Nonlinear Design of Geophysical Surveys and Processing Strategies

Thomas Guest

Thesis submitted for the Degree of
Doctor of Philosophy

The University of Edinburgh

2010

To Mum & Dad

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Thomas Guest
August 2010

Abstract

The principal aim of all scientific experiments is to infer knowledge about a set of parameters of interest through the process of data collection and analysis. In the geosciences, large sums of money are spent on the data analysis stage but much less attention is focussed on the data collection stage. Statistical experimental design (SED), a mature field of statistics, uses mathematically rigorous methods to optimise the data collection stage so as to maximise the amount of information recorded about the parameters of interest. The uptake of SED methods in geophysics has been limited as the majority of SED research is based on linear and linearised theories whereas most geophysical methods are highly nonlinear and therefore the developed methods are not robust. Nonlinear SED methods are computationally demanding and hence to date the methods that do exist limit the designs to be either very simplistic or computationally infeasible and therefore cannot be used in an industrial setting.

In this thesis, I firstly show that it is possible to design industry scale experiments for highly nonlinear problems within a computationally tractable time frame. Using an entropy based method constructed on a Bayesian framework I introduce an iteratively-constructive method that reduces the computational demand by introducing one new datum at a time for the design. The method reduces the multidimensional design space to a single-dimensional space at each iteration by fixing the experimental setup of the previous iteration. Both a synthetic experiment using a highly nonlinear parameter-data relationship, and a seismic amplitude versus offset (AVO) experiment are used to illustrate that the results produced by the iteratively-constructive method closely match the results of a global design method at a fraction of the computational cost. This new method thus extends the class of iterative design methods to nonlinear problems, and makes fully nonlinear design methods applicable to higher dimensional industrial

scale problems.

Using the new iteratively-constructive method, I show how optimal trace profiles for processing amplitude versus angle (AVA) surveys that account for all prior petrophysical information about the target reservoir can be generated using totally nonlinear methods. I examine how the optimal selections change as our prior knowledge of the rock parameters and reservoir fluid content change, and assess which of the prior parameters has the largest effect on the selected traces. The results show that optimal profiles are far more sensitive to prior information about reservoir porosity than information about saturating fluid properties. By applying ray tracing methods the AVA results can be used to design optimal processing profiles from seismic datasets, for multiple targets each with different prior model uncertainties.

Although the iteratively-constructive method can be used to design the data collection stage it has been used here to select optimal data subsets post-survey. Using a nonlinear Bayesian SED method I show how industrial scale amplitude versus offset (AVO) data collection surveys can be constructed to maximise the information content contained in AVO crossplots, the principal source of petrophysical information from seismic surveys. The results show that the optimal design is highly dependant on the model parameters when a low number of receivers is being used, but that a single optimal design exists for the complete range of parameters once the number of receivers is increased above a threshold value. However, when acquisition and processing costs are considered I find that, in the case of AVO experiments, a design with constant spatial receiver separation is close to optimal. This explains why regularly-spaced, 2D seismic surveys have performed so well historically, not only from the point of view of noise attenuation and imaging in which homogeneous data coverage confers distinct advantages, but also as providing data to constrain subsurface petrophysical information. Finally, I discuss the implications of the new methods developed and assess which areas of geophysics would benefit from applying SED methods during the design stage.

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Andrew Curtis for all the support and guidance he has given me over the past three (and a wee bit) years. When times got tough Andrew was always there to help me through them. He also throws a pretty decent party!!!

I doubt this thesis would have been completed if Emanuel Winterfors had not been willing to teach me the basics of Bayesian statistics. When the maths was starting to go over my head, Emanuel was always there to answer my questions, however trivial, on a level that I would understand.

I would like to thank NERC and Schlumberger for supporting me financially.

Of course, none of this would have ever been possible without the support of all my family. I will be forever grateful to my Mum and Dad, as I'm positive that if they had listened to the teachers at my primary school I wouldn't have passed any exams let alone reached university.

I have made some great new friends whilst in Edinburgh. Adam, Craig, David, Heather, Simon, and Mohammad - I hope you realise that you have all played an important part in the last four years and I hope we will continue to be friends wherever we all might end up. In particular I would like to thank Heather for breaking up the monotony of day-to-day life in the Attic and letting me rant about my problems.

Last, but certainly not least, I would like to thank Debs. Over the last four years she has put up with me "working from home", taking over our flat with paperwork, and generally lacking in interesting conversation. I hope she knows how lucky I feel to have had her support throughout my time in Edinburgh.

Tom Guest
10^{th} August 2010

Contents

List of Figures

Chapter 1

Introduction

Large sums of money are invested every year in geophysical surveys and experiments both by academia and industry. The principle task of all geophysical surveys and experiments is to increase our knowledge about the subsurface through the process of data acquisition and analysis so that our initial model can be refined, updated, and improved. Although the stage of data analysis is a significant area of research the data collection and planning stage is one that receives much less attention (Maurer and Boerner, 1998a; Curtis and Maurer, 2000). Since the data acquisition stage ultimately defines the information content of the data, no amount of time or money spent on improving the data analysis can compensate for lack of information.

The aim of the survey planning stage is to maximise the amount of target information we expect to record whilst also taking into consideration any physical and logistical constraints that define bounds on the types of experiments that are feasible. An ideal design would be one where the estimates on the subsurface parameters of interest are only limited by the geophysical method used and not by any inappropriate survey layout or insufficient data coverage (Maurer et al., 2010). Maximising the amount of information we expect to record often trades off with minimising the cost of the survey. For this reason optimising the design of a survey in terms of cost, logistics, and the amount of information the survey is expected to provide becomes of critical importance to maximising return on investment (Maurer and Boerner, 1998a; Curtis and Maurer, 2000). In general, geophysical survey designs are based on relatively inaccurate heuristics (rule-of-

thumb) derived from theoretical studies, repeated simulation, simple numerical models and interpretations and experience from previous surveys (Maurer et al., 2010). When heuristic design methods are used there is usually no quantifiable measure of how appropriate a chosen design may be for the given target of interest and how a survey design should be adapted for changes in logistical and economical constraints between surveys in different environments.

In this thesis I describe more rigorous methods for designing geophysical surveys. Using statistical methods it is shown how surveys can be designed taking into consideration prior geological knowledge, nonlinear mathematical relationships, complex subsurfaces and survey cost functions. In this way geophysical surveys can be designed with the explicit knowledge of the extra information expected to be recorded compared to a standard design.

1.1 Statistical Experimental Design

Statistical experimental design (SED) is a mature field of statistics that focusses on the development of methods to design experiments so as to maximise information. Typically, this is achieved by minimising the expected post-experimental uncertainties on parameters of interest whilst satisfying other necessary constraints. The first major work on SED was by Fisher (1935) in the context of agricultural field studies. During the 1950s and 1960s G.E.P. Box was one of the main authors associated with the mathematical development of SED methods (Box and Wilson, 1951; Box and Draper, 1959; Box and Lucas, 1959; Box and Hunter, 1965a,b) for the purpose of tackling applied problems. Simultaneous to the work of Box, similar studies into the mathematical and algorithm development of SED were taking place in the USSR. This work is reviewed in the book of Fedorov (1972).

SED methods were popularised by Taguchi (1987) who developed and implemented design processes to improve the quality of manufactured goods produced in Japanese industries. His main contribution to SED methods was not in their mathematical formulation but in their simplicity. Since the introduction of SED methods the main applications have been predominately applied to studies in agriculture and the biological sciences, mainly focussing on clinical trials (Atkinson and Bailey, 2001). In more recent years the use of SED methods have become

widespread with applications in research areas including (but not limited to) chemical engineering (e.g. Froment, 1975), environmental sciences (e.g. Muller and Zimmerman, 1999), oceanography (e.g. Barth and Wunsch, 1990), atmospheric sciences (e.g. Berliner et al., 1999), computer science (e.g. Belue et al., 1997), food science (e.g. Saguy and Karel, 1980), civil engineering (e.g. Mous, 1993), medical imaging (e.g. Delforge et al., 1989), and nutritional studies (e.g. Vanmilgen and Baumont, 1995). Within geophysics, where enormous sums of money are spent on data collection, formal SED theory has only been applied in a limited number of cases.

1.2 SED Methods Within Geophysics

The qualitative reason to perform SED methods to design geophysical experiments is to record data that contains more information about the subsurface than a survey designed using heuristics for the same cost, or alternatively to record the same amount of information for a lower overall cost. Figure 1.1 shows this relationship qualitatively. Both the heuristic design (dashed line) and the optimal design (bold line) incur an initial startup cost before any data is acquired. If, for example, a surface seismic experiment is performed these costs would reflect the costs of equipment purchases and hiring of field crews etc. If a fixed number of geophones are used and therefore a constant fixed cost may be assumed then Figure 1.1 shows that the optimal survey records more information about the subsurface than a standard design. Alternatively if a specific set of information is needed from the survey then using the optimal design over the standard design might require less capital investment. Both surveys however show the concept of "diminishing returns" where as more money is invested the relative gain in returns is reduced, as shown by the reduction in the gradient as the costs incurred increase. For a design based on heuristics it is impossible to predict exactly at what point the return on investment falls below a certain threshold. For a survey designed using SED methods this threshold is far easier to calculate and therefore surveys can be designed which incorporate a specific cost function to ensure maximum return on investment.

In the following section the results and conclusions of the major and prominent areas of research into applying SED methods in geophysics are presented. The

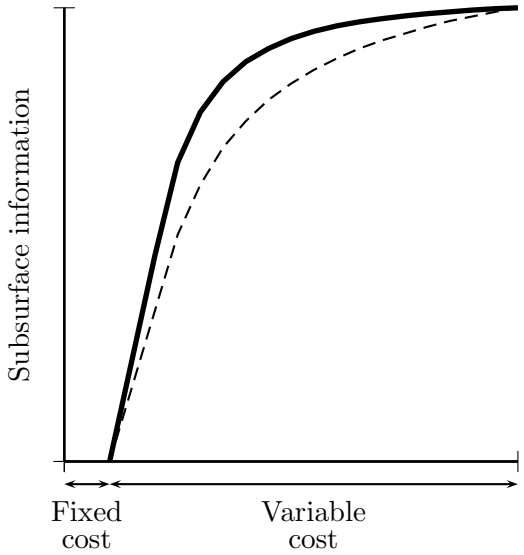

Figure 1.1: Information acquired during an experiment as a function of cost for a standard design (dashed line) and an optimal design (bold line). Both surveys incur the same fixed startup costs but subsequently the optimal design provides more information for a set cost, or the same amount of information for less cost. Both surveys exhibit the concept of "diminishing returns" where for an additional increase in cost the amount of additional information returned decreases.

mathematical frameworks and algorithms developed in the works cited are omitted in the descriptions. Relevant methods are presented in Chapter 2.

1.2.1 Earthquake Monitoring Surveys

An optimal seismicity monitoring network should provide the best possible precision for locating earthquake hypocentres. The first use of statistical experimental design methods within geophysics were the works by Kijko (1977a,b) in which it was shown that planning a distribution of seismic stations in complex situations by intuition alone is either impossible, or if networks are planned heuristically the results can lead to highly sub-optimal configurations.

Kijko (1977a) introduced an algorithm for finding the optimum positions of n seismic stations to best locate earthquake focus coordinates within a specified region. The algorithm required as *a priori* information the potential seismicity of different areas of the region and a subsurface velocity model so that travel

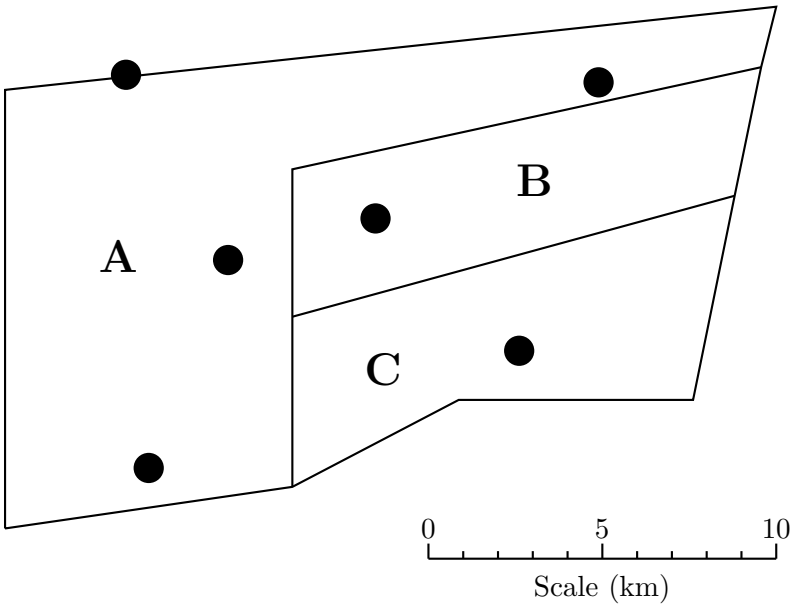

Figure 1.2: Optimum distribution of seismic stations (black circles) for optimally locating earthquake focus coordinates in the Lublin Coal Basin. The prior knowledge of seismicity of the region is divied into three areas. 'A' represents the area of highest seismicity as it includes several faults, 'C' has the lowest seismicity, and area 'B' is an area of intermediate seismicity. (Adapted from Kijko (1977a)).

times from potential earthquakes to receiver positions could be calculated. Kijko (1977a) applied the algorithm to the Lublin Coal Basin in Poland to find the optimum distribution of six surface seismic stations. Figure 1.2 shows the resulting distribution of seismic stations for the Lublin Coal Basin scenario. The prior information was split into three areas of differing expected seismicity. Area 'A' represents a region containing known faults and therefore has a high expected associated seismicity. Area 'C' represents a region of the thinnest coal seams and an associated low seismicity value, and area 'B' represents an intermediate area. Kijko (1977a) created a highly flexible method that could be used in any geographic location where, *a priori*, the information about the seismicity of the area could be defined and a velocity model assumed.

After the publication of Kijko (1977a,b) several studies used the proposed algorithms (or variants thereof) to design optimal networks at a regional scale (e.g Ghalib et al., 1984) and networks at a more local scale (García-Fernandez et al., 1988), in this case to monitor seismic activity attributed to Mount Teidi, Tenerife. The work of Kijko (1977a,b) was later extended by Kijko and Sciocatti (1995) to

plan optimal distributions of seismic stations in mines, thus increasing the spatial design space from two dimensions to three.

Whereas the work of Kijko (1977a,b) focussed on methods that could be used to calculate an optimal seismic network given a prior distribution of seismic activity, Rabinowitz and Steinberg (1990) introduced an algorithm to locate the optimal distribution of multiple seismic receivers for a single seismic source with known location. Rabinowitz and Steinberg (1990) assessed how the optimal distribution would differ from intuitive designs, and how efficient algorithms for determining optimality could be used to design optimal seismic station networks. Like Kijko (1977a,b), Rabinowitz and Steinberg (1990) used variations of D-optimality (see Section 2.3.2) to optimally design the receiver network. Rabinowitz and Steinberg (1990) extracted the following characteristics from their results:

1. All receivers should be located on concentric circles around the epicentre, with stations on each circle equidistant from each other.
2. For a two layer model with a hypocentre in the halfspace a four-receiver optimal design places a single station over the hypocentre and the remaining three stations at equidistant points on a circle about the epicentre.
3. For a hypocentre in the layer above the halfspace an optimal design locates a receiver at the epicentre and at least three receivers equidistant on each of two concentric circles about the epicenter. The radii of the two circles are such that waves arriving at the near circle are direct waves and the waves arriving at the far circle reach the stations via refraction paths. The use of a single circle for a source in the top layer results in an inefficient design.

To construct the optimal networks the use of the DETMAX algorithm (see Section 2.4.1) was used which works by adding new optimally located receivers to, and removing non-optimal receivers from, an existing network so that the final design represents as well as possible the globally optimal design rather than a locally optimal design. The algorithm also allows additional optimal receivers to be added to an already deployed network designed using non-SED methods.

Figure 1.3 shows three of the main results from Rabinowitz and Steinberg (1990). For each setup, receivers were only allowed to be located on grid nodes, the earthquake epicentre is represented by a star (*) and the optimal receivers by circles.

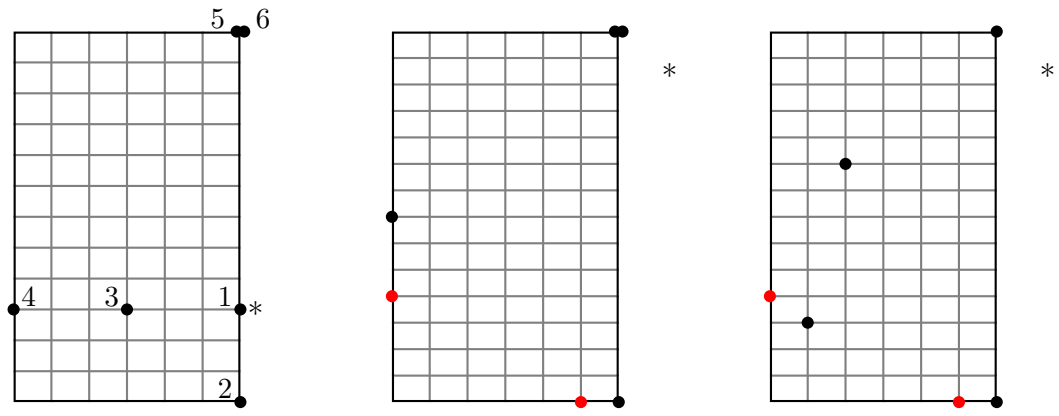

Figure 1.3: Optimal six receiver seismic network distributions as calculated using the Rabinowitz and Steinberg (1990) modified DETMAX algorithm. Plot (a) shows the optimum distribution of receivers for an epicentre located at the edge of the design region. Plots (b) and (c) show how the optimum distribution changes when recording errors are uncorrelated (plot (b)) and correlated (plot (c)). In each plot receivers can only be located at grid nodes within the 30x70 km design region. Receivers are shown as circles. In plots (b) and (c) black circles represent receivers that record direct arrivals and red circles represent receivers that record refracted arrivals. The earthquake epicentre is represented by a stars (*). (Adapted from Rabinowitz and Steinberg (1990)).

In each case a four layer homogeneous subsurface model is used. Receivers 1, 2, and 3 in Figure 1.3 (a) would be chosen intuitively by a seismologist, especially receiver 1 which is located as close to the hypocentre as possible. Receivers 4 and especially 5 and 6 would intuitively by labeled as non-optimal. In this case locating two of the receivers at a single position is chosen because the assumed errors are not correlated. Figure 1.3 (b) shows results for uncorrelated errors for a different source location and again shows two receivers being placed at the same location. The black circles represent receivers that record a direct first arrival and the red circles receivers that record a refracted first arrival. As indicated above recording both direct arrivals and refracted arrivals can make an improvement to the resolution of the network, a phenomenon automatically recognised by the DETMAX algorithm, but might be overlooked by a seismologist working on intuition alone. Figure 1.3 (c) shows the results when errors are correlated and therefore placing receivers at the same location provides no extra information about the earthquake epicentre. A similar configuration to Figure 1.3 (b) is seen but with the error correlation added, the receivers are more spread over the

design region. The two receivers recording refracted arrivals remain, similar to case (b).

Using the modified DETMAX algorithm of Rabinowitz and Steinberg (1990), Steinberg et al. (1995) expanded the work to account for multiple sources, since seismic sources are often typically distributed along fault lines, and most networks are designed to monitor systems of faults rather than specific point sources (Steinberg et al., 1995). The resulting algorithms could then be applied to global, regional, and local network design. In expanding the work of Rabinowitz and Steinberg (1990), Steinberg et al. (1995) found that the algorithm developed by Kijko (1977a,b) was flawed and would add extra receivers to monitor sources that already had good coverage and were therefore the most resolved sources, and would ignore sources that were difficult to monitor. In contrast the method of Rabinowitz and Steinberg (1990) guaranteed that the optimal network would provide minimal coverage for all potential sources.

Figure 1.4 shows an optimal network configuration using the Steinberg et al. (1995) algorithm when three point sources are located in the top layer of a four layer velocity model. The point sources are represented by stars (*) with the circles indicating the optimal locations of the receivers. The plot shows results for a six-station network (blue circles), a eight-station network (red circles) and a ten-station network (green circles). The receivers, as in the Rabinowitz and Steinberg (1990) examples, can only be located at specific locations represented here by the grid nodes. The results show some of the characteristics proposed by Rabinowitz and Steinberg (1990) with receivers surrounding the sources (where possible) and although not explicitly shown, several of the receivers record refracted waves as well as direct waves. Figure 1.5 shows the optimality measure of the overall design (thick solid line) and the optimality measures for each of the individual sources (thin lines). As expected the addition of more receivers results in larger overall optimality measures with each individual source measure also increasing indicating that the algorithm does not favour any one source. Since source 3 (dotted line) is located within the design region the algorithm is able to surround the source with receivers resulting in a larger optimality measure than the other two sources which are located outside the design region.

Steinberg et al. (1995) found that using only a few point sources was enough to represent a fault: adding many more sources did not significantly improve results

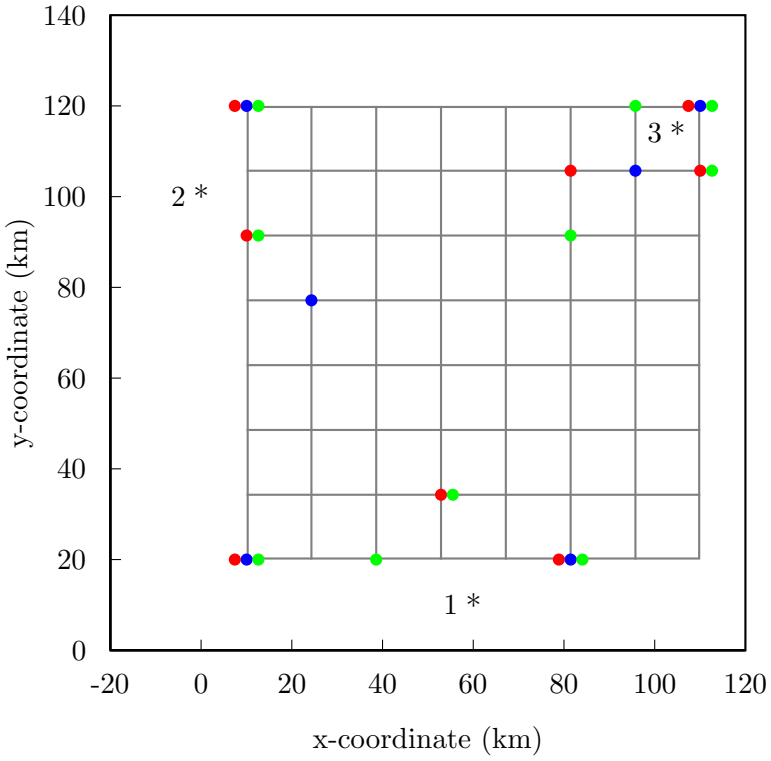

Figure 1.4: Optimum network of seismic stations for monitoring three point sources (indicated by stars (*)) and the possible locations for seismic stations are on the nodes of the gray grid. The chosen sites are indicated by circles with the optimal six locations shown in blue, eight locations shown in red and ten locations shown in green. (Adapted from Steinberg et al. (1995)).

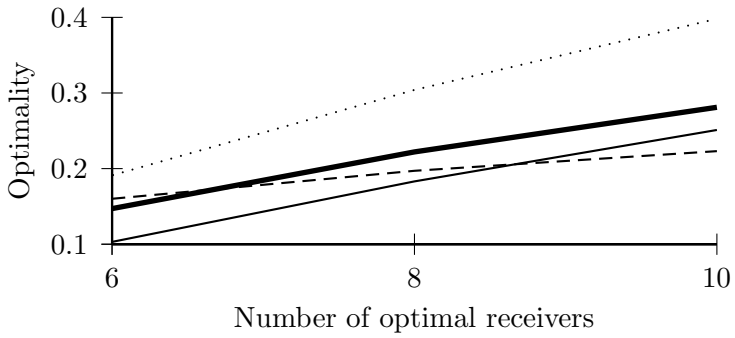

Figure 1.5: D-optimality measures for the receiver networks shown in Figure 1.4 as a function of the number of receivers used. The thick solid line represents the overall optimality measure whilst the thin solid line represents the optimality of the network for resolving the location of source one, the dashed line source two, and the dotted line source three.

but using fewer sources did significantly decrease the computational time required to find the optimal design. To conclude their work, Steinberg et al. (1995) used the methods and algorithms developed to suggest how from an already deployed seismic network, which receivers should be removed if funding were to be reduced so that only a subset of receivers could be maintained. By modelling areas using multiple point sources, the optimality measures (as in Figure 1.5) of the viable designs can be considered so that the optimal network configuration given the cost limitations can be deduced.

In 1996 The Comprehensive Test Ban Treaty (CTBT), an international agreement banning nuclear tests, was introduced with seismic monitoring being the most important method used to verify compliance with the treaty. If an event is suspected to be in violation of the CTBT then a mobile seismic network is allowed to be deployed to provide much more accurate estimates of the epicenters as well as event depths. Rapid development of a optimal network is critical to achieve high resolution power for hypocentre locations (Rabinowitz and Steinberg, 2000). The best discriminator to decide whether a source is nuclear in origin or natural is to calculate the depth of the hypocentre. The works of Bartal et al. (2000); Rabinowitz and Steinberg (2000); Steinberg and Rabinowitz (2003) showed how the DETMAX algorithm could be altered so that the optimal design is most sensitive to estimating the source depth. By using such methods a mobile network can be quickly designed and deployed to best constrain the hypocentre in order to constrain the type of source.

1.2.2 Resistivity Surveys

Resistivity methods are based on the use of two pairs of surface electrodes. The first pair introduce a current into the subsurface and simultaneously the potential gradient is measured across the other pair. Each potential gradient measure gives insight into the electrical properties of the subsurface materials. Given that the electrode pairs are located on the surface the measurements are more sensitive to the properties of shallow subsurface materials than deeper materials (Reynolds, 1997). In general the measurements recorded with larger electrode spacings contain more information about the deeper subsurface than recordings taken with smaller electrode spacings. It is thought that the use of a large number of measurements with different electrode spacings and configurations will lead to

a more accurate representation of the subsurface when the data are inverted (Furman et al., 2003).

With the introduction of versatile and multielectrode acquisition systems (e.g. Griffiths et al., 1990) and efficient inversion schemes (e.g. Loke and Barker, 1996a,b) it has become possible to perform resistivity surveys without having to use one of the standard electrode array designs. However, for n equally spaced colinear electrodes there exist

$$nd = n \times (n-1) \times (n-2) \times (n-3)\,/8 \tag{1.1}$$

non-reciprocal, four-point electrode configurations (Noel and Xu, 1991). For a scenario consisting of 50 potential electrode locations there are over 690,000 possible configurations. The recording and subsequent inversion of such a comprehensive dataset is impractical.

The first work to use statistical methods to identify datasets that would record optimal data given *a priori* geological information was by Cherkaeva and Tripp (1996). The method (based on work previously performed in the biomedical sciences) involved adjusting the intensity distribution of injected currents to maximise the response of a given target. Maurer and Boerner (1998b) introduced an alternative method to determine optimal designs for electromagnetic sounding surveys by formulating the design as a global optimisation problem and solving it using a genetic algorithm (see Section 2.4.1). Maurer et al. (2000) showed how similar techniques could be applied to design resistivity surveys, however, the computational requirements of performing such a global optimisation prevented these methods from being used for realistic problems when large numbers of possible electrode locations are considered.

Stummer et al. (2004) introduced the first design procedure to take full advantage of the capabilities of modern resistivity acquisition methods. The algorithm is initiated by recording a small dataset using a standard electrode configuration (e.g. dipole-dipole array, Wenner array). The remaining possible configurations are ranked according to a quality function which measures which new configurations will provide the most information. The quality value is essentially a measure of how sensitive the specific electrode configuration being tested is to a small perturbation to the subsurface model parameters. The quality is weighted to account

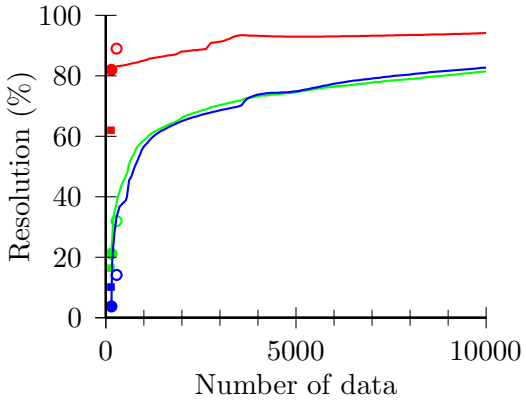

Figure 1.6: Resolution measure as a function of the number of data measurements for depth ranges of 2.1m to 3.4m (red), 15.4m to 18.4m (green), and 25.7m to 30.0m (blue). The solid lines represent the results from the addition of optimal data measurements. The filled circles represent the result for a 147-point dipole-dipole experiment, the open circle a 282-point Wenner-dipole-dipole experiment, and the square a 135-point Wenner experiment. (Adapted from Stummer et al. (2004)).

for the sensitivity decrease with depth so that the best electrode configurations do not focus on the parameters representing the shallow section of the model but are equally sensitive to all sections of the subsurface model. The new electrode configuration is only accepted if previously recorded data measurements have not been sensitive to the same sections of the subsurface. In this way optimal configurations that are added are most sensitive to the given model and also sample the entire model.

Figure 1.6 shows how well the subsurface model is resolved using different numbers of electrodes and designs compared to using the comprehensive data set calculated at three subsurface depths. The solid lines represent the results when additional electrodes are progressively added to the 147-point dipole-dipole experiment using the ranking algorithm proposed by Stummer et al. (2004). Results for a depth range of 2.1m to 3.4m (shown in red) show that a small Wenner-dipole-dipole configuration produces better results than an optimal configuration using 2500 data points which enforces the heuristic result that a Wenner-dipole-dipole configuration is best suited to resolving shallow sections of the subsurface (Reynolds, 1997). The results for a depth range of 15.4m to 18.4m (shown in green) and 25.7 to 30.0m (shown in blue) best show the advantages of using the optimisation algorithm. In the depth range 15.4m to 18.4m an optimal data set

of 282 data measurements has a resolution value of 40% compared to 21% for a Wenner-dipole-dipole array of the same size. For the depth range 25.7m to 30.0m the advantages of using the optimisation algorithm are even clearer. An optimal 282 data point set results in a 34% resolution measure in comparison to 14% for the standard Wenner-dipole-dipole array. These results show that the algorithm proposed by Stummer et al. (2004) accurately selects the optimal data sets to add to the original dipole-dipole configuration and that these data sets are being selected to maximally increase the resolution measure at all depths and not just at near surface depths. The results can also be used to assess the trade-off between increasing survey complexity and the associated increase in cost. Figure 1.6 shows that the addition of data measurements results in a large resolution increase until approximately 2000 data measurements are used. At this point the rate decreases and the addition of extra data measurements has a smaller increase effect on the resolution. The curve can therefore be used to estimate at which point the cost of acquiring more data outweighs the increase in information provided.

Nyquist et al. (2007) assessed the validity of using the Stummer et al. (2004) algorithm by performing a field comparison of 2D resistivity data collected using both the traditional dipole-dipole array and an optimised array over a karst terrain. The field tests showed that the two arrays imaged the shallow karst equally well as expected since the optimal designs are most effective at improving the resolution of deeper targets. The optimal design was however, able to resolve a crosscutting bedrock fracture system (confirmed by drilling) which was not visible in the conventional dataset. The optimal dataset required three times as many measurements per line and therefore Nyquist et al. (2007) concluded that optimal arrays are only preferable over classic arrays when resolution improvement at depth is more important than rapid data collection.

Wilkinson et al. (2006) later proposed two new optimisation strategies based on the work of Stummer et al. (2004). The first method selected new electrode configurations by calculating the resulting increase in resolution for each candidate configuration, a method that is highly computationally demanding. The second approach is more closely related to the original quality function of Stummer et al. (2004) but results in an increased resolution whilst simultaneously adding configurations that are sensitive to sections of the model which have not been previously sampled. The results showed that their first method produced the

closest to optimal subsurface resolution but that the algorithm is very slow and scales unfavourably with model complexity. The second method, as with the original Stummer et al. (2004) algorithm produced results an order of magnitude faster than the first method and results in resolution values much closer to the first method than the Stummer et al. (2004) method (Wilkinson et al., 2006). Wilkinson et al. (2006) concluded that their first method would only be suitable in the preparation stages of a survey, whereas the second method is fast enough to be used for real-time array optimisation. Wilkinson et al. (2006) also proposed that the second method could therefore be used in a time-lapse setting were inversion solutions could be used to weight sections of the subsurface which require a higher density of data sampling and conversely disregard sections of the subsurface model that are of little interest.

Using the methodologies developed in Furman et al. (2004), Furman et al. (2007) presented a practical method in which data provided from other geophysical methods can be incorporated into designing an optimised time-dependent resistivity survey. Furman et al. (2007) gave the example where continuous monitoring using a small resistivity dataset in conjunction with water sampling in a borehole could trigger a more complete optimised dataset to be recorded if a specific signature in the low resolution dataset is found. The combination of the low resolution resistivity dataset and the borehole measurements would allow the algorithm to focus attention onto subsurface areas of interest which have been seen to have the largest spatial and temporal changes.

One of the main drawbacks of using several of the mentioned optimisation algorithms to determine the best electrode configuration is that they are computationally expensive and are therefore too slow to be practical for many real problems (Coles and Morgan, 2009). The latest development in optimising resistivity surveys by Coles and Morgan (2009) is the use of a greedy algorithm. Critically, update formulae were introduced that minimise the computational expense in a greedy, sequential framework. Whereas the DETMAX algorithm can both add and remove elements to the design space the greedy method of Coles and Morgan (2009) only allows the addition of new elements. The main advantage of such a method is that it executes far more quickly than global optimisation algorithms. The results of the method introduced are shown in Figure 1.7. The gradient of the line passing through the points of the greedy algorithm is far steeper than that of the global algorithm indicating that the quality of the surveys designed using

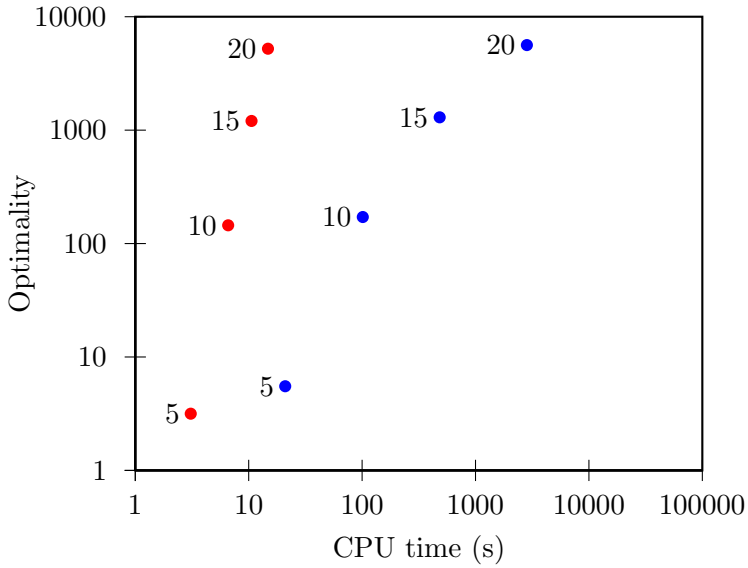

Figure 1.7: Optimality as a function of computational time for the greedy algorithm of Coles and Morgan (2009) (red) and a commonly used global optimisation algorithm (blue). Numbers next to the circles indicate the number of observations for which the experiments has been designed. (Adapted from Coles and Morgan (2009)).

the greedy algorithm increase more rapidly per unit computational time than the standard global optimisation algorithm. The plot also shows that the disparity in the optimality measure reduces as the number of observations increases, leading to the conclusion that the larger the experiment the closer the greedy algorithm becomes to the globally optimal result. Coles and Morgan (2009) concluded that resistivity surveys using their algorithm could be designed in real time.

1.2.3 Seismic Borehole Tomography

Borehole seismic tomography involves the measurement and inversion of the travel times of seismic raypaths between two boreholes. One borehole houses the seismic sources and the other the receivers. Measurement of arrival times for each shot, at each position in the receiver borehole, results in a network of overlapping raypaths which can then be inverted in order to constrain an image of the seismic velocity of the intervening subsurface.

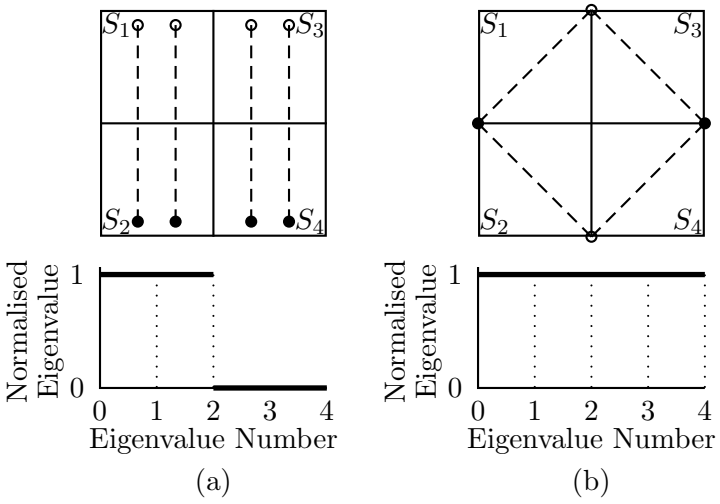

Figure 1.8: Comparison of the eigenvalue spectra for calculating the constant slownesses $(S_1 \dots S_4)$ within each cell given the traveltime data along each of the dashed paths. Twice as much independent information is given by using the raypath geometry in experiment (b) then in design (a) which is shown in the corresponding eigenvalue plots. (Adapted from Curtis (1999a)).

The location of the sources and receivers will therefore have a large impact on which areas of the subsurface are best constrained. For example, if all sources and receivers are located at the bottom of the boreholes then the velocity structure near the surface will be unresolved in the inversion. Currently the most commonly used experiment designs place regularly spaced sources and receivers down boreholes with no sources or receivers located on the surface (Curtis, 1999a).

Curtis (1999a) introduced the problem of optimising cross-well designs using a simple experiment and an information measure based on the corresponding eigenvalue spectrum (see Section 2.3.2). The aim of the experiment shown in Figure 1.8 is to design a source (filled circles) and receiver (open circles) geometry so that the traveltime data recorded along the raypaths (dashed lines) can be inverted to provide the maximum information about the four constant slownesses $(S_1 \dots S_4)$. The geometry in plot (a) only produces two independent pieces of information, the average slownesses $(S_1 + S_2)/2$ and $(S_3 + S_4)/2$. The differences in the slownesses $(S_1 - S_2)$ and $(S_3 - S_4)$ can not be constrained and therefore the individual slownesses can not be constrained. The amount of information provided is represented by the corresponding eigenvalue spectrum of the inverse problem. The two positive eigenvalues represent the fact that two independent

pieces of information are resolvable and the two zero values show that two pieces of information cannot be resolved. Plot (b) shows a geometry in which each slowness can be uniquely constrained which is shown by the corresponding eigenvalue spectrum.

Curtis (1999a) defined a vector ($\mathbf{\Theta}$) containing five design quality measures each based on a measure of the eigenvalues (λ_i).

$$
\begin{aligned}
\Theta_1 &= \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_1} \quad \text{(it can be shown that all } \lambda_i \geq 0) \\
\Theta_2 &= \log \lambda_k \quad \text{for pre-defined fixed index } k \\
\Theta_3 &= k \quad \text{such that } \lambda_k > \delta \text{ for some pre-defined tolerance } \delta \\
\Theta_4 &= \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta} \\
\Theta_5 &= \sum_{i=1}^{N} \gamma_i \quad \text{where} \quad \gamma_i = \begin{cases} \log \lambda_i & \text{if } \lambda_i \geq \delta \\ \text{penalty} & \text{if } \lambda_i < \delta \end{cases}
\end{aligned}
$$

It is assumed that the eigenvalues are listed in order of decreasing magnitude. The quality measure Θ_1 was first introduced in Curtis and Snieder (1997), Θ_2 in Barth and Wunsch (1990), Θ_4 in Maurer and Boerner (1998b), and Θ_5 in Kijko (1977a) and Rabinowitz and Steinberg (1990) and was used as the quality measure for optimising earthquake monitoring surveys (Section 1.2.1). The damping parameter (δ) in measures $\Theta_3 \ldots \Theta_5$ reduces the sensitive to eigenvalues with magnitude approximately less than δ. Θ_1 is the most efficiently calculated measure requiring approximately N^2 operations to calculate compared to approximately N^3 operations to calculate the other measures where N is the number of model parameters (Curtis, 1999a).

Using the five measures and a global optimisation algorithm, an optimal experimental design can be found by maximising the optimality measure

$$\Phi = \sum_{i=1}^{5} \beta_i \Theta_i \tag{1.2}$$

where β_i are weights such that the design measure gives some desired sensitivity to the eigenvalue spectrum or designed to reduce the computational time to reach a solution.

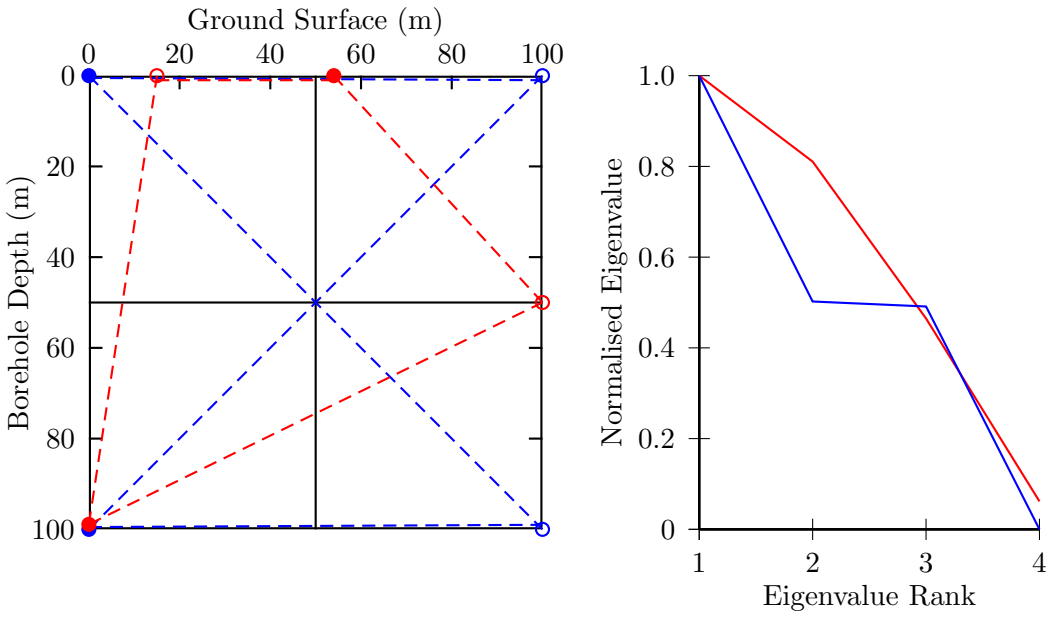

Figure 1.9: (a) Raypaths for a commonly used cross borehole experiment (blue lines) and the raypaths for the optimal experiment (red lines) found by maximising Θ_1 allowing sources and receivers to be placed in either boreholes and on the ground surface. (b) Normalised eigenvalue spectrum for the standard survey design (blue) and the optimal design (red). (Adapted from Curtis (1999a)).

