International Journal of Machine Learning and Cybernetics (2021) 12:2747–2759
https://doi.org/10.1007/s13042-021-01337-1

ORIGINAL ARTICLE

Compressed sensing using generative models based on fisher information

Meng Wang[1] · Jing Yu[2] · Zhen-Hu Ning[2] · Chuang-Bai Xiao[2]

Received: 5 April 2019 / Accepted: 21 April 2021 / Published online: 3 August 2021

Abstract
In compressed sensing applications, self-learning generative models have attracted increasing attention because they provide guarantees that are similar to those of standard compressed sensing without employing sparsity. However, improving the performances of a generative model is challenging. In this paper, we improve the recovery performances of generative models (generative adversarial networks) by making use of prior knowledge about the support of the vector of the original signal in the relevant domain. We demonstrate the advantage of using a parametric model with the Fisher distance metric for the exploitation of a distribution over the support when constraints on the distribution have been specified. We combine the generative model with the Fisher distance to study the recovery of sparse signals that satisfy a distribution for the purpose of improving the recovery performance of the model when there are some constraints on the distribution. Finally, we present the results of extensive experiments conducted on simulated signals and imaging signals.

Keywords Compressed sensing · Fisher information · Generative adversarial networks

1 Introduction

Compressed sensing (CS) is an efficient method for recovering sparse signals of interest from a substantially reduced number of samples, and it has been widely used in many fields [1–3]. The main objective of CS is to recover the original signals from the following acquisition system [4]:

$$y = \Phi x + e$$

where $x \in R^N$ is the original signal, $y \in R^M$ is the compressed signal, $\Phi = (\varphi_{ij})_{M\times N}$ is the measurement matrix, e is the noise term, and $M << N$ are integers.

To uniquely recover x given y and Φ, x must be inherently sparse or sparse in a specified basis Ψ, that is, $\Psi^{-1}x$ must be sparse. A signal is referred to as S-sparse if it has at most S nonzero elements.

The ideal procedure for reconstructing x is based on the following l_0-minimization problem:

$$\text{P0} \quad \tilde{x} = \arg\min_{x\in R^N} \|x\|_{l_0} \quad s.t. \quad y = \Phi x$$

which is an NP-hard problem [5]. However, under condition [6], the l_0-norm is equivalent to the l_1-norm. Using the l_1-minimization procedure, the problem becomes

$$\text{P1} \quad \tilde{x} = \arg\min_{x\in R^N} \|x\|_{l_1} \quad s.t. \quad y = \Phi x$$

To address the above questions, recovery frameworks for CS have been extensively studied in the literature [7–10]. Orthogonal matching pursuit (OMP) [11] is a typical greedy algorithm that selects the column at each step that is most correlated with the current residuals. In a variant of the OMP algorithm, namely, random refined OMP for sparse solutions, a false discovery rate control mechanism is employed to generate several sparse representations of the original signal [12]. Another classic recovery framework is basis pursuit (BP), which is a strategy for decomposing a signal into an "optimal" superposition of dictionary elements, where optimal refers to the decomposition having the smallest l_1-norm for its coefficients among all such decompositions [13]. In 2014, Bazerque

✉ Zhen-Hu Ning
ningzhenhu@126.com

[1] Beijing Information Science and Technology University, Beijing, China

[2] Beijing University of Technology, Beijing, China

et al. [14] concluded that the prism of reproducing kernel Hilbert spaces promotes nonparametric BP as the overarching framework for kernel-based learning approaches that leverage sparse linear regression, nuclear-norm regularization, and dictionary learning. In 2010, Baron et al. [15] performed asymptotically optimal Bayesian inference using belief propagation decoding; this technique represents the CS encoding matrix as a graphical model. Fast computation is realized by reducing the size of the graphical model using sparse encoding matrices. Furthermore, the statistical structure of the wavelet coefficients is exploited explicitly in wavelet-based Bayesian compressed sensing [16] to recover the compressible data from a wavelet basis. Recent advances in deep learning [17] have yielded encouraging results with respect to CS [18–20].

To learn a statistical structure automatically, generative adversarial networks (GANs) [21] have been introduced into CS. GANs are a class of machine learning systems that play a mini-max game to build a generative model. Instead of relying on sparsity, Bora et al. [22] presented an algorithm that uses GANs for CS. It captures the structure of vectors in the relevant domain of the original signal. The theoretical results demonstrate that gradient descent yields a satisfactory approximate solution for CS problems. Considering image diagnostic quality, Mardani et al. [23] proposed a novel CS framework that utilizes the advantages of a GAN to train a (low-dimensional) manifold of diagnostic-quality images from historical patients by leveraging a mixture of least-squares GANs and pixel-wise l_1-costs. Refine GAN [24] is a variant of the fully residual convolutional autoencoder and regular GANs that is designed for the CS-MRI (magnetie resonance imaging) formulation; it employs deep generator and discriminator networks with cyclic data consistency losses to realize a faithful interpolation of the specified under-sampled k-space data. In 2018, a conditional GAN-based model was proposed that bridges a substantial gap between conventional nonlearning CS methods using only data from a single image [25].

Some prior information about signals has been investigated for the purpose of improving the reconstruction performances of various models. CS has been widely used in many applications, such as medical imaging, monitoring, and communication [26–28]. The analysis of various applications enables the extraction of prior information about the distribution over the support set of a given signal's sparse representation to improve the performance of a model during the reconstruction stage [29–31].

However, there is no recovery framework for GANs embedded with prior information. To combine GANs for CS with prior information, we propose a new method in this paper. We describe the prior knowledge as the distribution over the support set, which is represented as a parametric function. In addition, we apply a parametric model (GAN) with the Fisher distance [32] to exploit the distribution when it is constrained. We propose a Fisher distance-based generative adversarial networks (FD-GANs) reconstruction algorithm. With Fisher information theory, parametric measurement models are defined as fundamental geometric objects. The constraints are exploited with the help of geometric properties.

The remainder of this paper is organized as follows: In Sect. 2, we describe our main method. In Sect. 3, we present the results of experiments conducted on simulated signals and imaging signals. We draw conclusions about our main theorem in Sect. 4.

2 Materials and methods

2.1 Overview of the proposed method

We present the FD-GANs reconstruction algorithm for CS, which uses GANs based on Fisher information. Similar to most GANs, the FD-GANs operates in two modes: generation and discrimination. The generative model, which is denoted as G, is trained with a novel loss function that contains prior knowledge about the support of the signal of interest for capturing the true distribution over the signal. It is used to generate the recovered signals as mimics of the original signals. In contrast, the discriminative model, which is denoted as D, is used to differentiate between the real signal and the mimics. After the system with G and D is balanced, gradient descent is utilized to optimize the input noise of G for generating the recovered signals. The complete procedure is described in Algorithm 1.

Algorithm 1.

