

doi: 10.11720/wtyht.2022.1458

王德英, 张凯, 李振春, 等. 基于分段采样和 MCA 的地震数据重建[J]. 物探与化探, 2022, 46(5): 1214–1224. <http://doi.org/10.11720/wtyht.2022.1458>

Wang D Y, Zhang K, Li Z C, et al. Seismic data reconstruction based on segmented random sampling and MCA[J]. Geophysical and Geochemical Exploration, 2022, 46(5): 1214–1224. <http://doi.org/10.11720/wtyht.2022.1458>

基于分段采样和 MCA 的地震数据重建

王德英¹, 张凯², 李振春², 张医奎³, 许鑫¹

(1. 中国石油勘探开发研究院西北分院, 甘肃 兰州 730030; 2. 中国石油大学(华东) 地球科学与技术学院, 山东 青岛 266580; 3. 物华能源科技有限公司, 陕西 西安 710067)

摘要: 数据重建是地震资料处理中一项重要的前期工作。压缩感知(compress sensing, CS)已经在数据重建领域取得了很好的应用。CS 的关键是采样的随机性, 随机采样将常规欠采样引起的互相干假频转化为较低能量的不相干噪声。一方面, 传统的随机采样方法缺乏对采样点的约束, 导致产生过多的噪声干扰, 分段随机采样可有效地控制采样点之间的距离。另一方面, 单一的数学变换会导致信号的不完全稀疏表达, 影响数据重建效果, 形态分量分析(morphological component analysis, MCA)将信号分解成几个具有显著特征的成分以逼近数据复杂的内部结构。本文在 MCA 框架下找到了一个新的字典组合(Shearlet+DCT), 并使用块坐标松弛(block coordinate relaxation, BCR)算法得到最优解, 从而获得理想重构结果。对实际资料的实验表明, 该方法在重建分段随机采样数据时具有较好效果。

关键词: 压缩感知; 分段随机采样; 形态分量分析; 数据重建

中图分类号: P631.4

文献标识码: A

文章编号: 1000-8918(2022)05-1214-11

0 引言

在实际地震勘探中, 数据体内容复杂且容量巨大, 加之野外地形条件和经济成本的限制, 地震勘探数据在空间方向上经常进行不规则采样。在陆地上这种情况可能起因于地震道的布置受到建筑、湖泊、禁采区等复杂地形限制, 另外人工放炮的不充分激发也可能起因于坏道或者严重污染道的出现, 从而导致采集到的地震数据不规则、不完整或出现空间假频。在海上地震勘探中, 则可能由于电缆的水平偏移等引起, 这种稀疏分布的地震数据难以满足高分辨地震勘探的需要^[1]。

压缩感知数据恢复的 3 个关键因素分别是: 对待恢复数据的稀疏表示、对完整数据的随机采样和最优优化差值重建算法。其中基于压缩感知的数据采

样通常是完全随机采样, 这种采样方式会导致采样点间隔过于稀疏或密集, 当采样点过于稀疏时会导致重要信息丢失, 影响恢复效果; 采样点过于密集则造成多余采样, 违背压缩感知节约经济成本的原则。唐刚^[2]引入泊松碟采样方式, 提高了重建效果, 但该采样方式由于事先固定采样, 也不能根据野外地质条件灵活调整, 使其在实际应用中有一定局限。Leneman 提出了 Jitter 采样方式^[3], 首先应用于图像处理。Hennenfent G 等^[4]指出在利用非线性稀疏性促进求解方法时, 随机采样相比等间距的规则采样可以获得更好的数据重建效果。Hennenfent G 等^[4]、唐刚^[2]将其引入到地震数据重建中, 使采样点分布相对均匀, 在一定程度上能够根据地质条件灵活调整。Herrmann F J 等^[5]使用 Jitter 采样进行了基于压缩感知的地震数据重建。Mosher C C^[6]提出了 NUOS 采样方案并用于在感知成像技术(compressed

收稿日期: 2021-11-25; 修回日期: 2022-05-18

基金项目: 中国石油勘探与生产分公司科技项目“薄储层全频处理方法研究与目标精细刻画技术攻关试验”(2022KT1503)

第一作者: 王德英(1997-), 男, 2019 年毕业于防灾科技学院(现应急管理大学)并获得地球物理学学士学位, 同年进入中国石油大学(华东)就读硕士研究生, 主要从事压缩感知在地震勘探中的应用、速度建模、全波形反演等方面的研究工作, 现在中国石油勘探开发研究院西北分院从事高分辨率处理和机器学习等方面的研究工作。Email: 498987895@qq.com

sensing imaging, CSI) 中寻找炮点和接收点的最佳位置。Yang 等^[7]在 2015 年完成 CSI 技术在陆地、海洋和 OBN 勘探中的生产和应用,提出了一种分段随机采样方法,并对其进行了理论分析。以上研究结果表明,与随机采样和 Jitter 采样相比,相同次数下分段随机采样的重建效果优于其他两种方法。

压缩感知理论的一个重要前提是信号的稀疏性。由于地震数据通常不是稀疏的,因此如何找到合适的稀疏字典,使得信号在该字典下的系数是稀疏的,即能对信号进行稀疏表示,就显得尤为重要。现今已有多种数学变换被用于地震数据重建,如 Fourier 变换^[8]、Curvelet 变换^[5]、Wavelet 变换^[9]、Dreamlet 变换^[6]、Shearlet 变换^[10]等。由于地震数据通常由不同类型的元素组成,单一变换往往不能充分、有效地将其全部表达,李海山等^[11]提出形态分量分析(morphological component analysis, MCA)方法。该方法主要利用不同信号之间成分的差异对信号进行分离,最早由 Starck 等提出并将其应用于图像的去噪与修复,且取得良好效果,目前被广泛应用于信号的去噪、重建、分隔、修复以及融合等多种领域。李海山等^[11]将 MCA 方法引入地震数据重建领域,取得了较好效果。周亚同等^[12]在 MCA 框架下定量评价了多种不同稀疏字典组合的数据重建效果,判定离散余弦变换(discrete cosine transform, DCT)与 Curvelet 字典组合具有最高的重建精度。刘成明等^[10]和张良等^[13]指出 Shearlet 变换作为一种新型的多尺度几何变换,具有比 Curvelet 变换更好的方向性以及其自身所具有的各向异性特点,从而拥有更突出的稀疏表达能力,认为基于 Shearlet 变换的重建算法可望取得精度更高的重建结果。

受上述研究的启发,将 Shearlet 变换应用到 MCA 中,它能准确地描述数据的局部特征,因此,有必要寻求一种全局型变换来与之配合。经过反复测试,确定 Shearlet+DCT 的重建效果优于单一数学变换和其他字典组合。此外,我们还比较了不同采样方法的重建结果,证明了 SRS(segmented random sampling)更有利于 CS 的重建。本文采用 SRS 和 Shearlet+DCT 对实际资料进行处理,均取得理想结果。

1 基本理论

1.1 压缩感知理论

缺失地震数据的恢复可以看作一个压缩感知问题,一个不满足奈奎斯特采样定理的信号可以通过

某些稀疏变换在特定的稀疏域来恢复。地震数据的重建问题可以描述为由一组不完整数据通过线性算子的作用恢复出完整的数据,假设线性正演模型:

$$\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{e}, \quad (1)$$

其中: $\mathbf{y}$ 代表采集的地震数据; $\mathbf{x}$ 表示待恢复的数据; $\mathbf{R}$ 表示随机采样矩阵; $\mathbf{e}$ 代表随机噪声。

