

# 基于 MP 方法的地震信号快速分解算法

陈发宇, 杨长春

(中国科学院地质与地球物理所, 北京 100029)

**摘要** Matching Pursuits 算法是信号分解的一种具体实现方法, 针对地震信号在过完备库中进行分解时原子字典索引和计算量均非常庞大的问题, 提出了一种原子字典索引的快速生成算法. 首先根据地震信号的频带特征缩小字典索引的频率范围; 然后同样依据地震信号的波形特征缩小字典索引的尺度范围, 从而利用对地震信号的先验知识, 收缩原子字典索引的扫描范围, 实现基于 Matching Pursuits 算法的地震信号的快速分解.

**关键词** 字典, 原子, 匹配追踪, 快速算法

中图分类号 P315

文献标识码 A

文章编号 1004-2903(2007)06-1962-06

## Seismic signal's decomposition based on matching pursuits method

CHEN Fa-yu, YANG Chang-chun

(Institute of Geology and Geophysics, Chinese academy of Sciences, Beijing 100029, China)

**Abstract** Matching Pursuits algorithm is a specific implementing method for signal decomposition. It's a problem to decompose or reconstruct seismic signal upon over-completed dictionary since the index of the dictionary and the amount of the computation are very huge. So a fast algorithm is proposed in which the dictionary index is shortened in frequency scope and in scale scope according to the frequency characteristic and the waveform characteristic of the seismic signal. Thus, by means of using known knowledge about the seismic signal, shrinkage of scan scope of atom dictionary index is achieved to accomplish the seismic signal's fast decomposition based on Matching Pursuits algorithm.

**Keywords** dictionary, atom, Matching Pursuits, fast algorithm

### 0 引 言

MP(Matching Pursuits)匹配追踪方法<sup>[1]</sup>的实质是将信号在过完备库中进行分解. MP 算法的计算量主要体现在三个方面, 原子字典的索引方式、原子的生成和信号的分解过程. 原子字典的索引确定了信号进行匹配追踪计算时扫描原子的方式, 对原子的匹配搜索确定了信号分解的方向, 字典索引则确定了对原子进行匹配搜索的密度和循环迭代的程度. 同时原子生成的计算过程同样构成了庞大的计算量, 不同原子字典索引所控制的原子(参数)在离散化时均需要经历多次的重复计算. 最后在进行信号与原子的相关性计算时也会涉及大量的重复计算.

对于原子生成的算法, 可以依据原子的尺度参数, 将其划分为小原子和大原子<sup>[3]</sup>. 对于小原子, 因为其能量集中在较小的范围, 所以用小范围生成的局部原子代替整个原子; 对于大原子, 先生成相应的较小原子, 然后通过插值方法生成大原子. 这样通过算术插值计算来替代幂和乘积计算, 从而提高算法的计算速度.

对于循环迭代过程, 可以充分利用前一次计算的中间结果<sup>[1]</sup>, 将每一步骤的计算紧密地结合上一个步骤的计算, 避免多次重复计算从而提高计算速度. 由于原子的控制参数之间存在相互关联关系<sup>[2]</sup>, 因而可以通过尺度、位移和频率参数来获得最优的相位参数, 从而可以避免对相位参数的扫描而提高计算速度.

收稿日期 2007-05-23; 修回日期 2007-08-20.

基金项目 国家 973 项目(2005CB422104)资助.

作者简介 陈发宇, 男, 1965 年出生. 现为中国科学院地质与地球物理所攻读固体地球物理专业博士研究生, 主要从事地震信号处理的学习和研究工作. (E-mail: chenfayu@mail.iggcas.ac.cn, chenfayu@126.com)

对于地震信号的 MP 分解和重构,可以充分利用地震信号的频率和波形特性,从原子字典索引的角度出发,缩小频率参数和尺度参数的扫描范围,从而实现计算速度的提高.

### 1 匹配追踪算法

MP 算法是将信号在过完备库中进行分解,过完备则意味着信号的分解目标在信号所组成的空间中是足够的密集,因此无法再保持相互正交的基的特性,所以信号分解为原子的组合.

