Th SBT3 03

An Automatic Arrival Time Picking Method Based on RANSAC Curve Fitting

L. Zhu (Georgia Institute of Technology), E. Liu* (Georgia Institute of Technology) & J.H. McClellan (Georgia Institute of Technology)

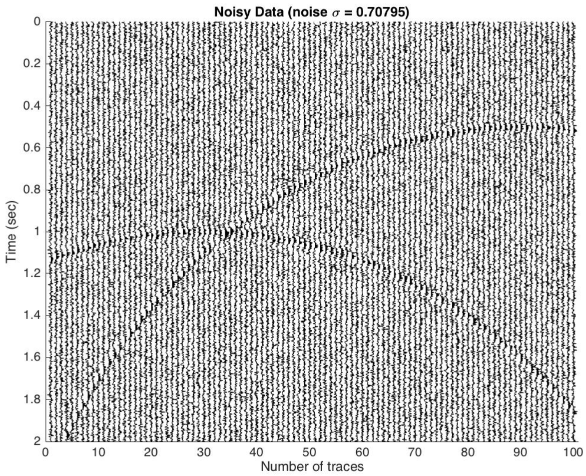
SUMMARY

Arrival time picking is useful in both active and passive seismic processing problems. Many current time picking methods suffer the problem of high false picking rate under low SNR cases. The random noise add wrong picking points that are far from the true moveout curves. In this study, we propose a new automatic arrival time picking method based on RANSAC curve fitting algorithm. Synthetic example indicates that the propose method is robust against high noise and can be used in multiple events scenario.

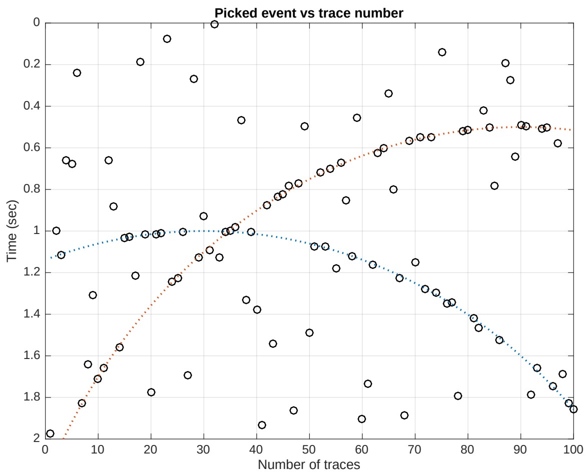
Introduction

The arrival time of microseismic events provide valuable information for predicting the fracturing and increasing the production rate of the unconventional reservoirs. Microseismic monitoring can be performed using downhole, surface, or near-surface arrays (Duncan and Eisner, 2010). The microseismic data received by a sensor array typically have relative weak signals in the presence of noise and interference, especially for the surface array. To estimate the arrival times of the direct waves in a seismogram with low signal-to-noise ratio (SNR) is difficult but desirable for a number of localization algorithms (Maxwell et al., 2010; Sabbione and Velis, 2010; Al-shuhail et al., 2013). Special treatment is needed to enhance and pick the arrival times in the low SNR scenario (Coppens, 1985; Mousa et al., 2011).

Many methods have been developed to solve this time-picking problem, including short-time-average over long-time-average (STA/LTA) algorithm (Earle and Shearer, 1994), cross-correlation based algorithms (Grandi and Oates, 2009), and subspace method (Harris, 2006). All those methods suffer high false alarm rate issue in low SNR cases. Figure 1 is a synthetic example of time-picking under high noise. The seismic signal is generated by Ricker wavelet whose peaks are normalized to 1. Figure 1b shows the picked arrival times by cross-correlation method from the noisy data in Figure 1a. In low SNR case, many of the picked arrival time points lie on or close to the true curve (red and blue lines) as inliers; however, a considerable portion of those points are still far from the true curve as outliers. The lower SNR environment the data were collected in, the more outliers will occur in the time-picking results. Using the picked time in Figure 1 directly in stacking or localization algorithm will significantly harm the results. Instead of trying to improve the accuracy of time-picking algorithm or developing a more robust algorithm for single trace, a curve-fitting algorithm can correct the false detected points and improve the overall time-picking results. In many microseismic applications, the true Earth can be approximated by a layered model. The moveout curve of a event passing through layered velocity model is roughly a hyperbola curve. Thus, we propose to use RANSAC (Fischler and Bolles, 1981) to fit a hyperbola curve with large amount of outliers.