压缩感知技术的第一条件是信号可以在某一稀疏域被稀疏表示,假设数据 $\mathbf{x}'$ 是数据 $\mathbf{x}$ 在稀疏域 $\mathbf{F}$ 的稀疏表示,则方程(1)可以改写为:

$$\mathbf{y} = \mathbf{A}\mathbf{x}' + \mathbf{eA} = \mathbf{R}\mathbf{F}^H, \quad (2)$$

其中, $\mathbf{F}^H$ 是稀疏变换 $\mathbf{F}$ 的共轭转置矩阵,如果 $\mathbf{x}'$ 中有 k 个非零元素且 $k \ll L$,我们视 $\mathbf{x}'$ 是稀疏的,它可以通过凸优化问题来恢复^[12]。

$\langle \mathbf{x} \rangle = \arg \min \| \mathbf{x}' \|_1$ s. t. $\| \mathbf{y} - \mathbf{A}\mathbf{x}' \| \leq \sigma$ (3)
式中: $\langle \mathbf{x} \rangle$ 是 $\mathbf{x}'$ 的恢复结果; l_1 范数 $\| \mathbf{x}' \|_1$ 代表 $\mathbf{x}'$ 中所有元素绝对值的总和; σ 代表规定最大恢复误差。

通过求解式(2)中的范数 $\| \mathbf{x}' \|_1$,找到可以满足误差 σ 的最稀疏解,然后通过逆稀疏变化得到恢复数据 $\mathbf{y}'$,如式(4):

$$\mathbf{y}' = \mathbf{F}^T \langle \mathbf{x} \rangle. \quad (4)$$

1.2 基于形态分量分析的地震数据重建

MCA 方法认为,信号是由形态不同的多个分量线性组合而成,每个分量都能找到一个字典,该字典只能稀疏表示对应的形态分量,而无法稀疏表示其他分量。

给定二维信号 X ,假设它由 K 个不同形态的分量线性组合而成,即

$$\mathbf{X} = \sum_{k=1}^K \mathbf{X}_k = \sum_{k=1}^K \mathbf{D}_k \mathbf{a}_k, \quad (5)$$

式中: $\mathbf{X}_k$ 为 K 个形态分量,它可被字典 $\mathbf{D}_k$ 稀疏表示; $\mathbf{a}_k$ 为对应的稀疏系数。

地震数据的重建可表示为:

$$\mathbf{Y} = \mathbf{R}\mathbf{X}, \quad (6)$$

式中: $\mathbf{X}$ 表示理想的完整地震数据; $\mathbf{Y}$ 表示实际采集的含缺失道地震数据; $\mathbf{R}$ 表示采样矩阵。

在 MCA 框架下,式(6)可被表示为:

$$\mathbf{Y} = \mathbf{R}\mathbf{X} = \mathbf{R} \sum_{k=1}^K \mathbf{X}_k = \mathbf{R} \sum_{k=1}^K \mathbf{D}_k \mathbf{a}_k, \quad (7)$$

将式(7)改写为无约束最优化问题,求解得:

$$\mathbf{a}_k = \arg \min \| \mathbf{a}_k \|_1 + \lambda \left\| \sum_{k=1}^K \mathbf{D}_k \mathbf{a}_k \right\|_2^2, \quad (8)$$

式中: λ 为拉格朗日乘子,用来平衡 l_1 范数项和 l_2 范数项所占比重。

1.3 MCA 求解方法

对于恢复算法的求解,Sardy 等^[9]结合块坐标松

弛 (Block coordinate relaxation, BCR) 算法给出如下基于形态分量的求解方法。

输入: 采样矩阵 R , 字典组合 $D=[D_1, \cdots, D_K]$, 缺失地震数据 Y , 总迭代次数 N

输出: 重建地震数据 X

初始化: 每个形态分量 $X_i^{(0)}=0, i=1, \cdots, K$

- 1) for: $n=1:N$;
- 2) 残差 $r^{(n)}=Y-R \cdot \text{sum}(X_1, \cdots, X_K)$
- 3) for: $k=1:K$
- 4) $\alpha_k^{(n)}=D_k(X_k^{(n)}+r^{(n)})$
- 5) $X_k^{(n)}=D_k^{-1}T_\lambda(\alpha_k^{(n)})$
- 6) 应用阈值模型减小 λ
- 7) end
- 8) end
- 9) $X=\text{sum}(X_1, \cdots, X_K)$

其中: D^{-1} 表示字典 D 对应的逆变换; T_λ 为阈值函数, 其公式为:

$$T(x, \lambda) = x \cdot \exp\left[-\left(\frac{\lambda}{|x|}\right)^{2-p}\right] \quad (9)$$

除了阈值 λ 外, 我们还有另一个独立参数 p 可以灵活选择, 以获得更好的性能。在泰勒级数的基础上, 指数阈值可被认为是一个光滑的 0 范数约束。当 $|x| \gg \lambda$ 时, 它是 p 阈值算子的近似, 可简化为 $p=0$ 时的 Stein 阈值算子和 $p=1$ 时的软阈值。在求解算法时, 需要调整拉格朗日乘子来得到最优解。步骤如下: 首先选取较大的变换域系数作为阈值, 得到稀疏近似解。其次, 不断减小该值以包含更多的变换域系数, 并通过连续迭代逼近最优解。阈值选择策略称为阈值模型, 它影响算法的速度和精度。标准阈值模型包括线性模型、指数模型和数据驱动模型。本文采用指数模型, 其形式为:

$$\lambda_n = \left(\frac{\lambda_{\min}}{\lambda_{\max}}\right)^{\frac{n-1}{N-1}} \cdot \lambda_{\max}, n=1, 2, \cdots, N$$
$$\lambda_{\max} = q_{\max} \cdot \max_i \{ | (D * Y) |_i \},$$
$$\lambda_{\min} = q_{\min} \cdot \max_i \{ | (D * Y) |_i \}, \quad (10)$$

式中: $\lambda_{\max}$ 、 $\lambda_{\min}$ 分别为所选最大、最小正则化参数; $q_{\max}$ 和 $q_{\min}$ 是用户定义的百分比; N 是迭代次数。

1.4 稀疏字典的选取

稀疏字典的选取是 MCA 方法的核心问题, 不同的字典对稀疏表示的效果影响很大。周亚同等^[12]已对 Curvelet 字典的不同组合方案进行了实验, 最终认定离散余弦变换 (DCT) 与 Curvelet 字典组合具有最高重建精度。Curvelet 具有刻画多尺度、多方向各向异性的能力, 能够很好地表达地震信号中的

曲线状边缘特征; 但与 Shearlet 变换相比, 其稀疏表示能力相对弱一些。Shearlet 变换的数学结构较简单, 它通过对一个函数进行伸缩、平移、旋转来生成基函数, 该特点与小波类似, 却正是 Curvelet 变换所欠缺的, 所以 Shearlet 字典非常适用于图像恢复重建。DCT 变换是一种在信号和图像处理中常用的数学变换, 它具有一个很重要的特性——能量集中, 即经 DCT 变换后信号能量非常集中, 且大多数系数都集中在低频部分。经过多次的对比实验得出利用 MCA 框架下 Shearlet 和 DCT 字典组合的方法进行数据重建的效果最好。