假设  $D$  为进行信号分解的过完备库,地震信号为  $f$ ,信号长度为  $N$ .  $D$  中的元素满足如下的关系:  $D = \{g_\gamma; \gamma \in \Gamma\}$  且  $\|g_\gamma\| = 1$ . 匹配追踪算法是通过把  $f$  函数垂直投影到字典  $D$  的元素上来进行重复迭代估算的. 假设  $g_{\gamma_0} \in D$ , 则  $f$  函数可以表示为:

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + Rf, \tag{1}$$

其中  $Rf$  是把  $f$  函数在  $g_{\gamma_0}$  上进行近似后的残值,为了使残值尽可能的小,就必须使内积计算项  $|\langle f, g_{\gamma_0} \rangle|$  尽可能大. 很显然,  $g_{\gamma_0}$  与  $Rf$  是正交的,因此:

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + \|Rf\|^2. \tag{2}$$

假设  $R^0 f = f$ , 且我们进行了  $n$  轮的迭代 ( $n \geq 0$ ) 得到残值  $R^n f$ . 此时我们再选择一个原子  $g_{\gamma_n} \in D$ , 使其匹配追踪这个残值  $R^n f$ . 因此:

$$R^n f = \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^{n+1} f, \tag{3}$$

其中  $R^{n+1} f$  就是进行了  $n+1$  次迭代得到的残值. 匹配追踪算法就是这样一个重复迭代的过程. 如果重复这样的迭代过程  $m$  次,即可将  $f$  表示为如下的形式:

$$f = \sum_{n=0}^{m-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^m f. \tag{4}$$

因此经过  $m$  次迭代分解计算后,原始信号可用  $m$  个原子的合成来近似表示,其误差为第  $m-1$  次迭代计算后的残差.

### 2 Gabor 字典索引

假设过完备库  $D$  中的元素为 Gabor 原子  $g_\gamma$ <sup>[1]</sup>.  $g_\gamma$  是由经过调制的高斯窗函数构成. 高斯窗函数定义为:

$$g(t) = e^{-\pi t^2}. \tag{5}$$

Gabor 原子  $g_\gamma$  定义为:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\omega t},$$

$$g_\gamma(t) = \frac{K_\gamma}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \cos(\xi t + \varphi),$$

$$g_\gamma(n) = g_s(n-p) \cos\left(i \frac{2\pi k}{N} + \varphi\right), \tag{6}$$

其中,  $\gamma = (s, u, \xi, \varphi)$ ,  $s$  为尺度参数,  $u$  为位移参数,  $\xi$  为频率参数,  $\varphi$  为相位参数. 这样, MP 算法的实质便是在每次迭代计算中,依照原子字典的索引方式扫描原子的控制参数从而得到此次迭代中与分析信号最为相关的原子,然后通过从分析信号中剔除此最相关原子的能量来获得此次迭代计算后的残值信号. 通过控制迭代次数或残值信号的能量可以控制匹配追踪的计算过程.

对于 Gabor 原子所组成的字典, Mallat 索引算法的时频参数是按以下方法离散化:

$$(\gamma, \varphi) = (a^j, pa^j \Delta u, ka^{-j} \Delta \xi, i \Delta \varphi),$$

其中  $a = 2$ ,  $\Delta u = 1/2$ ,  $\Delta \xi = \pi$ ,  $\Delta \varphi = \pi/6$ ,  $0 < j < \log_2 N$ ,  $0 \leq p < N2^{-j+1}$ ,  $0 \leq k < 2^{j+1}$ ,  $0 \leq i < 12$ . 信号的采样点数为  $N$ , 分频基数为 2, 则随着倍频程  $j$  值的递增, 位移参数的扫描间隔在不断稀疏, 而频率参数的扫描间隔在不断密集.

