Approximating vector quantisation by transformation and scalar quantisation

Lei Yang[1], Pengwei Hao[2], Dapeng Wu[1]

[1]*Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA*
[2]*Department of Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK*
E-mail: wu@ece.ufl.edu

Abstract: Vector quantisation provides better rate-distortion performance over scalar quantisation even for a random vector with independent dimensions. However, the design and implementation complexity of vector quantisers is much higher than that of scalar quantisers. To reduce the complexity while achieving performance close to optimal vector quantisation or better than scalar quantisation, the authors propose a new quantisation scheme, which consists of transformation and scalar quantisation. The transformation is to decorrelate and raise the dimensionality of the input data, for example, to convert a two-axis representation in two-dimensional into a tri-axis representation; then scalar quantisation is applied to each of the raised dimensions, for example, along three axes. The proposed quantiser is asymptotically optimal/suboptimal for low/high rate quantisation, especially for the quantisation with certain prime number of quantisation levels. The proposed quantiser has $O(N^2)$ design complexity, whereas the design complexity of VQ is $O(N!)$, where N is the number of quantisation levels per dimension. The experimental results show that the average bit-rate achieves 0.4–24.5% lower than restricted/unrestricted polar quantisers and rectangular quantisers for signals of circular and elliptical Gaussian and Laplace distributions. It holds the potential of improving the performance of the existing image and video coding schemes.

1 Introduction

Quantisation is a critical technique for analogue-to-digital conversion and signal compression [1]. Scalar quantisation is simple, fast and easily amenable to a hardware implementation, whereas vector quantisation in high dimension could achieve smaller mean-square error (MSE) and better rate-distortion (R-D) performance, by jointly considering all the dimensions [2, 3], but at the cost of an exponentially increasing quantiser design time and more quantisation computations, that is, at the cost of more codebook design and look-up time.

To reduce the cost, a lot of research has focused on two-dimensional (2D) random variables, especially those of circular Gaussian distributions, since circular distributions have a lot of elegant closed-form expressions [4, 5]. The earliest work referred to Huang and Schultheiss's method [6], which quantises each dimension of random variables in Gaussian distributions with separate 1D Lloyd–Max quantisers [7]. It is efficient and effective, but definitely could be improved. Later, Zador [8] and Gersho [9] studied quantisation by using companders with a large number of quantisation levels theoretically. They used a compressor to transform the data into a uniform distribution, and then applied the optimal quantisers for the uniform distribution, and then transformed the data with an expander. However, this scheme does not work well under a small number of quantisation levels. Another major method for designing quantisers for circular distributions uses polar coordinates,

termed as polar quantisation. Polar quantisation includes separable magnitude quantisation and phase quantisation. Uniform polar quantisation was studied by Moo and Neuho [10] with uniform magnitude and phase quantisation. The optimal ratio between the number of magnitude quantisation levels and the number of phase quantisation levels was studied by Pearlman [11] and Bucklew and Gallagher [12, 13], and an minimum mean square error (MMSE) restricted polar quantiser is implemented by using a uniform quantiser for the phase angles and a scaled Lloyd–Max quantiser of Rayleigh distribution for the magnitude. However, their MMSE scheme does not always consider the centre of a circular distribution as a quantisation level and is restricted by the number of quantisation levels, thus, its MSE performance is sometimes worse than rectangular quantisers and other lattice quantisers, and it does not work well for elliptical distributions neither. Wilson [14] proposes a series of non-continuous quantisation lattices which provide almost the optimal performance among the existing polar quantisation. It is a kind of unrestricted polar quantisation with arc boundaries. Swaszek and Thomas [15] improved Wilson's scheme by replacing arc boundaries with Dirichlet boundaries. He showed the optimal circularly symmetric quantisers for circular Gaussian distributions with a small number of quantisation levels.

Most of the previous work focuses on Gaussian distributions, and provide numerical results only for Gaussian distributions. Gaussian source is considered as the 'worst case' source for data compression. The quantisers for

Guassion distributions are instructive to construct robust quantisers for other distributions [16], but the quantisers for Gaussian distributions are far from the optimal quantisers for other distributions. They did not consider the elliptical distributions neither, whose optimal quantisers are obviously different from those for circular distributions. Also, they did not provide a unified framework for arbitrary distributions. Therefore the optimal quantisers for other distributions such as circular Laplace distributions, elliptical Gaussian and Laplace distributions need investigation under a unified framework.

To address these problems, we propose a unified quantisation system to approach the optimal vector quantisers by using transforms and scalar quantisers. The function of transforms, especially unitary transforms and volume-preserving scaling transforms, on signal entropy and distortion is discussed. The optimal decorrelation transform is illustrated which turns an arbitrary memory source into a memoryless source in a mixed distribution model. Then we focus on the scalar quantiser design for memoryless circular and elliptical sources. The tri-axis coordinate system is proposed to determine the quantisation lattice, that is, the positions of quantisation levels, inspired by the well-known optimal hexagonal lattice for 2D uniformly distributed signals [17]. It provides a unified framework for both circular and elliptical distributions [A multivariate distribution is said to be elliptical if its characteristic function is of the form $e^{it'\nu}\phi(t'\Sigma t)$ for a specified vector ν, positive-definite matrix Σ and characteristic generator ϕ. When $\Sigma = I$, it is circular distribution.], and encompasses the polar quantisation as a special case. The proposed quantiser is also a kind of adaptive elastic lattice quantiser. We will present the simple design methodology, which utilises the Lloyd–Max quantisers for the corresponding 1D distributions. The merits of this scheme are verified on elliptical/circular Gaussian and Laplace distributions. The methodology description and experiments are focused on the bivariate random variables, and the extension to high-dimensional random variables is also discussed.

The advantages of our scheme include the following:

1. It provides an elegant quantisation lattice for arbitrary number of quantisation levels, especially for prime numbers.
2. It almost always has smaller MSE than the other quantisers.
3. It considers both memoryless and memory sources of arbitrary distributions, which include circular distributions, elliptical distributions and mixed distributions.
4. It is in a unified framework of a tri-axis coordinate system.
5. It has small design and implementation complexity.

The rest of the paper is organised as follows. Section 2 describes the system architecture of transform plus scalar quantisation to approximate the optimal vector quantiser. The preprocessing with transforms is discussed in Section 3 to decorrelate signals. In Section 4, we present a tri-axis coordinate system, and the methodology to design the optimal scalar quantiser for both circular and elliptical distributions in detail. Experimental results are shown in Section 5. Finally, Section 6 concludes the paper.

2 System architecture

2.1 Quantisation for compression

The general coding system usually includes transform, quantisation and entropy coding as shown in Fig. 1. The optimal transform could simplify vector quantisation scheme into scalar quantisation, and even replace variable-length entropy coding in the coding system with fixed length coding. R-D code is an optimal code proposed by Shannon listed in [3]. It is an optimal vector code when block length $n \to \infty$. Only is known its existence, but not its design in general. Vector quantisation has the ability to approach R-D bound when the number of quantisation levels $N \to \infty$, but is overwhelmed by the exponentially increasing complexity. Therefore vector quantisation is desired to be replaced by transformation followed by scalar quantisation with the same R-D performance but much less design and implementation complexity, as adopted by a general transform compression system shown in Fig. 1. Therefore an optimal transform plus an optimal scalar quantiser gives us a new promising guideline to achieve R-D bound as presented in the next sections.

2.2 Theorem and system framework

Theorem 1: The MMSE vector quantisation of uniform distribution sources could be achieved by transformation followed by scalar quantisation.

It will be proved in Section 4.

