2019 IEEE Symposium Series on Computational Intelligence (SSCI)
December 6-9 2019, Xiamen, China

Two Objective Matching Pursuit for Seismic Data Decomposition

Yongqing Li, Jun Wang, Hui Li, and Peng Ren
College of Control Science and Engineering
China University of Petroleum (East China),
Qingdao 266580, P. R. China
Email: lyq_upc@163.com; slwangjun@upc.edu.cn; lihuiupc@126.com; pengren@upc.edu.cn

***Abstract*—Matching pursuit is an effective method for seismic data decomposition. The traditional matching pursuit has one single objective that successively maximizes the inner products between signal residuals and atoms at different approximating levels. However, such inner product strategy does not guarantee the consistency of the peak of an atom and that of its corresponding signal residual. To address this limitation, we propose a two objective matching pursuit strategy that not only optimizes the traditional matching pursuit but also minimizes peak distance. Compared with the traditional matching pursuit with one single objective, the two objective matching pursuit has higher search ability. In this scenario, the genetic algorithm (GA) used for solving the traditional matching pursuit is not suitable for the two objective matching pursuit. Therefore, we propose to solve the two objective optimization problem by exploiting a nondominated sorting genetic algorithm (NSGA-II) algorithm, which not only increases the diversity of searching optimization and but also reduces the reconstruction error over the GA. Experiments for seismic data validate the advantages of the proposed two objective matching pursuit.**

Keywords—matching pursuit; nondominated sorting genetic algorithm; multiobjective optimization.

I. INTRODUCTION

Matching pursuit is one of the earliest sparse methods which have been widely used because of its simplicity and effectiveness. Matching pursuit for seismic data is an essential strategy that analyses data structure and extracts geological information. Wang [1] acquired seismic time-frequency spectral decomposition by matching pursuit. Chen et al. [2] proposed a fast matching pursuit method by shrinking the scan scope of atom dictionary. Zhou et al. [3] employed a fast orthogonal matching pursuit algorithm to seismic noise attenuation. The large scale of an atom dictionary makes matching pursuit computationally expensive. With the development of evolutionary computation, matching pursuit with genetic algorithm (GA) was applied to accelerating the process of searching for optimal solution. Wang et al. [4] conducted multichannel matching pursuit with GA. Xiong et al. [5] used structure-oriented filtering method and GA to decompose and reconstruct seismic data.

However, conventional matching pursuit with one single objective does not achieve the consistency between the peak of an atom and that of its corresponding signal residual. To address this limitation, we propose a two objective matching pursuit method that is based on multiobjective optimization. In this scenario, GA based on one single objective is unsuitable for solving the two objective matching pursuit. We exploit a nondominated sorting genetic algorithm (NSGA-II) algorithm to solving the two objective optimization problem. NSGA-II is the typical algorithm of multiobjective optimization [6]. The outputs of NSGA-II are pareto solutions that provide more choices for finding an optimal solution [7]. We combine the two objective matching pursuit with NSGA-II in order to improve the optimization process and accelerate the iteration. Instead of acquiring one optimal solution in each iteration of GA, NSGA-II provides multiple solutions used for choice and mutation, which greatly increases the diversity of local search. Therefore, the two objective matching pursuit exports the optimal atom that matches the signal residual better. The decomposition results of seismic data demonstrate the effectiveness and efficiency of the algorithm proposed in this paper.

II. SEISMIC DATA DECOMPOSITION WITH MATCHING PURSUIT

A seismic record is normally formulated as the convolution of seismic wavelets and reflection coefficients, with added noise. For the time t, let $f(t)$, $a(t)$, $\delta(t)$ and $n(t)$ denote the seismic record, the seismic wavelet, the reflection coefficient and the noise, respectively. The seismic record is formulated as follows:

$$f(t) = \sum_{\tau=0}^{\infty} a(\tau)\delta(t-\tau) + n(t) \tag{1}$$

where τ is the summation variable for the convolution.

In order to analyze the signal structure, an essential method is to decompose the seismic record into a series of atoms that best match the signal structure with respect to a given dictionary. Matching Pursuit decomposes a signal into a linear expansion of waveforms that are selected from a redundant dictionary of functions [8], which is widely used in seismic data decomposition.