2 采样方法

压缩感知的关键技术之一是随机欠采样方法, 因为它可以把混淆真实频率的相干噪声转化成容易滤除的不相干噪声, 即使这些随机噪声和真实频谱相互叠合在一起, 仍然可以通过迭代的方式检测出真实频率。因此, 相干噪声的程度很自然地就成为衡量一种欠采样方法效果的重要标准之一。然而单纯的随机采样会造成采样点过于聚集或者过于分散的情况, 难以达到理想的重建效果, 同时也不能根据野外实际地质条件灵活调整, 因此引入 Jitter 欠采样方式, 以控制采样间隔, 更加适合地震数据重建。Jitter 采样方法首先将待处理区域划分成若干个子区域, 然后在每个子区域内都随机地强制采一个点。由于每个子区域都有采样点, 相邻缺失地震道之间的间隔也就得到了控制, 同时也保持采样点的随机性, 可将假频转化成低幅度噪声, 使真实频率更容易检测^[14-23]。但是 Jitter 欠采样过程中会存在剩余采样点, 不容易使 Jitter 参数与离散随机变量相等, 导致无法达到最优 Jitter 采样。分段随机采样在保证采样点的规则性与随机性的同时, 很好地克服了这一缺点。

分段随机采样的设计主要分为两个步骤: 第一步是根据项目要求及工区实际地质条件, 设计一个充分完整规则的观测系统。压缩感知的作用与目的就是尽可能地减少炮点和检波点的数量, 不仅可以节约成本也可以扩大地震勘探区域或提高空间分辨率^[21, 24-26]。第二步则是对第一步生成的完整观测系统进行优化。非规则观测系统设计步骤如下:

- 1) 根据地震勘探项目与客观条件确定欠采样缺失因子 P 。
- 2) 假设一条有 N 个检波器的接收道可以表示为:

$$X = [x(1), x(2), \cdots, x(N)]。$$
 (11)

3) 给定一个分段长度 L , X 被分成若干个长度为 L 的 S_i , 在 S_i 内根据欠采样缺失因子 P 确定采样点数量, 则实际的接收点可以表示为:

$$X_r = x(h(m)), X_0 = 0,$$
 (12)

其中, $h(m)$ 代表随机采集 m 个点, 0 代表未采集的点。

- 4) 炮点的不规则化方法与采样点相同。
- 5) 将不规则的炮点和检波点输出到 SPS 文件中。

图 1 是分段随机采样的示意, 表现了规则采样点不规则化的过程。从图中可以看出几个采样点共同位于一个小段 ($P=50%$) 中, 黑色代表采样点, 白色代表未被采样的点。每一小段的长度 $L=4$, 欠采样缺失因子。沿着这条检波道, 每 4 个点被分成一段, 段中 50% 的点被随机选择。每小段的采样点个

数为 $L \times P=2$, 此检波道的最大采样间隔为 4。在选择分段长度 L 时也要考虑欠采样缺失因子的大小, 尽量使 $L \times P$ 的结果是整数^[5], 最大采样间隔也是由这 2 个参数来决定。如果 $L=4, P=25%$, 则采样间隔是 6。从图 2f、g、h 可以看出, 最后 3 种采样方法所采集的缺失数据没有假频, 频谱中只存在部分噪声。相比之下, 分段随机采样所产生的噪声最小。从时间域图像分析, 分段随机采样能更好地保留有效信息。

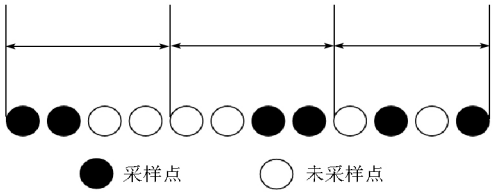
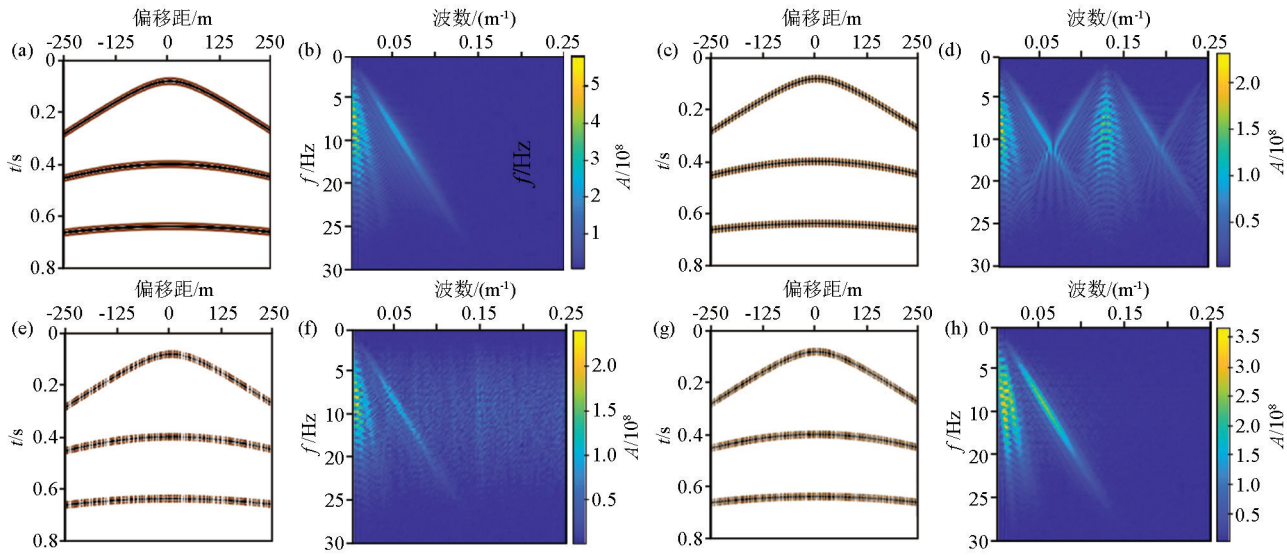


图 1 分段随机采样示意

Fig. 1 Schematic diagram of segmented random sampling



a—合成地震记录;b—合成地震记录频谱;c—规则欠采样数据;d—规则欠采样数据频谱;e—完全随机采样数据;f—完全随机采样数据频谱;g—分段随机采样数据;h—分段随机采样数据频谱
a—synthetic seismic record;b—spectrum of synthetic seismic record wavenumber;c—synthetic seismic data after regular sampling;d—spectrum of synthetic seismic data after regular sampling wavenumber;e—synthetic seismic data after random sampling;f—spectrum of synthetic seismic data after random sampling wavenumber;g—synthesized seismic record after SRS;h—spectrum of synthesized seismic record after SRS

图 2 欠采样数据及其频谱

Fig. 2 Sampled data and spectrum

3 数值实验

3.1 评价参数

为了对重建结果进行定量描述, 本文引入两个评价参数, 其定义如下:

1) 信噪比 $R_{S/N}$

$$R_{S/N} = 10 \times \lg \left(\frac{\|X\|_2^2}{\|\hat{X} - X\|_2^2} \right),$$
 (13)

其中, X 表示完整地震数据; $\hat{X}$ 表示重建地震数据。

2) 峰值信噪比 $R_{S/N}^p$

$$R_{S/N}^p = 10 \times \lg \left[\frac{\max(X)^2}{M_{SE}} \right],$$
 (14)

