Successive Refinement of Gaussian Projections

Giel J. Op 't Veld Michael C. Gastpar
École Polytechnique Fédérale de Lausanne
Dept. I&C, LINX
Lausanne, Switzerland
giel.optveld@epfl.ch michael.gastpar@epfl.ch

Abstract

Successive refinement is a technique to encode a source iteratively, improving the distortion at each step. If the code rates of all steps combined can be as efficient as coding only for the final distortion in one go, then the source is said to be successively refinable. We turn our attention to vector sources and we propose to replace the high-distortion first stage by a projection to low-dimensional space. The goal remains the same: code the original, high-dimensional source afterwards by only a small addendum. We look at Gaussian sources subject to an MSE criterion as a first step in this direction. A coding scheme for this problem is developed by turning the projection into an estimator of the source. It is shown that Gaussian vector sources coded via a projection can be successively refinable. We derive a parametric description of distortion pairs that can achieve this state; this region exists for any projection matrix, but its size may vary.

1 Introduction

With the rapid growth of data, it becomes increasingly difficult to assess in advance whether a data set is going to be of value. To prevent any unnecessary processing, one can evaluate either a fragment or a lower quality thumbnail first to decide if one wants to have the entire set. Equitz and Cover pitched the idea of successive refinement as an improvement on efficiency: can one optimally code a random variable X at a low distortion D_1 and then use the bits such that only an addendum is required to code a second representation at a better distortion $D_2 < D_1$ [1]? Moreover, can such two stages combined be as efficient as the rate-distortion optimal rate $R_X(D_2)$? The answer turned out to be affirmative, be it under strict conditions and for specific sources.

We extend this idea to involve projections. First, a user requests only a low-dimensional projection of the original data. This projection reflects his interests: he may value certain components more than others in his assessment of the data's value. Can one still use this first low-dimensional projection to code a final high-dimensional representation by only a small addendum and if so, can one do it without the loss of rate? In a continuous setting, we adopt the mean squared error as our distortion measure. Gaussian sources, as maximizers of differential entropy, form our choice for a start to the problem.

The amount of classic results on successive refinement is vast, featuring amongst others [1, 2]. Both [3] and [4] also considered the successive coding of two *correlated* variables X and Y; our problem is in principle encapsulated in their models. Both papers also go into the specific case of a *pair* of jointly Gaussians. We let the source X and the projection $U = \mathbf{A}^T X$ be of arbitrary dimensions, N and M respectively. Other work by Nayak and Tuncel involved the extension to vector sources as well, but focused on accommodating individual distortion criteria [5].

We present a sequential coding scheme for Gaussian sources with a projection on the first stage in Section 3. The strategy is to, if the projection was satisfactory, turn it into an estimator of the original source and use it as side information on the

refinement. The algebraic nature of Gaussian sources allows us to find the estimator in closed form and we show that the refinement boils down to a rank-M downdate of the source statistics in Section 4. Furthermore, we will derive a lemma on the rates and comment on the successive refinability of the problem afterwards in Section 5. We find a parametric expression for a distortion level on the projection that marks a threshold in successive refinability. The value of this threshold depends on the projection $\mathbf{A}$, but it always exists, provided the required distortion on the second stage is not too large.

2 Preliminaries

Let X be a random variable over some alphabet $\mathcal{X}$ and let $\hat{X}$ be its reconstruction, whose precision one evaluates by some distortion measure $d(X, \hat{X}) : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$. A classic, two stage successive refinement problem codes X twice, first at a distortion D_1 and then at a $D_2 \leq D_1$. The scheme is such that the second stage only codes an addendum to the bits of the first so as to meet $d(X, \hat{X}_2) \leq D_2$. In particular, a source is called *successively refinable* if it achieves the rate distortion limit on both stages, i.e., $R_1 = R(D_1)$ and $R_1 + R_2 = R(D_2)$. Equitz and Cover showed that Gaussians subject to an MSE criterion are successively refinable [1].

