Proceedings of the Project Review, Geo-Mathematical Imaging Group (Purdue University, West Lafayette IN), Vol. 1 (2011) pp. 57-64.

GEOMETRIC SEGMENTATION USING GAUSSIAN WAVE PACKETS

FREDRIK ANDERSSON* AND PETTER STRANDMARK†

Abstract. We segment seismic data based on the geometric information provided by sparse gaussian wave packet representations. We use this information to for estimates an abundance of candidate curves for the identification of the wave fronts. A graph cut methodology is then employed to reduce the number of curves such that the data is well described by only few wave fronts.

1. Introduction. We investigate the possibility to do (seismic) data segmentation using only the geometric information that sparse gaussian wave packet representation provides. Each Gaussian wave packet has information about localization, orientation, scale and frequency content. In this work, we will only make use of the localization and orientation parameters. A data set represented by Gaussian wave packets will thus be regarded as a set of points with associated directions. A crucial ingredient for this work is it is possible to obtain very sparse data representations using Gaussian wave packets, which in principle only picks out the elements needed in order to provide good data approximations. The sparse decomposition problem has been addressed in [2].

The sparse decompositions described in [2] is based on (very) redundant discrete representations. A redundant discrete representations using functions φ_γ are characterized by the fact that their *Gram matrix*, $G(\gamma_1, \gamma_2) = \langle \varphi_{\gamma_1}, \varphi_{\gamma_2} \rangle$, is invertible but with (very) high condition number. In this way we are able to represent, for instance functions which essentially have support inside two discs, one in the spatial domain and one in the frequency domain, respectively. We use variable selection techniques to select a subset of relevant functions $\{\varphi_\gamma\}_\gamma$ to use in our representation. In this process, we will have some restrictions on possible centers and orientations of the wave packets, and may therefore obtain representations that do not align exactly with the wave front sets of the data. We are then faced with the problem of detecting wave front sets given some perturbation in center location and orientation.

Given a set of points in the plane with associated directions, we want to find a set of curves such that they pass through the points and such that their tangent in normal to the wave packet orientation (i.e. so that the wave packets are aligned with the curve). A problem of simpler nature is the fitting of lines through points in $\mathbb{R}^2$. The Radon transform (Hough transform) can be a useful tool in order to detect which lines that are highly correlated with the data. A similar tool used in seismology is the parabolic Radon transform [1]; used for the detection of parabolas that are highly correlated with the data, and the usage of parabolic Radon transforms for the detection of parabolas. Since a parabola has more degrees of freedom (four in the general case and three when considered a graph at a certain orientation), it becomes more computationally heavy to compute the correlation of the data to all possible curves. A common trick is to enforce some sort of restriction on the set of parabolic curves used.

For the case where the data is decomposed using wave packets, the set of curves that needs to be investigated can be reduced substantially by using only the geometric data provided by the decomposition instead of the raw data. For instance, we can tell whether two wave packets are probable to belong to the same curve by trying to fit a quadratic polynomial to them. Since this curve have only three degrees of freedom, while there are four parameters provided by the two wave-packets, a good fit can only be obtained for some pairs of wave packets. We will use this to detect candidates of neighboring pairs. Another feature that is significant for relating neighboring wave packets is that the inner product of two wave packets decay rapidly if they are far apart from each other, either in space or in frequency, cf. [2].

By analyzing the neighborhood structure we can identify which points that should be associated with the same curves. We can construct a graph that describe the (possible) connection between

*Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden
†Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden

different pairs of wave packets. The connected components of such a graph will then contain subsets of wave packets that belong to either one or several curves. Typically, the simple process of identifying connected components will cause an undersegmentation, i.e. subsets of wave packets that belong to several different graphs. However, the initial set of all wave packets are composed into smaller sets which may be easier to work with.