Figure 1.9 shows a standard survey design and an optimal survey design with the corresponding quality measures for a two source and two receiver cross-borehole experiment. In this case the optimised quality measure is equal to Θ_1 with the sources and receivers permitted down the boreholes and on the ground surface. Although five quality measures were presented, Curtis (1999a) focused on the results produced by Θ_1 as it represents the most computationally efficient measure to calculate. Figure 1.9 (a) shows the standard design in blue and the optimal design in red. Filled circles represent source locations, open circles receiver locations and the dashed lines the raypaths. The optimal design indicates that both a source and a receiver should be placed on the ground surface to record maximum information. The corresponding eigenvalue plot (Figure 1.9 (b)) shows that the optimal design removes the zero eigenvalue of the regular design and greatly increases the second eigenvalue, showing that in this case all four velocities may be constrained.

Figure 1.10 illustrates the optimal result found for a cross-borehole survey when

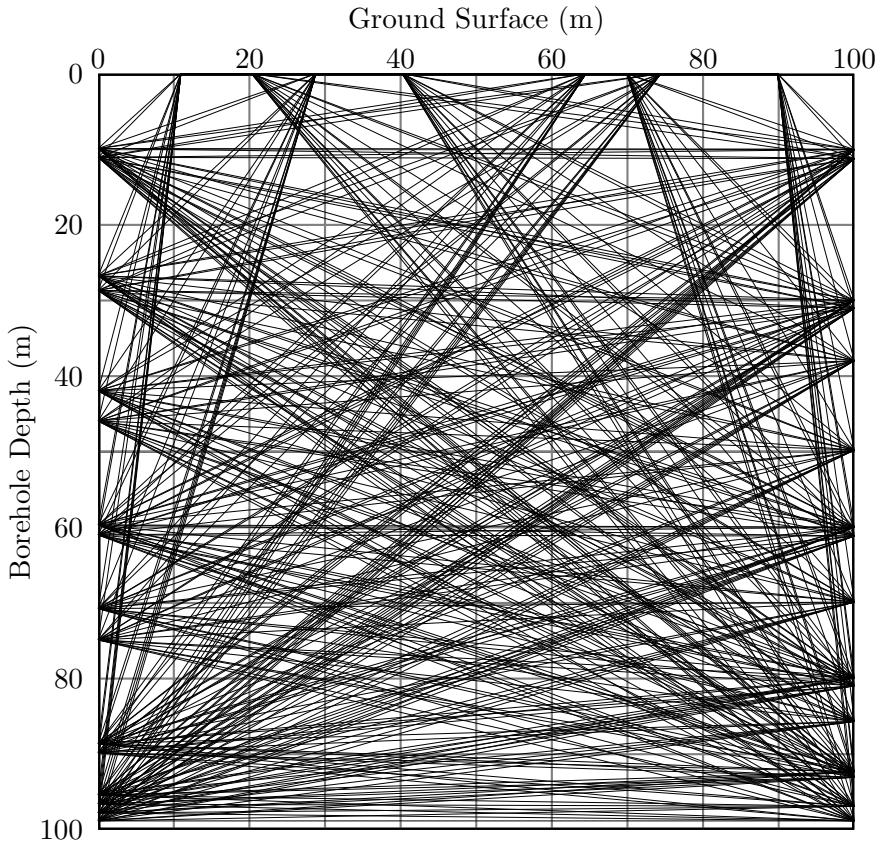

Figure 1.10: Optimal design calculated by Curtis (1999a) for a cross-borehole survey discretised into 100 equal sized cells. The black lines represent raypaths linking optimal sources and receivers. The optimal design places both sources and receivers on the ground surface. (Adapted from Curtis (1999a)).

placing twenty sources and receivers when maximising Θ_1. The subsurface area between the borehole is divided into 100 equal sized cells and the sources and receivers are allowed to be placed down both the boreholes and on the ground surface. Since the final design is dependant on the starting design and the optimisation algorithm (see Section 2.4.1) the optimal survey design shown may not represent the globally optimal design. However, Curtis (1999a) extracted several attributes common to all designs which gave high-quality measures;

1. Placing sources and receivers at the surface significantly improves designs.
2. Source and receiver densities increase steadily down the length of each borehole.
3. Surface source and receiver densities increase slightly around the central

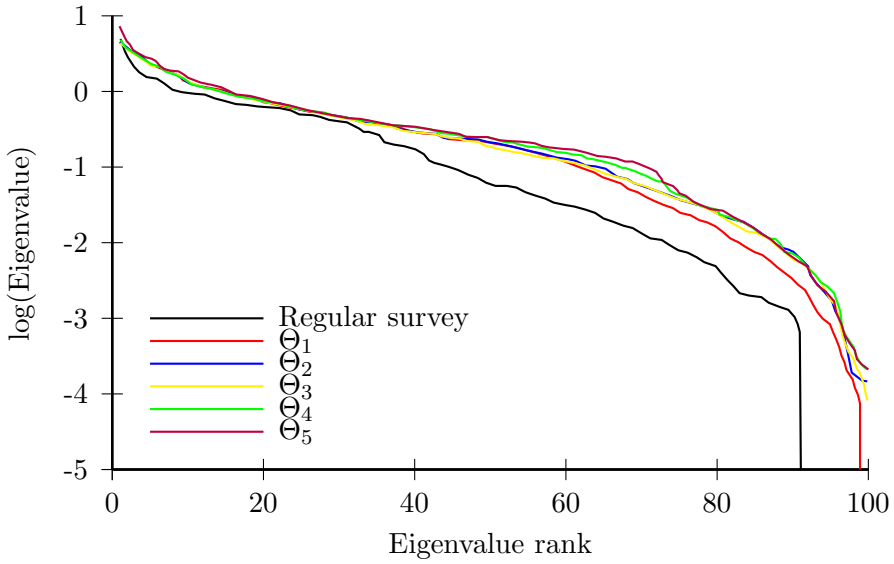

Figure 1.11: Log eigenvalue spectra of the regular survey geometry and optimal geometries found using each of the Θ quality measures (equations (2.50)) for a 20 source and 20 receiver cross-borehole tomography survey. In each measure involving a δ term, $\delta = 0.001$ was used and for Θ_2, $k = 90$ was used. (Adapted from Curtis (1999a)).

point between the boreholes.

4. Average source and receiver densities on the ground surface is lower than that down each well.

Figure 1.11 shows the corresponding eigenvalue spectra of the optimal design for each optimality measure (Θ) for the borehole setup shown in Figure 1.10. The black line represents the result obtained for a standard design (equal source and receiver separation positioned down the boreholes only) and the red line represents the result for the design shown in Figure 1.10. Curtis (1999a) found that using any of the Θ functions to design an optimal survey would result in far better conditioned inverse problems than the regular design and that all the optimal designs place multiple sources and receivers on the surface, a feature not used in the standard design. Figure 1.11 shows that the efficient design measure, Θ_1, unlike the other measures is unable to constrain all the model parameter. The choice of the weights in equation (1.2) can therefore be seen as defining a trade-off between maximising information and minimising the computational time to reach a solution.

Curtis (1999a) concluded that the design process confirmed intuitive borehole tomography designs. Model cells at depth are least constrained by surface sources and receivers and hence an increased density of sources and receivers with depth is required. Model cells near the centre of the space between the boreholes are least constrained by the sources and receivers in the boreholes and therefore the inversion can be better constrained by placing sources and receivers on the surface. The SED method presented by Curtis (1999a) provides additional quantitative information that cannot be obtained using intuition alone. The methods introduced are also ideally suited to situations where irregular geometries are encountered and heuristic methods fail.

Whereas the algorithms and examples presented by Curtis (1999a) optimise the design to best constrain all the model parameters (the complete discretised velocity structure) Curtis (1999b) introduced a method to design focused surveys which reduce post-survey ambiguity in a particular region of the subsurface. Curtis (1999b) decomposed the quality measure Φ (equation (1.2)) into two parts

$$\Phi = \mu\Phi^G + (1-\mu)\,\Phi^F \tag{1.3}$$

where Φ^G is a measure of the global information and Φ^F is the amount of information focussed on a subregion of the model and μ is a weighting factor. A value of $\mu = 0$ will result in a design that only constrains the subregion of interest whilst a value of $\mu = 1$ will result in a globally optimal result with no increased focussing on any specific region.

Although the focussing algorithm proposed could be used in any SED setting, Curtis (1999b) gave examples from a borehole tomography setting. Figure 1.12 shows three optimal survey designs, each consisting of six sources and six receivers, which have been designed to focus on the model region shaded in yellow using μ values of (a) 0.9, (b) 0.5 and (c) 0.1. The plot clearly shows that as the value of μ decreases the design becomes more focussed on the subsurface region of interest. By focussing the survey, however, areas of the model which do not have any raypaths intersecting them are unable to be constrained. The method of focussing is therefore an ideal algorithm to use in a time-lapse scenario where only a specific area of the model is of interest and either the rest of the subsurface has been constrained in a previous survey or further constraint of the velocity model has no impact on future reservoir developments.

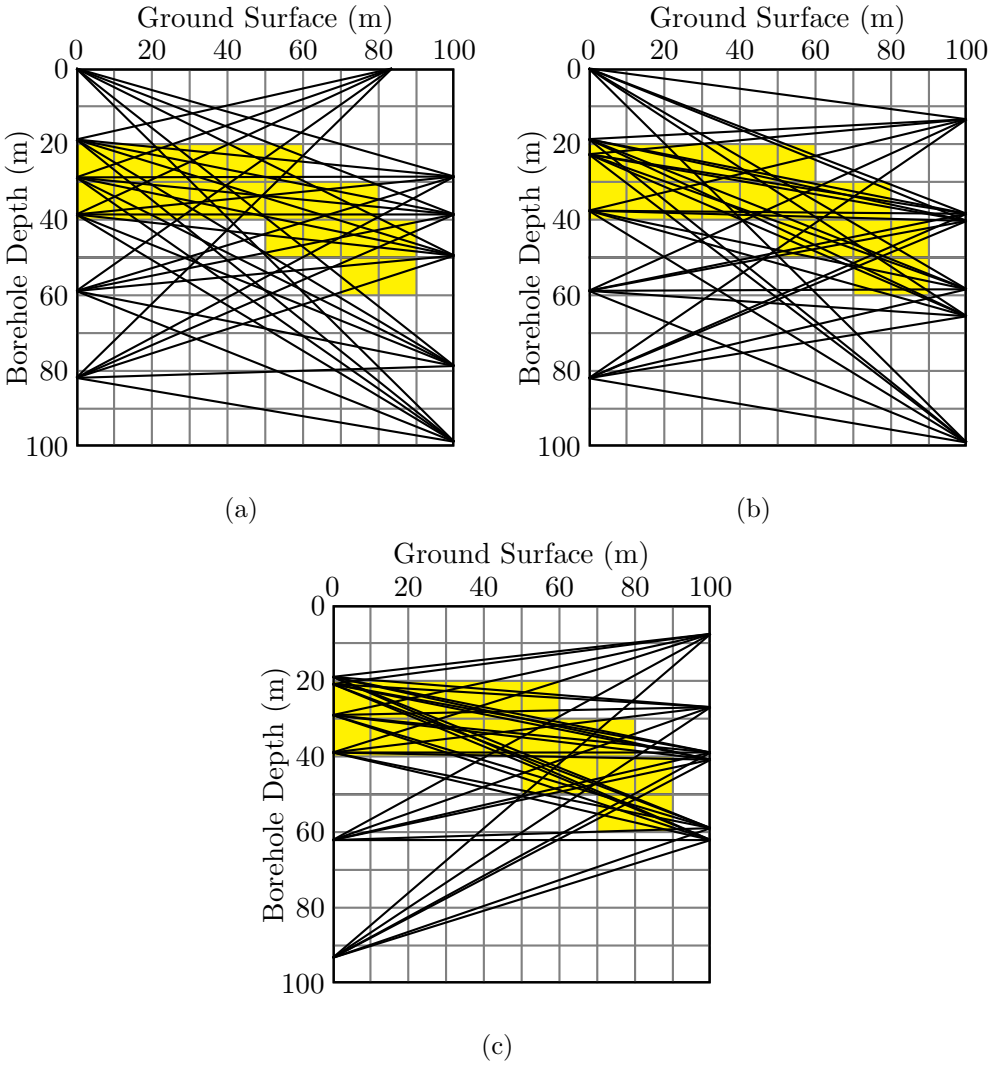

Figure 1.12: Optimal cross-borehole designs found using the Curtis (1999b) method to place six sources and six receivers with the ability to focus the design on a specific region of the model shown here in yellow. Plot (a) shows the result when $\mu = 0.9$, (b) when $\mu = 0.5$ and (c) when $\mu = 0.1$. (Adapted from Curtis (1999b)).

The methods used thus far to calculate the optimal borehole tomography designs are taken from a class of algorithms termed stochastic (see Section 2.4.1) and produce different results every time the algorithm is executed. The inherent random nature of stochastic algorithms means that there is no guarantee that they will perform well given computational limitations.

Curtis et al. (2004) presented a deterministic algorithm (based on observations of Sabatier (1977)) with the advantage of the results being reproducible and that the algorithm guarantees that an experimental design will be obtained within a fixed length of time for a problem of finite size. As with the stochastic methods however, the final design may not represent the globally optimal design. The design algorithm of Curtis et al. (2004) starts by considering an unrealistic experimental setup. In the case of a tomography experiment, one borehole would be filled at every possible position with sources and the other borehole with receivers. The Curtis et al. (2004) algorithm then considers each of the source and receivers and assigns a value quantifying how much information it contributes to constraining the model. The source or receiver that supplies the least amount of information is removed. The algorithm then updates all other values to reflect this change and the procedure is repeated until a pre-defined number of sources and receivers is reached (see Section 2.4.2). The algorithm allows both the model parameters and the sources and receivers to be weighted. Therefore, as in Curtis (1999b) the design can be made to focus on subregions of the model and can also force specific sources and receivers to be present in the final design.

Figure 1.13 shows an example presented in Curtis et al. (2004) to demonstrate the algorithm. Plot (a) represents the starting design for a simple borehole tomography problem consisting of ten sources and ten receivers. In this example the traveltime from a given source is only recorded at a single receiver. Using the deterministic algorithm with no focussing or weighting, the sources and receivers are removed in the following order:

13,14,16,15,3,4,7,8,6,5,10,9,1,2,12,11,18,17,20,19.

Plot (b) shows the optimal design after eight sources or receivers have been removed. The results show that raypaths that are duplicates and therefore provide no linearly independent information about the model are removed first. This makes sense intuitively with the algorithm only removing data for which the sen-

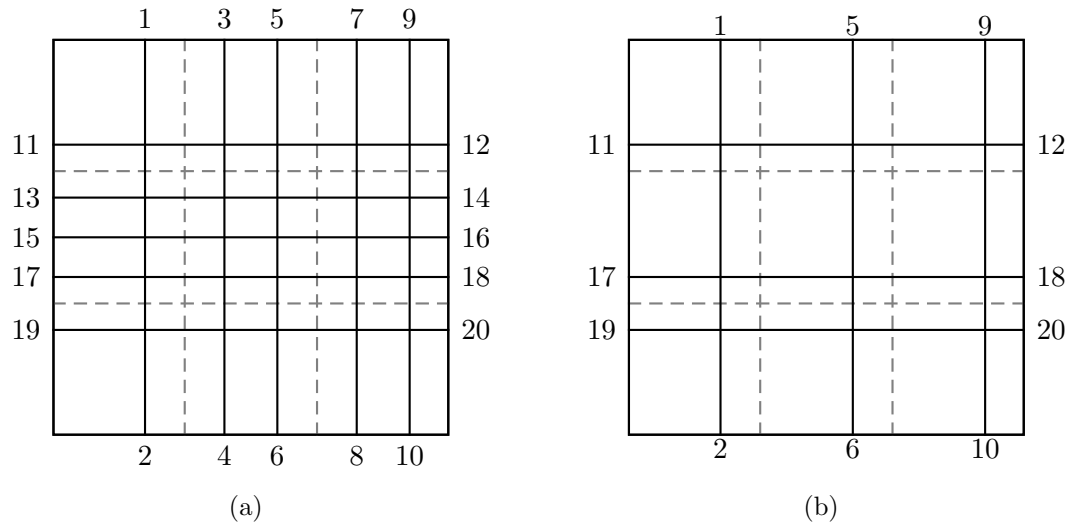

Figure 1.13: Geometry of a simple tomography scenario. The grey dashed squares represent model discretisations. Bold lines represent raypaths joining a single source and receiver denoted by the numbers ranging from 1 to 20. Plot (a) shows an initial starting model and plot (b) an optimal design after eight sources or receivers have been removed. (Adapted from Curtis et al. (2004)).

sitivities to the model parameters are most linearly dependent on other data from the experiment. The plots in Figure 1.14 show the eigenvalue spectra after one, six, eleven and sixteen sources and receivers have been removed for both an optimal design (solid line) and the average of twenty random realisations (dashed line) of source and receiver removals. The plots show that the deterministic algorithm produces better constrained results than randomly removing sources and receivers.

Curtis et al. (2004) applied the same algorithm to a more complex cross-borehole tomography example and a microseismic monitoring example and found that the pattern of source and receiver removal was consistent with intuition. Although the results agreed with intuitive ideas in the simple examples presented, the deterministic method provides quantitative results which would provide robust surveys in scenarios where intuition alone would not be possible.

Ajo-Franklin (2009) studied the use of SED methods for designing time-lapse traveltime tomography experiments. The main issue with performing repeat experiments and then trying to compare the resulting data sets is that logistical constraints hamper the repeatability of surveys (e.g. replicating source/receiver geometry, changes to overburden properties), so being able to perform the exact

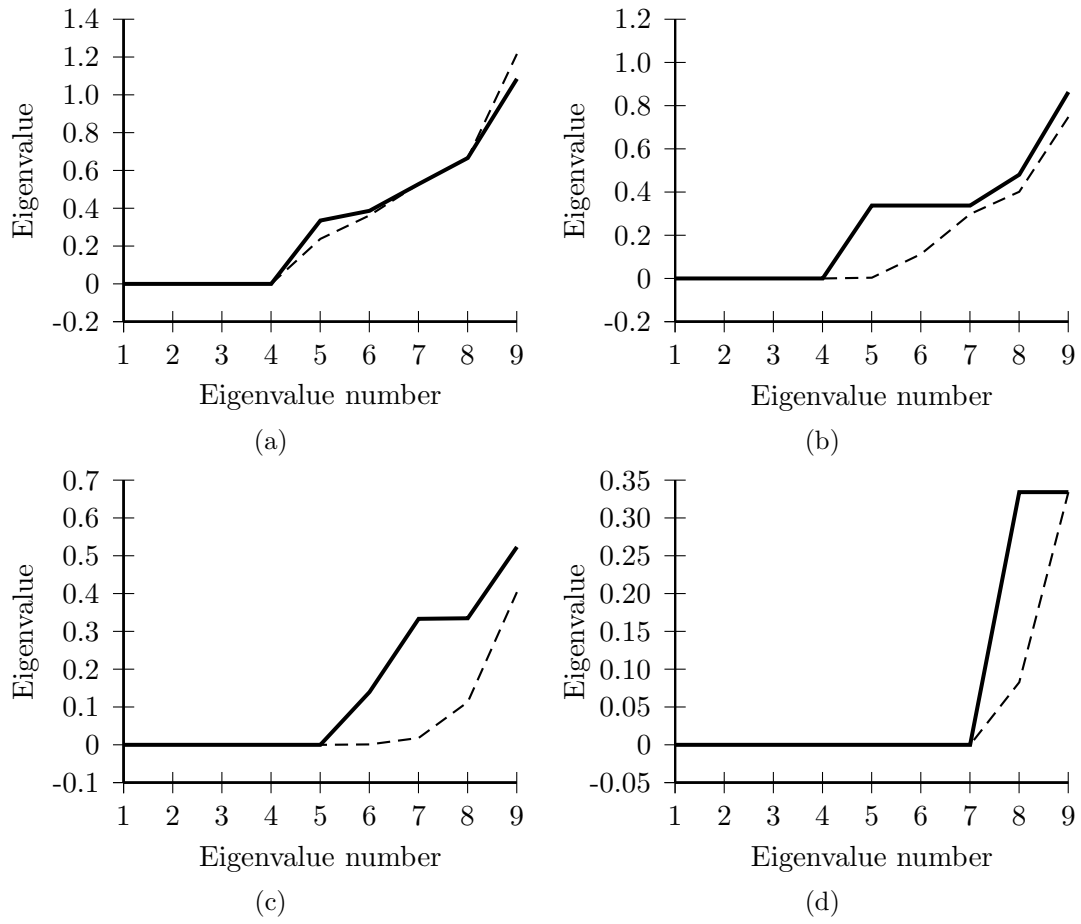

Figure 1.14: Plots (a), (b), (c) and (d), respectively, show eigenvalue spectra after one, six, eleven and sixteen sources or receivers have been removed from the tomography problem shown in Figure 1.13 using the method developed by Curtis et al. (2004). Dashed lines show the average spectra after twenty random realisations of removing sources and receivers and bold lines show the result produced by the design algorithm . (Adapted from Curtis et al. (2004)).

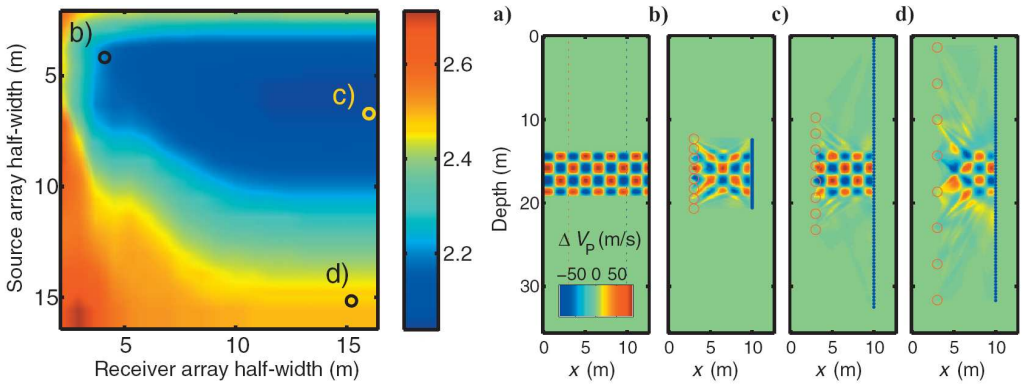

Figure 1.15: The left hand plot shows the penalty function measure for the cross-borehole experiment as a function of source and receiver half-width. Cold colours represent optimal designs and warm colours sub-optimal. Plot (a) shows the velocity model for which the experiment is being designed to image. The three remaining figures show the results for designs labelled in the left hand plot. The red circles represent source positions and the blue dots receiver positions. (Reproduced from Ajo-Franklin (2009)).

same processing flow is impossible. Instead, permanent sensor arrays are now being deployed (Blanco et al., 2006; Daley et al., 2007) to minimise the repeatability issues. Because permanent arrays are often unable to be reconfigured once deployed, so their design becomes much more important compared to a standard survey where mistakes can potentially be corrected during a later deployment (Ajo-Franklin, 2009).

The approach introduced in Ajo-Franklin (2009) was to use a reduced parameterisation. Whereas previous works allowed sources and receivers to be located at arbitrary locations the algorithm introduced by Ajo-Franklin (2009) optimises a secondary set of descriptive parameters such as array width, centre location or orientation to locate the individual sources and receivers. Although this method removes the ability to design highly complex surveys it does allow large search problems to become computationally viable. To demonstrate the method Ajo-Franklin (2009) applied it to designing an optimal cross-borehole experiment to locate eight sources and eighty receivers around a subsurface region of interest. To locate the sources and receivers using previously mentioned methods would require an eighty-eight dimensional design space to be searched. Ajo-Franklin (2009) parameterised the problem using only two hyper-parameters, the half-width or extent to which the sources and receivers are spread above and below the region of interest.

The first plot in Figure 1.15 shows the optimality measure for the cross-borehole designs as a function of the source and receiver half-widths to provide the best image of the known subsurface velocity distribution shown in plot (a). For this case Ajo-Franklin (2009) found that the optimal design would locate the sources close to the region of interest and the receivers spread over a larger depth range. Plots (b), (c) and (d) show the results for each of the points labelled in the optimality plot. Plot (c) (the optimal design) clearly shows the best results for replicating the checker board velocity structure.

1.2.4 Amplitude-Versus-Offset (AVO) Surveys

All of the geophysical survey design papers cited above rely on linear or linearised design theory to quantify the amount of information that a particular survey provides, or to represent the physical relationship linking the model parameters and recorded data. One of the main reasons that SED theory has not gained general acceptance in the Geosciences is that applying linearised design theory is not robust as the relationships can be highly nonlinear.

Van den Berg et al. (2003, 2005) introduced a fully nonlinear design algorithm and applied it in a reflection seismology setting where the recorded seismic amplitude (which varies with the distance from the source to the receiver) of a reflected wave contains information about the subsurface (see Chapter 3). Data that reflects from the same subsurface location but which is recorded at different source receiver offsets provides differing amounts of information. Selecting an optimal source-receiver separation reduces the uncertainty on the subsurface properties of interest. The algorithm developed by Van den Berg et al. (2003, 2005) to select this optimal separation was constructed in a Bayesian framework were values are represented by probability distributions and the optimality of a specific design is calculated by measuring the entropy of a given distribution (see Section 2.3.1). In this way *a priori* information can be represented by distributions such as Gaussian, Cauchy or Uniform rather than being given discrete values. Applying this theory, Van den Berg et al. (2003, 2005) showed that applying linearised SED theory to highly nonlinear problems can result in designs which are severely sub-optimal.

Because of the numerical implications of using fully nonlinear design methods Van

den Berg et al. (2003, 2005) was only able to locate a single optimal receiver and therefore the results produced are not able to be used in practical applications. Winterfors and Curtis (2008) also introduced a fully nonlinear design method, however, the algorithm proposed was only used to design a two-dimensional parameter vector (see Section 2.3.1) and therefore has not been demonstrated to design a full scale industrial AVO (or other geophysical) survey design.

1.3 Thesis Aims

With linearised methods shown not to be robust in geophysical problems and the proposed nonlinear methods currently unable to produce survey designs that resemble full scale experiments, there is a need to create nonlinear algorithms capable of designing surveys on a scale similar to those produced using linearised theory. The work in this thesis fills this gap by producing industrial scale surface seismic AVO survey designs using totally nonlinear (and non-linearised) methods. The algorithms introduced are not seismic survey specific with the main concepts applicable to any geophysical method. It is hoped that the work in this thesis will thus increase the exposure of nonlinear SED methods in geophysics and as a result increase the uptake of SED methods within the Geosciences.

1.4 Thesis Outline

In Chapter 2 the mathematical theory used to design optimal experiments is introduced. The methods used are based on inverse theory which is presented from a nonlinear perspective built around a Bayesian framework. Since the majority of the references cited rely on linearised theory, classical linear (and linearised) methods are also included. It is shown how the linear and nonlinear design measures are related. Finally in Chapter 2 standard methods used to search and locate the optimal design without the need to search the entire design space are presented.

In Chapter 3 the basics of Amplitude Versus Offset (AVO) theory are introduced. The theory outlined is used in all subsequent chapters as the basis of linking the

subsurface model parameters of interest to the data recorded at the Earth's surface. AVO methods are widely used in the exploration industry and therefore designing optimal surveys using this theory should further demonstrate the importance and possibilities of using SED in planning industrial scale surveys.

Chapter 4 (published in JGR-Solid Earth as Guest and Curtis (2009)) introduces a novel stochastic algorithm for designing a survey consisting of up to ten parameters using fully nonlinear methods without any linearisation.

Chapter 5 (published in Geophysics as Guest and Curtis (2010a)) uses the methods presented in Chapter 4 to produce processing designs for a range of possible hydrocarbon-bearing reservoir models. It is shown how the prior knowledge of the reservoir porosity has the largest effect on the processing designs, and how optimal designs consisting of only ten receivers can be upscaled to produce approximately optimal designs consisting of any number of receivers.

Chapter 6 (submitted to GJI as Guest and Curtis (2010b) uses hyper-parameters to design for the first time a full scale surface seismic AVO experiment accounting for AVO processing techniques that are commonplace in the exploration industry. It is shown that the optimal survey design is highly dependant on the scale of the survey being designed.

Chapters 7 and 8 conclude the thesis with an overall discussion and conclusions of what has been achieved in this program of work, and how the SED methods presented can now, and should in future be used routinely to design geophysical surveys.

Chapter 2

Mathematical Framework

When planning a geophysical experiment it is important that the objectives for which the experiment is being designed are accurately defined. The objectives can be separated into two distinct classes, geological and operational. Geological objectives define the subsurface parameters of interest and can range from being specific (for example, locating a subsurface interface) to being vague (for example, image the volume between two boreholes). Operational objectives relate to the cost and logistics of performing the experiment and include factors such as the types of equipment to be used, the cost of manufacturing and deploying the equipment, and the cost of recording and storing the data. Operational factors are generally based on the desire to minimise the overall survey cost. Once all parameters are defined they are combined into a single mathematical objective function which is typically chosen to have a minimum when the survey design best fits the survey objectives (Maurer et al., 2010). To design experiments optimally therefore requires an understanding of the physical relationships linking the recorded data and post-experimental parameter uncertainties (Box and Lucas, 1959; Atkinson and Donev, 1992). Calculating which design corresponds to a minimum can be achieved by a number of mathematical algorithms.

Extracting geological information about the subsurface from geophysical data is usually performed by inversion or inference techniques. Figure 2.1 illustrates this process for a simple tomography experiment. A single traveltime recorded from the source receiver pair in Figure 2.1 (a) along a straight raypath (dashed line) is inverted to calculate the subsurface velocity structure which has been discretised

into two equal sized sections. Figure 2.1 (b) shows that the inversion result is a range of slowness (reciprocal of velocity) combinations with each combination resulting in the correct recorded traveltime. In this scenario it is impossible to constrain the exact velocity structure without additional information. If the same experiment is performed again using the setup shown in Figure 2.1 (c) with two traveltime measurements there exists a unique slowness solution (as shown by the intersection of the two lines in Figure 2.1 (d)) that corresponds to the correct subsurface velocity structure.

In normal inverse problems the experimental geometry and equipment used has already been chosen and the data has already been recorded. The inversion results (Figure 2.1 (b) and (d)) are therefore already fixed. In the experimental design stage the geometries are flexible and the data is yet to be recorded. With knowledge of how the experimental parameters will affect the *expected* data to be recorded, and how the inversion process will map the recorded data into parameter uncertainties, the experimental geometry can be adapted until a measure of the *expected* post-experimental parameter uncertainties are minimised.

The plots in Figure 2.1 assume that the recorded traveltimes are error free and that the raypaths are exactly known and therefore only a single pair of slownesses fit the recorded data. In practice all geophysical measurements have associated errors. If a more realistic dataset is recorded and inverted then the resulting inversion solution (like that depicted in Figure 2.1 (d)) would not intersect at a single point but over an area of model space leading to an ambiguous solution.

Rather than a single data value, it is therefore more appropriate to quantify measurements in terms of probability density functions (pdfs) (Tarantola, 2005) which specify a range of possible values for underlying 'ideal' or error-free data, each with different probabilities of accuracy. Figure 2.2 (a) shows the results of Figure 2.1 (d) with added uncertainty to account for traveltime measurement error. Figure 2.2 (b) shows the intersection of the inversion results now shown as a pdf over possible slowness values that are consistent with the two traveltime data measurements. In this scenario the maximum probability occurs where the error free lines (red lines) intersect but there also exists a range of viable solutions with varying probabilities that lie on neither line.

Inverting a complete realistic dataset with many data would result in a more

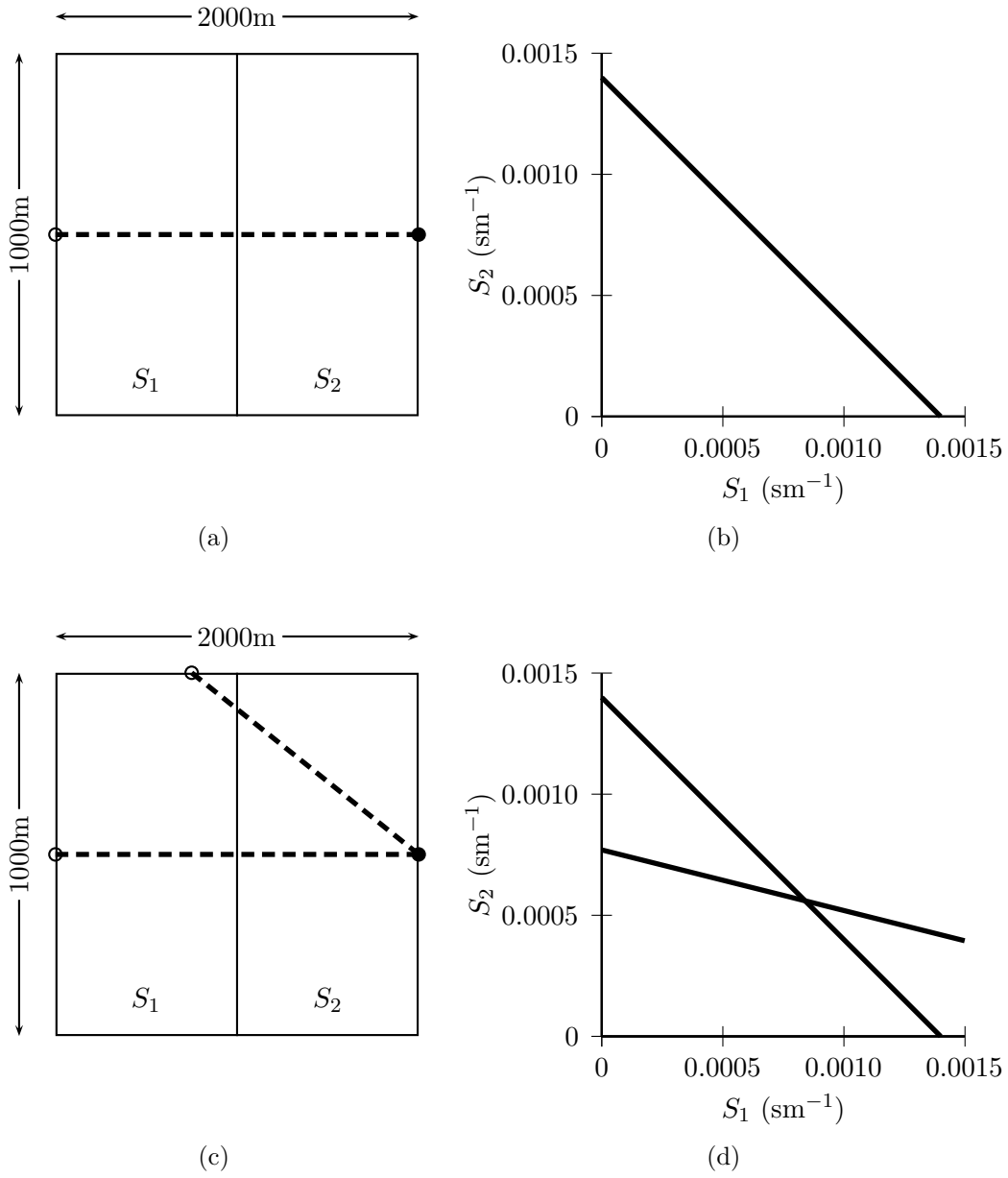

Figure 2.1: Plot (a) shows the geometry for a single source receiver tomography experiment where the subsurface has been discretised into two constant slowness regions. Plot (b) shows the possible slowness combinations that result in traveltimes that are equal to the recorded traveltime. Plot (c) shows the same experiment with two source receiver travel paths. Plot (d) shows the slowness combinations that give results equal to the recorded traveltimes in Plot (c). The intersection of the two lines corresponds to the only solution which fits both traveltime data.

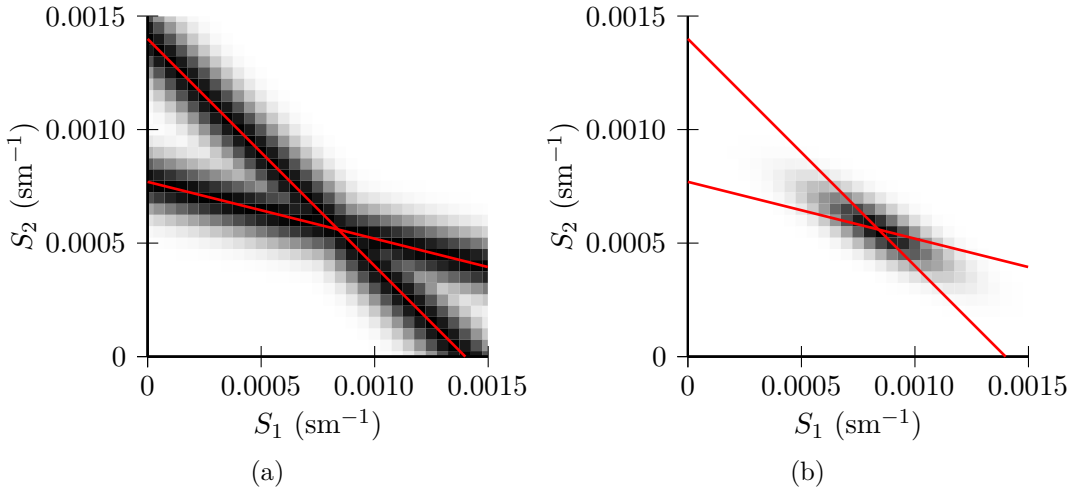

Figure 2.2: Plot (a) shows a probabilistic representation for each of the possible solutions to the traveltimes recorded using the geometry in Figure 2.1 (c). The red lines represent the error free solutions whilst the shading represents probabilistic results when Gaussian data error is added. Darker shading represents a more probable solution to the inverse problem. Plot (b) shows the intersection of the results and therefore the probabilistic solution to the inverse problem when noisy traveltime data are considered.

complex final pdf. By varying the experimental setup the properties of the pdf will change resulting in different measures of the model parameter uncertainties. The theory underling SED methods therefore requires an understanding of probabilistic inverse theory and how the amount of information within a pdf can be quantified.

2.1 Inverse Theory

2.1.1 Nonlinear

After a survey has been completed a data set $\mathbf{d}$ is available for geological interpretation to constrain the subsurface model $\mathbf{m}$. In the above example the traveltimes would be represented by the vector $\mathbf{d}$ and the slowness parameters by $\mathbf{m}$. Inverse theory is the branch of mathematics that studies the process of using the information in the dataset $\mathbf{d}$ to constrain estimates of the parameters $\mathbf{m}$.

In inverse theory (and SED methods) it is essential to be able to estimate the dataset $\mathbf{d}$ that will be recorded for a given model $\mathbf{m}$. Let function $\mathbf{F}$ represent the relationship between parameters $\mathbf{m}$ and data $\mathbf{d}$, such that if we ignore measurement error for now,

$$\mathbf{d} = \mathbf{F}(\mathbf{m}) \tag{2.1}$$

would be recorded if the parameters $\mathbf{m}$ correctly represented the true subsurface of the Earth.

As previously mentioned, one of the principal reasons that SED methods have not gained general acceptance in the Geosciences (other than the lack of awareness) is that most research efforts on SED in the statistical community has focussed on developing methods that assume a linear or linearised relationship $\mathbf{F}$ between parameters and data, while in geophysical application the parameter-data relationship is commonly significantly nonlinear.

As introduced above, it is useful in inverse theory to represent uncertainties in both data $\mathbf{d}$ and model parameters $\mathbf{m}$ with probability density functions (pdfs). The pdf of a continuous random variable ($\mathbf{X}$), represented here as $P(\mathbf{x})$, is the function that describes the relative likelihood of $\mathbf{X}$ taking any specific value $\mathbf{x}$. Pdfs are always positive or zero and have the following relationships to probabilities ($Pr(\ldots)$)

$$Pr(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} P(\mathbf{x})\, d\mathbf{x} \tag{2.2}$$

$$Pr(-\infty \leq \mathbf{X} \leq \infty) = \int_{-\infty}^{\infty} P(\mathbf{x})\, d\mathbf{x} = 1 \tag{2.3}$$

for real vectors $\mathbf{a}$ and $\mathbf{b}$, where $a_i < b_i$ for all i up to the dimensionality of $\mathbf{X}$, and where the integrals run over each dimension of the vector $\mathbf{x}$. The equations show that the pdf contains all the information required to calculate the probability of any particular event occurring. The properties also shows that pdfs are normalised so that the total probability that $\mathbf{X}$ takes any value is 1.

If $\mathbf{X}$ and $\mathbf{Y}$ are two random variables, then the joint random variable that describes both together is given by $(\mathbf{X}, \mathbf{Y})$ with pdf $P(\mathbf{x}, \mathbf{y})$. With knowledge of the joint distribution two further distributions can be defined. The marginal distribution of $\mathbf{X}$ is given by

$$P(\mathbf{x}) = \int_{-\infty}^{\infty} P(\mathbf{x}, \mathbf{y})\, d\mathbf{y} \tag{2.4}$$

and is obtained by integrating the joint pdf over all possible values of $\mathbf{Y}$. The marginal distribution represents the pdf of the value of $\mathbf{X}$ given no knowledge at all of the value of $\mathbf{Y}$. Conversely, the conditional distribution of $\mathbf{X}$ given knowledge of the value of $\mathbf{Y}$ (represented as $(\mathbf{X}|\mathbf{Y})$ is given by

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x},\mathbf{y})}{P(\mathbf{y})}. \tag{2.5}$$

This provides the pdf of $\mathbf{X}$ with exact knowledge of the value $\mathbf{y}$ assumed by random variable $\mathbf{Y}$.

From equation (2.5) it also follows that

$$P(\mathbf{x},\mathbf{y}) = P(\mathbf{y}|\mathbf{x})\,P(\mathbf{x}), \tag{2.6}$$

and therefore

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})\,P(\mathbf{x})}{P(\mathbf{y})}. \tag{2.7}$$

Equation (2.7) is known as Bayes' theorem with the pdf $P(\mathbf{x}|\mathbf{y})$ representing all information available about model $\mathbf{x}$ given the data $\mathbf{y}$ and represents the probabilistic solution to the inverse problem. If the data recorded during an experiment is represented by the vector $\mathbf{d}$ and the model parameters of interest by vector $\mathbf{m}$ then equation (2.7) can be written in the more informative form

$$\sigma(\mathbf{m}|\mathbf{d}) = \frac{\theta(\mathbf{d}|\mathbf{m})\,\rho(\mathbf{m})}{\sigma(\mathbf{d})} \tag{2.8}$$

where $\sigma(\mathbf{m}|\mathbf{d})$ is the posterior pdf of information about the values of parameters $\mathbf{m}$ given the recorded data $\mathbf{d}$. The pdf $\rho(\mathbf{m})$ contains the information on parameters $\mathbf{m}$ without any knowledge of the data $\mathbf{d}$, and hence represents the state of knowledge about $\mathbf{m}$ prior to the data being inverted. Prior knowledge may come from previous experiments, well data, general geological knowledge, etc. Although this is not necessary for expression (2.8), for practical purposes a Gaussian distribution is commonly used to represent an approximation to the prior knowledge about the parameters (Tarantola, 2005),

$$\rho(\mathbf{m}) = k_1 \exp\left[-\frac{1}{2}(\mathbf{m}-\hat{\mathbf{m}})^{\mathrm{T}}\,\mathbf{C}_{\mathbf{m}}^{-1}\,(\mathbf{m}-\hat{\mathbf{m}})\right] \tag{2.9}$$

where $\hat{\mathbf{m}}$ is the mean of the distribution, $\mathbf{C}_{\mathbf{m}}$ is the covariance matrix which

contains the prior model uncertainties, and k_1 is a normalisation constant defined by equation (2.3). The prior distribution should also contain constrains so that physical laws are not broken. For example, the region of the pdf representing negative seismic velocities should be defined as having a probability of zero.