(a)



(b)

Figure 1 *Synthetic example: a)two seismic events with additive white Gaussian noise (noise $\sigma = 0.708$); b)picked cross-correlation maxima (solid line indicate the true moveout curve).*

Method

Modelling a curve as a hyperbola

A hyperbolic curve is a special case of conic section. It can be represented by a general quadratic equation in two variables (x, y) as shown in (1).

$$Ax^2 + Bxy + Cy^2 + Dx + Ey = 1 \tag{1}$$

when $\Delta = B^2 - 4AC > 0$, we have a hyperbola. Notice that (1) takes rotation into account so that a tilted and layered velocity model can be represented by this method as well. When $A = B = C = 0$, (1) degenerates into a linear equation, so this model can also be used to fit linear moveout events.

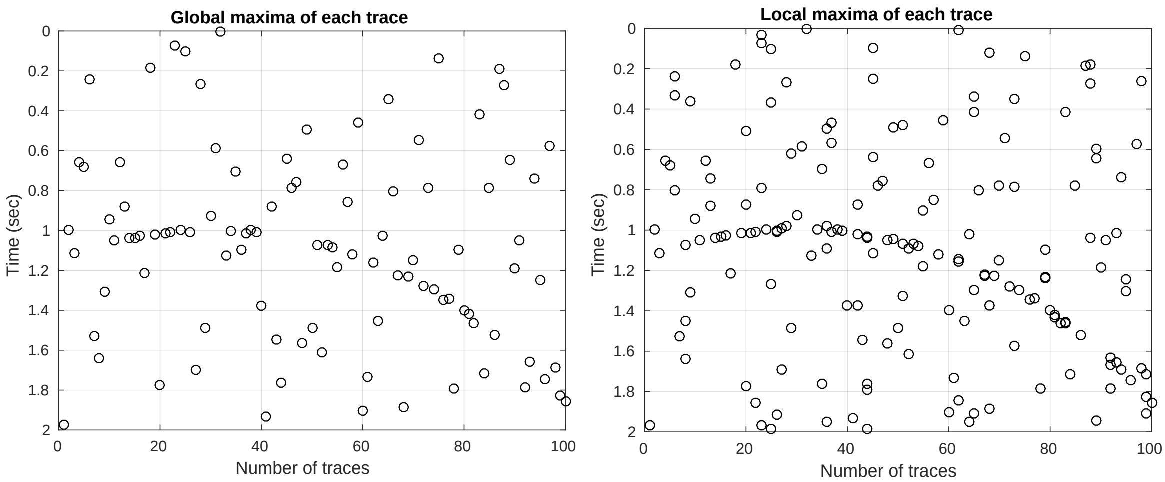
Equation (1) has five parameters, so we need at least five (x,y) pairs to uniquely determine a hyperbolic curve. With noisy data, more (x,y) pairs will improve the fitting results. Given n data points $(x_1,y_1)\cdots(x_n,y_n)$, a data matrix A, constant vector b and parameter vector x can be formed as shown in (2).

$$A = \begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_ny_n & y_n^2 & x_n & y_n \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \text{ and } \quad x = \begin{bmatrix} A \\ \vdots \\ E \end{bmatrix} \tag{2}$$

so that $Ax = b$ in the noise-free case. Assuming additive white Gaussian noise, we can solve for the parameter vector x by least squares approximation to obtain $x = (A^TA)^{-1}A^Tb$. Least squares error minimization is the optimal solution of this curve-fitting problem. However, it is shown in Figure 3a that least-squares type of curve-fitting scheme can easily fail when many outliers are present in the data.

