

基于固定点算法的地震数据降噪

张晓峰¹, 许江浩²

1. 成都理工大学沉积地质研究院, 成都 610059

2. 成都理工大学; 数学地质四川省高校重点实验室, 成都 610059

摘要 结合改进的固定点算法, 解决了噪声环境下的 ICA 问题。根据噪声分布特性, 分两个阶段去除不同类型的随机噪声。在预处理阶段去除了加性高斯白噪声, 预处理后的数据采用改进的固定点算法, 盲分离出有效信号和非高斯随机噪声。提出了对固定点算法迭代过程中设定较精确的初始值问题的算法, 该方法能较为准确地设置初始值, 使算法能提取有效信号。通过仿真实验和对实际地震数据的处理, 得到了满意的分离结果, 较好地恢复了有效信号。此外, 当实际地震数据加载了较强噪声, 信噪比降低时, 采用本文算法进行盲分离, 同样取得了良好的效果, 再次验证了本文算法具有良好的稳健性和适应性。将盲分离算法应用到实际地震数据处理方面的研究, 有助于地震资料的解释, 同时这种处理技术的研究也能够促进盲分离技术的发展。

关键词 地震资料; 随机噪声; 固定点算法

中图分类号 TU352

文献标识码 A

doi 10.3981/j.issn.1000-7857.2011.16.006

Method of Denoising Seismic Random Data Based on Improved FastICA

ZHANG Xiaofeng¹, XU Jianghao²

1. Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China

2. Key Laboratory of Geomathematics of Sichuan Province; Chengdu University of Technology, Chengdu 610059, China

Abstract An improved fixed-point algorithm is used to solve the ICA problem accompanied with noise. According to the noise distribution, it takes two phases to eliminate random noises of different types in preprocessing. The additive white Gaussian noise is removed at first, then the Improved FastICA algorithm is used to process the preprocessed data and to blindly separate the effective signal from non-Gaussian random noise. It might be a problem to set a good starting value in the iterative process of Improved FastICA. In this way, one can accurately set the starting value to make the algorithm recover the effective signal. The satisfactory separation results and better recovery of the effective signal are achieved as shown by the simulation experiments and real seismic data processing. Furthermore, in the case of the strong seismic noise with actually loaded and reduced SNR, this algorithm of blind separation also produces good results. This verifies that the algorithm has good robustness and adaptability. Using the algorithm of blind separation to do the seismic data enhancement can help to better interpret the seismic data, and promote the development of blind separation technology.

Keywords seismic exploration data; random noise; FastICA

0 引言

盲信号分离技术是当今现代数学和信号领域研究的一个重要热点。盲源分离(合成信号中找出不同的信号)早在 20 世纪中期就已提出, Tong 和 Liu^[1]分析了盲源分离问题的可辨识性和不确定性。Cardoso 于 1993 年提出基于高阶统计的联

合对角化盲源分离方法, 并应用于波束形成^[2]。1995 年 Bell 和 Sejnowski^[3]提出了基于熵最大思想的盲源分离方法。Amari^[4]于 1998 年提出了自然梯度算法, 大大加快了算法的收敛速度。随着盲源分离的深入研究, 研究范围不断拓展, 一些学者开始研究相关源信号的盲分离问题^[5-6], 提出了时-频分析方法。

收稿日期: 2011-02-28; 修回日期: 2011-05-19

基金项目: 四川省教育厅自然科学重点项目(08ZA105)

作者简介: 张晓峰, 博士研究生, 研究方向为沉积与储层, 电子信箱: zxf@cdu.edu.cn

2004 年 Yuan 和 ErkkiOja 提出了一种非独立分量分析的 FastICA 算法(也称为固定点算法)^[7]。中国科学院的惠晓宇^[8]于 2006 年对盲信号分离技术在地学中的应用做了论述和展望。近年来,状态空间法逐步发展,它从控制工程的角度重新描述混合矩阵和解混合矩阵,转换了盲源分离的研究思路。在未知信号源个数的情况下有效提取出源信号,盲信号提取技术得到了较快的发展。本文结合改进的 FastICA 算法,解决了噪声环境下的 ICA 问题,提出了对 FastICA 算法迭代过程中设定较精确的初始值问题的算法。将盲分离算法应用到实际地震数据处理方面的研究,有助于地震资料的解释,并且这种处理技术的研究也能够促进盲分离技术的发展。

