Outlier Detection Using Generative Models with Theoretical Performance Guarantees

Jirong Yi and Duc Anh Le
Department of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
Email: {jirong-yi,ducanh-le}@uiowa.edu

Tianming Wang
Department of Electrical and Computer Engineering
University of Texas, Austin, Texas, 78712
Email: tw27479@utexas.edu

Xiaodong Wu and Weiyu Xu
Department of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
Email: {xiaodong-wu,weiyu-xu}@uiowa.edu

I. ABSTRACT AND INTRODUCTION

This paper considers the problem of recovering signals from compressed measurements contaminated with sparse outliers. In this paper, we propose a generative model neural network approach for reconstructing the ground truth signals under sparse outliers. We propose an iterative alternating direction method of multipliers (ADMM) algorithm for solving the outlier detection problem via \( \ell_1 \) norm minimization, and a gradient descent algorithm for solving the outlier detection problem via squared \( \ell_1 \) norm minimization. We establish the recovery guarantees for reconstruction of signals using generative models in the presence of outliers, and give an upper bound on the number of outliers allowed for recovery. Our results are applicable to both the linear generator neural network and the nonlinear generator neural network with an arbitrary number of layers. We conduct extensive experiments using variational auto-encoder and deep convolutional generative adversarial networks, and the experimental results show that the signals can be successfully reconstructed under outliers using our approach. Our approach outperforms the traditional Lasso and \( \ell_2 \) minimization approach.

II. PROBLEM STATEMENT

Let the generative model \( G(\cdot) : \mathbb{R}^k \to \mathbb{R}^n \) be a \( d \)-layer neural network. Define the bias terms in each layer as \( b^{(1)} \in \mathbb{R}^{n_1}, \ldots, b^{(d)} \in \mathbb{R}^{n_d} \), and the weight matrix \( W^{(1)} \) in the \( i \)-th layer as

\[
W^{(1)} = [w_{1}^{(1)}, \ldots, w_{n_1}^{(1)}]^T \in \mathbb{R}^{n_1 \times k}, w_j^{(1)} \in \mathbb{R}^k, j = 1, \ldots, n_1,
\]

\[
W^{(i)} = [w_{1}^{(i)}, \ldots, w_{n_i}^{(i)}]^T \in \mathbb{R}^{n_i \times n_{i-1}}, w_{j}^{(i)} \in \mathbb{R}^{n_{i-1}}, j = 1, \ldots, n_i,
\]

\[
W^{(d)} = [w_{1}^{(d)}, \ldots, w_{n_d}^{(d)}]^T \in \mathbb{R}^{n_d \times n_{d-1}}, w_{j}^{(d)} \in \mathbb{R}^{n_{d-1}}, j = 1, \ldots, n,
\]

where \( 1 < i < d \). The element-wise activation functions \( a(\cdot) \) in different layers are defined to be the same, and some commonly used activation functions include the leaky ReLU, i.e., \( a(x) = x, \) if \( x_i \geq 0 \) and \( a(x) = h x_i, \) if \( x_i < 0, \) where \( h \in (0, 1) \) is a constant. Now for a given input \( z \in \mathbb{R}^k \), the output will be

\[
G(z) = a \left( W^{(d)} \cdots a \left( W^{(1)} z + b^{(1)} \right) \right) \cdots + b^{(d)}.
\]

Given the measurement vector \( y \in \mathbb{R}^m \), i.e., \( y = M x_0 + e \), where the \( M \in \mathbb{R}^{m \times n} \) is a measurement matrix, \( x_0 \) is a signal to be recovered, and the \( e \) is an outlier vector due to the corruption of measurement process. The \( e \) and \( x_0 \) can be recovered via the following \( \ell_0 \) "norm" minimization if \( x \) lies within the range of \( G \) (II.3), and due to the NP hardness of it, we solve the \( \ell_1 \) minimization (II.4), i.e.,

\[
\min_{z \in \mathbb{R}^k} ||MG(z) - y||_0, \quad \text{(II.3)}
\]

\[
\min_{z \in \mathbb{R}^k} ||MG(z) - y||_1. \quad \text{(II.4)}
\]

We assume that for a true signal \( x_0 \in \mathbb{R}^n \), there is a low dimensional \( z_0 \in \mathbb{R}^k \) such that \( G(z_0) = x_0 \). Once we get \( z_0 \) by solving (II.4), the \( G(z_0) \) will be the estimation of true signal \( x_0 \).

III. MAIN RESULTS

The main results of this work can be summarized as Theorem III.1. More details of the theorem and its proof can be found in [1].

**Theorem III.1.** Let the generative model \( G(\cdot) : \mathbb{R}^k \to \mathbb{R}^n \) be implemented by a \( d \)-layer neural network in (II.2). Let the weight matrix in each layer be defined in (II.1) with \( n_i \geq k, i = 1, \ldots, d-1 \) and \( k \leq n - 2(d-1)+1 \) according to the standard Gaussian distribution \( \mathcal{N}(0, 1) \). Let each entry in each weight matrix \( W^{(i)} \) of the well-trained generator be distributed randomly according to the standard Gaussian distribution \( \mathcal{N}(0, 1) \). Let the weight matrix \( M \in \mathbb{R}^{m \times n} \) be a random matrix with all entries i.i.d. random according to the Gaussian distribution and \( \min(n, m) > l \). Then with high probability, the \( z_0 \) can be recovered by solving (II.3).

We present numerical results to show the recovery performance of the proposed approach in real data sets using commonly used generative models. For MNIST data set [2], [3], we apply the variational auto-encoder structure [4], and the result is shown in Figure 3. It shows that the proposed generative-model-based approaches, i.e., VAE-based \( \ell_1 \) minimization solved by ADMM algorithm (VAE \( \ell_1 \) ADMM) and VAE-based squared \( \ell_1 \) minimization solved by gradient descent algorithm (VAE (\( \ell_1^2 \)) GD) can achieve successful recovery in Figure 3a. While the VAE-based approach (Lasso) and VAE-based squared \( \ell_2 \) minimization solved by gradient descent algorithm (VAE (\( \ell_2^2 \)) GD) cannot successfully recover the images in Figure 3b. For CelebA data set [5], we apply the deep convolutional generative adversarial network [6], and the results are shown in Figure 4 and 5. The proposed approaches can achieve successful recovery while the other methods fail.

IV. CONCLUSIONS

The proposed generative-model-based approach can outperform traditional outlier detection methods. The theoretical recovery guarantees for \( \ell_0 \) minimization can be established as long as the weights of a well-trained network satisfies mild conditions.
REFERENCES


