DOI:10.13718/j.cnki.xsxb.2023.01.011

基于生成模型的块稀疏偏差建模①

斯那雨追， 余鹏彬， 王建军

西南大学 数学与统计学院，重庆 400715

摘要：研究了基于生成模型的块稀疏偏差模型(Block Sparse-Gen). 依靠块 RIP 条件和块 REC 条件，从理论上给出了最优解码的重构误差以及高概率恢复生成函数的测量次数. 实验数值也验证了 Block Sparse-Gen 的有效性.

关 键 词：生成模型；块稀疏；压缩感知；块稀疏偏差模型

中图分类号：TN911.73 **文献标志码**：A **文章编号**：1000-5471(2023)01-0081-09

Block Sparse Deviation Modeling Through Generative Models

SINA Yuzhui， YU Pengbin， WANG Jianjun

School of Mathematic and Statistics，Southwest University，Chongqing 400715，China

Abstract：In this paper，we concentrate on studying the compressed sensing block sparse deviation model based on the generative model(Block Sparse-Gen). Relying on the Block RIP condition and the Block REC condition，we theoretically gave the reconstruction error for the optimal decoding and the number of measurements for the high probability recovery generating function. Moreover，the experimental values verifies the effectiveness of Block Sparse-Gen.

Key words：generative model；block sparse；compressed sensing；block sparse-gen

传统的信号采样方式是基于 Nyquist 采样框架实现的，根据香农采样定理，若要从采样的离散信号中无失真地恢复模拟信号，则要求采样速率必须达到信号带宽的两倍以上[1-2]. 然而，在现实世界中，信号采集的成本很高. 为了解决这一问题，文献[3]提出了压缩感知理论，其核心思想是将压缩与采样相结合，突破了香农采样定理的瓶颈，使得高分辨率信号的采集成为可能. 从数学角度出发，压缩感知的核心思想为一个线性测量过程. 选取测量矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ $(m \ll n)$，可以得到信号 $\boldsymbol{x}$ 的测量信号 $\boldsymbol{y}$. 数学表达式如下：

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon \tag{1}$$

在实际应用中，存在一类信号，其非零元素呈现出特殊的结构，如块状结构[4]. 从数学的角度来讲，给定分块

① 收稿日期：2022-02-07
基金项目：国家自然科学基金项目(12071380).
作者简介：斯那雨追，硕士研究生，主要从事机器学习与数据挖掘的研究.
通信作者：王建军，博士，教授.

$\tau=\{d_1, d_2, \cdots, d_N\}$ 后，任意向量 $\boldsymbol{x} \in \mathbb{R}^n$ 都可以被描述为 $\boldsymbol{x}=(\underbrace{x_1 \cdots x_{d_1}}_{\boldsymbol{x}[1]}, \underbrace{x_{d_1+1} \cdots x_{d_1+d_2}}_{\boldsymbol{x}[2]}, \cdots, \underbrace{x_{n-d_N+1} \cdots x_n}_{\boldsymbol{x}[N]})^{\mathrm{T}}$，$n=\sum_{i=1}^N d_i$[5]. 如果向量 $\boldsymbol{x}$ 在分块 τ 下至多有 s 个非零块，那么我们称该向量为块 s-稀疏信号，记为 $\|\boldsymbol{x}\|_{2.0} \leqslant s$，其中 $\|\boldsymbol{x}\|_{2.0}=\sum_{i=1}^N I(\|\boldsymbol{x}[i]\|_2)$[6-7](此处 I 表示符号函数). 使用标准的凸规划来研究块稀疏信号忽视了信号的结构特性，不能很好恢复稀疏信号. 大量研究表明，在处理这样的块稀疏信号时，混合的 $\frac{\ell_2}{\ell_1}$ 最小化问题优于 ℓ_1 最小化问题.

为了使得欠定方程有唯一解，即便是在无噪声的情况下，通常需要对未知向量 $\boldsymbol{x}$ 进行一些结构性假设，最常见的结构性假设是 $\boldsymbol{x}$ 是稀疏的. 如果信号在某一个正交空间具有稀疏性，就能以较低的频率采样该信号，并可能以高概率精确重建该信号. 经典的变换方法包括离散余弦变换(DCT)[8]、傅里叶变换(FFT)[9]、离散小波变换(DWT)[10] 等. 在上述环境中，如果矩阵 $\boldsymbol{A}$ 满足某些条件，如限制等距性质(RIP)[11] 或限制特征值条件(REC)[12]，则可以保证 $\boldsymbol{x}$ 从 $\boldsymbol{y}$ 中有效恢复.

尽管 $\boldsymbol{x}$ 上的稀疏性假设是最常见的选择，但也可以从数据的特征中提取结构性假设. 基于生成模型的方法已经被广泛应用于压缩感知中，并表现出优异的结果[13]. 上述方法的局限性在于，将要恢复的信号限制在生成器函数 G 的范围内. 因此，如果检测到的真实信号不在 G 的范围内，则即使 $m \gg n$，算法也无法将重构误差驱动为 0. 为了克服这一缺陷，一种称为稀疏生成(Sparse-Gen)的框架被提出[14]. 具体来说，该框架允许恢复的信号在生成器 G 的范围内添加稀疏偏差，恢复的信号一般具有 $G(\hat{\boldsymbol{z}})+\hat{\boldsymbol{\omega}}$ 的形式，其中 $\hat{\boldsymbol{\omega}} \in \mathbb{R}^n$ 是一个稀疏向量. 则有以下的优化问题：

$$
\begin{aligned}
&\min_{\boldsymbol{z}, \boldsymbol{\omega}} \|\boldsymbol{\omega}\|_1 \\
&\text{s.t.} \quad \boldsymbol{A}(G(\boldsymbol{z})+\boldsymbol{\omega})=\boldsymbol{y}
\end{aligned} \tag{2}
$$

其中 $\|\boldsymbol{\omega}\|_1=\sum_{i=1}^n |\boldsymbol{\omega}_i|$. 鉴于 ℓ_1 是 ℓ_0 范数的松散近似，基于 ℓ_1 最小化问题得到的解一般是次最优的，也有相关文章采用 ℓ_q 范数对稀疏偏差进行约束，提出了基于生成模型的非凸稀疏偏差模型[15].

