

## Short Note

# A modified Lloyd algorithm for characterizing vector fields

*Robert G. Clapp*

### INTRODUCTION

In geophysics we often encounter problems where we must balance between cost and accuracy. A specific class of these problems involve characterizing a field or a vector field with few points, as accurately as possible. An example is in Fourier-based downward continuation based migration Claerbout (1995). These methods are only accurate in a  $v(z)$  medium, but have been extended to work in variable velocity (Stoffa et al., 1990; Ristow and Ruhl, 1994; Le Rousseau and de Hoop, 1998) by using one, or several, reference velocities and then applying a correction based on the difference between the reference velocity and the true velocity at a given location. The larger the difference, the less accurate the wave propagation at high angles. As a result, choosing appropriate reference velocities can have a meaningful impact in image quality. In anisotropic wave equation migration, choosing appropriate reference velocities becomes even more problematic. Instead of a single value, velocity, at a given location, you now have multiple values (e.g. velocity,  $\delta$ , and  $\epsilon$ ) at each point. The dimensionality of the space that must be spanned, and potentially the cost, is now cubed. Efficiently spanning this 3-D vector space becomes essential.

Electrical engineers and image processors face similar problems. In speech compression, it is important to accurately describe a signal in as few bytes as possible. In image processing, it is often important to reduce the number of colors in an image with as minimal loss in image quality as possible. These problems have led to the field known as quantization. One family of method often employed is based on Lloyd's method (Lloyd, 1982), an iterative technique that allows for variable rate quantization. In previous work (Clapp, 2003), I demonstrated a modified version of the 1-D Lloyd's method for velocity selection.

In this paper I extend the ideas in Clapp (2003) to the case of vector fields. I show how the method can automatically identify a series of locations that best characterize the vector field. In a companion paper Tang and Clapp (2006) the methodology is applied to anisotropic migration problem.

## LLOYD'S METHOD

Lloyd's 1-D method attempts to identify the major features in a signal's histogram as accurately as possible, with as few points as possible. The selected points are referred to as the signal's "codebook". The methodology is iterative in nature. It starts from an initial code book, then repeats

- 1 Find for every point the closest codebook entry. These points form a cell.
- 2 Find the average of all points associated with a given codebook entry. The average becomes the new codebook entry.
- 3 Remove codebook entries that have no, or few, points associated with them.

Figure 1 demonstrates a single iteration of this approach. Steps 1-3 are repeated until the solution no longer changes significantly.

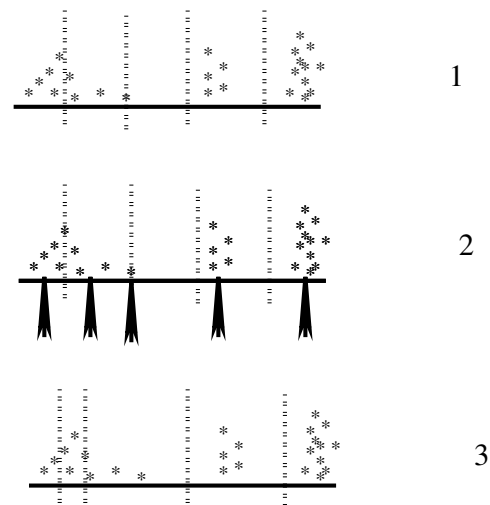


Figure 1: In step 1, the data is dividing into cells. In step 2, the average of each cell is calculated. In the third step new cell boundaries are calculated between the centroids. If a cell contains no points, the cell is removed. [bob2-ref] [NR]

Lloyd's method is attempting to solve a non-linear problem by an iterative scheme, so local minima can be problematic. To avoid prejudicing the solution the initial model is often a random set of vectors. To solve the problem, several methods have been developed (Linde et al., 1980). These methods often rely on replacing an empty cell by splitting regions with high variance. Clapp (2002) followed a modified version of this approach for the velocity selection problem. Figure 2 shows a velocity model, the selected reference velocities, and the percentage error between the two using the methodology described in Clapp (2002). The methodology proved effective in selecting appropriate reference velocities with two noticeable drawbacks. At each depth step a new non-linear estimate is performed. At some depths steps the solution gets stuck in local-minima. Note the striping in the right panel of Figure 2. In addition, the methodology will sometimes pick too few reference velocities to adequately describe a layer by getting stuck in a local minima. Both of these problems can be minimized by using as a starting solution the reference velocities at the previous depth step and adding a

random component to the cell splitting. Figure 3 shows the selected reference velocities and the percentage error in the selected reference velocities.