Following Theorem 1, we propose a system architecture as shown in Fig. 2. A vector quantiser is implemented by a transform and a scalar quantiser. The transform we focus on can be a linear transform with high decorrelation ability. We will discuss the unitary transforms, volume-preserving scaling transforms and the optimal decorrelation transforms in Section 3. The scalar quantiser is implemented in the 'tri-axis coordinate system' which will be described in detail in Section 4. The transformation plus scalar quantisation has the advantage of possible small complexity and good R-D performance. The system still has a tradeoff
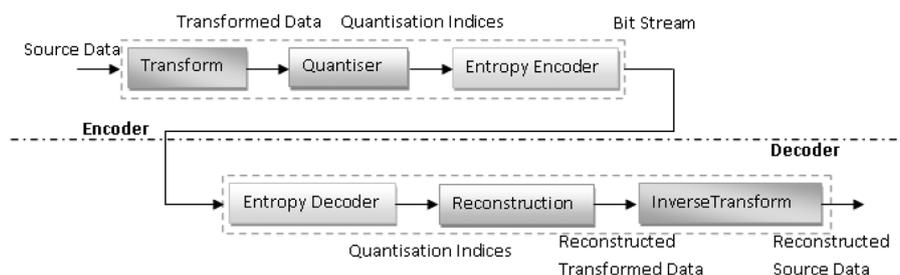

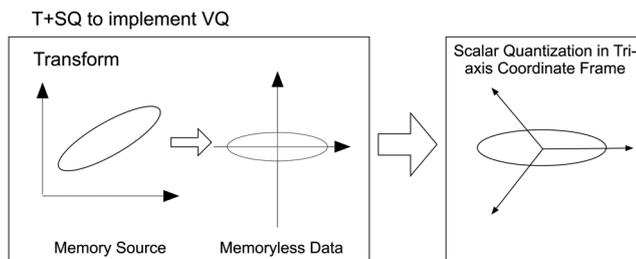**Fig. 1** *General encoding and decoding pipeline with transforms and scalar quantisation*

T+SQ to implement VQ

Fig. 2 *System architecture to implement VQ with a transform plus scalar quantisation*

issue between complexity (C), rate (R) and distortion (D). Therefore the system design should comprise the C–R–D theory. Best R-D performance with least complexity is desired.

The proposed system is flexible, in which the companding technique could also be plugged in as shown in Fig. 3. As we will show later, the companding technique is asymptotically optimal, but may not work well in low rate situation. However, our method with the tri-axis coordinate system works almost universally.

3 Preprocessing with transforms

Transformation is helpful for quantisation. Any mapping is a transform. Non-linear transforms introduce undesired non-linear error after quantisation. Therefore linear transform is considered in this section. To preserve signal energy constant, linear transforms are focused on, such as unitary transforms and volume-preserving scaling transforms, represented by matrices with unitary determinant. Such transforms can also be implemented lossless with PLUS factorisation [18].

Unitary transform is highlighted in our system, due to MSE invariance and high decorrelation ability.

Lemma 1: MMSE vector quantisation of random vectors transformed with a unitary transformation is equivalent to MMSE vector quantisation of the untransformed vectors and unitary transformation of quantised vectors.

Usually the MSE and energy of signals change after non-trivial scaling transformation or scaling transformation with unitary determinant. The rate-distortion theory requires the MSE be uniformly distributed for every component of the random vector, if the MSE does not exceed the variance of that component. Therefore the MMSE vector quantiser for an elliptical distribution could not be obtained from the MMSE vector quantiser for a circular distribution by a simple scaling. That is why the existing works seldom consider elliptical distributions, or independently consider

the circular and elliptical distributions. However, they are unified in our system.

4 Optimal scalar quantisers in tri-axis coordinate system

After transformation, we can obtain random vectors with independent components. For Theorem 1, 1D vector quantisation is scalar quantisation. No transform is needed. It is a trivial case. For 2D vector quantisation, we will prove this theorem for uniform distributions in a tri-axis coordinate system. For high-dimensional vector quantisation, a multi-axis coordinate system is needed.

4.1 Tri-axis coordinate system

Definition 1: The tri-axis coordinate system in a 2D space has three axes X, Y and Z, and the angles are 120° between the three axes X, Y and Z.

The tri-axis coordinate system is shown in Fig. 4. Every point in the 2D space can be represented by a point in this tri-axis coordinate system.

4.2 Orthonormal property of tri-axis coordinate system

Proposition 1: Any point in the tri-axis coordinate system (x, y, z) satisfies $x + y + z = 0$.
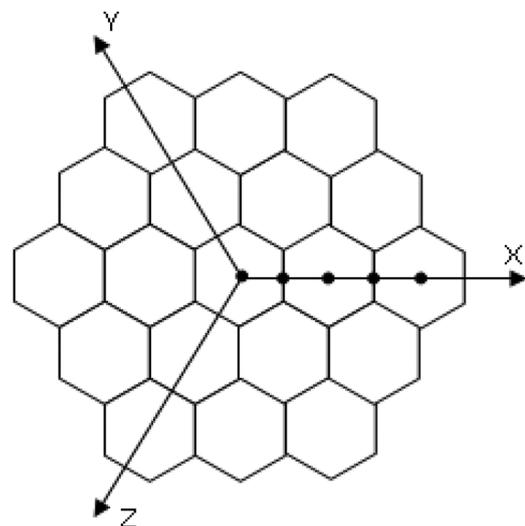

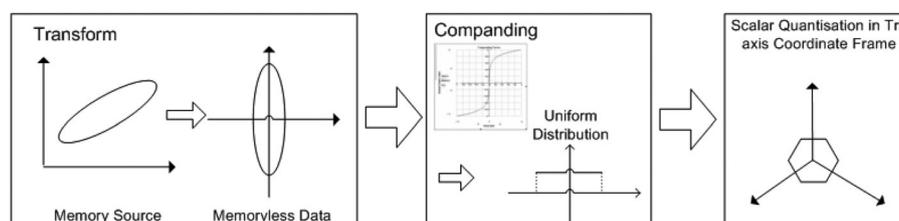**Fig. 4** *2D tri-axis coordinate system*

Fig. 3 *Pipeline of transform plus scalar quantisation with companding technique*

Proof: Unit vectors along three axes X, Y and Z could be $\boldsymbol{X} = (1,0)^T$, $\boldsymbol{Y} = (-1/2, \sqrt{3}/2)^T$ and $\boldsymbol{Z} = (-1/2, -\sqrt{3}/2)^T$. A point $\boldsymbol{P} = (u, v)$ in Cartesian coordinate system could be represented by (x, y, z) in tri-axis coordinate system with $x = P \cdot X$, $y = P \cdot Y$ and $z = P \cdot Z$. Then $x + y + z = P \cdot X + P \cdot Y + P \cdot Z = P \cdot (X + Y + Z) = P \cdot 0 = 0$. □

The big advantage of the tri-axis system is that the coordinate lattice is hexagonal, which is what we need for optimal vector quantisation, whereas bi-axis systems can only give quadrilateral coordinate lattices. Linear transforms preserve linearity and parallelism, therefore if a linear transform is applied, the coordinate lattice is still hexagonal and with three pairs of parallel edges although the angles are not necessarily 120°. The optimal vector quantiser for a uniform distribution can be perfectly represented by the tri-axis coordinate system as shown in Proposition 2, but not by polar coordinate system.