在本文中，我们研究了基于生成模型的压缩感知块稀疏偏差模型. 在该方法中，我们使用 $\ell_{2.1}$ 范数来约束偏差向量，其中 $G: \mathbb{R}^k \longrightarrow \mathbb{R}^n$ 是生成函数，$\boldsymbol{\omega} \in \mathbb{R}^n$，$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ 是测量矩阵，$\boldsymbol{y} \in \mathbb{R}^m$ 是观测向量. 本文关心的模型是

$$
\begin{aligned}
&\min_{\boldsymbol{z}, \boldsymbol{\omega}} \|\boldsymbol{\omega}\|_{2.1} \\
&\text{s.t.}\ \boldsymbol{A}(G(\boldsymbol{z})+\boldsymbol{\omega})=\boldsymbol{y}
\end{aligned} \tag{3}
$$

其中 $\|\boldsymbol{\omega}\|_{2.1}=\sum_{i=1}^N \|\boldsymbol{\omega}[i]\|_2$，$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{\varepsilon}$. 基于上述问题的拉格朗日方程，考虑如下针对块稀疏生成的最终无约束优化问题：

$$
\min_{\boldsymbol{z}, \boldsymbol{\omega}} \|\boldsymbol{\omega}\|_{2.1}+\lambda\|\boldsymbol{A}(G(\boldsymbol{z})+\boldsymbol{\omega})-\boldsymbol{y}\|_2^2 \tag{4}
$$

其中 λ 是拉格朗日常数. 我们给出了理论结果和仿真，在理论上，本文首先提出了针对块稀疏信号的块约束等距性质(B-RIP)和块有限等距性质(B-S-REC)，如果测量矩阵具有这两个性质，则最优解码的重构误差存在上界. 最后推导出在生成函数条件下以高概率成功恢复所需的测量次数. 在实验方面，为了进一步验证本文提出的 Block Sparse-Gen 的有效性和优越性，使用两个数据集(MNIST 和 CelebA)和两个生成模型(VAE 和 DCGAN)进行了一系列实验. 在测量次数相对较少的情况下，该方法的重建误差远小于基于 LASSO 的恢复、基于生成模型的恢复和稀疏生成.

1 相关定义

定义 1 $\mathrm{BS}_l(0)=\{\boldsymbol{x}: \|\boldsymbol{x}-\boldsymbol{0}\|_{2.0} \leqslant l\}$，其中 $\|\boldsymbol{x}\|_{2.0}=\sum_{i=1}^N I(\|\boldsymbol{x}[i]\|_2)$.

定义 2　(Block-RIP) 对于参数 $\alpha \in (0, 1)$ 和一个给定的分块 $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，如果对于所有的 $\boldsymbol{x} \in \mathrm{BS}_\ell(0)$ 满足下面的不等式，则我们说矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 满足 B-RIP(ℓ, α)，

$$(1-\alpha)\|\boldsymbol{x}\|_2^2 \leqslant \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leqslant (1+\alpha)\|\boldsymbol{x}\|_2^2 \tag{5}$$

定义 3　(Block-REC) 对于参数 $\gamma > 0$ 和一个给定的分块 $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，如果对于所有的 $\boldsymbol{x} \in \mathrm{BS}_\ell(0)$ 满足下面的不等式，则我们说矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 满足 B-REC(ℓ, γ)，

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 \geqslant \gamma\|\boldsymbol{x}\|_2 \tag{6}$$

定义 4　(Block-S-REC) 令 $S \subseteq \mathbb{R}^n$，对于参数 $\gamma > 0$，$\delta \geqslant 0$ 和一个给定的分块 $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，如果对于所有的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ 满足下面的不等式，则我们说矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 满足 B-S-REC(S, γ, δ)，

$$\|\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2 \geqslant \gamma\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 - \delta \tag{7}$$

2　主要结论及证明

引理 1　对于一个给定的分块 $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，假设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 满足 B-S-REC$(S_{\frac{1.5l}{d}, G}, (1-\delta), \tau)$ 和 B-RIP$\left(\dfrac{2l}{d}, \delta\right)$，其中参数 $\delta \in (0, 1)$，$l > 0$. 给定一个函数 $G: \mathbb{R}^k \longrightarrow \mathbb{R}^n$ 和测量噪声 $\boldsymbol{\varepsilon}$，$\|\boldsymbol{\varepsilon}\|_2 \leqslant \varepsilon_{\max}$. 则存在一个解码器 $\Delta: \mathbb{R}^m \longrightarrow \mathbb{R}^n$ 满足

$$\|\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})\|_2 \leqslant \left(\frac{2l}{d}\right)^{-\frac{1}{2}} C_0 \sigma_{\frac{l}{d}, G}(x) + C_1\varepsilon_{\max} + \tau' \tag{8}$$

对于所有的 $\boldsymbol{x} \in \mathbb{R}^n$ 都成立. 其中 d 是指每块所含元素的个数，这是个输入变量. $N = \lceil \frac{n}{d} \rceil$表示分块数. $C_0 = 2(\sqrt{(1+\delta)}(1-\delta)^{-1} + 1)$，$C_1 = 2(1-\delta)^{-1}$，$\tau' = \tau(1-\delta)^{-1}$.