Inputs:

y the measurement of the signal of interest;

p_z the distribution over the input noise item z;

p_{true} the distribution over the original signal;

k the number of steps;

$train\_iter_{\max}$ the maximal iteration of training;

$up\_iter_{\max}$ the maximal iteration of updating in gradient descent;

for $train\_iter_{\max}$ iterations do

for k steps do

Sample a batch of samples $\{z^j\}$ (for $j \in \{1,2,\ldots,L\}$) from p_z;

Sample a batch of true signals $\{x^j\}$ (for $j \in \{1,2,\ldots,L\}$) from p_{true};

Update the discriminator by the maximization problem:

$$\max_D E_{x\sim p_{true}(x)}[D(x)] + E_{z\sim p_z}[1 - D(G(z))]$$

end

Sample a batch of samples $\{z^j\}$ (for $j \in \{1,2,\ldots,L\}$) from p_z;

Update the generator by the minimization problem:

$$\min_G E_{x\sim p_{true}(x)}[D(x)] + E_{z\sim p_z}[1 - D(G(z))] + \alpha E_{z\sim p_z}\left\{d_F\left[p(\xi|\theta), p(\xi|\tilde{\theta})\right]\right\}$$

end

Optimize the input noise $\tilde{z}$ by the minimization problem:

$$\min_G \alpha E_{z\sim p_z}\left\{d_F\left[p(\xi|\theta), p(\xi|\tilde{\theta})\right]\right\} + \beta E_{z\sim p_z}\left[\left\|\Phi G(\tilde{z}) - y\right\|^2\right]$$

Output:

$G(\tilde{z})$ the recovered signal.

The whole procedure is divided into two parts: the modelling of the GANs and the reconstruction of the signal. Next, we describe these two phases in Sects. 2.2 and 2.3.

2.2 Modelling

2.2.1 Probability model

We describe the prior information, which is represented by the distribution over the indexes in the support set of the signal's sparse representation $x \in R^N$, as a probability density function (PDF). That is

$$p(t = i) = p_i \tag{1}$$

where t is the index and $i \in \{0, 1, \ldots, N-1\}$. As a discrete distribution, its PDF can be rewritten in the form of an exponential family [33] as

$$p(\xi|\theta) = \exp\left\{\sum_{i=1}^{N-1} \theta_i \xi_i - \psi(\theta)\right\} \tag{2}$$

where

$$\xi_i = \delta_i(t) = \begin{cases} 1, & t = i \\ 0, & t \neq i \end{cases} \tag{3}$$

$$\theta_i = \log \frac{p_i}{p_0} \tag{4}$$

Then,

$$p_0 = 1 - \sum_{i=1}^{N-1} p_i = 1 - p_0 \sum_{i=1}^{N-1} e^{\theta_i} \tag{5}$$

From Eq. (5), we obtain

$$\psi(\theta) = \log\left(1 + \sum_{i=1}^{N-1} e^{\theta^i}\right) \tag{6}$$

2.2.2 Generative model

Based on the above discussion about probability, we state the details about how the generative model G captures the distribution followed by the original signal x in this section.

G maps noise vectors z with a predefined distribution p_z to reconstruct the data. For the recovery of the j th signal, namely, x^j, the mapping becomes

$$\tilde{x}^j = G(z^j) \tag{7}$$

for $j \in \{1, 2, \ldots, L\}$, where L is the size of the batch of noise vectors z.

Then, we acquire a group of samples with G. We describe the distribution over the samples as a PDF since the conditional distributions of the stochastic units in neural networks can be effectively approximated with a distribution from Table 1. In the output layer of G, each unit corresponds to an index regarding the position of an element in x. Thus, for a pair containing an input z^j and an output $\tilde{x}^j$, the conditional distribution of the entry in the i th position can be represented as $f\left(\tilde{x}_i^j|\tilde{\theta}(z^j)\right)$, where $\tilde{x}_i^j$ is the i th entry of the output $\tilde{x}^j$ and $\tilde{\theta}$ denotes the estimated parameters conditioned on z^j (i.e., the parameters ϖ in the Table 1). Additional details are provided in Ref. [34].

With the definition of the support, the PDF of position t learned by the GANs is calculated as follows

$$\tilde{p}(t = i|z^j) = P\{\tilde{x}_i^j|\tilde{x}_i^j > b\} = 1 - \int_{\tilde{x}_i^j \le b} f\left(\tilde{x}_i^j|\tilde{\theta}(z^j)\right) d\tilde{x}_i^j \tag{8}$$

where b is the lower bound of the magnitude of the nonzero elements in the support set and i is the index of the entry in a signal $x \in R^N$ for $i \in \{0, 1, ..., N-1\}$. As an estimator of $p(t = i)$, $\tilde{p}(t = i|z^j)$ is also a discrete distribution. Thus, it can be rewritten in the form of an exponential family through the method used in Sect. 2.2.1. The rewritten function is denoted as $p\left(\xi|\tilde{\theta}(z^j)\right)$. We denote the parameter estimated by the GAN as $\bar{\tilde{\theta}}$. Then,

$$\begin{aligned} p(\xi|\bar{\tilde{\theta}}) &= \int p\left(\xi|\tilde{\theta}(z^j)\right) p_z(z^j) dz^j \\ &= \int p\left(\xi|\tilde{\theta}(z^j)\right) p_z(z^j) dz^j \\ &= E_{z\sim p_z}\left[p\left(\xi|\tilde{\theta}(z^j)\right)\right] \\ &= \frac{1}{L}\sum_{j=1}^{L} p\left(\xi|\tilde{\theta}(z^j)\right) \end{aligned} \tag{9}$$

where L is the size of the training set.

2.2.3 Generative adversarial model

To improve the performance of the model during reconstruction, G is trained to follow the prior known distribution over the support of the signal's sparse expression. To embed the prior knowledge about the support into G, we minimize the difference between $p(\xi|\bar{\tilde{\theta}})$ and $p(\xi|\theta)$. A classic parametric space for the exponential family of PDFs is

$$H = \{\theta \in R^N\} \tag{10}$$

For distributions that are represented in a non-Euclidean plane, the Fisher information matrix (FIM) is analogous to a Riemannian metric tensor [35]. With Fisher information theory, parametric measurement models are defined as fundamental geometric objects that have reparameterization-invariant