The coordinates (x, y, z) are highly correlated. For some symmetrical distributions, two axes or one axis is sufficient. For example, the optimal 2D vector quantisation for uniform distributions could be determined by two axes X, Y of 120° in spanning a hexagonal lattice. The optimal 2D vector quantisation for circular distributions also needs two axes, one of which determines the magnitude quantisation, another determines the phase quantisation. It is rotation-invariant for circular distributions, but not for elliptical distributions. We will show them in next subsections.

4.3 Tri-axis coordinate system for uniform distribution

It is well known that the optimal vector quantiser for uniform distributions in a 2D space is regular honeycomb [17], which is from the geometry of numbers, also from discrete geometry in the Euclidean space. We will implement it with scalar quantisation in tri-axis system as shown in Fig. 5.

Proposition 2: Hexagonal lattice in tri-axis system is still R-D optimal for quantisation of uniform distributions.

Proof:

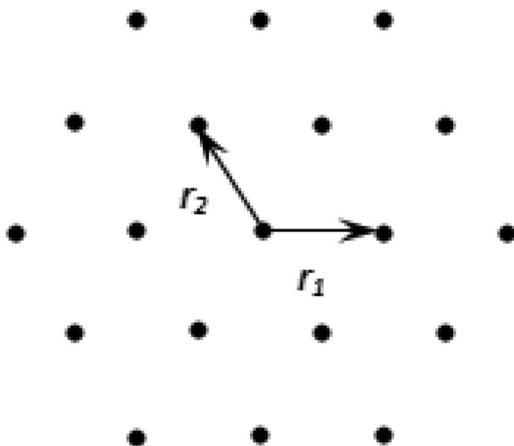1. Vector quantisation levels are the centroids of the hexagons. The centroid of each hexagon of the optimal

Fig. 5 *2D optimal uniform vector quantiser*

quantiser could be represented by a fixed length code. $R = \log_2 N$, where N is the number of quantisation levels.

2. Every point in the 2D space could be represented by the vector $\boldsymbol{r} = c_1 \boldsymbol{r}_1 + c_2 \boldsymbol{r}_2$, as shown in Fig. 5, where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the basis vectors of VQ. c_1 and c_2 are integers and uniformly distributed if the centroid is uniformly distributed.

3. Scalar quantisers compose of two independent scalar quantisers along two axes $\boldsymbol{r}'_1$ and $\boldsymbol{r}'_2$. Every point could be represented by two indices in the codebook. The indices are obtained by projecting the point to the nearest code on the axes. For example, a point representation is $\boldsymbol{r} = x \boldsymbol{r}'_1 + y \boldsymbol{r}'_2$ which is quantised to $\boldsymbol{r} = x_m \boldsymbol{r}'_1 + y_n \boldsymbol{r}'_2$.

4. Let $\boldsymbol{r}_1 = \boldsymbol{r}'_1$ and $\boldsymbol{r}_2 = \boldsymbol{r}'_2$, then we have $c_1 = x_m$ and $c_2 = y_m$. The sum of square error of SQ is the inner product of $(x - x_m)\boldsymbol{r}'_1$ and $(y - y_m)\boldsymbol{r}'_2$, and the sum of square error of VQ is the inner products of $(x - c_1)\boldsymbol{r}_1$ and $(y - c_2)\boldsymbol{r}_2$. Thus, VQ and SQ have the same distortion.

5. For uniform distribution, only one codebook is needed for two axes $\boldsymbol{r}'_1$ and $\boldsymbol{r}'_2$. The indices could be coded with a fixed length. $R_i = \log_2 N_i$, where N_i is the number of quantisation levels along axis i ($i = 1, 2$).

6. $N_1 \times N_2 = N$ asymptotically, that is, $R_1 + R_2 = \log_2 N_1 + \log_2 N_2 = \log_2 N = R$, VQ and SQ have the same rate.

7. Therefore SQ and VQ are with the same R-D performance.

Therefore the optimal MMSE vector quantisation could be achieved by a transform (an identity transform) followed by a uniform scalar quantiser for a 2D ideal uniform distribution. □

In this way, SQ and VQ have the same R-D performance, while the codebook size of SQ is around the square root of that of VQ, because $N_1 \times N_2 = N$. As the reduced complexity provided by SQ, we try to use SQ to replace VQ in this paper. Another point needed to mention is that the number of quantisation levels for hexagon lattice are prime numbers 1, 7, 19, 37, …, growing along circles with larger and larger radius, which could be found in Fig. 4. Therefore it is easy to design quantisers of prime number of quantisation levels with hexagonal lattices, whereas it is difficult and inferior with rectangular lattices for circular distributions.

4.4 Tri-axis coordinate system for circular and elliptical distributions

How about the distribution is not uniform, what will the optimal quantisation lattices be? This is a problem of finding the transform from a non-uniform distribution to a uniform distribution.

4.4.1 Elastic quantisation lattices for circular and elliptical distributions: From the optimal hexagonal lattice for a uniform distribution, we state that the optimal vector quantiser for a circular distribution forms an expanded hexagonal lattice, as shown in Figs. 6 and 7. The expansion ratio between the optimal lattice for a 2D circular distribution and that of a 2D uniform distribution along the radius direction may approximately follow the expansion ratio between the Lloyd–Max quantiser for the corresponding 1D distribution and that of a 1D uniform distribution, that is, the Lloyd–Max quantiser for the corresponding 1D distribution.

4.4.2 Design methodology: We firstly focus on the positions of quantisation levels of a 2D vector quantiser.
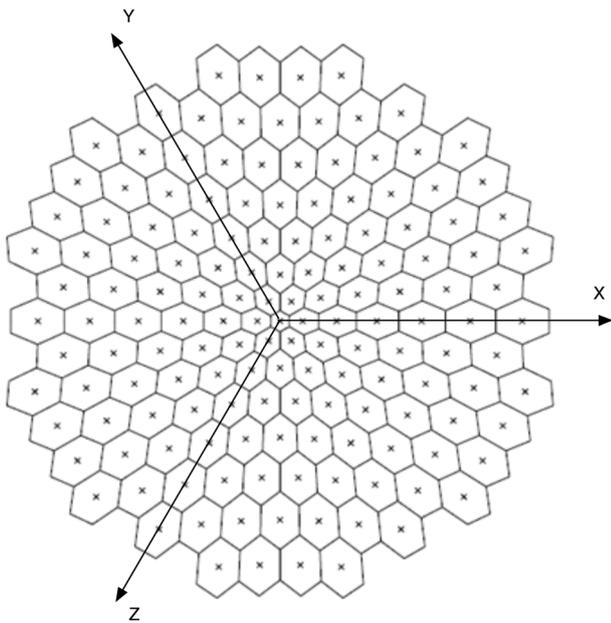
Fig. 6 *Circularly expanded hexagon lattice for 2D circular Gaussian distribution*

The lattice patterns of the proposed quantiser are determined beforehand, as shown in Figs. 9 and 10. The quantisation levels approximately fall on the centroids of the lattice, which are uniformly distributed in each annulus. We restrict them to be on the same circle for simplicity, and each quantisation region of lattice does not need to be hexagonal. In different annuluses, quantisation levels are staggered arranged similar to those of the rotated polar quantisation [15]. The optimal distance between the quantisation levels and the origin for the magnitude quantisation is determined by weighting the Lloyd–Max quantiser of the corresponding 1D distribution with the unitary-variance. For more precise locations of MMSE magnitude quantisation levels, they are further searched outwards in radial directions for MMSE.

To be specific, for a 2D circular distribution, its pdf could be separately represented in the polar coordinate system as $f(r, \theta) = f_1(r) \cdot f_2(\theta)$, while it is not for elliptical distributions.