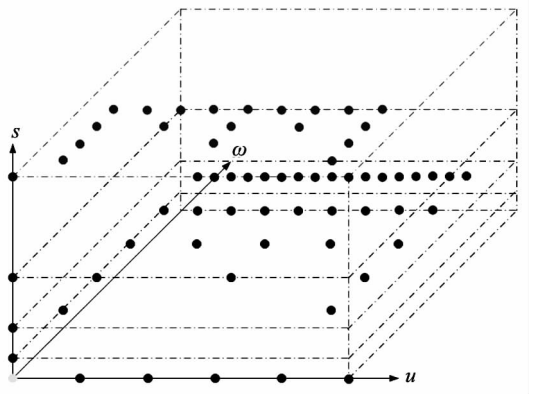


图 1 Gabor 原子的离散算法

Fig. 1 Discrete method of Gabor atom

从图 1 可知,随着倍频程  $j$  值的递增(即尺度参数  $s$  的增加),位移参数的扫描间隔在增加,而频率参数的扫描间隔则在减小. 图中的红点和兰点的坐标位置即代表了时频原子控制参数的数值. 因此在同一次迭代过程中,根据倍频程参数先确定各个时频参数的取值范围和步距,再依次扫描时频参数来选取最相关的原子参数. 所以扫描原子的计算量是十分庞大的, MP 算法的效率和成本都体现于此.

### 3 地震信号的 MP 的快速算法

由 Gabor 原子的计算公式(6)可知,尺度参数控制了谐波信号的衰减程度.尺度越小,谐波信号衰减越快,适合描述较高频率的信号;而尺度越大,谐波信号衰减越慢,适合描述较低频率的信号.

在进行 MP 计算时所选取的 Gabor 原子是在控制衰减的包络线内,能够出现谐波信号(一个或几个周期)的完整波形,因此在包络与谐波信号的频率之间便建立了一定的对应关系.为了描述高频的 Gabor 原子,使其在包络之内出现一个或多个完整的波形,就需要选取较小的尺度参数;为了描述低频的 Gabor 原子,使其在包络之内出现一个或多个完整的波形,就需要选取较大的尺度参数.如图 2 所示.

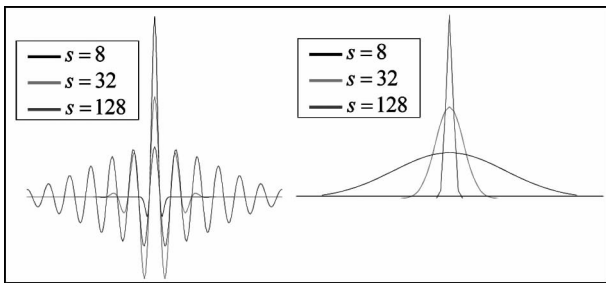


图 2 尺度参数对 Gabor 原子的影响

Fig. 2 The function of scale parameter to Gabor atom

在频率确定的情况下,我们选取出现一个至四个完整谐波信号波形时的包络即尺度参数作为尺度的扫描区间,即尺度参数的选择是使 Gabor 原子的波形在包络中仅出现一个周期至四个周期.这样明确和缩小了尺度参数的有效搜索范围,提高了计算效率和精度.

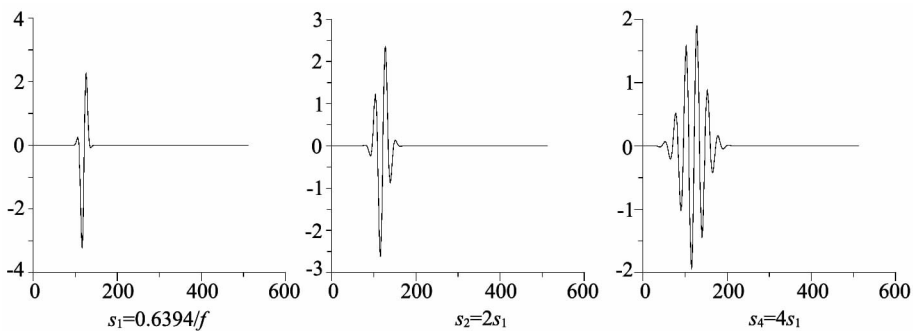


图 3 不同尺度的 Gabor 原子

Fig. 3 Gabor atoms with different scale

高斯正态分布函数为:

$$P(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-x^2/(2\sigma^2)},$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx. \tag{7}$$

因为,  $\Phi(1.96) = 0.975$ ,  $\Phi(-1.96) = 0.025$ , 所以,  $\Phi(1.96) - \Phi(-1.96) = 0.95$ , 即高斯分布的对称轴两侧  $\pm 1.96\sigma$  范围内的分布概率为: 95%. 因此, 我们可以认定  $\pm 1.96\sigma$  就是高斯函数衰减包络线的有效宽度(占实际的 95%). 进而通过包络线的有效宽度来确定和控制衰减的程度参数, 即尺度参数.