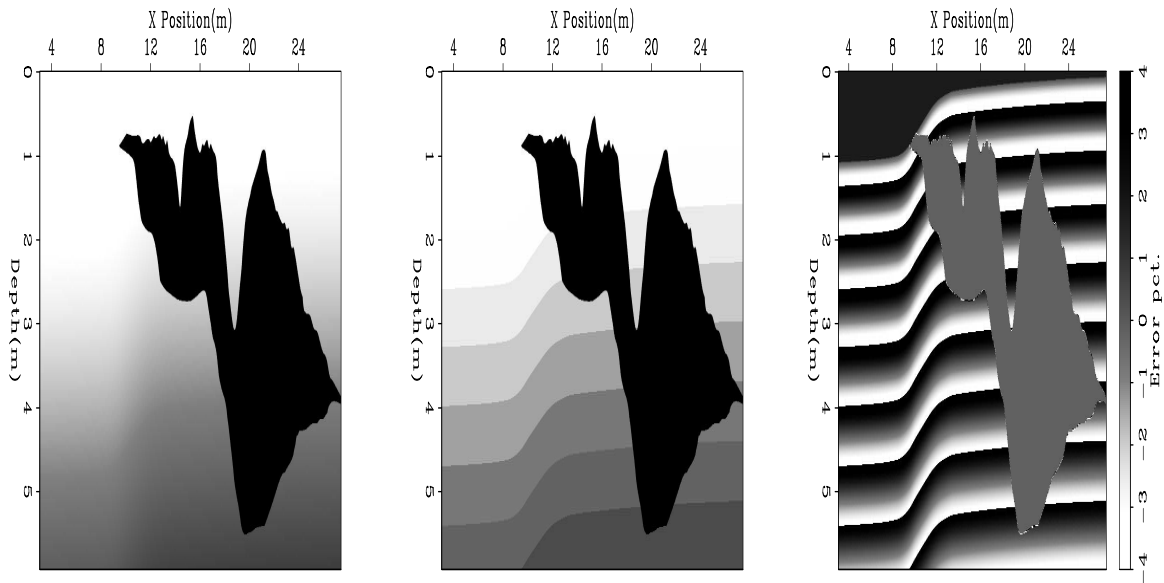


Figure 2: The left panel shows a velocity model from a 2-D synthetic. The center panel shows the reference velocities selected using the Lloyd's approach described in Clapp (2002). The right panel shows the percent error between the selected reference velocity and the the true velocity. Note the striping due to local minima. `bob2-error1` [ER]

### EXTENSION TO N-D

Lloyd's method would not lead to an interesting result for a signal that has a uniform distribution. The result would be a set of evenly-spaced points that span the range of the signal. Lloyd's method is effective for a signal that tends to cluster around a few selected values (what turns into the codebook). The velocity selection problem is effectively handled by the Lloyd approach because the earth is composed of geologic layers. The velocity range within a geologic layer tend to be relatively constant. Lloyd's algorithm automatically identifies these clusters.

In some problems we have several parameters at every model location. If these parameters are correlated, a clustering techniques such as Lloyd's approach can be effective. Extending the 1-D approach described above to a 3-D case is relatively simple:

- The codebook entries become vectors rather than scalars.
- To find the nearest point, we do a vector distance calculation.
- Instead of averaging the points attached to a given codebook, we find the centroid of the attached points.