This source coding problem has been extended to involve different, yet correlated sources on the first and second stage (X_1, X_2) by most notably [4, 3]. One first codes X_1 and subsequently, X_2 can be described by fewer bits if one exploits the correlation with $\hat{X}_1$. Viswanathan and Berger adopt the terminology *sequential coding* to refer to a situation in which the first encoder only has access to the first source, while the second has access to both [4]. Nayak and Tuncel allowed both encoders access to both sources and dubbed their problem *successive coding* [3]. Of these two, we adopt the sequential coding approach.

In this paper, X is an N-dimensional vector and we replace the first stage by a projection to M-dimensional space for $M < N$. We label it

$$U = \mathbf{A}^T X, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{N \times M}$ is an arbitrary matrix with orthonormal columns. In contrast to U, we label the vector corresponding to the refinement in the second stage as V. One could view upon the source and its projection as two correlated sources, X and U. Their correlation is of course special, since $p(X, U) = p(X)\delta(U - \mathbf{A}^T X)$. All in all, we seek the following encoder-decoder pairs:

$$\begin{cases} f_1 : & \mathcal{U}^n \to \{1, \cdots, 2^{nR_U}\} \\ f_2 : & \mathcal{X}^n \to \{1, \cdots, 2^{nR_V}\} \\ g_1 : & \{1, \cdots, 2^{nR_U}\} \to \hat{\mathcal{U}}^n \\ g_2 : & \{1, \cdots, 2^{nR_U}\} \times \{1, \cdots, 2^{nR_V}\} \to \hat{\mathcal{X}}^n, \end{cases} \tag{2}$$

where $\hat{U} = g_1(f_1(U))$ and $\hat{X} = g_2(f_1(U), f_2(X))$. A schematic is depicted in Figure 1.

We investigate the impact of these projections for a specific case: Gaussians under mean squared error distortion measure. The source thus follows $X \sim \mathcal{N}(0, \Sigma_X)$, with $\Sigma_X = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N)$, ordered such that $\lambda_1 \geq \cdots \geq \lambda_N$. The same ordering and numbering applies to the eigenvalues of any matrix. We adopt scalar distortions (D_U, D_X) to evaluate $\hat{U}$ and $\hat{X}$ respectively. Since we are working with vectors, an encoder will allocate rate to each vector component such that their individual distortion levels satisfy the end distortion constraint. To accommodate that procedure with terminology, $\hat{D}_{X,i}$ refers to the individual distortion of the i'th component of X. For

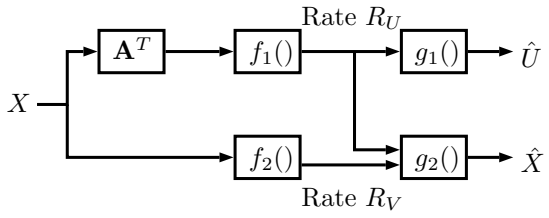

Figure 1: Schematic of successive refinement with a projection.

convenience, it is sometimes shorter to aggregate the distortion profile into a diagonal matrix $\hat{\mathbf{D}}_X = \mathrm{diag}_i(\hat{D}_{X,i})$.

The achievable rate region of our problem is contained in the results of [4].Thus if we reformulate, the achievable rate region is already known to be:

Theorem 1. *The rate distortion quadruple (R_U, R_V, D_U, D_X) is achievable if and only if there exists a joint distribution $p(X, \hat{X}, \hat{U})$ such that*

$$I(U;\hat{U}) \leq R_U \tag{3}$$

$$I(X;\hat{X}|\hat{U}) \leq R_V \tag{4}$$

$$\mathbb{E}[||U - \hat{U}||^2] \leq D_U \tag{5}$$

$$\mathbb{E}[||X - \hat{X}||^2] \leq D_X \tag{6}$$

In the following sections, we will introduce a sequential coding scheme for Gaussian sources, whose optimality we will evaluate with respect to the following definition:

Definition 1. *We say the source X is successively refinable via a projection U at (D_U, D_V) if the quadruple (R_U, R_V, D_U, D_X) is achievable and it holds that $R_U = R_U(D_U)$ and $R_U + R_V = R_X(D_X)$.*

3 Coding Strategy

Let us first emphasize that Definition 1 states that successive refinability requires *both* stages to be rate-distortion optimal. Hence, any possible code for a Gaussian source that we wish to be fully successively refinable can be constructed by minimizing the rate of the first stage first; the second stage will follow suit. A trivial first observation is that if X is Gaussian then so is $U = \mathbf{A}^T X$. Hence, the rate $R_U \geq R_U(D_U)$ is uniquely (see [6, Theorem 12.1.1]) minimized by the straightforward procedure of reversed water filling [6, p. 314], which leads to a $\hat{U}$ that is Gaussian as well.