Table 1 Stochastic unit, conditional distribution, and approximation

Unit name	$p(\varpi)$	$f(x\|\varpi)$
Sigmoid unit	$(1+e^{-\omega})^{-1}$	$\exp\{\varpi x - \log(1+\exp(\varpi))\}$
Noisy Tanh unit	$(1+e^{-\varpi})^{-1} - \frac{1}{2}$	$\exp\{\varpi x - \log(1+\exp(\varpi)) + ent(x) + g(x)\}$
Linear (Gaussian) unit	ϖ	$\exp\left\{\varpi x - \frac{1}{2}\varpi^2 - \frac{1}{2}x^2 - \log(\sqrt{2\pi})\right\}$
Softplus unit	$\log(1+e^{\varpi})$	$\exp\left\{\varpi x - 2Li_2(-e^{\varpi}) - x\log(1-e^x) + y\log(e^{\varpi}-1) + g(x)\right\}$
Rectified linear unit	$\max(0, \varpi)$	$N(f(\varpi), \sigma(\varpi > 0)\varpi)$
Exponential unit	e^{ϖ}	$N(e^{\varpi}, e^{\varpi})$
Sinh unit	$\frac{1}{2}(e^{\varpi} - e^{-\varpi})$	$N(\sinh(\varpi), \cosh(\varpi))$
Poisson unit	e^{ϖ}	$\exp\{\varpi x - e^{\varpi} - y!\}$

properties. The Fisher distance arises from the FIM, and it is a measure of the amount of distinct information contained between two distributions. Furthermore, it is more accurate than some popular divergence measures, such as Kullback–Leibler divergence [36]. For the exponential family, the Fisher distance is calculated as

$$d_F\big(p(x|\theta_1), p(x|\theta_2)\big) = \psi(\theta_1) + \phi(\eta_2) - \theta_1 \cdot \eta_2 \tag{11}$$

where x is a random variable, while η_2 denotes the dual coordinates of θ_2 [37]. In addition,

$$\eta_2 = \frac{\partial \psi(\theta_2)}{\partial \theta_2} \tag{12}$$

and

$$\phi(\eta_2) = \int p(x|\theta_2) \log p(x|\theta_2) dx \tag{13}$$

Substituting $p\big(\xi|\tilde{\theta}(z^j)\big)$ and $p(\xi|\theta)$ into Eq. (11), we obtain

$$d_F\big(p(\xi|\theta), p\big(\xi|\tilde{\theta}(z^j)\big)\big) = \psi(\theta) + \phi\big(\tilde{\eta}(z^j)\big) - \theta \cdot \tilde{\eta}(z^j) \tag{14}$$

d_F is a convex function. From Eqs. (9), (14) and Jensen's inequality for convex function [38], we obtain an upper bound for $d_F\Big(p(\xi|\theta), p(\xi|\bar{\tilde{\theta}})\Big)$. It is described as follows

$$\begin{aligned} & d_F\Big(p(\xi|\theta), p(\xi|\bar{\tilde{\theta}})\Big) \\ &= d_F\Big(p(\xi|\theta), E_{z\sim p_z}\big[p\big(\xi|\tilde{\theta}(z^j)\big)\big]\Big) \\ &\le E_{z\sim p_z}\big\{d_F\big(p(\xi|\theta), p\big(\xi|\tilde{\theta}(z^j)\big)\big)\big\} \\ &= \frac{1}{L}\sum_{j=1}^{L} d_F\big(p(\xi|\theta), p\big(\xi|\tilde{\theta}(z^j)\big)\big) \end{aligned} \tag{15}$$

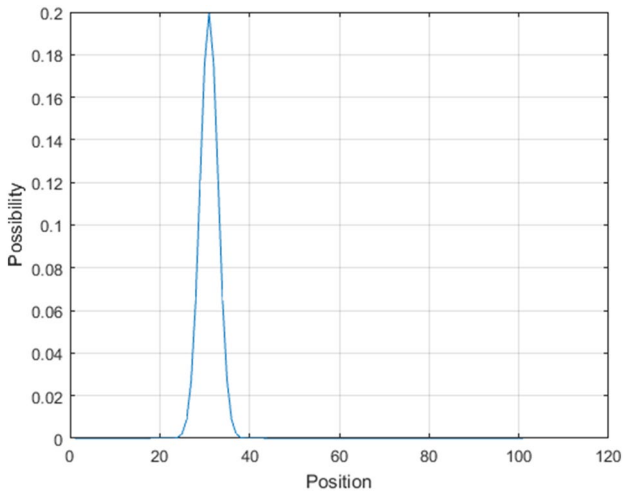

Fig. 1 PDF for the positions of the nonzero entries of the original signal $x \in R^{100}$

A discriminative network D distinguishes the candidates that are produced by G from the true data distribution. The procedure for training the GANs is summarized as the following two-player mini-max game defined by loss functions:

Generation:

$$\begin{aligned} & \min_G \; E_{x\sim p_{true}(x)}[D(x)] \\ & + E_{z\sim p_z}[1 - D(G(z))] \\ & + \alpha E_{z\sim p_z}\Big\{d_F\Big(p(\xi|\theta), p(\xi|\tilde{\theta})\Big)\Big\} \end{aligned} \tag{16}$$

where the sum of the first two terms is the traditional objective function in GANs, while the third term is the Fisher distance in Eq. (15). In practice, we minimize the last term in Eq. (16) via the optimization of the mean in Eq. (15). The main objective of the optimization procedure is to build a generative model for the signal.

Discrimination:

$$\max_D \; E_{x\sim p_{true}(x)}[D(x)] + E_{z\sim p_z}[1 - D(G(z))] \tag{17}$$

This function is the same as the objective function in traditional GANs. The main objective of the optimization procedure is to build a discriminative model for the signal.

The loss functions discussed above are optimized by the traditional stochastic gradient descent.

2.3 Reconstruction

Because the goal of training is to find a recovered signal $\tilde{x}$ such that the corresponding vector $\tilde{y} = \Phi\tilde{x}$ matches the observed measurements, the mean measurement loss is defined as

$$Loss_{MML} = E[||y - \tilde{y}||_2^2] \tag{18}$$

After training, a signal built by the generative model with any input noise follows the structure and distribution of the original signal. With the stable recovery condition of the signal [39], the unique solution is the signal with zero measurement loss. To make the condition tighter, a loss term for the Fisher distance in Eq. (15) is added to the mean measurement loss in Eq. (18). The minimization of a loss function can be expressed as

$$\min_G \; \alpha E_{z\sim p_z}\big\{d_F\big(p(\xi|\theta), p(\xi|\tilde{\theta})\big)\big\} + \beta E_{z\sim p_z}\big[\|\Phi G(z) - y\|^2\big] \tag{19}$$

Here, gradient descent is utilized to optimize the input noise z. Then, the optimized input noise is denoted as $\tilde{z}$. Then, we obtain the final output $G(\tilde{z})$, namely, the recovered signal.