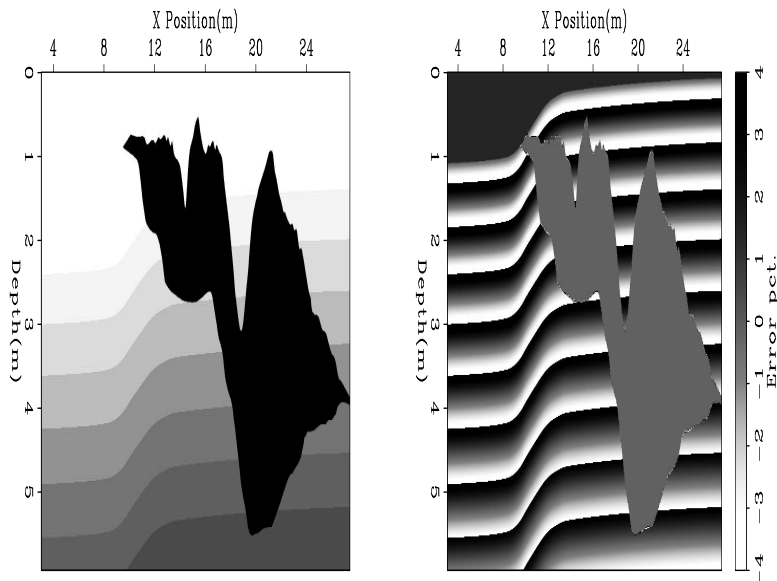


Figure 3: The left panel shows the selected reference velocity using an improved selection scheme. The right panel shows the error percentage error between the selected reference and the true velocity. Note the reduced error compared to the right panel of Figure 2. bob2-error2 [ER]

These changes are all making an assumption that the range of each element of our vector field is approximately the same. If this assumption isn't true, for example when our vector field consists of velocity (ranging from 1500-4500 m/s) and  $\delta$  (ranging from 0-1), we need to scale the  $\delta$  value so their range is approximately equal to velocity. If we do not, the  $\delta$  values will have such a small effect in our calculations that the algorithm reduces to a 1-D velocity solution.

### EXAMPLE

One example of vector quantities with spatial correlation is in migrating from topography using Riemannian coordinate systems (Shragge and Sava, 2004; Shragge, 2005). The approach attempts to migrate from a coordinate system that is consistent with the recording surface. The resulting coordinate system has therefore variable spatial density. Figure 4 shows both the scaling of the orthogonal wave number (squeezing and stretching of the coordinate system) and a velocity component for the 2-D problem.

At first glance these two fields do not show an obvious correlation. If we form a 2-D histogram (shown in Figure 5) another story emerges. The parameters are correlated, and concentrated in top portion of the histogram. Figure 6 shows the histogram from several different depths steps. The correlation and clustering of the parameters becomes even more obvious. Figure 7 shows the same depths overlaid by the N-D Lloyd codebook. Note how the histograms features are easily identified. Figure 8 shows the error percentage in the parameters

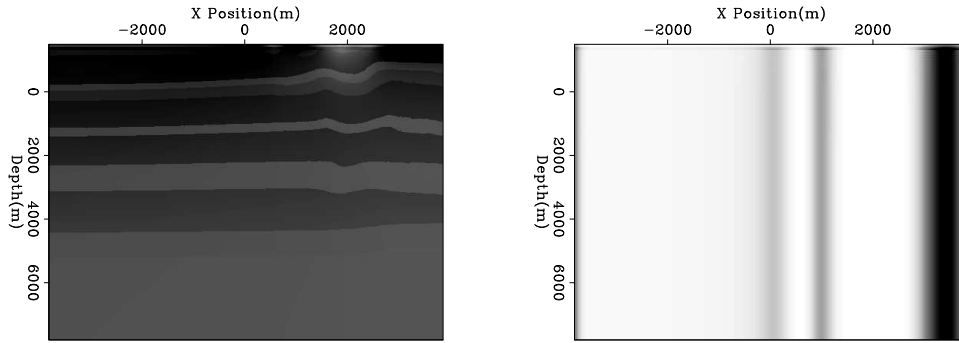
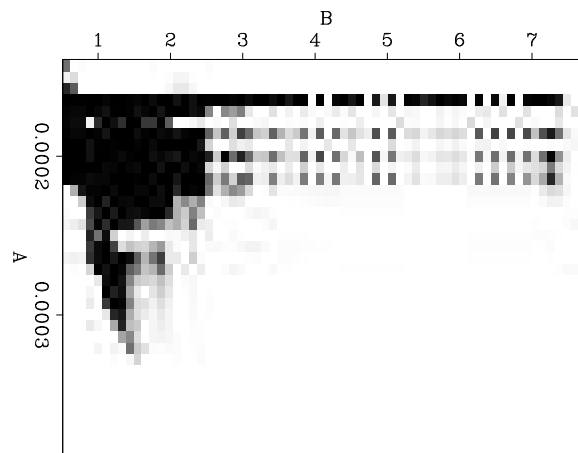