Feature extraction

Obtaining the (x,y) picks is a feature extraction step which can be done with methods such as the STA/LTA ratio or cross-correlation. All of the existing methods generate curves whose magnitudes represent the likelihood of an event. Common practice is to use the peak locations of these curves as the picked arrival time. However, in low SNR cases such as Figure 1, those peaks may be far from the true moveout curve since they may be created by the noise instead of a real event. Real events may create a peak that is comparable to but smaller than the noise peaks which would then be ignored by simply choosing the global maxima of those curves. Therefore, we propose picking all local maxima that have peak values comparable to the maximum peak in a curve. By doing so we will increase the probability of detecting peaks due to real events while not increasing the probability of false peaks due to noise which would come into the dataset as outliers.



(a) (b)

***Figure 2** Feature extraction(noise $\sigma = 0.708$) using a) maximum time of each traces and b)local maxima with threshold on each trace.*

Figure 2 compares the two feature extraction methods on single event example. The proposed method (Figure 2b) has more sample points on the true moveout curve while not introduces many outliers. Since the peaks due to noise are less likely to form a moveout curve over all traces, increasing number of randomly distributed outliers will not cause too much trouble while including more peaks that are close to the true moveout curve will help to improve the curve-fitting results.

Curve fitting

To overcome the problem with a large number of outliers, we use the random sample consensus (RANSAC) method. Given a dataset that contains both inliers and outliers, RANSAC uses a voting scheme to find the optimal fitting result. The key concept is to generate a collection of curves by randomly sampling data points and then select the curves with the largest set of inliers. There are only two essential steps in the RANSAC algorithm which are performed iteratively:

1. Fit least number of random samples using simple fitting algorithm;
2. Determine if the fitted curve in previous step has enough inliers.

For a fixed number of iterations, a collection of feasible curves will be generated. The one with the most inliers will be chosen as the optimal curve that fits the given dataset. To further improve the result, least-squares approximation is performed on the inliers of the best fit curve provided by RANSAC. The detailed steps of proposed method are summarized in Algorithm 1.

Algorithm 1: RANSAC finds the best curve that fits a hyperbola out of (x,y) samples

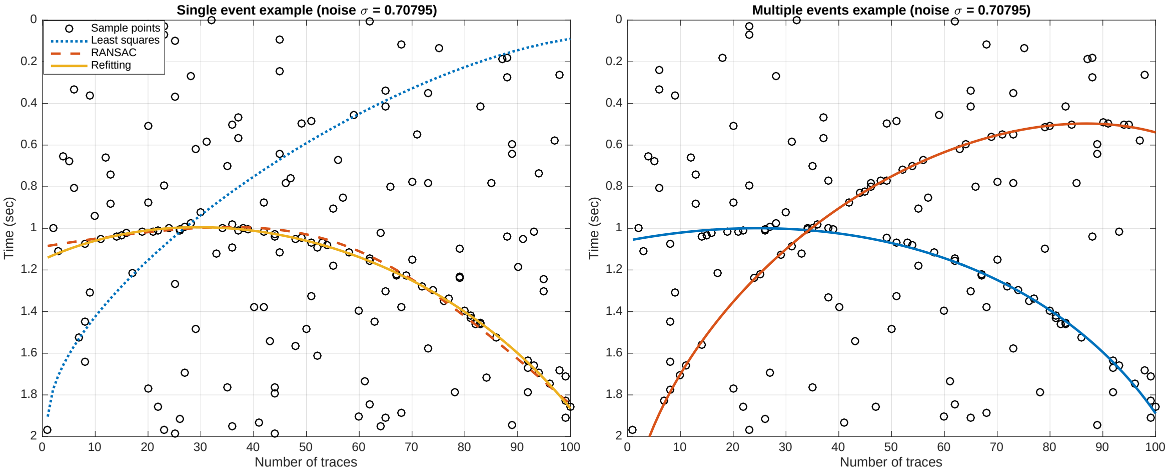
---

Input: Data points from feature extraction
Output: Parameters vector x^* for quadratic equation in (1)
Best fit curve $x^* \leftarrow [0,0,0,0,0]^T$
Constant vector $b \leftarrow [1,1,1,1,1]^T$
Number of inliers of best fit curve $n^* \leftarrow 0$
for $i \leftarrow 1$ ***to*** *# of iteration* **do**
  Randomly select five sample points and form data matrix A as in (2)
  Least-squares fitting: $x_i \leftarrow (A^TA)^{-1}A^Tb$
  Count number of inliers n_i based on distance between fitted curve and all sample points
  if $n_i > n^*$ **then**
    Save all inliers
Use saved inliers to form data matrix A^* as in (2)
Use least-squares approximation to fit all saved inliers: $x^* \leftarrow (A^{*T}A^*)^{-1}A^{*T}b$
return x^*

