Jansen, B. J. Analyzing demographic bias in artificially generated facial pictures. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2020.

Santurkar, S., Tsipras, D., Tran, B., Ilyas, A., Engstrom, L., and Madry, A. Image synthesis with a single (robust) classifier. *arXiv preprint arXiv:1906.09453*, 2019.

Shah, V. and Hegde, C. Solving linear inverse problems using gan priors: An algorithm with provable guarantees. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr 2018a. doi: 10.1109/icassp.2018.8462233. URL http://dx.doi.org/10.1109/ICASSP.2018.8462233.

Shah, V. and Hegde, C. Solving linear inverse problems using gan priors: An algorithm with provable guarantees, 2018b.

Song, G., Fan, Z., and Lafferty, J. Surfing: Iterative optimization over incrementally trained deep networks. *NeurIPS*, 2019.

Sudakov, V. N. Gaussian measures, cauchy measures and $\varepsilon$-entropy. In *Soviet Math. Dokl*, volume 10, pp. 310–313, 1969.

Sun, W. and Chen, Z. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Transactions on Image Processing*, 29:4027–4040, 2020. ISSN 1941-0042. doi: 10.1109/tip.2020.2970248. URL http://dx.doi.org/10.1109/TIP.2020.2970248.

Sun, Y., Liu, J., and Kamilov, U. S. Block coordinate regularization by denoising. *NeurIPS*, 2019.

Sun, Y., Liu, J., and Kamilov, U. S. Block coordinate regularization by denoising. *IEEE Transactions on Computational Imaging*, 6:908–921, 2020. ISSN 2573-0436. doi: 10.1109/tci.2020.2996385. URL http://dx.doi.org/10.1109/TCI.2020.2996385.

Tan, S., Shen, Y., and Zhou, B. Improving the fairness of deep generative models without retraining, 2020.

Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., and Lin, C.-W. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, Nov 2020. ISSN 0893-6080. doi: 10.1016/j.neunet.2020.07.025. URL http://dx.doi.org/10.1016/j.neunet.2020.07.025.

Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Tripathi, S., Lipton, Z. C., and Nguyen, T. Q. Correction by projection: Denoising images with generative adversarial networks. *arXiv preprint arXiv:1803.04477*, 2018.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Wu, Y., Rosca, M., and Lillicrap, T. Deep compressed sensing, 2019.

Xiao, C., Li, B., Zhu, J.-Y., He, W., Liu, M., and Song, D. Generating adversarial examples with adversarial networks. *arXiv preprint arXiv:1801.02610*, 2018.

Yang, C., Ding, L., Chen, Y., and Li, H. Defending against gan-based deepfake attacks via transformation-aware adversarial faces. *arXiv preprint arXiv:2006.07421*, 2020.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21 (12):3106–3121, 2019.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5505–5514, 2018.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. Free-form image inpainting with gated convolution. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019. doi: 10.1109/iccv.2019.00457. URL http://dx.doi.org/10.1109/ICCV.2019.00457.