For an arbitrary elliptical distribution, the data could be transformed by unitary transforms into a distribution whose principal axes are parallel to the coordinate systems, and then translated to the origin. After such transformation, the
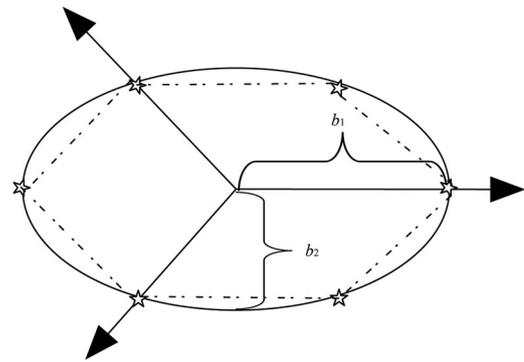

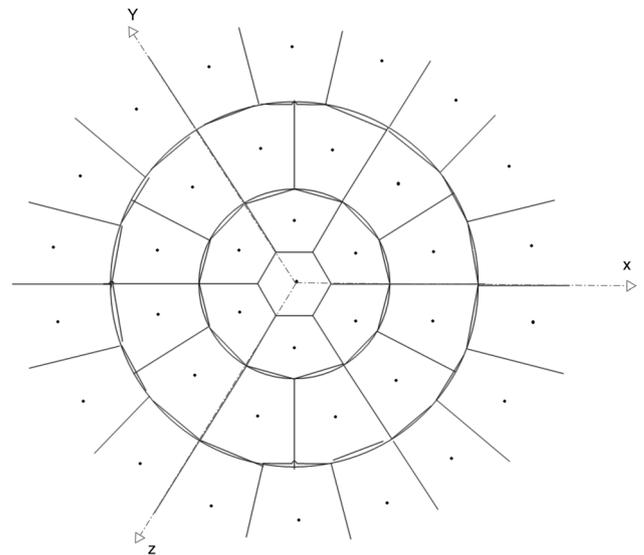**Fig. 8** *Tri-axis frame for a general 2D elliptical distribution*

Fig. 9 *Expanded hexagonal lattice for 2D circular Gaussian distribution*
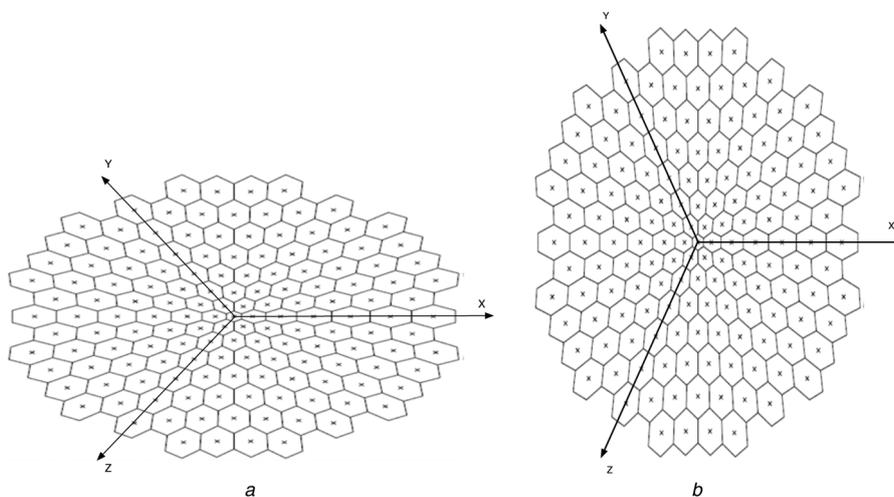

Fig. 7 *Elliptically expanded hexagon lattice for 2D elliptical Gaussian distribution*
a Horizontal elliptical hexagonal lattice
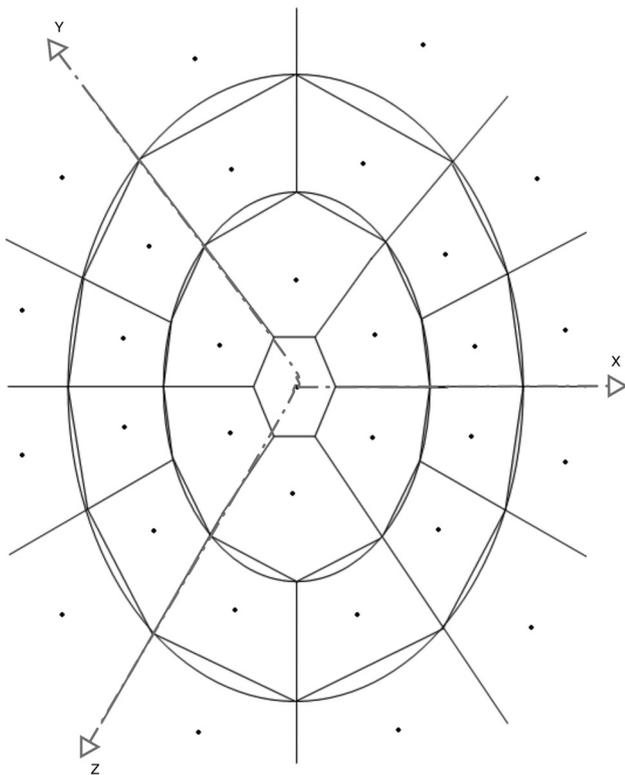*b* Vertical elliptical hexagonal lattice

Fig. 10 *Expanded hexagonal lattice for 2D elliptical Gaussian distribution*

equal-probability contours of distributions could be uniformly represented by the following equation in Cartesian coordinate system

$$\sum_{i=1}^{n} \frac{x_i^2}{b_i^2} = 1 \tag{1}$$

$b_1 = b_2 = \cdots = b_n = b$ for circular distributions; b_is are not all equal for elliptical distributions. The weighting effect from b_1 and b_2 for 2D elliptical distributions is important. As if the quantisers of circular distributions are used for elliptical distributions, the resultant 'MSE per dimension' has a ratio of b_1^2/b_2^2. Whereas, from Shannon R-D theory (i.e. reverse water filling), we know that if the MSE is less than the variance of each component, the bit-rate should be allocated such that the MSE per dimension is nearly equal. Therefore we should use b_1 and b_2 to weight quantisation levels towards this for elliptical distributions.

The magnitude quantisation is non-uniform. For both circular and elliptical distributions, the 2D quantisation levels fall on each circle or oval could be represented by the coordinates $(c \cdot b_1 \cdot \cos\theta, \ c \cdot b_2 \cdot \sin\theta)$ shown as stars in Fig. 8c increases non-uniformly in radial directions. c could be determined by searching outward starting from Lloyd–Max quantisation for Gaussian distribution in radial directions.

The uniform phase quantisation is optimal for circular distributions, but may not for elliptical distributions. We take uniform phase quantisation for both kinds of distributions, since the optimal phase quantisation for elliptical distributions is a little perturbation from the uniform phase quantisation. We will show its suboptimality for elliptical distribution in experiments. As shown in Fig. 9, the number of quantisation levels in each annulus is

1, 6, 12 and 18, similar to that of the regular hexagonal lattice. Within each magnitude annulus, the k phase regions all have equal size, whose boundaries are represented as follows

$$(j-1)\frac{2\pi}{k} \le \theta < j\frac{2\pi}{k} \tag{2}$$

where $j = 1, 2, \ldots, k$.

The boundaries of quantisation intervals are obtained by the nearest neighbour scheme based on the fixed quantisation levels. The resultant quantisation regions are not necessarily hexagonal. The optimal VQ for any distributions, including uniform distributions, in finite regions is deformed from hexagonal lattices in this way (Fig. 10).