The matching pursuit is a greedy algorithm that finds a waveform best adapted to the approximate part of the signal at each iteration [8]. Matching pursuit is commenced by constructing an atom dictionary. In our work, we use a Ricker wavelet as an atom. Ricker wavelets approximately simulate the actual seismic wavelets. Let u, ξ, s and φ denote time delay, frequency, scale and phase, respectively. The waveform of the Ricker wavelet is decided by the parameter vector $\gamma = [u, \xi, s, \varphi]^T$. For the time t, the Ricker wavelet with four parameters is given as follows:

$$g_\gamma(t) = \left\{1 - 2\left[\pi\xi\left(t-u\right)\right]^2\right\} e^{-\left[\pi\xi(t-u)/s+\varphi\right]^2}. \tag{2}$$

We use Ricker wavelets with different parameters to construct the dictionary $D = \left\{g_\gamma\right\}_{\gamma\in\Gamma}$, where the parameter set Γ consists of parameter vectors generated by discreting u, ξ, s and φ. g_γ denotes an atom and satisfies $\left|\left|g_\gamma\right|\right|_2 = 1$ by atom normalization.

Matching pursuit is conducted by successive approximations of the residual signal with orthogonal projections on the elements of D. Let f and $R^n f$ denote the initial signal and the residual signal of the nth iteration, respectively. The first residual vector $R^0 f$ is equal to f. The purpose of the (n+1)th iteration is to search for an atom g_{γ_n} that best matches $R^n f$. Its objective function is formulated as follows:

$$v_1(\gamma) = \left|\left\langle R^n f, g_\gamma\right\rangle\right|. \tag{3}$$

Here, we compute g_{γ_n} that maximizes the objective function (3). Let $R^{n+1} f$ denote the residual vector after approximating $R^n f$ in the direction of g_{γ_n}. The vector $R^n f$ is decomposed into:

$$R^n f = \left\langle R^n f, g_{\gamma_n}\right\rangle g_{\gamma_n} + R^{n+1} f. \tag{4}$$

Clearly, g_{γ_n} is orthogonal to $R^{n+1} f$, and hence:

$$\|R^n f\|^2 = \left\|\left\langle R^n f, g_{\gamma_n}\right\rangle\right\|^2 + \left\|R^{n+1} f\right\|^2. \tag{5}$$

With the increase of n, f is decomposed into:

$$f = \sum_{n=0}^{\infty} \left\langle R^n f, g_{\gamma_n}\right\rangle g_{\gamma_n}. \tag{6}$$

However, n is finite in the actual operation. The number of iterations is an important parameter for matching pursuit. We use the adjacent residual ratio threshold as the termination condition of iterations. Let $E(\cdot)$ represent mathematical expectation. We compute variable α as follows:

$$\alpha = \sqrt{E(\|R^{n+1} f\|^2)/E(\|R^n f\|^2)}. \tag{7}$$

The threshold σ is computed as follows:

$$\sigma = \frac{\|R^{n+1} f - \alpha R^n f\|^2}{\|\alpha R^n f\|^2}. \tag{8}$$

σ is used to terminate the greedy iteration and avoid over-matching caused by noisy signals.

III. TWO OBJECTIVE MATCHING PURSUIT

Conventional matching pursuit only has one objective that maximizes inner product between the selected atom and the signal residual. Its objective function is formulated as the equation (3). However, such single objective strategy does not guarantee the consistency of the peak of an atom and that of its corresponding signal residual.

To address the limitation of the single objective strategy, we propose to incorporate another objective that minimizes the peak distance between the atom and the signal residual. Let $p(g_\gamma)$ and $p(R^n f)$ denote the peak position of g_γ and that of $R^n f$ corresponding to the support part of g_γ, respectively. The peak distance is given as follows:

$$w(\gamma) = \left|p(R^n f) - p(g_\gamma)\right|. \tag{9}$$

The objective function based on $w(\gamma)$ is represented as follows:

$$v_2(\gamma) = \frac{1}{w(\gamma)}. \tag{10}$$

The two objective matching pursuit takes both $v_1(\gamma)$ and $v_2(\gamma)$ into consideration. The function model based on two objectives is formulated as follows:

$$\boldsymbol{V}(\gamma) = \left[v_1(\gamma),\, v_2(\gamma)\right]. \tag{11}$$

The purpose of multiobjective optimization is to maximize the objective function (11). The multiobjective optimization contributes to increasing the diversity of choices and accelerating the convergence speed of the algorithm. However, $v_1(\gamma)$ and $v_2(\gamma)$ do not reach the maximum at the same γ. Traditional genetic algorithm (GA) based on single objective is incapable of searching for the optimal solutions of multiobjective problems. In the subsequent section, we will investigate how to solve the two objective optimization problem.

IV. OPTIMIZATION BASED ON NSGA-II

In order to solve the optimization problem established by the two objective function (11), we propose to use a nondominated sorting genetic algorithm (NSGA-II) to perform computation. NSGA-II extends the capability of traditional GA with solving multiobjective optimization. NSGA-II not only has less computational complexity, but also prevents the loss of good solutions once they are found. Therefore, we employ NSGA-II to find the optimal solutions of two objective matching pursuit.