It is not immediately clear that the refinement, which we dubbed V, is best coded in a Gaussian fashion as well. After the first stage, the decoder has a copy of $\hat{U}$ at a distortion no more than D_U and he may decide he is also interested in a copy of the high-dimensional X. Before even starting to code, the decoder already knows something on X, namely he can compute $\mathbb{E}[X|\hat{U}]$ as its MMSE-estimate. Hence, the second stage only revolves around the residual, that which is not yet known on X:

$$V = X - \mathbb{E}[X|\hat{U}], \tag{7}$$

so that coding this leads to a final estimate on X afterwards as:

$$\hat{X} = \hat{V} + \mathbb{E}[X|\hat{U}]. \tag{8}$$

Conveniently, the distortion measures on X and the residual V match as follows:

$$\mathbb{E}[||X - \hat{X}||^2] = \mathbb{E}[||(V + \mathbb{E}[X|\hat{U}]) - (\hat{V} + \mathbb{E}[X|\hat{U}])||^2] = \mathbb{E}[||V - \hat{V}||^2]. \tag{9}$$

In other words, coding V for a distortion D_X after having received $\hat{U}$ ensures one of also having a copy of $\hat{X}$ at the same distortion. In this setup, the jointly Gaussian nature of (U, V) follows from the following:

$$\begin{aligned} I(U;\hat{U}) + I(X;\hat{X}|\hat{U}) &\geq R_U(D_U) + h(X|\hat{U}) - h(X|\hat{X},\hat{U}) && (10)\\ &= R_U(D_U) + h(X - \mathbb{E}[X|\hat{U}]|\hat{U}) - h(X - \hat{X}|\hat{X},\hat{U}) && (11)\\ &\overset{\dagger}{\geq} R_U(D_U) + h(V) - h(V - \hat{V}) && (12)\\ &\geq \frac{1}{2}\log\left(\frac{|\Sigma_U|}{|\hat{\mathbf{D}}_U|}\right) + \frac{1}{2}\log\left(\frac{|\Sigma_V|}{|\hat{\mathbf{D}}_V|}\right), && (13) \end{aligned}$$

where $\hat{\mathbf{D}}_U$ is the reversed water filling solution, computed as $\hat{D}_{U,i} = \min(\theta_U, \lambda_i(\Sigma_U))$ with θ_U chosen such that $\sum_{i=1}^{n} \hat{D}_{U,i} = D_U$ is satisfied. The same goes for $\hat{\mathbf{D}}_V$, but with respect to end distortion constraint D_X. (†) Follows from the fact that via this construction, $\hat{U}$ is jointly Gaussian with X and hence $X - \mathbb{E}[X|\hat{U}] \perp \hat{U}$. With this set-up in check, it remains to be found what the covariance of the residual, Σ_V, is and what impact it has on the sum-rate $R_U + R_V$ when compared to $R_X(D_X)$.

4 Estimator and Refinement Algebra

Thanks to the Gaussian nature of both stages, rates can be computed in closed form. To that end, we first derive the estimator $\mathbb{E}[X|\hat{U}]$ and the residual statistics Σ_V. We already concluded that R_U is minimized by applying reversed water filling to the principal components of covariance Σ_U, which we find as

$$\Sigma_U := \mathbb{E}[UU^T] = \mathbf{A}^T \Sigma_X \mathbf{A}. \tag{14}$$

Without loss of generality, we define our projections $\mathbf{A}$ actually as follows:

