高技术通讯 2021 年 第31 卷 第9 期:925-933

doi:10.3772/j.issn.1002-0470.2021.09.003

基于时空约束压缩感知的地震数据重建①

石 敏②* 朱震东* ** 路 昊* ** 朱登明③** 周 军***

(* 华北电力大学控制与计算机工程学院 北京 102206)

(** 中国科学院计算所前瞻研究实验室 北京 100190)

(*** 中国石油集团测井有限公司 西安 710077)

摘 要 在实际勘探中,由于环境、设备或人为因素的影响,采集的地震数据中有很多丢失的数据,严重影响了数据的解释工作。针对这一问题,根据地震数据的时空相关性,提出了一种基于时空约束压缩感知的地震数据重建方法。该方法使用内核奇异值分解(K-SVD)字典学习算法训练超完备字典作为稀疏变换基,进而利用改进的稀疏自适应匹配追踪算法(SAMP)完成重建。通过初始稀疏性估计和变步长策略,减少了 SAMP 中收敛所需的迭代次数。利用真实的地震数据和微电阻率成像数据进行实验,将所提出的方法与压缩感知重建算法进行了比较,不仅提高了重建数据的准确性,而且缩短了执行时间。

关键词 地震数据重建;时空相关性;压缩感知;字典学习

0 引 言

地震勘探是油气勘探的重要手段。随着油气勘探的不断深入,勘探目标区域的结构和周围环境越来越复杂,使得收集的数据不规则甚至不完整,从而对后续分析解释及油气判断产生了影响。同时,这也意味着更高的勘探成本。如果能够基于低密度采样数据重建出准确的高密度地震数据,便能够改善现有勘探方法,有效地利用获得的数据集可为估算地下石油储量的形成提供更可靠的支持,并从一定程度上降低勘探成本。地震数据在变换域中呈现稀疏分布,并且在时域和空域上也显示出很强的相关性,这为地震数据重建提供了可能。

在这项工作中,针对不完整地震数据的压缩感知重建方法,本文分析了地震勘探和测井勘探的数据特点,研究了一种时空约束模型,对传统压缩感知模型进行了改进。本文通过内核奇异值分解(kernel singular value decomposition, K-SVD)算法来训练超完备字典,并使用改进的稀疏自适应匹配追踪算法(sparsity of adaptive matching pursuit, SAMP)解决相应的优化问题,从而完成压缩感知重建。最终,进行了大量的对比实验,验证了本文算法的重建效果及效率。

总而言之,本文的贡献如下。

(1) 分析了地震数据的特征,改进了经典的压缩感知模型,添加时空相关信息作为压缩感知模型的约束。

(2) 改进了稀疏自适应匹配追踪算法,增加了初始稀疏性估计和可变步长的策略,确保了重建精度的同时,提高了算法效率。

(3) 在真实地震数据和微电阻率成像数据上实现了本文算法,证明了其出色的重建能力和泛化能力。

1 相关工作

基于地震数据的重建方法,通常被划分为以下

① 国家科技重大专项(2017ZX05019005)资助项目。

② 女,1975 年生,博士,副教授;研究方向:计算机视觉,服装动画;E-mail: shi_min@ncepu.edu.cn

③ 通信作者,E-mail: mdzhu@ict.ac.cn

(收稿日期:2020-09-01)

3 种:第一种方法是基于预测滤波的,通常通过高斯窗口对不规则样本数据进行插值,这种方法导致很多错误出现[1-2]。第二种方法基于波动方程,通过正反演算子解决一个反问题。这种方法结合地下结构的先验信息,并依据波传播的物理特性实现重建,这种方法通常计算量很大[3]。第三种是基于变换的方法,首先对地震数据进行变换,进而在变换域中通过迭代求解等方法实现地震数据的重建。由于具备稳定性和可解释性,该方法得到了较为广泛的发展和应用[4-5]。Abma 等人[6]在地震重建领域引入了压缩感知算法[7-8]。Wen 等人[9]指出, 在基于压缩感知方法的重建中,影响地震数据重建效果及效率的 3 个主要因素是稀疏变换,迭代算法以及阈值模型。Bora 等人[10]提出了一种基于生成网络的压缩感知算法,该算法实现了基于更少的数据进行重建,但是带来了诸如训练困难和精度不足的问题。

稀疏变换对重建效果有很大的影响,其包括离散余弦变换(DCT),傅立叶变换(Fourier)和超完备字典[11]等。在压缩感知的应用领域中,DCT 和 Fourier 较为常见,但这两种变换都难以有效地识别局部特征。短时傅立叶变换改善了这一问题,但是对于诸如地震数据之类的复杂信号,在不同时间的波形变化很大,短时傅立叶变换的时频局部化能力仍然有限。小波分析采用并完善了这种局部化思想[12],但是不能识别方向。直到后来,提出了曲波(Curvelet)变换[13-14],在保证多尺度识别能力情况下优化了多方向识别能力。因此在地震数据重建领域,曲波变换作为最佳稀疏变换方法之一而被经常应用[15]。剪切波变换[16]对多方向识别能力进一步发展,可以使地震信号的稀疏表示更加稀疏。

目前,基于上述变化的压缩感知理论已被用于地震数据重建[17-19],取得了良好的数据重建效果。但是,上述变换方法有一个共同点,变换基都是预先选择的,并不一定能够适合特定场景数据本身的特征。由于实际的地震数据一般非常复杂,通常会涉及多种类型的元素,预设的稀疏基很难对复杂的地震数据进行有效变换。在相关研究中,使用内核奇异值分解算法来学习超完备字典,能够结合场景数据,训练得到更适应数据特征的字典作为稀疏变换基[20-22]。在最近的研究中,基于字典学习的地震数据处理展示了比传统字典更强的稀疏重建能力[23-25]。