Gabor 原子的定义为:

$$g_{(\gamma, \varphi)}(n) = K_{(\gamma, \varphi)} \exp\left(-\pi \frac{(n-u)^2}{s^2}\right) \times \cos\left(\frac{2\pi f}{N}n + \varphi\right),$$

令:  $s^2 = 2\pi\sigma^2$ , 则有

$$g_{(\gamma, \varphi)}(n) = K_{(\gamma, \varphi)} \exp\left(-\frac{(n-u)^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi f}{N}n + \varphi\right), \tag{8}$$

此时包含 Gabor 原子 95% 能量的有效包络宽度为:  $\pm 1.96\sigma$ ; 假设此时所包络的谐波信号为一个周期  $t_0$ , 则有:  $t_0 = 2 \times 1.96\sigma$ , 因此,

$$s_1 = \sqrt{2\pi}\sigma = \sqrt{2\pi}t_0 / (2 \times 1.96) = 0.6394t_0.$$

如果此时所包络的谐波信号为四个周期  $4t_0$ , 则有:  $4t_0 = 2 \times 1.96\sigma$ , 因此,

$$s_2 = \sqrt{2\pi}\sigma = \sqrt{2\pi} \times 4t_0 / (2 \times 1.96) = 2.5578t_0.$$

这样便可由频率来确定尺度参数的扫描范围. 例如: 频率  $f$  参数为 10 赫兹, 位移参数  $u$  为 120, 相位参数  $\Phi$  为 0 时, 尺度参数  $s$  的扫描范围如下图所示.

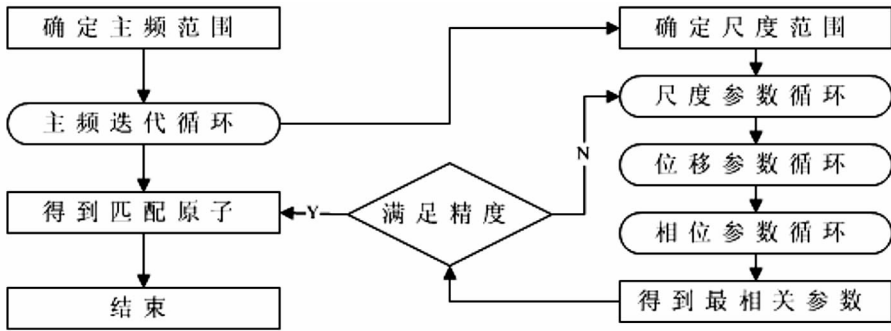


图 4 Gabor 原子的索引流程

Fig. 4 Gabor atom indexing process

地震信号的主频范围非常明确,通常分布在 30 Hz~100 Hz 之间. 因此对地震信号进行 MP 分解计算时,可以围绕频率参数来实现 Gabor 原子的快速索引和参数表达. 首先确定频率参数的扫描范围和扫描间隔;再依据上述频率与包络之间的对应关系,确定与频率相关的尺度参数的扫描范围;然后对位移参数和相位参数进行循环扫描,从而实现对 Gabor 原子的离散计算.

这种离散算法的有点是充分利用了对地震信号的先验知识,将通常意义下对尺度参数的全局离散操作,转化为对与特定范围内频率相关的尺度参数的局部离散操作. 这样不仅使得离散过程紧紧围绕地震主频来进行,突出信号分解过程中的频率因素;同时也因为建立了频率参数与尺度参数的直接关联而降低了循环迭代的计算维度.

## 4 计算实例

### 4.1 仿真地震信号

我们用  $n$  个 Gabor 原子来合成一个信号来仿真地震信号. 假设采样间隔为 2 ms, 采样点数为 1024, 则记录长度为 2 s. Gabor 原子的计算参照公式(6).