The pdf $\theta(\mathbf{d}|\mathbf{m})$ in equation (2.8) describes the probability distribution of data $\mathbf{d}$ that might be recorded given a fixed set of model parameters $\mathbf{m}$. Equation (2.8) implies that the theory relating data and parameters does not impose any constraint on the model parameters $\mathbf{m}$ but only on the data $\mathbf{d}$ (Tarantola and Valette, 1982). In most inverse problems the forward function (equation (2.1)) is used to create synthetic data (Bard, 1974). If actual data measurements are represented by the vector $\mathbf{d}_{\text{obs}}$ then when $\mathbf{d} = \mathbf{d}_{\text{obs}}$ is substituted into $\theta(\mathbf{d}|\mathbf{m})$ then the result, interpreted as a function of $\mathbf{m}$ is called the likelihood function (Duijndam, 1988)

$$L(\mathbf{m}) = \theta(\mathbf{d} = \mathbf{d}_{\text{obs}}|\mathbf{m}) \tag{2.10}$$

and gives a measure of how good a model $\mathbf{m}$ is at explaining the measured data (Tarantola, 2005).

If the errors in the forward function are represented by a Gaussian function, described by mean $\mathbf{d}_0$, covariance matrix $\mathbf{C}_\mathbf{d}$ and normalisation constant k_2, then

$$L(\mathbf{m}) = k_2 \exp\left\{-\frac{1}{2}\left[\mathbf{d}_0 - \mathbf{F}(\mathbf{m})\right]^{\mathrm{T}} \mathbf{C}_\mathbf{d}^{-1}\left[\mathbf{d}_0 - \mathbf{F}(\mathbf{m})\right]\right\}. \tag{2.11}$$

The pdf $\sigma(\mathbf{d})$ in equation (2.8) represents the marginal distribution over data $\mathbf{d}$ given no knowledge of the values of $\mathbf{m}$, hence represents knowledge about data that are likely to be measured prior to conducting the survey or experiment. This is given by

$$\sigma(\mathbf{d}) = \int \theta(\mathbf{d}|\mathbf{m})\,\rho(\mathbf{m})\,d\mathbf{m}. \tag{2.12}$$

In Bayes' theorem (equation (2.8)) the pdf $\sigma(\mathbf{d})$ is also a necessary normalisation factor to ensure that the posterior pdf over parameters $\mathbf{m}$ has unit volume.

If the number of model parameters is low and the computational expense required to calculate $\sigma(\mathbf{m}|\mathbf{d})$ is low then a complete grid search over all model space values of $\mathbf{m}$ can performed to fully solve the inverse problem and find the likelihood of how well a representative set of all possible model parameter values

fit the observed data. If the number of model parameters is too large then an exhaustive search can not be performed. As a result the normalisation constant $\sigma(\mathbf{d})$ can not be calculated (since by equation (2.12) it requires the integration over all model space). The final posterior distribution $\sigma(\mathbf{m}|\mathbf{d})$ is therefore by definition not a pdf since it can not be normalised. In such cases the product of the prior $\rho(\mathbf{m})$ and the non-normalised likelihood $L(\mathbf{m})$ usually suffices as sufficient information about the inverse problem solution (Maurer et al., 2010). A probabilistic solution to the inverse problem can then be constructed by combining the information represented by the prior information on the parameters and the likelihood function

$$\sigma(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{m}) L(\mathbf{m}). \tag{2.13}$$

It is commonly the case that the forward function is nonlinear or the distribution $\rho(\mathbf{m})$ is non-Gaussian. The posterior distribution $\sigma(\mathbf{m}|\mathbf{d})$ is therefore likely to be highly irregular with multiple local maxima. If an exhaustive search is not possible then algorithms used to search the posterior distribution to find the global maximum value must be robust. Search methods often used include pseudo-random ("Monte Carlo") methods, genetic algorithms and simulated annealing methods (see Section 2.4).

If using such algorithms do not make the problem tractable, the prior information with which to constrain model parameter estimates is too limited, or the forward function is highly nonlinear, then the best method to characterise a local sub-volume of the posterior distribution is to approximate the nonlinear forward function with a linear function, linearised around a prior estimate of parameter value, $\mathbf{m}_0$ (Maurer et al., 2010). The sub-volume characterised will then be local to $\mathbf{m}_0$.

2.1.2 Linearised

Although the methods presented in this thesis do not require any linearisation of the forward function many of the geophysical SED problems mentioned in Section 1.2 were solved using linear methods. This section on linear and linearised methods has been included for completeness.

Let the model be parameterised into a model vector $\mathbf{m}$ of length P and let the recorded data be contained in a vector $\mathbf{d}$ of length D. In linear inverse theory the relationship between the data and model vectors can be written as

$$\mathbf{d} = \mathbf{A}\mathbf{m} \tag{2.14}$$

where $\mathbf{A}$ is a $D \times P$ matrix of coefficients that are independent of both the data and the model (Gubbins, 2004). The matrix $\mathbf{A}$ can be written as

$$A_{ij} = \frac{\partial d_i}{\partial m_j} \tag{2.15}$$

where d_i and m_j are elements of vectors $\mathbf{d}$ and $\mathbf{m}$ respectively. The i^{th} row of $\mathbf{A}$ describes how the i^{th} datum depends on the model vector.

In the case of an overdetermined problem there are more linearly independent data than parameters and the least squares solution is given by

$$\mathbf{m} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{d}. \tag{2.16}$$

Let $\mathbf{C_d}$ represent the data covariance matrix which describes all measurement uncertainties in the data $\mathbf{d}$. The least squares solution is then modified to be given by

$$\mathbf{m} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C_d}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C_d}^{-1}\mathbf{d} \tag{2.17}$$

where $\left(\mathbf{A}^{\mathrm{T}}\mathbf{C_d}^{-1}\mathbf{A}\right)^{-1}$ describes how the measurement errors have been mapped into the calculated model parameters (Menke, 1989).

If the number of linearly independent data is equal to the number of parameters then the solution reduces to

$$\mathbf{m} = \mathbf{A}^{-1}\mathbf{d}. \tag{2.18}$$

If the number of linearly independent constraints is less than the number of unknown parameters then the inversion is deemed under determined and no unique solution exists. One solution to the inverse problem can be written as

$$\mathbf{m} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C_d}^{-1}\mathbf{A} + \mathbf{C_m}^{-1}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C_d}^{-1}\mathbf{d} \tag{2.19}$$

where $\mathbf{C}_\mathbf{m}^{-1}$ is the prior covariance matrix and allows constraints to be placed on the model parameter regularisation such as defining model smoothness, complexity etc.

When the forward function is not completely linear the elements of the matrix $\mathbf{A}$ are derived from the nonlinear formula $\mathbf{F}(\mathbf{m})$ for an initial reference model $\mathbf{m}_0$

$$(\mathbf{A}_0)_{ij} = \left.\frac{\partial F_i}{\partial m_j}\right|_{\mathbf{m}=\mathbf{m}_0}. \tag{2.20}$$

The reference model $\mathbf{m}_0$ is usually either the mean or maximum likelihood model from the prior model pdf (Maurer et al., 2010). The inversion method seeks to find a small improvement in the model $\delta\mathbf{m}_0$ by performing a linearised inversion of the data residuals

$$\delta\mathbf{d}_0 = \mathbf{A}_0\delta\mathbf{m}_0 \tag{2.21}$$

using, for example, equation (2.16). The improved solution

$$\mathbf{m}_1 = \mathbf{m}_0 + \delta\mathbf{m}_0 \tag{2.22}$$

should better fit the data $\mathbf{d}$. If $\mathbf{m}_1$ does not provide a good enough solution then the iteration procedure is repeated with starting model $\mathbf{m}_1$ rather than $\mathbf{m}_0$. In general the model update for the I^{th} iteration is

$$\delta\mathbf{m}_I = \left(\mathbf{A}_I^{\mathrm{T}}\mathbf{C}_\mathbf{d}^{-1}\mathbf{A}_I + \mathbf{C}_\mathbf{m}\right)^{-1}\left(\mathbf{A}_I^{\mathrm{T}}\mathbf{C}_\mathbf{d}^{-1}\delta\mathbf{d} - \mathbf{C}_\mathbf{m}^{-1}\mathbf{m}_I\right). \tag{2.23}$$

A solution is reached when $\delta\mathbf{m}_I$ falls below a certain threshold and no significant update to the model is being made with each iteration.

The quality of an inversion for a truly linear forward function (equation (2.14)) can be calculated by assessing the resolution matrix $\mathbf{R}$

$$\mathbf{R} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C}_\mathbf{d}^{-1}\mathbf{A} + \mathbf{C}_\mathbf{m}^{-1}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}_\mathbf{d}^{-1}\mathbf{A} \tag{2.24}$$

which relates the estimated model to the true model

$$\mathbf{m}_{\text{est}} = \mathbf{R}\mathbf{m}_{\text{true}}. \tag{2.25}$$

Resolution is classed as perfect when $\mathbf{R} = I$, the identity matrix, and there is no difference between the true model and the estimate. Diagonal elements of

$\mathbf{R}$ close to zero indicate poorly resolved model parameters whilst non-zero off-diagonal elements of the resolution matrix indicate a trade-off between the model parameters (Gubbins, 2004).

The posterior model parameter covariance matrix

$$\mathbf{C} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathbf{d}}^{-1}\mathbf{A} + \mathbf{C}_{\mathbf{m}}^{-1}\right)^{-1} \tag{2.26}$$

translates data uncertainties $\mathbf{C_d}$ into the model parameters and combines them with the prior model parameter uncertainties $\mathbf{C_m}$ to estimate the posterior covariances $\mathbf{C}$. Variances of the individual parameters are represented by the diagonal elements with small values indicating well resolved parameter estimates. Non-zero off diagonal elements indicate the degree to which parameter estimates are correlated post-inversion (Maurer et al., 2010).

If a linearised method has been used and the forward function is not truly linear then the resolution matrix and the parameter covariance matrix are valid only locally around the final solution. In such cases the overall quality of the estimates of the parameters must be quantified by sampling the posterior pdf (if computationally possible).

2.2 Design Theory

From the inverse theory methods presented above it is clear that the quality of the estimates on the parameters ($\mathbf{m}$) represented by the posterior pdf ($\sigma(\mathbf{m}|\mathbf{d})$) or the resolution matrix ($\mathbf{R}$) is constrained by recorded data ($\mathbf{d}$), the prior model pdf ($\rho(\mathbf{m})$) and the forward function (equation (2.1)). SED theory in geophysics consists of methods used to select the optimal experimental setup so that the data expected to be recorded will best constrain the parameters of interest in the inversion process.

The forward function (equation (2.1)) can be written as

$$\mathbf{d} = \mathbf{F}_{\xi}(\mathbf{m}) \tag{2.27}$$

where the subscript ξ in the forward function $\mathbf{F}_{\xi}$ indicates that the parameter-

data relationship is dependant on the experimental design ξ where ξ is a vector representing, for example, source and receiver types and locations, or any other pertinent aspects of the design. Bayes' theorem can also be given as

$$\sigma(\mathbf{m}|\mathbf{d},\xi) = \frac{\theta(\mathbf{d}|\mathbf{m},\xi)\,\rho(\mathbf{m})}{\sigma(\mathbf{d}|\xi)}. \tag{2.28}$$

The goal of the design procedure is to select the elements of ξ such that the information recorded by the experiment best constrains the model parameter information in the posterior pdf $\sigma(\mathbf{m}|\mathbf{d},\xi)$.

The SED problem is therefore an optimisation problem, where prior to any data collection, the inverse problem expected to be solved is optimised (Maurer et al., 2010). If logistics and costs are ignored then the objective function that is maximised is a measure of the information expected to be provided in an experiment for a given survey design

$$\Phi(\xi) = E_{\mathbf{d}}\{I[\sigma(\mathbf{m}|\mathbf{d},\xi)]\} \tag{2.29}$$

where $E_{\mathbf{d}}$ is the expectation (average) operator over all possible datasets $\mathbf{d}$ and $I[\sigma(\mathbf{m}|\mathbf{d},\xi)]$ is a measure of the information content provided by the posterior distribution for a given design. To calculate $I[\sigma(\mathbf{m}|\mathbf{d},\xi)]$ requires that the distribution in equation (2.28) can be evaluated, and in most cases in Geophysics this involves solving the inverse problem of constraining parameter $\mathbf{m}$ given data $\mathbf{d}$ and forward function $\mathbf{F}_\xi$ (e.g. Tarantola, 2005). In nonlinear problems this can be computationally demanding even for a single measured data set $\mathbf{d}$. In the experimental design situation, however, no data has yet been measured; hence, this problem has to be solved for all possible data sets $\mathbf{d}$ that might be recorded during the experiment so that the information I that would be obtained can be calculated for each and thus the objective function (equation (2.29)) can be estimated. The computational cost of such an approach thus typically becomes extraordinarily high. For this reason all the design papers introduced in Section 1.2 except those by Van den Berg et al. (2003, 2005); Winterfors and Curtis (2008, 2010) rely on the linearised inverse theory approximations shown above to design the surveys.

2.3 Information Measures

2.3.1 Nonlinear

Shannon Information

In order to maximise the value of $\Phi(\xi)$ in equation (2.29), the amount of information represented by a pdf needs to be quantified. The entropy of any probability distribution $P(\mathbf{x})$ is related to Shannon's measure of information (Shannon, 1948) as

$$Ent(\mathbf{X}) = -I[P(\mathbf{x})] + c = -\int_{\mathbf{x}} P(\mathbf{x}) \log[P(\mathbf{x})]\, d\mathbf{x} = -E\{\log[P(\mathbf{x})]\} \quad (2.30)$$

where Ent is the entropy function, $P(\mathbf{x})$ is the pdf of the random variable $\mathbf{X}$, I is the information measure as defined by Shannon (1948), c is a constant, and E is the expectation operator. Figure 2.3 shows three pdfs for Gaussian distributions with mean values of 0 and different standard deviations. The analytical expression for Shannon's Information content for a Gaussian distribution is given by

$$I = -\frac{1}{2}\log(2\pi e\tau) \quad (2.31)$$

where e is the exponent of the natural logarithm and τ is the standard deviation. The information content is therefore greatest for the distribution with the smallest standard deviation which makes sense intuitively since post-inversion the best result should have the tightest constraint on the parameter estimate. The definition in equation (2.30) extends the concept of information to pdfs of any form, in which case entropy can be seen to be a measure of uncertainty.

A fully nonlinear quality measure can therefore by defined (after Lindley, 1956) as

$$\Phi(\xi) = -\int_D Ent[\sigma(\mathbf{m}|\mathbf{d},\xi)]\, \sigma(\mathbf{d}|\xi)\, d\mathbf{d} \quad (2.32)$$

where $-Ent[\sigma(\mathbf{m}|\mathbf{d},\xi)]$ represents the amount of information (up to a constant value c in equation (2.30)) contained in the posterior pdf $\sigma(\mathbf{m}|\mathbf{d},\xi)$ about the parameters $\mathbf{m}$ given a particular data measurement $\mathbf{d}$ recorded using the design ξ. The quality measure is calculated taking the expectation of the entropy function

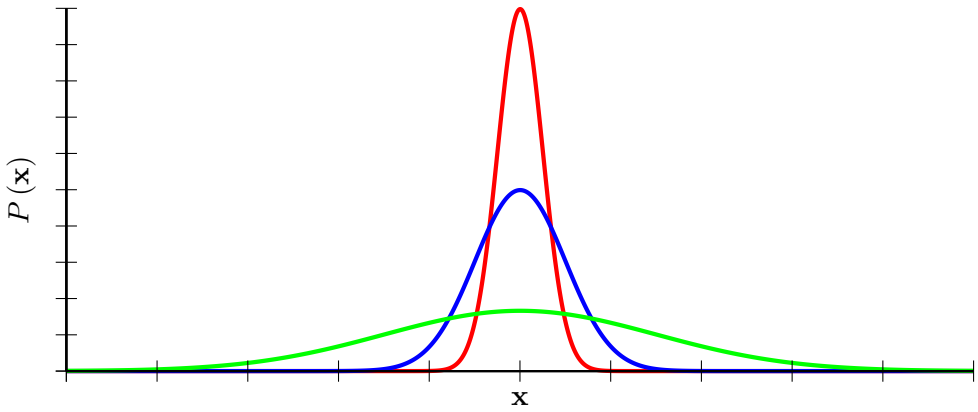

Figure 2.3: Pdf functions for three Gaussian distributions. All three distributions have the same mean value but differing standard deviations. The information content as defined by equation (2.30) is greatest for the distribution with the smallest standard deviation and lowest for the distribution with the largest standard deviation.

with respect to the marginal distribution (equation (2.12)) over all possible data measurements. To evaluate $-Ent\left[\sigma\left(\mathbf{m}|\mathbf{d},\xi\right)\right]$ requires that the distribution in equation (2.28) can be evaluated. As shown above this represents finding the solution to an inverse problem which in nonlinear problems can be extremely demanding computationally.

Shewry and Wynn (1987) introduced an alternative approach. If the random variable, $\mathbf{X}$ in equation (2.30) is partitioned into two parts $\mathbf{X} = [\mathbf{X_S}, \mathbf{X_{\bar{S}}}]$ where $\mathbf{S}$ represents the set of indices that are in the first part and $\bar{\mathbf{S}}$ in the second part then using equation (2.5) to expand $P\left(\mathbf{X_S}, \mathbf{X_{\bar{S}}}\right)$ and inserting this into equation (2.30) gives

$$I\left(\mathbf{X}\right) = I\left(\mathbf{X_S}\right) + E_{\mathbf{X_S}}\left[I\left(\mathbf{X_{\bar{S}}}|\mathbf{X_S}\right)\right] \tag{2.33}$$

where the expectation is taken by integrating over $\mathbf{X_S}$ only. The order of the partitions in equation (2.33) is arbitrary, thus reversing the order gives

$$I\left(\mathbf{X}\right) = I\left(\mathbf{X_{\bar{S}}}\right) + E_{\mathbf{X_{\bar{S}}}}\left[I\left(\mathbf{X_S}|\mathbf{X_{\bar{S}}}\right)\right]. \tag{2.34}$$

Vectors $\mathbf{D}$ and $\mathbf{M}$ can be defined to represent the random variables to represent the data measurements and model parameters respectively. Let $X = [\mathbf{D}, \mathbf{M}]$ represent the joint random vector. The dependance of the experimental design on the variables can be explicitly added to the joint distribution and then equations

(2.33) and (2.34) can be given in pdf form as

$$I\left[\sigma\left(\mathbf{d},\mathbf{m}|\xi\right)\right] = I\left[\sigma\left(\mathbf{d}|\xi\right)\right] + E_{\mathbf{d}}\left\{I\left[\sigma\left(\mathbf{m}|\mathbf{d},\xi\right)\right]\right\} \tag{2.35}$$

and

$$I\left[\sigma\left(\mathbf{d},\mathbf{m}|\xi\right)\right] = I\left[\rho\left(\mathbf{m}\right)\right] + E_{\mathbf{m}}\left\{I\left[\sigma\left(\mathbf{d}|\mathbf{m},\xi\right)\right]\right\}. \tag{2.36}$$

Combining equations (2.35) and (2.36) gives an expression for the expected posterior information on model parameter given recorded data as

$$E_{\mathbf{d}}\left\{I\left[\sigma\left(\mathbf{m}|\mathbf{d},\xi\right)\right]\right\} = I\left[\rho\left(\mathbf{m}\right)\right] + Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right] - E_{\mathbf{m}}\left\{Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right]\right\} \tag{2.37}$$

where the term $I\left[\rho\left(\mathbf{m}\right)\right]$ can be treated as a constant. Equation (2.37) shows that the inverse problem in equation (2.32) need not be solved to design an optimal survey design. A nonlinear design measure $\Phi\left(\xi\right)$ can therefore be defined (after Shewry and Wynn, 1987) as

$$\Phi\left(\xi\right) = Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right] - \int Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right]\rho\left(\mathbf{m}\right)d\mathbf{m}. \tag{2.38}$$

Shewry and Wynn (1987) showed that this represents a measure of the parameter information expected to be gained by performing the experiment. The design measure combines the uncertainty embodied in the marginal distribution $\sigma\left(\mathbf{d}|\xi\right)$ (the first term on the right) which represents the pdf of the data expected to be recorded given a specific design, and a measure of the average data uncertainty $Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right]$ (second term) over all possible models given the same specific survey design. In cases where the data error is not design dependent, the second integral term in equation (2.38) can be assumed constant.

Essentially Shewry and Wynn (1987) showed that the data space uncertainty as defined by equation (2.38) is directly related to the model space information: maximising the former with respect to design ξ also maximises the latter. Most importantly though, to calculate $\Phi\left(\xi\right)$ in equation (2.38) only requires that the prior information on parameters $\rho\left(\mathbf{m}\right)$ is projected through the physical relationship $\mathbf{F}_{\xi}\left(\mathbf{m}\right)$(to calculate $\theta\left(\mathbf{d}|\mathbf{m},\xi\right)$ and $\sigma\left(\mathbf{d}|\xi\right)$). Maximising $\Phi\left(\xi\right)$ thus only requires that the forward function (rather than the inverse problem) be evaluated.

In general the pdfs required to calculate the entropy are not known analytically

and must therefore be evaluated numerically. To calculate a numerical approximation for $Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right]$, samples of $\rho\left(\mathbf{m}\right)$ are generated and projected through the physical relationship

$$\mathbf{d} = \mathbf{F}_{\xi}\left(\mathbf{m}\right) + \epsilon \tag{2.39}$$

into a discretised data space. The physical relationship used incorporates a random associated measurement error ϵ. The resulting discretised data space histogram of all projected samples is normalised to have unit volume, wherefore it represents a numerical approximation to the posterior pdf $\sigma\left(\mathbf{d}|\xi\right)$. The quantity $Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right]$ can then be approximated as

$$Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right] \approx \sum_{i} \hat{\sigma}\left(\mathbf{d}_i\right) \log\left[\hat{\sigma}\left(\mathbf{d}_i\right)\right] \tag{2.40}$$

where $\mathbf{d}_i$ represents the centre of the ith discretised bin in the data space.

To accurately estimate the pdf $\theta\left(\mathbf{d}|\mathbf{m},\xi\right)$, multiple realisations of the same model and experimental design are required, and for each realisation many data values (with different random errors) are calculated and histogrammed in a discretised data space. The resulting data space represents an estimate of the uncertainty in calculating the data value due to the measurement noise for the given experimental setup. The histogram is normalised to represent a numerical approximation to the pdf $\theta\left(\mathbf{d}|\mathbf{m},\xi\right)$. A numerical approximation can therefore be calculated for the integral term in equation (2.38) as

$$\int Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right]\rho\left(\mathbf{m}\right)d\mathbf{m} \approx \frac{1}{\mathrm{M}}\sum_{i} Ent\left[\theta\left(\mathbf{d}|\mathbf{m}_i,\xi\right)\right] \tag{2.41}$$

where M is the total number of models sampled from the prior distribution $\rho\left(\mathbf{m}\right)$.

For each of the distributions the approximation is affected by the number of parameter samples, the data space discretisation bin size and the measurement error.

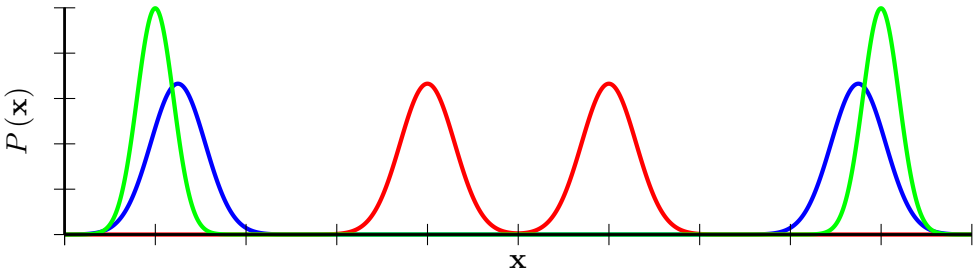

Figure 2.4: Pdf functions for three multimodal Gaussian distributions. Although the distribution shown in red has the smallest difference in possible maximum and minimum values, the distribution in green has the largest Shannon information measure. The distributions shown in blue and red result in identical information values.

Variance

The variance of a random variable $\mathbf{X}$ is a measure of the amount of variation within the values of that variable and is defined as

$$V(X) = \int_X E\left[(X - E(X))^2\right] dX. \tag{2.42}$$

In geophysics, inverse solutions are often deemed better when the variance (or spread) of the result spans a smaller range of parameter values.

The main limitation of using the Shannon information measure to design a survey is that it does not discriminate between designs that result in different variance values in the expected inversion solutions (Maurer et al., 2010). Figure 2.4 shows three multimodal Gaussian pdfs. The distribution shown in red would be classed as the optimal result in terms of minimum variance and the distribution in green would as the worst result of the three possible distributions. An optimal design based on variance therefore aims to minimise the distance between all permissible solutions. Using the Shannon information as a discriminator, however, results in the distribution in green being selected as the optimal result with the other two distributions having identical information values. An optimal solution as defined by Shannon's measure of information minimises the total volume of possible solutions in model space.

An alternative to maximising the Shannon entropy in equation (2.29) is therefore to minimise the expected variance of the posterior pdf $\sigma(\mathbf{m}|\mathbf{d}, \xi)$. Optimal-

ity measures based on Shannon's measure of information and variance measures have both advantages and disadvantages. Although variance measures opt for unimodel solutions rather than multimodal solutions located throughout model space the optimal solution may have a large overall permissible solution range. Although a solution based on Shannon's measure of information may be highly multimodal the optimal solution will have tighter constraints on the parameter estimates.

Even though the design problem of measuring variance is less computationally demanding than calculating the Shannon information the method is still prone to the numerical challenge labelled "the curse of dimensionality" which states that the computational effort required to calculate the integral in equation (2.42) increases geometrically as the dimensionality of the integrand increases (Curtis and Lomax, 2001).

Winterfors and Curtis (2008, 2010) introduced alternative measures related to variance which are efficient enough to be used to design micro-seismic monitoring surveys. In nonlinear problems multiple different sets of model parameters can result in the same data measurement. This leads to increased parameter uncertainties since the data can not discriminate between these various parameter values. Given two different model points ($\mathbf{m}_1$, $\mathbf{m}_2$) a quantitative measure of how much their respective posterior data space pdfs overlap can be given by the "bifocal" measures of Winterfors and Curtis (2008, 2010). One example of a bifocal measure is

$$S\left(\mathbf{m}_1, \mathbf{m}_2, \xi\right) = \int \sigma\left(\mathbf{d}|\mathbf{m}_1, \xi\right) \sigma\left(\mathbf{d}|\mathbf{m}_2, \xi\right) dD. \tag{2.43}$$

This measure simultaneously focuses on two points ($\mathbf{m}_1$, $\mathbf{m}_2$) in model space instead of only one which is the most common approach. However, the measure in equation (2.43) has two distinct disadvantages. First, if $\mathbf{m}_1$ is equal to $\mathbf{m}_2$ then the bifocal measure $S\left(\mathbf{m}_1, \mathbf{m}_2, \xi\right)$ is large even though this case does not contribute to uncertainty in estimates of the model parameters $\mathbf{m}$. Second, the distribution $S\left(\mathbf{m}_1, \mathbf{m}_2, \xi\right)$ has the same units as $\sigma\left(\mathbf{d}\right)$ implying that $S\left(\mathbf{m}_1, \mathbf{m}_2, \xi\right)$ will increase with decreased observational uncertainty, which is the opposite of what would make a useful measure of parameter uncertainty.

Winterfors and Curtis (2010) solved the first problem by multiplying the bifocal

measure by the squared distance between the model parameters $\mathbf{m}_1$ and $\mathbf{m}_2$ ($d^2(\mathbf{m}_1, \mathbf{m}_2)$). The second problem can be addressed by dividing equation (2.43) by a measure of the average observational pdf

$$D_0 = \int \sigma(\mathbf{d}|\xi)\,\sigma(\mathbf{d}|\xi)\,dD \tag{2.44}$$

resulting in the ambiguity measure (after Winterfors and Curtis, 2010)

$$R(\mathbf{m}_1, \mathbf{m}_2, \xi) = \frac{d^2(\mathbf{m}_1, \mathbf{m}_2)}{D_0} S(\mathbf{m}_1, \mathbf{m}_2, \xi). \tag{2.45}$$

A global measure of the ambiguity can then be assessed by taking the expectation of $R(\mathbf{m}_1, \mathbf{m}_2, \xi)$ over all possible model space pairs with respect to the prior model $\rho(\mathbf{m})$

$$W(\xi) = \int\int \rho(\mathbf{m}_1)\,\rho(\mathbf{m}_2)\,R(\mathbf{m}_1, \mathbf{m}_2, \xi)\,dM dM \tag{2.46}$$

The expected observational ambiguity $W(\xi)$ is thus a measure of the average ambiguity of all possible observations, given a specific survey design.

Winterfors and Curtis (2010) showed that the observational ambiguity measure $W(\xi)$ is the expected variance of the posterior pdf ($\sigma(\mathbf{m}|\mathbf{d}, \xi)$) weighted by the marginal distribution $\sigma(\mathbf{d}|\xi)$. In linear problems it is equal to the variance. Similar to the advantages introduced by Shewry and Wynn (1987), calculation of the ambiguity measure (equation (2.46)) only requires the forward function to be calculated, rather than the posterior distribution which is required to calculate the variance directly.

Minimising the Winterfors and Curtis (2010) ambiguity measure $W(\xi)$ requires an integration over model space M rather than an integration over data space D, and hence does not seem to fit within the general framework of equation (2.29). To optimise with respect to the ambiguity measure, we have to replace $\Phi(\xi)$ by $W(\xi)$ in equation (2.29). However, as noted above, $W(\xi)$ can be thought of as a re-weighted alternative to the variance V over the model space, and the variance V does fit into the framework of equation (2.29) by simply setting $I = V$ (Maurer et al., 2010).

2.3.2 Linearised

In linear and linearised SED methods, the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ (equation (2.16)) can be decomposed numerically into its eigenvalues and eigenvectors. In the context of SED positive eigenvalues correspond to independent pieces of information obtained from the survey, and zero eigenvalues to pieces of information that are unobtainable. The magnitude of the eigenvalues relate to how well each piece of information can be estimated from the data with the exact piece of information being represented by the eigenvector (Curtis, 2004a).

The eigenvalue and eigenvector decomposition of the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, referred to as the information matrix for model parameters (Atkinson and Donev, 1992), thus analyses exactly how many, and which pieces of information are expected to be resolved by a survey (Curtis, 2004a). Optimal experimental design methods based on linearised inversion methods (Section 2.1.2) are concerned with maximising quality measures of the information matrix to minimise post survey parameter uncertainties. The model uncertainty in the post-survey solution in the direction of eigenvector $\mathbf{e}_k$ is inversely proportional to the corresponding eigenvalue λ_k. Hence, the quality measures used are often (weighted) measures of the magnitude of the various eigenvalues λ_k.

A commonly used optimal measure used is the so called D-optimality measure in which the experimental setup (ξ) is designed so that the determinant of the information matrix (the product of all eigenvalues) is maximised. Maximising the determinant is directly equivalent to minimising expected post survey model parameter uncertainties. If errors are Gaussian then the determinant is inversely proportional to the area of the posterior error ellipse. A common measure used in linear problems with Gaussian uncertainties is

$$\Phi_{\mathrm{linear}}(\xi) = k \det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \tag{2.47}$$

where k is a constant. For Bayesian linearised problem, if all posterior distributions are Gaussian then the classical nonlinear estimate for the quality of an experimental design can be used (Atkinson and Donev, 1992; Box and Lucas, 1959; Chaloner and Verdinelli, 1995; Ford et al., 1989):

$$\Phi_{\mathrm{linearised}}(\xi) = \int_M k \det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \rho(\mathbf{m})\, d\mathbf{m}. \tag{2.48}$$

Rather than using a single set of parameter values, the expected design quality measure is calculated over the prior distribution of parameter values. Therefore equation (2.48) represents an average measure of linearised design quality over the prior parameter pdf.

Steinberg et al. (1995) introduced a variation of the D-optimality criteria for designing an optimal seismic network for multiple sources (see Figure 1.4). Steinberg et al. (1995) proposed that for k point sources the measure

$$\Phi_{\text{steinberg}}(\xi) = \sum_{1}^{k} a_j \ln\left[\det\left(\mathbf{A}_j^{\mathrm{T}}\mathbf{A}_j\right)\right] \tag{2.49}$$

be maximised where a_j reflects the relative importance attached to source j. The weighted log determinant criteria was first proposed by Lindley (1956) who derived it from the context of Bayesian design methods. As such the weights a_j are interpreted as representing the prior pdfs on the location of the source hypocentres and must sum to unity (Steinberg et al., 1995).

Two of the other main criteria widely used in the SED literature (e.g. Atkinson and Donev, 1992) include A-optimality and E-optimality. A-optimality seeks to minimise the inverse of the trace of the information matrix, which results in minimising the average variance of the estimates of the model parameters whilst E-optimality seeks to maximise the minimum eigenvalue of the information matrix.

As introduced in Section 1.2.3, Curtis (1999a) introduced a quality measure vector (Θ) based on the eigenvalues of the information matrix. In the notation here it is assumed that the eigenvalues are listed in order of deceasing magnitude (λ_1 largest, λ_N smallest).

$$
\begin{aligned}
\Theta_1 &= \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_1} \quad \text{(it can be shown that all } \lambda_i > 0) \\
\Theta_2 &= \log \lambda_k \quad \text{for pre-defined fixed index } k \\
\Theta_3 &= k \quad \text{such that } \lambda_k > \delta \text{ for some pre-defined tolerance } \delta \\
\Theta_4 &= \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta} \\
\Theta_5 &= \log \left| \mathbf{A}^{\mathrm{T}} \mathbf{A} \right|_{\delta} = \sum_{i=1}^{N} \gamma_i \quad \text{where} \quad \gamma_i = \begin{cases} \log \lambda_i & \text{if } \lambda_i \geq \delta \\ \text{penalty} & \text{if } \lambda_i < \delta \end{cases}
\end{aligned}
$$

Measure Θ_1 seeks to maximise the area under the eigenvalue curve (Figure 1.9). The normalisation by λ_1 ensures that only relative magnitudes of the eigenvalues effects the measure and that the measure cannot be maximised by maximising λ_1 alone.

Barth and Wunsch (1990) suggested that the eigenvalue spectrum might be increased by maximising the eigenvalue λ_k at a pre-defined rank k, Θ_2. Generally eigenvalues below this threshold will be damped or removed in the inversion process. The level which is chosen depends on prior knowledge about the model and data noise levels (Matsu'ura and Hirata, 1982). In theory, a survey should be designed to optimise the eigenvalues about this threshold value. Therefore the value of k should be chosen to be the same or larger than the threshold value.

The practical difficulties involved in applying quality measure Θ_2 can be avoided by using Θ_3 (Curtis, 1999a). The tolerance δ should be chosen *a priori* to reflect the expected data uncertainty. As long as δ is not specified too low Θ_3 is best suited to an overdetermined inverse problem. The measure Θ_4 is based on a measure introduced by Maurer and Boerner (1998b). The measure is most sensitive to eigenvalues equal to δ and therefore has similar sensitivity features to Θ_3 although Θ_4 is also sensitive to the larger eigenvalues.

Whereas the standard D-optimal measure cannot be used in under determined inverse problems, quality measure Θ_5 uses a penalty term for eigenvalues below the chosen threshold value to extend the use of the D-optimality criteria to be used in under determined problems.

2.4 Search Algorithms

Ideally the entire design space would be searched to locate the globally optimal design. In the simple case presented by Rabinowitz and Steinberg (1990), locating six optimal receivers in a coarsely discretised grid (as shown in Figure 1.3) would require the optimality of 91^6 designs to be quantified which is computationally intractable. Locating only three receivers still allows approximately 800,000 unique designs. As a result, two classes of search algorithms have been developed to efficiently locate an optimal, or approximately optimal design. By not searching every possible design the search algorithms are not guaranteed to locate the globally optimal design but should find a locally optimal design.

2.4.1 Stochastic Methods

The random nature of a stochastic design method means that the final design that is chosen will be different each time the same algorithm is run. Stochastic methods, however, have been shown to converge on a local maximum very quickly and efficiently. Several stochastic methods have been used extensively in geophysical SED methods. Two of the most important are DETMAX (Mitchell, 1974) and the Genetic Algorithm (Holland, 1975).

DETMAX Algorithm

The DETMAX method introduced by Mitchell (1974) is an algorithm for constructing D-Optimal designs and has been used in geophysical SED works including Rabinowitz and Steinberg (2000) and Steinberg et al. (1995). The DETMAX algorithm can either design a survey from scratch or be used to augment a previously fixed design. In terms of a receiver location problem the algorithm generates an initial network by selecting N sites at random from the possible source and receiver locations. The algorithm sequentially adds and removes single sources and receivers so as to maximise and increase the D-optimality measure. When the increase in optimality falls below a certain threshold the algorithm increases the number of sources and receivers it removes or adds in a single iteration to ensure (as best as possible) that the solution found is not a local maxima. The

algorithm is also allowed to make "excursions" at random in which networks of various sizes are constructed by adding (or removing) stations to the previous network and eventually returning to an N site network. If no improvement is made during the excursion then the constructed networks are added to a set of "failure networks" called F. The set F is then used to guide the next excursion. The addition of allowing excursions is to enable the algorithm to escape from local maxima. Mitchell (1974) recommended that several starting random designs are used to increase the probability of locating the globally optimal design.

In the case of a multiple source location problem (e.g. Steinberg et al., 1995) the DETMAX algorithm can be altered to so that the design quality proposed in equation (2.49) is maximised rather than the standard determinant measure.

Genetic Algorithm

Genetic algorithms, first introduced by Holland (1975) have an analogy with biological evolution and are closely related to simulated annealing methods. Initially a starting "population" of designs are chosen. The initial population is encoded normally into a binary string. From the starting population a set of Q parents are chosen where the selection is dependant on an optimality measure of the population. From this "parent" population a new "child" generation is created by combining the information from pairs of parents. For any given parent combination the children may either be a direct clone of one of the parents or a combination of the two parents. The probability of the child being a combination of the two parents and the exact location at which the two binary strings are split and combined are called crossover parameters. The child generation may then be liable to a "mutation" where none or several of the binary bits are switched. A new parent population is then chosen from the child population. As successive generations evolve the aspects of the design that contribute to optimal designs tend to remain. The relative amounts and probabilities of crossovers and mutations occurring is problem-dependent and is set by the user. In the genetic algorithm case, the selection of the parents is analogous to the notion of survival of the fittest, the crossover step allows mixing and sharing of information between designs, and the mutation step allows "diversity" to remain in the population (Sambridge and Drijkoningen, 1992). Genetic algorithms have been successively used by Curtis (1999a,b) to design optimal tomography surveys.

2.4.2 Deterministic Methods

Deterministic algorithms have no inherent randomness and therefore produce the same result every time they are performed. The design to which a deterministic algorithms converges therefore often depends on the chosen starting model. Deterministic algorithms are usually constructive or destructive in nature, or a combination of the two. Constructive algorithms initially start with a base design of a low number of measurement points, and extra optimal measurement points are added until a predefined number of data points is reached. In destructive methods the design is initiated with a very large set of measurement points. Data measurements expected to provide minimum information are then removed in turn until the final solution is found. As with stochastic methods the final design is only guaranteed to be locally optimal.

Two of the main references in Chapter 1 to introduce deterministic methods are those by Curtis et al. (2004) and Stummer et al. (2004) both of which are based on linearised inverse theory.

The method of Stummer et al. (2004) is based on a constructive method which is initiated by selecting a design with very few data measurements. In the resistivity case studied by Stummer et al. (2004) this initial design would be a standard design such as a dipole-dipole or Wenner array. Each possible data measurement that could be added to the design is then ranked according to a goodness function based on how much extra information it adds to the resolution matrix and how linearly independent the data is compared to the data measurements already selected. In this way a survey consisting of a predefined number of data measurements can be constructed.

Curtis et al. (2004) approached the problem from the other extreme starting with a complete dataset and removed measurements that add the least information. A penalty function

$$\alpha_i = \sum_{j=1, j\neq i}^{N} \left[\frac{\mathbf{a}^{(i)} \cdot \mathbf{a}^{(j)}}{\|\mathbf{a}^{(i)}\| \|\mathbf{a}^{(j)}\|} \right]^2 \tag{2.50}$$

can be assigned to each row of the $\mathbf{A}$ matrix (equation (2.14)) where $\mathbf{a}^{(i)}$ is the ith row of matrix $\mathbf{A}_j$ and N is the total number of data measurements. The penalty function is a measure of how much linearly independent information each measurement is expected to provide to the experiment compared to the

other measurements. In this way data that do not provide linearly independent information can be removed from the survey (as shown in Figure 1.13). In the case of a seismic survey, Curtis et al. (2004) extended the function to define a penalty value for each source and receiver rather than each data measurement. The algorithm was also modified so the penalty function allows surveys to be focussed on specific model parameters (see Figure 1.14).

2.5 Thesis Objectives

In this thesis, experiments are designed using the nonlinear quality measure defined by Shewry and Wynn (1987) (equation (2.38)). In Chapters 4 and 5 errors are assumed offset-independent and as such the optimal designs are found by maximising the entropy of the marginal data pdf. In Chapter 6 offset-dependent errors are introduced into the design and as such the integral term in equation (2.38) is also included.

Both of the deterministic methods presented above rely on optimising a linearised design measure. Therefore when optimising a nonlinear problem the majority of search methods use stochastic algorithms. In Chapter 4 a deterministic algorithm is presented which is applicable to both linear and nonlinear design measures. The algorithm presented is an iteratively constructive algorithm in which the need to search the full multi-dimensional design space for an optimal solution is reduced to multiple searches of a single-dimension design space.

Chapter 3

Amplitude Versus Offset

3.1 Fundamentals

When seismic waves travel through the earth and encounter a subsurface boundary of contrasting velocity and density properties, the incident energy is partitioned into reflected and refracted energy (Figure 3.1). The amplitude of the seismic wave (initially assumed to be of unit amplitude) which is reflected from the boundary at depth is a function of the incident angle of the wave at the boundary, density ρ_i, and the elastic media properties summarised by the P-wave velocity α_i, and S-wave velocity β_i, for a medium with isotropic layers $i = 1, 2$ above and below the boundary respectively.