迭代算法不仅会影响到整体的重建效果,还极大程度上决定了算法的效率。由于早期迭代算法(如内点算法[26]和梯度投影法[27])的缺点, 提出了重建贪婪算法,包括传统的贪婪算法,例如匹配追踪算法和正交匹配追踪算法(OMP)[28-29],以及基于传统算法改进得到的分段正交匹配追踪[30]和规则化正交匹配追踪[31]等,但是这些方法通常只应用于信号具有较低的稀疏度的场景。子空间追踪算法(SP)[32]和压缩采样匹配追踪算法(CoSaMP)[33]是两种具有良好性能的相似算法。Blumensath[34]提出了迭代硬阈值(IHT)及其改进算法,进一步提升了重建效果。但是,这些传统的贪婪算法都需要预先获得信号稀疏性。对此,Thong 等人[35]提出了一种自适应估计信号稀疏性的迭代算法,即稀疏自适应匹配追踪(SAMP)算法。解决了估计信号稀疏性的问题,但是并没有得到最佳的重建精度,并且算法效率较低。

综上所述,研究基于压缩感知的地震数据重建,主要在于研究如何基于地震数据的相关特征,构造稀疏效果更好的变换基以及该场景下更高效快速的迭代算法。另外,传统的压缩感知框架下的重建是针对单帧数据进行的,而地震数据在时域和空域上具有相关性,在重建地震数据时,仅考虑单帧数据的信息,而忽略了连续帧之间的相关信息,这将会从一定程度上影响地震数据的重建效果。因此本文也对此也进行了研究。

2　基于时空约束的压缩感知

2.1　地震数据的时空相关性

地震信号是复杂的,甚至瞬态特征也是不稳定的[36]。由于岩层的密度差异的存在,在不同岩层的交界处会存在地震波反射和地震波折射的现象,因此地震波数据与时间域相关,不同深度的地质信息可以由探测器在不同时间点获取地震波来表示。而震波在同一层的岩石中连续传递,因此地震波数据也与空间域有关,不同空间位置的地质信息可以由

波器延测线的水平分布来表示。

目前,传统的压缩感知算法能够基于不完整的地震数据 $y_m \in R^M$,通过重建得到完整的地震数据 $f_m \in R^N (M < N)$。然而,传统的压缩感知算法仅满足了单帧数据重建的合理性,而忽略了地震数据相邻帧之间的相关性。如图 1 所示,利用传统的压缩感知算法,基于单帧信息进行数据重建,重建结果中的高频区域具有较为明显的“失真”现象。因此,在利用压缩感知算法对地震数据进行重建时,有必要利用数据相邻帧之间的相关信息对重建进行优化或约束。

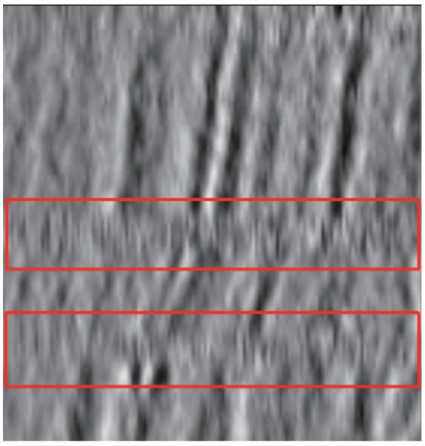

图 1　单帧数据重建结果

2.2　基于时空约束的压缩感知算法

压缩感知算法使用观测矩阵用于描述采样数据和完整地震数据之间的关系。

$$y_m = \boldsymbol{\Phi}_m f_m \tag{1}$$

其中,$\boldsymbol{\Phi}_m \in R^{M\times N}$ 是观测矩阵,而 m 是地震道号。利用稀疏变换基 φ 进行稀疏,完整的地震数据 f_m 可以进而表示为

$$f_m = \varphi \boldsymbol{x}_m \tag{2}$$

其中,$\boldsymbol{x}_m$ 是稀疏向量,将两式结合在一起,可以得到压缩感知的基本表示形式:

$$y_m = \boldsymbol{\theta}_m \boldsymbol{x}_m \tag{3}$$

其中,传感矩阵 $\boldsymbol{\theta}_m = \boldsymbol{\Phi}_m \varphi$,且 $\boldsymbol{\Phi}_m$ 与 φ 不相关。进而可以通过求解 $\boldsymbol{x}_m$ 的最小 L0 范数得到估计值 $\hat{x}_m$,最终获得重建结果 $\hat{f}_m$,即:

$$\begin{aligned} &\hat{x}_m = \operatorname{argmin} \| \boldsymbol{x}_m \|_0 \\ &\text{s.t.}\quad y_m = \boldsymbol{\theta}_m \boldsymbol{x}_m \end{aligned} \tag{4}$$

$$\hat{f}_m = \varphi \hat{x}_m \tag{5}$$

在传统的压缩感知算法基础上进行了修改,增加时空相关性约束,使其能够更好地重建连续多帧地震数据。设 $\{f_{m-n}, f_{m-n+1}, f_{m-n+2}, \cdots, f_m\}$ 为重建的连续 $n+1$ 个数据帧,并定义 R 为连续帧之间变化的能量损失。

$$R = \sum_{i=1}^{n} \| f_{m-i} - f_{m-i+1} \|_1 \tag{6}$$

可以看出,R 可以衡量帧之间的差异,并与相关性呈负相关。将式(2)带入式(6),可以得到:

$$\begin{aligned} R &= \sum_{i=1}^{n} \| \varphi x_{m-i} - \varphi x_{m-i+1} \|_1 \\ &= \sum_{i=1}^{n} \| \varphi (x_{m-i} - x_{m-i+1}) \|_1 \end{aligned} \tag{7}$$