1 信噪盲分离算法

FastICA 算法基于非高斯性最大化原理,使用固定点(fixed-point)迭代理论寻找 $\mathbf{w}^T \mathbf{x}$ 的非高斯最大值,该算法采用牛顿迭代算法对观测变量 $\mathbf{x}$ 的大量采样点进行批处理。该算法的非高斯型度量函数为^[9]

$$J(\mathbf{y}) \propto [E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\}]^2 \quad (1)$$

其中, $\propto$ 表示正比, $E(\cdot)$ 为期望, $G(\cdot)$ 为一个非线性函数, $\mathbf{v}$ 为与 $\mathbf{y}$ 具有同样方差的零均值高斯变量。

为了减少算法需要估计的参数,简化算法的计算,在运行 FastICA 算法之前,假定已经进行了 ICA 预处理。

1.1 目标函数负熵的计算

非高斯性度量的重要方法是估计负熵。由于负熵难于直接计算,实际中使用负熵的近似计算。

本文使用一个非二次函数 $G(\cdot)$ 计算负熵近似^[10]:

$$J(\mathbf{y}) \approx \sum_{i=1}^n k_i [E\{G_i(\mathbf{y})\} - E\{G_i(\mathbf{v})\}] \quad (2)$$

式中 k_i 为一正的常量。此处如果函数 $G(\cdot)$ 不会增长太快就可以得到鲁棒性较好的负熵估计。

1.2 FastICA 算法原理

常用的最大化目标函数的方法是随机梯度搜索法。这实质是一种自适应算法,其优点是能在不稳定的情况下进行快速的调节,缺点是收敛速度慢,需要选定好迭代步长,如果迭代步长选择错误会使算法失去收敛性。由于在实际计算中,常常很重视算法的收敛速度和稳定性,所以在 FastICA 算法中没有使用随机梯度搜索法而采用了固定点迭代法。在该算法的每次迭代中,采样数据是成批使用的。该算法为并行分布式,计算简单,需要内存较少,具有很好的收敛性,在不需要快速自适应的情况下是最好的选择。

基于非高斯性最大化的独立分量分析算法判断分量之间是否独立等价于求取各分量之间的非高斯性的最大值。对于非高斯性最大化方法中, FastICA 算法利用基于公式(1)的负熵表达式估计独立分量。为了找出第 i 个独立分量,或求出 $\mathbf{y}_i = \mathbf{w}_i^T \mathbf{x}$ 的投影方向,使式(1)的值最大化:

$$J_G(\mathbf{w}) = [E\{G(\mathbf{w}^T \mathbf{x})\} - E\{G(\mathbf{v})\}]^2 \quad (3)$$

其中, $\mathbf{w}$ 为 m 维变量。由于 $\mathbf{v}$ 和 $\mathbf{x}$ 具有相同均值和协方差矩

阵的高斯变量,在预处理后,式(3)的最大化问题可以转化为 $E\{G(\mathbf{w}^T \mathbf{x})\}$ 的优化问题。按照 Kuhn-Tucker 条件, $E\{G(\mathbf{w}^T \mathbf{x})\}$ 的优化问题在条件 $E\{(\mathbf{w}^T \mathbf{x})^2\} = \|\mathbf{w}\|^2 = 1$ 下可以通过求解得到牛顿迭代法的优化迭代为^[10]

$$\mathbf{w}_{k+1} = \frac{\mathbf{w}_k - [E\{\mathbf{x}G(\mathbf{w}_k^T \mathbf{x})\} - \beta \mathbf{w}_k]}{E\{G'(\mathbf{w}_k^T \mathbf{x})\} - \beta} \quad (4)$$

式中 $\beta = E\{\mathbf{w}_k^T \mathbf{x} G(\mathbf{w}_k^T \mathbf{x})\}$ 。

为提高算法的稳定性,在每次迭代后,对 $\mathbf{w}_{k+1}$ 进行归一化处理:

$$\mathbf{w}_{k+1} = \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$$

将式(4)两边乘以 $\beta - E\{G'(\mathbf{w}_k^T \mathbf{x})\}$, 得到算法的迭代公式

$$\mathbf{w}_{k+1} = E\{\mathbf{x}G(\mathbf{w}_k^T \mathbf{x})\} - E\{G'(\mathbf{w}_k^T \mathbf{x})\} \mathbf{w}_k \quad (5)$$