为了证明引理 1，首先定义如下记号. 考虑到要在我们的分析中考虑测量噪声，我们将矩阵 $\boldsymbol{A}$ 的 ε 型管集定义为

$$T_{\boldsymbol{A}}(\varepsilon) = \{\boldsymbol{\omega}: \|\boldsymbol{A}\boldsymbol{\omega}\|_2 \leqslant \varepsilon\} \tag{9}$$

定义一个差分函数 $G': \mathbb{R}^k \times \mathbb{R}^k \longrightarrow \mathbb{R}^n$ 如 $G'(\boldsymbol{z}_1, \boldsymbol{z}_2) = G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2)$，我们可以得到一个差分集 $\mathrm{BS}_{\ell, G'}$

$$\begin{aligned}\mathrm{BS}_{\ell, G'} = {} & \mathrm{BS}_\ell(G') = \\ & \textstyle\bigcup_{z_1, z_2} \mathrm{BS}_\ell(G'(\boldsymbol{z}_1, \boldsymbol{z}_2)) = \\ & \{\boldsymbol{x}: \|\boldsymbol{x} - G'(\boldsymbol{z}_1, \boldsymbol{z}_2)\|_{2,0} \leqslant l, \boldsymbol{z}_1, \boldsymbol{z}_1 \in \mathrm{Dom}(G)\} = \\ & \{\boldsymbol{x}: \|\boldsymbol{x} - G(\boldsymbol{z}_1) + G(\boldsymbol{z}_1)\|_{2,0} \leqslant l, \boldsymbol{z}_1, \boldsymbol{z}_1 \in \mathrm{Dom}(G)\}\end{aligned} \tag{10}$$

记

$$\sigma_{\ell, G'}(\boldsymbol{x}) = \inf_{\hat{\boldsymbol{x}} \in \mathrm{BS}_{\ell, G'}} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_{2,1} \tag{11}$$

为了证明引理 1，我们先陈述并证明下面的引理 2 和引理 3.

引理 2　对于一个给定的分块 $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，测量矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，测量噪声 $\boldsymbol{\varepsilon}$ 满足 $\|\boldsymbol{\varepsilon}\|_2 \leqslant \varepsilon_{\max}$ 和一个生成器函数 $G: \mathbb{R}^k \longrightarrow \mathbb{R}^n$. 在块稀疏向量集合 $\mathrm{BS}_{\frac{l}{d}}$ 中，我们定义一个解码器 $\Delta: \mathbb{R}^m \longrightarrow \mathbb{R}^n$ 满足接下来的$(\ell_2, \ell_{2,1})$ 混合范数保证

$$\|\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})\|_2 \leqslant C_0\left(\frac{l}{d}\right)^{-t} \sigma_{\frac{l}{d}, G}(\boldsymbol{x}) + C_1\varepsilon_{\max} + \tau \tag{12}$$

对于某些给定的常数 $C_0, \tau, t \geqslant 0$. 这样的解码器存在的充分条件由下式给出

$$\|\boldsymbol{\eta}\|_2 \leqslant \frac{C_0}{2}\left(\frac{l}{d}\right)^{-t} \sigma_{\frac{2l}{d}, G'}(\boldsymbol{\eta}) + C_1\varepsilon_{\max} + \tau, \ \forall\, \boldsymbol{\eta} \in U_{\boldsymbol{A}}(2\varepsilon_{\max}) \tag{13}$$

我们将其称为$(\ell_2, \ell_{2,1})$混合范数空间属性.

证 为了证明零空间条件的充分性，我们定义一个如下的解码器$\Delta: \mathbb{R}^m \longrightarrow \mathbb{R}^n$,

$$\Delta(\boldsymbol{y}) = \underset{\hat{\boldsymbol{x}}: \|\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{y}\|_2 \leqslant \varepsilon_{\max}}{\arg\min} \sigma_{\frac{l}{d}, G}(\boldsymbol{x}) \tag{14}$$

给定$(\ell_2, \ell_{2,1})$混合范数空间属性，我们将证明该解码器满足混合范数保证. 根据Δ的定义有

$$\|\boldsymbol{A}(\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}))\|_2 \leqslant \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon} - \boldsymbol{A}\Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})\|_2 + \varepsilon_{\max} \leqslant 2\varepsilon_{\max} \tag{15}$$

这意味着

$$\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}) \in T_{\boldsymbol{A}}(2\varepsilon_{\max})$$

结合混合范数空间属性，我们有

$$\begin{aligned}
&\|\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})\|_2 - \tau - C_1\varepsilon_{\max} \leqslant \\
&\frac{C_0}{2}\left(\frac{l}{d}\right)^{-t} \sigma_{\frac{2l}{d}, G'}(\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})) \leqslant \\
&\frac{C_0}{2}\left(\frac{l}{d}\right)^{-t} (\sigma_{\frac{l}{d}, G}(\boldsymbol{x}) + \sigma_{\frac{l}{d}, G}(\Delta(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}))) \leqslant \\
&C_0\left(\frac{l}{d}\right)^{-t} \sigma_{\frac{l}{d}, G}(\boldsymbol{x})
\end{aligned} \tag{16}$$

倒数第二步可由三角不等式$\sigma_{\frac{2l}{d}, G'}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \leqslant \sigma_{\frac{l}{d}, G}(\boldsymbol{x}_1) + \sigma_{\frac{l}{d}, G}(\boldsymbol{x}_2)$得到，最后一步由解码器是使得$\sigma_{\frac{l}{d}, G}(\boldsymbol{x})$达到最小的工具得到.

引理3 对于整数$a, b, l > 0$和函数$G: \mathbb{R}^k \longrightarrow \mathbb{R}^m$，如果矩阵$\boldsymbol{A} \in \mathbb{R}^{m \times n}$满足B-S-REC$(\mathrm{BS}_{\frac{(a+b)l}{2d}, G}, 1-\delta, \tau)$和B-RIP$\left(\frac{bl}{d}, \delta\right)$，则对任意向量$\boldsymbol{\eta} \in T_{\boldsymbol{A}}(\varepsilon)$有，

$$\|\boldsymbol{\eta}\|_2 \leqslant \left(\frac{bl}{d}\right)^{-\frac{1}{2}} (C_0 + 1)\sigma_{\frac{al}{d}, G'}(\boldsymbol{\eta}) + C_1\varepsilon + \tau' \tag{17}$$

其中$C_0 = (1-\delta)^{-1}\sqrt{(1+\delta)}$, $C_1 = (1-\delta)^{-1}$, $\tau' = \tau(1-\delta)^{-1}$.