目标是找到 R 的最小值,这等同于求解目标 $\min \sum_{i=1}^{n} \| x_{m-i} - x_{m-i+1} \|_1$。对于稀疏数据,最小 L1 范式和最小 L0 范式从一定程度上是等效的。因此,可以将目标转换为最小 L0 范式的解。

$$\min R \rightarrow \min \sum_{i=1}^{n} \| (x_{m-i} - x_{m-i+1}) \|_0 \tag{8}$$

从而可以为式(4)添加时空相关信息 R,使重建的数据具有尽可能小的变化损失:

$$\min\left(\| x_m \|_0 + \lambda \sum_{i=1}^{n} \| x_{m-i} - x_{m-i+1} \|_0 \right) \tag{9}$$

其中,λ 代表时空相关损失的比例。还可以在式(9)上进行等效转换,以使加入时空信息的模型与压缩感知的表示相同,便于使用迭代算法来重建数据。本文将 λ 设为 1,然后进行等效转换得到:

$$\begin{aligned} &\min \left\| \begin{pmatrix} \boldsymbol{x}_m^{\mathrm{T}},\ (x_{m-1} - x_m)^{\mathrm{T}},\ (x_{m-2} - x_{m-1})^{\mathrm{T}}, \\ \cdots,\ (x_{m-n} - x_{m-n+1})^{\mathrm{T}} \end{pmatrix} \right\|_0 \\ &\text{s.t.} \begin{bmatrix} y_m \\ y_{m-1} \\ \vdots \\ y_{m-n} \end{bmatrix} = \mathbf{A}_m^n \begin{bmatrix} x_m \\ x_{m-1} - x_m \\ x_{m-2} - x_{m-1} \\ \vdots \\ x_{m-n} - x_{m-n+1} \end{bmatrix} \end{aligned} \tag{10}$$

其中 $\boldsymbol{A}_m^n = \begin{bmatrix} \theta_m & 0 & \cdots & 0 \\ \theta_{m-1} & \theta_{m-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{m-n} & \theta_{m-n} & \cdots & \theta_{m-n} \end{bmatrix}$ 为时空传感矩

阵。

3　地震数据重建

本节对地震数据重建算法进行实现。地震数据的特征比较复杂,不同地质环境下的地震数据也有很大差异。非自适应变换基不能很好地表示地震特征。因此,选择利用 K-SVD 算法来训练超完备字典,以此用重建所需的稀疏变换基。本文还通过添加初始稀疏性估计和变步长更新策略来改进 SAMP 算法。基于第 3 节中的优化模型,使用改进的 SAMP 算法来求解稀疏矩阵。将稀疏矩阵和超完备字典相乘以获得重建数据。完整的流程图如图 2 所示。

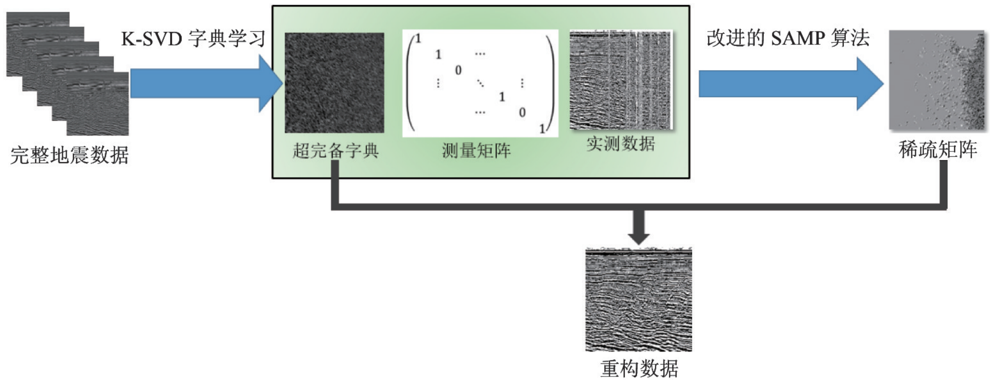

图 2　地震数据重建流程图

3.1　基于 K-SVD 的超完备词典学习

K-SVD 算法的本质是迭代思想,每次迭代使用范数稀疏约束跟踪计算稀疏系数,并利用奇异值分解算法更新字典原子,使得稀疏系数与字典能够得到同步更新,最终可以根据稀疏约束来自适应地训练出超完备字典。

假设,矩阵 $\boldsymbol{Y} = \{y_i\}_{i=1}^{N}$ 是 N 个完整的地震数据切片的集合,代表了字典学习过程中的训练数据,矩阵 $\boldsymbol{X} = \{x_i\}_{i=1}^{N}$ 是 N 个与训练数据 $\boldsymbol{Y} = \{y_i\}_{i=1}^{N}$ 所对应的稀疏表示系数向量集合,矩阵 $\boldsymbol{\varphi} \in R^{n \times K}$ 表示超完备字典。那么超完备字典训练的过程就可以表示为一个优化问题:

$$\min_{\varphi, X}\{\ \| Y - \varphi X \|_F^2\}\ \forall i,\ \ \| x_i \|_0 \leqslant T_0 \tag{11}$$