Within a certain subset we can now propose a set of curve candidates. We can then use *graph cut* techniques to label the different wave packets to plausible curve candidates. Such a labeling can then be used to refine the set of curve candidates. On this new set, we can then apply a labeling using a graph cut formulation.

2. Gaussian wave packet representations. In two dimensions, a Gaussian wave packet aligned with the coordinate axes, and with center at the origin have the form

$$\varphi_{\alpha,\beta,\eta}(x) = \rho_{\alpha,\beta,\eta}\, \mathrm{e}^{2\pi i \eta x_1}\, \mathrm{e}^{-x^t \Lambda(\alpha,\beta,\eta) x}, \tag{2.1}$$

where

$$\Lambda(\alpha,\beta,\eta) = \ln(16) \begin{pmatrix} \eta^2/\alpha^2 & 0 \\ 0 & \eta^2/\alpha^2\beta^2 \end{pmatrix}$$

describes the decay coefficients. The parameter $\alpha > 0$ describes the number of oscillations within the half-width of the Gaussian and controls the shape of the wave packet, while the parameter η determines the frequency of the oscillations and also plays the role of scale parameter. The factor $\ln(16)$ is needed to make the half-width of the Gaussian equal to α oscillations with frequency η. The parameter $\beta > 0$ describes the ratio between the half-widths of the Gaussians in the x_1 and x_2 directions. The normalization factor

$$\rho_{\alpha,\beta,\eta} = \sqrt{\frac{2\ln(16)}{\pi}}\, \frac{|\eta|}{\alpha\sqrt{\beta}}$$

is chosen so that $\varphi_{\alpha,\beta,\eta}$ has unit L^2-norm.

To account for rotations and translations, we introduce the operators $T_y(x) = x - y$ and

$$\Theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

A generic Gaussian wave packet can then be written

$$\varphi_\gamma(x) = \varphi_{\alpha,\beta,\Theta^{-1}(\eta,0)^t}(\, \Theta^{-1}\, T_y(x)\,); \quad \gamma = ((y_1,y_2),\eta,\alpha,\beta,\theta). \tag{2.2}$$

The parameters that define the Gaussian wave packets have the following physical interpretation; y - Central location; η - Frequency/scale; θ - Orientation of axes; α - Number of oscillations within a half-width; β - Ratio of the half-widths along and perpendicular to the direction of oscillation.

Given some function f and some set of Gaussian wave packets φ_γ, $\gamma \in \Gamma$, we try find an approximation

$$f \approx \sum_{\gamma\in\Gamma'} c_\gamma \varphi_\gamma, \tag{2.3}$$

for some $\Gamma' \subset \Gamma$, where the number of elements $|\Gamma'|$ is as small as possible. Of coarse there is a tradeoff between the accuracy of the approximation compared to the number of terms $|\Gamma'|$ used. A description about how to make that variable selection (choose Γ') is given in [2]. In the construction of Γ we will choose the localization parameters y and the orientation parameters θ to align on discrete sets. This will naturally force the locations and orientations to deviate from the "true" wave front set due the the finite resolution in the discretization. We will use the notation $\gamma(y), \gamma(\theta)$, etc, to describe the individual wave packet parameters.