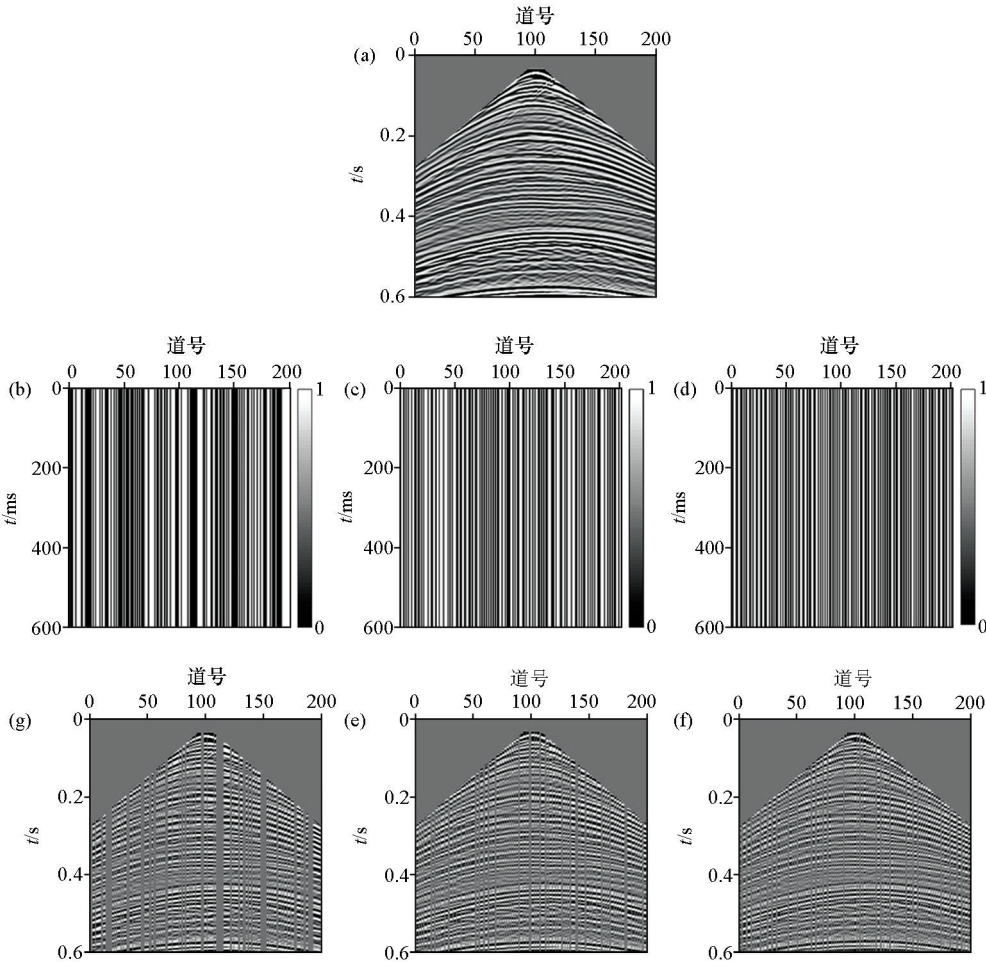
其中, M_{SE} 表示均方差, 其计算公式如下:

$$M_{SE} = \frac{1}{mn} \sum_{i=0}^m \sum_{j=0}^n \|x(i,j) - \hat{x}(i,j)\|_2^2. \quad (15)$$

3.2 实际数据模型试算

为了充分说明分段随机采样与 MCA 重建技术的优越性,选用 3 种随机采样方法(完全随机采样、Jitter 采样和分段随机采样)对原始数据进行采样,得到 3 种缺失数据,比较它们单一 DCT 字典、单一 Shearlet 字典和 MCA 框架下 Shearlet 和 DCT 字典组合的重建效果。本文选择了某地区真实的炮记录进行实验(为便于观察重建质量,数据已经过去噪处

理),其时间采样频率为 1 ms,含有 200 道,每道有 600 个采样点。图 3b、c、d 分别是与数据大小相等的随机采样矩阵、Jitter 采样矩阵和分段随机采样矩阵,采样点的缺失率均为 50%;图 3e、f、g 分别是图 3a 经过 3 种采样矩阵作用得到的 3 种缺失数据,可以看出图 3e 中存在多连续道缺失,过多的连续数据缺失是影响重建效果的主要原因。相反,图 3f、g 保留了缺失道的随机性也控制了采样点的最大间隔,对比图 3c、d 可以看出分段随机采样的这种控制能力更强。



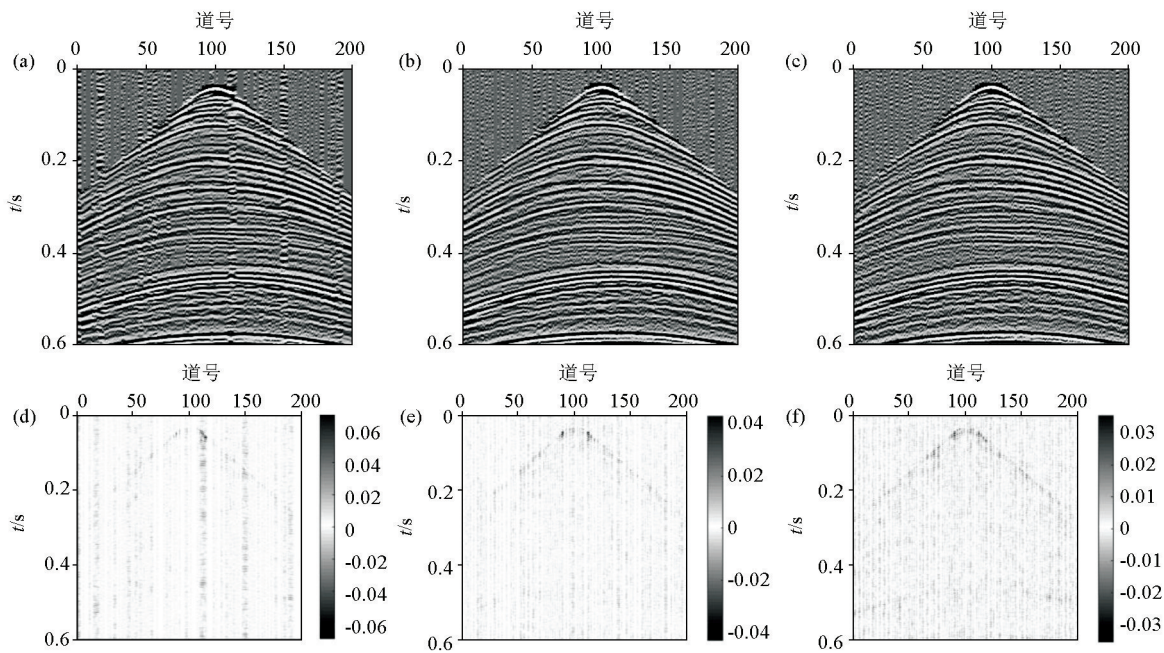
a—原始数据;b—随机采样矩阵;c—Jitter 采样矩阵;d—分段随机采样矩阵;e—完全随机采样数据;f—Jitter 采样数据;g—分段随机采样数据
a—original data;b—Random sampling matrix;c—Jitter sampling matrix;d—segmented random sampling matrix;e—missing data after random sampling;f—missing data after Jitter sampling;g—missing data after segmented random sampling

图 3 原始数据、采样矩阵和缺失数据

Fig. 3 Original data, sampling matrix and missing data

此实验在使用单一稀疏域进行重建时选用 DCT 字典和 Shearlet 字典,它们分别通过 DCT 变换和 Shearlet 变换生成,求解算法为 FPOCS 算法,其迭代次数均是 120 次。图 4 是利用单一 DCT 字典的重建结果和所对应的差值剖面。差值剖面通过重建数据与原始数据作差来得到,在没有造影参与的情