The recorded amplitudes of the reflected and transmitted waves (after accounting for geometrical spreading effects during propagation) are given by the solution to the Zoeppritz equations,

$$\begin{aligned}
\cos i_1 A_1 + \frac{\alpha_1}{\beta_1}\sin j_1 B_1 + \frac{\alpha_1}{\alpha_2}\cos i_2 A_2 - \frac{\alpha_1}{\beta_2}\sin j_2 B_2 &= \cos i_1 \\
-\sin i_1 A_1 + \frac{\alpha_1}{\beta_1}\cos j_1 B_1 + \frac{\alpha_1}{\alpha_2}\sin i_2 A_2 + \frac{\alpha_1}{\beta_2}\cos j_2 B_2 &= \sin i_1 \\
-\cos 2j_1 A_1 - \sin 2j_1 B_1 + \frac{\rho_2}{\rho_1}\cos 2j_2 A_2 - \frac{\rho_2}{\rho_1}\sin 2j_2 B_2 &= \cos 2j_1 \\
\sin 2i_1 A_1 - \frac{\alpha_1^2}{\beta_1^2}\cos 2j_1 B_1 + \frac{\rho_2}{\rho_1}\frac{\beta_2^2}{\beta_1^2}\frac{\alpha_1^2}{\alpha_2^2}\sin 2i_2 A_2 + \frac{\rho_2}{\rho_1}\frac{\alpha_1^2}{\beta_1^2}\cos 2j_2 B_2 &= \sin 2i_1
\end{aligned} \tag{3.1}$$

where for a horizontal boundary and incident P-wave, i_1 is the P-wave angle

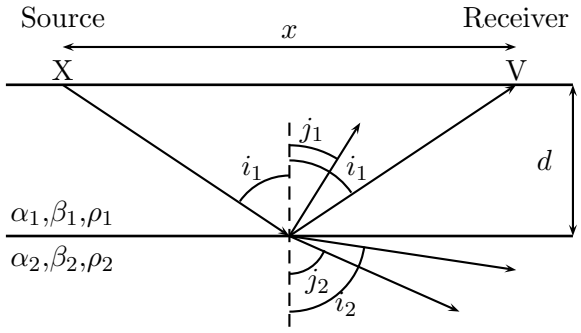

Figure 3.1: Geometry of an AVO experiment with a single interface. The distance between the source (X) and the receiver (V) is called the offset (x). The depth to the interface is d. At the interface, the incident P-wave energy is split into a reflected P-wave and P-S wave conversion, and is also transmitted into the second layer as a P-wave and P-S wave conversion. The amplitudes of each wave are given by equations (3.1). The properties of the subsurface are given by the density (ρ), P-wave velocity (α) and the S-wave velocity (β) in each layer.

of incidence and reflection, i_2 the P-wave angle of refraction, j_1 the converted P-S wave angle of reflection and j_2 the converted P-S wave angle of refraction (Zoeppritz, 1919). The amplitudes of the waves are represented by A_1 for the reflected P-wave, A_2 for the refracted P-wave, B_1 for the reflected S-wave and B_2 for the refracted S-wave. Although the Zoeppritz equations can be solved for each of the reflected and refracted P- and S-waves, the interest in exploration seismology is largely confined to the angle-dependency of the P-to-P reflection coefficient given by A_1 (Yilmaz, 2001). From the reflected P-wave amplitude it is possible, given values for the other elastic properties, to estimate the P-wave velocity of the lower layer (α_2)

$$A_1 = f_x(\alpha_2)\,. \tag{3.2}$$

Figure 3.2 shows the P-wave reflection coefficient as a function of incident angle calculated from equations (3.1) for an upper layer with properties $\alpha = 3000\text{ms}^{-1}$, $\beta = 1200\text{ms}^{-1}$, and $\rho = 1900\text{kgm}^{-3}$ and a lower layer with properties $\alpha = 3400\text{ms}^{-1}$, $\beta = 2200\text{ms}^{-1}$, and $\rho = 2700\text{kgm}^{-3}$. The dotted line in Figure 3.2 indicates the angle at which the critical angle is reached. Around this area the solution to the Zoeppritz equations are highly nonlinear.

The Zoeppritz equations (equations (3.1)) are non-intuitive in terms of their prac-

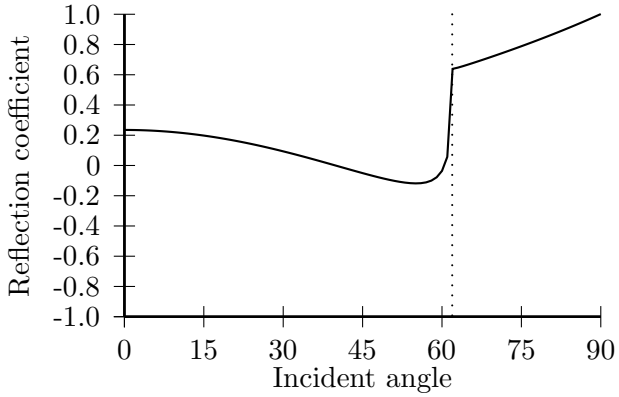

Figure 3.2: P-wave reflection coefficient as a function of incident angle for a boundary with overburden properties $\alpha = 3000\text{ms}^{-1}$, $\beta = 1200\text{ms}^{-1}$, and $\rho = 1900\text{kgm}^{-3}$ and a lower layer with properties $\alpha = 3400\text{ms}^{-1}$, $\beta = 2200\text{ms}^{-1}$, and $\rho = 2700\text{kgm}^{-3}$ as calculated from equations (3.1). The dotted line represents the critical angle for the given interface properties.

tical use as it is hard to understand how changes in the rock parameters and incident angle will effect the reflection coefficient. Bortfeld (1961) simplified the Zoeppritz equations and gave an approximation for the P-to-P reflection coefficient as

$$A_1(i_1) = \frac{1}{2}\ln\left(\frac{\alpha_2\rho_2\cos i_2}{\alpha_1\rho_1\cos i_1}\right) + \left[2 + \frac{\ln\left(\frac{\rho_2}{\rho_1}\right)}{\ln\left(\frac{\alpha_2}{\alpha_1}\right) - \ln\left(\frac{\alpha_2\beta_1}{\alpha_1\beta_2}\right)}\right]\frac{\beta_1^2 - \beta_2^2}{\alpha_1^2}\sin^2 i_1. \quad (3.3)$$

An important aspect of equation (3.3) is that it provides insight into the relationship between the rock properties and the corresponding reflection coefficient. The first term is called the fluid-fluid reflection coefficient (since it depends on the P-wave velocities and P-waves can be supported by a fluid) whilst the second term is called the rigidity term as it also depends on the S-wave velocity (S-waves can not be supported by a fluid). Although equation (3.3) is a simplified version of the Zoeppritz equations for a P-to-P reflection it does not explicitly indicate angle dependence on the reflection amplitudes and is rarely used in practical implementations of AVO analysis (Yilmaz, 2001).

As we are only interested in the P-P reflection coefficient the conventional AVO notation shall subsequently be used with $R(\theta)$ replacing A_1 as the angle-dependent reflection amplitude.

The Bortfeld (1961) equation was further simplified by Aki and Richards (1980) such that the reflection coefficient is derived in a form which is expressed in terms of velocity and density differences across the interface as

$$\begin{aligned} R(\theta) = & \left[\frac{1}{2}\left(1+\tan^2\theta\right)\right]\frac{\Delta\alpha}{\alpha} \\ & - \left[4\frac{\beta^2}{\alpha^2}\sin^2\theta\right]\frac{\Delta\beta}{\beta} \\ & + \left[\frac{1}{2}\left(1-4\frac{\beta^2}{\alpha^2}\sin^2\theta\right)\right]\frac{\Delta\rho}{\rho} \end{aligned} \tag{3.4}$$

where α is the average P-wave velocity $((\alpha_1+\alpha_2)/2)$ and $\Delta\alpha$ is the P-wave velocity contrast $((\alpha_2-\alpha_1))$ with the other quantities defined similarly. $\theta = (i_1+i_2)/2$ represents the average of the P-wave incidence and transmission angles. The Aki and Richards (1980) approximation is split into three terms, each of which represent the contribution to the reflection coefficient as a function of the change in each of the parameters across the interface. In reality, these separate effects are not independently observed, we only observe changes in the reflection amplitude as a function of angle of incidence. Shuey (1985) rearranged and simplified the Zoeppritz equations further to

$$\begin{aligned} R(\theta) = & \left[\frac{1}{2}\left(\frac{\Delta\alpha}{\alpha}+\frac{\Delta\rho}{\rho}\right)\right] \\ & + \left[\frac{1}{2}\frac{\Delta\alpha}{\alpha} - 4\frac{\beta^2}{\alpha^2}\frac{\Delta\beta}{\beta} - 2\frac{\beta^2}{\alpha^2}\frac{\Delta\rho}{\rho}\right]\sin^2\theta \\ & + \left[\frac{1}{2}\frac{\Delta\alpha}{\alpha}\right]\left(\tan^2\theta - \sin^2\theta\right) \end{aligned} \tag{3.5}$$

where each term is related to successive ranges of incidence angle: the first term represents the normal incidence P-to-P reflection coefficient, the second term characterises the reflection coefficient at intermediate angles and the third term becomes significant at large angles. The third term increases as θ^4 and so it does not normally contribute for angles less than 30° (Shuey, 1985). Therefore, at intermediate angles $(0 < \theta < 30)$ the third term can be dropped, leading to the well known Shuey two-term AVO equation

$$R(\theta) = A + B\sin^2\theta \tag{3.6}$$

where A and B are defined as

$$\begin{aligned} A &= \left[\frac{1}{2}\left(\frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho}\right)\right] \\ B &= \left[\frac{1}{2}\frac{\Delta\alpha}{\alpha} - 4\frac{\beta^2}{\alpha^2}\frac{\Delta\beta}{\beta} - 2\frac{\beta^2}{\alpha^2}\frac{\Delta\rho}{\rho}\right] \end{aligned} \tag{3.7}$$

which are commonly known as the AVO intercept and gradient respectively. For a given common midpoint gather the AVO response can therefore be defined by these two parameters, which in turn can be inverted to extract constraints on α, β and ρ for both layers.

The majority of AVO analysis involves gas saturated sandstone reflections since the elastic property contrasts are such that the reflection coefficient amplitudes and gradients are easily detectable (Castagna and Smith, 1994). Rutherford and Williams (1989) introduced three classes of gas saturated sandstones based on the normal incidence P-wave reflection coefficient and how the amplitude varies with increasing offset (Figure 3.3). Class 1 occurs when the normal-incidence P-wave reflection coefficient is strongly positive and shows a strong amplitude decrease with offset, Class 2 has a very small P-wave normal-incident P-wave reflection coefficient and shows a large percentage change in reflection amplitude with offset, and Class 3 have large negative coefficients at normal-incidence which become more negative with offset. In all three cases Rutherford and Williams (1989) stated that the AVO gradient would be negative, however the absolute amplitude can increase with offset as shown by Classes 2 and 3. Castagna et al. (1998) found that some Class 3 anomalies can also show a decrease in amplitude (a positive gradient) and therefore defined a Class 4 to include these.

3.2 AVO Crossplot

Castagna and Swan (1997) introduced the notion of AVO crossplotting, where the values of A and B in equation (3.7) are plotted against each other resulting in an AVO response curve being plotted as a single point in the crossplot (Figure 3.4). The single point in Figure 3.4 (b) represents the result from a single common midpoint gather (Figure 3.4 (a)).

The 4 Classes defined are now represented by areas of the crossplot with Class

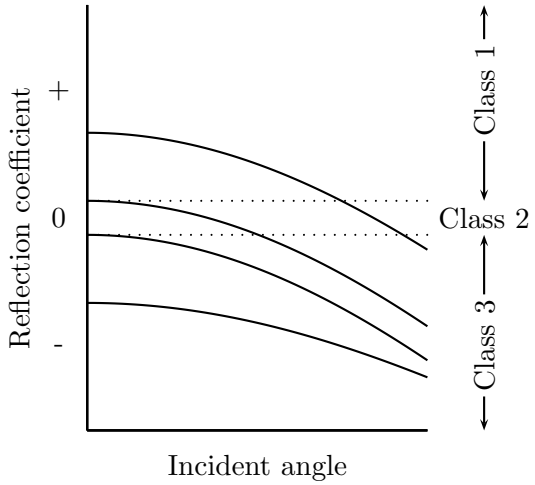

Figure 3.3: Classification of AVO responses based on the zero-offset reflection coefficient, as defined by Rutherford and Williams (1989).

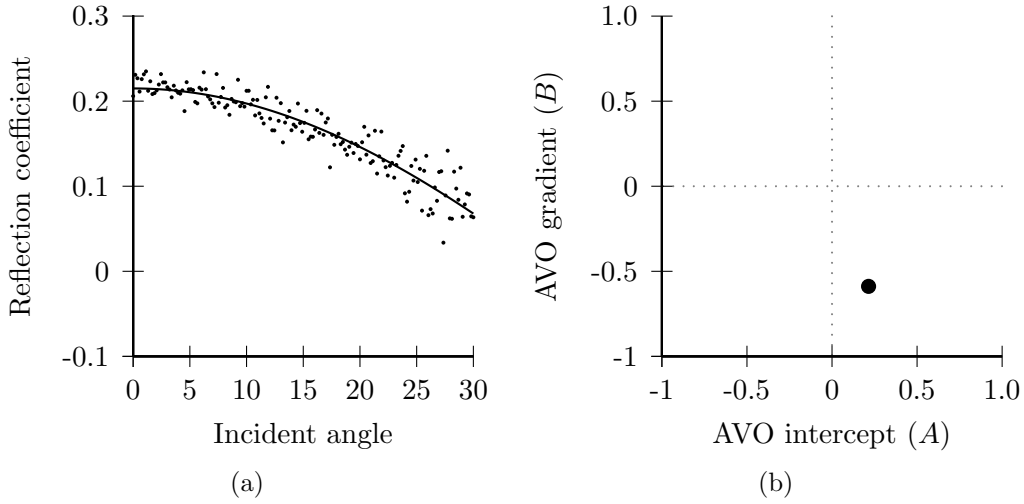

Figure 3.4: (a) Shuey two-term best-fit regression line fitted to discrete amplitude measurements over a 30° offset range. (b) Regression line plotted as a single point in a crossplot of AVO intercept A against AVO gradient B (equation (3.7)).

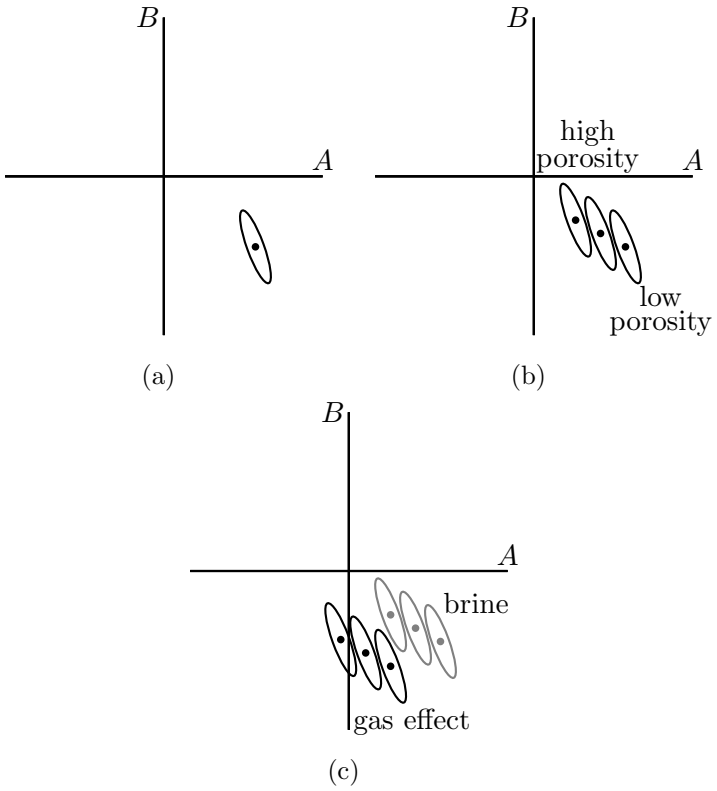

Figure 3.5: Factors that effect AVO crossplots. Plot (a) shows how offset dependent errors in a common midpoint gather appear as an elipse in an AVO crossplot, plot (b) shows how increasing the porosity of a Class 1 sand reduces both the gradient and intercept measurements, and plot (c) shows how changing the saturating fluid from brine to gas results in a mainly lateral shift in the AVO intercept measurement.

1 signatures appearing in the lower right quadrant, Class 2 in the lower middle section on the crossplot, Class 3 signatures in the lower left quadrant and finally Class 4 in the upper left quadrant. By presenting the data in this way trends which would otherwise be missed in the standard coefficient versus offset plots (Figure 3.3) are easily observed.

The crossplot is also an ideal way to examine how fluid and lithologic variations are represented in AVO signatures (Simm et al., 2000). Although the AVO crossplot contains information about the properties of the rock interface, it is not usually detectable in terms of trends directly seen within the crossplot due to the effects of noise (Simm et al., 2000). Figure 3.5 (a) shows the average value and the error associated with processing the same common midpoint gather of a Class 1 reflector multiple times, each time adding random offset-dependent noise to the data. The error is elliptical due to the sensitivity of the gradient estimation to noise (Hendrickson, 1999). Cambois (2000) suggested that the gradient of the noise is related to the P-wave traveltime, velocity structure and offsets over which the AVO gradient and intercept are estimated. Figure 3.5 (b) shows how a change in porosity of the lower rock unit appears in the AVO crossplot. Increasing the porosity has the effect of reducing the amplitude of both the gradient and intercept. Figure 3.5 (c) shows the effect of replacing the brine in the reservoir with gas. As the percentage change from brine to gas increases so does the reduction in the AVO intercept value (Hendrickson, 1999).

3.3 AVO Interpretation

Through the Shuey (1985) two-term equations (3.6) and (3.7) the data displayed in the AVO crossplot contains information about the elastic properties of the layers either side of the subsurface boundary at which the recorded energy was reflected. Prior to the introduction of AVO crossplots, Smith and Gidlow (1987) developed a method for inverting AVO data into rock property estimates by using "weighted stacking". The method applies a least squares fitted curve to a reflection coefficient versus offset angle for multiple common midpoint gathers (as in the AVO crossplot method). The parameters estimates of A and B can then be inverted using either a two-term or three-term approximation. Castagna et al. (1998) introduced a framework for AVO crossplot interpretation. Given

a reasonable petrophysical assumption a linear background trend in the AVO crossplot that passes through the origin can be calculated. Deviations from this background can then be indicative of hydrocarbons (Castagna et al., 1998)

Whilst the two-term inversion is growing in popularity through the acceptance of AVO crossplots as an interpretation tool, it is often overlooked that the two-term Shuey approximation (equation (3.6)) is invalid beyond about 30° angle of incidence whereas using a three-term inversion (equation (3.5)) is valid up to incidence angles of approximately 50° (Downton and Ursenbach, 2006). Whichever interpretation method is used the information provided is nonunique since the elastic properties of different rock units overlap, and fitting a curve to noisy data (Figure 3.4 (a)) results in errors propagating into the AVO crossplots (Figure 3.5).

AVO interpretation is normally performed in a heuristic setting (e.g. Ostrander, 1984; Castagna et al., 1998). In the absence of a hydrocarbon-bearing subsurface the crossplot of AVO intercept A and gradient B often forms a well-defined "background trend". Deviations from this background trend can then be used as a hydrocarbon indicator. The classification of rock types and pore fluids from the AVO intercept and gradient is usually associated with large uncertainties and therefore performing the inversion in a statistical framework seems ideal (Mavko et al., 1998).

Although many different methods have been developed and studied to invert AVO data (e.g. Fatti et al., 1994; Simmons and Backus, 1996; Santoso et al., 1996; Jin et al., 2000; Russell et al., 2003; Duffaut and Landrø, 2007) inversion methods that employ the use of Bayesian statistics (e.g. Houck, 2002; Eidsvik et al., 2004; Larsen et al., 2006; Buland et al., 2008) best fit into the context of this work.

Houck (2002) treated uncertainty in the inversion procedure as the product of two sources: geological uncertainty and measurement uncertainty. Geological uncertainty, as mentioned above arises since different rock units can have the same elastic properties and therefore the interpretation of seismic amplitudes is generally nonunique, and measurement uncertainties arise because observed AVO responses may be influenced by processing workflows, noise or other geological boundaries above the interface of interest. Whereas Mukerji et al. (1998) and Houck (1999) only consider geological uncertainty, Houck (2002) considers both

geological and measurement uncertainty in a Bayesian framework to look at the probability that hydrocarbons are present given AVO data.

Given a seismic observation of AVO ($\mathbf{d}$) which might be, for example, the peak reflection amplitude over a range of offsets, Houck (2002) stated that the post-observation conditional probability that the reflector lithologies L_i produced the given response is given by $P(L_i|\mathbf{d})$. Using the Shuey (1985) two-term approximation, estimates of the AVO intercept (A) and gradient (B) can be calculated from the recorded data (Figure 3.4). Because of measurement uncertainty the values of A and B that would be calculated from the elastic properties of the layers above and below the reflector will be different from those estimated from the seismic data. Therefore associated with the observed data $\mathbf{d}$ is a distribution of possible true values of A and B, $P(A, B|\mathbf{d})$ that quantifies the measurement uncertainty with estimates coming from well ties and seismic data modelling (Houck, 2002).

Even if A and B are known without errors, geological uncertainty does not allow a unique set of lithologies to be determined and therefore there is a separate distribution $P(L_i|A, B)$ relating the probability that lithology $i = 1, \ldots, M$ would result in values A and B being recorded. Applying Bayes' rule, Houck (2002) stated that the probability of the lithology L_i being correct is given by

$$P(L_i|A, B) = \frac{P(L_i)\, P(A, B|L_i)}{P(A, B)} \tag{3.8}$$

where $P(L_i)$ is the prior probability that lithology L_i is correct and represents the main source of prior geologic information, $P(A, B|L_i)$ is the probability of observing AVO intercept A and gradient B given lithology L_i and represents the combination of forward modelling and measured data (see Chapter 6) and $P(A, B)$ is the marginal probability of selecting a reflector with the intercept-gradient pair A,B from all the possible lithologies and is given by

$$P(A, B) = \sum_{j=1}^{M} P(A, B|L_j)\, P(L_j)\,. \tag{3.9}$$

The post-observational probability that is of importance is given by

$$P(L_i|\mathbf{d}) = \int_A \int_B P(L_i|A, B)\, P(A, B|\mathbf{d})\, dA dB. \tag{3.10}$$

Substituting equations (3.8) and (3.9) into equation (3.10) gives the probability

of each possible lithology being true.

By performing an inversion in this way the AVO Class system first introduced by Rutherford and Williams (1989) is no longer required as the information about the lithologies that give rise to the AVO attributes are extracted statistically.

The method introduced by Houck (2002) was further expanded by the works of Eidsvik et al. (2004); Larsen et al. (2006); Buland et al. (2008) amongst others to add more complexity to the Bayesian inversion method. Eidsvik et al. (2004) included well observations and seismic reflection times into the inversion to better constrain the results. The method developed by Larsen et al. (2006) introduced a recursive algorithm. From AVO data recorded at multiple depths the algorithm is able to build a lithology model that predicts all the layers rather than only the standard two layer model. Buland et al. (2008) introduced a method to invert seismic AVO data for 3D lithology as well as a feasibility analysis which calculates the expected posteriori probability of lithology i given that lithology j is true and can be used assess the nonuniqueness of a solution.

3.4 Thesis Objectives

Although all the methods referenced above are different they are all based on the principal of inverting seismic AVO data to infer the lithology of the subsurface boundaries. From a survey design context the aim is to make the inversions more robust by ensuring the recorded data minimise the nonuniqueness or uncertainty in the inversion results. Fundamentally, SED methods can be used to maximise the information content in seismic data (calculated synthetically using the Zoeppritz equations) so that any inversion methods will benefit and this is what is acheived in the methods introduced in Chapters 4 and 5. Alternatively, as shown in Chapter 6, SED methods can be used to design a survey so that the information content provided by the AVO crossplot method is maximised so that interpretation methods such as Houck (2002) will provide more reliable results.

Chapter 4

Nonlinear Design Algorithm

4.1 Introduction

In Section 2.2 it was shown that the forward function $\mathbf{F}_\xi$ representing the relationship between parameters $\mathbf{m}$ and data $\mathbf{d}$ is given by,

$$\mathbf{d} = \mathbf{F}_\xi(\mathbf{m}) \tag{4.1}$$

where the subscript ξ in the forward function indicates that the parameter-data relationship is dependant on the experimental design ξ where ξ is a vector representing, for example, source and receiver types and locations, or any other pertinent aspects of the design.

As previously stated, one of the main reasons that SED theory has not gained general acceptance in the Geosciences is that the relationship $\mathbf{F}_\xi$ is commonly highly nonlinear. The results of applying linear SED theory are therefore not necessarily robust, while applying fully nonlinear design theory is too computationally costly. Consequently, of all the works cited thus far, only the papers of Van den Berg et al. (2003, 2005) and Winterfors and Curtis (2008, 2010) apply non-linearised design theory, and these only addressed one and two dimensional parameter vectors respectively. Since naturally occurring design problems in many other fields are also nonlinear, I suspect that the uptake of SED techniques will be similarly hampered. There is thus a great need for new methods of SED that are applicable

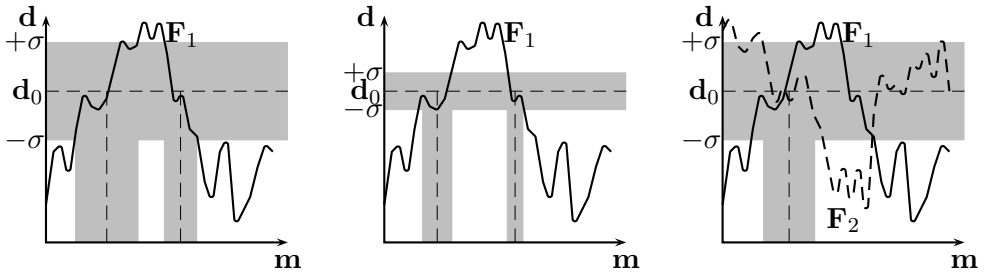

Figure 4.1: Parameter-data relationships for three experimental designs. Designs in (a) and (b) use the same experimental setup so the forward function $\mathbf{F}_\xi$ is identical. The first experiment (a) only uses a single measurement to predict the parameter and hence has a large uncertainty ($\pm\sigma$). The second experiment (b) represents repeated measurements resulting in a reduced data uncertainty and hence a smaller parameter uncertainty. In (c) two data are recorded, both with value $\mathbf{d}_0$, but each with a different forward function (solid and dashed lines).

to multi-dimensional, nonlinear problems without linearisation.

Experimental error and data noise always result in uncertainty on the measurement of data $\mathbf{d}$. Figure 4.1 (a) and (b) show schematically the results of two experiments characterised by the same nonlinear parameter-data forward function. Experiment (a) collects a single data measurement $\mathbf{d}_0$ with associated uncertainty σ, whilst experiment (b) collects multiple identical measurements that reduce the data uncertainty. The aim is to invert equation (4.1) to infer the post-experimental (*a posteriori*) parameter range. Figure 4.1 (a) and (b) show that both inversions result in a bimodal parameter estimate, however experiment (b) is better constrained. Figure 4.1 (c) shows an estimate of the parameters that could be inferred from an experiment using two different parameter-data relationships. Since both experiments are used to constrain the same parameters the results must be consistent. The resultant parameter range is given by the intersection of the two inversion solutions and is shown in Figure 4.1(c) to be unimodal.

Consider designing an experiment that allows $n = 10$ data measurements to be recorded so as to minimise post-experimental uncertainties. Say 20 unique designs are available for each datum, all of which have the same cost. This results in 20^{10} possible designs, and assuming that the optimality of the overall experimental design can be quantified in some way, finding the globally optimal design

out of so many possibilities is generally computationally infeasible. Ryan (2003) proposed a method to calculate the optimal design using Markov chain Monte Carlo techniques without any parameter or data space discretisation (which is in contrast to my method), whilst Muller and Parmigiani (1996) used the assumption that the quantifiable design space is a smooth surface to reduce the computational cost. As noted by Hamada et al. (2001), when the dimension of the data space becomes greater than three the numerical integration required to compute the optimality measure becomes computationally infeasible.

As shown in Section 2.4.2, Curtis et al. (2004) introduced a deterministic, iterative design algorithm that was applied to linearised design of seismic tomography and microseismic monitoring surveys. Their method starts with the design containing the maximum number of receivers conceivable, then in turn, the receiver that provides the least post-experimental information is iteratively removed until a final design with an affordable number of data measurements is obtained. I refer to this kind of algorithm as "iteratively-destructive". Curtis et al. (2004) also described an iteratively-constructive version of their algorithm where at each iteration the datum that provides the most additional information to some minimal design is added until a cost constraint is reached. Such an algorithm was applied by Coles and Morgan (2009), Stummer et al. (2002, 2004), and Wilkinson et al. (2006) to design linearised geoelectrical tomography surveys.

Since these methods are iterative they do not necessarily find globally optimal designs. Furthermore, these methods have only been developed for linearised design problems using linear algebraic results to compute information increases and losses cheaply.

In this chapter an iteratively-constructive method that works for linear or nonlinear problems is introduced, and which reduces the number of designs considered to nx where x is the number of possible single-datum designs, and n is the total number of data to be recorded. In the above example the total number of designs that need to be considered would be reduced to 200. In brief, the algorithm works by first calculating the optimal design for a single-datum measurement (x possible designs to consider). It then designs an experiment that records two data measurements, but instead of considering all x^2 possible two-data designs it fixes the first datum to that found in the previous step. There are therefore only x new designs to consider corresponding to possible designs for the second datum. This

iterative method is repeated for subsequent data one at a time, fixing previously found designs, until the required final number of data measurements has been attained or a cost threshold exceeded.

By not considering every one of the x^n possible design a locally optimal design is found rather than the globally optimal design. However, note first that considering the entire design space is generally infeasible, second it is shown that no linearisation of physics is necessary in iterative methods, and third the final design is found in a more scientifically rigorous manner than when using rules of thumb or other heuristic methods. Additionally, as shown below the designs found can be close to globally optimal, even in nonlinear problems.

It is important to distinguish "iterative design" methods discussed here, from so-called "sequential design" methods (Fedorov, 1972; Ford et al., 1989; Hu, 1998). In sequential methods the data collected from the previous experiment are used to estimate the parameter pdf which is then used to design the next experiment. Sequential design methods thus build on previous experiments. Iterative design methods on the other hand use a constant prior parameter pdf when adding successive data measurements, and are methods to find a (sub)optimal design for a single experiment when full global optimisation is infeasible or too costly to find.

The quality measure and example application (designing amplitude versus offset (AVO) experiments) is similar to the study of Van den Berg et al. (2003, 2005). However, that previous paper found globally optimal solutions and as a consequence was limited to designing a single-datum experiment due to the huge computational demand. In this chapter it is shown how the new, iteratively-constructive algorithm reduces the computation required for nonlinear design problems such that I can extend the range of application to many more data points. More generally, I believe that iteratively-constructive or -destructive methods may represent one of the best available options to extend fully nonlinear design methods to real-world, high dimensional scientific problems.

4.2 Iterative Design Theory

In Section 2.3.1 it was shown that a nonlinear design measure where the data error is design independent can be defined as

$$\Phi(\xi) = Ent\left[\sigma(\mathbf{d}|\xi)\right]. \tag{4.2}$$

In general, the pdf $\sigma(\mathbf{d}|\xi)$ required to calculate the entropy is not known analytically and must therefore be deduced numerically. The Monte Carlo method adopted here varies from that used by Van den Berg et al. (2003, 2005). Samples of the prior pdf $\rho(\mathbf{m})$ are generated and projected through the forward function

$$\mathbf{d} = f_{\xi}(\mathbf{m}) + \epsilon \tag{4.3}$$

into a discretised data space. The physical relationship used incorporates the random associated measurement error ϵ. The resulting data space histogram of all projected samples is normalised to have unit volume, whereafter it represents a numerical approximation to the posterior pdf $\sigma(\mathbf{d}|\xi)$. This approximation of $\sigma(\mathbf{d}|\xi)$ is then used to assess the quality of the experimental design by calculating the entropy,

$$Ent\left[\sigma(\mathbf{d}|\xi)\right] \approx \sum_{i} \hat{\sigma}(\mathbf{d}_i) \log\left[\hat{\sigma}(\mathbf{d}_i)\right] \tag{4.4}$$

where $\mathbf{d}_i$ represents the centre of the i$^{\text{th}}$ discretised bin in the data space. For any given experimental design, $\hat{\sigma}(\mathbf{d}|\xi)$ is affected by the number of parameter samples, the data space discretisation bin size, and the measurement error.

In the method of Van den Berg et al. (2003, 2005), the entropy is calculated for every possible design. The design that corresponds to the maximum entropy value is deemed the optimal design. In iteratively-constructive methods on the other hand,

$$\xi_j = \arg\max\{Ent\left[\sigma(\mathbf{d}|\xi_j)\right]\}, \text{ such that is } \xi_{j-1} \text{ fixed} \tag{4.5}$$

where $\xi_j = (\xi^1 \ldots \xi^j)$ and ξ^i is an element of the design vector. The new optimal design ξ_j combines the data measurements of the previous iteration ξ_{j-1} augmented by the single datum measurement that maximises the entropy of the posterior data pdf given that ξ_{j-1} remains fixed.

The same method can be used in an iteratively-destructive manner were all possible data are recorded and one at a time the datum that provides least information is removed until the final experimental design is found. Although this method is plausible in principle, as shown below it would prove to be computationally infeasible.

4.2.1 Design of Numerical Implementation

The ability to evaluate the right side of equation (4.5) constitutes the computational limiting factor in the new method. While the computation is greatly reduced compared to that required to find the global optimum of

$$\arg\max\left[Ent\left\{\sigma\left(\mathbf{d}|\xi\right)\right\}\right] \tag{4.6}$$

used in previous studies, it nevertheless increases significantly for increasing data space dimensions, as shown in examples below. In Section 4.5 it is shown that the introduction of a second design sub-problem, to design dynamically the number of data space samples required to design each new ξ_j accurately, greatly enhances the method's performance.

4.3 Synthetic Test: Sawtooth Functions

The performance of the new design method is now evaluated by applying it to a highly nonlinear, synthetic problem for which there is an analytic solution for the projected data space marginal distribution $\sigma\left(\mathbf{d}\right)$. Imagine that the experimental situation is such that by choosing different designs the forward function $f_\xi\left(m\right)$ in equation (4.3) can be altered to be a sawtooth function with an integer number ξ of periods in the range $m \in [0, 10]$ (Figure 4.2). Prior information on m is assumed to be uniform in this range. Van den Berg et al. (2003, 2005) illustrated that classical Bayesian (linearised) nonlinear measures (equation (2.48)) fail when applied to a one-dimensional sawtooth forward function. Such measures deliver as optimal design a sawtooth with as many periods as possible, corresponding to a forward function with the highest possible average gradient. This results in a post-experimental parameter space posterior pdf that is highly multi-modal, but

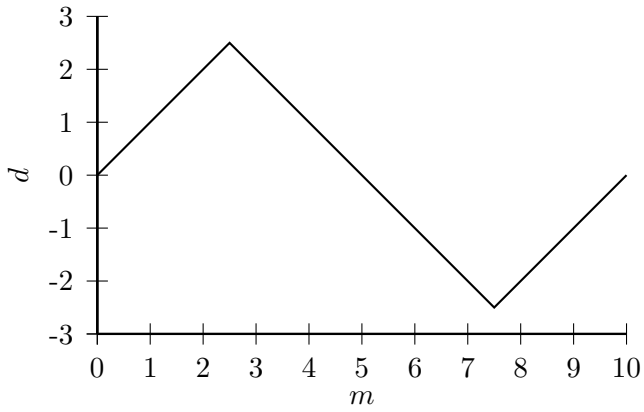

Figure 4.2: Single-period sawtooth function with maximum amplitude 2.5 in the parameter range $m \in [0, 10]$.

which in one dimension provides no Shannon information gain over a single-period sawtooth function.

Using the Shannon entropy value as the optimality measure the accuracy of the iteratively-constructive method is assessed against the globally optimal results for multi-dimensional sawtooth functions

$$\mathbf{f}_{\xi}(\mathbf{m}) = \{f_{\xi_1}(m), \ldots, f_{\xi_n}(m)\}. \tag{4.7}$$

The distribution $\theta(\mathbf{d}|\mathbf{m})$ is defined to have Gaussian error with standard deviation 0.1 around mean $\mathbf{f}_{\xi}(\mathbf{m})$, truncated at ± 3 standard deviations from the mean. The analytical value of $\sigma(\mathbf{d})$ for any of the sawtooth functions (without truncation of the Gaussian) is then given by

$$\sigma(\mathbf{d}) = 0.1\left[-erf\{5\sqrt{2}(-2.5+d)\} + erf\{5\sqrt{2}(2.5+d)\}\right] \tag{4.8}$$

where erf is the error function. Equation (4.8) is therefore used as a very close approximation to the truncated Gaussian distribution for any single sawtooth forward function.

Consider first a sawtooth function with a single period over the parameter range (Figure 4.2). The analytical value for the entropy of the single sawtooth function is equal to 1.645. $\sigma(\mathbf{d})$ is approximated numerically using different data space discretisation lengths and total number of parameter space samples. Figure 4.3 shows the numerical approximation for the entropy as a function of data space

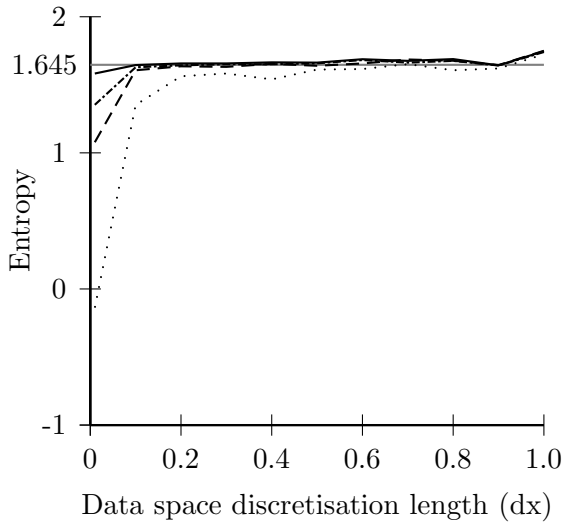

Figure 4.3: Entropy values for the single-period sawtooth function as a function of data space discretisation length (dx) for four experiments with different numbers of parameter space samples (dotted = 100, dashed = 500, dot-dashed 1000 and the solid 5000). The grey line represent the analytical value of 1.645.

discretisation length and number of parameter space samples for the single period sawtooth function. The plot shows that as the data space discretisation becomes small the total number of samples needed to accurately sample the data space becomes large. When the data space discretisation length becomes large then even several thousand parameter space sample points are insufficient to approximate equation (4.8) accurately. For the remaining sawtooth function examples we use a data space discretisation length of 0.1 with 500,000 random parameter space samples.

It is instructive to consider the simple problem of designing a two-datum experiment where each datum could result in sawtooth functions with either $\xi_i = 1$ or 2 periods between $m \in [0, 10]$. If the second function had $\xi_2 = 1$ period, say, this leaves two possible designs: either a repeat experiment $\xi_1 = 1$, or a combination experiment $\xi_1 = 2$ using a single instance of 1 and 2 periods. The same uniform prior parameter range and Gaussian errors on each datum are used as in the single-sawtooth function experiment, and Figure 4.4 shows the resulting $\sigma(\mathbf{d})$ functions. The repeat experiment (b) shows a linear trend since any given parameter value results in approximately the same datum for both sawtooth functions. The design using both sawtooth functions in (a) shows lower $\sigma(\mathbf{d})$ values than (b) because a larger proportion of the data space has been sampled using the two

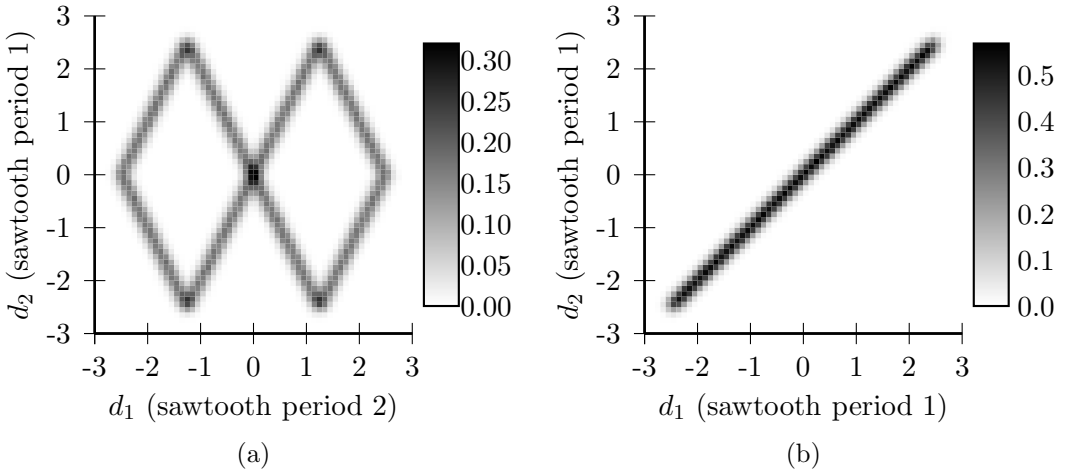

Figure 4.4: Plots of $\sigma(\mathbf{d})$ for (a) a design that uses both period sawtooth functions, and (b) a repeat experimental design, calculated using 500,000 parameter space samples and a data space discretisation length of 0.1. Scale bar represents the normalised $\sigma(\mathbf{d})$ value for each data space discretisation.

different period functions.

The entropy $Ent[\sigma(\mathbf{d}|\xi)]$ in equation (4.4) for the repeat experiment in (b) is 1.126 while the entropy for the experiment in (a) using both one-period and two-period functions is 2.184. The data expected to be recorded from repeat experiments will therefore always provide less post-experimental information on parameters than an experiment using different sawtooth functions for each datum.

The performance of the iteratively-constructive method is now assessed using this challenging forward function for each f_i between 1 and 10 periods in the range $m \in [0, 10]$. Since Van den Berg et al. (2003, 2005) showed that the entropy value is constant for any period of sawtooth function in the single-datum problem, I begin the iteratively-constructive method at the globally optimal design for a two-data scenario. Figure 4.5 shows the entropy of $\sigma(\mathbf{d})$ for all possible two-data experiments. The optimal design uses $\xi_1 = 9$ and $\xi_2 = 10$ periods. As was observed in the previous example, performing a repeat experiment (represented by cells on the diagonal) provides minimal post-experimental information compared with any other design.