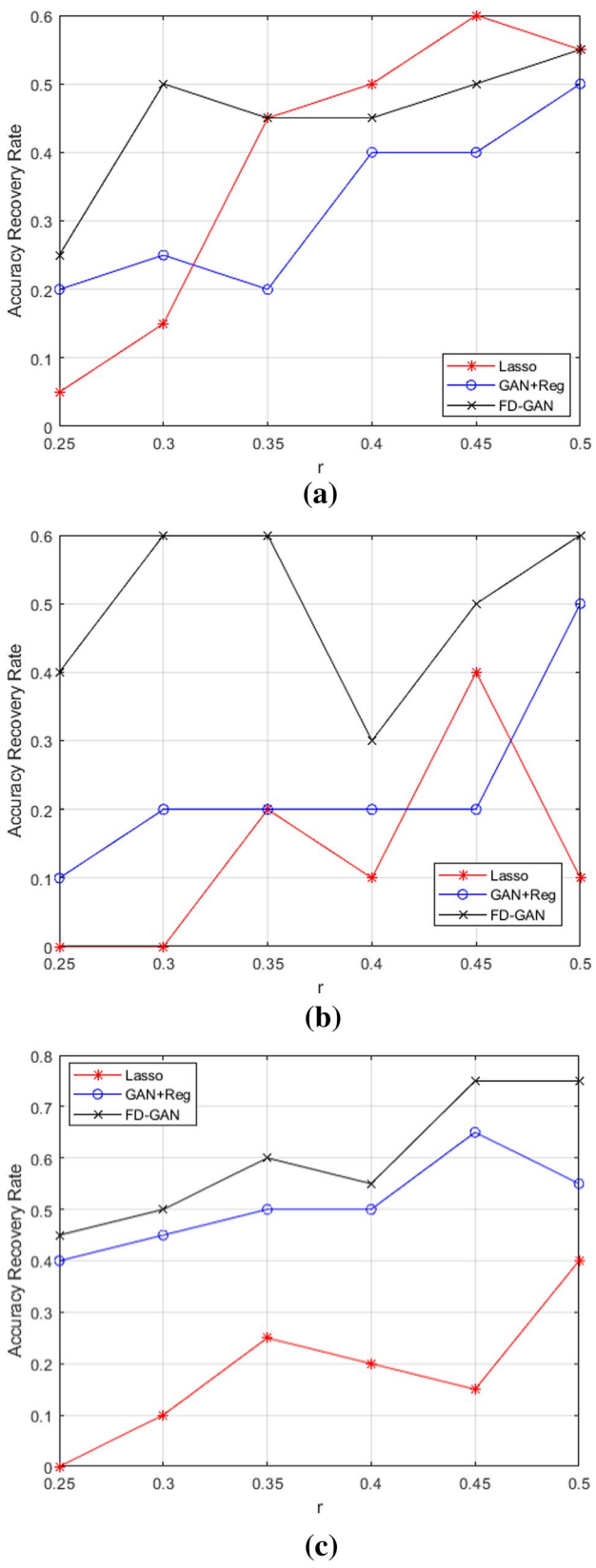

Fig. 2 Comparison among Lasso, GAN+Reg and FD-GAN with a specified distribution. **a** $S=15$. **b** $S=20$. **c** $S=25$

3 Results and discussion

3.1 Simulation

In this section, we compare the proposed networks to previous methods (GAN + Reg [22] and Lasso [40]) on a dataset following a specified distribution. The distribution underlying the support of the data is plotted in Fig. 1. The latitudes of the entries decay in a manner that is similar to a power law, as in [41].

N is set to 100. The number of nonzero entries in the signal S takes values in $\{15, 20, 25\}$, while the sampling rate r ($r = M/N$) ranges from 0.25 to 0.5 (at intervals of 0.05). For each case, we generate 1000 signals randomly, on which we conduct a series of experiments. The architecture of the generator is 20–100 and the architecture of the discriminator is 100–10–2. We determine the coefficients in Eqs. (16) and (19) experimentally: α is set to 0.1, while β is set to 1. We set the maximum number of training iterations, namely, $train\_iter_{\max}$, to 100, while the maximum number of update iterations, namely, $up\_iter_{\max}$, for the testing process is set to 50. The signal is recovered from a noisy measurement with $SNR_{mes} = 30dB$. We perform 10 random restarts with 50 steps per restart and select the reconstruction with the lowest 2-norm loss.

The accuracy recovery is defined as

$$\|x - \tilde{x}\|_2 / \|x\|_2 \leq 10^{-2} \tag{20}$$

The accuracy recovery rate is the ratio of the number of signals that satisfy Eq. (20) to the total amount. In Fig. 2, the results are plotted versus the accuracy recovery rate.

The results demonstrate that the FD-GAN outperforms GAN + Reg in all cases. Hence, the proposed method is successful in its GAN modification. When $S = 15$, the generative models and Lasso are almost the same. As S increases, the performance improvement of the generative models over Lasso increases. This is because the generative models provide stable reconstruction results without employing sparsity.

3.2 Real data

In this section, we test the proposed method with two popular datasets, including the MNIST dataset [42] and CelebA dataset [43].

3.2.1 Experiment with MNIST

We compare the proposed networks to previous methods (GAN + Reg and Lasso) using various numbers of measurements ($M = 50$, 100, 200, 300) on the MNIST dataset, which includes 60,000 images of handwritten digits of size

Table 2 (a) $Loss_{MML}$ for various numbers of measurements ($M = 50$, 100, 200, 300). (b) $Loss_{MNL}$ for various numbers of measurements ($M = 50$, 100, 200, 300)

(a)	Lasso	GAN + Reg	FD-GAN
50	26.578	20.548	16.548
100	21.805	15.177	13.033
200	18.727	14.627	13.616
300	20.920	13.8704	9.8768
(b)	Lasso	GAN + Reg	FD-GAN
50	0.103	0.073	0.035
100	0.089	0.020	0.016
200	0.077	0.015	0.010
300	0.075	0.011	0.009

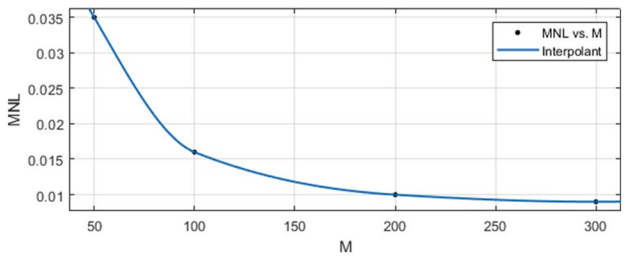

Fig. 3 Results of $Loss_{MNL}$ versus M. The dots represent the raw data and the curve is the corresponding interpolant

28×28. The support for the imaging signals in the discrete cosine transform (DCT) domain follows a specified distribution. In addition, our objective is to reconstruct these signals. After being vectorized, the signal of interest x is transformed into a vector with 784 dimensions. We set $S = 80$ (approximately 10% of the entries of the signal are nonzero). The architecture of the generator is 20–500–500–784, and the architecture of the discriminator is 784–30–2. We determine the coefficients in Eqs. (16) and (19) experimentally: α is set to 0.1, while β is set to 1. We set the maximal number of training iterations, namely, $train\_iter_{\max}$, to 500, and the maximal number of update iterations, namely, $up\_iter_{\max}$, for the testing process is set to 200. For each sample, there exists additional noise that follows a Gaussian distribution, namely, $N(0, 4)$. We perform 10 random restarts with 50 steps per restart and select the reconstruction that has the lowest measurement error in Eq. (19). The input is the compressed signal y. The output is the recovered signal $\tilde{x}$, i.e., the vector containing the DCT coefficients of the image.