况下可用来表示重建误差。从图中可以看出在单一 DCT 稀疏域中,无论随机采样、Jitter 采样和分段随机采样都获得不理想的重建效果。这是因为 DCT 字典的特征是能量集中且大多数系数都集中在低频,使其在重建较为复杂的实际资料时具有一定的局限性。对比图 4a、b、c 可以明显的看出,图 4c 的



a—随机采样的 DCT 重建结果;b—Jitter 采样的 DCT 重建结果;c—分段随机采样的 DCT 重建结果;d—随机采样的 DCT 差值剖面;e—Jitter 采样的 DCT 差值剖面;f—分段随机采样的 DCT 差值剖面
a—DCT reconstruction result of random sampling;b—DCT reconstruction result of Jitter sampling;c—DCT reconstruction result of segmented random sampling;d—DCT reconstruction result of segmented random sampling;e—DCT difference profile of Jitter sampling;f—DCT difference profile of segmented random sampling

图 4 DCT 字典重建数据与差值剖面

Fig. 4 Reconstruction results and difference profiles of DCT dictionary

同相轴最为光滑、连续,对比图 4d、e、f 也可以看出分段随机采样的重建误差的浮值(分段随机采样重建误差绝对值 0.03;Jitter 采样重建误差绝对值 0.04;随机采样重建误差绝对值 0.06)低于前两者,这也说明了分段随机采样的优越性。

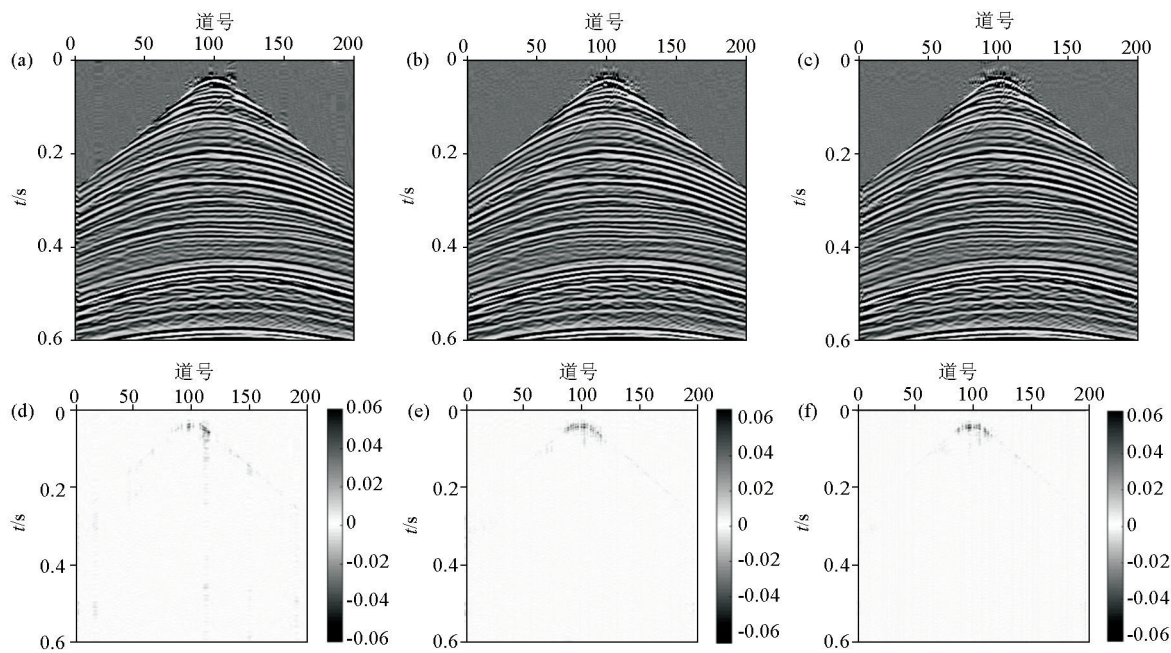
对比图 4 和图 5 可以发现单一 Shearlet 字典总体重建效果远好于单一 DCT 字典。这是因为 Shearlet 变换是基于具有合成膨胀的仿射系统构建的新型多尺度几何变换,用于对图像等高维数据进行更为稀疏的表示。相比于 DCT 变换以及其他稀疏变换,Shearlet 变换具有更简单的数学结构及更突出的稀疏表示能力。现阶段,综合时间成本与重建效果的双重因素,在恢复实际资料时 Shearlet 变换是非常实用的选择。从图 5 可以看出无论哪一种采样方法,Shearlet 变换都可以取得良好效果,但观察图 5d 会发现由于完全随机采样的缺点,在连续多道缺失的地方仍然会存在部分误差,相比之下 Jitter 采样和分段随机采样可以很好地解决这一问题。

图 6 展示了 Shearlet 与 DCT 字典组合的重建数据,可以看出重建效果优于单一 Shearlet 字典和单一 DCT 字典。虽然误差最大绝对值略高于单一 DCT 字典误差最小绝对值(Shearlet+DCT 重建误差

绝对值 0.04;DCT 重建误差绝对值 0.03;Shearlet 重建误差绝对值 0.06),但误差集中于直达波区域,对后期处理无过大影响,而字典组合的中深层误差基本为 0。对比 3 种字典的重建结果也可看出图 6 所示结果的图像更为干净,同相轴更为连续清晰。对比图 6e、f 可知,字典组合对重建 Jitter 采样数据和分段随机采样数据的效果都很理想,但在对数据边界处理的方面来看,分段随机采样仍具有一定优势,两者差别见图中红圈部分。

基于压缩感知的数据重建不只是一个找回丢失数据的过程,同样也是一个提高信噪比的过程。表 1 展示了 3 种采样数据在使用不同稀疏字典重建前后的信噪比和峰值信噪比,两者的计算方法在 3.1 节给出。从表中可以看出本文提出方法(分段随机采样与 Shearlet+DCT 组合)的重建结果无论信噪比和峰值信噪比均有最大的提升。

不同的拉格朗日乘子数值对恢复重建的效果不同。从图 7 中可以看出选取拉格朗日乘子为 0.95 时,重建后信噪比较 0.90 时具有更高值,图 7a、b 的迭代次数分别是 100 和 200。迭代次数越多,重建结果信噪比越高,但也要承担相应的时间成本。在本文的实验中,迭代次数选 120 次时,已经能够达到