其中, T_0 是稀疏表示系数中非零元素的最大数量。超完备字典训练的整体步骤如下。

步骤 1　利用某种变换基进行字典初始化。

步骤 2　根据已知字典 $\boldsymbol{\varphi}$,使用迭代算法求解每个样本 y_i 的稀疏系数向量 x_i, 即

$$\min_{x_i}\{\ \| y_i - \boldsymbol{\varphi} x_i \|_2^2\}\quad \| x_i \|_0 \leqslant T_0 \qquad (i = 1,2,\cdots, N) \tag{12}$$

步骤 3　更新字典 $\boldsymbol{\varphi}$。如果满足收敛条件抑或达到预先设置的迭代次数上限,则获得最终字典 $\boldsymbol{\varphi}$,否则转向步骤 2。设向量 d_k 为要更新的字典 $\boldsymbol{\varphi}$ 的第 k 列原子,此时样本集的分解形式可表示为

$$\begin{aligned} \| Y - \varphi X \|_F^2 &= \| ((Y - \sum_{j \neq k} d_j x_T^j) - d_k \boldsymbol{x}_T^k) \|_F^2 \\ &= \| \boldsymbol{E}_k - d_k \boldsymbol{x}_T^k \|_F^2 \end{aligned} \tag{13}$$

其中,向量 $\boldsymbol{x}_T^k$ 是与 d_k 相对应的 $\boldsymbol{X}$ 中第 k 行向量;矩阵 $\boldsymbol{E}_k$ 代表提取 d_k 之后的误差矩阵。

3.2　基于自适应动态步长 SAMP 算法的重建

在地震数据重建领域通常情况下待重建数据的稀疏性是未知的,而在这种情况下,SAMP 算法可以通过设置固定步长 s 动态更新稀疏度来达到最优。通常设定稀疏度为最小值,并在每次迭代中,根据残值判断是否需要增加稀疏度,逐渐逼近信号实际的稀疏度,从而得到稀疏度的最佳估计,最终实现重建结果。

由于 SAMP 算法中步长是固定的,如果将步长 s 设置太大,则信号的真实稀疏性可能会被跳过,得到的结果可能并非全局最优解,这将导致重建精度降低;如果步长 s 设置太小,迭代次数将会显著增加,严重影响了算法的效率。因此,本文对 SAMP 算法进行了两方面的改进,即初始稀疏度 K_0 估计和动态

步长更新策略。

根据文献[37]可知,当观测矩阵 $\boldsymbol{\theta}$ 以参数 (K, σ) 满足 RIP 性质时,若满足 $K_0 \geqslant K$,则存在 $\|\theta_{F_0}^{\mathrm{T}} y_2\| \geqslant \frac{1-\sigma}{\sqrt{1+\sigma}} \|y\|_2$,其中 F_0 为 $\theta^{\mathrm{T}} y$ 中元素绝对值最大的前 K_0 个索引。根据其逆否命题可得,当 $\|\theta_{F_0}^{\mathrm{T}} y\|_2 < \frac{1-\sigma}{\sqrt{1+\sigma}} \|y\|_2$ 时,$K_0 < K$。

由此可以得到初始稀疏度的估计方法,K_0 取初始值 1,若满足 $\|\theta_{F_0}^{\mathrm{T}} y_2\| < \frac{1-\sigma}{\sqrt{1+\sigma}} \|y\|_2$,则增大 K_0 直到不满足该不等式,便得到初始估计值。

具体的公式如下所示。

$$\hat{K}_0 = \underset{K_0}{\operatorname{argmin}} \left(\frac{1-\sigma}{\sqrt{1+\sigma}} \|y\|_2 - \|\theta_{F_0}^{\mathrm{T}} y\|_2 \right) \tag{14}$$

若算法迭代当前轮的稀疏系数估计 $\hat{x}^*$ 与当前最优的稀疏系数估计 $\hat{x}$ 差异较小,则此时稀疏度估计 K 已经接近收敛。因此,可以设定一个阈值 $\boldsymbol{\eta}$ 衡量稀疏系数估计的变化率,当变化率低于阈值时,减少步长,避免跳过最优解。

$$s_k = \begin{cases} \lceil \lambda s_{k-1} \rceil & \dfrac{\|\hat{x}_k^* - \hat{x}_k\|_2}{\|\hat{x}_k\|_2} < \eta \\ s_{k-1} & \dfrac{\|\hat{x}_k^* - \hat{x}_k\|_2}{\|\hat{x}_k\|_2} \geqslant \eta \end{cases} \tag{15}$$

其中 $\lambda \in (0,1)$ 为步长变化率。本文实验设定为 $\eta = 0.1$, $\lambda = 0.5$。改进后的 SAMP 算法完整步骤如下。

输入　传感矩阵 $\boldsymbol{\theta}$,相关度 n,观测向量 $\{y_1, y_2, \cdots, y_n\}$,迭代次数 M,稀疏系数变化率阈值 $\boldsymbol{\eta}$,初始步长 s,步长变化率 λ,邻接矩阵 $\boldsymbol{A}$。

输出　信号稀疏表示系数估计 $\hat{x}$。

步骤 1　根据式(10)构造时空传感矩阵 $\boldsymbol{A}$, $y = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$。

步骤 2　初始化:$g = \boldsymbol{A}^{\mathrm{T}} y$, $F_0 = \varnothing$, $K_0 = 1$。

步骤 3　取 g 中 K_0 个最大值的索引组成 F_0。

步骤 4　如果 $\|\boldsymbol{A}_{F_0}^{\mathrm{T}} y\|_2 < \frac{1-\sigma}{\sqrt{1+\sigma}} \|y\|_2$,则 $K_0 = K_0 + 1$,重复步骤 4;否则,到步骤 5。