$$\mathbf{A} = \tilde{\mathbf{A}}\mathbf{Q}, \tag{15}$$

in which $\mathbf{Q}$ are the eigenvectors of $\tilde{\mathbf{A}}^T \Sigma_X \tilde{\mathbf{A}}$. $\tilde{\mathbf{A}}$ can still be arbitrary, as long as its columns are orthonormal. Incorporating $\mathbf{Q}$ inside $\mathbf{A}$ has as an effect that Σ_U is already diagonalized into its eigendecomposition. Consequently, also the $\hat{\mathbf{D}}_U$ found by means of (13) is diagonal. Making this diagonalization implicit is without loss of generality and allows us to construct simpler equations. Namely, we build a simple Gaussian test channel $\hat{U} = B(U + W)$ (see, e.g., [6, p. 339]) where $U \perp W$, $W \sim \mathcal{N}(0, \Sigma_W)$ and

$$\begin{aligned} \Sigma_W &= \mathrm{diag}_{i=1,\cdots,M}\left(\frac{\lambda_i(\Sigma_U)\hat{D}_{U,i}}{\lambda_i(\Sigma_U) - \hat{D}_{U,i}}\right) &&= \Sigma_U \hat{\mathbf{D}}_U (\Sigma_U - \hat{\mathbf{D}}_U)^{-1} && (16)\\ B &= \mathrm{diag}_{i=1,\cdots,M}\left(\frac{\lambda_i(\Sigma_U) - \hat{D}_{U,i}}{\lambda_i(\Sigma_U)}\right) &&= (\Sigma_U - \hat{\mathbf{D}}_U)\Sigma_U^{-1}. && (17) \end{aligned}$$

Since X and $\hat{U}$ are jointly Gaussian, the estimation of X follows as:

$$\mathbb{E}[X|\hat{U}] = \mathbb{E}[X\hat{U}^T]\mathbb{E}\left[\hat{U}\hat{U}^T\right]^{-1}\hat{U}, \tag{18}$$

of which the unknown expectations on the right hand side are found as:

$$\mathbb{E}\left[X\hat{U}^T\right] = \Sigma_X \mathbf{A}\mathbf{B} \tag{19}$$

$$\mathbb{E}\left[\hat{U}\hat{U}^T\right] = \mathbf{B}\left(\Sigma_U + \Sigma_W\right)\mathbf{B} = \Sigma_U - \hat{\mathbf{D}}_U. \tag{20}$$

Observe that (20) confirms that the channel follows the desired distribution, i.e., $p(\hat{U}) \sim \mathcal{N}(0, \Sigma_U - \hat{\mathbf{D}}_U)$. This ultimately gives us the first estimation of X as

$$\mathbb{E}[X|\hat{U}] = \Sigma_X \mathbf{A}\mathbf{B}\left(\Sigma_U - \hat{\mathbf{D}}_U\right)^{-1}\mathbf{B}(\mathbf{A}^T X + W) \tag{21}$$

$$= \Sigma_X \mathbf{A}\Sigma_U^{-2}(\Sigma_U - \hat{\mathbf{D}}_U)(\mathbf{A}^T X + W). \tag{22}$$

The only hiccup in this analysis is that $(\Sigma_U - \hat{\mathbf{D}}_U)^{-1}$ might not be invertible, which happens when $\hat{D}_{U,i} = \lambda_i(\Sigma_U)$ for some i. This situation occurs when the reversed water filling procedure concludes that some principal components of U do not require coding to meet the distortion constraint D_U. These components do not partake in the coding and thus cannot contribute to the estimator and can therefore be safely excluded to make everything invertible. Notice, though, that the exclusion is implicit in (22).

As for the covariance of the residual $V = X - \mathbb{E}[X|\hat{U}]$,

$$\Sigma_V(D_U) := \mathbb{E}[VV^T] = \mathbb{E}[(X - \mathbb{E}[X|\hat{U}])(X - \mathbb{E}[X|\hat{U}])^T] \tag{23}$$

$$= \mathbb{E}[(X - \mathbb{E}[X|\hat{U}])X^T] \tag{24}$$

$$= \Sigma_X - \Sigma_X \mathbf{A}\Sigma_U^{-2}(\Sigma_U - \hat{\mathbf{D}}_U)\mathbf{A}^T\Sigma_X. \tag{25}$$

One term in (23) drops out by the orthogonality principle. The statistics of this residual V are thus a perturbation of the original source statistics by subtracting a real symmetric matrix of at most rank M, representing what one learned on the first stage. We denote this covariance $\Sigma_V(D_U)$ as a function of D_U to emphasize its dependency. Sometimes the function notation is dropped if the context does not benefit. On a side note, observe that a projection back would lead one back to $\mathbf{A}^T\Sigma_V\mathbf{A} = \hat{\mathbf{D}}_U$.