选取  $n = 8$ , 并按照如下的参数计算样本信号:

$$s = [128, 16, 8, 32, 64, 32, 16, 256]^T,$$

$$u = [100, 240, 300, 380, 490, 590, 680, 900]^T,$$

$$\xi = [15, 30, 35, 25, 20, 25, 30, 10]^T,$$

$$\varphi = [0, 1.13, 0.13, 0.34, 0.23, 0.13, 0.38, 0.27]^T.$$

### 4.2 混合 Gauss 噪音的仿真地震信号

我们用符合 Gauss 分布的随机函数作为噪音混入,以生成新的模拟信号. 噪音的方差值决定了噪音的强弱,我们选择的噪音方差为 0.05, 经过匹配追踪计算分别得到了对原始信号的重构信号.

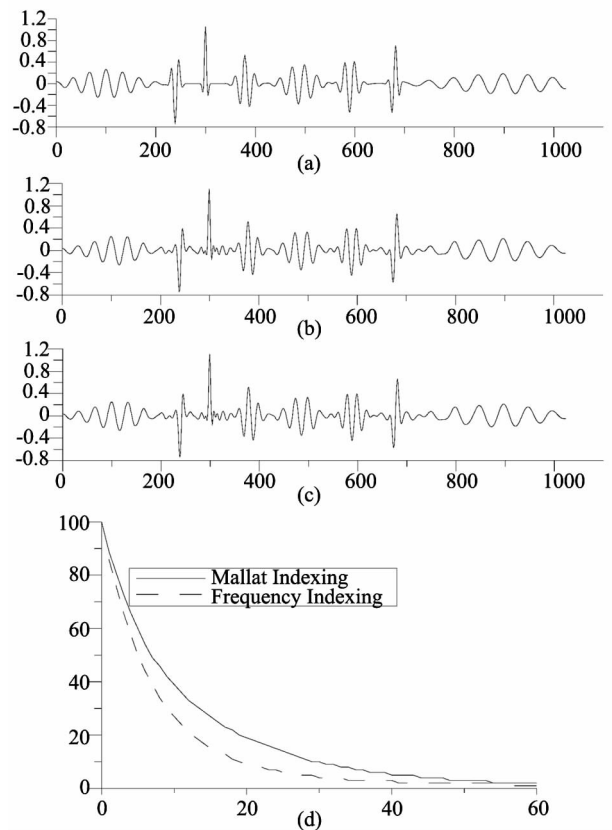


图 5 仿真地震信号 MP 索引算法比较

(a) 原始信号; (b) 重构信号; Mallat 索引算法;

(c) 重构信号; 频率索引算法

(d) 算法对比: 纵坐标是信号能量横坐标是原子数目

Fig. 5 Emulating seismic signal's MP indexing methods comparing

## 5 讨论

从图 5 (d)和图 6 (d)的算法对比可知,无论是对纯粹的仿真信号还是对混合了噪音的仿真信号,以频率为主导的 MP 索引方式均比 Mallat 索引方式具有更快的收敛速度,图中表示频率索引的虚线

总是比表示 Mallat 索引的实线更快地衰减,表明在对原始信号进行近似重构时可以用更少量的原子来实现。

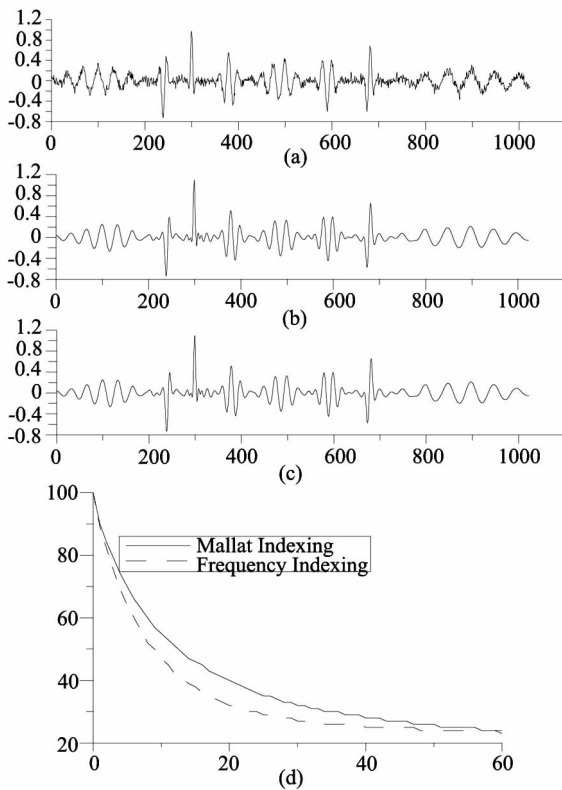


图6 混合噪声的仿真地震信号 MP 索引算法比较

(a) 原始信号; (b) 重构信号: Mallat 索引算法;