According to our iterative method the optimal design for a three-sawtooth function setup is calculated by first fixing the periods of the two-sawtooth design at

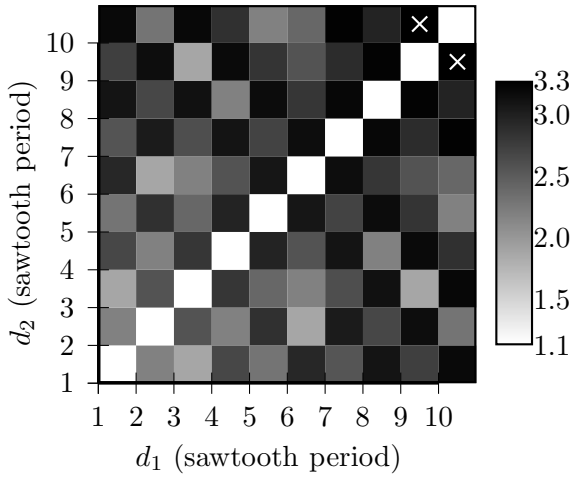

Figure 4.5: Two-data entropy values of entropy of $\sigma(\mathbf{d})$ for the complete design space for pairs of sawtooth functions with integer periods ranging from 1 to 10 over the parameter range 0 to 10. For each experiment, the data space discretisation length is 0.1 and 500,000 samples are drawn at random from a Uniform prior distribution over the parameter range. The white crosses indicate the optimal two-datum experimental design.

$\xi = [9, 10]$. The only sawtooth period to be designed is that of the third function, ξ_3. The entropy is calculated for the 10 possible resulting three-sawtooth experiments, and the maximum was found to correspond to period $\xi_3 = 8$. To compare this to the globally optimal result, the entropy of all possible 10^3 designs using three sawtooth functions was calculated. In this case, the locally optimal design found using the iterative method exactly matches the global optimum.

The iterative method is used again to calculate the optimal four-sawtooth design, and predicts that the functions with periods 8, 9, 10 and 10 are optimal. In this case the optimal design uses a repeated data measurement rather than using another function with a different period. The globally optimum design was also located from the possible 10^4 designs. Again the local optimum of the iterative method exactly matches the global optimum whilst only searching a 10-element design space, a saving of three orders of magnitude in computing power.

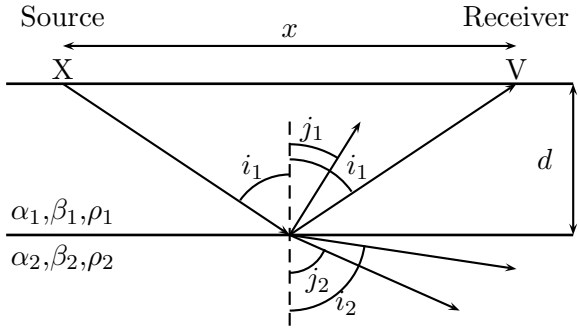

Figure 4.6: Geometry of an AVO experiment with a single interface. The distance between the source (X) and the receiver (V) is called the offset (x). The depth to the interface is d. At the interface, the incident P-wave energy is split into a reflected P-wave and P-S wave conversion, and is also transmitted into the second layer as a P-wave and P-S wave conversion. The amplitudes of each wave are given by equations (3.1). The properties of the subsurface are given by the density (ρ), P-wave velocity (α) and the S-wave velocity (β) in each layer.

4.4 Geophysical Application: Multiple AVO Receivers

Having gained confidence in the new method, I now apply it to the practical geophysical problem addressed by Van den Berg et al. (2003, 2005). The computational savings offered by our method when applied to placing multiple receivers allows us to extend their results to far more complex designs than they found computationally tractable.

The data will be the amplitudes of planar seismic waves generated with unit amplitude at the ground surface, reflected from a subsurface boundary between two geological layers at depth d, and recorded when they arrive again at the ground surface (Figure 4.6).

Each is a function of the incident angle of the wave at the boundary, density ρ_i and the elastic media properties (summarised by the P-wave velocity α_i, and S-wave velocity β_i, for an isotropic medium) of both layers $i = 1, 2$. The recorded amplitude A_1 is given in full by the solution to the Zoeppritz equations (equations (3.1)). From the reflected P-wave amplitude it is possible, given values for the other elastic properties, to estimate the P-wave velocity of the lower layer (α_2)

$$A_1 = f_x(\alpha_2)\,. \tag{4.9}$$

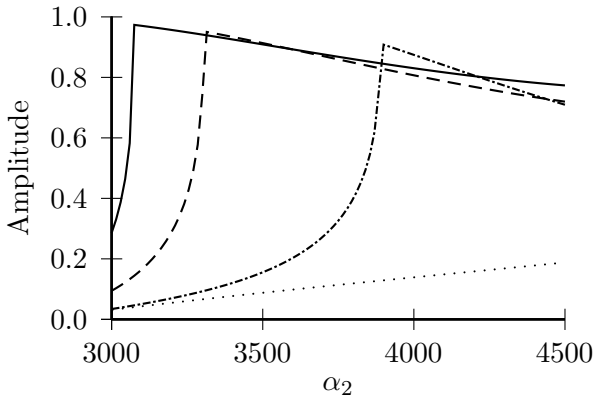

Figure 4.7: Reflected P-wave amplitude data as a function of parameter α_2 for four different offsets, $x = 500$m (dotted line), $x = 1000$m (dot-dash line), $x = 1500$m (dashed line) and $x = 2000$m (solid line). The velocity of the top layer is 2750 ms^{-1} and the depth to the reflecting boundary is 500m.

The initial design problem is to find the single offset (x) between one source and one receiver that best constrains the value of α_2, given the P-wave reflection coefficient A_1 measured at that offset.

I assume that both layers are represented by a Poisson medium in which $\beta = c\alpha$ where $c = 1/\sqrt{3}$, and that there is no significant density contrast between the layers. Figure 4.7 shows the P-wave reflection amplitudes as a function of the parameter α_2 for four different offsets using a depth of 500m for the reflecting boundary and a P-wave velocity of 2750ms^{-1} for the top layer. The discontinuities in the curves occur when the critical angle of incidence is reached for the given velocity structure (the angle at which the refracted P-wave becomes horizontal).

For the design problem the prior parameter information about α_2 is specified by a uniform distribution ranging from 3000ms^{-1} to 4500ms^{-1}, and $\theta(d|m)$ is given by the numerical solution of equations (3.1) with added Gaussian uncertainty of standard deviation 0.01 to simulate measurement uncertainties.

Entropies are calculated for offsets ranging from 0m to 5000m at 25m intervals. For each offset the data discretisation length is 0.01 and 500,000 samples are drawn at random from the uniform parameter space. Van den Berg et al. (2003, 2005) used the same prior information and geometrical setup to find the optimal receiver offset. Figure 4.8 shows the entropy values as a function of offset for my results and the most densely sampled results of Van den Berg et al. (2003, 2005).

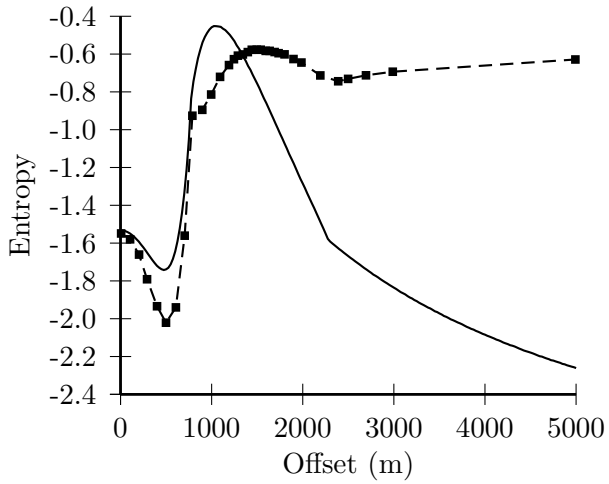

Figure 4.8: Entropy values as a function of offset for a single receiver. The most densely-sampled results of Van den Berg et al. (2003, 2005) are represented by the dashed line (data space discretisation length of 0.001 using 300,000 samples). The solid line represents the results using a data discretisation length of 0.01 using 500,000 samples.

The discrepancy between the results is due to different forward functions being used to calculate the reflection coefficient. Van den Berg et al. (2003, 2005) used the Aki and Richards (1980) approximation to the solution to the Zoeppritz equations (equation (3.4)). Assuming a Poisson medium ($\beta = c\alpha$, where $c = 1/\sqrt{3}$) and no density contrast between the layer ($\Delta\rho = 0$) simplifies the equation to that used by Van den Berg et al. (2003, 2005):

$$R_P = \left(\frac{1}{2}\left(1 + \tan^2 i\right) - 4c^2 \sin^2 i\right) \frac{\Delta\alpha}{\alpha} \tag{4.10}$$

The approximation of Aki and Richards (1980) assumes that the angles i_1, i_2, j_1 and j_2 (Figure 4.6) are real and close to 90°. This assumption holds for small offsets (less than 1000m) and low layer-2 P-wave velocities (less than $3500\,\mathrm{ms}^{-1}$). At values greater than these the approximation fails and only the Zoeppritz equations correctly predict the reflection amplitude coefficient. Therefore the full solution to the Zoeppritz equations are used to calculate the results.

The new results show that the optimal location for a single receiver is at an offset of 1050m. This is in contrast to the Van den Berg et al. (2003, 2005) optimal

offset of approximately 1500m. Both methods produce a local minimum at an offset of 500m, a distance equal to the interface depth. A standard rule of thumb approximation for an AVO experiment is to use offsets between one and three times the depth to the interface under study. For the single receiver case this rule is confirmed.

Using the Zoeppritz equations I now consider a two-receiver experiment. Using both the iteratively-constructive method and calculating the entropy values across the complete design space I can see how closely the two resulting experimental designs match. For the iterative method the first receiver is fixed at an offset of 1050m. The second receiver is varied between offsets ranging from 0m to 5000m at 25m intervals. The same data space discretisation and number of parameter space samples are used as in the single receiver experiment.

When calculating the single receiver optimal design, model space samples are projected through the forward function (equation (4.3)) into the discretised data space (Figure 4.9 (a)). The optimal design corresponds to the data space with the largest associated entropy value. To locate the optimal second location, a two dimensional design space is constructed for each possible design (Figure 4.9 (b)). In every instance the offset represented by d_1 is fixed at 1050m, the optimal offset found at the previous iteration. When placing the third receiver, the model space samples are used to construct a new three dimensional data space. Although the model space and design space have single dimensions, the data space populated increases by one dimension at each iteration. Although this method increases the computational expense required to calculate the optimal solutions at each iteration, the solution is guaranteed to be as close to optimal as possible.

Figure 4.10 shows the entropy values as a function of offset for the two-receiver design. The optimal design would therefore place the second receiver at an offset of 1350m, resulting in a two-receiver design with offsets of 1050m and 1350m. Similarly to the single receiver experiment there is a local entropy minimum at an offset of approximately 500m and a decreasing entropy trend with increasing offsets greater than 2500m. The entropy notch at 1050m represents a repeat experiment; although this would provide extra information about velocity α_2, the extra information obtained by placing the second receiver at a slightly larger or smaller offset significantly outweighs this advantage.

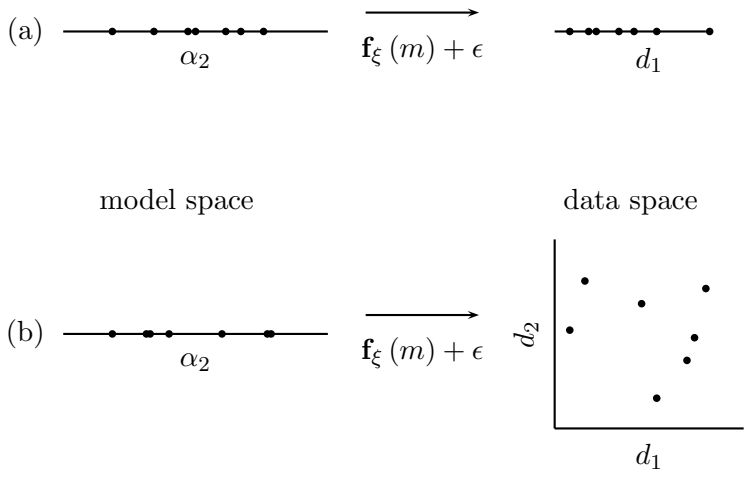

Figure 4.9: (a) For each possible design, samples (dots) are projected from the one dimensional model space, through the forward function, into a one dimensional data space. (b) When locating the second receiver, samples are projected from the one dimensional model space into a two diemensional data space. For each possible design the data space represented by d_1 corresponds to solution found in the previous iteration.

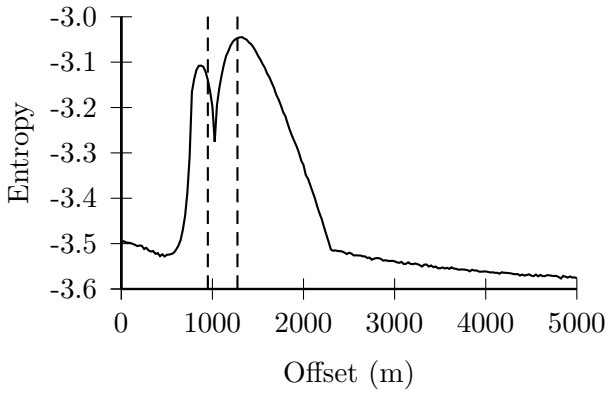

Figure 4.10: Entropy values as a function of offset for a two-receiver experimental design using the iteratively-constructive method with the first receiver fixed at 1050m. The entropy at each offset is calculated using 500,000 samples and a data space discretisation of 0.01. Entropies are calculated at offsets ranging from 0m to 5000m at 25m intervals. The dashed lines represent the optimal offsets found by searching the complete design space (Figure 4.11).

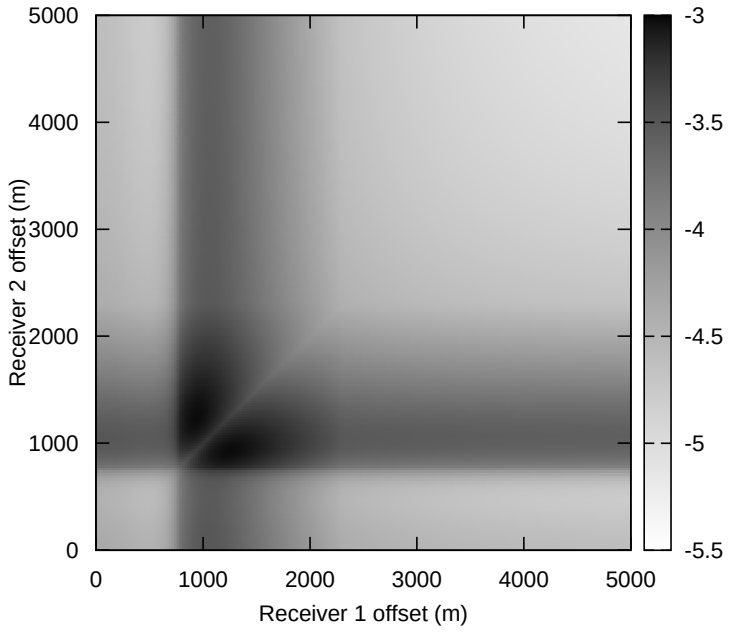

Figure 4.11: Entropy values for every possible two-receiver experimental design. The plot consists of 40,401 entropy values. The globally optimal receiver locations are at offsets of 1275m and 950m.

Figure 4.11 shows the entropy values obtained for every possible two-receiver configuration. The globally optimal receiver locations are at offsets of 1275m and 950m. These do not coincide with the locations found using the iterative method as the offsets lie either side of the optimal single-receiver offset. However, the entropy map shows several global features that are also seen in Figure 4.10. First, placing either or both receivers at offsets less than 500m (equal to the depth of the reflector) or at offsets greater than 2500m results in an experiment that is expected to record relatively little information that can be used to constrain the parameter. Also a repeat experimental design in the offset range 750m-2250m provides less post-experimental information than having a small offset separation. The information gain expected by performing the experiment is given by

$$Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right] - Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right] \tag{4.11}$$

where $Ent\left[\sigma\left(\mathbf{d}|\xi\right)\right]$ is the calculated optimal entropy value, and $Ent\left[\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\right]$ represents the entropy of the measurement noise and is described by a Gaussian. The difference in the information gain found using the iteratively-constructive method compared to that found by performing a complete design space search is 1.7%, while the global search takes 201 times longer to perform.

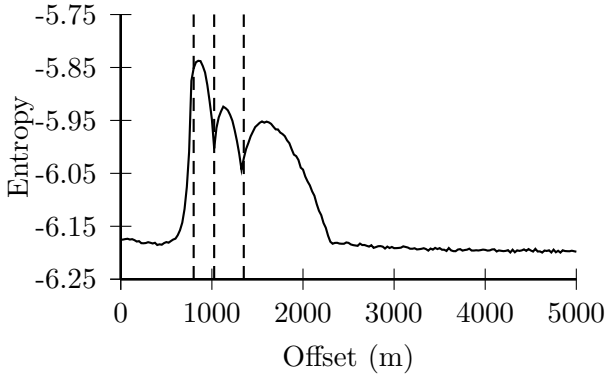

Figure 4.12: Entropy values as a function of offset for a three-receiver experimental design using the iteratively-constructive method with the first two receivers fixed at offsets of 1050m and 1350m. The dashed lines represent the globally-optimal 3-receiver design (Figure 4.13).

Figure 4.12 show the entropy values obtained for the three-receiver optimal experiment calculated using the iteratively-constructive method. In the two-receiver plot (Figure 4.10) there were two entropy peaks of similar magnitude. The three-receiver experiment clearly locates the third receiver at an offset of 850m. The local entropy minima seen at approximately 500m offset in both the single and two-receiver setup is now less prominent. The entropy decline with offsets greater than 2500m is also less steep than in the previous designs, indicating that any offset greater than 2500m will provide the same, minimal amount of post-experimental information. The entropy signal at large offsets also shows small amounts of numerical noise, indicating poor sampling in the data space. In the single-receiver case 500,000 parameter samples populated a maximum of 100 data space discretisations, whereas in the three-receiver case the same number of samples are used to populate a data space with 1,000,000 (100^3) discretisations. The entropy signal-to-noise ratio within the offset range 600m to 2300m is nevertheless still high as indicated by the smooth entropy function. The two entropy notches at offsets of 1050m and 1350m represent offsets where a receiver is already located from previous iterations.

I also calculated the globally-optimal three-receiver experiment by searching the entire experimental design space. Calculating every possible experimental design at 25m offset spacing was too expensive computationally, so Figure 4.13 shows an entropy isosurface for the three-receiver design produced using 100m offset spacings. Since the maximum seems to occur in the offset range 500m to 2000m

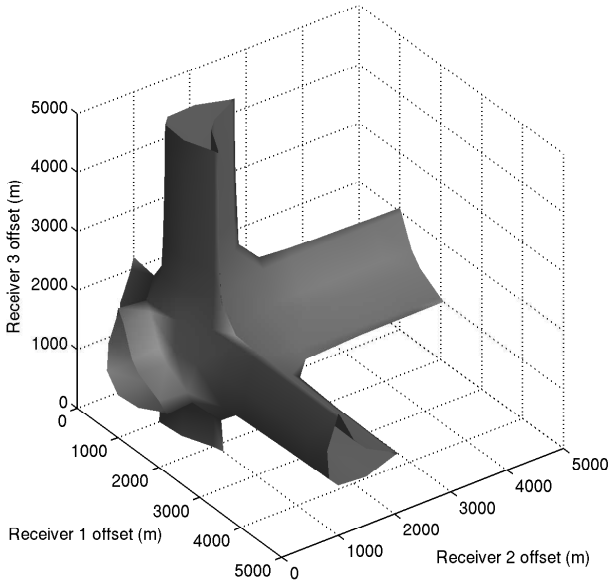

Figure 4.13: Entropy isosurface for a three-receiver experimental design using global sampling. The isosurface represents entropy values equal to 75% of the maximum.

a more detailed search at 25m spacings was performed over that interval. The globally-optimal experimental design uses offsets of 800m, 1025m and 1350m. This is almost an exact replication of the design produced by our iterative method.

A total of 359,632 entropy values were calculated to locate the three globally-optimal receiver offsets. Only 603 entropies were calculated using the iteratively-constructive method, a computational saving of over 99%. The same general features of the two-receiver design (Figure 4.11) are seen in the three-receiver case (Figure 4.13). Offsets greater than 2500m result in experiments with a low associated entropy value, as do experiments with offsets smaller than 500m.

The iteratively-constructive method was then used to place ten receivers in total. For each iterative design the same number of parameter space samples and data space discretisation lengths have been used. Figure 4.14 shows the entropy results for successive experimental designs. Clearly the last few plots in Figure 4.14 are contaminated by numerical noise, a point to which I return below.

The final ten-receiver experimental design is illustrated in Figure 4.15. Note that

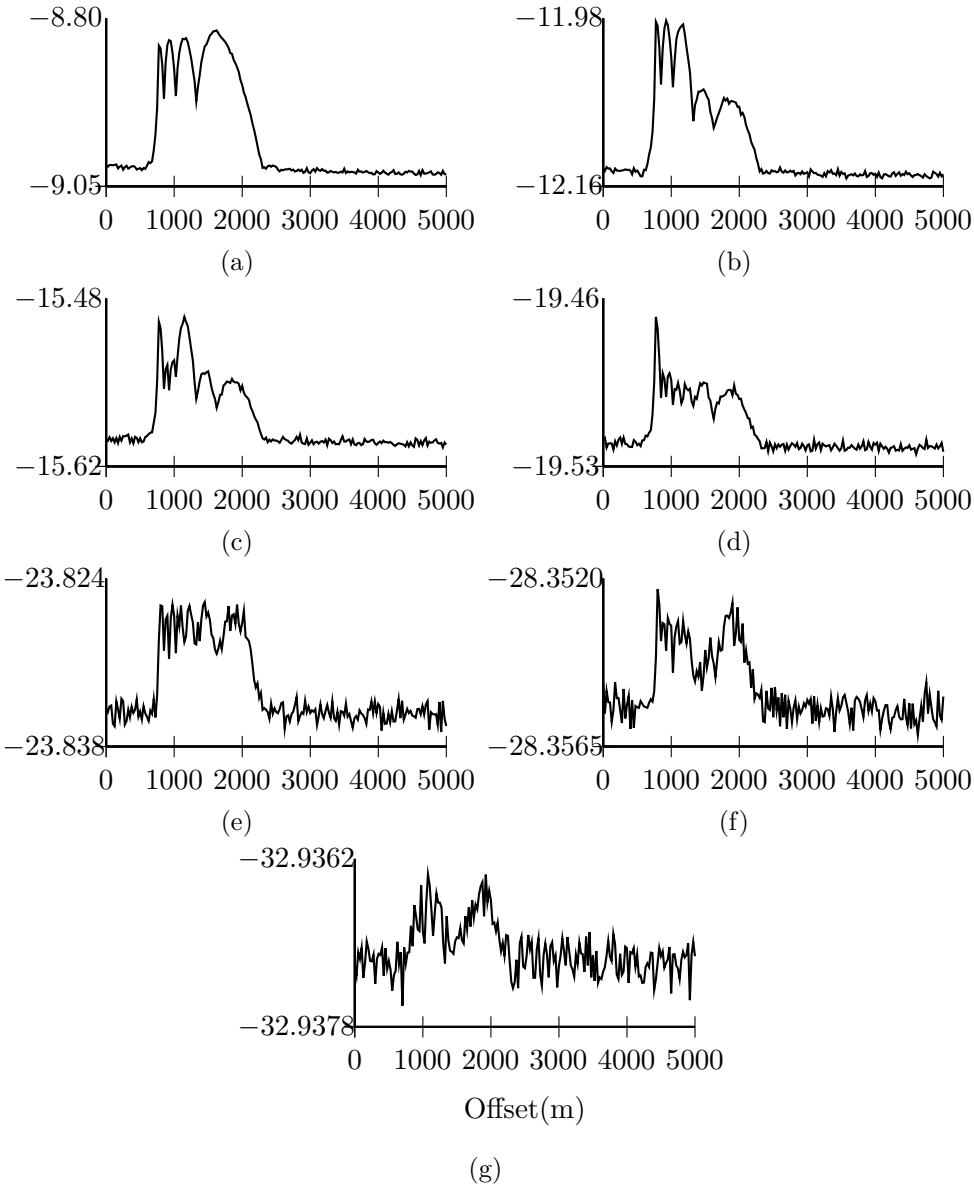

Figure 4.14: Entropy plots for successive receiver locations using the iteratively-constructive method. Plot (a) represents the entropy curve used to place the fourth receiver, and plot (g) the curve used to place the tenth receiver. For each plot 500,000 parameter samples are used and the discretisation length in each data space dimension is 0.01.

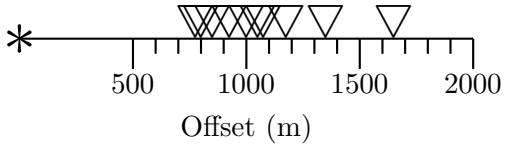

Figure 4.15: Experimental design using ten receivers placed using the iteratively-constructive method and 500,000 parameter space samples. The source is denoted by the $*$ symbol and each receiver by a $\bigtriangledown$.

nine of the ten receivers are placed within the range 500m to 1500m (between one and three times the depth to the reflector), the range often used in AVO rule-of-thumb design methods. However, one of the receivers lies outside of this range, corresponding to the fourth receiver. The fourth receiver offset (Figure 4.14 (a)) is located with high numerical accuracy indicating that the classical heuristic rules fail to match the mathematics-based design criterion in this simple problem.

As the number of receivers increases so does the noise level in the entropy signal. I can locate the seventh receiver (Figure 4.14 (d)) offset with some certainty but for the eighth, ninth, and tenth receivers the poor data space sampling causes the noise level to be too large to locate each receiver offset accurately. However, there is a clear trend seen in the first nine receiver plots that places all the receivers at offsets between 500m and 2000m. Also, in the numerically-accurately placed receivers (one to seven) a repeat experiment is never the optimal solution.

4.5 Design of an Efficient Numerical Sampling Scheme

Accurately locating more than seven receivers using the new method requires a more densely sampled data space than used above to reduce the noise level in the entropy signal. The number of samples required to locate each successive receiver is therefore a trade-off between the desired signal-to-noise ratio of the entropy response and the amount of computation used. I have approached this trade-off as requiring a separate design procedure that is performed as the number of samples is increased.

The minimum required number of parameter space samples in the iteratively-constructive method is determined by assessing the standard deviation of the entropy estimates. It should be noted that the use of 'standard deviation' is only used as one possible variation measure in order to assess whether sufficient samples have been drawn, and not in the classical sense of characterising a pdf. As more samples are drawn from the parameter space the entropy asymptotes to a constant value (Figure 4.3). I evaluate this convergence by considering the standard deviation of the entropy signature at offsets of 1000m, 2000m, and 3000m which approximately spans the range of offsets likely to be optimal given the above tests. At each of these offsets the entropy value is calculated after each subsequent 200 samples have been drawn from the parameter space. The standard deviations are calculated over each consecutive set of 100 data space entropy values. The standard deviation measures the variation in 100 entropy values over the addition of 20,000 parameter space samples. Hence, the standard deviation should be observed to decrease with every successive set of 20,000 samples as the entropy estimate converges.

Figure 4.16 (a) shows how the entropy estimate changes as samples are drawn from the parameter space for the AVO experiment that locates the second optimal receiver using the iteratively-constructive method. The standard deviation is calculated for each offset profile. The maximum standard deviation from the three profiles is retained as the value used in the algorithm below.

Figure 4.16 (b) shows the resulting entropy profile (grey line) estimated using 20,000 samples. I determine if sufficient samples have been drawn from the parameter space by assessing the entropy signature between the global maximum (indicated by the solid black line) and two standard deviations from this value (dot-dash line). If no other local maxima are found in this range I conclude that with 95% confidence I have found the true maximum value and no more samples are required. If, as in this example, there are local maxima in the two standard deviation range another 20,000 samples are added at each offset. The standard deviation is re-calculated as above and the entropy profile re-evaluated. This process is repeated until a single peak is found within the two standard deviation lower entropy limit. To reduce the computational time I discard offsets that have associated entropy estimates that fall below three standard deviations (dashed line) from the maximum entropy value when adding extra samples these offsets are not considered.

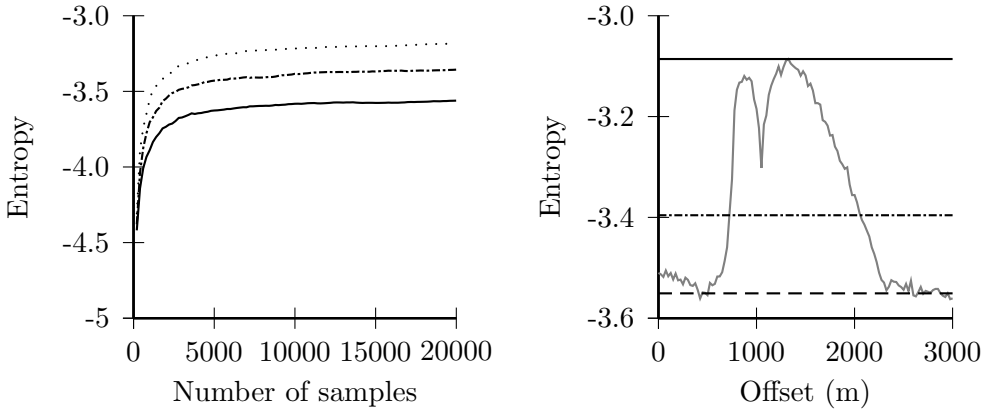

Figure 4.16: Plot (a) illustrates how the entropy estimates for offsets of 1000m (dotted), 2000m (dot-dash) and 3000m (solid) evolve as increasing number of samples are drawn from the parameter space. Plot (b) shows the resulting entropy profile (grey line) produced from 20,000 samples. The solid line indicates the maximum entropy value, the dot-dash line two standard deviations, and the dashed line three standard deviations from the maximum.

This process creates a dramatic reduction in the computational time required to place multiple receivers. I re-calculated the optimal ten-receiver design using the iteratively-constructive method and the variable sample size using the above algorithm. The final design is shown in Figure 4.17 (top). Comparing Figure 4.17 to Figure 4.15, the new results locate two receivers beyond 1500m instead of only one. Harder to see is that the experiments that required fewer than 500,000 samples all located the optimal receiver at the same offset as the previous experiment, showing that we had needlessly over sampled the data space in the previous results. Conversely, in the experiments that required more than 500,000 samples, the receivers were positioned in different locations indicating that the data space was not previously adequately sampled to reveal the true global maximum due to numerical noise.

Figure 4.18 shows the actual number of samples required to accurately locate the receivers. In the previous method I had over sampled the first five receiver locations but under sampled the remaining five. In fact, using the variable sample size method also shows that only 40% of the samples were required to place the first five receivers when compared to the iteratively-constructive method that

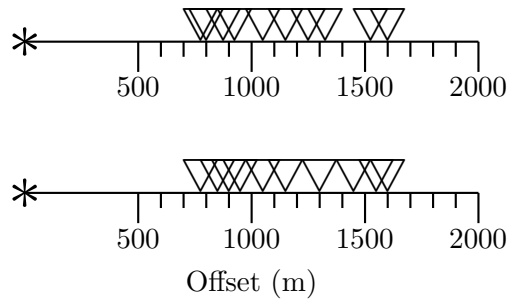

Figure 4.17: Experimental design using ten receivers placed using the iteratively-constructive method with the variable sample size test. Top design represents optimal ten-receiver design to 95% confidence, lower design to 66% confidence.

used a static 500,000 samples for each receiver.

The requirement of 95% confidence in finding global entropy maxima for successive receivers is arbitrary. Figure 4.17 (bottom) also shows the design obtained if we are satisfied with 66% confidence (one standard deviation), while Figure 4.18 also shows the number of samples in this case. Nine of the ten receivers placed with 66% confidence match within ±25m of the receivers offsets found to 95% confidence. Placing the ninth and tenth receivers with 95% confidence requires approximately 2.5×10^6 more samples than with 66% confidence.

4.6 Discussion of AVO Results

The final designs for the AVO application appear to contradict accepted heuristic design methods by placing one or more receivers at offsets greater than three times the depth to the reflecting layer. However, rather than this being a general, design-independent phenomenon, the final design can be shown to depend strongly on the detailed subsurface Earth structure above the reflecting boundary, on the prior information about parameters, and on the data uncertainty. While the two-layer structure in Figure 4.6 might be representative of a shallow geophysical experiment, a more representative subsurface structure for experiments incorporating deeper layers would typically contain several layers above the target interface, each with different properties, and velocity would usually increase with depth. At every interface the raypath of the seismic energy would be re-

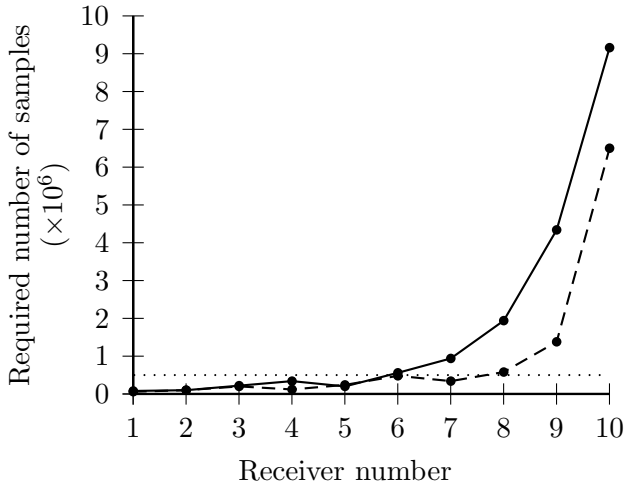

Figure 4.18: The number of samples required to accurately locate the optimal receivers using the iteratively-constructive method to 95% (solid line) and 66% (dashed line) confidence. The dotted line shows the number of samples used in the previous example (Figure 4.15): I had over sampled the first five experiments but under sampled the final five.

fracted towards the horizontal at increasing depths in the model. The refraction process results in a reduced offset at the surface for the same angle of incidence and hence reflection coefficient at the subsurface interface (Figure 4.19). That is, the angular range at the bottom reflecting boundary created using the optimal designs in Figure 4.17 could also be obtained with receivers placed at smaller offsets.

Reducing the size of the prior parameter range both decreases the computational power required to place each successive receiver and changes the final design. Figure 4.20 represents the final ten receiver design after the parameter range has been reduced from [3000ms^{-1}, 4500ms^{-1}] to [3375ms^{-1}, 4125ms^{-1}]. The design shows that all receivers now conform to the heuristic design by falling within the offsets 500m to 1500m, and receiver density decreases with increasing offset.

The final simplification used was to have a uniform data uncertainty with increasing offset. In many realistic situations uncertainty might increase with offset. An increased data uncertainty results in a larger volume of the data space being sampled in the entropy calculation. As seen in Figure 4.4 this results in reduced $\sigma(\mathbf{d})$ values and larger entropy; optimal receiver locations would then tend to-

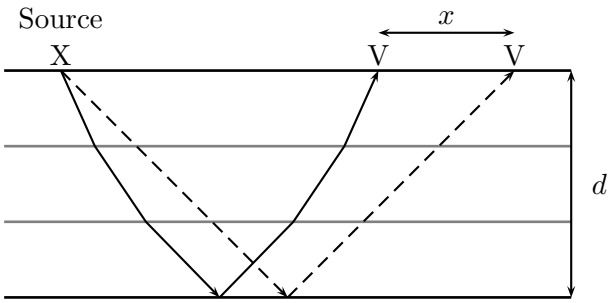

Figure 4.19: Raypath differences caused by having a different subsurface model. The dashed line represents a model with no refraction before the targeted interface, whereas the solid line refracts at two interfaces (grey lines) both prior and post reflection. Both raypaths have the same angle of incidence and reflection at the lower-most interface.

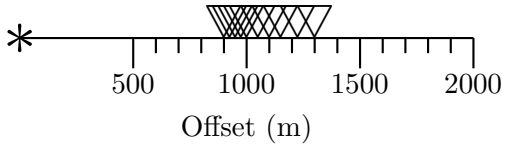

Figure 4.20: Experimental design using ten receivers placed using the iteratively-constructive method with the variable sample size test and a reduced prior parameter range. The source is denoted by the $*$ symbol and each receiver by a $\bigtriangledown$.

wards longer offsets than in Figure 4.20.

From the three factors given above it is clear that the heuristic design statement is over-simplified for AVO design problems. Each situation must be designed independently to optimise information, incorporating as much prior information about subsurface structure and expected data uncertainty as possible.

4.7 Conclusions

An iteratively-constructive design method has been presented which is applicable to nonlinear, multidimensional design problems. The method is based on a Bayesian framework which is suited to nonlinear scenarios where classical design methods, based on forward function gradients, fail.

For both the synthetic sawtooth and practical AVO examples, locally optimum experimental designs produced by the iteratively-constructive method closely resemble globally-optimal designs, in low dimensional cases where the later designs can be calculated. However, I have also produced a multi-receiver AVO experimental design using the iterative method that would have been computationally intractable using grid-search nonlinear design methods. The parameter space sample size has been treated as both a fixed parameter, and a variable which requires optimisation at each iteration. This extra optimisation process increases the efficiency of the new method significantly. A simple search of the possible offsets at uniform intervals through the design space has been used to locate the optimal receiver offset. To reduce computation, the only limiting factor to the number of possible data space dimensions, it is highly recommended that more efficient methods are used to search the already-limited design space and to select the optimal discretisation length. Given more efficient algorithms, designing experiments for more complex applications which use both the P- and S-wave reflection coefficient for a multi-layered subsurface now promises to be computationally feasible.

Chapter 5

Optimal AVO/AVA Processing Designs

5.1 Introduction

In Chapter 4 I introduced a novel iteratively-constructive algorithm for designing nonlinear, multidimensional design problems. Whereas the AVO example given was described in terms of survey design and data collection, optimal experimental design methods can also be used in the data processing phase where only the recorded data that provide significant information about the parameters of interest are selected and actually processed. The term "optimal design" used subsequently can therefore refer either to designing the survey or designing the data subset for processing.

In this chapter I focus on applying the methods developed in Chapter 4 to designing optimal trace selection profiles for processing amplitude versus angle (AVA) surveys for sand-clay reservoir models where some, perhaps limited prior petrophysical information about the subsurface is known. I examine how the optimal trace selection designs change as our prior knowledge of the rock parameters and reservoir fluid content changes, and assess which of the prior parameters has the largest effect on the trace selection designs and are therefore deemed the most important to bound as tightly as possible prior to the survey. The resulting trace selections can be used either to select a subset of traces to process or, if all traces

are to be processed, then to decide which traces deserve most of the processor's attention to reduce noise and which should be given less attention. An overview of the method developed in Chapter 4 in the context of optimal trace selection is given below.

5.2 Experimental Design Method

The method developed in Chapter 4 can be divided into three parts. First a Bayesian quality metric is defined that quantifies the information provided by any given trace-subset (design), second an iterative method sequentially increases the complexity of the selected trace-subset so as to maximally increase the quality, and third there is a numerical sampling scheme that minimises the computational time required to calculate the best design at each iteration. Each is presented briefly below from the point of view of selecting optimal data subsets for processing.

5.2.1 Bayesian Design Method

The optimal trace-subset corresponds to the trace design that maximises the information *expected* to be contained in the posterior parameter pdf

$$\sigma\left(\mathbf{m}|\mathbf{d},\xi\right)=\frac{\theta\left(\mathbf{d}|\mathbf{m},\xi\right)\rho\left(\mathbf{m}\right)}{\sigma\left(\mathbf{d}|\xi\right)}. \tag{5.1}$$

In the AVA case considered here, ξ will then define the set of selected traces that provide the most information about the particular subsurface reflector of interest. In selecting which data to process we wish to maximise the information content on the parameters that is expected to be contained within the posterior pdf $\sigma\left(\mathbf{m}|\mathbf{d},\xi\right)$.

For nonlinear problems Shewry and Wynn (1987) showed (see Section 2.3.1) that a design measure where the data error is design independent can be defined as

$$\Phi\left(\xi\right)=Ent\left\{\sigma\left(\mathbf{d}|\xi\right)\right\}. \tag{5.2}$$

In reality the measurement noise may not be completely design independent but the assumption is often made nevertheless, as in many practical cases it is assumed that the parameter-data relationship is of the form

$$\mathbf{d} = f_{\xi}(\mathbf{m}) + \epsilon \tag{5.3}$$

where ϵ is a vector of independent random errors which do not depend on either the parameters or on the trace selection design.

Altering the trace selection ξ changes the properties of the forward function (equation (5.3)), the properties of the pdf $\sigma(\mathbf{d}|\xi)$, and therefore its entropy value. The selected trace combination that corresponds to the maximum entropy value also corresponds to maximum expected information about $\mathbf{m}$ and hence to the optimal experimental design.

5.2.2 Sequential Design

When there are multiple design parameters (in our case traces or angles of incidence), searching the entire design space (the space of all permissible combinations of design parameters) is computationally infeasible. Chapter 4 introduced a novel iteratively-constructive design strategy which sequentially increases the number of elements within the design vector ξ by one element at each iteration,

$$\xi_j = \arg\max\left[Ent\left\{\sigma(\mathbf{d}|\xi_j)\right\}\right], \text{ such that } \xi_{j-1} \text{ is fixed} \tag{5.4}$$

where $\xi_j = \{\xi^1, \ldots, \xi^j\}$ and ξ^i is an element of the trace selection design vector. The new optimal trace selection design ξ_j combines the data defined at the previous iteration ξ_{j-1}, augmented by the single datum defined by ξ^j that maximises the entropy of the posterior data pdf given that ξ_{j-1} remains fixed. In this way the total work required to design an experiment with ten selected traces is reduced from searching a ten-dimensional design space to ten separate searches of a one-dimensional design space. Although the trace selection design at each iteration is then only a locally optimal design, Chapter 4 showed that for AVO/AVA problems the locally optimal design is almost identical to the globally optimal design in cases where the latter was computationally viable to calculate.

5.2.3 Numerical Sampling Scheme

In general, the pdf required to calculate the entropy is not known analytically and must therefore be calculated numerically. Samples of $\rho(\mathbf{m})$, the prior pdf describing what is known about the subsurface before the experiment, are generated and projected through the physical relationship (equation (5.3)), which also incorporates measurement error (ϵ), to produce samples of $\sigma(\mathbf{d}|\xi)$ in the (discretised) data space. The resulting histogram is normalised to have unit volume, whereafter it represents a numerical approximation to the pdf $\sigma(\mathbf{d}|\xi)$. This numerical approximation of $\sigma(\mathbf{d}|\xi)$ can then be used to assess the quality of the experimental design by calculating the entropy,

$$Ent\{\sigma(\mathbf{d}|\xi)\} \approx \sum_i \hat{\sigma}(\mathbf{d}_i) \log\{\hat{\sigma}(\mathbf{d}_i)\} \tag{5.5}$$

where $\mathbf{d}_i$ represents the centre of the i^{th} discretised bin in the data space. For any given experimental design, the estimate of $\hat{\sigma}(\mathbf{d}|\xi)$ is affected by the number of parameter samples, the data space discretisation bin size, and the measurement error.