4.4.3 Number of quantisation levels in each annulus:
How many quantisation levels should we assign to each annulus? Previously, for the restricted polar quantisation [11], quantisation levels N is factorised into $N = N_\theta \cdot N_r$, where N_θ is the number of quantisation levels in each annulus, and N_r is the number of annuluses. Although the optimal ratio between N_θ and N_r is studied, some numbers of N cannot be perfectly factorised, not to mention a prime number. This difficulty also lies in the unrestricted polar quantisation [14]. The non-continuity of quantisation patterns exists in all the previous works. It is also an imperfection in our schemes. We have two schemes to arrange magnitude quantisation levels against phase quantisation levels. Our quantiser design and optimisation methodology is much simpler than that of the unrestricted polar quantisation.

The first scheme allows freedom in the number of phases assigned at each magnitude level. The optimal patterns are derived from experiments, which are coincident with Wilson's scheme [14] but with better performance and Dirichlet boundaries, as shown in Fig. 11.

The second one is the progressive quantisation scheme [19] as shown in Fig. 12. The number of annuluses L increases with the number of quantisation levels $N = 1, 7, 19, \ldots, 1 + 6 \cdot (1 + 2 + 3 + \cdots)$. Define set $N_L = \{1, 7, 19, \ldots, 1 + 3 \, l(l+1)\}$ and $N_L(l)$ is the lth element in set N_L. That is the number of annuluses L is determined by

$$L = \begin{cases} \inf\{l : N - 3l \le N_L(l)\}, & N \le 4 \\ \inf\{l : N + 7 - 6l \le N_L(l)\}, & o/w \end{cases} \tag{3}$$

where inf is the infimum. Therefore the quantiser could be implemented progressively with the increase of N. The previously located quantisation levels need not change their relative positions, only their magnitudes should be shrinked a little as suggested by the Lloyd–Max quantiser of 1D Gaussian. Or hierarchically, we could further quantise each existing quantisation region with our scheme.

Comparing the two schemes of quantisation lattice patterns, we can see that the quantitative descriptions of the first optimal scheme are difficult to provide. For small N, the second scheme has performance close to the first scheme, although with possible different lattice patterns; for large N, their performance difference decreases, and the quantisation patterns of the second scheme asymptotically approach those of the first scheme. Scheme one and scheme two have a lot of common quantisation lattice patterns.
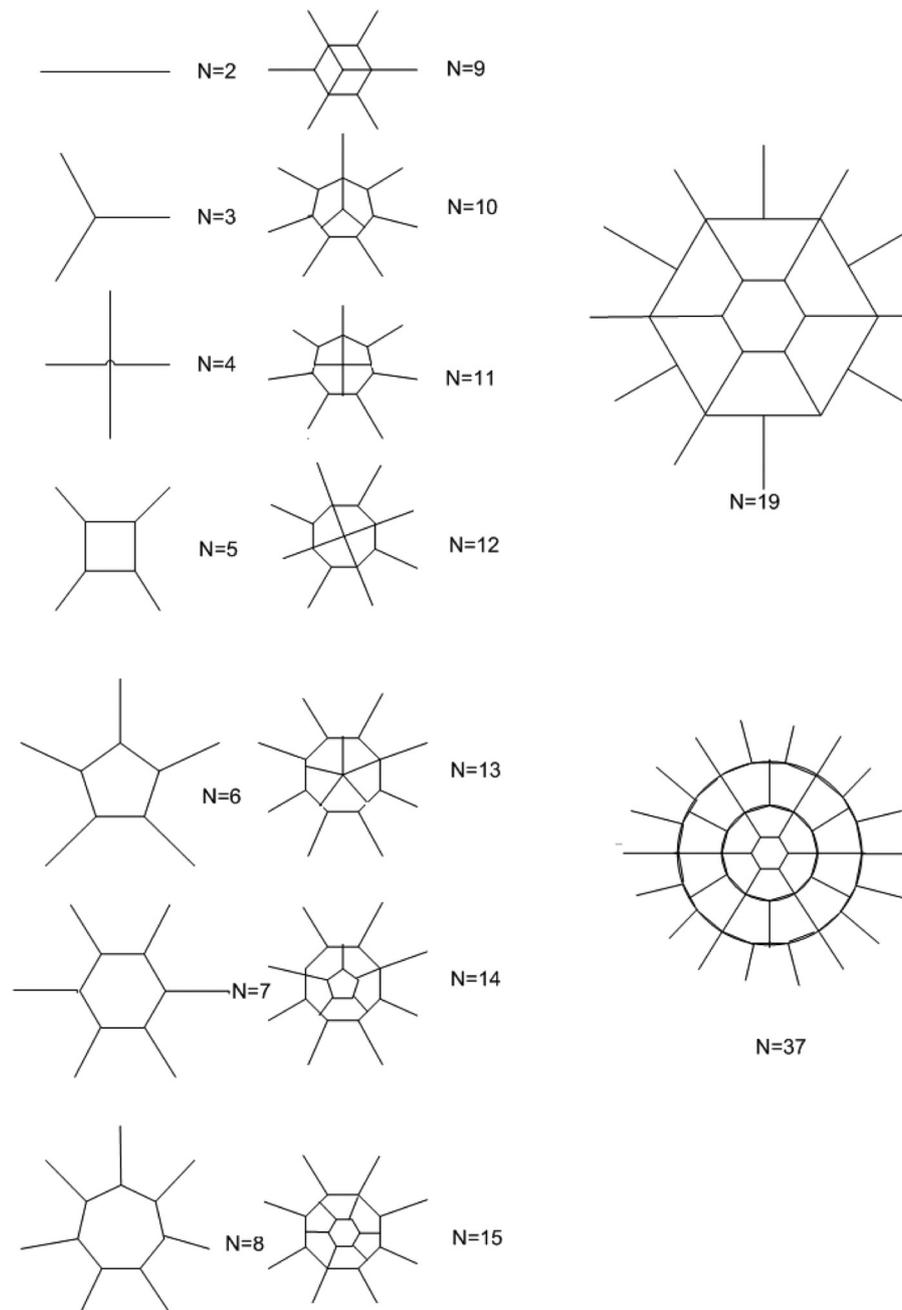
Fig. 11 *First optimal quantisation scheme*

4.4.4 Expansion rule: How far away are the quantisation levels found by the expansion rule along radial direction for 2D distributions from those found by the Lloyd–Max quantiser for the corresponding 1D distributions? Take Gaussian distribution for example. For 2D Gaussian with joint pdf given by

$$P_X(x_1, x_2) = \frac{1}{2\pi} \exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\} \qquad (4)$$

where $-\infty < x_1, x_2 < \infty$. Its polar coordinate representation is

$$P_{R,\Theta}(r, \theta) = \frac{r}{2\pi} \exp\{-r^2/2\} \qquad (5)$$

where

$$0 \le r < \infty, \quad 0 \le \theta < 2\pi \quad r = \left(x_1^2 + x_2^2\right)^{1/2} \quad \theta = tan^{-1}\left(\frac{x_2}{x_1}\right)$$ The number of annuluses L of the quantiser for 2D circular Gaussian has the following relationship with the number N_1 of quantisation levels of 1D Gaussian distribution

$$N_1 = \begin{cases} 2L, & L = 1 \\ 2L - 1, & L \ge 2 \end{cases} \qquad (6)$$

Then the expansion rule for r in (5) with L annuluses is found in table of the Lloyd–Max quantiser for 1D Gaussian with N_1 quantisation levels. For example, $N = 7$, $L = 2$ case as shown in Fig. 11. It corresponds to $N_1 = 3$ of the quantiser for 1D Gaussian, that is, $r_1 = 0$, $r_2 = 1.2240$.