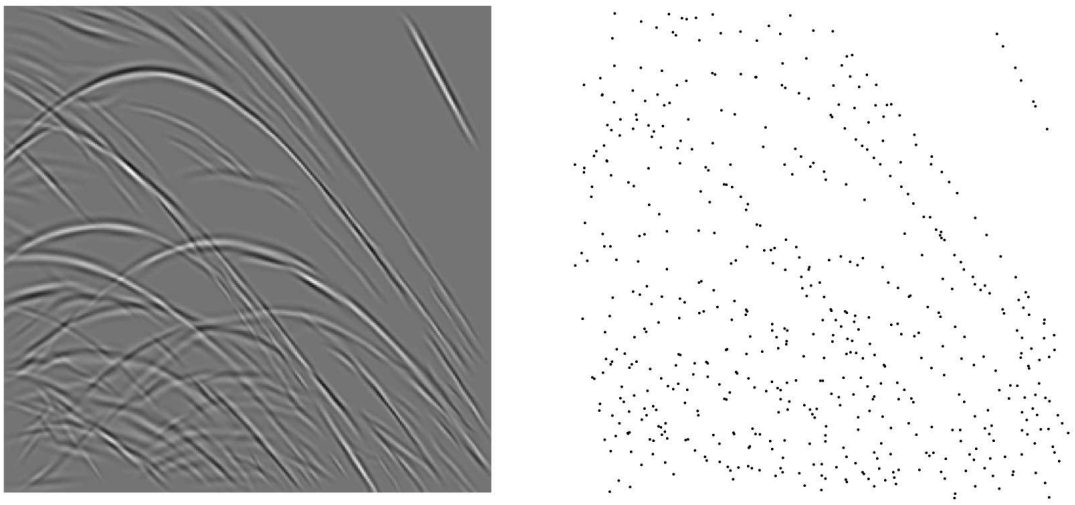


Fig. 1: (a) Representation of a data set using Gaussian wave packets; (b) wave packet locations

3. Extracting geometric information. The first step for clustering wave packets belonging to the same wave front, is to check them pair wise for parabolic consistency. By the latter we mean that given to wave packets with parameters γ and γ', we try to fit a second order polynomial through the points $\gamma_1(y)$ and $\gamma_2(y)$ such that there is a second order polynomial $p(t) = p_2t^2 + p_1t + p_0$ such that

$$p(\gamma(y_1)) \approx \gamma(y_2), \quad p'(p(\gamma(y_1)) \approx \cot(\gamma(\theta)), \quad \gamma = \gamma_1, \gamma_2.$$

The equations above are overdetermined, and hence a fit can only be obtained for wave packets that are close to being aligned on second order polynomials. A suitable rule to determine whether two wave packets could belong to the same event is by requiring that

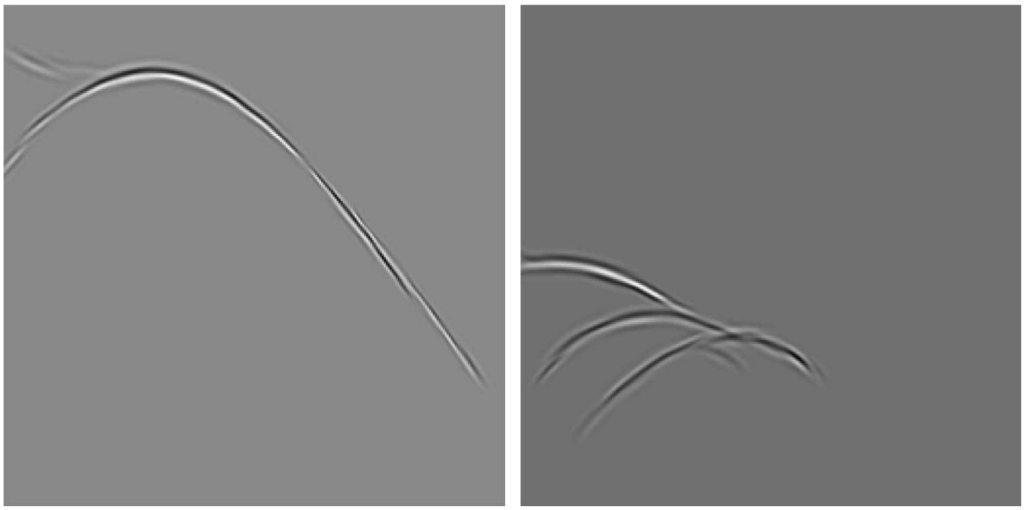
$$|p(\gamma(y_1)) - \gamma(y_2)| < \gamma(\eta)/\gamma(\alpha), \tag{3.1}$$