Figure 4: The left panel shows velocity field, the right panel shows the stretching of the coordinate system from Cartesian. `bob2-jeff` [ER]

Figure 5: A 2-D histogram of the parameters shown in Figure 4. The vertical axis is velocity (A), the horizontal axis is the stretching parameter (B). The amplitude shows the number of times a given (A,B) combination exists in the data shown in Figure 4. Note the correlation in the parameters. `bob2-jeff.histo` [ER]



by using the N-D Lloyd approach and by evenly partitioning the parameter range at each depth step. Not only is the evenly spaced partitioning scheme eight times more expensive, the difference between the true parameters at each model location and the selected parameters are significantly higher. Tang and Clapp (2006) use a 3-D version of Lloyd's algorithm for selected  $v$ ,  $\delta$  and  $\epsilon$  parameters for anisotropic migration.

## CONCLUSIONS

In this paper I showed a modified version of Lloyd's multi-dimensional algorithm that can effectively and efficiently characterize vector fields. The method was applied with success on several different vector fields resulting in a higher order of accuracy with fewer parameters.

## REFERENCES

Claerbout, J. F., 1995, Basic Earth Imaging: Stanford Exploration Project.



Figure 6: 2-D histograms (black and grey indicates the number of parameters with a given value) at three different depths. Note the correlation of the parameters. `bob2-jeff.histos` [ER]

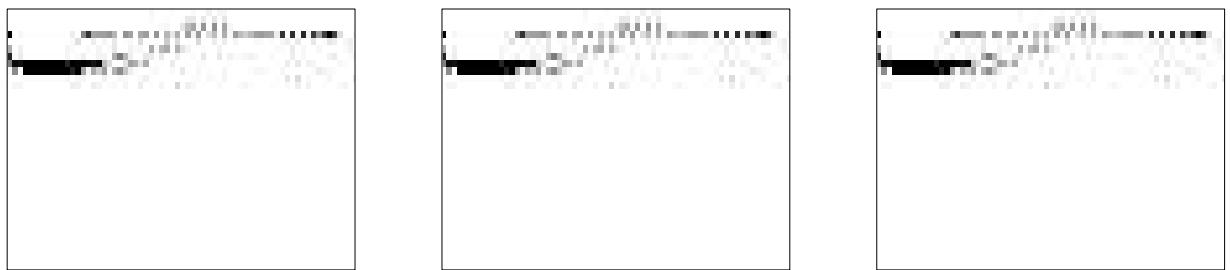


Figure 7: 2-D histograms (black and grey indicates the number of parameters with a given value) at three different depths overlaid by the Lloyd codebook (\*). Note how well the features are detected. `bob2-jeff.overlays` [ER]

Clapp, R. G., 2002, Reference velocity selection by a generalized Lloyd method: SEP-111, 215–225.

Clapp, R. G., 2003, Semblance picking: SEP-113, 323–330.

Le Rousseau, J. and M. de Hoop, 1998, Modeling and imaging with the generalized screen algorithm: 68th Ann. Internat. Meeting, Soc. Expl. Geophys., 1937–1940.

Linde, Y., A. Buzo, and R. M. Gray, 1980, An algorithm for vector quantizer design: IEEE Transactions on communications, **1**, no. 1, 84–95.