The two main algorithms used in NSGA-II are fast nondominated sorting and crowded comparison. The procedure of nondominated sorting requires $O(MN^2)$ computations, where M is the number of objectives and N is the population size [6]. Through this sorting approach, the population is sorted into different nondomination levels. In addition, the utilization of crowded-comparison operator is helpful for maintaining the diversity among population members [9].

The flowchart of NSGA-II is described in Fig. 1. We introduce the important steps of the flowchart in details as follows.

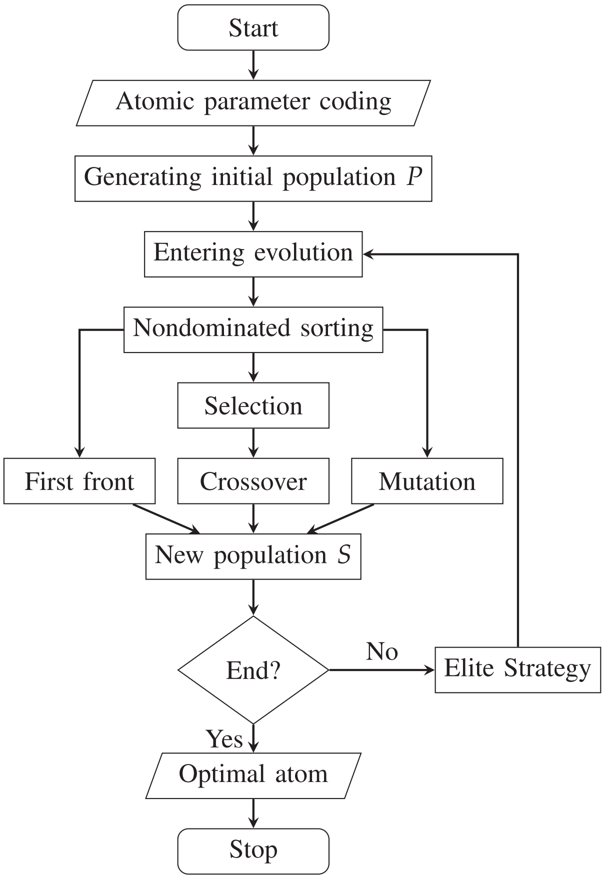

Fig. 1. The flowchart of NSGA-II.

Atomic parameter coding: The process of coding is to represent atoms by parameter vectors, and each atom is corresponding to four parameters.

Generating initial population P: The initial population is the first generation that is chosen from the atom dictionary based on random sampling.

Nondominated sorting: According to the objective function (11), atoms in a population are ranked by the nondominated sorting algorithm. Each rank contains more than one atom, which is different from the traditional fitness function based on one single objective.

First front: The solutions in the first nondominated front have their domination count as zero, and these atoms enter the next population directly.

Selection: Except for the atoms in first front, rest atoms participate the competition in order to select the excellent individuals. The principle of selection is based on nondominated levels and crowded distance, and the selection method is usual binary tournament. If two atoms have different nondominated ranks, we prefer the one with a lower rank. Otherwise, if two atoms belong to the same rank, we prefer the atom that is located in a less crowded region.

Crossover: When the operation of selection is completed, the atoms selected from competition are recombined by exchanging parameters randomly, which increases the possibility of generating approximate optimization.

Mutation: The operation of mutation is based on the atoms of the first front. These atoms are relatively optimal in a whole population. The mutation of these atoms contributes to searching for optimal solution.

New population S: The new population consists of atoms that are from the first front, selection and mutation. The number of new population is the same as initial population.

Elite Strategy: When the number of iterations is not over, we combine the parent population and the offspring population S. The combined population is sorted according to nondomination. We obtain an elite population based on nondomination ranks and crowded distance.

Optimal atom: After all iterations, we obtain a set of pareto solutions. We choose an optimal solution from first front according to the fastest energy attenuation of the signal residual.

V. EXPERIMENTAL EVALUATION

In this section, we compare the performance between the two objective matching pursuit and the traditional matching pursuit. We use the synthetic record and actual seismic record to verify the effectiveness and efficiency of the proposed algorithm. In Fig. 2, we present a synthetic seismic record that contains 8 Ricker wavelets with different parameters.