$$|\cot^{-1}(p'(\gamma(y_1))) - \gamma(\theta)| < \Delta\theta, \tag{3.2}$$

for $\gamma = \gamma_1, \gamma_2$, and where $\Delta\theta$ denotes the angular resolution from the discrete Gaussian wave packet transform. The first condition (3.1) says that that the approximation difference is less than the half-width of the Gaussian wave packet. There is clearly also a choice about how to weigh the position and orientation parts when doing the fit. First, there is a choice how to weigh position versus orientation. Secondly, it is natural to minimize $p'(\gamma_j(y_1))$ against $\cot(\gamma_j(\theta))$, $j = 1, 2$, rather than the corresponding angle because of linearity, but the non-uniform effect for different angles should be accounted for in the least squares fit.

In addition to checking for parabolic alignment of wave packets, we also require that they are close to each other in terms of that the inner product $\langle\varphi_{\gamma_1}, \varphi_{\gamma_2}\rangle$, i.e. by inspection of the elements in the Gram matrix. These elements decay very quickly if the wave packets are far away from each other, e.g., spatially.

In order to make an initial segmentation, we form an adjacency matrix where we indicate that two wave packets may belong to the same wave front if align parabolically, and if the corresponding element of the Gram matrix is small in absolute value. We can then look for connected components in the adjacency graph. In many cases this will provide a good initial decomposition. However, as illustrated in Figure 2 (b), this procedure may result in undersegementation.



(a) (b)

Fig. 2: Initial segmentation by finding connected components in the adjacency matrix. The right figure the undersegmentation that may happen when only considering connected components

4. Graph cuts and Energy Minimization. We let $\mathbb{B} = \{0, 1\}$. Let $E : \mathbb{B}^n \to \mathbb{R}$ be an *energy function* of the form

$$E(\boldsymbol{a}) = \sum_{1 \le j \le n} E_j(a_j) + \sum_{1 \le j < k \le n} E_{j,k}(a_j, a_k), \tag{4.1}$$

where $E_j : \mathbb{B} \to \mathbb{R}$ and $E_{i,j} : \mathbb{B} \times \mathbb{B} \to \mathbb{R}$. To minimize E is in general NP-hard, but if its components satisfies the *submodularity* condition

$$E_{j,k}(0,0) + E_{j,k}(1,1) \le E_{j,k}(1,0) + E_{j,k}(0,1), \quad \forall j, k, \tag{4.2}$$

it is possible to minimize E in polynomial time by computing the minimum cut in a graph. [5, 6]

While optimization problems with a finite number of possible values for each variable can be converted into Boolean problems by binary encoding, it is often not a preferred method as the resulting optimization problems tend to be very hard to solve. Instead, an approximate minimization method known as α-*expansion* can be employed. This is an iterative process where given some estimate $\boldsymbol{a}^{(t)} \in \{1 \dots L\}^n$ for a problem with L labels, an improved estimate $\boldsymbol{a}^{(t+1)}$ is computed. The improvement is computed by selecting a $\alpha \in \{1 \dots L\}$ and computing the optimal *expansion* with respect to α.

DEFINITION 4.1. *A vector* $\boldsymbol{a}^{(t+1)}$ *is an* α-expansion *of an image* $\boldsymbol{a}^{(t)}$ *if for every index* i*:*

$$a_j^{(t+1)} = \alpha \qquad \text{or} \qquad a_j^{(t+1)} = a_j^{(t)}. \tag{4.3}$$

The complete algorithm is to perform expansion moves iteratively for different indices until convergence:

$\boldsymbol{a}^{(0)} \leftarrow \mathbf{1}$
$t \leftarrow 0$
while *function value is improved* **do**
  foreach $\alpha \in \{1 \dots L\}$ **do**
    Set $\boldsymbol{a}^{(t+1)}$ to the optimal α-expansion of $\boldsymbol{a}^{(t)}$
    $t \leftarrow t + 1$
  end
end