(c) 重构信号: 频率索引算法

(d) 算法对比: 纵坐标是信号能量横坐标是原子数目

Fig. 6 Noised emulating seismic signal's MP indexing methods comparing

本文介绍了一种针对地震信号的 MP 匹配追踪的频率索引快速算法,充分利用了对地震信号频率的先验知识,通过建立频率与尺度之间的关联关系,压缩了频率参数和尺度参数的扫描范围,从而加快了全局参数寻优的计算速度;同时由于强调了频率参数的作用,扫描过程紧紧围绕频率来进行,因此匹配追踪的收敛速度也有所提高,可以实现以较为少量的迭代循环或原子个数来近似追踪原始信号。

致 谢 感谢导师杨长春研究员的悉心指导。

## 参 考 文 献 (References):

[1] Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries[J]. IEEE Trans. Signal Process., 1993, 41(12):

3397~34215.

- [2] Ferrando S E, Kolas L A, Kovacevic N. A flexible implementation of matching pursuit for Gabor dictionaries.
- [3] Durka P J. Stochastic Time-frequency dictionaries for matching pursuit[J]. IEEE Trans. Signal Processing, 2001, 49(3): 507~510.
- [4] Qian S, Chen D. Signal representation via adaptive normalized Gaussian functions[J]. Signal Processing, 1994, 36: 1~11.
- [5] Liu J, Wu Y, Han D, Li X. Time-frequency decomposition based on Ricker wavelet: 74th Annual International Meeting [J]. SEG, Expanded Abstracts, 2004, 1937~1940.
- [6] Gribonval R. Fast matching pursuit with a multiscale dictionary of Gaussian Chirps[J]. IEEE Trans. on Signal Processing, 2001, 49(5): 994~1001.
- [7] Liu Q, Wang Q, Wu L. "Size of the dictionary in matching pursuit algorithm"[J]. IEEE Trans. Sig. Proc., 2004, 52(12): 3403~3408.
- [8] Arthur P L, Philips C L, Voiced/unvoiced speech discrimination in noise using Gabor atomic decomposition[A]. Proc of IEEE ICASSP[C]. Hong Kong, IEEE, 2003, I(4): 820~828.
- [9] Gratkowski M, Haueisen J, Arendt-Nielsen L, Chen A C, Zanow F. Time-frequency filtering of MEG signals with matching pursuit[J]. J Physiol Paris, 2006, 99(1): 47~57.
- [10] Wang Y H. Seismic time-frequency spectral decomposition by matching pursuit[J]. Geophysics, 2007, 72(1), 13~20.
- [11] Durka P J, Blinowska K J. Analysis of EEG transients by means of matching pursuit[J]. Ann Biomed Eng. 1995, 23(5): 608~611.
- [12] Cui J, Wong W, Mann S. Time-frequency analysis of visual evoked potentials by means of matching pursuit with chirplet atoms[J]. Conf Proc IEEE Eng Med Biol Soc. 2004, 1: 267~70.
- [13] 尹忠科, 王建英, Vandergheynst P. 在低维空间实现的图像稀疏分解算法[J]. 电讯技术, 2004, 44(3): 12~15.
- [14] 傅霆, 尧德中. 稀疏分解的加权迭代方法及其初步应用[J]. 电子学报, 2004, 32(4): 567~570.
- [15] 尹忠科, 王建英, 邵君. 基于原子库结构特性的信号稀疏分解[J]. 西南交通大学学报, 2004, 40(2): 173~178.
- [16] 刘光鼎. 我国油气资源勘探开发中存在的主要问题及对策[J]. 地球物理学进展, 2005, 20(1): 1~3.
- [17] 殷文, 印兴耀. 基于 MPI 的时频分布的改进及应用[J]. 地球物理学进展, 2005, 20(1): 165~169.
- [18] 李永兵, 陈绪松, 刘建明, 等. 利用 VBA 和 Excel 实现快速处理国产甚低频电磁仪 DDS-1 测量数据[J]. 地球物理学进展, 2005, 20(1): 225~230.
- [19] 刘光鼎, 陈洁. 中国前新生代残留盆地油气勘探难点分析及对策[J]. 地球物理学进展, 2005, 20(2): 273~275.
- [20] 崔若飞, 武旭仁, 陈同俊. 煤矿地震数据管理系统的开发[J]. 地球物理学进展, 2005, 20(2): 374~376.
- [21] 吴钦宝, 王秀东, 宁方助. 煤田三维地震勘探的应用及效果分析[J]. 地球物理学进展, 2005, 20(2): 377~380.
- [22] 陈相府, 安西峰, 王高伟. 浅层高分辨地震勘探在采空区勘测