Then how far away are $r_1 = 0$, $r_2 = 1.2240$ from the optimal r_1^*, r_2^*? Consider an upper bound of the difference between r_1
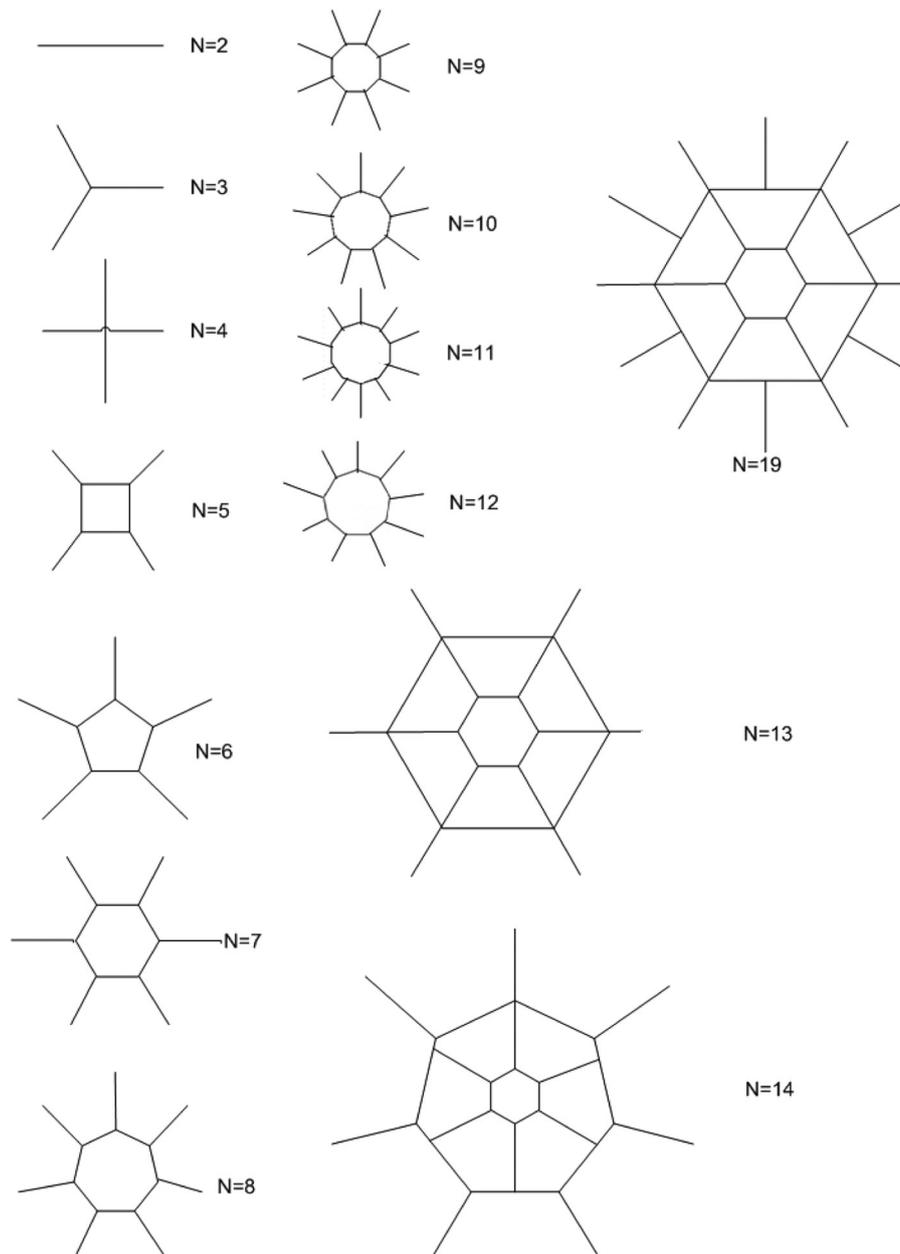
Fig. 12 *Second progressive quantisation scheme*

and r_1^* when $L = 1$. The radial expansion follows the rule for Rayleigh distribution. However, there is no quantisation level at origin for Rayleigh distribution, so we have to utilise the quantiser for Gaussian distribution for our quantiser

$$\frac{\int_0^\infty \int_{\theta_0}^{\theta_1} \frac{r^2}{2\pi} e^{-(r^2/2)} \, d\theta \, dr}{\int_0^\infty \int_{\theta_0}^{\theta_1} \frac{r}{2\pi} e^{-(r^2/2)} \, d\theta \, dr} = \sqrt{\frac{\pi}{2}}$$

The Lloyd–Max quantiser for 1D Gaussian distribution is as follows

$$\frac{\int_0^\infty \frac{x}{\sqrt{2\pi}} e^{-(x^2/2)} \, dx}{\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \, dx} = \sqrt{\frac{2}{\pi}}$$

Then the upper bound of the difference is around $0.46 \left(= \sqrt{\frac{\pi}{2}} - \sqrt{\frac{2}{\pi}} \right)$ for unit variance distributions. As N

obtains larger, the difference becomes smaller. These are also the maximal searching ranges to find the optimal magnitude quantisers. The Lloyd–Max quantiser for one-dimension Gaussian is a good initial for finding the optimal magnitude quantisation for 1D distributions.

The magnitude quantisation is almost independent of the phase quantisation. It means that when phase quantisation changes, the magnitude quantisation suffers a little perturbation at most. As the number of quantisation levels goes larger, the perturbation turns smaller.

5 Experimental results and discussions

In this section, we first show the basic properties of the proposed scalar quantiser. Then we show experimental results with a 2D memoryless source of unitary circular Gaussian and Laplace distributions, and elliptical Gaussian and Laplace distributions with $b_1 = 2$, $b_2 = 1$.

We will compare MSE and the R-D performance of our proposed quantisers based on the first scheme with optimal quantisation lattices, the unrestricted polar quantisers [14] (indicated by 'UPQ'), the restricted polar quantisers [11] (indicated by 'PQ'), the rectangular quantisers [6] (indicated by 'rectangular'). The rate here is defined as $\log_2 N/2$. The distortion is shown with MSE per dimension. The benchmark is R-D function for a Gaussian memoryless source

$$R(D) = \frac{1}{2}\log_2\left(\frac{1}{D}\right) \qquad (7)$$

where $0 < D \leq 1$. Each quantiser is tested for its best performance, with the corresponding optimal quantisation levels, and the optimal ratio between the number of phase quantisation levels and the number of magnitude quantisation levels. For example, the rectangular quantisers are almost tested with n^2 quantisation levels for circular distributions, that is, each dimension is quantised by a Lloyd–Max quantiser with n quantisation levels, and $2n \times n$ for elliptical distributions, that is, data are quantised by a Lloyd–Max quantiser with $2n$ and n quantisation levels, respectively, applied to the two dimensions. We also show the results of the vector quantisers found by Linde-Buzo-Gray (LBG) algorithm [20]. Since LBG is highly initial dependent and the results with bad initialisation are much worse than those of our proposed quantisers. Thus we use our proposed quantiser as initialisation of LBG algorithm, to see how much LBG could improve on top of our algorithm.