Example

Figure 3a shows a simple case with synthetic data for a single event using the proposed method. This case has a large number of outliers, so direct least-squares fitting (blue dotted line) fails to pick a curve that is anywhere close to the true moveout curve. After using RANSAC with 100 iterations, we obtain the red dashed curve that approximates the true curve well. However, there are still several places where the RANSAC curve deviates from the true moveout curve. This misfit is due to the small amount of random noise in the sample points used to fit the RANSAC curve. Since it only takes five sample points to fit the final curve, any small random noise in those chosen sample points has significant effect on the fitted curve. Thus, instead of using the curves from RANSAC, we refit the inliers recognized by RANSAC using least squares. Since RANSAC eliminates most of the outliers, the least squares method will find the optimal curve that fit the refined data. The refined result is shown in the orange line in Figure 3a which fits the true moveout curve perfectly.

The proposed method can be extended to finding the arrival times for multiple events simultaneously. RANSAC can fit a collection of curves to a dataset that includes all the events. After choosing the first event curve with the highest number of inliers, all these inliers are removed from the dataset and then redo RANSAC to find the next curve. The number of events can be specified as an input parameter, or determined automatically by RANSAC since RANSAC will stop looking for new curve if none of the possible curves has enough number of inliers. Figure 3b shows an example of two events; this dataset was generated by adding a different event to the previous example under the same SNR environment.



(a) (b)

Figure 3 *Curve fitting results of a)single event:least square (blue dotted line), RANSAC (red dashed line), and refitting with RANSAC inliers (orange solid line) and b)multiple events: blue and red curves indicate two distinct events found by RANSAC.*

Conclusion

We proposed a RANSAC based curve fitting scheme to correct false picking in common arrival time picking algorithms. Synthetic data shows that the proposed method is robust against noise and can be used for multiple events scenarios.

Acknowledgment

This work is supported by the Center for Energy and Geo Processing (CeGP) at Georgia Tech and by King Fahd University of Petroleum and Minerals (KFUPM). We are grateful to Prof. SanLinn Isma'il Kaka and Prof. Azzedine Zerguine for helpful discussions.

References

Al-shuhail, A.A., Kaka, S.I. and Jervis, M. [2013] Enhancement of passive microseismic eevent using seismic interferometry. *Seismological Research Letters*, **84**(5), 781–784.

Coppens, F. [1985] First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophys. Prospect*, **33**, 1212–1231.

Duncan, P.M. and Eisner, L. [2010] Reservoir characterization using surface microseismic monitoring. *Geophysics*, **75**(5), 75A139–75A146.

Earle, P.S. and Shearer, P.M. [1994] Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, **84**, 366–376.

Fischler, M.A. and Bolles, R.C. [1981] Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381–395.

Grandi, S. and Oates, S. [2009] Microseismic event location by cross-correlation migration of surface array data for permanent reservoir monitoring. In: *71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009*.

Harris, D.B. [2006] *Subspace detectors: theory*. United States. Department of Energy.

Maxwell, S., Rutledge, J., Jones, R. and Fehler, M. [2010] Petroleum reservoir characterization using downhole microseismic monitoring. *Geophysics*, **75**(5), 75A129–75A137.

Mousa, W.A., Al-Shuhail, A.A. and Al-Lehyani, A.F. [2011] A new technique for first-arrival picking of refracted seismic data based on digital image segmentation. *Geophysics*, **76**, V79–V89.

Sabbione, J.I. and Velis, D. [2010] Automatic first-breaks picking: New strategies and algorithms. *Geophysics*, **75**(4), V67–V76.