证 对于任意的$\boldsymbol{\eta} \in T_{\boldsymbol{A}}(\boldsymbol{\varepsilon})$和$G(\boldsymbol{z}_1), G(\boldsymbol{z}_2)$，令$\boldsymbol{v} \in \mathrm{BS}_{\frac{al}{d}}$是使得$\|\boldsymbol{\eta} - G(\boldsymbol{z}_1) + G(\boldsymbol{z}_2) - \boldsymbol{v}\|_{2,1}$最小化的向量. 根据$\mathrm{BS}_{\frac{al}{d}}$是一个闭集合，我们可以找到满足条件的$\boldsymbol{v}$. 取$\boldsymbol{\eta} - G(\boldsymbol{z}_1) + G(\boldsymbol{z}_2)$这个向量的前最大$\dfrac{al}{d}$块(即对块内元素求和后最大的块)，其余为零. 给定一个n维矢量的索引集I，我们用I_c表示不在I中的索引集. 注意到$\boldsymbol{v}$对应于$\boldsymbol{\eta}' = \boldsymbol{\eta} - G(\boldsymbol{z}_1) + G(\boldsymbol{z}_2)$的最大$\dfrac{al}{d}$块，令其坐标对应的索引为$\mathcal{T}_0$. 将$\mathcal{T}_1$作为下一个最大$\dfrac{bl}{d}$块对应的索引. 类似的，将$\mathcal{T}_2, \cdots, \mathcal{T}_s$作为下一个最大$\dfrac{bl}{d}$块元素对应的索引. 设$\mathcal{T}_0 \cup \mathcal{T}_1 = \mathcal{T}$，用$\boldsymbol{x}_1$表示令集合$I_c$所有索引中$\boldsymbol{x}$的值归零而获得的向量. 将$\boldsymbol{\eta}_{\mathcal{T}} + (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}_c}$写为$\boldsymbol{\eta}_{\mathcal{T}} - (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}} + (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))$，其中$\boldsymbol{\eta}_{\mathcal{T}}, (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}} \in \mathrm{BS}_{\frac{(a+b)l}{d}}$. 将$\boldsymbol{\eta}_{\mathcal{T}} - (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}}$写为$\boldsymbol{s}_1 - \boldsymbol{s}_2$，其中$\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathrm{BS}_{\frac{(a+b)l}{2d}}$. 综上所述，我们可以将$\boldsymbol{\eta}_{\mathcal{T}} + (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}_c}$写为$G(\boldsymbol{z}_1) + \boldsymbol{s}_1 - (G(\boldsymbol{z}_2) + \boldsymbol{s}_2)$，其中$G(\boldsymbol{z}_1) + \boldsymbol{s}_1, G(\boldsymbol{z}_2) + \boldsymbol{s}_2 \in \mathrm{BS}_{\frac{(a+b)l}{d\,2}, G'}$.

接下来根据$\boldsymbol{A}$满足B-S-REC$(\mathrm{BS}_{\frac{(a+b)l}{2d}, G}, 1-\delta, \tau)$的事实，有

$$\begin{aligned}
&\|\boldsymbol{\eta}_{\mathcal{T}} + (G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2))_{\mathcal{T}_c}\|_2 = \\
&\|G(\boldsymbol{z}_1) + \boldsymbol{s}_1 - (G(\boldsymbol{z}_2) + \boldsymbol{s}_2)\|_2 \leqslant \\
&(1-\delta)^{-1}\|\boldsymbol{A}(G(\boldsymbol{z}_1) + \boldsymbol{s}_1 - (G(\boldsymbol{z}_2) + \boldsymbol{s}_2))\|_2 + (1-\delta)^{-1}\tau =
\end{aligned}$$

$$(1-\delta)^{-1}\|\boldsymbol{A}(\boldsymbol{\eta}_{\mathcal{T}}+(G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2))_{\mathcal{T}_c})\|_2+(1-\delta)^{-1}\tau \tag{18}$$

可以将 $\boldsymbol{\eta}$ 写为 $\boldsymbol{\eta}=\boldsymbol{\eta}_{\mathcal{T}}+\boldsymbol{\eta}_{\mathcal{T}_2}+\cdots+\boldsymbol{\eta}_{\mathcal{T}_s}$. 再根据 $\boldsymbol{\eta}\in T_{\boldsymbol{A}}(\varepsilon)$，将 $\boldsymbol{A}\boldsymbol{\eta}_{\mathcal{T}}$ 表达为 $\boldsymbol{A}\boldsymbol{\eta}_{\mathcal{T}}=-\boldsymbol{A}(\boldsymbol{\eta}_{\mathcal{T}_2}+\cdots+\boldsymbol{\eta}_{\mathcal{T}_s})+\gamma$，其中 $\|\gamma\|_2\leqslant\varepsilon$. 因此，

$$\begin{aligned}&\|\boldsymbol{A}(\boldsymbol{\eta}_{\mathcal{T}}+(G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2))_{\mathcal{T}_c})\|_2=\\&\|\boldsymbol{A}((\boldsymbol{\eta}-G(\boldsymbol{z}_1)+G(\boldsymbol{z}_2))_{\mathcal{T}_2}+\cdots+(\boldsymbol{\eta}-G(\boldsymbol{z}_1)+G(\boldsymbol{z}_2))_{\mathcal{T}_s})-\gamma\|_2=\\&\|\boldsymbol{A}\boldsymbol{\eta}'_{\mathcal{T}_2}+\cdots+\boldsymbol{A}\boldsymbol{\eta}'_{\mathcal{T}_s}-\gamma\|_2\leqslant\\&\sum_{j=2}^{s}\|\boldsymbol{A}\boldsymbol{\eta}'_{\mathcal{T}_j}\|_2+\|\gamma\|_2\leqslant\\&\sqrt{(1+\delta)}\sum_{j=2}^{s}\|\boldsymbol{\eta}'_{\mathcal{T}_j}\|_2+\varepsilon\end{aligned} \tag{19}$$