步骤 5　初始化:$\hat{x} = 0$, $r_0 = y$, $I = K_0$, $k = 1$。

步骤 6　计算并选取 $|\boldsymbol{A}^{\mathrm{T}} r_{k-1}|$ 中 I 个最大值,将这些值对应 A 的序列号 j 构成集合 S_k。

步骤 7　$C_k = F_{k-1} \cup S_k$, $\boldsymbol{C}_k$ 中所有的序号所对应的 $\boldsymbol{A}$ 的列向量构成 $\boldsymbol{A}_k$。

步骤 8　求 $y = \boldsymbol{A}_k x_k$ 的最小二乘解,$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \|y - \boldsymbol{A}_k x_k\|_2 = (\boldsymbol{A}_k^{\mathrm{T}} A_k)^{-1} \boldsymbol{A}_k^{\mathrm{T}} y$。

步骤 9　从 $\hat{x}_k$ 中选取出绝对值最大的 I 项记为 $\hat{x}_{kI}$, $\boldsymbol{A}_k$ 中对应的 I 项记为 $\boldsymbol{A}_{kI}$,对应的 $\boldsymbol{A}$ 的序列号记为 F。

步骤 10　更新残差 $r = y - \boldsymbol{A}_{kI}(\boldsymbol{A}_{kI}^{\mathrm{T}} A_{kI})^{-1} \boldsymbol{A}_{kI}^{\mathrm{T}} y$。

步骤 11　重建所得 $\hat{x}^*$ 在 F 处的非零项,其值为 $\hat{x}_{kI}$。

步骤 12　如果满足终止条件(r 小于跳出阈值或者达到最大迭代次数),$\hat{x} = \hat{x}^*$,输出 $\hat{x}$。

步骤 13　如果 $\|r\|_2 \geqslant \|r_{k-1}\|_2$,那么转至步骤 14,否则转至步骤 15。

步骤 14　如果 $\frac{\|\hat{x}^* - \hat{x}\|_2}{\|\hat{x}\|_2} < \eta$,那么 $s = \lceil \lambda * s \rceil$, $I = I + s$,重复步骤 6。

步骤 15　$F_k = F$, $r_k = r$, $k = k+1$, $\hat{x} = \hat{x}^*$,重复步骤 6。

4　实验结果和分析

4.1　量化指标

本文在真实地震数据进行量化实验,来验证本文算法的可行性和有效性。衡量数据重建效果的量化指标是信噪比(SNR)以及峰值信噪比(PSNR)。

$$SNR = 10 \lg \frac{\|f\|_2^2}{\|\hat{f} - f\|_2^2} \tag{16}$$

$$PSNR = 10 \lg \frac{\max(y)^2}{MSE} \tag{17}$$

其中,f 是原始数据,$\hat{f}$ 是重建的数据,MSE 是原始数据和重建数据的均方误差。

4.2　稀疏变换矩阵对地震数据重建的影响

本文中的数据集来自实际中收集的完整地震数据。为了进行实验分析和验证,将其设为 601×626 切片数据,用作原始数据集,如图 3(左)所示为原始地震切片数据。为了验证重建算法的有效性,对其

进行了 50% 高斯随机抽样,如图 3(右)所示。为了更好地显示重建结果,计算重建前后的 SNR 和 PSNR,计算结果如表 1 所示。

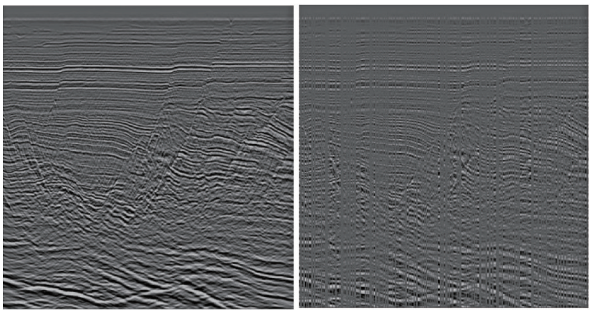

图 3　原始地震数据和 50% 采样地震数据

表 1　重建前后 SNR 和 PSNR 对比

	SNR	*PSNR*
重建前	3.003	45.289
重建后	10.514	49.044

为了验证本文中使用的 K-SVD 词典学习在地震数据上的稀疏重建能力,设置了一种生成式压缩感知(GCS)[10]算法进行比较。在数据恢复之前,使用 K-SVD 算法训练得到 601 × 1052 的超完备字典,然后将学习到的超完备字典用作变换基,利用本文提出的迭代算法对 50% 采样数据进行重建,重建结果如图 4(右)所示。使用生成式压缩感知算法获得的重建结果如图 4(左)所示。

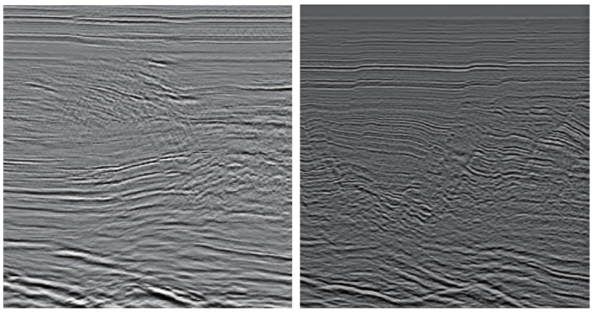

图 4　使用 GCS(左)或使用超完备字典(右)作为转换基础的重建结果

从图 5 的直观重建结果可以看出,尽管生成式压缩感知的重建结果可以获得具有良好图像质量的地震数据,但是重建的地震数据与原始数据中的高频信息并不十分吻合。这反映了基于神经网络的生成式压缩感知重建方法的不确定性,不适用于地震数据重建。另一方面,不同地貌的地震数据差异很大,基于网络训练的模型无法普遍使用,导致该方法在地震数据重建领域的局限性。