5.1 Basic optimal properties

1. The property of optimal solutions is considered as follows. Assume MMSE per dimension is the objective. For $N = 7 = 1 + 6$ with two magnitude levels (i.e. the first magnitude quantisation level is quantised with one phase quantisation level, the second magnitude quantisation level with six phase quantisation levels), MSE per dimension performance on uni-variance circular Gaussian is shown in Fig. 1 with respect to different radius of circles where lies the second magnitude quantisation level. The radius shown in Fig. 13 starts from the second quantisation level of Lloyd–Max quantiser for univariate Gaussian around 1.224. Then MSE per dimension decreases with the increase of radius, reaches
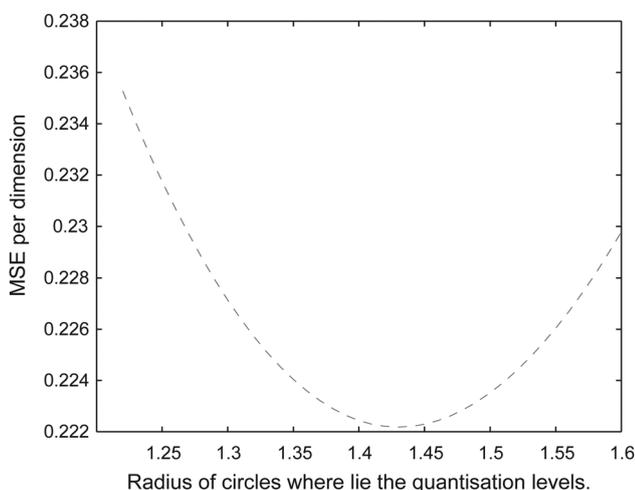
Fig. 13 *MSE per dimension for quantisation of 10 000 samples from uni-variance circular Gaussian distribution*
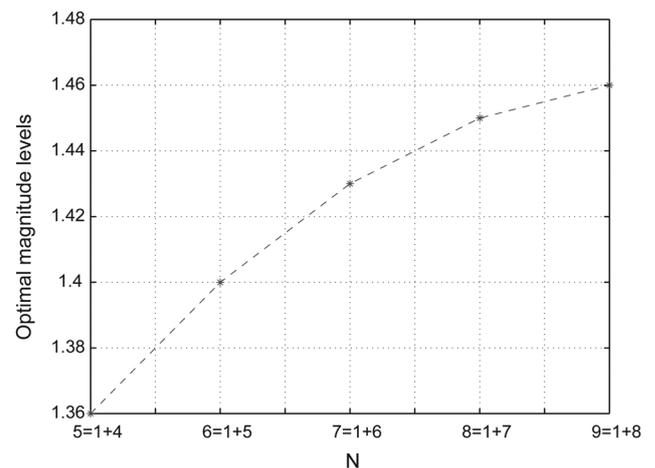
Fig. 14 *Optimal magnitude levels for different number of quantisation levels for uni-variance circular Gaussian distribution*

its unique minimum about 1.43, and then increases with the increase of radius. With more magnitude levels and more than one radiuses needed to be tuned for the optimal performance, there are definitely local minima. However, the optimal radiuses could be easily and quickly found with values starting from the quantisation levels of Lloyd–Max quantisers for univariate Gaussian.

2. For the quantisation lattices with the same number of magnitude quantisation levels, the radiuses of the optimal magnitude levels increase with the number of quantisation levels N, and are saturated with relatively large N. The optimal radiuses of the second magnitude quantisation level are shown by the vertical coordinates of points in Fig. 14 corresponding to the number of quantisation levels $N = 5(= 1 + 4)$, $6(= 1 + 5)$, $7(= 1 + 6)$, $8(= 1 + 7)$ and $9(= 1 + 8)$. They increase with N, and gradually slow down. This gives us a guidance on how to tune the optimal magnitude quantisation levels.

5.2 Circular Gaussian distribution

We show the R-D performance of different quantisers on a uni-variance circular Gaussian distribution in Fig. 15. From Fig. 15, we can see that the R-D performance of our proposed quantisers is always a little better than that of UPQs, and much better than that of PQs and that of rectangular quantisers. They have the same R-D performance when $N = 4$, because of we assign them the same quantisation levels. Rectangular quantisers may have better performance than PQs with some n^2 (such as $n = 3$) quantisation levels. Lloyd–Max quantiser with LBG on top of our proposed algorithm can improve performance a little in some cases to match random input better.

5.3 Elliptical Gaussian distribution

We show the R-D performance of different quantisers on an elliptical Gaussian distribution in Fig. 16. From Fig. 16, we can see that UPQs do not consider the different variances among different random vector components, thus do not perform well. Our proposed quantisers almost always perform better than rectangular quantisers, except when $N = 8$. Since $N = 8 = 4 \times 2$ is the best factorisation for the
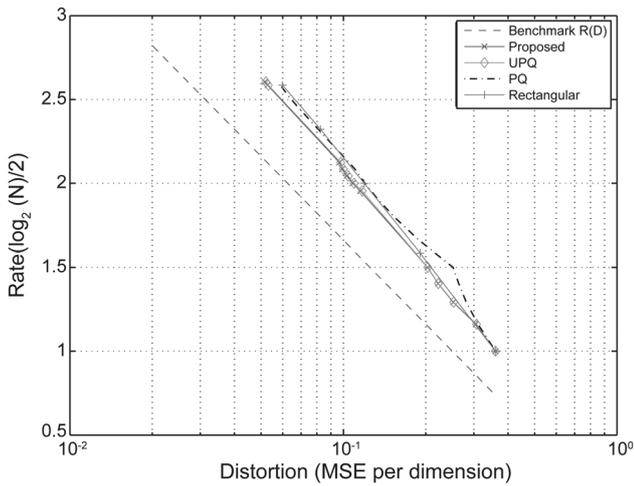
Fig. 15 *R-D comparison among different quantisers for circular Gaussian distribution*

Fig. 17 *R-D comparison among different quantisers for circular Laplace distribution*

rectangular quantiser on elliptical distributions when the ratio of data component variances equals 2. Whereas, for other N non-factorable, rectangular quantisers perform much worse than polar quantisers and the proposed quantisers as expected, although we did not plot it in the figure. Also Lloyd–Max quantiser with LBG on top of our proposed algorithm can improve performance a little in some cases to match random input better.

5.4 Circular Laplace distribution

We show the R-D performance of different quantisers on a uni-variance circular Laplace distribution in Fig. 17. It indicates in Fig. 17 that our proposed quantisers always perform a little better than UPQs, and much better than PQs and rectangular quantisers. PQ performs better than rectangular when $N = 4$, because of Laplace distribution is peak at the centre. However, PQ following Guassian distribution does not perform well here. Lloyd–Max quantiser with LBG on top of our proposed algorithm almost performs similar to our proposed quatisation with a small improvement.
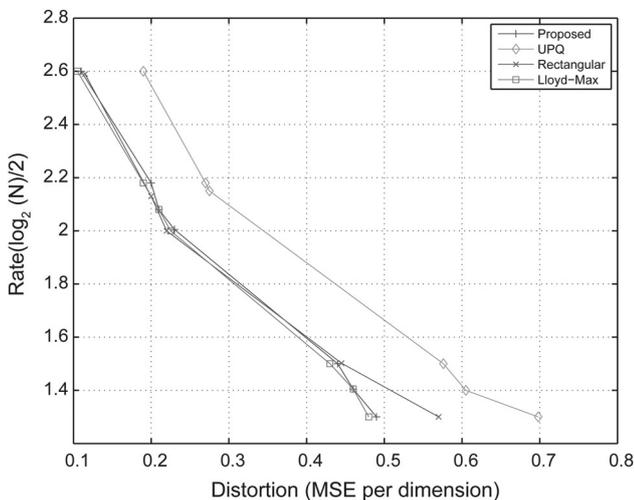
5.5 Elliptical Laplace distribution

We show the R-D performance of different quantisers on an elliptical Laplace distribution in Fig. 18. It tells that our proposed quantisers always perform better than UPQs, and better than rectangular quantisers except when $N = 8 = 4 \times 2$. Our proposed quantisers have predominant advantages when $N = 7, 19, 37, \ldots$. Lloyd–Max quantiser with LBG on top of our proposed algorithm has a general improvement over our proposed quatiser.