We measure the convergence of the compared algorithms with the mean measurement loss defined in Eq. (18) because

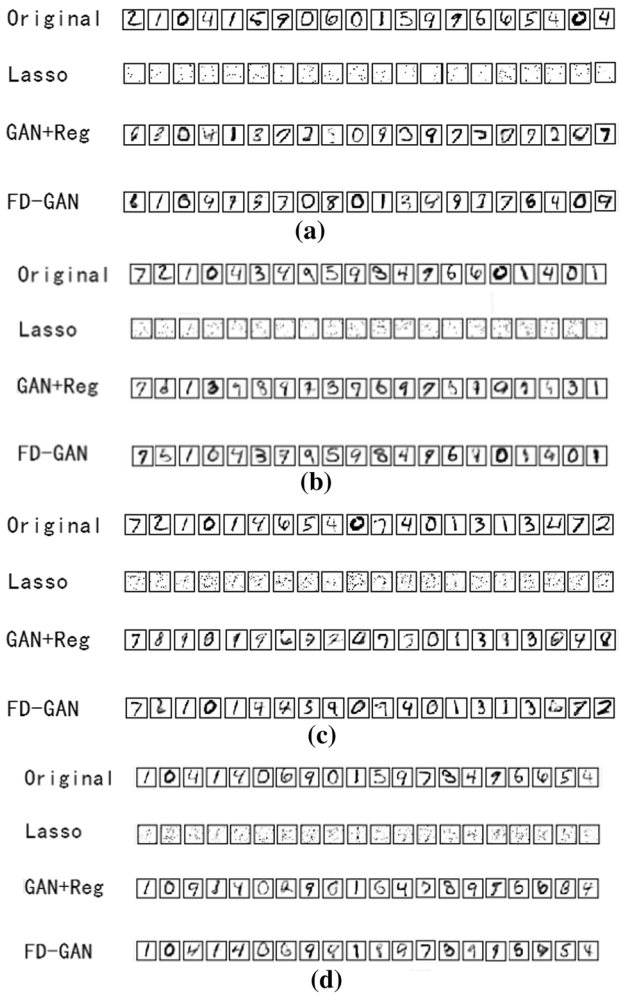

Fig. 4 Illustration of the recovered images for various numbers of measurements ($M = 50$, 100, 200, 300). **a** $M = 50$. **b** $M = 100$. **c** $M = 200$. **d** $M = 300$

Table 3 (a) $Loss_{MML}$ for various numbers of measurements ($M = 19{,}000$, 22,000, 30,000). (b) $Loss_{MNL}$ for various numbers of measurements ($M = 19{,}000$, 22,000, 30,000)

(a)	Lasso	GAN + Reg	FD-GAN
19,000	34.536	16.434	12.134
22,000	11.950	10.941	3.520
30,000	52.326	18.947	12.905
(b)	Lasso	GAN + Reg	FD-GAN
19,000	0.746	0.571	0.222
22,000	0.221	0.119	0.057
30,000	0.174	0.066	0.025

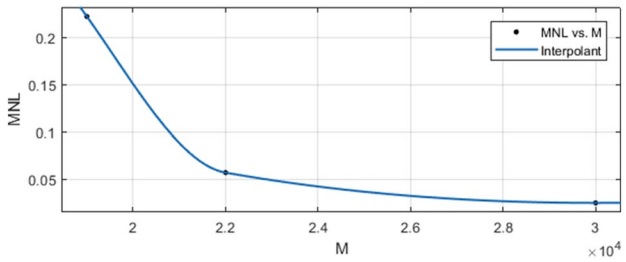

Fig. 5 Results of $Loss_{MNL}$ versus M. The dots represent the raw data, and the curve is the corresponding interpolant

all of the algorithms in the experiments aim to minimize the loss. The mean 2-norm loss ratio is another indicator that reflects the difference between the original signal and the recovered signal. It is defined as

$$Loss_{MNL} = E\left[||x - \tilde{x}||_2^2 / ||x||_2^2\right] \tag{21}$$

It is associated with the reconstructed error.

In Table 2, the performances on the test set are specified numerically.

M influences the performance of the FD-GAN. We plot the results as a function of M in Fig. 3. The raw data of the FD-GAN in Table 1b are fitted with an interpolant curve. The recovered images are shown in Fig. 4.

3.2.2 Experiment with CelebA

We compare the proposed networks to previous methods (GAN + Reg and Lasso) using various numbers of measurements ($M = 19{,}000$, 22,000, 30,000) on the CelebA dataset, which includes 202,599 images of faces of size 218×178. It is known that the support for the imaging signals in the DCT domain follows a specified distribution. In addition, our objective is to reconstruct these signals. After being vectorized, the signal of interest x is transformed into a vector with 38,804 dimensions. We set $S = 3800$ (approximately 10% of the entries of the signal are nonzero). The architecture of the generator is 1000–38,804, and the architecture of the discriminator is 38804–280–2. We determine the coefficients in Eqs. (16) and (19) experimentally: α is set to 0.05, while β is set to 1. We set the maximal number of training iterations, namely, $train\_iter_{\max}$, to 500, and the maximal number of update iterations, namely, $up\_iter_{\max}$, for the testing process is set to 200. For each sample, there exists additional noise that follows a Gaussian distribution, namely, $N(0, 4)$. We perform 10 random restarts with 50 steps per restart and select the reconstruction that has the lowest measurement error in Eq. (19). The input is the compressed signal y. The output is the recovered signal $\tilde{x}$, i.e., the vector containing the DCT coefficients of the image.

We measure the convergence of the algorithms with the mean measurement loss. The mean 2-norm loss is another indication that reflects the difference between the original signal and the recovered signal. In Table 3, the performances on the test set are specified numerically.

M influences the performance of the FD-GAN. We plot the results as a function of M in Fig. 5. In Fig. 5, the raw data of the FD-GAN in Table 3b are fitted with an interpolant curve. Some of the recovered images are shown in Figs. 6, 7 and 8.

3.2.3 Discussion

In Tables 2 and 3, the proposed method numerically outperforms the previous methods. In Tables 2a and 3a, the $Loss_{MML}$ of the FD-GAN is lower than those of the previous methods under different conditions. Thus, the FD-GAN has a better convergence property than those of the previous methods. In Tables 2b and 3b, the $Loss_{MNL}$ of the FD-GAN is lower than those of the previous methods under different conditions. Thus, the FD-GAN yields better reconstruction quality than the previous methods.