为了验证 K-SVD 字典学习方法在地震数据上优于传统的稀疏变换方法,本文使用地震数据重建中常用的 4 种常见稀疏变换矩阵进行数据重建作为对比实验,它们分别是傅立叶变换、离散余弦变换、小波变换和曲波变换。表 2 列出了每个稀疏基的重建结果的平均 SNR 和 PSNR。从表中可以看出,傅立叶变换矩阵的重建精度最差,而超完备字典是其中最佳的稀疏基。

通过以上实验证明,本文选择的 K-SVD 字典学习方法对地震数据的稀疏重建能力较强。

表 2　5 种稀疏基重建地震数据的 SNR 和 PSNR 对比

稀疏基种类	*SNR*	*PSNR*
傅里叶变换	9.263	47.486
离余弦变换	9.426	48.500
小波变换	9.949	48.845
曲波变换	10.182	48.989
超完备字典	10.514	49.044

4.3　迭代算法对地震数据重建的影响

为了比较不同迭代算法之间的性能差异, 使用超完备字典作为数据重建的稀疏变换基, 并使用不同算法重建 50% 的采样数据 10 次并计算平均值。重建效果和运行时间如表 3 所示。

表 3　不同迭代算法对比

	SNR	*PSNR*	时间/s	预设 *K*
OMP($K = 50$)	5.979	46.776	14.194	是
SP($K = 50$)	6.190	46.882	3.741	是
CoSaMP($K = 50$)	5.892	46.733	3.746	是
IHT($K = 50$)	6.804	47.189	2.295	是
IRLS	8.336	47.955	23.882	否
SAMP($s = 5$)	7.881	47.727	18.711	否
本文算法	10.514	49.044	7.063	否

从实验结果的分析可以看出,OMP 算法在重建精度及执行效率方面均不佳。SP、CoSaMP 和 IHT

算法的运行时间较短,但重建精度相对较低,且需要设置稀疏性。IRLS 和 SAMP 不需要设置稀疏度,并且具有更好的重建效果,但是运行时间相对较长。本文方法具有最佳的重建效果。并且与 IRLS 和 SAMP 算法相比,大大缩短了运行时间。

本文增设不同采样率实验,基于 10% ~70% 样本数据进行了重建。其中 OMP、SP、CoSaMP、IHT 设置固定稀疏度 $K = 50$,SAMP 算法设置步长 $s = 5$。不同算法的重建效果,见图 5 和图 6。

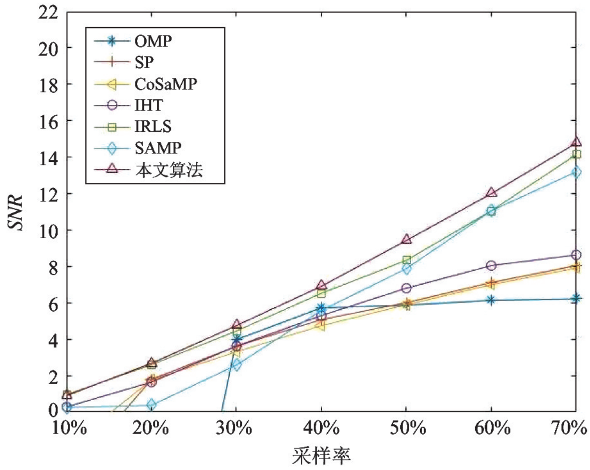

图 5　不同采样率下的 SNR 对照图

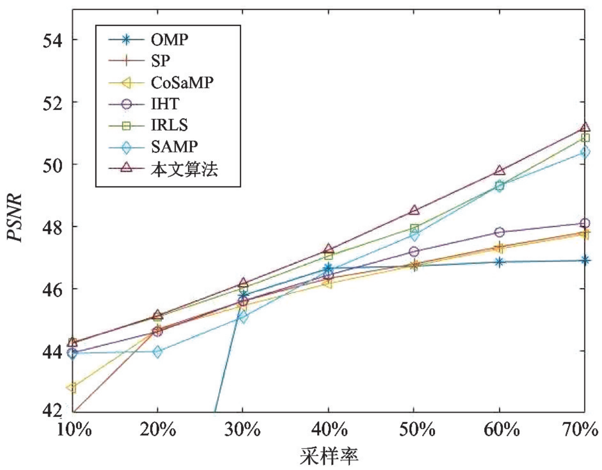

图 6　不同采样率下的 PSNR 对照图

从实验结果的分析来看,需要设置稀疏度的重建算法(OMP、IHT、SP、CoSaMP)无法很好地处理不同采样率情况。而无需设置稀疏度的算法(IRLS,SAMP 和本文方法)在多种采样率情况下效果相对较好,且高采样率情况重建效果更佳。但是,SAMP 算法需要设置步长,如果采样率较低且步长设置得太高,则会导致跳过最准确的稀疏度。当采样率较高时,如果步长较小,则将导致过多的迭代。在实验中,本文方法和 IRLS 可以随着采样率的增加而保持稳定的增长,但是本文方法所需的时间比 IRLS 所需的时间要短得多。

