FISHER INFORMATION DECISION DIRECTED QUANTISATION: A Fast Sub-Optimal Solution to Discrete Optimisation Problems with Applications in Wireless Communications [1]

Jonathan H. Manton and Ba-Ngu Vo

Department of Electrical and Electronic Engineering
The University of Melbourne, Victoria 3010, Australia.
j.manton@ee.unimelb.edu.au

Abstract

Discrete optimisation problems arise naturally in wireless communications because the source symbols come from a finite alphabet. However, solving these discrete optimisation problems optimally is often computationally too expensive to be implemented in practice. This paper presents a fast sub-optimal algorithm, named "Fisher Information Decision Directed Quantisation", for solving a broad class of discrete optimisation problems. This algorithm is applied successfully to three problems of increasing complexity: integer linear regression, source symbol estimation in precoded transmissions and joint source and channel estimation in precoded transmissions.

1 Introduction

Optimally estimating the source symbols in a wireless communications system is an example of a discrete optimisation problem whose solution often cannot be computed in practice due to the high computational cost involved. It is therefore important to devise sub-optimal solutions which strike a satisfactory balance between accuracy and computational cost. This paper proposes a novel method for finding a sub-optimal solution to a wide class of discrete optimisation problems in signal processing. The method is called "Fisher Information Decision Directed Quantisation" because it uses the Fisher Information Matrix[2] to determine the best sequence in which to estimate the discrete parameters.

The idea behind the proposed method is now explained by way of example. Consider the integer linear regression problem of estimating the integer valued vector $\boldsymbol{s}$ given the observed vector $\boldsymbol{y}$ defined by $\boldsymbol{y} = A\boldsymbol{s} + \boldsymbol{n}$ where A is a (square or tall) real valued matrix and $\boldsymbol{n}$ denotes additive noise (a random variable). The optimal (least-squares) solution is to find the integer valued vector $\widehat{\boldsymbol{s}}$ which minimises $\|\boldsymbol{y} - A\widehat{\boldsymbol{s}}\|^2$. This can be computed using the algorithm in [1] and, in the context of communication systems, is referred to as sphere decoding. A very simple sub-optimal solution, but one often used in practice, is to compute the optimal solution $\widehat{\boldsymbol{s}} = (A'A)^{-1} A'\boldsymbol{y}$ of the non-integer-constrained linear regression problem and then quantise the elements of $\widehat{\boldsymbol{s}}$, thus forcing them to be integer valued. (Throughout, a prime is used to denote matrix transpose or complex-conjugate transpose, whichever is appropriate.)

[1] This work was supported by the Australian Research Council, the Special Research Centre for Ultra-Broadband Information Networks (CUBIN) and the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP). Presented at *International Conference on Optimisation Theory and Applications, Hong Kong, December 2001.*

[2] However, the method extends to deterministic problems for which the FIM is not defined; see Section 2.3.

The novel Fisher Information Decision Directed Quantisation method lies between these two extremes; although it is not optimal, it has a low computational complexity and it often produces significantly more accurate estimates than the popular per-symbol quantisation method. The basic idea behind the method is as follows. Consider the non-integer-constrained solution

$$\begin{aligned} \widehat{\boldsymbol{s}} &= \left(A'A\right)^{-1} A'\boldsymbol{y} \\ &= \boldsymbol{s} + \left(A'A\right)^{-1} A'\boldsymbol{n}. \end{aligned} \tag{1}$$

Due to the factor $\left(A'A\right)^{-1} A'$, certain elements of the error vector $\left(A'A\right)^{-1} A'\boldsymbol{n}$ are likely to be larger, on average, than other elements. Indeed, if $\boldsymbol{E}[\boldsymbol{n}] = 0$ and $\boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}'] = I$ (the identity matrix) then the error covariance matrix is

$$\boldsymbol{E}\left[\left(\left(A'A\right)^{-1} A'\boldsymbol{n}\right)\left(\left(A'A\right)^{-1} A'\boldsymbol{n}\right)'\right] = \left(A'A\right)^{-1} \tag{2}$$

showing that if the ith diagonal element of $\left(A'A\right)^{-1}$ is the smallest diagonal element then the ith element of $\boldsymbol{s}$ is the one which can be estimated most accurately[3]; note that the Fisher Information Matrix for this estimation problem is $A'A$. It is therefore proposed to solve initially the optimisation problem $\min_{\widehat{\boldsymbol{s}}} \|\boldsymbol{y} - A\widehat{\boldsymbol{s}}\|^2$ subject to the constraint that the ith element of $\widehat{\boldsymbol{s}}$ is integer valued but all other elements are real valued. The ith element of $\widehat{\boldsymbol{s}}$ found in this way is fixed and the process repeated for the remaining elements.