Lloyd, S. P., 1982, Least squares quantization in pcm: IEEE Transactions on Information Theory, , no. 28, 127–135.

Ristow, D. and T. Ruhl, 1994, Fourier finite-difference migration: Geophysics, **59**, no. 12, 1882–1893.

Shragge, J. and P. Sava, 2004, Incorporating topography into wave-equation imaging through conformal mapping: SEP-117, 27–42.

Shragge, J., 2005, Wave-equation migration from topography: Imaging husky: SEP-123.

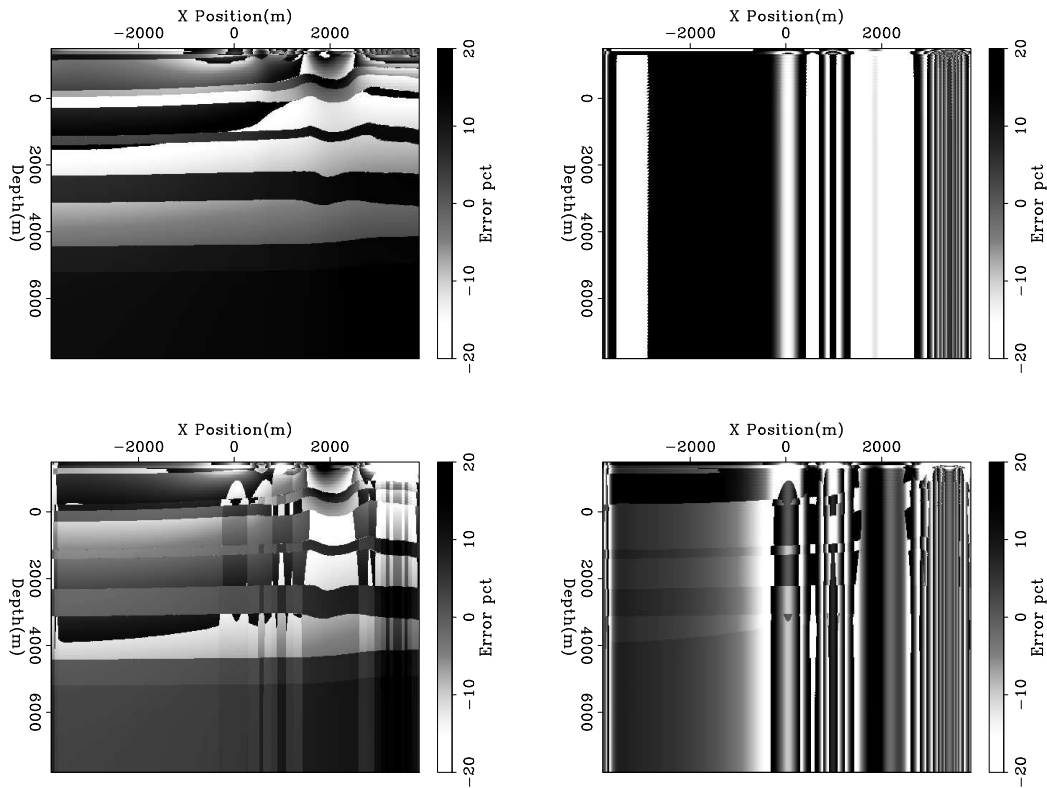


Figure 8: The percent error between the selected reference parameter and the true parameter. The top two panels show the result of using evenly spaced A,B parameters. The bottom two panels show the result of using the Lloyd approach to select the parameters. Note the significantly smaller error with the Lloyd's approach with 1/8 as many parameters. bob2-jeff.error [ER]

Stoffa, P. L., J. T. Fokkema, R. M. de Luna Freire, and W. P. Kessinger, 1990, Split-step Fourier migration: *Geophysics*, **55**, no. 4, 410–421.

Tang, Y. and R. G. Clapp, 2006, Lloyd's algorithm for selecting reference anisotropic parameters during wavefield extrapolation: SEP-124.