An expansion for every index value is performed until the function value does not decrease for any value of α. This usually happens within a few loops through the set of indices. While reaching the global optimum is not guaranteed, it is possible to prove [3, Theorem 6.1] that the α-expansion algorithm will yield an estimate that whose energy is bounded by a constant times the minimal energy. The bound constant depends on the type of pairwise interactions between the variables, and under favorable circumstances it is equal to two.
Computing the optimal α-expansion can often be done exactly. Let $\boldsymbol{z}$ be a Boolean vector and let

$$a_j^{(t+1)} = \begin{cases} a_j^{(t)}, & \text{if } z_j = 0 \\ \alpha, & \text{if } z_j = 1. \end{cases} \tag{4.4}$$

The problem of finding the optimal expansion is with this notation a Boolean optimization problem over $\boldsymbol{z}$. With the proper assumptions on the interactions between the variables, all pseudo-Boolean problems will be submodular and the minimization problem can be solved formulated as a minimum cut problems. The graph construction by [3, Section 5] uses additional nodes where $\alpha \neq a_j^{(t)} \neq a_k^{(t)} \neq \alpha$ and $k \in \mathcal{N}_j$, although the construction can be done without extra nodes [6].

In addition to α-expansion the related *α/β-swap* [3] may also be used. Instead of considering a single value α and expanding it as much as possible, the swap considers two possible labels $\alpha, \beta \in \{1 \dots K\}$ and computes the optimal swap between the two. It can handle more general functions than expansion while still ensuring submodularity of the subproblems. The main drawback is the lack of guarantee to obtain a solution close the optimum, and the fact that each full iteration now consists of solving $\binom{L}{2} = (L^2 - L)/2$ Boolean problems instead of L.

5. The labeling problem. We describe how to formulate the following labeling problem: Given a set of candidate curves and wave packets, assign each wave packet to one curve such that it is close to that curve, while keeping the total number of curves at a minimum. This problem can be solved by using graph cut techniques.

Let $1 \leq \gamma_l \leq L$ be a set of Gaussian wave packet parameters, cf. (2.2), and let C_j, $1 \leq j \leq J$ be a set of candidate curves that describe the wave fronts of the wave packets. Let $D(j, l)$ be a cost function that describes how well curve C_j fits wave packet γ_l. For our setup, we choose the cost variables binary in such a way that the cost is zero if (3.1) and (3.2) are satisfied. In addition to the this *label cost*, we impose an additional cost for assigning different labels to wave packets that are close to each other. We use the neighboring structure described in Section 3 for this cost. Finally, we introduce a cost for the number of curves that are actually used in the labeling. The setup can be described as

$$E(\boldsymbol{a}) = c_d \sum_{1 \leq j \leq n} D(j, a_j) + c_n \sum_{k \in \mathcal{N}(j)} \delta(a_j - a_k) + \sum_{1 \leq \ell \leq K} \mathbb{I}_\ell(\boldsymbol{a}). \tag{5.1}$$

Above, $\boldsymbol{a} \in \{1 \dots J\}^L$ is a vector of assignments; i.e. wave packet γ_l is assigned to curve C_j. The neighboring cost is described above the Kronecker delta function δ. Concerning the last term, $\mathbb{I}_\ell(\boldsymbol{a}) = 1$ if there exists an $a_l = j$ and zero otherwise, and the last sum then measures the number of different curves that are being used. This is important, since often fewer models are more desirable than many (Occam's razor). In our simulations we have used $c_d = c_n = 1$. There is a