根据(18)，(19) 两个不等式可以得到

$$\|\boldsymbol{\eta}_{\mathcal{T}}+(G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2))_{\mathcal{T}_c}\|_2-\tau'\leqslant(1-\delta)^{-1}\sqrt{(1+\delta)}\sum_{j=2}^{s}\|\boldsymbol{\eta}'_{Tj}\|_2+(1-\delta)^{-1}\varepsilon \tag{20}$$

在两边加上 $\|\boldsymbol{\eta}'_{\mathcal{T}_c}\|_2$ 并根据三角不等式可以得到

$$\begin{aligned}&\|\boldsymbol{\eta}\|_2\leqslant\|\boldsymbol{\eta}_{\mathcal{T}}+(G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2))_{\mathcal{T}_c}\|_2+\|\boldsymbol{\eta}'_{\mathcal{T}_c}\|_2\leqslant\\&((1-\delta)^{-1}\sqrt{(1+\delta)}+1)\sum_{j=2}^{s}\|\boldsymbol{\eta}'_{\mathcal{T}_j}\|_2+\tau'+C_1\varepsilon\end{aligned} \tag{21}$$

对于任意的 $i\geqslant 1$，$j_1\in\mathcal{T}_{i+1}$ 和 $j_2\in\mathcal{T}_i$，我们有 $\|\boldsymbol{\eta}'_{j1}\|_2\leqslant\|\boldsymbol{\eta}'_{j2}\|_2$，意味着 $\|\boldsymbol{\eta}'_{j1}\|_2\leqslant\left(\frac{bl}{d}\right)^{-1}\|\boldsymbol{\eta}'_{\mathcal{T}_i}\|_{2.1}$. 对于 $\mathcal{T}_i$ 和 $\mathcal{T}_{i+1}$ 内的所有索引平方并且加上不等式可以得到

$$\|\boldsymbol{\eta}'_{\mathcal{T}_{i+1}}\|_2\leqslant\left(\frac{bl}{d}\right)^{-\frac{1}{2}}\|\boldsymbol{\eta}'_{\mathcal{T}_i}\|_{2.1} \tag{22}$$

结合(21) 式可得

$$\begin{aligned}&\|\boldsymbol{\eta}\|_2-\tau'-C_1\varepsilon\leqslant\left(\frac{bl}{d}\right)^{-\frac{1}{2}}((1-\delta)^{-1}\sqrt{(1+\delta)}+1)\sum_{j=1}^{s}\|\boldsymbol{\eta}'_{\mathcal{T}_j}\|_{2.1}=\\&\left(\frac{bl}{d}\right)^{-\frac{1}{2}}(C_0+1)\|\boldsymbol{\eta}'_{\mathcal{T}_0}\|_{2.1}\end{aligned} \tag{23}$$

引理 1 的证明　结合引理 2 和引理 3，令 $a=2$，$b=1$，我们可以直接推导出引理 1.

引理 4　设 $G:B^k(r)\longrightarrow\mathbb{R}^n$ 为一个 L-Lipschitz 函数，其中 $B^k(r)=\{\boldsymbol{z}\mid\boldsymbol{z}\in\mathbb{R}^k,\ \|\boldsymbol{z}\|_2\leqslant r\}$ 是 $\mathbb{R}^k$ 上的 ℓ_2 范数球. 对于 $\delta\in(0,1)$，如果

$$m=O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\left(k\ \log\left(\frac{Lr}{\tau}\right)+\frac{l}{d}\log\left(\frac{en}{l}\right)\right)\right) \tag{24}$$

则对于一个给定的分块 $\boldsymbol{\tau}=\{\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\cdots,\boldsymbol{\tau}_N\}$，伴随有独立同分布项 $\boldsymbol{A}_{ij}\sim\mathcal{N}\left(0,\frac{1}{m}\right)$ 的随机高斯矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 以概率 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 满足 B-S-REC$(S_{\frac{1.5\,l}{d},G},1-\delta,\tau)$ 和 B-RIP$\left(\frac{2\,l}{d},\delta\right)$.

为了证明引理 4，我们先阐述下面的引理 5 和引理 6.

引理 5[14,16]　设 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 是一个带有独立同分布项 $\mathcal{N}\left(0,\frac{1}{m}\right)$ 的随机高斯矩阵，$\delta\in(0,1)$，如果

$$m=O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\right) \tag{25}$$

则 $\boldsymbol{A}$ 至少以 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 的概率满足 B-RIP(l,δ).

引理 6 设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是带有独立同分布项 $\mathcal{N}\left(0, \frac{1}{m}\right)$ 的随机高斯矩阵，G：$\mathbb{R}^k \longrightarrow \mathbb{R}^n$ 为一个 L-Lipschitz 函数，$B^k(r)=\{\boldsymbol{z}: \|\boldsymbol{z}\|_2 \leqslant r\}$ 为 ℓ_2 范数球. 如果

$$m = O\left(\frac{k}{\delta^2}\log\left(\frac{Lr}{\tau}\right)\right) \tag{26}$$

则 $\boldsymbol{A}$ 至少以 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 的概率满足 B-S-REC$(G(B^k(r)), 1-\delta, \tau)$.