5.6 Bit-rate saving

We also evaluate the average bit-rate saving of our quantisers compared to other quantisers. Average bit-rate is calculated by using Bjontegaard's method [21, 22] with fitting polynomials of degree 3. Bit-rate saving is evaluated based on relative average bit-rate in percentage as shown in the following equation
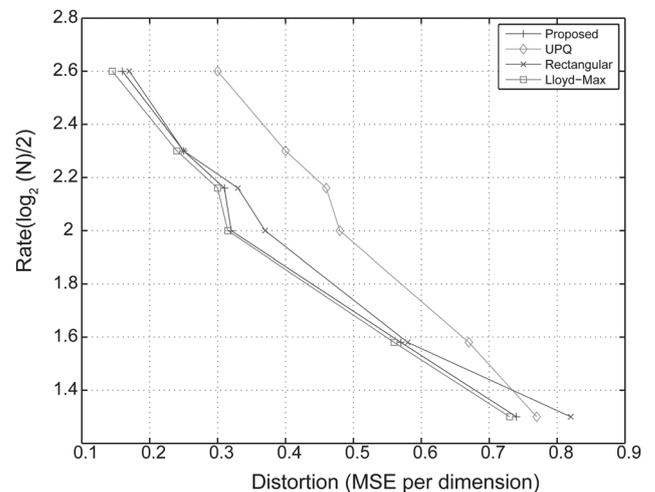
$$\frac{R_c - R_p}{R_p} \times 100\% \tag{8}$$

Fig. 16 *R-D comparison among different quantisers for elliptical Gaussian distribution*

Fig. 18 *R-D comparison among different quantisers for elliptical Laplace distribution*

Table 1 Average bit-rate saving of the proposed quantisers over other quantisers

	UPQ, %	PQ, %	Rectangular, %
circular Gaussian	0.36	6.78	3.22
elliptical Gaussian	22.4	—	16.9
circular Laplace	0.94	24.5	5.62
elliptical Laplace	19.8	—	6.32

where R_c is the average bit-rate of the compared quantisers, and R_p is the average bit-rate of the proposed quantisers.

From Table 1, we can see that the proposed quantiser saves 0.4–24.5% bit-rate on average, compared with unrestricted polar quantisers, restricted polar quantisers and rectangular qantisers. We did not list the average bit-rate gain over restricted polar quantisers for elliptical distributions, which is even higher than that over unrestricted polar quantisers.

6 Conclusions

In this paper, we proposed a scheme to use transformation plus scalar quantisation to replace the optimal vector quantisation. The unitary transforms rather than scaling transforms were needed for the optimal vector quantiser approximation. After transformation, scalar quantisation for both circular and elliptical distributions was studied in the proposed tri-axis coordinate system. The optimal quantisation levels were found in the elastic hexagonal lattices, which include the optimal and the progressive quantiser lattice patterns. The experimental results showed that our proposed quantisers almost always had better performance than UPQs, PQs and rectangular quantisers on both Gaussian and Laplace distributions, especially with a prime number of quantisation levels. We achieved $O(N^2)$ design complexity and 0.4–24.5% bit-rate saving, where N is the number of quantisation levels per dimension. Therefore we claimed that transforms plus scalar quantisers could approximate the optimal vector quantisers in terms of R-D performance but with much less computational complexity. Our future work will focus on the optimal vector quantiser approximation in high-dimensional spaces, and the applications in image and video coding.

The future work will focus on optimal quantisers for arbitrary distribution, which could be modelled by Gaussian mixture model or Laplacian mixture model, and develop algorithms for higher dimentional signals.

7 References

1 Gray, R.M., Neuhoff, D.L.: 'Quantization', *IEEE Trans. Inf. Theory*, 1998, **44**, (6), pp. 2325–2383
2 Gersho, A., Gray, R.: 'Vector quantization and signal compression' (Kluwer, 1993)
3 Cover, T., Thomas, J.: 'Elements of information theory' (Wiley, 2006)
4 Bryc, W.: 'The normal distribution: characterizations with applications' (Springer-Verlag, 1995)
5 Cover, T., Thomas, J., Wiley, J., *et al.*: 'Elements of information theory' vol. 306 (Wiley, 1991), Online Library
6 Huang, J., Schultheiss, P.: 'Block quantization of correlated Gaussian random variables', *IEEE Trans. Commun. Syst.*, 1963, **11**, (3), pp. 289–296
7 Max, J.: 'Quantizing for minimum distortion', *IRE Trans. Inf. Theory*, 1960, **6**, (1), pp. 7–12
8 Zador, P.: 'S. U. D. of statistics, development and evaluation of procedures for quantizing multivariate distributions' Dept. of Statistics, Stanford University, 1963
9 Gersho, A.: 'Asymptotically optimal block quantization', *IEEE Trans. Inf. Theory*, 1979, **25**, (4), pp. 373–380
10 Moo, P., Neuho, D.: 'Uniform polar quantization'. Proc. IEEE Int. Symp. of Information Theory, Citeseer, 1998
11 Pearlman, W.: 'Polar quantization of a complex Gaussian random variable', *IEEE Trans. Commun.*, 1979, **27**, (6), pp. 892–899
12 Bucklew, J., Gallagher Jr., N.: 'Quantization schemes for bivariate Gaussian random variables', *IEEE Trans. Inf. Theory*, 1979, **25**, (5), pp. 537–543
13 Bucklew, J., Gallagher Jr., N.: 'Two-dimensional quantization of bivariate circularly symmetric densities', *IEEE Trans. Inf. Theory*, 1979, **25**, (6), pp. 667–671
14 Wilson, S.: 'Magnitude/phase quantization of independent Gaussian variates', *IEEE Trans. Commun.*, 1980, **28**, (11), pp. 1924–1929
15 Swaszek, P., Thomas, J.: 'Optimal circularly symmetric quantizers', *J. Franklin Inst.*, 1982, **313**, (6), pp. 373–384
16 Chen, Q., Fischer, T.: 'Robust quantization for image coding and noisy digital transmission'. Proc. on Data Compression Conf. 1996, 1996, pp. 3–12
17 Jain, A.: 'Fundamentals of digital image processing' (Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989)
18 Yang, L., Hao, P., Wu, D.: 'Stabilization and optimization of plus factorization and its application in image coding', *J. Vis. Commun. Image Represent.*, 2011, **22**, (1), pp. 9–22
19 Ravelli, E., Daudet, L.: 'Embedded polar quantization', *IEEE Signal Process. Lett.*, 2007, **14**, (10), pp. 657–660
20 Graf, S., Luschgy, H.: 'Foundations of quantization for probability distributions' (Springer-Verlag, New York, Inc. Secaucus, NJ, USA, 2000)
21 Bjontegaard, G.: 'Calculation of average psnr differences between rd-curves (vcegm33)'. Proc. VCEG Meeting (ITU-T SG16 Q. 6), 2001, pp. 2–4
22 Pateux, S., Jung, J.: 'An excel add-in for computing bjontegaard metric and its evolution'. Proc. VCEG document VCEG-AE07, 31st VCEG Meeting, Marrakech, MA, 2007, pp. 15–16