1.3 FastICA 算法实现步骤

利用式(5)可以得到固定点迭代算法的基本形式:

(1) 随机选择初始权值矢量 $\mathbf{w}_0, k=0$;

(2) 利用式(5)更新权值矢量 $\mathbf{w}_{k+1}$;

(3) 归一化 $\mathbf{w}_{k+1} = \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$;

(4) 如果 $\|\mathbf{w}_{k+1} - \mathbf{w}_k\| > \varepsilon$, 算法不收敛,返回(2),否则 FastICA 算法估计出一个独立分量,算法结束。

对于多分量的提取,重复使用 FastICA 算法的基本形式进行多次分离即可。为了确信每次提出不同的分量,在每次提取一个分量后,从观测信号中去除提取出来的独立分量,重复执行,直到所有需要的独立分量提取出来为止。去除已经提取的独立分量的方法利用式(6)实现:

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_{k+1} - \sum_{j=1}^k \mathbf{w}_{k+1}^T \mathbf{w}_j \mathbf{w}_j \\ \mathbf{w}_{k+1} &= \mathbf{w}_{k+1} / \sqrt{\mathbf{w}_{k+1}^T \mathbf{w}_{k+1}} \end{aligned} \quad (6)$$

2 对 FastICA 算法的改进

两个信号 $\mathbf{x}$ 与 $\mathbf{y}$ 的相似性是指 $\mathbf{x} \approx \lambda \mathbf{y}$ (λ 为参数, $\mathbf{x}$ 为比较的基准,近似相等“ $\approx$ ”是指能量误差 $Q = \|\mathbf{x} - \lambda \mathbf{y}\|^2$ 较小)。当设参数 λ_0 使 $\mathbf{x} \approx \lambda_0 \mathbf{y}$ 成立,即能量误差为

$$Q = \|\mathbf{x} - \lambda \mathbf{y}\|^2 = \sum (\xi_n - \lambda_0 \eta_n)^2$$

为了使能量误差达到最小值,令 $dQ/d\lambda_0 = 0$, 于是得

$$\lambda_0 = \frac{\sum \xi_n \eta_n}{\sum \eta_n^2}$$

此时,对应的最小能量误差为

$$\begin{aligned} Q_{\min} &= \sum (\xi_n - \lambda_0 \eta_n)^2 \\ &= \sum \xi_n^2 - \frac{\left(\sum \xi_n \eta_n\right)^2}{\sum \eta_n^2} \end{aligned}$$

最小相对能量误差为

$$\frac{Q_{\min}}{\|x\|} = 1 - \frac{(\sum \xi_n \eta_n)^2}{\sum \xi_n^2 \sum \eta_n^2} = 1 - R_{xy}^2$$

其中

$$R_{xy} = \frac{\sum \xi_n \eta_n}{\sqrt{\sum \xi_n^2 \sum \eta_n^2}}$$

根据内积定义和 Cauchy-Schwarz 不等式可知

$$|R_{xy}| = \frac{\sum \xi_n \eta_n}{\sqrt{\sum \xi_n^2 \sum \eta_n^2}} \leq 1$$

且 $|R_{xy}|$ 越大, $Q_{\min}$ 越小, 表明信号 x, y 越相似。

根据上述的推导, 有如下结论: 设 $x = [\xi_1, \xi_2, \dots]$ 和 $y = [\eta_1, \eta_2, \dots]$ 为两个离散信号, 则波形相似系数就是 R_{xy} , 它表征 x, y 的相似性, 且 $|R_{xy}|$ 越大, 信号 x, y 越相似。这样, 可用 R_{xy} 衡量两个信号的相似性大小, 在每次盲分离后输出两个信号, 计算它们和输入信号的相似性大小, 即可判断哪个是有效信号, 哪个是噪声, 从而提取出有效信号。