引理 4 证明 我们在 $B^k(r)$ 构建一个 $\frac{\tau}{L}$ 网 M，则存在一个网络满足如下不等式，

$$\log(|M|) \leqslant k\ \log\left(\frac{Lr}{\tau}\right) \tag{27}$$

因为这个网是 $B^k(r)$ 的一个 $\frac{\tau}{L}$ 覆盖，那么 $G(M)$ 是 $G(B^k(r))$ 的一个 τ 覆盖. 任取两点 $\boldsymbol{z}_1, \boldsymbol{z}_2 \in B^k(r)$，我们可以找到对应的两个点 $\boldsymbol{z}'_1, \boldsymbol{z}'_2 \in M$ 满足

$$\|G(\boldsymbol{z}_1) - G(\boldsymbol{z}'_1)\|_2 \leqslant \tau$$
$$\|G(\boldsymbol{z}_2) - G(\boldsymbol{z}'_2)\|_2 \leqslant \tau$$

现在考虑索引集为 I 的块 $\frac{l}{d}$ 稀疏向量，索引集以外的元素都为零. 由三角不等式，我们可以得到

$$\begin{aligned}\|G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2)+\boldsymbol{v}\|_2 \leqslant\ & \|G(\boldsymbol{z}_1)-G(\boldsymbol{z}'_1)\|_2 + \|G(\boldsymbol{z}'_1)-G(\boldsymbol{z}'_2)+\boldsymbol{v}\|_2 + \|G(\boldsymbol{z}'_2)-G(\boldsymbol{z}_2)\|_2 \leqslant \\ & \|G(\boldsymbol{z}'_1)-G(\boldsymbol{z}'_2)+\boldsymbol{v}\|_2 + 2\tau\end{aligned} \tag{28}$$

再次由三角不等式，可以得到

$$\begin{aligned}&\|\boldsymbol{A}G(\boldsymbol{z}'_1) - \boldsymbol{A}G(\boldsymbol{z}'_2) + \boldsymbol{A}\boldsymbol{v}\|_2 \leqslant \\ &\|\boldsymbol{A}G(\boldsymbol{z}'_1) - \boldsymbol{A}G(\boldsymbol{z}_1)\|_2 + \|\boldsymbol{A}G(\boldsymbol{z}_1) - \boldsymbol{A}G(\boldsymbol{z}_2) + \boldsymbol{A}\boldsymbol{v}\|_2 + \|\boldsymbol{A}G(\boldsymbol{z}_2) - \boldsymbol{A}G(\boldsymbol{z}'_2)\|_2\end{aligned} \tag{29}$$

由文献[13] 引理 8.3，我们有以概率 $1-\mathrm{e}^{-\Omega(m)}$ 满足

$$\|\boldsymbol{A}G(\boldsymbol{z}'_1) - \boldsymbol{A}G(\boldsymbol{z}_1)\|_2 = O(\tau)$$
$$\|\boldsymbol{A}G(\boldsymbol{z}_2) - \boldsymbol{A}G(\boldsymbol{z}'_2)\|_2 = O(\tau)$$

将这一事实应用于(29) 式，则可以得到

$$\|\boldsymbol{A}G(\boldsymbol{z}'_1) - \boldsymbol{A}G(\boldsymbol{z}'_2) + \boldsymbol{A}\boldsymbol{v}\|_2 \leqslant \|\boldsymbol{A}G(\boldsymbol{z}_1) - \boldsymbol{A}G(\boldsymbol{z}_2) + \boldsymbol{A}\boldsymbol{v}\|_2 + O(\tau) \tag{30}$$

对于固定的 $\boldsymbol{z}'_1$ 与 $\boldsymbol{z}'_2$，$\boldsymbol{v}$ 随着索引集的变化而变化. 由引理 6，当 $m = O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\right)$ 时，至少以 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 的概率满足如下不等式，

$$(1-\delta)\|G(\boldsymbol{z}'_1) - G(\boldsymbol{z}'_2) + \boldsymbol{v}\|_2 \leqslant \|\boldsymbol{A}G(\boldsymbol{z}'_1) - \boldsymbol{A}G(\boldsymbol{z}'_2) + \boldsymbol{A}\boldsymbol{v}\|_2 \tag{31}$$

最后我们对 $\boldsymbol{z}'_1, \boldsymbol{z}'_2$ 和索引集 I 的选择进行联合约束(从 $\frac{n}{d}$ 个指标中选择 $\frac{l}{d}$ 个). 设选择的总个数为 N. 并且根据不等式 $\begin{pmatrix}\frac{n}{d}\\ \frac{l}{d}\end{pmatrix} \leqslant \left(\frac{ne}{d}\right)^{\frac{l}{d}}$，我们可以得到：

$$\begin{aligned}\log(N) \leqslant 2\ \log(|M|) + \frac{l}{d}\log\left(\frac{en}{l}\right) \leqslant \\ 2k\ \log\left(\frac{Lr}{\tau}\right) + \frac{l}{d}\log\left(\frac{en}{l}\right)\end{aligned} \tag{32}$$

则可以得到对于所有的 $\boldsymbol{z}_1, \boldsymbol{z}_2 \in B^k(r)$ 和 $\boldsymbol{v} \in \mathrm{BS}_{\frac{l}{d}}$，当

$$m = O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\left(k\ \log\left(\frac{Lr}{\tau}\right) + \frac{l}{d}\log\left(\frac{en}{l}\right)\right)\right) \tag{33}$$

以下不等式

$$(1-\delta)\|G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2)+\boldsymbol{v}\|_2 \leqslant \|\boldsymbol{A}(G(\boldsymbol{z}_1)-G(\boldsymbol{z}_2)+\boldsymbol{v})\|_2+O(\tau) \tag{34}$$

至少以 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 的概率成立. 引理 4 得证.

结合引理 1 和引理 4，我们可以得到如下所述定理 1 的结果.