4.4　基于微电阻率成像数据的重建实验

为了更好地验证算法的泛化能力,增加了基于微电阻率成像数据的重建实验。微电阻率成像数据也具有空间相关性的特征,但不同于地震数据的不规则特征。微电阻率成像数据的不完整部分相对较宽,具有一定的规律性。为了证明本文中的算法也可以解决此类问题,本文使用多种算法基于真实的微电阻率成像数据进行重建,原始数据大小为 360 ×1000。重建结果显示在图 7 中。对局部进行放大,显示结果如图 8 所示。

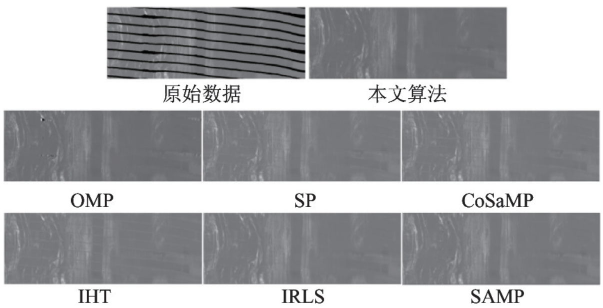

图 7　原始电成像数据及不同算法重建结果

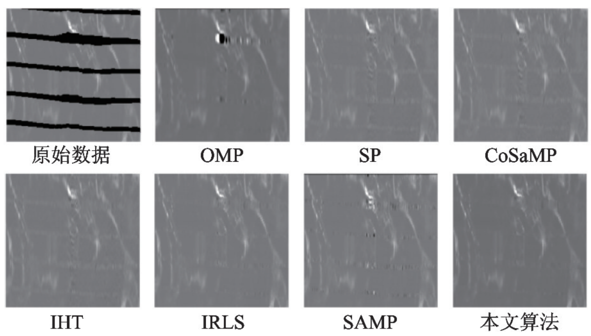

图 8　局部放大后的结果

从恢复效果可以看出,除去本文算法,其他算法的重建结果都有不同程度的“失真”。由于 OMP、SP、CoSaMP 和 IHT 算法需要预设稀疏度,稀疏度设置不佳,重建结果会稍微模糊。在实际生产过程中,具体的稀疏度通常未知,因此这些算法不适用于该

类问题。IRLS和SAMP算法重建结果细节较好,但是所需的计算时间太长。而且,这6种重建方法重建结果的缺陷部位均有明显的重建痕迹。与其他算法相比,本文算法利用了空间相关信息,使重建结果细节更清晰,水平过渡更平滑,并且相比IRLS和SAMP大大减少了操作时间。证明了本文算法具有较强的稀疏重建能力和泛化能力。

5 结论

本文提出了一种基于时空约束压缩感知的地震数据重构算法。该算法对传统的压缩感知理论模型进行了修改,增加了时空约束,并使用K-SVD算法基于现有地震数据训练并获得超完备字典来代替传统变换。然后,对SAMP算法进行了改进,提出初始稀疏估计和可变步长更新策略,从而大幅减少了迭代次数,提高了重构精度。将生成式压缩感知和传统压缩感知方法作为本文算法的对照实验,并在真实地震数据及微电阻率成像数据上进行实验,结果表明了本文提出的重建方法相比其他方法具有较强的重建能力。此外,本文方法与其他方法一样,在低采样率情况下无法很好地重建。生成式压缩感知方法在本文实验中没有能够展示出很好的结果,这是由于该方法约束不足而具有很强的不确定性,仍然将基于神经网络的重建方法作为未来的研究点,这可能将会是实现基于更低采样率重建的新突破口。

参考文献

[1] Naghizadeh M, Sacchi M D. Multistep autoregressive reconstruction of seismic records[J]. *Geophysics*, 2007, 72(6):111-118

[2] Liu G C, Chen X H. Seismic data interpolation using frequency-domain complex nonstationary autoregression[J]. *Geophysical Prospecting*, 2018, 66(3):478-497

[3] 李学聪, 刘伊克, 常旭, 等. 基于F-K偏移和反偏移的地震道插值方法研究[J]. 地球物理学进展, 2009, 24(6): 2020-2029

[4] Zwartjes P M, Sacchi M D. Fourier reconstruction of nonuniformly sampled, aliased data[J]. *Geophysics*, 2007, 72(1): 21-32

[5] Naghizadeh M, Sacchi M D. Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data[J]. *Geophysics*, 2010, 75(6): 189-202

[6] Abma R, Kabir N. 3D interpolation of irregular data with a POCS algorithm[J]. *Geophysics*, 2006, 71(6): 91-97

[7] Donoho D. Compressed sensing[J]. *IEEE Transaction on Information Theory*, 2006, 52(4):1289-1306

[8] Candes E, Wakin M. An introduction to compressive sampling[J]. *IEEE Signal Processing Magazine*, 2008, 25(2): 21-30

[9] 温睿, 刘国昌, 冉扬. 压缩感知地震数据重建中的三个关键因素分析[J]. 石油地球物理勘探, 2018, 53(4): 682-693

[10] Bora A, Jalal A, Price E, et al. Compressed sensing using generative models[J]. *arXiv*:1703.03208, 2017

[11] Anvari R, Mohammadi M, Kahoo A R, et al. Random noise attenuation of 2D seismic data based on sparse low-rank estimation of the seismic signal[J]. *Computers and Geosciences*, 2020, 135: 104376

[12] Daubechise I. Ten Lectures on Wavelets[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1992

[13] Candes E J, Demanet L, Donoho D L, et al. Fast discrete curvelet transforms[J]. *Multiscale Modeling and Simulation*, 2005, 5: 861-899

[14] Candes E J, Donoho D L. New tight frames of curvelets and optimal representations of objects with C2 singularities[J]. *Communications on Pure and Applied Mathematics*, 2004, 57(2): 219-266