With each iteration of equation (5.4) the data space increases by one dimension. For example, in a seismic AVA survey, the reflection coefficient recorded for a given interface at each receiver represents a single dimension. As successive receivers are selected to be processed, the number of data space dimensions increases and hence so does the number of model samples required to accurately approximate the posterior pdf $\sigma(\mathbf{d}|\xi)$. As more samples are added to approximate equation (5.5),the calculated entropy value becomes more accurate and asymptotes to a final value. In Chapter 4 I introduced a scheme that measures this change in entropy value for several different trace selection designs which span the entire design space. For each of the selected designs, the standard deviation of the entropy value is calculated over a specific range of additional samples with the maximum standard deviation value becoming our so-called test value. The optimal trace selection design result is reached when only one entropy maximum in the design space is found within the error range of the test value. Using the same notion, trace selection designs that have an entropy value lower than a predefined threshold, also derived from the test value, are removed from the current iteration to reduce the size of the design space that must be searched.

Figure 5.1 shows a flowchart of how the design methods are combined to select an additional design parameter. First the prior model pdf, the range of possible designs to be considered, and the variable 'ns' representing the maximum number of model evaluations that is computationally tractable at each iteration, are selected. For each possible design 'ns' random samples of the prior model pdf are projected into the discretised data space. The density of the resulting data space samples for each design are normalised so that they represent numerical approximations of the pdf $\sigma(\mathbf{d}|\xi)$. The entropy is then calculated for each possible design using equation (5.5). A statistical test is performed to calculate whether the result is likely to be optimal, in the sense that it represents a statistically significant maximum above the level of numerical random-sampling noise. If so, the design corresponding to that maximum is deemed optimal and the process is stopped. If it is not deemed optimal then designs that fall below a pre-defined entropy threshold are removed and an additional set of 'ns' random samples are projected from the prior model pdfs into the discretised data spaces. The process is repeated until a statistically significant optimal design has been found.

5.3 Reservoir Application: Selecting Optimal Traces for AVA Processing

5.3.1 Reservoir Model

To calculate the posterior pdf $\sigma(\mathbf{d}|\xi)$ where $\mathbf{d}$ represents the P-wave reflection coefficient requires prior distributions over α, β and ρ for both geologic layers. As in Chapter 4 I assume the overburden is a shale with known $\alpha = 3048\text{m/s}, \beta = 1244\text{m/s}$ and $\rho = 2400\text{kg/m}^3$, the same overburden model used by Ostrander (1984) to analyse plane-wave reflection coefficients as a function of incidence angle for a reservoir model. The corresponding values for the reservoir (the lower geological layer) are unknown. In other situations, the values for the overburden might also be varied without otherwise changing the methods used here. In this study, however, we wish to highlight how the optimal trace selection designs depend on expected reservoir properties, so we hold the overburden fixed.

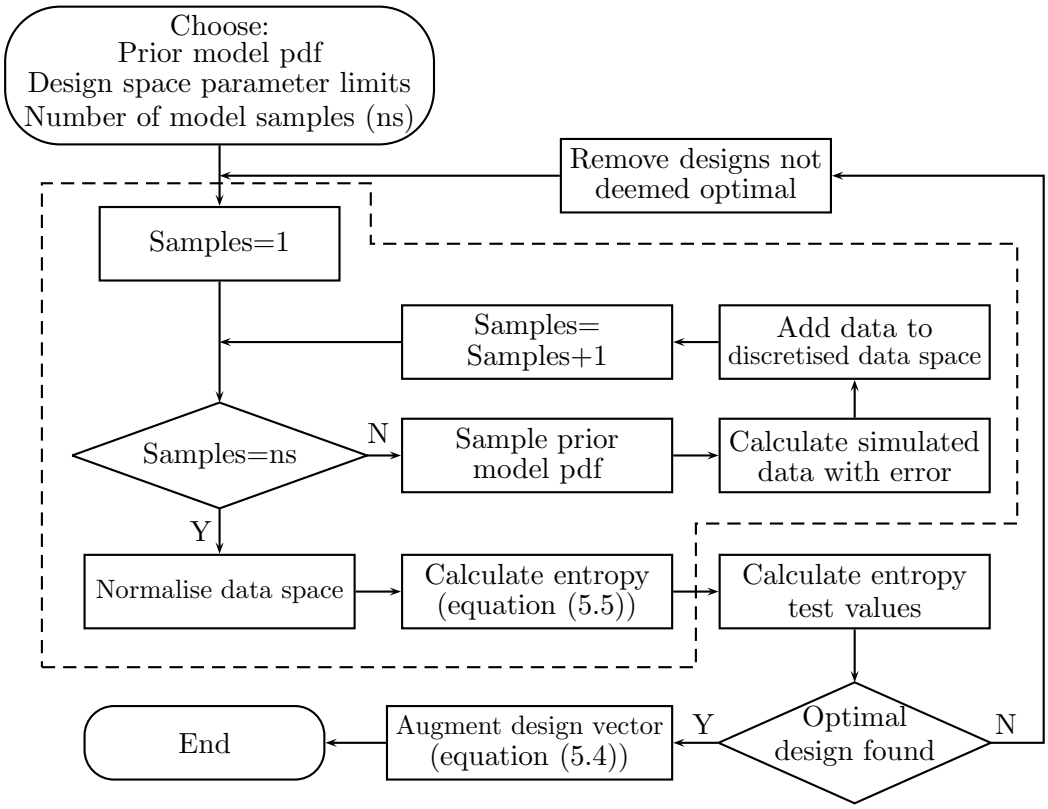

Figure 5.1: Flowchart showing how an additional design parameter is chosen. The workflow inside the dotted line is performed simultaneously for all possible designs. With each iteration that the entropy test value is calculated the design space is reduced by removing designs that are deemed never to be optimal. The required value of 'ns', the number of times the prior model pdf should be sampled before the data space is normalised is dependent on the forward function and the prior model pdf The initial design space prior model parameter ranges should be identical for each run through the sequential design process.

Parameter	Range
Sand bulk modulus (GPa)	36-43
Sand shear modulus (GPa)	33-46
Sand density (kg/m^3)	2640-2650
Clay bulk modulus (GPa)	20-34
Clay shear modulus (GPa)	7-19
Clay density (kg/m^3)	2350-2680
reservoir porosity (%)	10-40
Clay content (%)	20-50

Table 5.1: Rock parameters required for the Goldberg and Gurevich (1998) model. The ranges represent the extreme values of the uniform prior pdfs used to create velocity and density models. Extreme values are taken from Marion et al. (1992); Mavko et al. (1998); Carcione et al. (2003); Chen and Dickens (2009). The values shown represent grain properties rather than dry matrix properties.

I build a reservoir model by assuming that the semi-empirical petrophysical model of Goldberg and Gurevich (1998) is valid for sand-shale reservoirs with different saturating fluids. For any particular reservoir rock physics properties (e.g. sand and shale proportions, porosities, fluid content etc, see Tables 5.1 and 5.2) the Goldberg and Gurevich (1998) model calculates corresponding P-wave and S-wave velocities and density values. These can then be used in conjunction with the Zoeppritz equations (equations (3.1)) to calculate a P-wave reflection coefficient. By assuming prior uncertainty ranges over the properties in Tables 5.1 and 5.2, prior parameter pdfs of the P- and S-wave velocity and density can be constructed and used to determine the optimal trace selection design, given the various prior uncertainty ranges. The prior uncertainty ranges for the matrix bulk properties are shown in Table 5.1. I consider three different fluid properties (Table 5.2) independently to observe how the optimal trace selection design changes for different saturating fluids. I assume uniform pdfs over all parameter ranges. The method also allows the prior information ($\rho(\mathbf{m})$) to be represented by any type of distribution such as Gaussian or Cauchy.

Figure 5.2 shows crossplot histograms of the models produced from 1,000,000 random samples from these uniform distributions, using the properties of brine ((a)-(c)), oil ((d)-(f)) and gas ((g)-(i)) as the saturating fluid. Figure 5.2 illustrates that from the uniform ranges in Tables 5.1 and 5.2, a large variety of velocity and density models can be created. Figure 5.3 shows the P-wave reflection coefficient histogram for the brine saturated reservoir model generated from

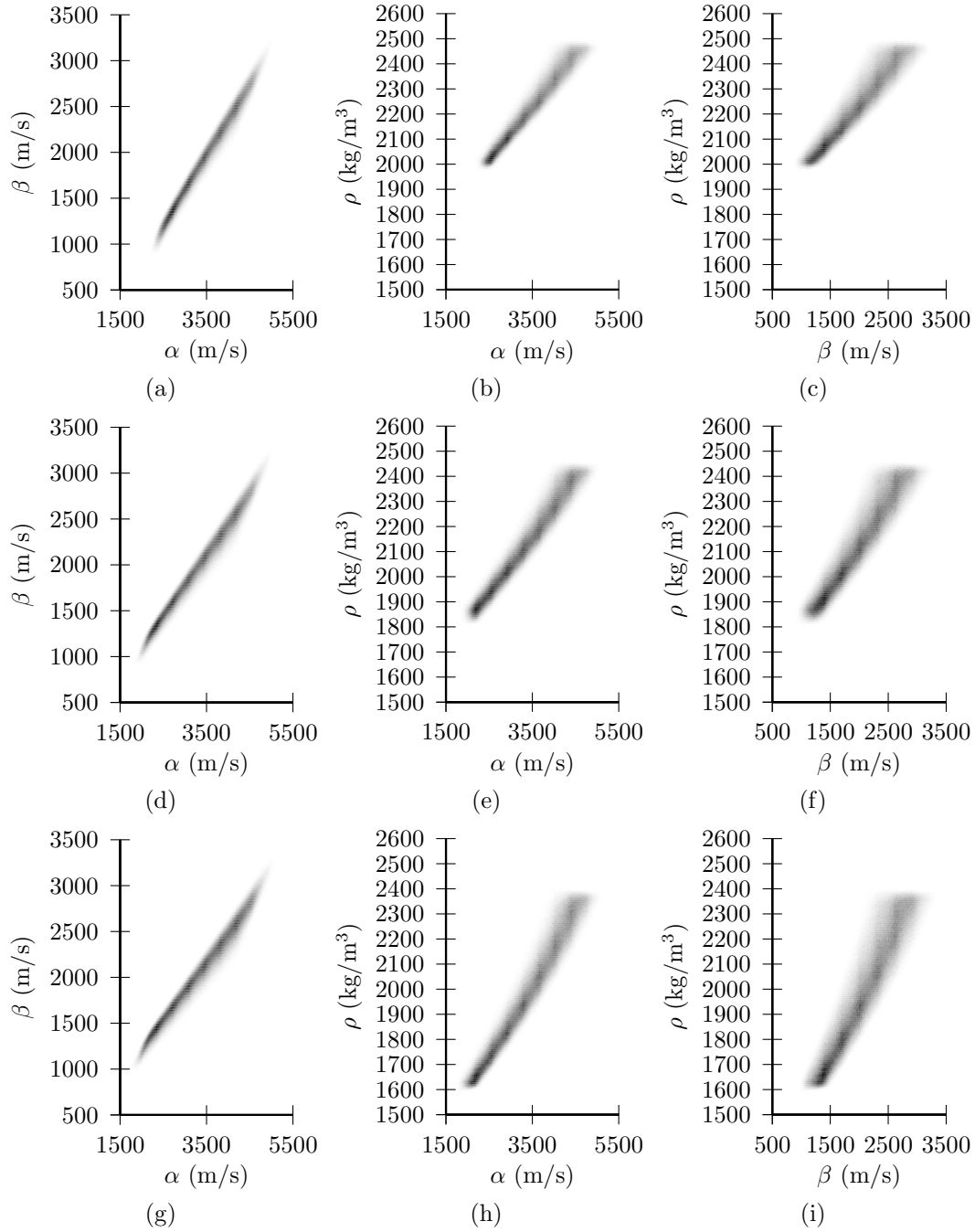

Figure 5.2: Velocity and density histograms produced for a brine filled reservoir ((a)-(c)), an oil filled reservoir ((d)-(f)) and a gas filled reservoir ((g)-(i)) using the Uniform parameter ranges in Tables 5.1 and 5.2. The shading represents the histogram frequency.

Parameter	Range
Brine bulk modulus (GPa)	2.4-3.2
Oil bulk modulus (GPa)	0.5-0.75
Gas bulk modulus (GPa)	0.01
Brine density (kg/m^3)	1040-1090
Oil density (kg/m^3)	616-738
Gas density (kg/m^3)	100

Table 5.2: Fluid parameters required for the Goldberg and Gurevich (1998) model. The ranges represent the extreme values of the uniform prior pdfs used to create velocity and density models. Values are taken from Clark (1992); Carcione et al. (2003); Chen and Dickens (2009).

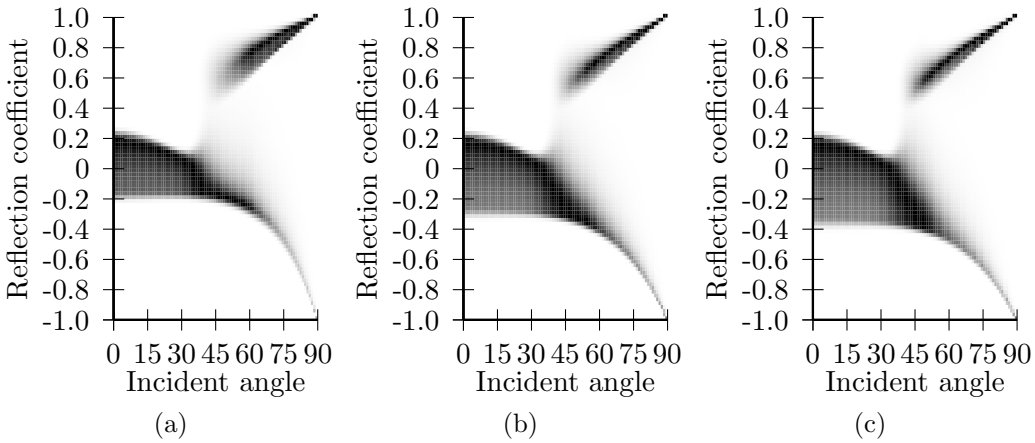

Figure 5.3: P-wave reflection coefficient histogram corresponding to the brine (a), oil (b) and gas (c) saturated reservoir histograms in Figure 5.2. The overburden is assumed to be a shale with $\alpha = 3048$m/s, $\beta = 1244$m/s and $\rho = 2400$kg/m^3. Darker shading corresponds to more likely reflection coefficients.

1,000,000 samples at incident angles ranging from 0° to 90° calculated using the Zoeppritz equations. For AVA modeling the Zoeppritz equations are relatively cheap to compute. If a more complex forward function is used it may be computationally infeasible to use as many as 1,000,000 samples. Clearly the posterior pdf $\sigma(\mathbf{d}|\xi)$, and therefore the corresponding entropy value, will be different for each incident angle at which the amplitude is measured. The incident angle that corresponds to the maximum entropy value represents the optimal angle at which to select data for processing in order to maximally reduce the expected uncertainties on parameters α_2, β_2 and ρ_2. In turn, these can be used to constrain the bulk rock and fluid properties of the reservoir.

Assuming different saturating fluid parameters (Table 5.2) results in different velocity and density profiles and therefore different reflection coefficient signatures. As one would predict, changing the saturating fluid from brine to oil to gas (Figure 5.2) results in a progressively lower reservoir density. For all three fluids however, the overall P-wave and S-wave velocity ranges remain very similar. Figure 5.3 shows the corresponding P-wave reflection coefficient histograms for the brine, oil, and gas reservoirs.

In the above reservoir models a uniform porosity ranging from 10% to 40% has been used. To assess the importance of accurately estimating the porosity of the target reservoir when selecting which traces provide the maximum information, I separately consider a low-porosity oil reservoir with a uniform porosity prior pdf distribution ranging from 10%-20%, a medium-porosity oil reservoir with a uniform prior pdf ranging from 20%-30%, and a high-porosity oil reservoir with a uniform prior pdf ranging from 30%-40%. Figure 5.4 shows the velocity and density histograms for the oil saturated reservoir models with varying porosity ranges whilst keeping the other parameters within the same ranges as those given in Tables 5.1 and 5.2. The porosity range has a large impact on the velocity and density profiles. As the porosity increases both the velocity and the density of the reservoir decrease. The velocity and density ranges over which the histograms are spread are much smaller than those in Figure 5.2, clearly demonstrating that the combination of all of the narrow-porosity range results creates the overall distributions seen in Figure 5.2 (d)-(f). As a result the corresponding reflection coefficient histograms shown in Figure 5.5 show reduced ranges of P-wave reflection coefficients, in particular showing that for a high-porosity reservoir the critical angle is never reached in any survey. When combined these results represent the overall porosity histogram shown in Figure 5.3 (b).

5.4 Results

5.4.1 Saturating Fluid

Figure 5.6 shows the entropy of $\sigma(\mathbf{d}|\xi)$ and hence the expected information about model $\mathbf{m}$ as a function of incident angle when selecting the first angle (trace) to process. This is plotted for each reservoir model (Tables 5.1 and 5.2) correspond-

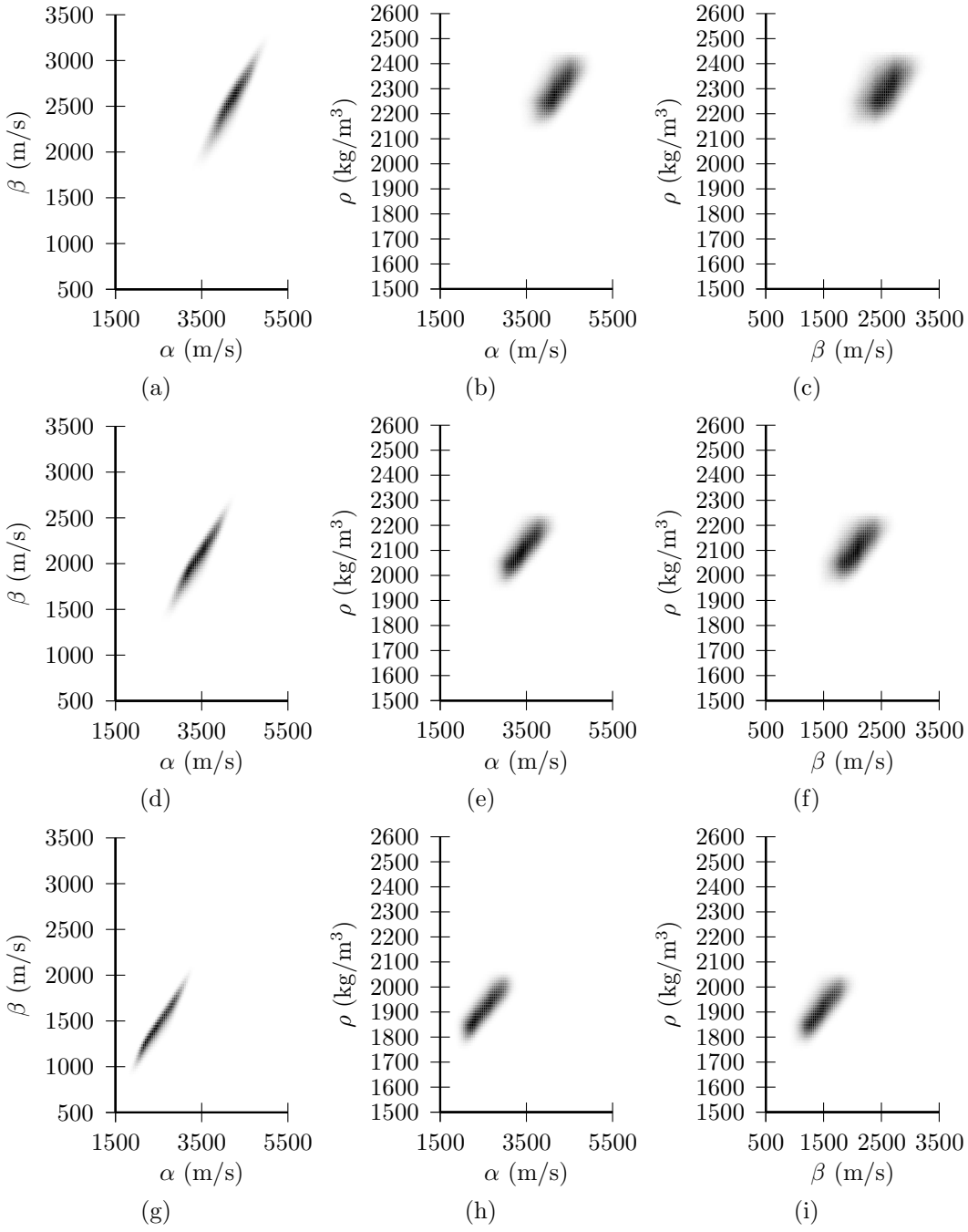

Figure 5.4: Velocity and density histograms for an oil-filled reservoir with varying Uniform prior porosity ranges. Plots (a)-(c) represent a reservoir with a porosity range of 10-20%, (d)-(f) 20-30% and (g)-(i) 30-40%. The shading represents the histogram frequency.

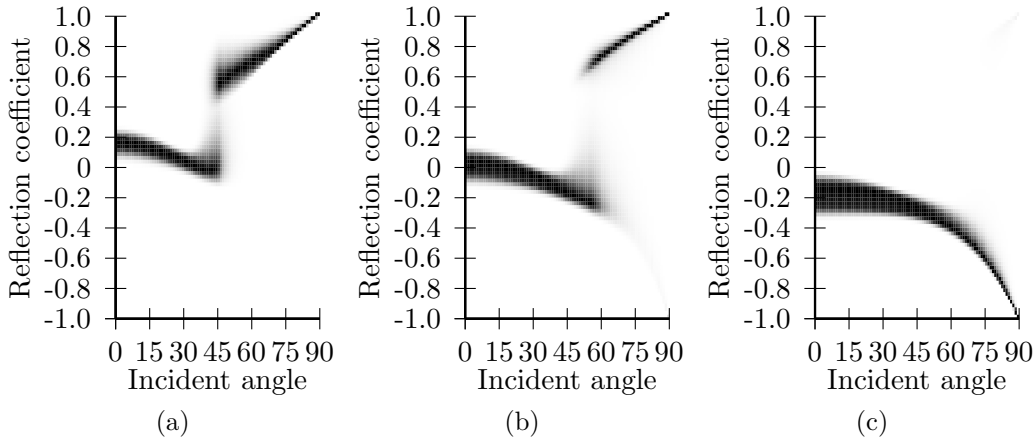

Figure 5.5: P-wave reflection coefficient histograms for oil-filled reservoir velocity-density models with the varying Uniform prior porosity ranges in Figure 5.4. Plot (a) represents the low-porosity reservoir (10%-20%), (b) the medium-porosity reservoir (20%-30%) and (c) the high-porosity reservoir (30%-40%).

ing to each of three different saturating fluids. For incident angles from 0° to 90° at 0.5° separation, the information value was calculated using 500,000 model space samples. As expected, since all the reflection coefficient curves were similar (Figure 5.3), the information curves all show the same general features. All three information curves have a maximum value at approximately 50° equating to the optimal location of the first angle to process. For angles over 70° the information value decreases quickly showing that receivers at large offset angles provide relatively little information relating to the subsurface in this case. Very small incident angles (less than 20°) correspond to large information values, whilst for all three fluids there is a local minimum at approximately 30°. Figure 5.7 shows the extension from a single-angle to a design that selects ten angles as calculated using the iteratively-constructive method developed in Chapter 4.

As expected all three plots show optimal trace designs with fairly similar features. In each case, the majority of angles selected for processing are placed between 0°-10° and 40°-70° incident angle. This trend fits with the information profiles in Figure 5.3 where for the single angle an information minimum was located between 20°-40°, and at angles greater than 70° the information declines to the global minimum seen at 90°. It is also evident that in the incident angle ranges were angles are selected for processing, the angles are not evenly spaced. In the 40°-70° range in all cases the density decreases with incident angle.

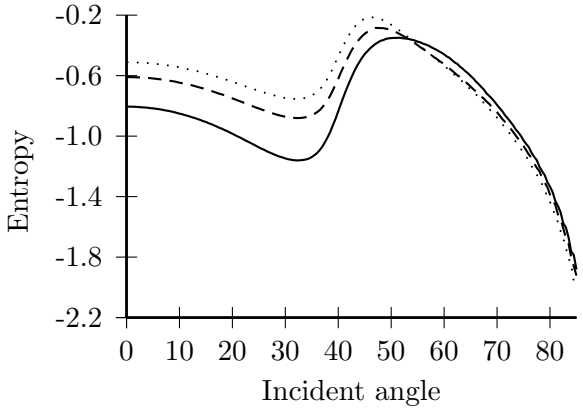

Figure 5.6: Entropy values as a function of incident angle for the general reservoir properties (Tables 5.1 and 5.2) estimated using 500,000 model parameter samples. The brine saturated reservoir is shown by the solid line, oil by the dashed line and gas by the dotted line. In each case a porosity range of 10%-40% has been used.

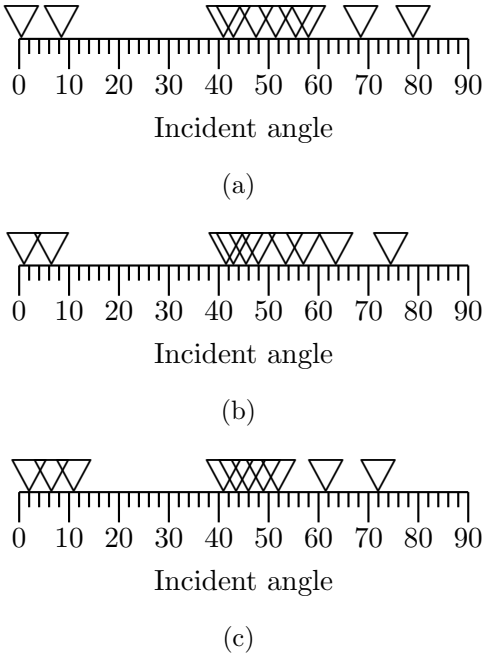

Figure 5.7: Processing designs using ten receivers placed using the method developed in Chapter 4. Plot (a) represents the final design for a brine filled reservoir, (b) an oil filled reservoir and (c) a gas filled reservoir.

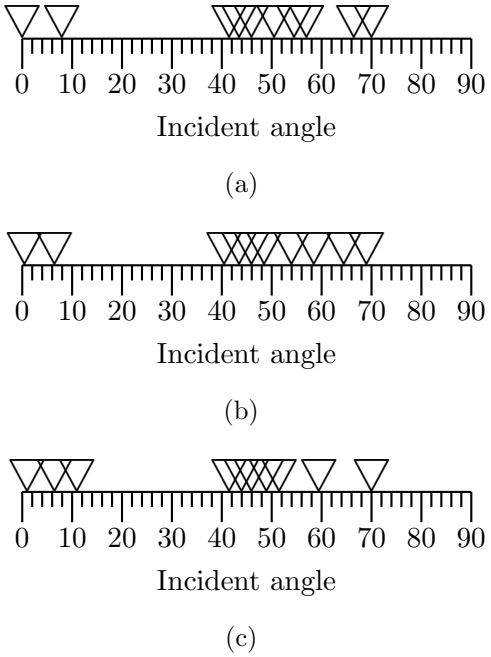

Figure 5.8: Processing design using ten receivers placed using the method developed in Chapter 4 allowing only incident angles up to 70°. (a) represents the final design for a brine filled reservoir, (b) an oil filled reservoir and (c) a gas filled reservoir.

In AVA inversion and processing, techniques exist and are implemented that use recorded data at post-critical angles (e.g. Macdonald et al., 1987; Williams et al., 2001; Liu and Schmitt, 2003; Downton and Ursenbach, 2006; Skopintseva and Nefedkina, 2006). In all the models shown, the critical angle occurs at less than 70°. Beyond this angle the actual source-receiver offset generally becomes prohibitively large, and as seen in our results little information about the subsurface would be recorded. Since only a single angle in each scenario in Figure 5.7 is at an angle greater than 70°, it is expected that reducing the allowed incident angle range will not have a profound impact on the overall design. Figure 5.8 shows the ten-angle design when only incident angles between 0° and 70° are considered.

The optimal trace designs show the same characteristics as before. The angle selected to be processed outside the 0°-70° range has been relocated to fit in with the general trend of decreasing density within the range 40°-70°.

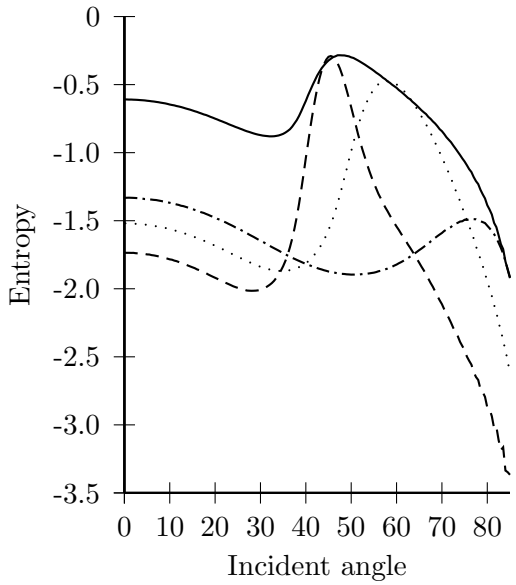

Figure 5.9: Entropy as a function of incident angle for selecting the first trace to process in the different porosity oil reservoirs. The solid line represents the oil reservoir with the complete prior porosity range (10%-40%), the dashed line 10%-20% porosity, the dotted line 20%-30% porosity, and the dot-dash line 30%-40% porosity.

5.4.2 Porosity

Figure 5.9 shows the information values used to select the first angle to process for the oil reservoir models with different porosities. The maximum and hence the optimal angle selected is very different for each of the three porosity models. As the porosity increases the optimal incident angle also increases until a certain porosity threshold is reached. At porosity values close to this threshold value the information provided at either large angles ($> 60°$) or near vertical incidence ($0°$) are comparable. Beyond this value near offset angles provide more information than large angles, in stark contrast to the low porosity result. From Figure 5.4 it can be seen that the pre-threshold increase in the optimal incident angle is related to an overall decrease in the P-wave velocity, S-wave velocity and density. The threshold corresponds to the point at which the reflection coefficients cease to exhibit the critical angle threshold as observed in Figures 5.4 (a) and 5.4 (b) but not in Figure 5.4 (c). This shows that in cases where the critical angle phenomenon angle can be sampled, maximum information is obtained by doing so.

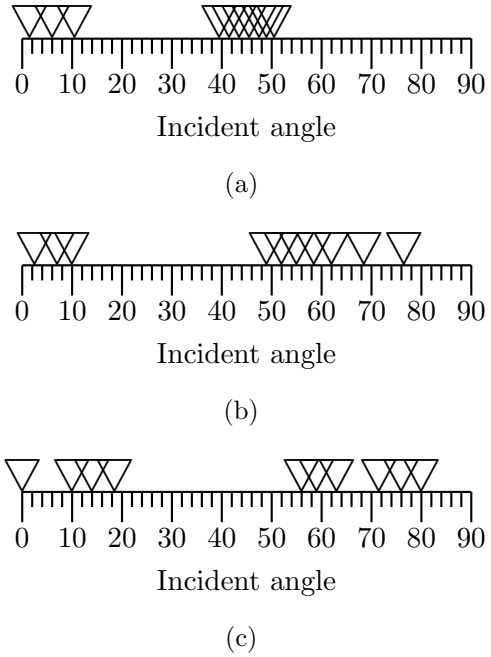

Figure 5.10: Processing designs for reservoir models with varying prior porosity ranges using ten traces selected using the method developed in Chapter 4. Plot (a) represents the final design for the 10% to 20% porosity reservoir, (b) the 20% to 30% porosity reservoir, and (c) the 30% to 40% porosity reservoir. Receivers are allowed to be selected between 0° and 90° incident angle.

The solid line in Figure 5.9 represents the reservoir model with a prior porosity range from 10% to 40%. This information curve encompasses the peaks of the three separate models. By increasing the prior porosity range the relative information provided by the small incident angles (relative to the large angles) increases when compared to the reservoir models with the porosity ranges 10% to 20% (dashed line) and 20% to 30% (dotted line).

Figure 5.10 shows the optimal design when ten angles are selected for processing at incident angles ranging from 0°-90° at 0.5° intervals for the three porosity ranges. As the porosity increases, so does the range of angles spanned by the selected angles.

This increase in angle spread is also apparent in Figure 5.9 where an increase in porosity broadens the information peak relating to the optimal angle. For the low porosity model (Figure 5.10 (a)) all angles are located within 0° to 51° with each of the two angle clusters located within 11° envelopes. As the prior

porosity range increases so does the maximum incident angle at which an angle is selected (76.5° for the 20-30% porosity reservoir, and 80° for the 30-40% porosity reservoir), as does the range over which angles are selected. As seen in Figure 5.9 the 20% to 30% porosity reservoir has the broadest information peak and this is reflected in Figure 5.10 (b) by having the largest incident angle range spanning 49.5°-76.5°. Note that this latter range almost precludes that of the low porosity design, again illustrating the importance of prior porosity information in AVA survey design. The relatively high importance of the near angles for the 30%-40% porosity reservoir is shown by angles being selected between 0° and 20° instead 0° and 10° as seen in the lower porosity reservoirs.

Although these processing designs were found using purely computational methods they make sense intuitively. Comparing Figures 5.5 and 5.10 we see that the optimal trace selection designs use angles so as to span (i) near-vertical incident angles, and (ii) the critical angle depicted in Figure 5.5, or (iii) the highest-curvature region of Figure 5.5 if no critical angle effects are observed (e.g. Figure 5.5 (c)). Thus the selected angles are sure to sample the ranges spanning the maximum variation in data values for the expected range of possible models. This 'strategy' produces the maximum data-model sensitivity, and is consistent with the expected behavior of optimal experimental design methods discussed in Curtis (2004a,b).

Figure 5.11 shows the trace selection design for the three varying porosity reservoir models if all ten selected angles are required to be located in the incident angle range 0°-70°. There is no change in the 10%-20% porosity reservoir since all selected angles in Figure 5.10 were located at angles less than 70°.

The two other trace selection designs show no change in the near offset selected angle positions. At large offsets the closest selected angle remains the same with the remaining angles more closely spaced with a single selected angle at the maximum offset of 70°. Although Figure 5.9 shows that small angles for a high porosity reservoir correspond to the largest entropy values for a single selected angle, the final ten angle design (Figure 5.11 (c)) locates the majority of the angles at large offsets, a result unobtainable by only considering the single angle entropy plot in Figure 5.9.

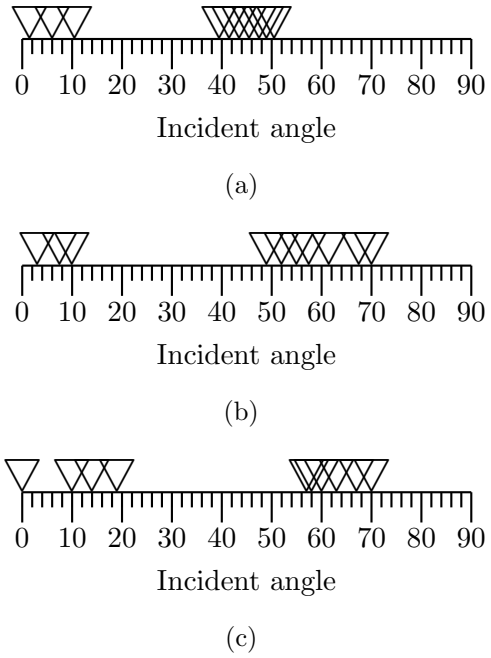

Figure 5.11: Processing design for the reservoir model with varying prior porosity ranges using ten selected traces using the method developed in Chapter 4. Plot (a) represents the final design for the 10% to 20% porosity reservoir, (b) the 20% to 30% porosity reservoir and (c) the 30% to 40% porosity reservoir. Receivers are allowed to be placed between 0° and 70° incident angle.

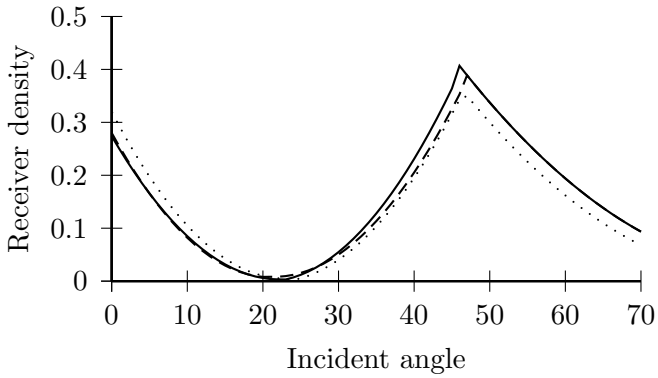

Figure 5.12: Receiver density plots calculated from the results in Figure 5.8. The regression lines shown have been fitted to 10° moving average results. The brine reservoir density is shown by solid line, the oil reservoir density by the dashed line and gas reservoir density by the dotted line.

5.5 Discussion

Although the ten-angle results shown do not constitute an industrial sized processing design, ten angles is sufficient to determine the receiver density profile along a full seismic survey to maximise the information recorded for any set number of selected receivers approximately. Figure 5.12 shows regression receiver-density curves fitted to the 10° moving average angle results calculated from Figure 5.8. As shown in Figure 5.8 the angle density plots are very similar for all three saturating fluids. The density value peaks around 50° corresponding to the information peak shown in the single-angle information curve (Figure 5.6). The density plot also shows a decreasing angle density pattern between 0° and 20°. The density is minimum at angles around 10°-35° which corresponds to the range void of selected angles in Figure 5.8.

Figure 5.13 shows the density result calculated from the angle positions in Figure 5.11 for the different porosity oil saturated reservoirs. It is clear that designing an optimal angle density processing profile for a reservoir is far more dependent on the reservoir porosity information than on information about the saturating fluid properties. For high porosity values (dotted line) a larger proportion of the selected angles to be processed should be placed at low incident angles, around 10°, and as the porosity decreases the highest density range also shifts to lower angles. For all cases, however, there is a density minimum around 30°.

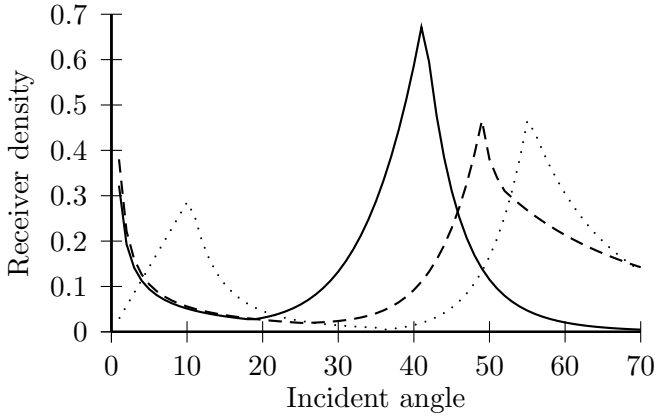

Figure 5.13: Receiver density plots calculated from the oil reservoir results in Figure 5.11. The regression lines shown have been fitted to 10° moving average results. The low porosity (10%-20%) oil reservoir density is shown by the solid line, the medium porosity (20%-30%) oil reservoir density by the dashed line and the high porosity (30%-40%) oil reservoir density by the dotted line.

Although the trace selection designs in Figure 5.10 are intuitive retrospectively given Figure 5.5, and indeed Figure 5.5 might be used to estimate the qualitative features of optimal trace selection designs, Figure 5.5 could not be used to create the exact processing designs themselves because the precise optimal angle-density curves (Figures 5.12 and 5.13) can not be estimated from Figure 5.5 directly. Hence, the sophisticated SED methods that we present are necessary.

These density profiles can be interpreted in several ways depending on how the recorded data is to be used. If all the traces are to be analysed using AVA/AVO methods then the density results indicate which traces require the most attention pre-analysis in order for example to remove sources of noise and multiples. Receivers with large density values require most effort since these traces will provided the most information on subsurface parameters. Traces corresponding to low densities provided little extra AVA information and time spent removing noise from the data might be proportionally reallocated to more important tasks. If only a subset of receivers are to be processed then the density indicates which traces should be selected to best constrain the subsurface parameters of interest. Rather than selecting single traces a more robust approach would be to select and process groups of traces in some small angle or offset range around each of the optimal locations found, perhaps combining the results of each group into a single AVA datum, using the stacking power of each group to reduce random

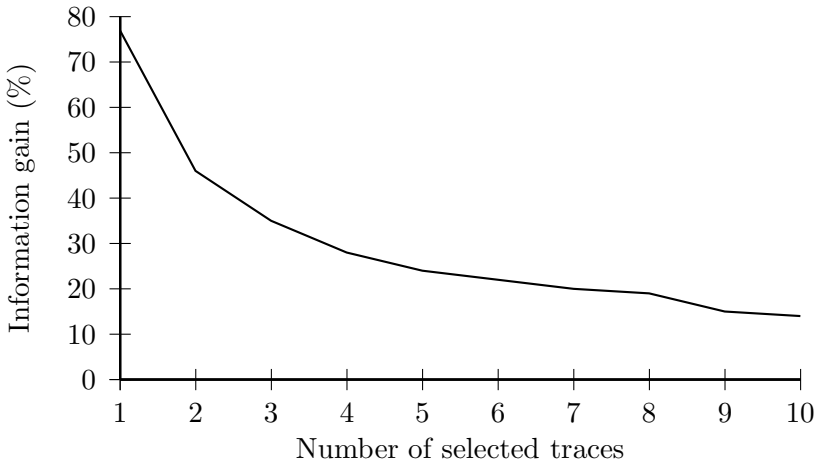

Figure 5.14: Information gain expected by processing the optimal trace selection compared to selecting evenly spaced receiver traces as plotted for a medium porosity oil reservoir at 500m depth as a function of the total number of receivers used.

noise in the datum. Both of these methods result in significant computational or time savings compared to standard methods. In the first method, the amount of time pre-analysis is reduced to a minimum by only concentrating on data that contains useful information. In the second method the processing time is reduced to a minimum by only considering an optimal subset of (groups of) the data.

Although optimal receiver selection profiles have been presented, the amount of extra information provided compared to a standard design has not been quantified. Figure 5.14 shows the expected increase in information gain on the subsurface parameters by using the optimal trace selection over the standard constant spatial receiver separation for a medium porosity oil reservoir located at 500m depth with a homogeneous overburden. Both the optimal trace selection design and the constant spatial separation trace selection designs place the majority of the receivers at large angles.

When only a few groups of receivers are used the expected information gain is very high. As more receiver groups are selected Figure 5.14 shows that the expected information gain falls but always remains greater than 14%. By using the notion that the results relate to receiver groups rather than individual traces a processing design using only three groups (which could consists of hundreds of individual traces) would produces expected gains of 35%.

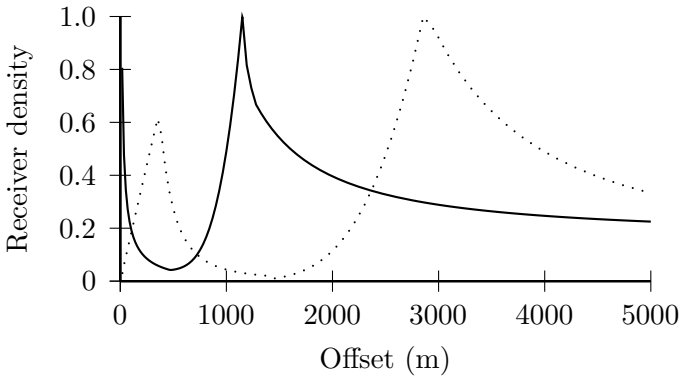

Figure 5.15: Normalized receiver density plots as a function of surface offset for a high porosity (30%-40%) reservoir at 1000m depth (dotted line) and a medium (20%-30%) oil reservoir at 500m depth (solid line).