3 改进的 FastICA 算法实现步骤

地震记录中包含着随机干扰信号。地震数据的负熵越大, 说明地震信号的不相关性越高。若设地震记录中的有效信号和干扰信号在统计上独立, 且服从非高斯分布, 那么根据独立分量分析的思想方法, 只需知道该道记录的两次观测结果, 或者近似取近邻的两道记录, 就可采用盲分离算法直接从地震道中分离出其中包含的有效信号和随机干扰, 从而达到去噪的目的。使用前述改进型 FastICA 算法进行地震勘探资料去噪方法的具体算法如下(图 1)。

(1) 输入观测信号, 相邻两道(1 和 2 道)地震数据, 输入 x_1 和 x_2 为抽取的邻近两道地震记录, 它们均是地震信号和随机噪声的混合。

(2) 对上述两道观测数据进行预处理, 将有效波与噪声都看作是源信号一部分, 根据盲分离应用的先决条件对观测的地震信号进行稳健的预处理。去均值后, 再利用稳健预白化方法对数据进行白化处理, 去除加性高斯白噪声。

(3) 利用适应地震勘探信号的 FastICA 算法对预处理后的数据进行处理, 根据建立的盲分离数学模型, 对实际的反射地震数据进行盲分离处理, 从而提高地震数据的信噪比, 盲分离出有效信号和非高斯分布的随机噪声。

(4) 根据信号的相似系数, 计算分离的两道信号与输入信号相似性的大小, 相似系数大的即为有效信号。

(5) 顺序保存有效信号到数据文件。

(6) 再依次输入 3 和 4 道, 返回(2)循环, 直到地震数据所有道处理完毕。

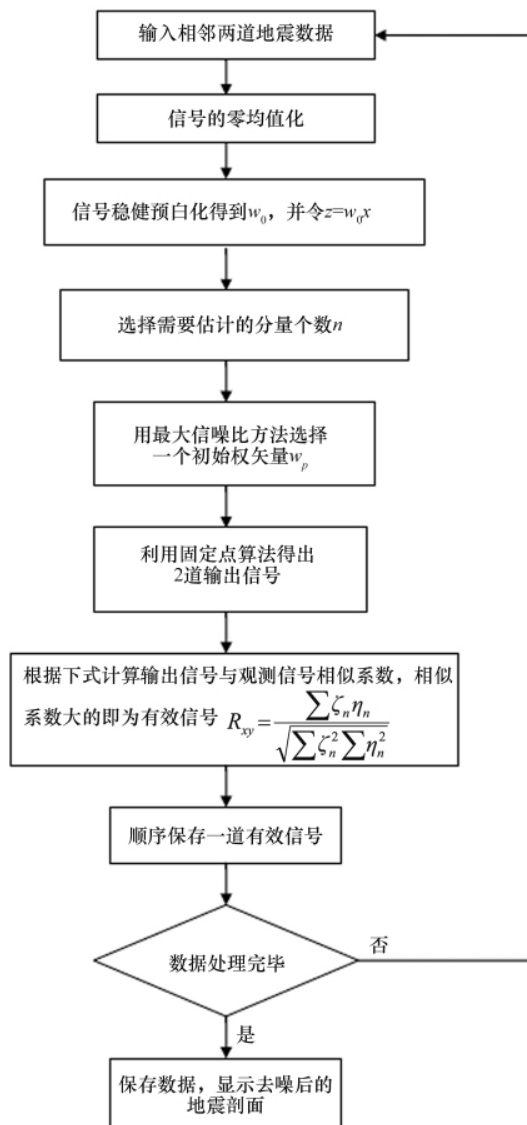


图 1 随机噪声盲分离处理流程

Fig.1 Flowchart of blind separation of random noise

4 人工合成地震记录去噪实验

利用人工合成的 Richer 子波的共炮点记录进行试验。设地震子波是 Richer 子波, 偏移距为 0, $t_0=0.4s$, 谱峰频率 $f_p=19Hz$, 不考虑距离衰减, 道间距 $d=50cm$, 道数 $N=40$, 叠加速度为 $v=1500m/s$ 。当对 Richer 子波和合成地震记录加噪时, 利用本文提出的盲分离算法进行处理的结果如图 2—图 5 所示。