[15] Ma J, Plonka G. A review of curvelets and recent applications[J]. *IEEE Signal Processing Magazine*, 2010, 27(2): 118-133

[16] 冯飞, 王征, 刘成明, 等. 基于Shearlet变换稀疏约束地震数据重建[J]. 石油物探, 2016, 55(5):682-691

[17] Gholami A. Non-convex compressed sensing with frequency mask for seismic data reconstruction and denoising[J]. *Geophysical Prospecting*, 2015, 62(6):1389-1405

[18] Wang X, Geng Y, Wu R S. Seismic data reconstruction in Dreamlet domain based on compressive sensing[J]. *Oil Geophysical Prospecting*, 2015, 50(3):399-404

[19] Gan S, Wang S, Chen Y, et al. Compressive sensing for seismic data reconstruction via fast projection onto convex sets based on seislet transform[J]. *Journal of Applied Geophysics*, 2016, 130:194-208

[20] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries[J]. *IEEE Transactions on Image Process*, 2006, 15(12): 3736-3745

[21] Sajjad M, Mehmood I, Baik S W, et al. Sparse coded image super-resolution using K-SVD trained dictionary based on regularized orthogonal matching pursuit[J]. *Bio-medical Materials and Engineering*, 2015, 26(51): 1399-1407

[22] Raja H, Bajwa W U. Cloud K-SVD: a collaborative dictionary learning algorithm for big, distributed data[J]. *IEEE Transactions on Signal Processing*, 2016, 64(1): 173-188

[23] Lingchen Z, Entao L, Mcclellan J H. Joint seismic data denoising and interpolation with double-sparsity dictionary learning[J]. *Journal of Geophysics and Engineering*, 2017, 14(4): 802-810

[24] Wu J, Bai M. Incoherent dictionary learning for reducing crosstalk noise in least-squares reverse time migration [J]. *Computers and Geosciences*, 2018,114(C): 11-21

[25] Sun Y, Jia R, Sun H, et al. Reconstruction of seismic data with missing traces based on optimized poisson disk sampling and compressed sensing[J]. *Computers and Geosciences*, 2018,114(C): 32-40

[26] Kim S J, Koh K, Lustig M, et al. An interior-point method for largescale regularized least squares[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1(4): 606-617

[27] Fiqueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1(4): 586-597

[28] Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit[J]. *IEEE Transactions on Information Theory*, 2007, 53(12): 4655-4666

[29] Wang R, Zhang J L, Ren S L, et al. A reducing iteration orthogonal matching pursuit algorithm for compressive sensing[J]. *Tsinghua Science and Tecnology*, 2016, 21(1): 71-79

[30] Donoho D L, Tsaig Y, Drori I, et al. Sparsesolution of underdetermined linear equations by stagewise orthogonal matchingpursuit[J]. *IEEE Transactions on Information Theory*, 2012, 58(2): 1094-1121

[31] Sun H M, Jia R S, Zhang X L, et al. Reconstruction of missing seismic traces based on sparse dictionary learning and the optimization of measurement matrices[J]. *Journal of Petroleumence and Engineering*, 2019, 175:719-727

[32] Dai W, Milenkovic O. Subspace pursuit for compressive sensing signal reconstruction[J]. *IEEE Transactions on Information Theory*, 2009, 55(5): 2230-2249

[33] Needell D, Tropp J A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples[J]. *Applied and Computational Harmonic Analysis*, 2009,26(3): 301-321

[34] Blumensath T. Accelerated iterative hard thresholding [J]. *Signal Processing*, 2012, 92(3):752-756

[35] Thong T D, Gan L, Nguyen N, et al. Sparsity adaptive matching pursuit algorithm for practical compressed sensing[J]. *Asilomar Conference on Signals, Systems, and Computers*, 2008, 10: 581-587

[36] Sun H, Jia R, Du Q, et al. Cross-correlation analysis and time delay estimation of a homologous micro-seismic signal based on the Hilbert-Huang transform[J]. *Computers and Geosciences*, 2016: 98-104

[37] Yang C, Feng W, Feng H, et al. A Sparsity adaptive subspace pursuit algorithm for compressive sampling[J]. *ACTA Electronica Sinica*, 2010, 38(8): 1914-1917

Seismic data reconstruction based on space-time constraint compressed sensing

Shi Min*, Zhu Zhendong* **, Lu Hao* **, Zhu Dengming**, Zhou Jun***

(* Control and Computer Engineering Institute, North China Electric Power University, Beijing 102206)

(** Prospective Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(*** China Petroleum Logging Co. Ltd, Xi'an 710077)

Abstract

In actual exploration, due to the influence of environment, equipment or human factors, there are a lot of missing data in the seismic data collected, which seriously affects the data interpretation work. Aiming at this problem, according to the space-time correlation of seismic data, a method of seismic data reconstruction based on space-time constrained compressed sensing is proposed. In this method, an over-complete dictionary as a sparse transform basis is trained using kernel singular value decomposition (K-SVD) dictionary learning algorithm. The reconstruction is accomplished using an improved sparsity of adaptive matching pursuit (SAMP). By incorporating an initial sparsity estimation step and adopting a variable step size strategy, the number of iterations needed for convergence in SAMP can be significantly reduced. Using real seismic data and micro-resistivity imaging data, the proposed novel method is compared with state-of-the-art compressive sensing reconstruction algorithms. The experimental results show that the accuracy of the reconstructed data is significantly improved, and the execution time is also reduced.

Key words: seismic data reconstruction, space-time correlation, compressed sensing, dictionary learning