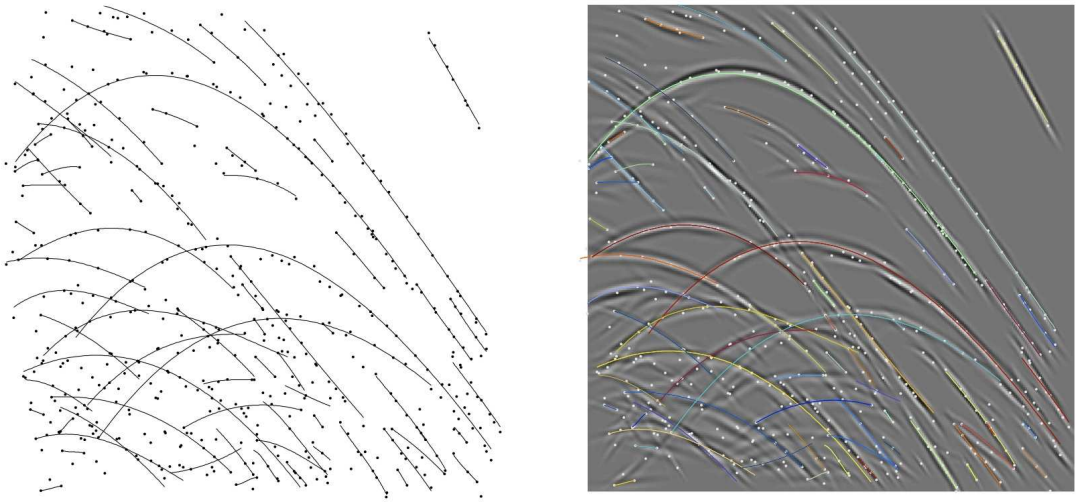


Fig. 3: (a) Wave packet locations (black dots) and estimated curves;(b) Wave packets and curves overlayed on wave packet reconstruction

clear resemblance between the functional (5.1) and that of (4.1). It is also readily verified that the submodular condition is satisfied by (5.1). It turns out that it is possible to extend the α-expansion framework to minimize (5.1). The minimizing procedure is described in detail in [4, 7].

We propose the following algorithm for the geometric segmentation using Gaussian wave packets.

1. Identify wave packets that are close to each other by fitting pairs of them to quadratic polynomials; and by requiring that their inner product is sufficiently large i magnitude.
2. Identify wave packet clusters by looking at connected components in the adjacency matrices. Estimate sets of curves using these clusters.
3. Reduce the (large) number of curves by minimizing (5.1).
4. Adjust curve estimates based on the remaining curves and repeat the previous step if necessary.

We show the result of the procedure in Figure 3. This is the result from two steps of the curve segmentation procedure above. Curves associated with single wave packets are not presented.

6. Conclusions. A method for segmentation of Gaussian wave packets along curves using only the geometric information provided by the wave packets has been presented. The method contains two principal steps. The first step is to identify neighboring wave packets by the ability to fit quadratic polynomials to pairs of wave packets along with the requirement that the inner product between the wave packets are small in magnitude. Given this information, a set of curves to fit the data set is proposed, and in the second step the number of curves are reduced by employing a graph cut technique.

7. Acknowledgments. This research was supported in part by the members, ConocoPhillips, ExxonMobil, PGS, Statoil and Total, of the Geo-Mathematical Imaging Group.

REFERENCES

[1] Brahim Abbad, Bjorn Ursin, and Milton J. Porsani. A fast, modified parabolic radon transform. *Geophysics*, 76(1):V11–V24, 2011.

[2] Fredrik Andersson, Marcus Carlsson, and Luis Tenorio. On the representation of functions with gaussian wave packets. *To appear in Journal of Fourier Analysis and Applications*, 2010.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. 23(11):1222–1239, Nov 2001.

[4] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *Conf. Computer Vision and Pattern Recognition*, 2010.

[5] P. L. (Hammer) Ivănescu. Some network flow problems solved with pseudo-boolean programming. *Operations Research*, 13(3):388–399, May 1965.

[6] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[7] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Graph cut based inference with co-occurrence statistics. In *European Conf. Computer Vision*, 2010.