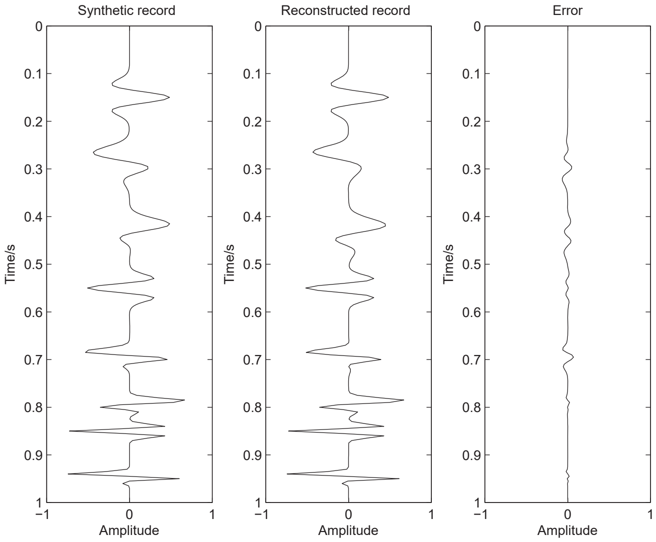

Fig. 2. A synthetic seismic record, its reconstructed result and error.

The synthetic seismic record is decomposed by two objective matching pursuit, and its decomposition results are shown in Fig. 3. The reconstructed record composed of decomposition atoms is illustrated in Fig. 2. The reconstruction error compared with traditional matching pursuit is marked in Table I. Obviously, the two objective matching pursuit with NSGA-II has less reconstruction error.

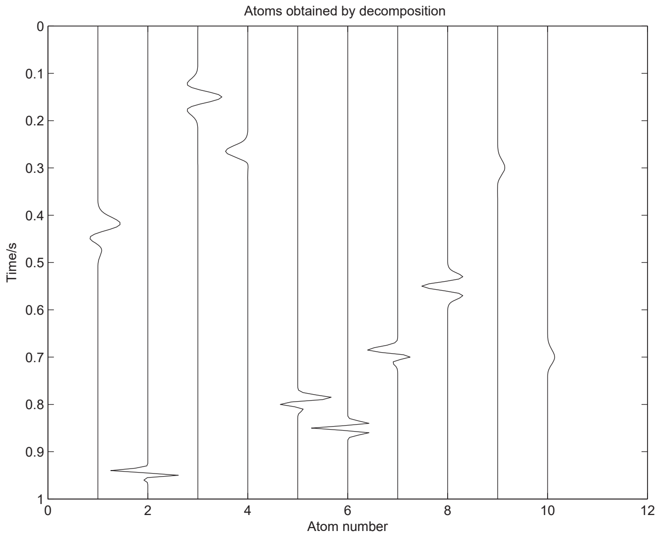

Fig. 3. Decomposition results of the synthetic seismic record.

TABLE I
RECONSTRUCTION ERRORS OF TWO METHODS.

Method	MSE
GA	3.4×10^{-3}
NSGA-II	1.8×10^{-3}

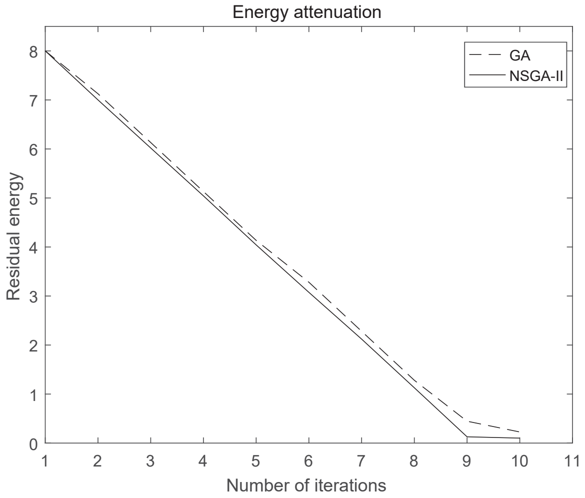

Fig. 4. Energy attenuation curves for the synthetic seismic record.

Genetic algorithms have randomness, and single experimental comparison is meaningless. In order to eliminate the randomness interference of genetic algorithms and compare the capability of searching optimization, we conduct 100 experiments and compute the energy of the residual signal after each iteration. The average value of energy attenuation is shown in Fig. 4. We observe from the tendency of curves and find that the two objective matching pursuit based on NSGA-II finds a more suitable atom in each iteration.