通过仿真结果可以发现, 利用本文提出的算法去噪后, 分离得到的有效波基本上完好地保留下来, 从而初步说明了本文所提出的算法的合理性和有效性。

模拟两道相邻地震信号, 对这两道分别添加高斯白噪声(噪声幅度分布服从高斯分布, 功率谱密度为均匀分布), 使这两道地震信号的信噪比分别为 8dB 和 10dB。然后利用本文的盲分离算法对加噪后的两道信号处理。分离结果如图 6 所示。

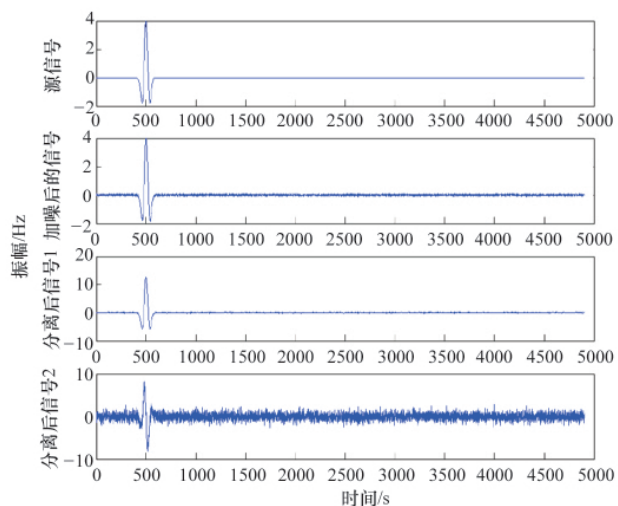


图2 Richer子波加噪后盲分离结果

Fig. 2 Result of blind separation of Richer wavelet added noise

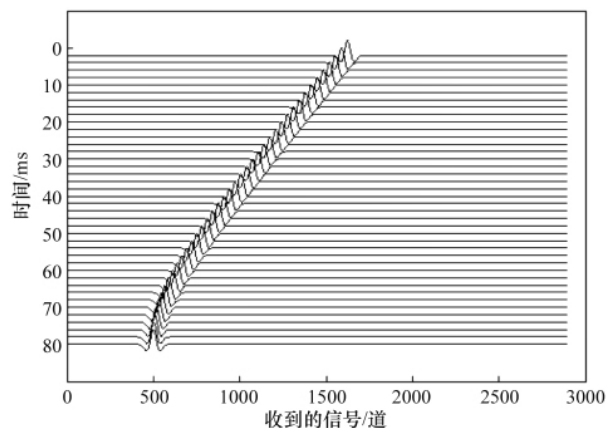


图3 不含噪声的人工合成地震记录

Fig. 3 Synthetic seismic records without noise

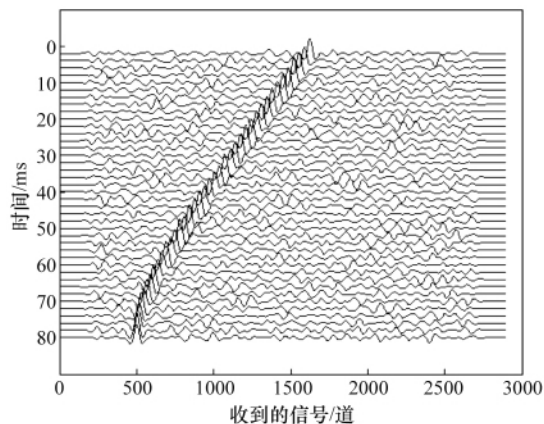


图4 含随机噪声的合成地震记录

Fig. 4 Synthetic seismic records with noise

由图6可以看出,原始信号的波形比较光滑,两道相邻地震信号被加入高斯噪声后,细节位置变得模糊,加噪