The recovered images in Figs. 4 and 6 validate this conclusion visually. In Fig. 4, the images that are reconstructed via the FD-GAN are most easily distinguishable. GAN + Reg makes some mistakes. For example, it generates the Arabic

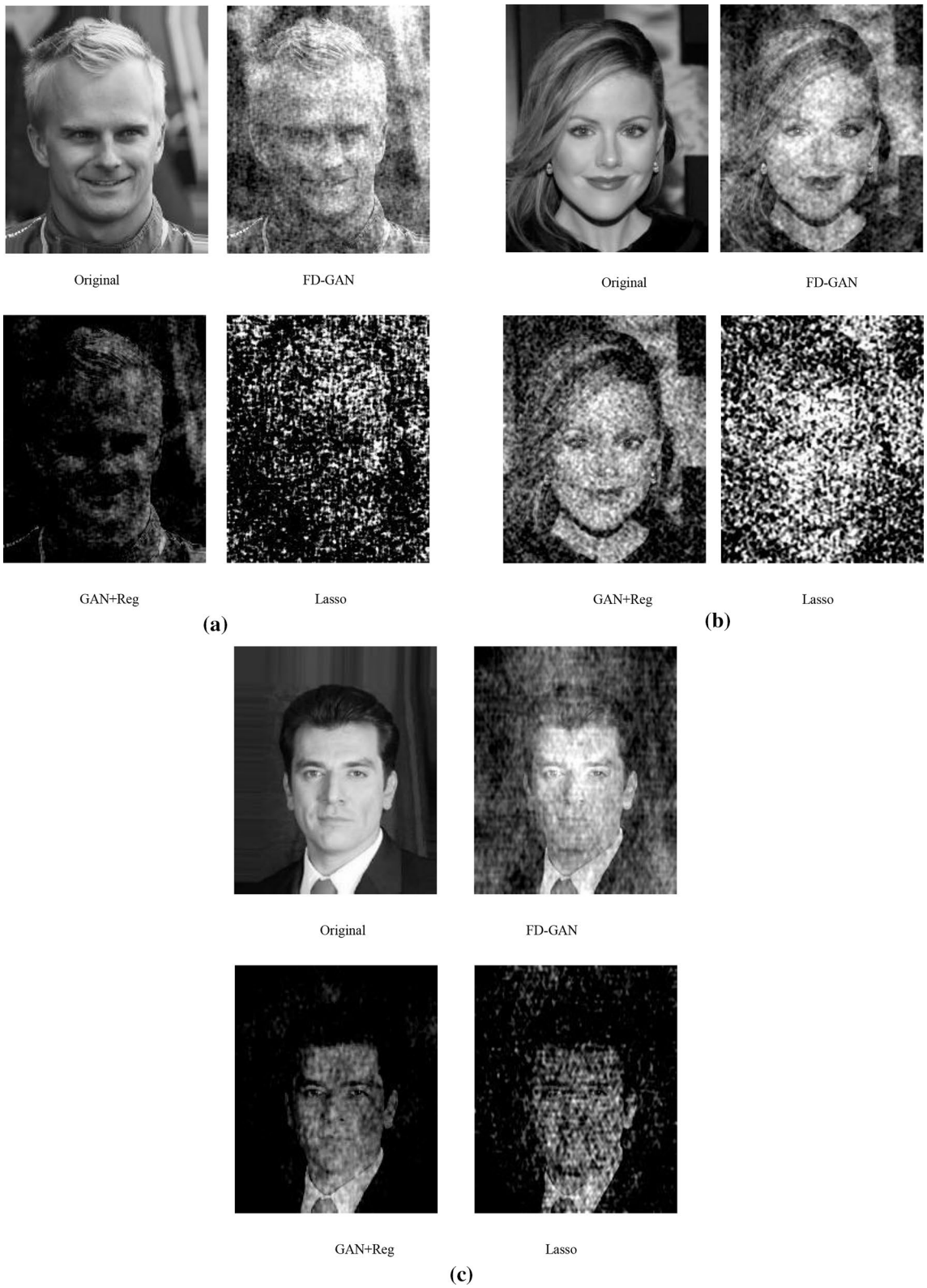

Fig. 6 Illustrations of the recovered images for $M = 19000$. **a** Original, FD-GAN, GAN + Reg, Lasso. **b** Original, FD-GAN, GAN + Reg, Lasso. **c** Original, FD-GAN, GAN + Reg, Lasso