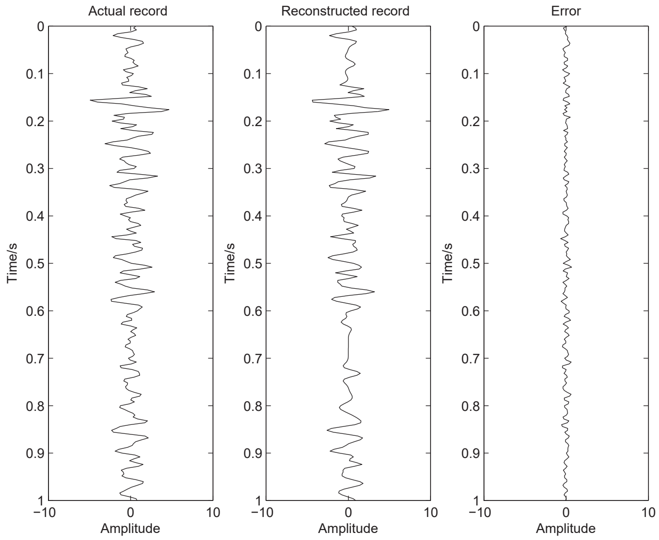

Fig. 5. An actual seismic record, its reconstructed result and error.

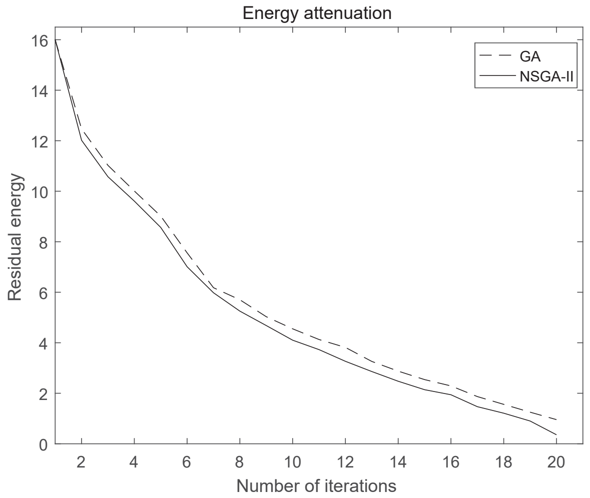

Fig. 6. Energy attenuation curves for the actual seismic record.

As shown in Fig. 5, the three figures contain an actual seismic record, a reconstructed record and reconstruction error, separately. To avoid the influence of randomness, we also conduct 100 experiments. Similar to Fig. 4, Fig. 6 illustrates the average energy attenuation of the actual seismic record along with the proceeding of iterations. The curves show that the performance of the two objective matching pursuit is better than traditional matching pursuit.

VI. CONCLUSION

Traditional matching pursuit has one single objective that maximizes the inner products successively, and does not consider the consistency of the peak of an atom and that of its corresponding signal residual. To address this shortcoming, we have proposed a two objective matching pursuit method that increases the possibility of searching for the optimal solutions. In this scenario, the traditional GA based on one single objective is not suitable for the two objective matching pursuit. In this paper, we have exploited NSGA-II to solving the two objective optimization problem. The two objective matching pursuit with NSGA-II has higher search capability and presents more accurate solutions. Experiments for the synthetic and actual seismic records have validated that the proposed algorithm is effective and highly efficient.

REFERENCES

[1] Y. Wang, "Seismic time-frequency spectral decomposition by matching pursuit," *Geophysics*, vol. 72, no. 1, pp. 13–20, 2006.

[2] F. Chen and C. Yang, "Seismic signal's decomposition based on matching pursuits method," *Progress in Geophysics*, vol. 22, no. 6, pp. 1692–1697, 2007.

[3] Y. Zhou, S. Li, J. Xie, D. Zhang, and Y. Chen, "Sparse dictionary learning for seismic noise attenuation using a fast orthogonal matching pursuit algorithm," *Journal of Seismic Exploration*, vol. 26, no. 5, pp. 433–454, 2017.

[4] J. Wang, L. Wang, and Y. Wang, "Seismic signal fast decomposition by multichannel matching pursuit with genetic algorithm," in *IEEE International Conference on Signal Processing*, 2012, pp. 1393–1396.

[5] Y. Xiong, D. Guan, C. Li, and Z. Jiang, "Application of fast matching pursuit decomposition technology based on anisotropic structure-oriented filtering method to complex fault block of bohai bay basin," in *SEG Technical Program Expanded Abstracts*, 2016, pp. 4725–4729.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[7] A. Khan and A. Baig, "Multi-objective feature subset selection using non-dominated sorting genetic algorithm," *Journal of Applied Research and Technology*, vol. 13, no. 1, pp. 145–159, 2015.

[8] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[9] V. Ghomsheh, M. Khanehsar, and M. Teshnehlab, "Improving the non-dominate sorting genetic algorithm for multi-objective optimization," in *IEEE International Conference on Computational Intelligence and Security Workshops*, 2007, pp. 89–92.