The reason why the method is called Fisher Information Decision Directed Quantisation is because the (inverse of the) Fisher Information Matrix is used to decide which element to quantise next, and moreover, the quantisation rule used at each stage depends on previous quantisation decisions. Indeed, it is shown in the next section that previous correct decisions improve the chances of the remaining elements being correctly quantised.

The following section states formally the Fisher Information Decision Directed Quantisation method and various extensions of the method. It also studies further the integer linear regression problem. Section 3 introduces a realistic non-linear estimation problem in wireless communications and demonstrates that the Fisher Information Decision Directed Quantisation method can be incorporated with good effect in practical systems. Section 4 concludes the paper.

2 Fisher Information Decision Directed Quantisation

This section states formally the proposed Fisher Information Decision Directed Quantisation method for solving discrete optimisation problems. A straightforward argument is given explaining why the method appears to perform well in practice. Several extensions of the method are also proposed.

2.1 The Proposed Method

Assume that the output $\boldsymbol{y}$ of a stochastic system is a function of a discrete valued vector $\boldsymbol{s}$, a continuous valued vector $\boldsymbol{h}$ and a continuous valued random noise vector $\boldsymbol{n}$, that is, $\boldsymbol{y} =$

[3] The expected value of $\left|\left(\left(A'A\right)^{-1} A'\boldsymbol{n}\right)_i\right|^2$ is given by the ith diagonal element of $\left(A'A\right)^{-1}$.

$f(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{n})$. For example, choosing $f(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{n}) = A\boldsymbol{s}+\boldsymbol{n}$ corresponds to the integer linear regression problem considered in Section 1. (For the precoded transmission problem considered in Section 3, the extra term $\boldsymbol{h}$ is required.) It is required to estimate $\boldsymbol{s}$ and $\boldsymbol{h}$ given $\boldsymbol{y}$.

It is also assumed that there is a cost function $g(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}}; \boldsymbol{y})$ which can be evaluated for continuous values of $\widehat{\boldsymbol{s}}$ and which achieves its minimum when $(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}})$ is an "optimal" estimate of $(\boldsymbol{s}, \boldsymbol{h})$. One such choice is the negative of the log likelihood function. For the aforementioned integer linear regression problem, $g(\widehat{\boldsymbol{s}}; \boldsymbol{y}) = \|\boldsymbol{y} - A\widehat{\boldsymbol{s}}\|^2$ for instance.

The only remaining ingredient is a measure of how accurately the estimator (obtained by minimising g) can estimate each element of $\boldsymbol{s}$. If the estimator is Fisher efficient then the inverse of the Fisher Information matrix [3], known as the Cramer-Rao Bound, provides a good measure of how accurately the elements of $\boldsymbol{s}$ can be estimated (see Section 1 for an example). Note that the Fisher Information matrix is a function of the true parameter vectors $\boldsymbol{s}$ and $\boldsymbol{h}$ in general.

The proposed Fisher Information Decision Directed Quantisation algorithm works as follows. With $\widehat{\boldsymbol{s}}$ allowed to take on continuous values, find the minimum $(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}})$ of the cost function $g(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}}; \boldsymbol{y})$, which serves as a first guess of $\boldsymbol{s}$ and $\boldsymbol{h}$. Using the Fisher Information matrix evaluated at $(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}})$ or otherwise, determine which element of $\widehat{\boldsymbol{s}}$ is likely to be the most accurate. Then, with this element of $\widehat{\boldsymbol{s}}$ constrained to take only discrete values but all other elements unconstrained, find the minimum[4] $(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}})$ of the cost function $g(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}}; \boldsymbol{y})$. (Note that here, the optimisation is over both $\widehat{\boldsymbol{s}}$ and $\widehat{\boldsymbol{h}}$. Therefore, at least in concept, the estimate $\widehat{\boldsymbol{h}}$ of $\boldsymbol{h}$ improves with each iteration due to the exploitation of the prior knowledge of the discrete nature of $\boldsymbol{s}$.) The constrained element of $\widehat{\boldsymbol{s}}$ is now locked into place and the above process repeated; first evaluate the Fisher Information matrix at the new point $(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}})$ so as to determine which remaining element of $\widehat{\boldsymbol{s}}$ to quantise next, then minimise $g(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}}; \boldsymbol{y})$ with the element of $\widehat{\boldsymbol{s}}$ to be quantised constrained to take discrete values only, with one or more elements of $\widehat{\boldsymbol{s}}$ fixed from previous stages, and with all other elements of $\widehat{\boldsymbol{s}}$ unconstrained.