定理 1　对于给定的分块 $\boldsymbol{\tau}=\{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_N\}$，假设 $G: B^k(r) \longrightarrow \mathbb{R}^n$ 是一个 L-Lipschitz 函数. 对于任意的 $\delta \in (0, 1)$，$l>0$，设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是一个带有独立同分布项 $\boldsymbol{A}_{i,j} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$ 的随机高斯矩阵，其中

$$m=O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\left(k\ \log\left(\frac{Lr}{\tau}\right)+\frac{l}{d}\log\left(\frac{en}{l}\right)\right)\right) \tag{35}$$

假设 Δ 是满足引理 1 的解码器，则我们至少以 $1-\mathrm{e}^{-\Omega(\delta^2 m)}$ 的概率有

$$\|\boldsymbol{x}-\Delta(\boldsymbol{Ax}+\boldsymbol{\varepsilon})\|_2 \leqslant \left(\frac{2l}{d}\right)^{-\frac{1}{2}} C_0\sigma_{\frac{l}{d},G}(x)+C_1\varepsilon_{\max}+\tau' \tag{36}$$

对于任意的 $\boldsymbol{x} \in \mathbb{R}^n$，$\|\boldsymbol{\varepsilon}\|_2 \leqslant \varepsilon_{\max}$，其中 C_0, C_1, γ, τ' 的定义参见引理 1.

3　实验

为了验证算法的有效性，本节将进行对比实验. 实验在 Intel(R) Xeon(R) Gold-5122 CPU(3.6GHz)，64GB 内存，Linux 操作系统和 python(3.7.4) 构成的实验环境中运行. 在接下来的实验中，将使用两个不同的数据集：手写数字数据集 MNIST 和名人脸属性数据集 CelebFaces Attributes Dataset(CelebA). MNIST 数据集为手写数字，包含 60 000 个训练样本和 10 000 个测试样本. CelebA 是一个大规模的面部属性数据集，其中包含 10 177 位名人身份的 202 599 张人脸图片，并且每个图像都标记有特征. 我们生成了一个高斯随机矩阵 $\boldsymbol{A}$，其每一项 $\boldsymbol{A}_{ij} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$. 在潜在变量生成模型(VAE 与 DCGAN) 中，潜变量 $\boldsymbol{Z}$ 为各项同性的高斯向量. 为了评价实验效果，我们用 $\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2$ 表示不同测量值 m 的重构误差.

3.1　实验设置

MNIST 数据集中每个图像的大小为 28×28 像素，并且每个像素值为 0 或者 1. 对于这个数据集，我们根据 Block Sparse-Gen 来训练 VAE，以恢复原始图像. 由于图像包含单个通道，因此输入尺寸为 $28\times 28=784$，学习率为 0.1，$\lambda=0.1$. 在 CelebA 数据集中，将人脸图像裁剪为 64×64 像素大小，使每个图像输入的尺寸为 $64\times 64\times 3=12\ 288$，并将每个像素值缩放为 $[-1, 1]$. 对于这个数据集，考虑根据 Block Sparse-Gen 模型训练一个 DCGAN 来恢复原始图像，同时会将结果与其他模型和算法进行比较. 对于 Block Sparse-VAE，使用 LASSO 作为基准，并将其与基于生成模型的算法(VAE) 和添加了稀疏偏差的生成模型(Sparse-VAE) 进行比较. 对于 Block Sparse-DCGAN，我们将结果与 LASSO 的结果进行了比较，LASSO 的结果包含两个变换域：二维离散余弦变换和小波变换. 类似地，还将结果与基于生成模型的算法(DCGAN) 和添加了稀疏偏差的生成模型(Sparse-DCGAN) 进行比较.

3.2　重建结果

为了探索每种算法的重建效果，对于 MNIST 数据集，我们在图 1 展示重建的均方误差结果. 可以看出，与理论结果类似，随着采样数的增加，均方误差明显减少，并且 Block Sparse-VAE 模型相比其他的模型能够更可靠地重构未知样本. 类似地，我们给出了 CelebA 数据集的恢复结果，如图 2 所示. 与 MNIST 数据集类似，本文算法明显优于 LASSO 等模型. 尤其是当测量次数较少时，具有独到的优势. 图 3 和图 4 展示了 MNIST 数据集在测量次数为 80 时的恢复效果以及 CelebA 数据集在测量次数为 1 000 时的恢复效果. 我们发现除 LASSO 外，其他方法恢复效果明显较好. 这足以说明一个与理论一致的结果，即在测量次数较少的情况下，基于生成模型的恢复方法的强先验完全优于基于 LASSO 的稀疏向量恢复方法. 另外可以发现我

们提出模型的恢复效果优于其他的模型，显示了所提分块方法的有效性和优越性.

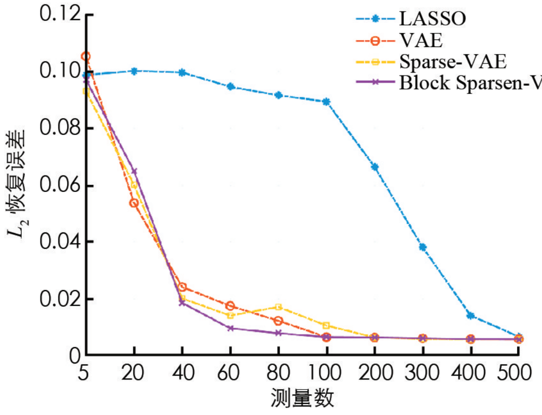

图 1　MNIST

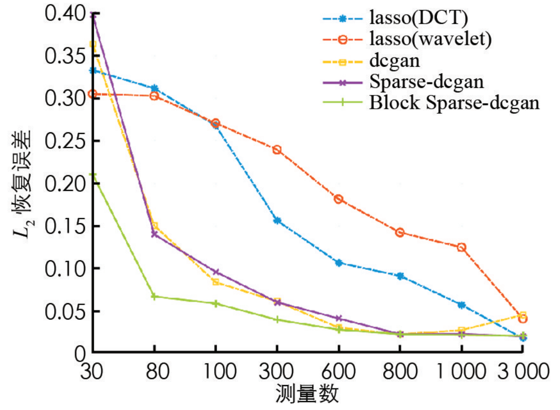

图 2　CelebA

图 3　MNIST 数据集恢复效果($m=80$)