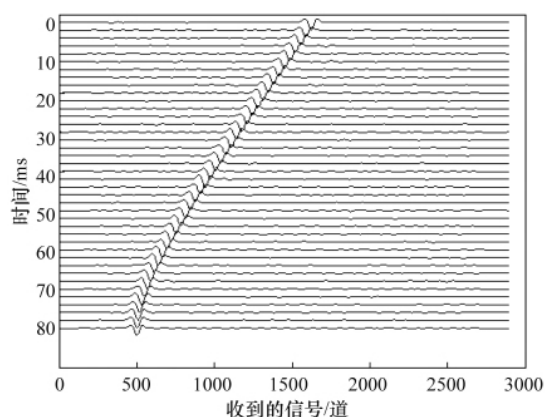


图5 合成地震记录盲分离去噪后的结果

Fig. 5 Result of blind separation of synthetic seismic records

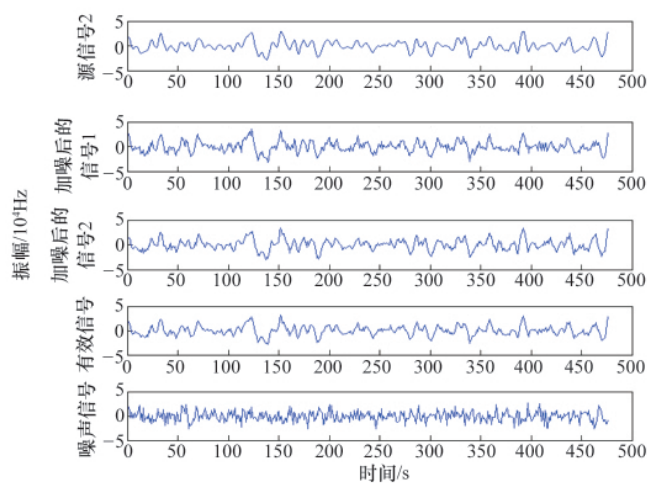


图6 仿真实验结果

Fig. 6 Result of simulation experiment

信号通过本文提出的盲分离算法进行处理后,信噪比提高,很好地恢复了源信号形态特征,从而说明了本文算法的有效性。

5 实际地震剖面去噪结果

利用一组二维叠后数据检验算法的效果。此地震数据是二进制格式,共有142道,每道476个采样点,采样率为2ms。每次输入相邻两道地震数据,经过分离后,提取出有效信号并保存,直到所有数据处理结束。对142道原始地震资料都添加了较强的随机噪声,其信噪比为5dB。原始剖面 and 去噪后的地震剖面显示如图7—图9所示。由图可知,盲分离算法可以很好地去除较强的随机噪声,信噪比由原来的5dB提高到13dB,去噪效果较好。而且从图9可以看出,本算法在有效去除的同时,分辨率也得到提高,验证了算法具有更广泛的适用范围和更强的稳健性。