Complexity: If $\boldsymbol{s}$ has n elements, and if each element can take one of m possible values, then the above method requires the cost function g to be minimised at most[5] nm times.

2.2 Justification of Method

In order to provide insight into how the method works, the integer linear regression problem considered in Section 1 is revisited. Since the cost function is quadratic, all the required minimisations are trivial to perform. Indeed, if just one element of $\widehat{\boldsymbol{s}}$ is constrained to be integer valued then it is readily proved that the minimum can be found by first computing the minimum without the constraint and then quantising (rounding to the nearest integer value) the element of $\widehat{\boldsymbol{s}}$ to be constrained.

Figures 1 and 2 help put the performance of the Fisher Information Decision Directed Quantisation method into perspective. Figure 1 shows the quantisation regions obtained for a discrete linear regression problem with two regressors, each of which can take the values -1 or 1 only. The ellipses represent equiprobable error distributions assuming the noise to be white Gaussian; recall that the error is given by the second term in (1). The centre subgraph shows the optimal

[4] If there are only a finite number of discrete values then this minimisation can always be achieved by exhaustive search; set the constrained element of $\widehat{\boldsymbol{s}}$ to an admissible value and minimise g with respect to the unconstrained elements, then repeat for all other discrete values.

[5] It is shown later that only n minimisations are required for the integer linear regression problem.

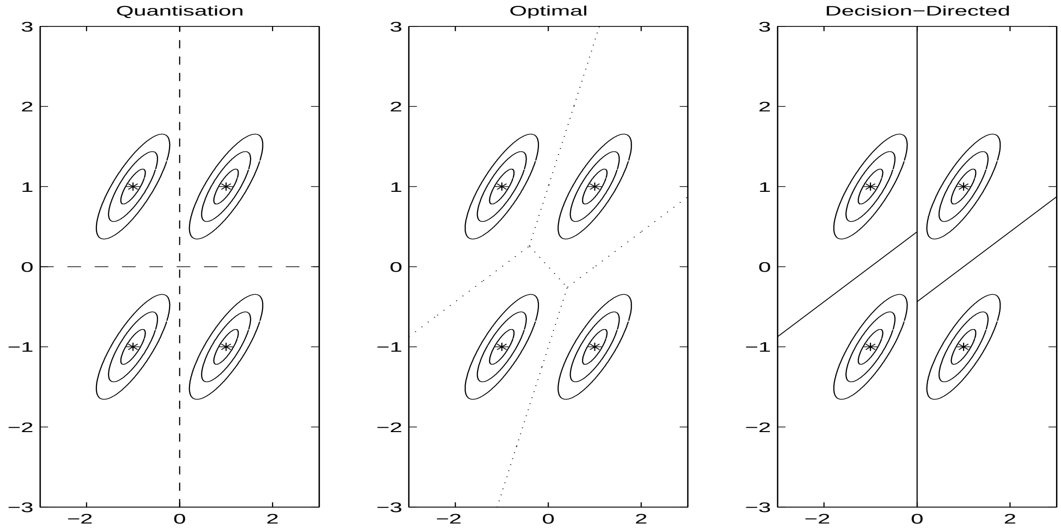

Figure 1: Decision regions for per-symbol quantisation, optimal quantisation and Fisher Information Decision Directed quantisation.

quantisation regions obtained by minimising the quadratic cost function. The left subgraph shows the standard per-symbol quantisation regions which match up poorly with the optimal regions. The Fisher Information Decision Directed quantisation regions, shown on the right, can be explained as follows. Since the error ellipses are taller than they are wider, the first regressor (graphed along the horizontal axis) can be estimated more accurately than the second regressor (graphed along the vertical axis). Therefore, the first regressor is the one to be quantised first. Its two quantisation regions are separated by the solid vertical line. The quantisation rule for the second regressor depends on the decision made about the first regressor; this is seen from the two different solid lines, one on the left and one on the right. Indeed, comparing the Fisher Information Decision Directed quantisation regions with the optimal regions shows that *if the first regressor is correctly quantised then the second regressor will be quantised optimally.* Conversely, if the first regressor is incorrectly quantised then the second regressor will be quantised poorly. This justifies the ordering, from most accurate to least accurate, in which the regressors are quantised.