Figure 5.14 agrees with the idea of "diminishing returns" which postulates that as the number of receivers increases, the relative advantage of using optimal designs over standard designs reduces (Coles and Morgan, 2009). The highest value niche for these design methods clearly occurs when focusing attention on relatively small portions of the data set, e.g. for rapid or low-cost AVA analysis. Indeed these design methods may significantly facilitate such applications.

Since the algorithm calculates the optimal incident angles at the caprock/reservoir boundary it is easy to transform the results into specific, more complex geometrical cases. Figure 5.15 shows the actual source-receiver offset profiles rather than incident angle which correspond to the optimal offset selection designs for different interface depths and subsurface properties (solid line represents a 20%-30% porosity interface at 500m depth and the dotted line a 30%-40% porosity interface at 1000m depth. In each case a homogeneous overburden model has been used).

Although the design may change if layers dip instead of being horizontal, the algorithm used to calculate optimal designs would still be robust. This is because I calculate the optimal distribution of incident angles at the reflector to be analysed. Whatever distribution of angles are found, these can be traced back to the surface through the overburden model as shown in Figure 5.16, in order to find optimal sensor locations on the surface.

In certain scenarios a limited design space may only be available to select traces

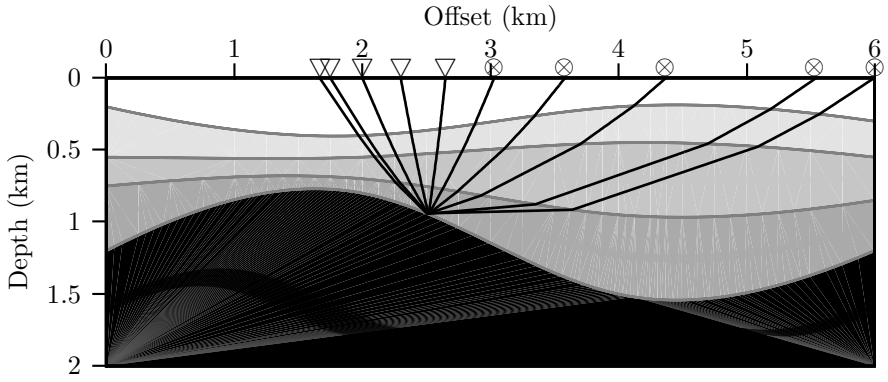

Figure 5.16: Raypaths calculated from five optimal angles located at a dipping interface. Tracing the optimal angles to the surface allows the properties of the complex overburden to be taken into consideration when transforming the optimal angles to optimal source-receiver offsets. Source locations are shown by $\otimes$ and receivers by $\bigtriangledown$. The gray shading represents velocity with darker shades corresponding to faster velocities.

from due to logistics of the data collection or methods used to process the data. In these cases the same methods as described above can be used by only calculating the information value over the relevant design space regions. It is also common practice to bin data into certain angle ranges, averaging the AVA response in each bin resulting in lower uncertainties on the mean AVA data within each bin. In such cases lower expected data uncertainties can be used in the methods above and the optimal angles found in the previous examples can be regarded as optimal bin centre locations.

In many realistic situations data uncertainty might increase with offset. In our methodology we only consider constant error with respect to offset. In future work it may be beneficial to add variable uncertainties so that information provided by far-offsets is weighted with respect to near-offsets which provide measurements with relatively low error. Alternatively the designs above are approximately correct if bin-widths increase with offset so that the uncertainties on amplitudes after stacking results within the bin are roughly equal for each bin.

Although our presentation of the new techniques has been targeted at how to optimise processing of AVO data by selecting which receivers from a regularly spaced data set should be processed to extract the maximum amount of information, the same principles can be used to design the actual data collection survey

with receivers only located at offsets which are expected to record maximum information about the subsurface. In this way unnecessary data is not recorded and all the data can be processed with the knowledge that the data set being processed is optimal.

5.6 Conclusions

Using the iteratively-constructive method developed in Chapter 4 in conjunction with the semi-empirical petrophysical model by Goldberg and Gurevich (1998), several optimal AVA trace selection processing designs have been presented for reservoirs with differing saturating fluids and porosity ranges. It has been shown, for a range of subsurface models, that reducing the prior uncertainty on the estimate of the reservoir porosity has a far greater influence on the resulting trace selection designs than has prior knowledge about the saturating fluid properties.

Using the optimal trace selection designs thus found to create receiver density profiles, large scale seismic survey data processing profiles can be specifically designed for subsurface horizons of interest. By running the trace selection algorithm for different targets of interest, optimal processing receiver density profiles can be created which can be applied to the same original data set to ensure that for each interface the maximum information is inferred for the lowest processing effort and hence cost.

Chapter 6

On standard and optimal designs of 2D surface seismic surveys

6.1 Introduction

In Chapter 5 I used the sequential design method developed in Chapter 4 to show how arbitrary prior petrophysical knowledge about the subsurface can be used to select optimal traces from a seismic dataset, post-survey, for AVO processing so that maximum information is inferred about the subsurface within some pre-defined limits on processing time and cost. Although the methods developed in Chapter 4 could be applied to pre-acquisition survey design the main focus of the work thus far has been on data selection.

The main drawback of using the methods developed in Chapter 4 to design pre-acquisition survey designs is that each source and receiver is defined by a separate dimension in the design space and consequently the problem becomes computationally intractable when more than ten sources or receivers are required. The approach in this chapter is similar to that used by Ajo-Franklin (2009) for linearised methods, where the design space is re-parameterised with a low number of hyper-parameters that control the receiver density, rather than individual receiver positions. Combining this technique with the nonlinear design method presented in Chapter 2 I can show that the problem of designing full 2D seismic surveys is computationally tractable for the first time using SED methods. I optimise the

information contained within the resulting AVO crossplots directly, bringing the acquisition design stage much closer to the seismic processing flow already used. I also integrate into our design method varying data errors which has not before been included in nonlinear Geoscientific design problems.

6.2 Method

Firstly I define precisely the type of parameters of interest in this chapter, the data type with which they are to be estimated, and then the forward function $\mathbf{F}_\xi$ relating the parameters ($\mathbf{m}$) and data ($\mathbf{d}$). I then specify how the amount of available information about the parameters can be measured or quantified. Next I describe our assumed model of a subsurface reservoir. Finally I introduce the design method used to optimise the design of the data acquisition survey so as to maximise information about the target parameters.

Whereas in Chapters 4 and 5 the Zoeppritz equations (equations 3.1) were used to link the model parameters ($\mathbf{m}$) and data ($\mathbf{d}$), in this Chapter I use the notion of AVO crossplotting (Section 3.2) to represent the data expected to be recorded during an AVO survey.

In this Chapter I apply nonlinear SED methods to design industrial scale seismic surveys by optimising receiver density profiles to maximise information about subsurface parameters A and B contained in AVO crossplots, so as to minimise velocity and density uncertainty after applying AVO inversion techniques.

In the method developed and used in Chapters 4 and 5 the design measure assumed that the errors where independent of the survey design. In this Chapter this restriction is removed and the full design measure

$$\Phi(\xi) = Ent\left\{\sigma(\mathbf{d}|\xi)\right\} - \int Ent\left\{\theta(\mathbf{d}|\mathbf{m},\xi)\right\}\rho(\mathbf{m})\,d\mathbf{m} \tag{6.1}$$

proposed by Shewry and Wynn (1987) (Section 2.3.1) is used.

The optimal survey design will be defined for a given prior Earth model pdf, $\rho(\mathbf{m})$. To be consistent with Chapter 5 the same reservoir model based on the Goldberg and Gurevich (1998) petrophysical model is used in conjunction with

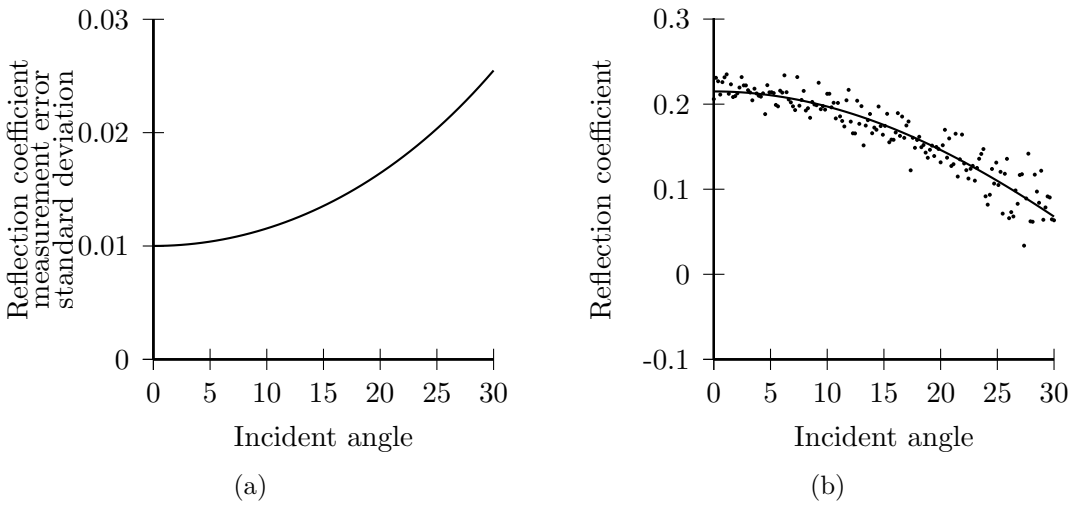

Figure 6.1: (a) Offset dependent reflection coefficient error. The error value represents the standard deviation of a Gaussian error. (b) Two-term Shuey equation (equation (6.1)) solution (solid line) calculated from simulated data (dots) for a specific survey design and reservoir model.

the parameter ranges defined in Tables 5.1 and 5.2.

6.2.1 AVA Design Method

In order to calculate the optimality of a specific experimental or survey design (equation (6.1)) I first construct an AVA crossplot based on prior information about the reservoir model described by $\rho(\mathbf{m})$, and the survey design ξ. Whereas in Chapters 4 and 5 I assumed a constant error with increasing offset and hence assumed that the integral term in equation (6.1) was constant and thus irrelevant from the perspective of survey design, I now consider offset-dependant errors and therefore also include the integral term. Figure 6.1(a) shows the standard deviation of the offset-dependant Gaussian error of the reflection coefficient that is used here, but any other such curve could be employed in the design method.

By formulating the problem using hyper-parameters, which define the position of individual receiver through a set of low-dimensional proxy parameters, I first consider the case where a total of 300 receivers are to be placed over a 30 degree offset range (the approximate range of valid angles of the two-term Shuey equation (equation (3.6))). Note that this is almost two orders of magnitude more receivers

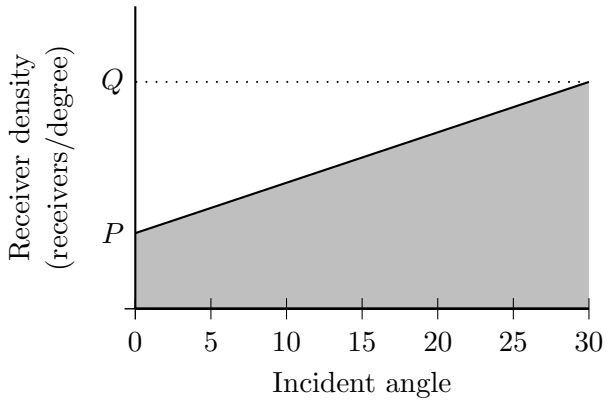

Figure 6.2: Receiver density profile (solid line) defined by the parameters P, the angular density at zero offset and Q, the angular density at maximum offset. The area of the shaded section is equal to the total number of receivers.

then have been designed previously using nonlinear design methods, and also that this approach allows almost any number of receivers to be located with approximate optimality. The work presented in Chapters 4 and 5 was limited by the design space dimensionality since the location of each selected receiver represented an additional dimension in ξ, and as a result the method suffered strongly from the "curse of dimensionality" (Curtis and Lomax, 2001). In this chapter I instead define the design vector ξ to describe the angular density of receiver locations. In the first place this is done using only two parameters,

$$\xi = [P, Q] \tag{6.2}$$

where P is the angular density of receivers at vertical incidence, and Q is the angular density at the maximum allowed angle (Figure 6.2). The density values at intermediate offsets are linearly interpolated. Any number of receivers can then be placed using this density distribution. Using this method allows almost unlimited numbers of receivers to be placed via two parameters. This formulation of the design problem is termed "hyper-parameterisation" in some fields, and "reduced parametrisation" by Ajo-Franklin (2009) who used it for optimising cross-borehole tomography surveys using linearised methods (Section 1.2.3). This approach does not allow each individual receiver to be placed at an arbitrary location; instead the optimal design is found via a set of hyper-parameters, in this case receiver density. The area under the receiver density plot (shaded area in Figure 6.2) represents the total number of receivers placed and therefore, for a given fixed

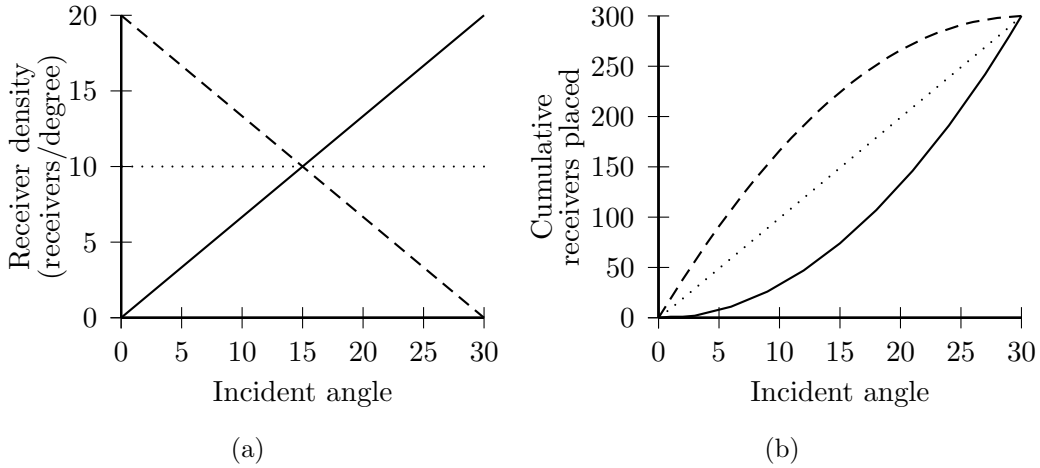

Figure 6.3: Receiver density profiles (a) for three survey designs and the corresponding cumulative number of placed receivers (b) as a function of incident angle.

maximum incident angle (I) and total number of receivers (N), the two design parameters are related by

$$Q = \frac{2N}{I} - P. \tag{6.3}$$

The extreme design parameter ranges considered here are $P = 0$, $Q = 20$ and $P = 20$, $Q = 0$, both shown in Figure 6.3(a). Figure 6.3(b) shows the cumulative number of receivers placed as a function of incident angle for the three example density profiles in Figure 6.3(a). It should also be stressed that a constant density with respect to angle at the interface does not equate to constant receiver separation in spatial receiver locations on the ground surface.

For each point in design space (possible combinations of the hyper-parameters, P and Q) the individual receivers can be located at the relevant offsets, and for a given reservoir model, the reflection coefficients at each of the placed receivers can be calculated by solving the Zoeppritz equation (equations (3.1)). For each receiver a Gaussian error is added to the reflection coefficient, with standard deviation shown in Figure 6.1(a). According to standard practice the two-term Shuey equation (equation (3.6)) is then fit in a least-squares sense to the resulting reflection coefficient data to determine the AVA gradient B and intercept A (equation (3.7)). Figure 6.1(b) shows the two-term Shuey equation solution calculated for one example of simulated data. This, however, only constitutes a single realisation of the data for one specific reservoir model.

To accurately estimate the pdf $\theta(\mathbf{d}|\mathbf{m},\xi)$, multiple realisations of the noisy data for the same reservoir model and receiver density distribution are required, and for each realisation a separate AVA intercept and gradient are calculated and histogrammed in a discretised AVA crossplot. The resulting crossplot represents an estimate of the uncertainty in calculating the intercept and gradient due to the measurement noise for the given receiver density profile. The histogram is normalised to have unit volume whereafter it represents a numerical approximation to the pdf $\theta(\mathbf{d}|\mathbf{m},\xi)$. A numerical approximation can therefore be calculated for the integral term in equation (6.1) as

$$\int \rho(\mathbf{m})\, Ent\left\{\theta(\mathbf{d}|\mathbf{m},\xi)\right\} d\mathbf{m} \approx \frac{1}{M}\sum_{i} Ent\left\{\theta(\mathbf{d}|\mathbf{m}_i,\xi)\right\} \tag{6.4}$$

where M is the total number of reservoir models sampled from the prior parameter distribution $\rho(\mathbf{m})$.

The marginal distribution $\sigma(\mathbf{d}|\xi)$ in equation (6.1) is represented by the normalised AVA crossplot histogram resulting from all of the realisations for *all* model parameter realisations (i.e., for all possible data that could be collected in the survey given the prior information on the possible range of reservoir models). Figure 6.4 shows the marginal distribution $\sigma(\mathbf{d}|\xi)$ for a design with constant receiver density calculated from 200,000 model samples from a gas-filled reservoir with 50 realisations of the data for each model. The crossplot shows that the majority of the reservoir models and data realisations result in both a negative gradient and intercept, and only a very small proportion exhibit a positive gradient.

For each survey design the expected information gain is then calculated using equation (6.1). The density profile that corresponds to the maximum value is the optimal survey design, since that design is expected to record data that will provide maximum information about, and hence most tightly constrain the subsurface reservoir parameters.

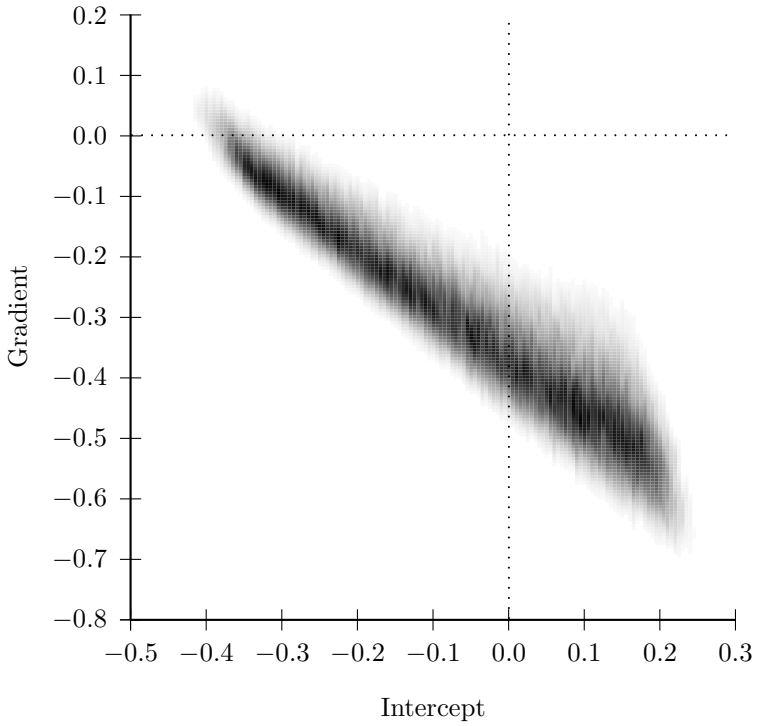

Figure 6.4: Probability distribution of the expected AVA crossplot for a gas saturated reservoir surveyed with a constant receiver density profile given available prior information about petrophysical parameters in Tables 5.1 and 5.2

6.3 Results

For all results crossplots have been discretised into 160 bins over the range -0.5 to 0.3 in the intercept dimension and 200 bins over the range -0.8 to 0.2 in the gradient dimension. For each survey design the reservoir model has been sampled 200,000 times, and for each particular reservoir model 50 realisation of the data have been produced. The receivers have been located using two hyper-parameters (P and Q) which represent the receiver density at zero offset and 30° respectively.

6.3.1 Porosity

Chapter 5 showed that the reservoir model parameter that has the largest effect on the survey design is the porosity. Figure 6.5 shows the information gain values as a function of P, the zero offset receiver density for gas saturated reservoirs with low (10%-20%) and high (30%-40%) uniform porosity ranges. The design

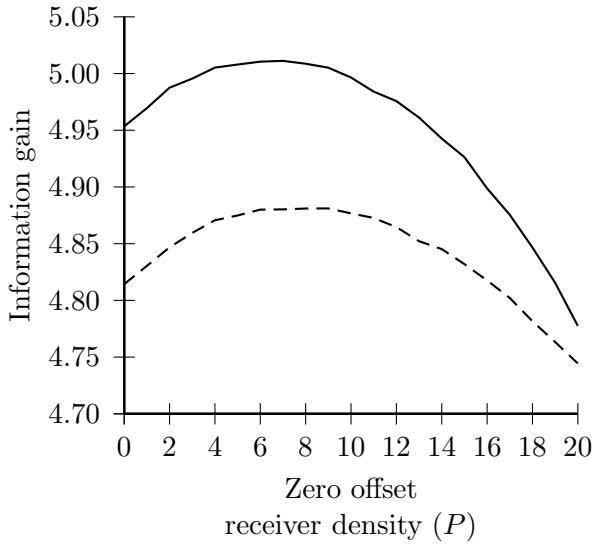

Figure 6.5: Information gain as a function of zero offset receiver density (P) for a gas reservoir with differing porosity ranges for a survey consisting of 300 receivers. All parameters are as in Table 5.1 except the solid line shows the results of the low porosity model (range restricted to 10%-20%) and the dashed line the high porosity model (30%-40%).

corresponding to the maximum information gain is the optimal survey design. For a maximum offset of 30 degrees and 300 receivers a P value of 10 receivers per degree equates to a constant angular receiver separation; values of P less than 10 result in more receivers being placed at larger offsets than near offsets, and P values greater than 10 result in more receivers located at near offsets than far offsets (Figure 6.3). The results show a small increase in the optimal zero offset receiver density as porosity increases: the optimal zero offset receiver density (P) for the low porosity model is 7 receivers per degree whereas the optimal receiver density for the high porosity model is 9 receivers per degree.

These results for the optimal designs are intuitive. For both of the cases in Figure 6.5 there is a larger proportion of receivers at far offsets compared to near offsets. Since the data error increases with offset, proportionally more receivers are required at large angles of incidence to constrain the crossplot gradient compared with fewer receivers required to constrain the reflection coefficient near zero offset. The end-member survey designs only constrain either the crossplot intercept ($P = 20$, $Q = 0$) or gradient ($P = 0$, $Q = 20$) resulting in the other parameter having a high associated uncertainty. However, although that much is intuitive, without performing the survey design algorithm the exact receiver density profile

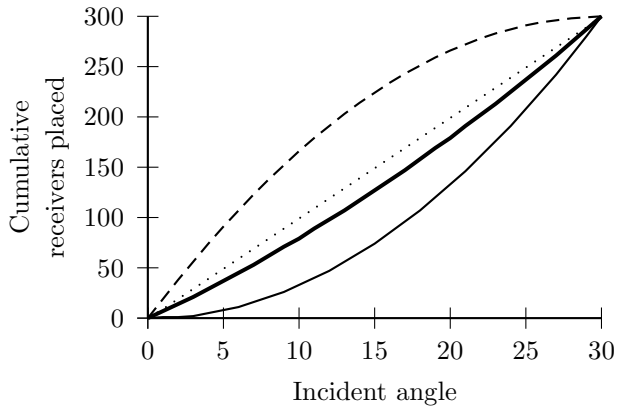

Figure 6.6: Optimal cumulative number of receivers placed (thick solid line) for the "one-size-fits-all" survey design of $P = 8$ when locating a total of 300 receivers. The dotted thin line represents the profile for a constant angular receiver density, and the two other profiles show the receiver density extremes considered herein (dashed line represents $P = 20$, $Q = 0$, and the thin solid line $P = 0$, $Q = 20$).

needed to ensure optimality would remain unknown.

The result shows that as the porosity increases the relative number of receivers required at smaller offsets compared to larger offsets increases. Although a difference in the optimal survey is seen, the large difference observed in Chapter 5 is not apparent. This is because the range of incident angles considered in Chapter 5 extended to 70° whereas in this work they never exceed the critical angle. Thus, I demonstrate that the nonlinearity in the forward function $\mathbf{F}_\xi$ around the critical angle creates strong dependence of the optimal design on the particular reservoir parameter ranges expected prior to conducting the survey. For pre-critical surveys on the other hand, there is (roughly speaking) a "one-size-fits-all" design that has an optimal P value of 8 when placing 300 receivers. Figure 6.6 shows the cumulative number of receivers placed with respect to angle for an optimal design of 8 receivers per degree at zero offset (P) increasing linearly to a value of 12 (Q) receivers per degree at an incident angle of 30 degrees.

6.3.2 Fluid Content

Figure 6.7 shows the information gain values as a function of P, the zero offset receiver density across the full prior porosity range (Tables 5.1 and 5.2) for all three general reservoir models relating to each of the three possible saturating

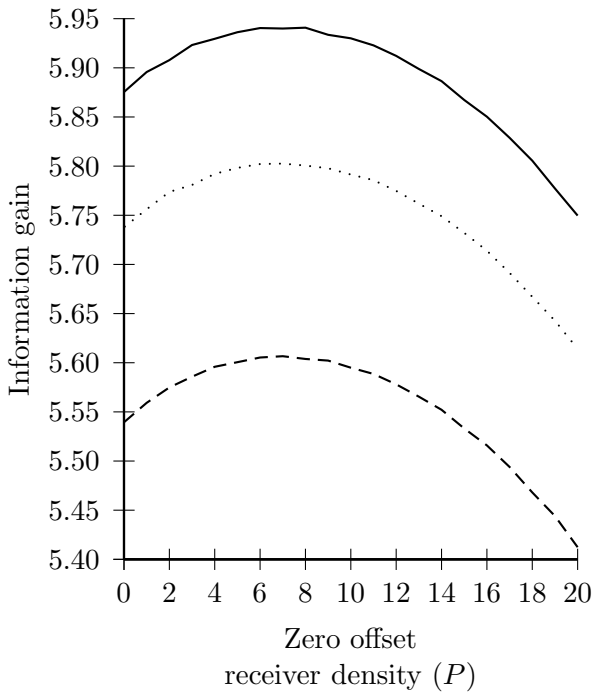

Figure 6.7: Information gain as a function of zero offset receiver density (P) for the general reservoir properties (Tables 5.1 and 5.2) when placing 300 receivers. The gas saturated reservoir is shown by the solid line, the brine reservoir by the dashed line and the oil reservoir by the dotted line.

fluids: oil, gas and brine. Although the three reservoir models result in different information gain values, the shape of the profiles are similar with optimal zero offset receiver densities of 7 receivers per degree for the oil and brine reservoirs and 8 receivers per degree for the gas saturated reservoir. Again, the varying fluid content results support the idea of a "one-size-fits-all" survey design for any typically-expected porosity and fluid content.

6.3.3 Number of Receivers

In all of the above optimal designs a total of 300 receivers have been placed using the hyper-parameters, P and Q. However, if there are cost or logistical constraints fewer (or more) receivers may be required. In the following results the receiver density values have been normalised so that $\hat{P}$ ranges from 0 to 1 and $\hat{Q}$ from 1 to 0, with a value of 1 corresponding to the maximum receiver density possible in each case. When $\hat{P} = \hat{Q} = 0.5$ there is a constant angular receiver

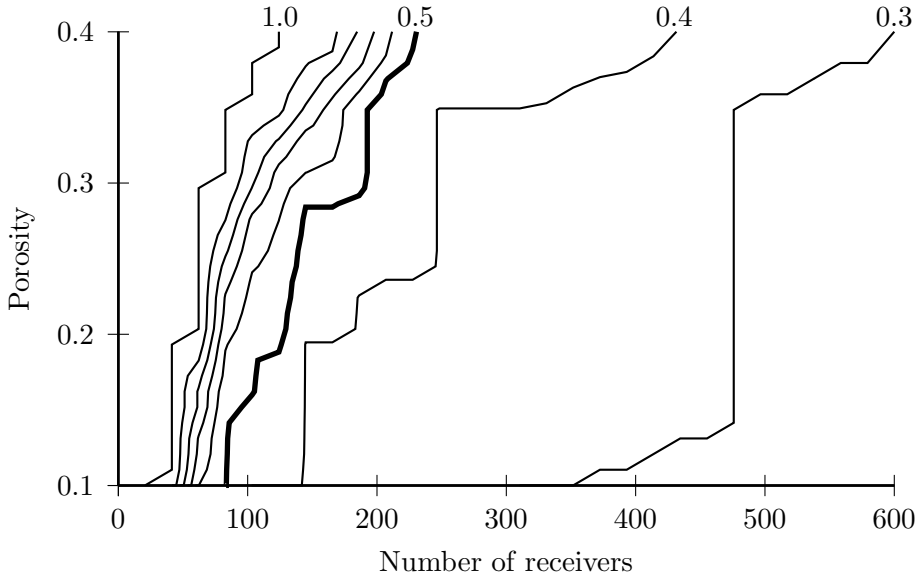

Figure 6.8: Optimal normalised $\hat{P}$ contours as a function of porosity and total number of receivers. The 0.5 contour represents a constant receiver separation with respect to angle. Values higher than 0.5 indicates a greater receiver density at near offset angles and values less than 0.5 a greater receiver density at far offsets.

separation.

Figure 6.8 shows the normalised optimal receiver density designs for the standard gas saturated reservoir for varying porosity values and total number of receivers located.

The plot shows that when fewer than 250 receivers are located using the hyper-parameter method, the optimal survey depends significantly on the number of receivers to be placed as shown by the high $\hat{P}$ gradient in the horizontal direction. When placing fewer than 250 receivers there is also a significant dependence on the prior porosity range, particularly for higher porosity reservoirs. For example, if only 100 receivers were to be placed, depending on the prior porosity value, an optimal design could have a $\hat{P}$ value ranging from less than 0.5 to 1.0.

When placing more than 250 receivers, the optimal $\hat{P}$ value is always less than 0.5 representing an increasing receiver density with offset and the dependence on the prior porosity is significantly reduced. Figure 6.8 also shows that the value of $\hat{P}$ never falls below 0.3 and although not shown, this result has also been observed

when placing up to a total of 5000 receivers. This result implies that once over a certain threshold of total-number-of-receivers-placed, the relative distribution of receivers remains constant and defined by $\hat{P} = 0.3$ so that both the crossplot gradient and intercept can be well constrained given the offset-dependent error (Figure 6.1(a)).

The results in Figure 6.8 imply that the "one-size-fits-all" design that was evident for the porosity and fluid content does extend to all porosity ranges and saturating fluids when the total number of receivers is greater than 250, but does not apply when the total number of receivers is less than 250.

6.4 Discussion

Although the results above show that the optimal designs for AVA experiments using a linear receiver density distribution can be expressed by a "one-size-fits-all" design once the total number of receivers surpasses a threshold value, no measure has yet been quantified of how much extra information about the subsurface parameters is provided by the optimal design when compared to the standard 2D seismic survey design of constant spatial receiver separation. We therefore compare the information gain provided by both a constant spatial receiver design and an optimal AVA design using the simple two-layer model in Figure 3.1. Scenarios are tested in which the interface is at two different depths and in which there is a different total number of receivers.

The first scenario has an interface depth of 1613m and a maximum offset of 1875m (which corresponds to an incident angle of 30°). Placing a total of 150 receivers results in a constant separation of 12.5m for a standard survey design. Figure 6.9 shows the number of cumulative receivers placed for both the standard design (dotted line) and the optimal design (solid line) with a zero offset receiver density (P) of 4 receivers per degree increasing to 6 receivers per degree at 30 degrees offset (Q). The two designs look very similar, and using a standard gas reservoir as the prior model with porosity and the other parameters potentially ranging across the full uniform bounds in Tables 5.1 and 5.2 results in only a 2% information gain by using the optimal survey design over the standard design.

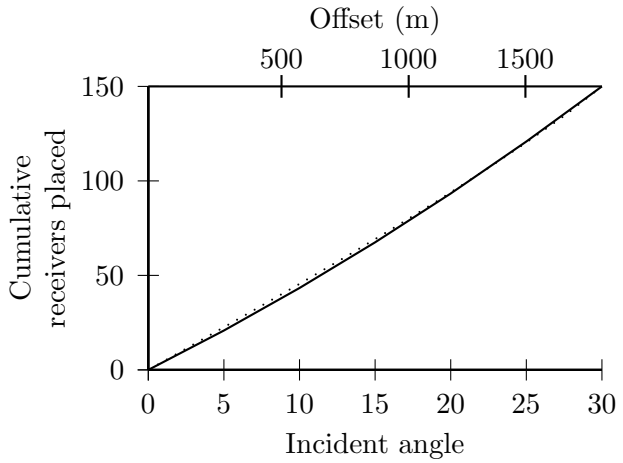

Figure 6.9: Cumulative number of receivers placed for a total of 150 receivers over a 30° range for a target depth of 1613m. The dotted line shows the constant spatial separation result and the solid line the optimal result.

The second scenario consists of a gas reservoir with the target interface at a depth of 500m which corresponds to a 30 degree maximum offset of 577m. Due to the small available offset range only 50 receivers are placed resulting in a constant separation of 10.7m for the standard survey design. Using the design algorithm a survey consisting of 50 receivers has the maximum possible number of receivers at small offsets (Figure 6.8). Figure 6.10 shows the comparison of the standard design (dotted line) and the optimal design (solid line).

While the two designs are now quite different, the expected increase in information gain from this survey is still only 5%.

Figure 6.11 shows the expected information gain values for an oil reservoir as a function of the total number of placed receivers when comparing the optimal design with a standard design of equal spatial receiver separation. The plot shows that for surveys consisting of fewer than around 50 receivers, the optimal design provides significantly more information than a regularly spaced design. However, the information gain provided by adding additional receivers to the optimal design compared to simply performing a standard design diminishes as the total number of receivers increases. This agrees with the idea of "diminishing returns" which postulates that as the number or receivers increases the relative advantage of using optimal designs reduces (Coles and Morgan, 2009). The large change in expected information gain for low numbers of receivers can also be

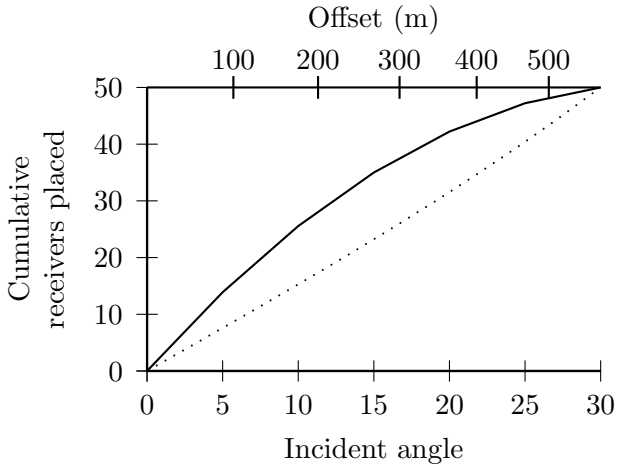

Figure 6.10: Cumulative number of receivers placed for a total of 50 receivers over a 30° range for a target depth of 500m. The dotted line shows the constant separation result and the solid line the optimal result.

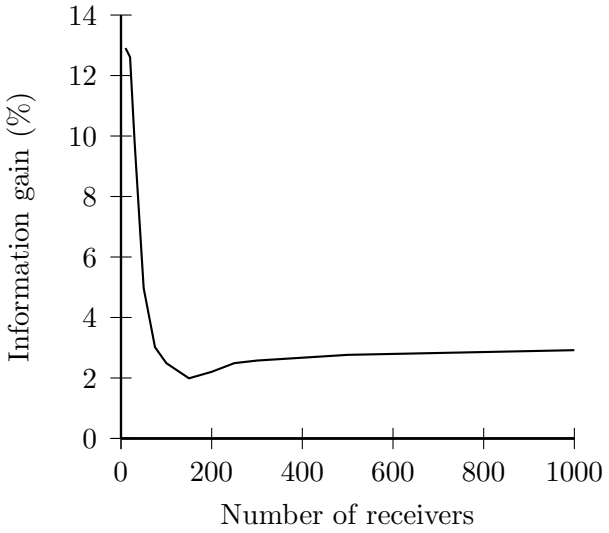

Figure 6.11: Information gain expected from using the optimal receiver distribution compared to a standard survey design as a function of the total number of receivers for a general oil filled reservoir.

explained by Figure 6.8: as the number of receivers used increases from 0 to 200, the optimal design quickly changes from one with maximum receivers at small offsets ($\hat{P} = 1.0$) towards a design which has equal angular receiver spacing ($\hat{P} = 0.5$) which occurs at 180 receivers. Therefore, as the number of receivers increases the optimal design tends to a design of equal angular receiver spacing and as a result it is expected that the relative information gain of the optimised survey will decrease. However, as the number of receivers increases beyond 180 the $\hat{P}$ value tends towards 0.3 corresponding to an optimal survey design that becomes less like the standard design. Figure 6.11 shows that this results in an increased information gain with increasing number of receivers, and although not shown we have tested that this remains true up to 5000 receivers. This is in contrast to the idea of "diminishing returns" which would still expect a decrease in information gain with increased number of receivers.

Although the methods used to locate the optimal receiver positions are different to those used in Chapter 4 which allowed receivers to be placed arbitrarily at any offset (in contrast to using hyper-parameters), the results should be approximately consistent since both methods maximise information about the same subsurface properties. I compare the results of both methods when used to place 10 receivers (around the order of the maximum number able to be placed using the method developed in Chapter 4 using a standard desktop PC (see Chapter 5) for a brine saturated reservoir within the angular range of 0 to 30°. For this comparison to be fair we have used a constant error with offset, to be consistent with that of Chapter 4.

Figure 6.12 shows the cumulative number of receivers placed as a function of incident angle for the method developed in Chapter 4 (solid line), a $\hat{P}$ value of 1.0 (dashed line) equating to an optimal survey when only 10 receivers are used, and a $\hat{P}$ value of 0.3 (dotted line) which reflects the optimal design when more than 100 receivers are placed and Figure 6.13 shows the actual receiver positions for each of these distributions. Figure 6.12 shows that the results calculated using the method developed in Chapter 4 in part match both results calculated using the linear receiver density method. Optimal receivers are located at both small offsets and large offsets to accurately estimate both the AVO gradient and intercept with a region void of receivers between 10 and 22 degrees offset. This is a result unobtainable in the examples above due to the relatively coarsely parameterised design space (linearly varying angular receiver density). As seen

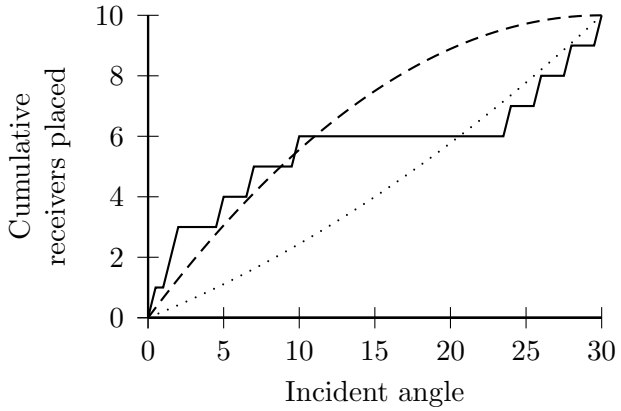

Figure 6.12: Cumulative number of receivers placed for a brine filled reservoir comparing the method developed in Chapter 4 with the receiver density method introduced above. The solid line represents the results found using the method developed in Chapter 4 when 10 receivers are placed, the dashed line the optimal receiver density result for a survey using 10 receivers and the dotted line an optimal survey when more than 100 receivers are placed.

in Figure 6.8 placing a low number of receivers results in the optimal design being located in a transition zone between a $\hat{P}$ value ranging between 0.3 and 1.0. Figures 6.12 and 6.13 show that the method developed in Chapter 4 spans both of these. Since the method developed in Chapter 4 is restricted to placing a maximum of around 10 receivers, it is impossible to say if additional receivers would make the optimal result tend towards the optimal result of $\hat{P} = 0.3$.

The above implies that a hyper-parameterisation where the receivers are located at offsets using only the receiver density control points at 0° and 30° offset (Figure 6.3) is too coarse for the purpose of this comparison. To increase the receiver placement complexity the design space can be increased to contain three hyper-parameters defining the receiver density at 0°, 15°, and 30°. This increase in complexity alters the location of placed receivers and therefore the reflection coefficient data (Figure 6.1(b)) used to calculate the AVO crossplot (Figure 6.4).

Figure 6.14(a) shows example normalised receiver density profiles and Figure 6.14(b) the corresponding normalised cumulative receiver plots that become possible when a third hyper-parameter ($\hat{M}$) is introduced to represent the receiver density at 15°, half the maximum incident angle, and when linear interpolation is used between 0° and 15°, and between 15° and 30°. Adding an extra hyper-parameter increases the design space by one dimension but allows more variation

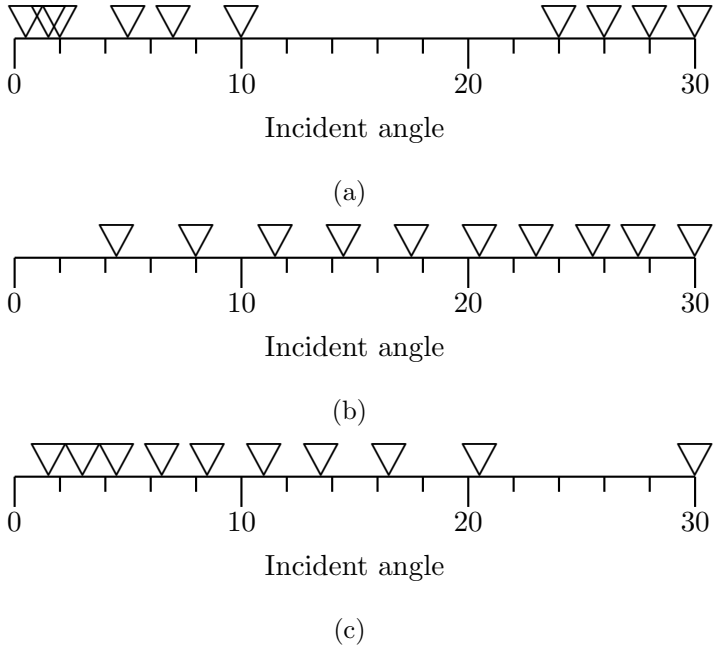

Figure 6.13: Optimal locations for placing 10 receivers for a brine saturated reservoir using (a) the method developed in Chapter 4, (b) a linear receiver density with a $\hat{P}$ value of 0.3 and (c) a $\hat{P}$ value of 1.0.