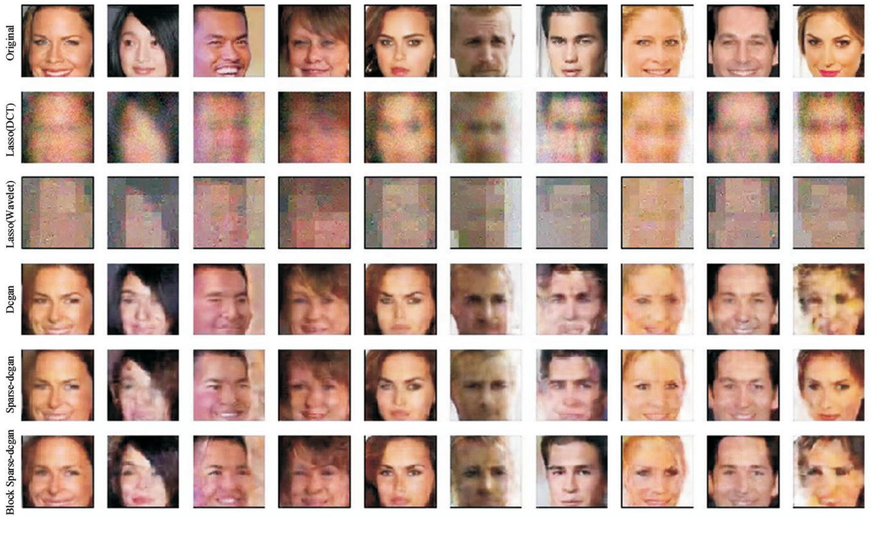

图 4　CelebA 数据集恢复效果($m=1\ 000$)

4　结论

本文对基于生成模型的稀疏偏差建模进行了推广，提出了 Block-Sparse Gen 模型. 针对此模型，我们提出了 Block-S-REC 条件，结合 Block-RIP 条件推导了在生成函数的稀疏偏差范围内最优解码器的误差上界，

并给出了原始信号高概率有效恢复的测量值次数

$$m=O\left(\frac{l}{d\delta^2}\log\left(\frac{n}{l}\right)\left(k\ \log\left(\frac{Lr}{\tau}\right)+\frac{l}{d}\left(\frac{en}{l}\right)\right)\right)$$

此结果在 $d=1$ 时退化为稀疏生成(Sparse-Gen)的情况. 在数值实验中，我们提出的模型减少了成功恢复的测量值条件，提高了恢复效果.

参考文献：

[1] NYQUIST H. Certain Topics in Telegraph Transmission Theory [J]. Transactions of the American Institute of Electrical Engineers，1928，47(2)：617-644.

[2] SHANNON C E. Communication in the Presence of Noise [J]. Proceedings of the IRE，1949，37(1)：10-21.

[3] 李树涛，魏丹. 压缩传感综述 [J]. 自动化学报，2009，35(11)：1369-1377.

[4] HE S Y，WANG Y，WANG J J，et al. Block-Sparse Compressed Sensing with Partially Known Signal Support via Non-Convex Minimisation [J]. IET Signal Processing，2016，10(7)：717-723.

[5] 刘春燕，李川，齐静. 基于扰动 BOMP 算法的块稀疏信号重构 [J]. 西南师范大学学报(自然科学版)，2020，45(7)：144-149.

[6] 王文东. 块压缩感知的 $L_2/L_q(0<q\leqslant1)$ 极小化算法研究 [D]. 重庆：西南大学，2014.

[7] 王文东，王尧，王建军. 一类光滑加权块 l_1 算法的收敛性分析与数值仿真实验 [J]. 西南大学学报(自然科学版)，2014，36(5)：72-77.

[8] AHMED N，NATARAJAN T，RAO K R. Discrete Cosine Transform [J]. IEEE Transactions on Computers，1974，C-23(1)：90-93.

[9] ALMEIDA L B. The Fractional Fourier Transform and Time-Frequency Representations [J]. IEEE Transactions on Signal Processing，1994，42(11)：3084-3091.

[10] SHENSA M J. The Discrete Wavelet Transform：Wedding the a Trous and Mallat Algorithms [J]. IEEE Transactions on Signal Processing，1992，40(10)：2464-2482.

[11] CANDÈS E J，et al. The Restricted Isometry Property and Its Implications for Compressed Sensing [J]. Comptes Rendus Mathematique，2008，346(9-10)：589-592.

[12] RASKUTTI G，WAINWRIGHT M，YU B. Restricted Eigenvalue Properties for Correlated Gaussian Designs [J]. J Mach Learn Res，2010，11：2241-2259.

[13] BORA A，JALAL A，PRICE E，et al. Compressed Sensing Using Generative Models [C]//9th International Conference on Machine Learning. New York：ACM Press，2017：537-546.

[14] DHAR M，GROVER A，ERMON S. Modeling Sparse Deviations for Compressed Sensing Using Generative Models [EB/OL]. (2018-05-16)[2021-11-05]. https：//arxiv. org/abs/1807. 01442.

[15] YANG Y X，WANG H L，QIU H Q，et al. Non-Convex Sparse Deviation Modeling via Generative Models [C]//ICASSP 2021-2021 IEEE International Conference on Acoustics，Speech and Signal Processing. New Yrok：IEEE Press，2345-2349.

[16] 黄建文. l_1 极小化框架下高维数据重构理论与算法研究 [D]. 重庆：西南大学，2019.

责任编辑 张枸