Even though earlier incorrect decisions increase the chance of subsequent incorrect decisions, this does not imply that the Fisher Information Decision Directed Quantisation method will not perform well in the presence of noise; if it makes a mistake quantising the most accurate element of $\widehat{\boldsymbol{s}}$ then it can be expected that per-symbol quantisation will make many mistakes since the other elements of $\widehat{\boldsymbol{s}}$ contain even larger errors. This argument is justifed in Figure 2 as well as in further simulation results in the next section. Figure 2 shows that Fisher Information Decision Directed Quantisation performs significantly better than per-symbol quantisation in integer linear regression problems.

2.3 Generalisations

Two generalisations are now discussed briefly. First, it is not necessary to quantise only one element of $\widehat{\boldsymbol{s}}$ at a time. The above method is trivially modified to choose the k most accurate elements and then minimise the cost function with these k elements constrained to take discrete values. As k increases, the accuracy of the method improves[6] but the computational complexity increases too. The second generalisation is if the noise is not random. That is, if it is desired to minimise a cost function $g(\boldsymbol{s},\boldsymbol{h})$ with $\boldsymbol{s}$ constrained to take discrete values then the above method can still be applied by replacing the Fisher Information matrix by the Hessian of $g(\boldsymbol{s},\boldsymbol{h})$ evaluated at the unconstrained minimum $(\widehat{\boldsymbol{s}},\widehat{\boldsymbol{h}})$. Specifically, if $(\widehat{\boldsymbol{s}},\widehat{\boldsymbol{h}})$ is the minimum of $g(\boldsymbol{s},\boldsymbol{h})$ then $g(\boldsymbol{s},\widehat{\boldsymbol{h}}) \approx g(\widehat{\boldsymbol{s}},\widehat{\boldsymbol{h}}) + (\boldsymbol{s}-\widehat{\boldsymbol{s}})'Q(\boldsymbol{s}-\widehat{\boldsymbol{s}})$ where Q is (half of) the Hessian of g and therefore gives an indication of which elements of $\boldsymbol{s}$ lead to the greatest change in the cost. It is interesting to note that for the linear regression problem, the Fisher Information matrix equals the Hessian matrix of the cost.

3 Precoded Transmissions

The Fisher Information Decision Directed Quantisation method is applied in this section to the problem of estimating the source symbols in a precoded transmission system. Simulations demonstrate the improvement over standard per-symbol quantisation.

3.1 Problem Formulation

Using a linear precoder (or a filterbank precoder) to encode the source symbols prior to transmission was first considered in [2] and has received significant attention since; see [10] and the references therein. A simple but important extension is to use an affine precoder [9]. Specifically, the following problem is considered.

Let $\boldsymbol{s} \in \mathbb{C}^p$ denote a vector of source symbols whose elements belong to a finite alphabet. (For example, in a BPSK system the elements are either -1 or 1 whereas in a QAM system the elements belong to the set $\{\pm e^{\jmath\pi/4}, \pm e^{\jmath 3\pi/4}\}$.) These elements are first precoded by an affine precoder [9] to form the transmitted vector $A\boldsymbol{s}+\boldsymbol{b}$ where $A \in \mathbb{C}^{n\times p}$ and $\boldsymbol{b} \in \mathbb{C}^n$ are known. (The purpose of the A matrix is to spread the symbols in the frequency domain so as to avoid channel spectral nulls [7] while the purpose of the vector $\boldsymbol{b}$ is to introduce a training sequence or pilot tones [5] to assist in identifying the channel.) The received symbols are a noise corrupted version of the transmitted vector convolved with the channel impulse response $\boldsymbol{h} = [h_0, \cdots, h_{L-1}] \in \mathbb{C}^L$, which in matrix form is

$$\boldsymbol{y} = H(A\boldsymbol{s} + \boldsymbol{b}) + \boldsymbol{n} \tag{3}$$

where H is the $(n-L+1)\times n$ upper triangular Toeplitz matrix whose first row is equal to $[h_{L-1}, \cdots, h_0, 0, \cdots, 0]$.