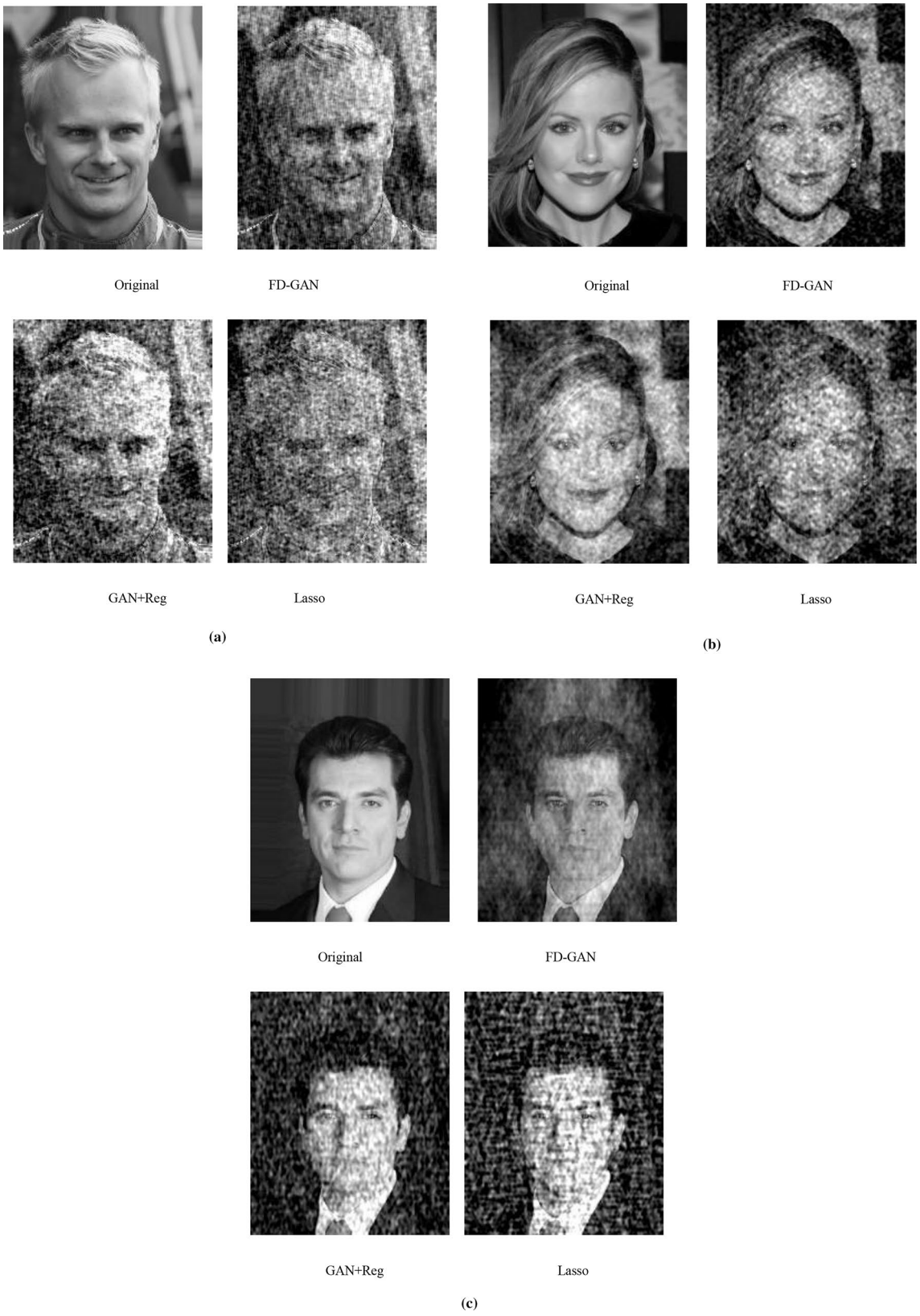

Fig. 7 Illustrations of the recovered images for $M = 22000$. **a** Original, FD-GAN, GAN + Reg, Lasso. **b** Original, FD-GAN, GAN + Reg, Lasso. **c** Original, FD-GAN, GAN + Reg, Lasso

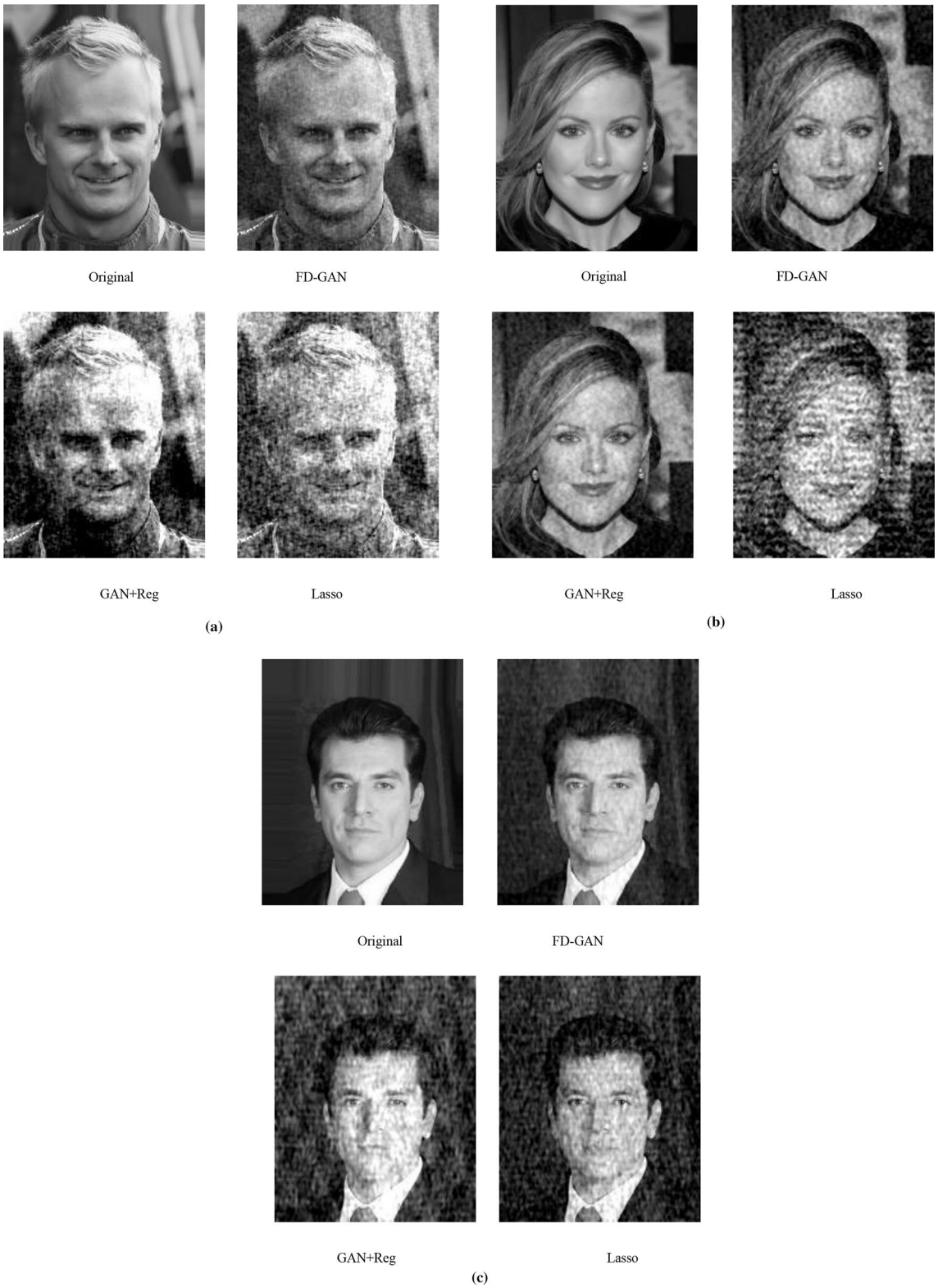

Fig. 8 Illustrations of the recovered images for $M = 30000$. **a** Original, FD-GAN, GAN + Reg, Lasso. **b** Original, FD-GAN, GAN + Reg, Lasso. **c** Original, FD-GAN, GAN + Reg, Lasso

numeral "3" when the original numeral is "0". In contrast, the FD-GAN rarely makes such mistakes. The results demonstrate that our algorithm outperforms previous methods under various numbers of measurements. In Figs. 6, 7 and 8, the images that are reconstructed via the FD-GAN are clearest. In particular, more facial details are shown in the images reconstructed via the FD-GAN than in those constructed by the competing methods.

In Figs. 3 and 5, we illustrate the correlation between the reconstruction quality and the number of measurements M. According to the curves, the losses tend to decline as M increases.

4 Conclusion

In this paper, we propose an FD-GAN reconstruction algorithm. We describe the prior knowledge as the distribution over the support, and we use the Fisher distance to exploit the distribution if the constraints on the distribution are known. We demonstrate that the FD-GAN algorithm outperforms the previous methods in terms of convergence and reconstruction quality under various numbers of measurements.