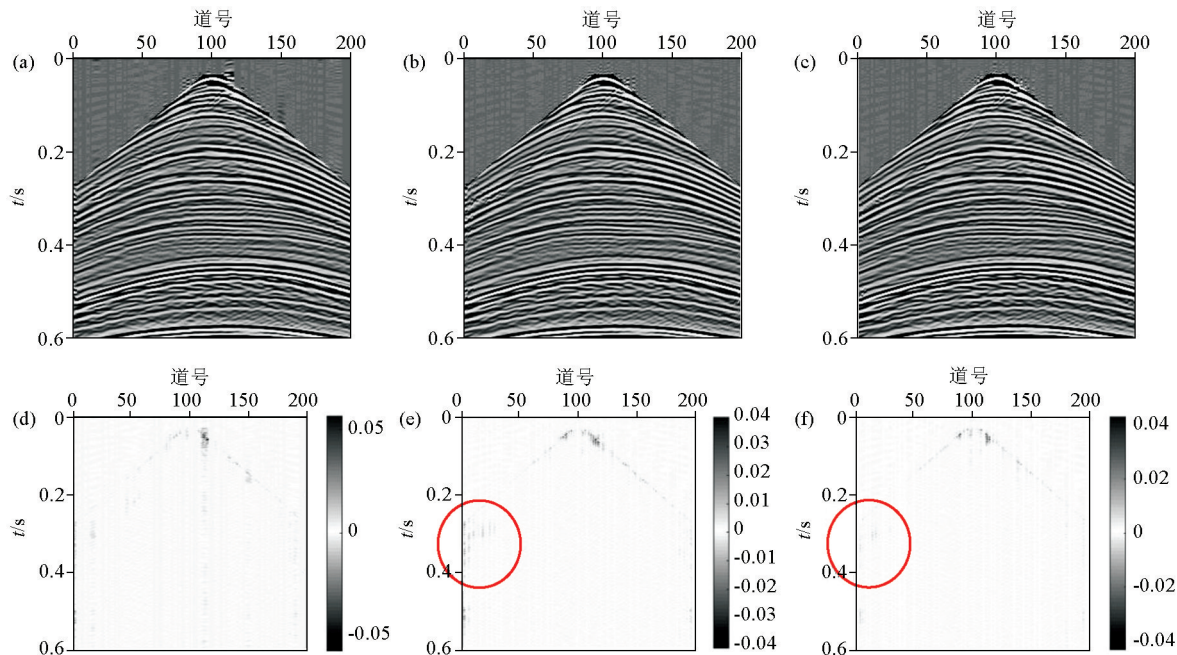


a—随机采样的 Shearlet 重建结果;b—Jitter 采样的 Shearlet 重建结果;c—分段随机采样的 Shearlet 重建结果;d—随机采样的 Shearlet 差值剖面;e—Jitter 采样的 Shearlet 差值剖面;f—分段随机采样的 Shearlet 差值剖面

a—Shearlet reconstruction result of random sampling;b—Shearlet reconstruction result of Jitter sampling;c—Shearlet reconstruction result of segmented random sampling;d—Shearlet reconstruction result of segmented random sampling;e—Shearlet difference profile of Jitter sampling;f—Shearlet difference profile of segmented random sampling

图 5 Shearlet 字典重建数据与差值剖面

Fig. 5 Reconstruction results and differences profile of Shearlet dictionary



a—随机采样的 Shearlet+DCT 重建结果;b—Jitter 采样的 Shearlet+DCT 重建结果;c—分段随机采样的 Shearlet+DCT 重建结果;d—随机采样的 Shearlet+DCT 差值剖面;e—Jitter 采样的 Shearlet+DCT 差值剖面;f—分段随机采样的 Shearlet+DCT 差值剖面

a—Shearlet+DCT reconstruction result of random sampling;b—Shearlet+DCT reconstruction result of Jitter sampling;c—Shearlet+DCT reconstruction result of segmented random sampling;d—Shearlet+DCT reconstruction result of segmented random sampling;e—Shearlet+DCT difference profile of Jitter sampling;f—Shearlet+DCT difference profile of segmented random sampling

图 6 字典组合重建结果及差值剖面

Fig. 6 Reconstruction results and difference profiles of Dictionary Combination

表 1 重建结果评价参数

Table 1 Evaluation parameters of reconstruction results

		缺失数据	DCT	Shearlet	DCT+Shearlet
完全随机采样	SNR/dB	2. 9288	-0. 5554	0. 5201	2. 6251
	PSNR/dB	28. 5367	24. 5367	25. 6122	27. 7172
Jitter 随机采样	SNR/dB	3. 3736	-0. 3296	7. 2277	10. 3023
	PSNR/dB	28. 4657	24. 7625	30. 0910	31. 3944
分段随机采样	SNR/dB	3. 0158	-0. 5368	8. 6044	12. 0367
	PSNR/dB	28. 1079	25. 5553	30. 6965	32. 1288

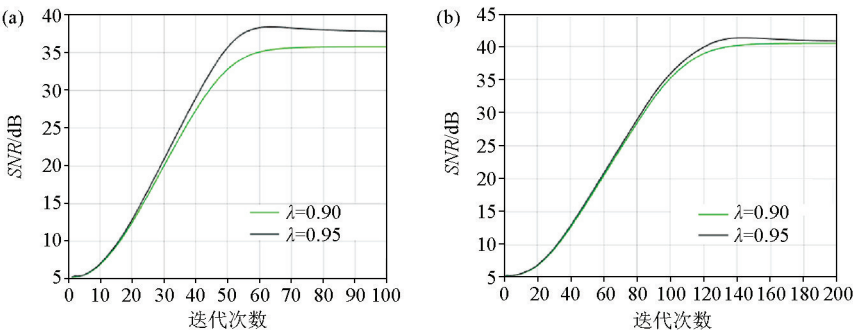


图 7 拉格朗日乘子及迭代次数与重建信噪比

Fig. 7 Lagrangian multiplier and number of iterations and reconstruction signal-to-noise ratio

很好的重建效果。

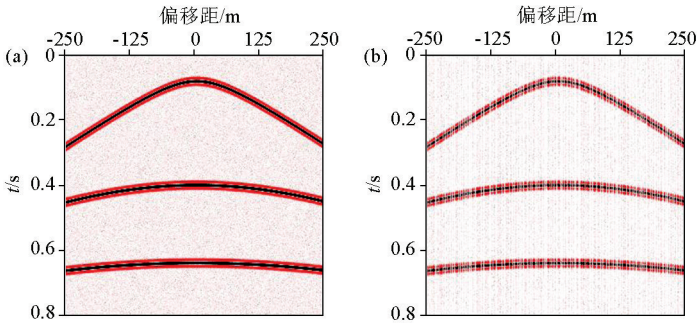
3.3 含噪模型试算

为更深地探究地震数据重建对信噪比的影响,对含噪合成地震数据进行欠采样及重建实验。为突出所提方法的优势,选择 DCT、Shearlet、Shearlet + DCT 进行实验,并对比 3 种方法的重建效果及定量评价价值。含噪合成地震记录共 500 道,每道含 800 个时间采样点。图 8 为数据的分段随机采集过程,图 8a、b 分别展示了原始数据及分段随机欠采样数据。图 9 展示了 3 种重建方法的重建结果,从图中可以看出本文所提方法(Shearlet+DCT 字典)的重建数据具有更少的噪声干扰。从表 2 中可以看出本文所提方法对原始数据的信噪比具有明显提升作用,

且提升量高于单一 DCT 字典与单一 Shearlet 字典。

3.4 连续缺失模型试算

以上采样方法基于压缩感知技术,压缩感知的理论基础是随机采样方式。但在实际生产中,由于实际工区地形限制,经常会出现地震道连续缺失的情况,难以满足随机采样的假设,因此我们添加复杂实际地震图像连续缺失重建实验。本节不设置对比实验,仅验证所提方法重建连续缺失地震数据的有效性。图 10 显示了分层介质的局部图像。为不失一般性,本文依次从原始图像中提取出 11 道和 22 道,然后使用所提出的方法进行重建。连续缺失数据、重建结果和重建误差分别如图 11 和图 12 所示。可以看出,两种情况下都能获得较好的重建结果,