Given the received vector $\boldsymbol{y}$, the aim is to estimate the source symbols $\boldsymbol{s}$ and the channel $\boldsymbol{h}$. (Note that this is possible provided there is enough redundancy introduced by the affine

[6]Choosing k equal to the length of $\widehat{\boldsymbol{s}}$ corresponds to optimal quantisation.

precoder; see [8, 6].) Specifically, it is required to minimise the cost function

$$g(\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{h}}; \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{y} - \widehat{H}(A\widehat{\boldsymbol{s}} + \boldsymbol{b})\|^2$$

where $\widehat{H}$ is obtained from $\widehat{\boldsymbol{h}}$ in the same way H is obtained from $\boldsymbol{h}$.

3.2 Source Estimation

If the channel $\boldsymbol{h}$ is known in advance (perhaps by using an alternative identification scheme) then (3) reduces to a discrete linear regression problem. The previous section demonstrated that the Fisher Information Decision Directed Quantisation method is superior to the standard approach of using a linear equaliser[7] followed by per-symbol quantisation.

3.3 Joint Source and Channel Estimation

It is reasonably straightforward to implement the Fisher Information Decision Directed Quantisation method for estimating $\boldsymbol{s}$ and $\boldsymbol{h}$. It can be shown that the inverse of the Fisher Information matrix (that is, the Cramer-Rao Bound) for estimating $\boldsymbol{s}$ is

$$\left(A'H'\left(I - Z\left(Z'Z\right)^{-1}Z'\right)HA\right)^{-1}$$

where

$$Z = [J_0(A\boldsymbol{s} + \boldsymbol{b}), \cdots, J_{L-1}(A\boldsymbol{s} + \boldsymbol{b})]$$

and $J_i = \frac{\partial H}{\partial h_i}$. Note that the Fisher Information Matrix for estimating $\boldsymbol{h}$ is not required since the elements of $\boldsymbol{h}$ are not restricted to being discrete.

Important Note: The minimisation of g cannot be performed reliably by standard minimisation algorithms due to a singularity of the cost function at $\boldsymbol{h} = 0$. However, the novel algorithm in [6] overcomes this problem by minimising an associated cost function on complex projective space.

3.4 Simulations

The performance of the Fisher Information Decision Directed Quantisation method applied to the precoded transmission problem is illustrated in Figure 3. The choice of precoder is such that the training sequence $0, \cdots, 0, 1, 0, \cdots, 0$ is first sent followed by the BPSK source symbols followed by trailing zeros; this is referred to as a Zero Padded system in the literature. (The importance of using either a cyclic prefix or zero padding is explained in [4].) The left subgraph shows the Bit Error Rate (BER) while the right subgraph shows the Packet Error Rate defined as follows. Each erroneously decoded source symbol counts as a bit error while the erroneous decoding of one or more source symbols counts as a single packet error. It is clear from Figure 3 that the Fisher Information Decision Directed Quantisation method significantly outperforms the standard per-symbol quantisation method.

Remark: Even better performance (at the expense of computational complexity) results if the source symbols are quantised in groups of two or more at a time; this will be investigated in subsequent work.

[7] The linear equaliser computes the optimal solution of the unconstrained estimation problem; c.f., Section 1.

4 Conclusion

Discrete optimisation problems arising in wireless communications due to the source symbols coming from a finite alphabet are often computationally prohibitive to solve optimally. Therefore, in practice, it is common to solve first a related continuous optimisation problem and then quantise each element of the result separately. This paper propounds a novel Fisher Information Decision Directed Quantisation method which is computationally inexpensive yet significantly outperforms the standard per-symbol quantisation approach, as evidenced by simulation results.

References

[1] U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Mathematics of Computation*, 44(170):463–471, April 1985.

[2] G. B. Giannakis. Filterbanks for blind channel identification and equalization. *IEEE Signal Processing Letters*, 4(6):184–187, June 1997.

[3] E. L. Lehmann. *Theory of Point Estimation*. John Wiley and Sons, Inc., 1983.

[4] J. H. Manton. Dissecting OFDM: The independent roles of the cyclic prefix and the IDFT operation. *IEEE Communications Letters*, 2001. Accepted.

[5] J. H. Manton. Optimal training sequences and pilot tones for OFDM systems. *IEEE Communications Letters*, 5(4):151–153, April 2001.

[6] J. H. Manton. A packet based channel identification algorithm for wireless multimedia communications. In *IEEE International Conference on Multimedia and Expo*, Tokyo, Japan, August 2001.