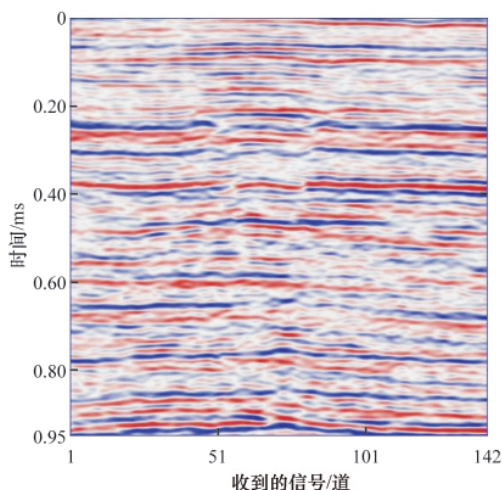


图7 原始地震剖面
Fig. 7 Raw seismic profile

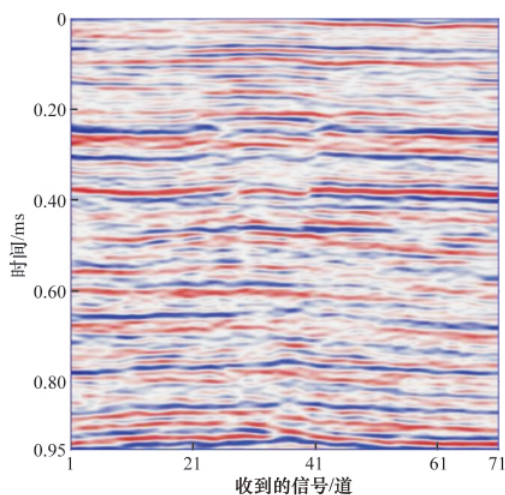


图8 FastICA 算法处理后的地震剖面
Fig. 8 Seismic profile processed by algorithm of FastICA

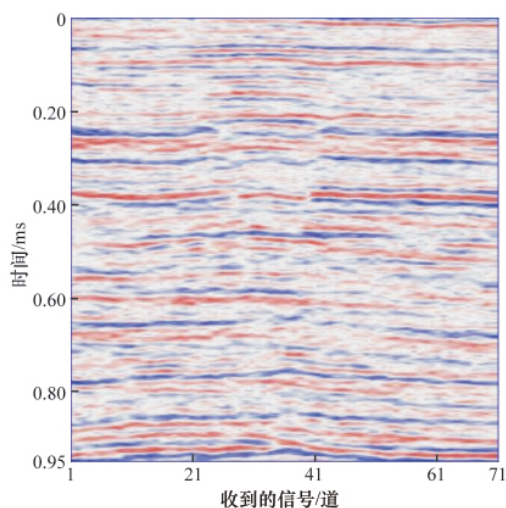


图9 改进后的 FastICA 算法处理结果
Fig. 9 Seismic profile processed by the improved algorithm

6 结论

本文提出了对 FastICA 算法迭代过程中设定较精确的初始值问题的算法,在实际地震数据加载了较强噪声,信噪比较低的情况下,采用本文算法进行盲分离,可以很好地去除较强随机噪声,且分辨率得到提高。将盲分离算法应用到实际地震数据处理方面的研究,有助于地震资料的解释,同时这种处理技术的研究也能够促进盲分离技术的发展。

参考文献 (References)

- [1] Tong L, Liu R, Soon V C, *et al.* Indeterminacy and identifiability of blind identification[J]. *IEEE Trans on Circuits and Systems*, 1991, 38(5): 499-506.
- [2] Cardoso J F. High-order contrasts for independent component analysis[J]. *Neural Computation*, 1999, 11(1): 157-192.
- [3] Bell M S, Jung A, Sejnowski T. Independent component analysis of electroencephalographic data [J]. *Advances in Neural Information Processing Systems*, 1996, 8(6): 145-151.
- [4] 杨福生, 洪波. 独立分量分析的原理与应用 [M]. 北京: 清华大学出版社, 2006.
Yang Fusheng, Hong Bo. Independent component analysis principle and application[M]. Beijing: Tsinghua University Press, 2006.
- [5] Amaris. Natural gradient works efficiently in learning [J]. *Neural Computation*, 1998, 10(2): 241-276.
- [6] 刘喜武, 刘洪, 李幼铭. 独立分量分析及其在地震信息处理中应用初探[J]. *地球物理学进展*, 2003, 18(1): 90-96.
Liu Xiwu, Liu Hong, Li Youming. *Progress in Geophysics*, 2003, 18(1): 90-96.
- [7] 张发启. 盲信号处理及应用 [M]. 西安: 西安电子科技大学出版社, 2006.
Zhang Faqi. Operation and application of blind signal [M]. Xi'an: Xi'an University of Electronic Science and Technology Press, 2006.
- [8] 惠晓宇, 刘洪. 主分量分析和独立分量分析方法的比较研究 [J]. *石油物探*, 2006, 45(5): 441-446.
Hui Xiaoyu, Liu Hong. *Geophysical prospecting for petroleum*, 2006, 45(5): 441-446.
- [9] 马建仓, 牛奕龙, 陈海洋. 盲信号处理 [M]. 北京: 国防工业出版社, 2006.
Ma Jiancang, Niu Yilong, Chen Haiyang. Blind signal processing[M]. Beijing: National Defense Industry Press, 2006.
- [10] Lu W, Mao F. Adaptive multiple subtraction using independent component analysis[J]. *The Leading Edge*, 2005, 24(3): 282-284.

(责任编辑 代丽)

《科技导报》“科技纵横捭阖”栏目征稿

“科技纵横捭阖”栏目收录对学术热点、前沿,学术争论、争端,科学与文化,科学人物介绍,海外科研、留学经历,科学史,科学渊源,科学决策、学术会议、科学活动,以及科研经费、科研项目申报、考试等方面的杂谈文章。每篇文章约 2200 字,要求求实、具体,行文深入浅出、言简意赅、逻辑清晰、有理有据、观点鲜明、切中要害,可读性强。栏目责任编辑:王芷,电子信箱:wangzhi@cast.org.cn;在线投稿:www.kjdb.org。