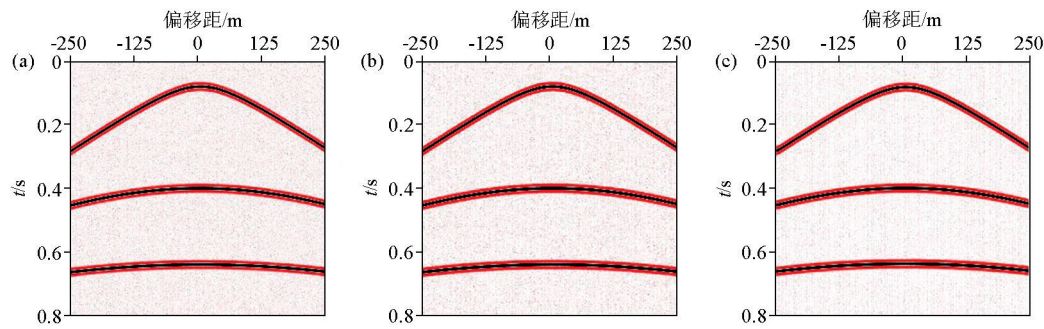


a—原始含噪数据;b—分段随机采样 50% 数据

a—original noise data;b—50% missing data

图 8 含噪合成地震数据欠采样

Fig. 8 Under sampling of noise synthetic seismic data



a—DCT 重建结果;b—Shearlet 重建结果;c—Shearlet+DCT 重建结果
a—DCT reconstruction result;b—Shearlet reconstruction result;c—Shearlet+DCT reconstruction result

图 9 三种重建结果

Fig. 9 Three reconstruction results

表 2 重建结果评价参数

Table 2 Evaluation parameters of reconstruction results

原始数据	欠采样数据	方法	重建数据	变换值
SNR=3. 0102dB PSNR=19. 8125dB	SNR=2. 246dB PSNR=17. 4515dB	DCT	SNR=13. 1747dB PSNR=29. 7691dB	+10. 1645
		Shearlet	SNR=15. 6185dB PSNR=30. 0910dB	+12. 6083
		Shearlet+DCT	SNR=17. 1428dB PSNR=33. 945dB	+14. 1326

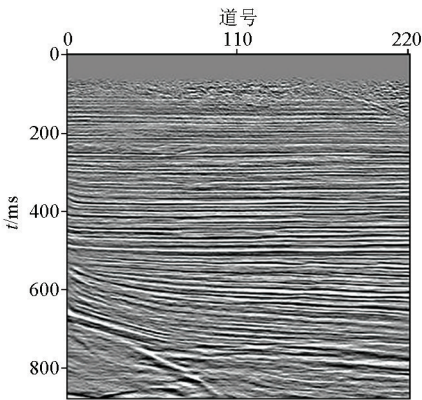


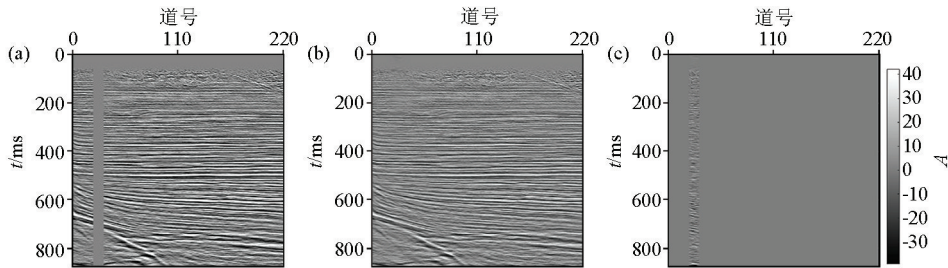
图 10 原始剖面
Fig. 10 Original profile

说明所提方法具有一定的空间表示能力,能够重建有限范围的连续缺失道。

4 结论及讨论

由于缺道坏道等问题,地震数据往往是不完整的。从反演的角度看,不完整图像重建是一个病态反问题。CS 技术利用较少的随机采样数据重构出完整信号,其主要思想是利用信号在某一变换域的稀疏性。

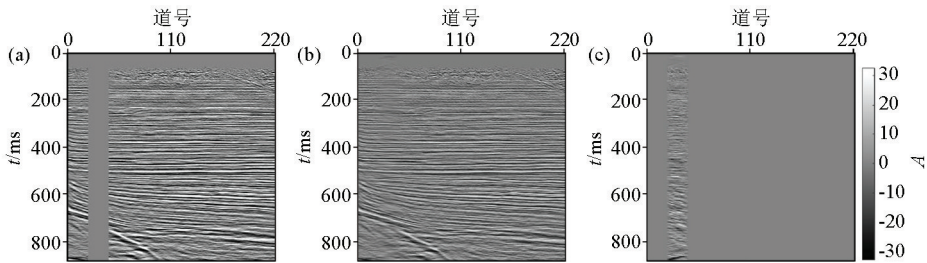
本文在 MCA 框架下提出了一种新的字典组合: Shearlet+DCT。Shearlet 变换是多尺度、多方向的,可



a—连续缺失 11 道图像;b—重建结果;c—绝对误差
a—11 consecutive missing images;b—reconstruction result;c—absolute error

图 11 连续缺失 11 道 Shearlet+DCT 重建

Fig. 11 Shearlet+DCT reconstruction of 11 consecutive missing



a—连续缺失 22 道图像;b—重建结果;c—绝对误差
a—22 consecutive missing images;b—reconstruction result;c—absolute error

图 12 连续缺失 22 道 Shearlet+DCT 重建

Fig. 12 Shearlet+DCT reconstruction of 22 consecutive missing

以得到更稀疏的系数,DCT 是一种用于重建平滑构造的全局型变换。因此,Shearlet+DCT 得到的信号系数更加稀疏,无论对局部细节还是整体图像趋势都有很好的重建效果。通过数值实验对 Shearlet、DCT 和 Shearlet+DCT 这 3 种方法进行了比较。结果表明,本文提出方法的重建结果的相对误差较小,可以很好地重建局部细节和深弱反射信号。但随着数据的复杂度的提升,数学变换的局限性会逐渐暴露出来。我们现在也在使用学习算法来训练自适应基函数,以获得更好的结果。

本文还证明了分段随机采样比 Jitter 采样更有利于数据重建。CS 的理论基础是随机采样,但在实际应用中,地震道往往在大范围内连续缺失,难以满足随机采样的假设。对于有限数量的连续缺失案例,本文提出的重建方法也是有效的。但对于大连连续缺失区域,例如超过 22 道连续缺失,重建结果并不是很理想。经反复实验,我们得出对于复杂实际资料来说,压缩感知可以接受原始数据 10% 以内的连续缺失,这一范围随数据量的大小、数据复杂程度及层位密集程度的整加而减小,随三者的降低而增加。因此,下一步的研究重点将是开发大连连续缺失信号的空间表达,利用有限的信息进行重建并开发更好的重建方法,如字典学习或机器学习重建方法等。

参考文献 (References):

[1] 张华,陈小宏. 基于 Jitter 采样和曲波变换的三维地震数据重建[J]. 地球物理学报,2013,56(5):1637-1649.
Zhang H,Chen X H. Seismic data reconstruction based on jittered sampling and curvelet transform [J]. Chinese J. Geophys.,2013,56(5):1637-1649.

[2] 唐刚. 基于压缩感知和稀疏表示的地震数据重建与去噪[D]. 北京:清华大学,2010.
Tang G. Seismic data reconstruction and denoising based on compressive sensing and sparse representation [D]. Beijing:Tsinghua University, 2010.

[3] Leneman O. Random sampling of random processes: Impulse response [J]. Information and Control,1966,9(2):347-363.

[4] Hennenfent G, Hermann F J. Simply denoise: Wavefield Reconstruction via jittered undersampling [J]. Geophysics, 2008, 73(3): V19-V28.

[5] Hermann F J,Wang D,Hennenfent G,et al. Curvelet-based seismic data processing: A multiscale and nonlinear approach [J]. Geophysics, 2008, 73(1): A1-A5.