Acknowledgements This research was funded by the Beijing Science and Technology Planning Program of China (Z171100004717001), Beijing Natural Science Foundation (4172002), and Natural Science Foundation of China (61701009).

References

1. Abdur RM, Simon D (2014) Energy-efficient sensing in wireless sensor networks using compressed sensing. Sensors 14(2):2822–2859
2. Wang DX, Mackie TR, Tome WA (2009) Proton computed tomography reconstruction using compressed sensing and prior image constrained compressed sensing. Med Phys 36(6):2443
3. Bo K, Gengxin Z, Wei Z, Dongming B, Zhidong X (2016) Data persistence in planetary surface network using raptor codes and probabilistic broadcasting Int. J Distrib Sens Netw 12(8):51–65
4. Candes EJ, Wakin MB (2008) An introduction to compressive sampling. IEEE Signal Proc Mag 25(2):21–30
5. Muthukrishnan S (2003) Data streams: algorithms and applications. Found Trends Theor Comput Sci 1(2):413–413
6. Candes EJ, Tao T (2005) Decoding by linear programming. IEEE Trans Inf Theory 51(12):4203–4215
7. Blumensath T, Davies ME (2009) Iterative hard thresholding for compressed sensing. Appl Comput Harmon Anal 27(3):265–274
8. Baron D, Sarvotham S, Baraniuk RG (2010) Bayesian compressive sensing via belief propagation. IEEE Signal Proc 58(3):269–280
9. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. SIAM J Sci Comput 20(1):33–61
10. Gilbert A, Guha S, Indyk P, Muthukrishnan S, Strauss M (2000) Near-optimal sparse Fourier representations via sampling.In: Proceedings of the 34th annual ACM symposium on theory of computing, pp 152–161
11. Cai TT, Wang L (2011) Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Trans Inf Theory 57(7):4680–4688
12. Li S, Fang L (2011) Signal denoising with random refined orthogonal matching pursuit. IEEE Instrum Meas Soc 23:26–34
13. Chen SS, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. SIAM Rev 43(1):129–159
14. Bazerque JA, Giannakis GB (2013) Nonparametric basis pursuit via sparse kernel-based learning: a unifying view with advances in blind methods. IEEE Signal Process Mag 30(4):112–125
15. Baron D, Sarvotham S, Baraniuk RG (2010) Bayesian compressive sensing via belief propagation. IEEE Trans Signal Process 58(1):269–280
16. He L, Carin L (2009) Exploiting structure in wavelet-based bayesian compressive sensing. IEEE Trans Signal Process 57(9):3488–3497
17. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
18. Palangi H, Ward R, Deng Li (2017) Convolutional deep stacking networks for distributed compressive sensing. Signal Process 131:181–189
19. Wang M, Xiao CB, Ning ZH, Li T, Gong B (2019) Neural networks for compressed sensing based on information geometry. Circuits Syst Signal Process 38(2):569–589
20. Merhej D et al (2011) Embedding prior knowledge within compressed sensing by neural networks. IEEE Trans Neural Netw 22(10):1638–1649
21. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inf Process Syst vol 2, pp 2672–2680
22. Bora A, Jalal A, Price E, Dimakis AG (2018) Compressed sensing using generative models. In: Proceedings of the 34th international conference on machine learning, PMLR, vol 70, pp 537–546
23. Mardani M, Gong E, Cheng JY, Vasanawala S, Xing L, Pauly JM (2017) Deep generative adversarial networks for compressed sensing (GANCS) automates MRI. In: 31st Conference on neural information processing systems (NIPS 2017), Long Beach, CA, USA
24. Quan TM, Nguyen-Duc T, Jeong W-K (2018) Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE Trans Med Imaging 37(6):1488–1497
25. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, Liu F, Arridge S, Keegan J, Guo Y, Firmin D (2018) DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans Med Imaging 37(6):1310–1321
26. Hao L, Dawei X, Xingwu Y, Xinyu Z (2019) Compressed sensing method for IGBT high-speed switching time on-line monitoring. IEEE Trans Ind Electron 66(4):3185–3195
27. Razzaque MA, Simon D (2014) Energy-efficient sensing in wireless sensor networks using compressed sensing. Sensors 14:2822–2859
28. Hong-An L, Zhan-Li L, Zhuo-Ming D (2017) A reconstruction method of compressed sensing 3D medical models based on the weighted 0-norm. J Med Imaging Health Inform 7(2):416–420
29. Majumdar A, Ward RK (2010) Compressive color imaging with group-sparsity on analysis prior. In: Proc. 17th IEEE Int. Conf. Image Process. (ICIP), Sep. 26–29, pp 1337–1340
30. Yu L et al (2011) Bayesian compressive sensing for clustered sparse signals. In: Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), May 22–27, pp 3948–3951
31. Vera E et al (2009) Bayesian compressive sensing of wavelet coefficients using multiscale Laplacian priors. In: Proc. IEEE/Signal

Process. 15th Workshop Statist. Signal Process. (SSP), Aug. 31–Sep. 3, pp 229–232
32. Amari S (1985) Differential-geometrical methods in statistics. Lect Notes Stat 28(5):168–178
33. Duong TV, Phung DQ, Bui HH, Venkatesh S (2006) Human behavior recognition with generic exponential family duration modeling in the hidden semi-Markov model. In: International conference on pattern recognition, pp 202–208
34. Ravanbakhsh S, Poczos B, Schneider J (2016) Stochastic neural networks with monotonic activation functions. In: Appearing in Proceedings of the 19th international conference on artificial intelligence and statistics (AISTATS) 2016, Cadiz, Spain. JMLR: W&CP vol 41
35. Rao CR (1945) Information and accuracy attainable in the estimation of statistical parameters. Bull Calcutta Math Soc 37:81–91
36. Costa SIR, Santos SA, Strapasson JE (2014) Fisher information distance: a geometrical reading. Discrete Appl Math S0166218X14004211
37. Arwini K, Dodson CTJ (2008) Information geometry: near randomness and near independence. Lecture Notes in Mathematics, Springer
38. Hazewinkel M (ed) (2001) Jensen inequality. In: Encyclopedia of mathematics. Springer Science + Business Media B.V./Kluwer Academic Publishers, ISBN 978-1-55608-010-4
39. Candès EJ, Romberg JK, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math 59(8):1207–1223
40. Slavche P, Venceslav K (2015) Asymptotic capacity lower bound for an OFDM system with lasso compressed sensing channel estimation for Bernoulli-Gaussian channel. IEEE Commun Soc 19(3):379–382
41. Candes EJ, Tao T (2006) Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans Inf Theory 52(12):5406–5425
42. LeCun M, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
43. Liu Z, Luo P, Wang X et al (2015) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 3730–3738