[7] J. H. Manton and Y. Hua. A frequency domain deterministic approach to channel identification. *IEEE Signal Processing Letters*, 6(12):323–326, December 1999.

[8] J. H. Manton and Y. Hua. Blind channel identifiability with an arbitrary linear precoder. In *Signal Processing Advances in Wireless and Mobile Communications*, volume 1, chapter 10, pages 339–366. Prentice-Hall, 2000.

[9] J. H. Manton, I. Y. Mareels, and Y. Hua. Affine precoders for reliable communications. In *IEEE Conference on Acoustics, Speech and Signal Processing*, volume V, pages 2749–2752, Istanbul, Turkey, June 2000.

[10] A. Scaglione, G. B. Giannakis, and S. Barbarossa. Redundant filterbank precoders and equalizers: Parts I and II. *IEEE Transactions on Signal Processing*, 47:1988–2022, July 1999.

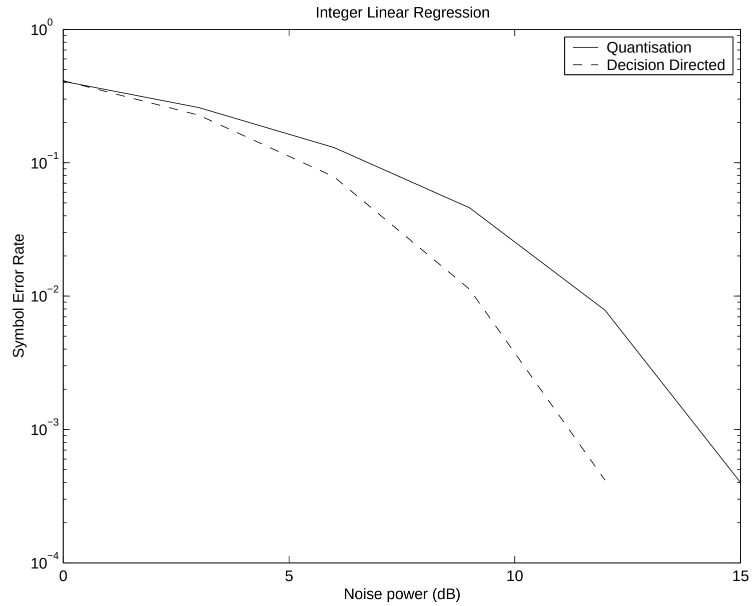

Figure 2: Graph of error rates of per-symbol quantisation and Fisher Information Decision Directed quantisation applied to a representative integer linear regression problem with six regressors. Note that the lower trace disappears because the Fisher Information Decision Directed quantisation algorithm made no estimation errors at 15dB in the simulation.

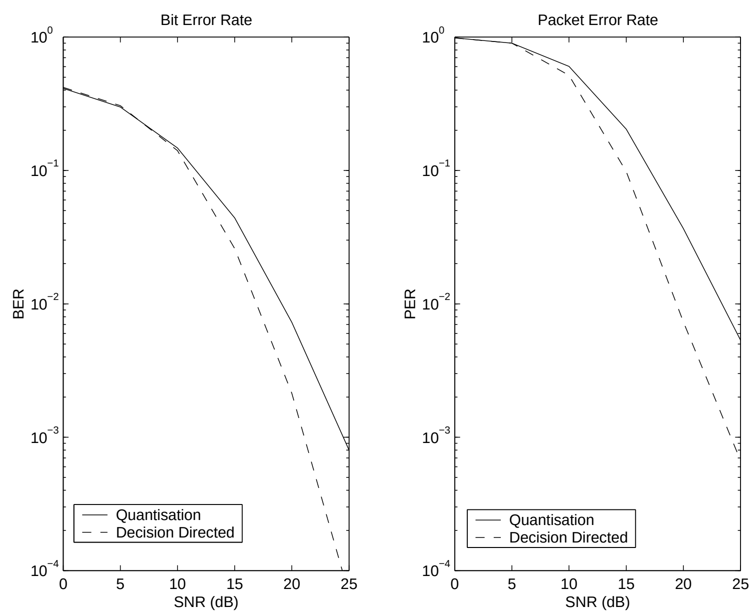

Figure 3: Graph of error rates of per-symbol quantisation and Fisher Information Decision Directed quantisation applied to the precoded transmission problem described in Section 3.