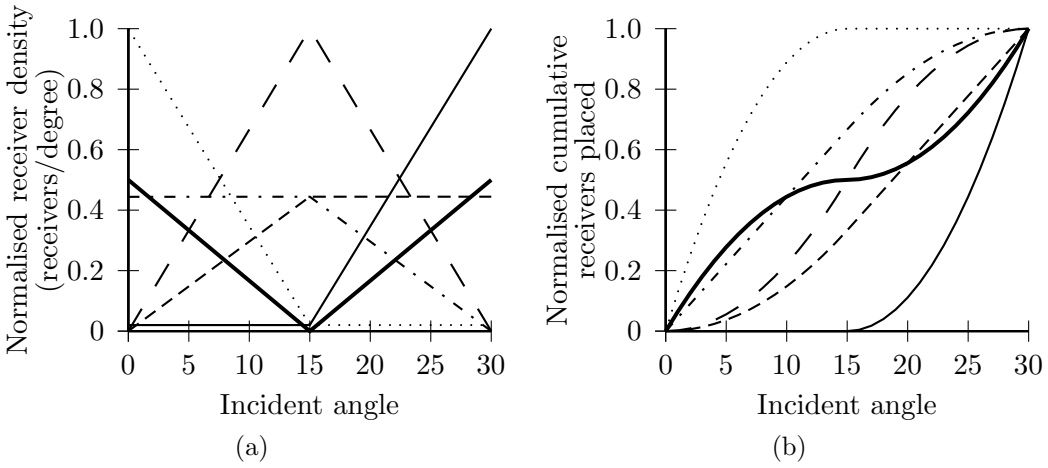

Figure 6.14: (a) Normalised receiver density profiles for possible survey designs using three hyper-parameters, and (b) the corresponding normalised cumulative number of placed receivers, both as a function of incident angle. Note that in (a) the solid and dotted lines before and after 15 degrees respectively have been shifted so that they are visible.

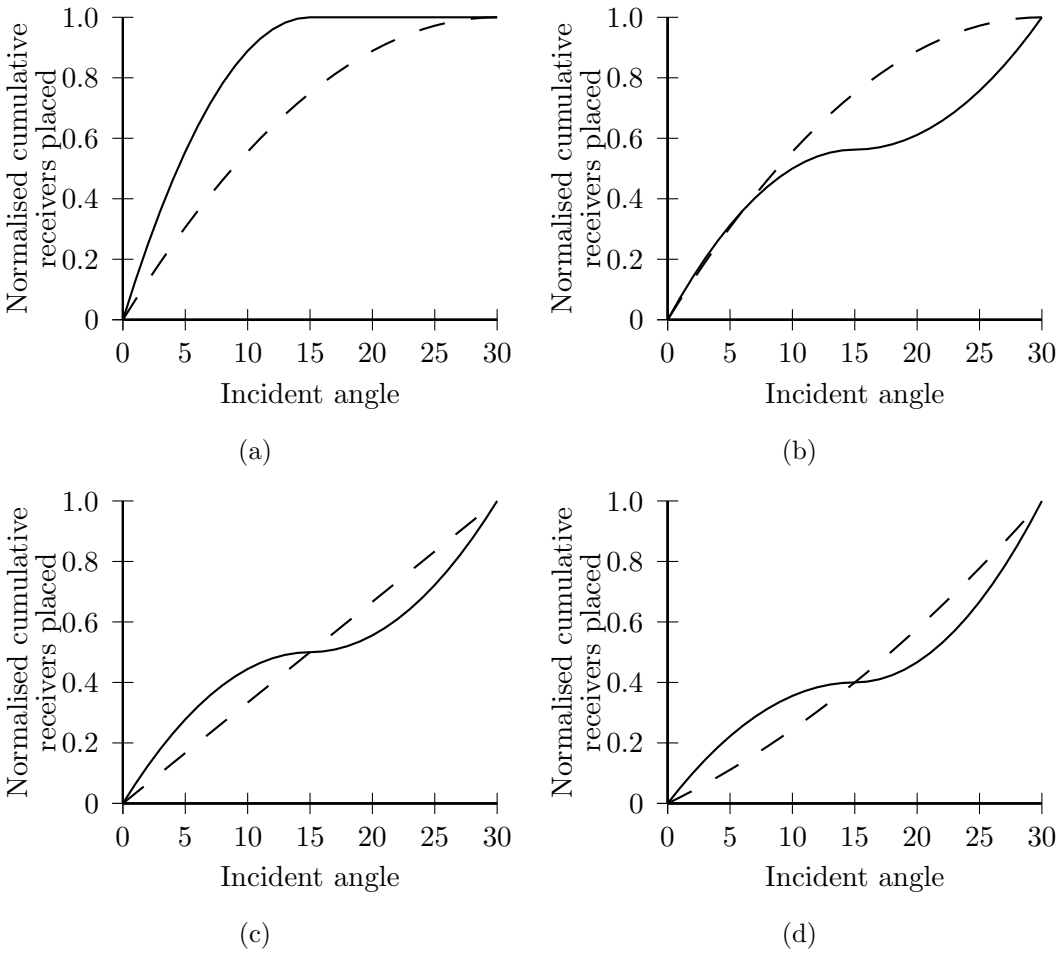

Figure 6.15: Normalised cumulative placed receiver profiles for optimal surveys consisting of (a) 10 receivers, (b) 20 receivers, (c) 100 receivers and (d) 600 receivers. In each plot the dashed line represents the 2 hyper-parameter result and the solid line the 3 hyper-parameter result.

in survey designs.

Optimal surveys that use a total of 10, 20, 100, and 600 receivers were calculated for a brine saturated reservoir (Tables 5.1 and 5.2) using the three hyper-parameter model. Figure 6.15 shows how the three hyper-parameter results (solid line) differ from the two hyper-parameter results (dashed line) for surveys consisting of (a) 10 receivers, (b) 20 receivers, (c) 100 receivers, and (d) 600 receivers. When only ten receivers are placed the two hyper-parameter result has the highest density of receivers at near offsets; with the three hyper-parameter result (Figure 6.15(a)) the same result is seen but now all receivers are located in the first 15° with no receivers between 15° and 30°. The value of $\hat{M}$ (the normalised receiver

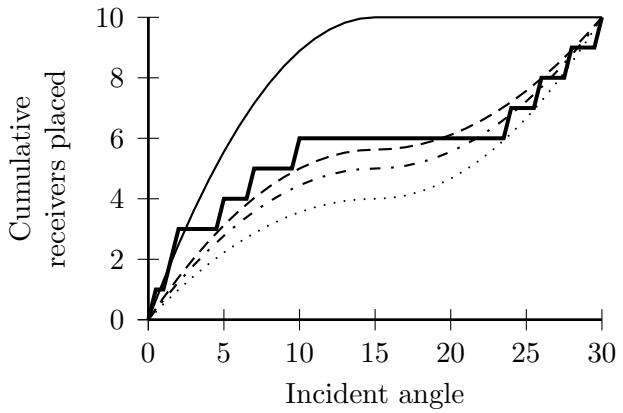

Figure 6.16: Cumulative number of receivers placed for a brine filled reservoir comparing the method described in Chapter 4 with the three hyper-parameter design method. The thick solid line represents the results found using the method described in Chapter 4 when 10 receivers are placed, the thin solid line represents the optimal receiver density result for a survey using 10 receivers, the dashed line an optimal survey when using 20 receivers, the dot-dashed line an optimal survey when using 100 receivers, and the dotted line when 600 receivers are placed.

density at 15°) for the 20, 100 and 600 receiver designs is 0 resulting in the inflection point seen in Figure 6.15 (b), (c) and (d). When using two hyper-parameters the result for the 10 and 20 receiver designs are identical. However, when using three hyper-parameters the results show a significant difference with the 20 receiver design resembling the 100 and 600 receiver designs. The addition of the extra hyper-parameter now produces optimal results that more closely resemble the result obtained using the method developed in Chapter 4.

Figure 6.16 shows that the ten receiver results using the method developed in Chapter 4 are best matched by the results found when placing twenty receivers using the three hyper-parameter method.

Although it might initially seem worrying that the ten-receiver results do not match using the two design methods, this is almost certainly because the forward function $\mathbf{F}_\xi$ considered here differs from that used in Chapters 4 and 5. The former chapters assumed that the recorded amplitudes of arriving waves at each receiver would be inverted directly for petrophysical parameters using equations (3.1) and the petrophysical model of Goldberg and Gurevich (1998). In this chapter, however, I assume that recorded amplitudes will be summarised by AVO intercept and gradient parameters as is standard practice in industry, and

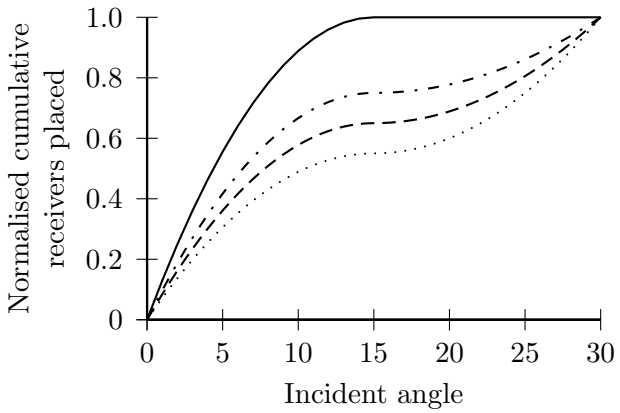

Figure 6.17: Normalised cumulative placed receivers for optimal three hyper-parameter surveys consisting of 12 receivers (solid line), 13 receivers (dot-dash line), 14 receivers (dashed line) and 15 receivers (dotted line).

that these AVO parameters will be inverted using equations (3.7). Hence, in each case the effective data sets inverted differ, and so do the forward functions. Nevertheless, the similarity between the bold and dashed lines in Figure 6.16 shows that the resulting designs in each case are strongly related, as we would hope to be the case if the standard industrial AVO workflow is robust. I find that using the method herein, the threshold at which the design shifts from that in Figure 6.15(a) to having an inflection point as in Figure 6.15(b) occurs at 13 receivers (Figure 6.17).

The information gains calculated using the new designs compared to a standard, equally spaced design result in values approximately 3% higher than those seen in Figure 6.11 for the three studied models. If the idea of "diminishing returns" applies in this case, adding a forth hyper-parameter would result in a further smaller information gain and a far larger design space to search for the optimal design.

Although the optimal surveys produced using the Bayesian design method result in information gains compared to standard constant spatial designs the actual gain values are relatively small, especially for large-scale industrial designs. In the analysis so far I have not taken into consideration the extra cost factors (for example acquisition and processing costs) introduced when using an optimal (non-regular) design, or the fact that surveys are generally designed to optimise noise attenuation and imaging and not solely to record data for AVO processing.

In practice additional cost factors should be applied to Figures 6.5 and 6.7 with a zero additional cost applied to the standard design and a non-zero cost to all other designs with a magnitude dependent on the extra costs expected to be incurred. In this way a true optimal design could be determined.

Extra costs associated with using optimal designs are likely to be significant. For marine seismics this would require that streamers are re-designed with an extremely high associated cost. For land seismics there would be significant extra expense due to the need to survey and lay geophones over wide areas according to non-standard spatial templates. In both cases there would be additional cost in adapting noise attenuation and imaging methods to non-uniform receiver densities. Thus, we conclude that in practice, if we balance the magnitude of the gains in information against the extra cost incurred, the best surveys to use for AVA studies will in fact almost always be regularly-spaced surveys. This is a somewhat surprising result, given that standard surveys have been designed to simplify and aid noise attenuation and imaging. However, it does explain why these standard designs have also been so successful for petrophysical inversion in the past. The relative drop in information resulting from designing for noise attenuation and imaging rather than for AVA is generally lower than 10%.

6.5 Conclusions

A Bayesian design method has been proposed which, when combined with a reservoir model and offset-dependent error measure, produces industrial scale, optimal AVA designs that are shown to decrease the expected uncertainty on the reservoir parameters compared to a standard design using the same number of receivers. Although the optimal designs are similar for different porosity values and saturating fluids, the total number of receivers in the survey has a large affect on the optimal design. However, once a particular threshold on the total number of receivers has been passed there exists a "one-size-fits-all" design that is optimal for any porosity, fluid content or number of receivers.

Although these optimal designs provide extra information, the two examples shown result in gains of up to only around 5% when compared to a standard survey with constant spatial receiver separation. Even when the hyper-parametrisation

is redefined to be more complex, these gains generally remain less than around 10%. When the cost of collecting and processing the new data is accounted for it is unlikely that this increase in information will represent value for money. For the given prior reservoir model and offset dependant error it is therefore concluded that although the "one-size-fits-all" result shown above is optimal, when the cost of data collection and processing are considered the current standard seismic survey design of constant spatial receiver separation is in fact optimal for pre-critical AVA surveys.

Chapter 7

Discussion

The work in this thesis has extended the scope of Geoscientific problems that can be addressed using nonlinear SED methods. As shown in Chapter 1 the majority of problems studied previously using SED algorithm have relied on linearised theory in order to make the problems computationally tractable. The methods proposed to tackle full nonlinear scenarios were limited to solving very simplistic problems.

In Chapter 4 a novel iteratively-constructive design algorithm was introduced that is capable of producing complex survey designs without any linearisation of the forward function or design metrics. The sequential part of the design algorithm belongs to the deterministic class of algorithms (although a stochastic Monte Carlo method was used to calculate the design metric at each iteration), and has been shown to produce close to globally optimum results where the optimum can be evaluated. The iteratively-constructive algorithm presented can be used in both linear and nonlinear design problems, although the computational benefits are mainly seen when studying nonlinear problems where linear design methods are shown to perform poorly.

It has been stated that the principal reasons that SED theories have not gained general acceptance in the Geosciences is that linearised methods, which have previously been the only available methods to design full scale surveys, are not robust in nonlinear problems and have therefore not been implemented. Hence as a consequence there is a lack of awareness that SED methods exist at all.

By introducing a nonlinear design algorithm capable of designing complex surveys it is hoped that the acceptance and uptake of nonlinear SED methods in the Geoscience will increase, and become an integral step in more survey design studies.

The algorithms presented in Chapter 5 introduced for the first time industrial-scale designs based on nonlinear SED methods. Although the results shown were presented for post-survey data selection problems the methodologies shown are equally applicable for survey design problems. The results show that when processing AVO data not all traces require (or should be allocated) the same amount of a processor's time or effort, as some traces provide far more information about the subsurface than others. By incorporating a petrophysical model it is shown that the traces that contain most information are highly dependent on the subsurface properties with the reservoir porosity having the largest effect. This result reinforces the need for nonlinear SED methods over heuristic design methods by showing that no single processing design is applicable to multiple targets. Given the limited amount of computational power required by the nonlinear algorithm it is feasible that a unique design should be created for every subsurface target. The results also show which areas of prior knowledge are most critical to bound as tightly as possible when producing the processing design.

Whereas implementing optimal designs at the data acquisition stage usually trades-off with a large increased associated cost, the cost of changing current processing workflows is far lower. With this in mind, and the fact that the algorithm is based on incident angle at the subsurface interface rather than actual source-receiver offset, a framework for optimal AVO processing has been introduced for which I see no major drawbacks in being implemented in most processing workflows, especially if the information gains shown in Chapter 5 are realised.

The algorithms presented in Chapters 4 and 5 are based on methods that are not commonplace in the seismic industry (e.g. direct inversion of the Zoeppritz equation). The design methods presented in Chapter 6 aim to bring SED methods more inline with current methodologies by building the design algorithm around AVO crossploting, a method commonly used in AVO interpretations. A distinct advantage of this is to show that SED algorithms can be incorporated into industry-standard methodologies. A complexity added to the design algorithm in

Chapter 6 is to incorporate offset-dependent errors; design-dependent data errors is a feature that has been absent from SED methods in the Geosciences. With the addition of offset-dependent errors and hence practically-applicable SED algorithms becoming more sophisticated, again it is hoped that the general acceptance of SED methods will increase.

One of the main conclusions of the work in Chapter 6 was that there exists a single design that is optimal when designing AVO surveys with more than 250 receivers over a 30° angle range. The optimal design, however, does not provide the same large expected information gains as seen in the processing designs presented in Chapter 5, because the upper limit of 30° in incidence angle is not sufficient to reach the critical angle (around which most information is to be gained). The additional costs associated with redesigning equipment are deemed to outweigh the additional benefits, and therefore current data acquisition methods are in fact optimal. This situation is reversed if the cost function is inverted (that is, redesigning the data collection system can make acquisition cheaper). Such a case is explored below.

While this work does not advocate a change in design of conventional surveys for AVO analysis, the algorithms developed in this thesis have been designed to be generally applicable to many SED problems. One of the objectives of this work was to raise awareness of the usefulness of SED methods in the Geosciences. I now expand on three areas that I think are ideally suited to SED design methods, and which should be the focus of future research.

7.1 Future Work

7.1.1 Wireless Land Acquisition

As concluded in Chapter 6 one of the main reasons that constant receiver spacing designs are currently optimal is because changing the receiver spacing, particularly for marine seismics has a high associated, positive cost function. However, in 2003 cableless land-based acquisition systems were introduced for seismic surveys (Williams and Hoenmans, 2006). An ideal cableless system consists of single receivers remotely connected to a central recording station. By removing the

need for cables, several design constraints are removed, namely the need to place sensors in gridded geometries with set separations (Hollis et al., 2005). Instead surveys can be designed for specific subsurface objectives and in principal can be tailored to any acquisition scenario.

One such acquisition system, developed by Ion, is called FireFly. The first commercial test of Firefly was performed in 2008 in northwestern Colorado (Williams and Drake, 2009). The survey consisted of over 10,500 receiver points and 7,000 shot points. One of largest costs associated with cableless designs is the transmission of the data from the individual receivers to the main recording station. However, one of the benefits of a cableless system is that every receiver can be remotely switched on and off for every shot. For every receiver that is turned off there is an associated cost benefit, since the amount of data that needs to be recorded, transmitted, and subsequently sorted and stored is reduced. Therefore the positive cost function associated with optimising AVO designs is replaced in this scenario by a negative cost function where by optimising the design the overall cost is reduced.

Utilising the iteratively-constructive design algorithm and the upscaling using the receiver density methods introduced in Chapters 4 and 5, respectively, are ideally suited to designing surveys for use with FireFly and other cableless design systems. For each shot location, an information "map" can be produced showing which receivers are expected to contain the most information about a subsurface target given prior knowledge. The information map can then be used to turn on and off as many receivers as is required to meet specific objectives such as a bounded total overall cost, receiver numbers and densities required for noise removal algorithms, etc.

7.1.2 OBS Survey Design

In standard marine seismic surveys, the seismic source and streamers of receivers are towed behind a vessel near the sea surface. In ocean bottom seismics (OBS), the receivers are located on the seafloor. As in the cableless systems discussed above, the ocean bottom receivers can be positioned as individual nodes or, more akin to conventional methods, the receivers can be connected via a cable. By placing receivers at the sea floor more information about how the seismic en-

ergy travels through the subsurface can be obtained since shear waves can also be recorded. Logistically, OBS surveys are more complex to perform and hence involve larger costs, however the extra information in terms of imaging and geological knowledge is significant (de Kok, 2006).

OBS surveys can range in scope from using over 800 receivers (e.g. Hays et al., 2008) down to as few as 6 receivers (e.g. Eguchi et al., 1998). In all cases however, since the receiver locations are fixed for multiple shots (unlike a streamer being towed), and the deployment costs are relatively high, it is imperative that the receiver are positioned in optimal locations. In cases where less than 10 receivers are to be located the iteratively-constructive method developed in Chapter 4 is an ideal algorithm to use to locate the optimal receiver locations. When more than 10 receivers are to be located the optimal positions can be found using the algorithm developed in Chapter 5.

For all survey scales, prior knowledge can be incorporated from the subsurface geology, but also from the ocean-floor topography so that locations that are easier to deploy a receiver to, and therefore have a lower operational cost, are favoured. Since the geometry between the source and receivers is less well defined compared to standard seismic surveys, SED methods can also be used to optimally locate the seismic source positions to best constrain subsurface properties of interest.

7.1.3 Wide Azimuth Survey Design

Conventional seismics, as used in this thesis, has the source located in-line with the streamer cable resulting in a very narrow range of available azimuths between the sources and receivers. Allowing a wide variation in azimuths between the source and receivers has been shown to improve subsalt illumination, signal-to-noise ratio, and attenuation of multiples, when compared to conventional narrow-azimuth surveys (Regone, 2006).

Although there are several variations of wide azimuth acquisition, most of them involve multiple source boats and shooting passes to achieve the desired offset-azimuth coverage. As a result, the cost of high density, wide azimuth acquisition is several times that of narrow azimuth acquisition (Ting and Zhao, 2009).

To reduce the cost, it is becoming common practice to perform synthetic data modeling to study the effect of shot density reduction schemes prior to acquisition (Ting and Zhao, 2009). Incorporating the methods presented in this thesis into the synthetic modelling stage would be an ideal way to assess the effects that receiver density has on the amount of information expected to be recorded. The extra cost of adding the SED algorithms would be a minimum since data modelling is already being performed. By extension the SED algorithms can then be used to locate the optimal source positions.

A further problem of having multiple vessels is that data that was expected to be recorded may not be recorded because of streamer feathering problems, airgun problems, misalignment of the source and receiver boats, etc. Currently it is impossible to assess how much extra information the missed data contains about the subsurface parameters. Using the nonlinear SED methods presented, multiple scenarios can be modelled prior to the survey being conducted to assess how important a missed section of data is given the recorded data. A decision can then be made in realtime as to whether it is cost effective to go back and collect any specific missing data.

Chapter 8

Conclusions

This thesis has successfully shown that nonlinear SED methods are now sophisticated and efficient enough to be used to plan industrial-scale data acquisition and processing designs.

An iteratively-constructive design method has been introduced which is applicable to nonlinear design problems. The method is based around a Bayesian framework and is ideally suited to scenarios where classical linearised design methods fail. Although the novel algorithm is only guaranteed to produce locally-optimal designs it is shown for both a highly nonlinear sawtooth function and an AVO design problem that the results closely resemble the globally-optimal results in low dimensional cases where the latter designs can be calculated for a fraction of the computational cost. In order to make the algorithm robust and as efficient as possible a separate design algorithm has been implemented so that the computational demand is reduced to a minimum. The algorithm is capable of designing experiments with up to ten independent design dimensions on a standard desktop computer.

Using the iteratively-constructive method in conjunction with a semi-empirical petrophysical model it has been shown that AVA trace selection processing designs can be produced for reservoirs with differing saturating fluids, porosity ranges or any other pertinent prior parameter range. By utilising the complete Zoeppritz equations, designs can be produced for incident angles up to 70 degrees. The results show that the prior uncertainty on the reservoir porosity range has the

largest influence on the trace selection designs than has the prior knowledge about the saturating fluid properties. By building the algorithm around optimal angles rather than offsets, designs can be produced for complex subsurfaces. Although selecting ten traces does not constitute a industrial-scale processing design the results can be upscaled to produce receiver density profiles of unlimited receivers. The results show that a processing design could be expected to return information gains of over 30% when compared to a standard design.

Utilising the methods of reduced-parameterisation a survey design algorithm capable of producing industrial-scale designs has been implemented which allows for design-dependent errors, a feature previously omitted from nonlinear design algorithms. The design algorithm is based around the notion of AVO crossplotting and therefore links the survey design stage much more closely with the data processing stage. Using the reduced-parameterisation method allows designs to be produced which consist of any number of receivers. By using the crossplotting method the maximum incident angle allowed is limited to 30 degrees. As a consequence, once a particular threshold on the total number of receivers has been passed there exists a "one-size-fits-all" design that is optimal for any porosity, fluid content or number of receivers. Although a reduced-parameterisation and simplified forward function is used the results are comparable with designs found using the iteratively-constructive method using the Zoeppritz equations. The extra information provided compared to standard designs is around 5% and therefore it concluded that once extra costs incurred are accounted for then standard equally-spaced receiver designs are in fact optimal.

Although one of the conclusions is that current marine seismic designs are already optimal, three cases have been highlighted that would benefit from the new algorithms presented. It is hoped that as more areas of the Geosciences begin to implement SED methods the amount of research into developing new algorithms will increase, and that SED methods will become commonplace in design and processing applications.

Bibliography

J. Ajo-Franklin. Optimal experiment design for time-lapse traveltime tomography. *Geophysics*, 74(4):Q27–Q40, 2009.

K. Aki and P. Richards. *Quantitative seismology*. University Science Books, 1980.

A. Atkinson and R. Bailey. One hundred years of the design of experiments on and off the pages of Biometrika. *Biometrika*, 88:53–97, 2001.

A. Atkinson and A. Donev. *Optimum experimental designs*. Oxford Science Publications, 1992.

Y. Bard. *Nonlinear parameter estimation*. Academic Press Inc., 1974.

Y. Bartal, Z. Somer, G. Leonard, D. Steinberg, and Y. Ben Horin. Optimal seismic networks in Israel in the context of the comprehensive test ban treaty. *Bulletin of the seismological society of America*, 90(1):151–165, 2000.

N. Barth and C. Wunsch. Oceanographic experiment design by simulated annealing. *Journal of Physical Oceanography*, 20(9):1249–1263, 1990.

L. Belue, K. Bauer, and D. Ruck. Selecting optimal experiments for multiple output multilayer perceptrons. *Neural Computation*, 9(1):161–183, 1997.

L. Berliner, Z. Lu, and C. Snyder. Statistical design for adaptive weather observations. *Journal of the Atmospheric Sciences*, 56:2536–2552, 1999.

J. Blanco, S. Knudsen, and F. Bostick. Time-lapse VSP field test for gas reservoir monitoring using permanent fibre optic seismic system. *SEG Expanded Abstracts*, pages 3447–3451, 2006.

R. Bortfeld. Approximation to the reflection and transmission coefficients of plane longitudinal and transverse waves. *Geophysical Prospecting*, 9:485–503, 1961.

G. Box and N. Draper. A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54(287):622–654, 1959.

G. Box and W. Hunter. Sequential design of experiments for non linear models. *Proc. IBM Sc. Comp. Symp. 1963*, 54:622–654, 1965a.

G. Box and W. Hunter. The experimental study of physical mechanisms. *Technometrics*, 7:23–42, 1965b.

G. Box and H. Lucas. Design of experiments in nonlinear situations. *Biometrika*, 46:77–90, 1959.

G. Box and K. Wilson. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society B*, 13:1–45, 1951.

A. Buland, O. Kolbjørnsen, R. Hauge, Ø. Skjæveland, and K. Duffaut. Bayesian lithology and fluid prediction from seismic prestack data. *Geophysics*, 73:C13–C21, 2008.

G. Cambois. Can P-wave AVO be quantitative? *The Leading Edge*, 19:1246–1251, 2000.

J. Carcione, H. Helle, N. Pham, and T. Toverud. Pore pressure estimation in reservoir rocks from seismic reflection data. *Geophysics*, 68(5):1569–1579, 2003.

J. Castagna and S. Smith. Comparison of AVO indicators: A modeling study. *Geophysics*, 59(12):1849–1855, 1994.

J. Castagna and H. Swan. Principles of AVO crossplotting. *The Leading Edge*, 16(4):337–344, 1997.

J. Castagna, H. Swan, and D. Foster. Framework for AVO gradient and intercept interpretation. *Geophysics*, 63(3):948–956, 1998.

K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

J. Chen and T. Dickens. Effects of uncertainty in rock-physics models on reservoir parameter estimation using seismic amplitude variation with angle and controlled-source electromagnetics data. *Geophysical Prospecting*, 57(1):61–74, 2009.

E. Cherkaeva and A. Tripp. Optimal survey design using focused resistivity arrays. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2):358–366, 1996.

V. Clark. The effect of oil under in-situ conditions on the seismic properties of rocks. *Geophysics*, 57(7):894–901, 1992.

D. Coles and F. Morgan. A method of fast, sequential experimental design for linearized geophysical inverse problems. *Geophysical Journal International*, 178 (1):145–158, 2009.

A. Curtis. Optimal experiment design: cross-borehole tomographic examples. *Geophysical Journal International*, 136:637–650, 1999a.

A. Curtis. Optimal design of focused experiments and surveys. *Geophysical Journal International*, 139:205–215, 1999b.

A. Curtis. Theory of model-based geophysical survey and experimental design: Part 1-linear problems. *The Leading Edge*, 23(10):997–1004, 2004a.

A. Curtis. Theory of model-based geophysical survey and experimental design: Part 2-nonlinear problems. *The Leading Edge*, 23(10):1112–1117, 2004b.

A. Curtis and A. Lomax. Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2):372–378, 2001.

A. Curtis and H. Maurer. Optimizing the design of geophysical experiments: Is it worthwhile? *The Leading Edge*, 19(10):1058–1062, 2000.

A. Curtis and R. Snieder. Reconditioning inverse problems using the genetic algorithm and revised paramterization. *Geophysics*, 62(4):1524–1532, 1997.

A. Curtis, A. Michelini, D. Leslie, and A. Lomax. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys. *Geophysical Journal International*, 157:595–606, 2004.

T. Daley, R. Solbau, J. Ajo-Franklin, and S. Benson. Continuous active-source seismic monitoring of CO_2 injection in a brine aquifer. *Geophysics*, 72:A57–A61, 2007.

R. de Kok. The ocean-bottom recording trade-off. *The Leading Edge*, 25(8): 928–933, 2006.

J. Delforge, A. Syrota, and B. Mazoyer. Experimental design optimization: Theory and application to estimation of receptor model parameters using dynamic positron emission tomography. *Physics in Medicine and Biology*, 34(4):419–435, 1989.

J. Downton and C. Ursenbach. Linearized amplitude variation with offset (AVO) inversion with supercritical angles. *Geophysics*, 71(5):E49–E55, 2006.

K. Duffaut and M. Landrø. V_p/V_s ratio differential stress and rock consolidation - a comparison between rock models and time-lapse AVO data. *Geophysics*, 72 (5):C81–C94, 2007.

A. Duijndam. Bayesian estimation in seismic inversion. part i: Principles. *Geophysical Prospecting*, 36:878–898, 1988.

T. Eguchi, Y. Fujinawa, E. Fujita, S. Iwasaki, I. Watabe, and H. Fujiwara. A real-time observation network of ocean-bottom-seismometers deployed at the Sagami trough subduction zone, central Japan. *Marine Geophysical Researches*, 20:73–94, 1998.

J. Eidsvik, P. Avseth, H. Omre, T. Mukerji, and G. Mavko. Stochastic reservoir characterization using prestack seismic data. *Geophysics*, 69:978–993, 2004.

J. Fatti, G. Smith, P. Vail, J. Strauss, and P. Levitt. Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the geostack technique. *Geophysics*, 59(9):1362–1376, 1994.

V. Fedorov. *Theory of optimal experiments*. Academic Press Inc., 1972.

R. Fisher. *Design of experiments*. Macmillan, 1935.

I. Ford, D. Titterington, and C. Kitsos. Recent advantages in nonlinear experimental design. *Technometrics*, 31(1):49–51, 1989.

G. Froment. Model discrimination and parameter estimation in heterogeneous catalysis. *AIChE Journal*, 21(6):1041–1057, 1975.

A. Furman, T. Ferré, and A. Warrick. A sensitive analysis of electrical resistivity tomography array types using analytical element modeling. *Vadose Zone Journal*, 2:416–423, 2003.

A. Furman, T. Ferré, and A. Warrick. Optimization of ERT surveys for monitoring transient hydrological events using perturbation sensitivity and genetic algorithms. *Vadose Zone Journal*, 3(4):1230–1239, 2004.

A. Furman, T. Ferré, and G. Heath. Spatial focusing of electrical resistivity surveys considering geologic and hydrologic layering. *Geophysics*, 72(2):F65–F73, 2007.

M. García-Fernandez, A. Kijko, J. Carracedo, and V. Soler. Optimum station distribution to monitor seismic activity of Teide volcano, Tenerife, Canary Islands. *Journal of Volcanology and Geothermal Research*, 35(3):195–204, 1988.

H. Ghalib, D. Russell, and A. Kijko. Optimal design of a regional seismological network for the Arab countries. *Pure and Applied Geophysics*, 122(5):694–712, 1984.

I. Goldberg and B. Gurevich. A semi-empirical velocity-porosity-clay model for petrophysical interpretation of P- and S-velocities. *Geophysical Prospecting*, 46:271–285, 1998.

D. Griffiths, J. Turnbull, and A. Olayinka. Two-dimensional resistivity mapping with a computer controlled array. *First Break*, 8(4):121–129, 1990.

D. Gubbins. *Time series analysis and inverse theory for geophysicists*. Cambridge University Press, 2004.

T. Guest and A. Curtis. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships. *Journal of Geophysical Research*, 114:B04307, 2009.

T. Guest and A. Curtis. Optimal trace selection for AVA processing of shale-sand reservoirs. *Geophysics*, 75(4):C37–C47, 2010a.

T. Guest and A. Curtis. On standard and optimal designs of industrial-scale 2D seismic surveys. *Geophysical Journal International*, 2010b.

M. Hamada, H. Martz, C. Reese, and A. Wilson. Finding near-optimal Bayesian experimental designs via genetic algorithms. *The American Statistician*, 55(3): 175–181, 2001.

D. Hays, K. Craft, P. Docherty, and F. Smit. An ocean bottom seismic node repeatability study. *SEG Technical Program Expanded Abstracts*, 27:55–59, 2008.

J. Hendrickson. Staked. *Geophysical Prospecting*, 47:663–705, 1999.

J. Holland. *Adaptation in natural and artificial systems*. University of Michigan Press, 1975.

J. Hollis, J. Iseli, M. Williams, and S. Hoenmans. The future of land seimic. Technical report, Hart Energy Publishing, November 2005.

R. Houck. Estimating uncertainty in interpreting seismic indicators. *The Leading Edge*, 18:320–325, 1999.

R. Houck. Quantifying the uncertainty in an AVO interpretation. *Geophysics*, 67 (1):117–125, 2002.

I. Hu. On sequential designs in nonlinear problems. *Biometrika*, 85(2):496–503, 1998.

S. Jin, G. Cambois, and C. Vuillermoz. Shear-wave velocity and density estimation from PS-wave AVO analysis: Application to an OBS dataset from the North Sea. *Geophysics*, 65(5):1446–1454, 2000.

A. Kijko. An algorithm for the optimal distribution of a regional seismic network - 1. *Pure and Applied Geophysics*, 115(4):999–1009, 1977a.

A. Kijko. An algorithm for the optimum distribution of a regional seismic network - 2. An analysis of the accuracy of location of local earthquakes depending on the number of seismic stations. *Pure and Applied Geophysics*, 115(4):1011–1021, 1977b.

A. Kijko and M. Sciocatti. Optimal spatial-distribution of seismic stations in mines. *International journal of rock mechanics and mining sciences & geomechanical abstracts*, 32(6):607–615, 1995.

A. Larsen, M. Ulvmoen, H. Omre, and A. Buland. Bayesian lithology/fluid prediction and simulation on the basis of a markov-chain model. *Geophysics*, 71: R69–R78, 2006.

D Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

Y. Liu and D. Schmitt. Amplitude and AVO responses of a single thin bed. *Geophysics*, 68(4):1161–1168, 2003.

M. Loke and R. Barker. Practical techniques for 3D resistivity surveys and data inversion. *Geophysical Prospecting*, 44:499–523, 1996a.

M. Loke and R. Barker. Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method. *Geophysical Prospecting*, 44:131–152, 1996b.

C. Macdonald, P. Davis, and D. Jackson. Inversion of reflection traveltimes and amplitudes. *Geophysics*, 52(5):606–617, 1987.

D. Marion, A. Nur, H. Yin, and D. Han. Compressional velocity and porosity in sand-clay mixture. *Geophysics*, 57(4):554–563, 1992.

M. Matsu'ura and N. Hirata. Generalized least-squares solutions to quasi-linear inverse problems with a priori information. *Journal of Physics of the Earth*, 30 (6):451–468, 1982.

H. Maurer and D. Boerner. Optimized design of geophysical experiments. *The Leading Edge*, 17(8):1119–1125, 1998a.

H. Maurer and D. Boerner. Optimized and robust experimental design: A non-linear application to EM sounding. *Geophysical Journal International*, 132: 458–468, 1998b.

H. Maurer, D. Boerner, and A. Curtis. Design strategies for electromagnetic geophysical surveys. *Inverse Problems*, 16:1097–1117, 2000.

H. Maurer, A. Curtis, and D. Boerner. Geophysical survey design - A quest for information. *Geophysics*, 2010.

G. Mavko, T Mukerji, and J Dvorkin. *The rock physics handbook*. Cambridge University Press, 1998.

W. Menke. *Geophysical data analysis: Discrete inverse theory*, volume 45 of *International Geophysics Series*. Academic Press Inc., Harcourt Brace Jovanovich Publishers, San Diego, 1989.

T. Mitchell. An algorithm for the construction of "D-Optimal" experimental designs. *Technometrics*, 16(2):203–210, 1974.

S. Mous. Identification of the movement of water in unsaturated soils - The problem of identifiability of the model. *Journal of Hydrology*, 143:153–167, 1993.

T. Mukerji, A. Jørstad, G. Mavko, and J. Granli. Applying statistical rock physics and seismic inversions to map lithofacies and pore fluid probabilities in a North Sea reservoir. *SEG Technical Program Expanded Abstracts*, 17:894–897, 1998.

P. Muller and G. Parmigiani. Numerical evaluation of information-theoretic measures. In D. Berry, K. Chaloner, and J. Geweke, editors, *Bayesian analysis in statistics and economics: Essays in honour of Arnold Zellner*, pages 397–406. Wiley, 1996.

W. Muller and D. Zimmerman. Optimal designs for variogram estimation. *Environmetrics*, 10(1):23–37, 1999.

M. Noel and B. Xu. Archaeological investigation by electrical resistivity tomography: A preliminary study. *Geophys. J. Int.*, 107:95–102, 1991.

J. Nyquist, J. Peake, and M. Roth. Comparison of an optimized resistivity array with dipole-dipole soundings in karst terrain. *Geophysics*, 72(4):F139–F144, 2007.

W. Ostrander. Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence. *Geophysics*, 49(10):1637–1648, 1984.

N. Rabinowitz and D. Steinberg. Optimal configuration of a seismographic network: A statistical approach. *Bulletin of the Seismological Society of America*, 80(1):187–196, 1990.

N. Rabinowitz and D. Steinberg. A statistical outlook on the problem of seismic network configuration. In C. Thurber and N. Rabinowitz, editors, *Advances in seismic event location*, volume 18 of *Modern Approaches in Geophysics*, chapter 3. Kluwer Academic Publishers, 2000.

C. Regone. Using 3D finite-difference modeling to design wide azimuth surveys for improved subsalt imaging. *SEG Technical Program Expanded Abstracts*, 25: 2896–2900, 2006.

J. Reynolds. *An introduction to applied and environmental geophysics*. Wiley, 1997.

B. Russell, K. Hedlin, F. Hilterman, and L. Lines. Fluid-property discrimination with AVO: A Biot-Gassmann perspective. *Geophysics*, 68(1):29–39, 2003.

S. Rutherford and R. Williams. Amplitude-versus-offset variations in gas sands. *Geophysics*, 54(6):680–688, 1989.

K. Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603, 2003.

P. Sabatier. On geophysical inverse problems and constraints. *J. Geophys.*, 43: 115–137, 1977.

I. Saguy and M. Karel. Modeling of quality deterioration during food-processing and storage. *Food Technology*, 32(2):78–85, 1980.

M. Sambridge and G. Drijkoningen. Genetic algorithms in seismic waveform inversion. *Geophys. J. Int.*, 109:323–342, 1992.

D. Santoso, S. Alam, L. Hendrajaya, Alfian, S. Munadi, and Purwoko. Reservoir fluid identification and elastic modulus determination using AVO. *Journal of Applied Geophysics*, 35:159–165, 1996.

C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, 1948.

M. Shewry and H. Wynn. Maximum entropy sampling. *Journal of Applied Statistics*, 14:165–170, 1987.

R. Shuey. A simplification of the Zoeppritz equations. *Geophysics*, 50(4):609–614, 1985.

R. Simm, R. White, and R. Uden. The anatomy of AVO crossplots. *The Leading Edge*, 19:150–155, 2000.

J. Simmons and M. Backus. Waveform-based AVO inversion and AVO prediction-error. *Geophysics*, 61(6):1575–1588, 1996.

L. Skopintseva and T. Nefedkina. AVO-inversion for postcritical reflections. In *EAGE 68th Conference and exhibition*, EAGE extended abstracts, page P058, 2006.

G. Smith and P. Gidlow. Weighted stacking for rock property estimation and detection of gas. *Geophysical Prospecting*, 35:993–1014, 1987.

D. Steinberg and N. Rabinowitz. Optimal seismic monitoring for event location with application to on site inspection of the comprehensive nuclear test ban treaty. *Metrika*, 58(1):31–57, 2003.

D. Steinberg, N. Rabinowitz, Y. Shimshoni, and D. Mizrachi. Configuring a seismographic network for optimal monitoring of fault lines and multiple sources. *Bulletin of the seismological society of America*, 85(6):1847–1857, 1995.

P. Stummer, H. Maurer, H. Horstmeyer, and A. Green. Optimization of DC resistivity data acquisition: Real-time experimental design and a new multi-electrode system. *IEEE Transactions on Geoscience and Remote Sensing*, 40 (12):2727–2735, 2002.

P. Stummer, H. Maurer, and A. Green. Experimental design: Electrical resistivity data sets that provide optimum subsurface information. *Geophysics*, 69(1):120–139, 2004.

G. Taguchi. *Systems of experimental design (vols 1 and 2, 1967 and 1977, with 1987 translation)*. UNIPUB, 1987.

A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

A. Tarantola and B. Valette. Inverse problems = quest for information. *Journal of Geophysics*, 50:159–170, 1982.

C. Ting and W. Zhao. A simulated wide azimuth simultaneous shooting experiment. *SEG Technical Program Expanded Abstracts*, 28:76–80, 2009.

J. Van den Berg, A. Curtis, and J. Trampert. Optimal nonlinear Bayesian experimental design: An application to amplitude versus offset experiments. *Geophysical Journal International*, 155:411–421, 2003.

J. Van den Berg, A. Curtis, and J. Trampert. Corrigendum. *Geophysical Journal International*, 161:265–265, 2005.

J. Vanmilgen and R. Baumont. Models based on variable fractional digestion rates to describe ruminal in-situ digestion. *British journal of nutrition*, 73(6): 793–807, 1995.

P. Wilkinson, P. Meldrum, J. Chambers, O. Kuras, and R. Ogilvy. Improved strategies for the automatic selection of optimized sets of electrical resistivity tomography measurement configurations. *Geophysical Journal International*, 167:1119–1126, 2006.

M. Williams and K. Drake. Wireless system images complex shall. Technical report, The American Oil & Gas Reporter, July 2009.

M. Williams and S. Hoenmans. Moving towards full-sampling in land 3D acquisition. *First Break*, 24:55–61, 2006.

R. Williams, G. Roberts, and K. Hawkins. Long offset towed streamer recording - A cheaper alternative to multi-component OBC for exploration? *Exploration Geophysics*, 32:316–319, 2001.

E. Winterfors and A. Curtis. Numerical detection and reduction of non-uniqueness in nonlinear inverse problems. *Inverse Problems*, 24(2):025016, 2008.

E. Winterfors and A. Curtis. A bifocal measure of expected ambiguity in Bayesian nonlinear parameter estimation. *J. Amer. Statist. Assoc., sub judice*, 2010.

Öz Yilmaz. *Seismic data analysis*. Society of Exploration Geophysicists, 2001.

K. Zoeppritz. On the reflection and propagation of seismic waves. *Göttinger Nachrichten*, 1:66–84, 1919.