- 中的应用[J]. 地球物理学进展, 2005, 20(2): 437~439.
- [23] 金雷, 李月, 杨宝俊. 用时频峰值滤波方法消减地震勘探资料中随机噪声的初步研究[J]. 地球物理学进展, 2005, 20(3): 724~728.
- [24] 李庆谋, 刘少华, 梁光和. GIS 环境下对地球物理地球化学资料奇异性的各向异性估算及其在矿床勘探中的应用[J]. 地球物理学进展, 2005, 20(4): 1015~1020.
- [25] 张军华, 吕宁, 田连玉, 等. 地震资料去噪方法、技术综合评述[J]. 地球物理学进展, 2005, 20(4): 1083~1091.
- [26] 刘喜武, 刘洪, 李幼铭, 等. 基于广义 S 变换研究地震地层特征[J]. 地球物理学进展, 2006, 21(2): 440~451.
- [27] 张军华, 吕宁, 田连玉, 等. 地震资料去噪方法技术综合评述[J]. 地球物理学进展, 2006, 21(2): 546~553.
- [28] 张进铎. 地震解释技术现状及发展趋势[J]. 地球物理学进展, 2006, 21(2): 578~587.
- [29] 徐亚, 郝天珧, 周立宏, 等. 位场小波变换研究进展[J]. 地球物理学进展, 2006, 21(4): 1132~1138.
- [30] 刘英明, 邹长春. 一种基于遗传算法的测井曲线高分辨率处理方法[J]. 地球物理学进展, 2006, 21(4): 1202~1207.
- [31] 李景叶, 陈小宏, 芮振华. 基于匹配滤波的多次波压制方法研究[J]. 地球物理学进展, 2007, 22(1): 200~206.
- [32] 陈赟, 高乐, 赵烽帆. 一种基于频率域偏振分析提高三分量地震资料信噪比的方法[J]. 地球物理学进展, 2007, 22(1): 255~261.
- [33] 苟学强, 张义军, 董万胜, 郗秀书. 0101 基于小波的地震首次回击辐射场的多重分形分析[J]. 地球物理学报, 2007, 50(1): 101~105.
- [34] 徐果明, 姚华建, 朱良保, 沈玉松. 中国西部及其邻域地壳上地幔横波速度结构[J]. 地球物理学报, 2007, 50(1): 193~208.
- [35] 凡友华, 陈晓非, 刘雪峰, 刘家琦, 陈小宏. Rayleigh 波的频散方程高频近似分解和多模式激发数目[J]. 地球物理学报, 2007, 50(1): 233~239.
- [36] 刘洪, 袁江华, 勾永峰, 张向君, 李幼铭. 地震逆散射波场和算子的谱分解[J]. 地球物理学报, 2007, 50(1): 240~247.
- [37] 高静怀, 周艳辉, 毛剑, 陈文超, 吴如山, 李幼铭. 局部角度域波传播步进算法研究[J]. 地球物理学报, 2007, 50(1): 248~259.
- [38] 刘定进, 印兴耀. 傅里叶有限差分法保幅叠前深度偏移方法[J]. 地球物理学报, 2007, 50(1): 268~276.