[6] Mosher C C. Generalized windowed transforms for seismic processing and imaging[C]//82nd Annual International Meeting Expanded Abstracts,SEG,2012.

[7] Yang P,Fomel S. Seislet-based morphological component analysis using scale-dependent exponential shrinkage [J]. Journal of Applied Geophysics,2015,118:66-74.

[8] Men Z,Ning H X,Zhang M G,et al. A method and application of irregular geometry design based on compressive sensing [C]//SEG Technical Program Expanded Abstracts,2019.

[9] Sardy S,Bruce A G,Tseng P. Block coordinate relaxation methods for nonparametric wavelet denoising [J]. Journal of Computational and Graphical Statistics,2000,9(2):361-379.

[10] 刘成明,王德利,王通,等. 基于 Shearlet 变换的地震随机噪声压制[J]. 石油学报,2014,35(4):692-699.
Liu C M,Wang D L,Wang T,et al. Random seismic noise attenuation based on the Shearlet transform [J]. Acta Petrolei Sinica, 2014,35(4): 692-699.

[11] 李海山,吴国忱,印兴耀. 形态分量分析在地震数据重建中的应用[J]. 石油地球物理勘探,2012,47(2):236-243.
Li H S,Wu G C,Yin X Y. Morphological component analysis in seismic data reconstruction [J]. Oil Geophysical Prospecting, 2012,47(2):236-243.

[12] 周亚同,刘志峰,张志伟. 形态分量分析框架下基于 DCT 与曲波字典组合的地震信号重建[J]. 石油物探,2015,54(5):560-568.
Zhou Y T,Liu Z F,Zhang Z W. Seismic signal reconstruction under the morphological component analysis framework combined with discrete cosine transform (DCT) and curvelet dictionary [J]. Geophysical Prospecting for Petroleum,2015,54(5):560-568.

[13] 张良,韩立国,许德鑫,等. 基于压缩感知技术的 Shearlet 变换重建地震数据[J]. 石油地球物理勘探,2017,52(2):220-225.
Zhang L,Han L G,Xu D X,et al. Seismic data reconstruction with

Shearlet transform based on compressed sensing technology [J]. Oil Geophysical Prospecting, 2017, 52(2): 220–225.

[14] 徐卫, 张华, 张落毅. 基于复值曲波变换的地震数据重建方法[J]. 物探与化探, 2016, 40(4): 750–756.

Xu W, Zhang H, Zhang L Y. A study of seismic data reconstruction based on complex-valued curvelet transform [J]. Geophysical and Geochemical Exploration, 2016, 40(4): 750–756.

[15] 石战战, 夏艳晴, 周怀来, 等. 一种基于 L_1 - L_1 范数稀疏表示的地震反演方法[J]. 物探与化探, 2019, 43(4): 851–858.

Shi Z Z, Xia Y Q, Zhou H L, et al. Seismic reflectivity inversion based on L_1 - L_1 -norm sparse representation [J]. Geophysical and Geochemical Exploration, 2019, 43(4): 851–858.

[16] 孔旭, 密文天, 莫雄, 等. 基于 MRAS 证据权重法的湖南怀化地区金矿成矿预测[J]. 物探与化探, 2016, 40(3): 467–474.

Kong X, Mi W T, Mo X, et al. Metallogenic prediction of gold deposits with weighting of evidence based on MRAS in Huaihua area, Hunan Province [J]. Geophysical and Geochemical Exploration, 2016, 40(3): 467–474.

[17] Wu R S, Geng Y, Ye L. Preliminary study on Dreamlet based compressive sensing data recovery [C]//SEG Technical Program Expanded Abstracts, 2013.

[18] 何真, 曹思远, 郝姍姍, 等. 基于自适应 K-SVD 的能量泄漏恢复研究[J]. 物探与化探, 2020, 44(2): 362–371.

He Z, Cao S Y, Hao H J, et al. Research on energy leakage recovery of adaptive K-SVD [J]. Geophysical and Geochemical Exploration, 2020, 44(2): 362–371.

[19] 张凯, 张医奎, 李振春, 等. MCA 框架下 Shearlet 和 DCT 字典组合地震数据重建[J]. 石油地球物理勘探, 2019, 54(5): 12.

Zhang K, Zhang Y K, Li Z C, et al. Seismic data reconstruction method combined with Discrete Cosine Transform and Shearlet dictionary under Morphological Component Analysis framework [J]. Oil Geophysical Prospecting, 2019, 54(5): 12.

[20] Li X, Guo M J, Li W H, et al. Sparsity promoting reconstruction with compressively acquired land data [C]//SEG Technical Program Expanded Abstracts, 2019.

[21] Zwartjes P, Gisolf A. Fourier reconstruction with sparse inversion [J]. Geophysical Prospecting, 2007, 5(2): 199–221.

[22] 郭奇, 曾昭发, 于晨霞, 等. 基于高精度字典学习算法的地震随机噪声压制[J]. 物探与化探, 2017, 41(5): 907–913.

Guo Q, Zeng Z F, Yu C X, et al. Seismic random noise suppression based on the high-precision dictionary learning algorithm [J]. Geophysical and Geochemical Exploration, 2017, 41(5): 907–913.

[23] Kumar R, Wason H, Herrmann F J. Source separation for simultaneous towed-streamer marine acquisition—A compressed sensing approach [J]. Geophysics, 2015, 80(6): WD73–WD88.

[24] Mosher C C, Kaplan S T, Janiszewski F D. Non-uniform optimal sampling for seismic survey design [C]//74th EAGE Conference and Exhibition, 2012.

[25] Neelamani R, Baumstein A, Gillard D, et al. Coherent and random noise attenuation using the curvelet transform [J]. The Leading Edge, 2008, 27: 240–248.

[26] 罗勇, 毛海波, 杨晓海, 等. 基于双重稀疏表示的地震资料随机噪声衰减方法[J]. 物探与化探, 2018, 42(3): 608–615.

Luo Y, Mao H B, Yang X H, et al. Seismic random seismic noise attenuation method on basis of the double sparse representation [J]. Geophysical and Geochemical Exploration, 2018, 42(3): 608–615.

Seismic data reconstruction based on segmented random sampling and MCA

WANG De-Ying¹, ZHANG Kai², LI Zhen-Chun², ZHANG Yi-Kui³, XU Xin¹

(1. Research Institute of Petroleum Exploration & Development-Northwest (NWGI), PetroChina, Lanzhou 730030, China; 2. School of Geosciences, China University of Petroleum, Qingdao 266580, China; 3. Wuhua Energy Technology Co., Ltd, Xi'an 710067, China)

Abstract: Data reconstruction is a critical preliminary work in the processing of seismic data. Compressed sensing (CS) has been well applied in data reconstruction. The key to CS is random sampling, which converts the mutual coherent alias caused by regular under-sampling into lower-amplitude incoherent noise. But traditional sampling methods lack constraints on sampling points, resulting in excessive noise interference. The segmented random sampling (SRS) method can effectively control the distance between sampling points. Furthermore, a single mathematical transformation will lead to incomplete sparse representation and impact data reconstruction. The morphological component analysis (MCA) can decompose a signal into several components with outstanding morphological features to approximate the complex internal structure of data. A new dictionary combination (Shearlet+DCT) has been found under the MCA framework, and the block coordinate relaxation (BCR) algorithm has been used to get the optimal solution to obtain desired reconstruction results. Tests of real data have proven that the proposed method can produce good effects when used to reconstruct the SRS data.

Key words: compressed sensing; segmented random sampling; morphological component analysis; data reconstruction

(本文编辑:叶佩)