Lemma 1. *For all $D_U \geq 0$, the eigenvalues of Σ_V satisfy*

$$\lambda_i(\Sigma_V) \geq \lambda_{i+M}(\Sigma_X) \qquad \text{for } 1 \leq i \leq N - M \tag{26}$$

$$\lambda_i(\Sigma_V) \geq 0 \qquad \text{for } N - M < i \leq N. \tag{27}$$

Proof. Observe that (25) is a subtraction of two real symmetric matrices. The subtracted matrix is of at most rank $M < N$. Consequently, we can apply a theorem by Weyl [7, Thm 4.3.6] that for any two $N \times N$ Hermitian matrices A, B, of which $\text{rank}(B) \leq M$, the following holds:

$$\lambda_k(A) \geq \lambda_{k+M}(A + B) \geq \lambda_{k+2M}(A). \tag{28}$$

Let $A = \Sigma_X$ and pick for B the perturbation matrix on the right of (25), including the minus sign. Conclude that the update matrix was positive semidefinite by construction and is thus now negative semidefinite if one indeed includes the subtraction as its sign. All $\lambda_i(\Sigma_V)$ can thus only decrease with respect to $\lambda_i(\Sigma_X)$, but Weyl's theorem now implies that in any case they can never drop below $\lambda_{i+M}(\Sigma_X)$. Hence, (26) holds.

The second line of the lemma is somewhat trivial, since Σ_V is a covariance matrix. However, to prove that this is indeed true, one could apply an argument similar to that of [8, Thm 1]. Due to space limitations, we leave it to a reference for now. □

5 Rates and Successive Refinability

Lemma 1 and (25) show that the eigenvalues of Σ_V are strictly non-decreasing in D_U. An example is drawn in the left plot of Figure 2. A trivial observation is that if $D_U \geq \mathrm{tr}(\Sigma_U)$ then $R_U(D_U) = 0$, implying that $\Sigma_V = \Sigma_X$; the entire first stage is skipped and the 'refinement' is the one-stage RD-coding problem. Furthermore, note that $\Sigma_X \succeq \Sigma_V$ and thus by the fact that both X and V are Gaussian, we have

$$R_V(D_X) \leq R_X(D_X), \tag{29}$$

with equality if and only if $R_U = 0$. Furthermore, the two-stage refinement can never produce a sum-rate lower than the one-step optimal coding of $R_X(D_X)$. Combining the latter observation with (29), we find the following bounds:

$$\max\{R_X(D_X), R_U(D_U)\} \leq R_U(D_U) + R_V(D_X) \leq R_U(D_U) + R_X(D_X). \tag{30}$$

An example is shown in the center plot of Figure 2. Evaluating the sum-rate is more convenient than one might expect, as becomes clear from the following lemma:

Lemma 2. * *For $\mathbf{0} \preceq \hat{\mathbf{D}}_U \preceq \Sigma_U$, it holds that*

$$\frac{1}{2}\log\left(\frac{|\Sigma_U|}{|\hat{\mathbf{D}}_U|}\right) = \frac{1}{2}\log\left(\frac{|\Sigma_X|}{|\Sigma_V|}\right). \tag{31}$$

Proof. Let us start with the determinant of the residual Σ_V by making use of (25).

$$|\Sigma_V| = |\Sigma_X - \Sigma_X \mathbf{A}\Sigma_U^{-2}(\Sigma_U - \hat{\mathbf{D}}_U)\mathbf{A}^T\Sigma_X| \tag{32}$$

$$= |\Sigma_X| \cdot |\mathbf{I}_N - \Sigma_X^{1/2}\mathbf{A}\Sigma_U^{-2}(\Sigma_U - \hat{\mathbf{D}}_U)\mathbf{A}^T\Sigma_X^{1/2}|. \tag{33}$$

The rest of the proof relies on Sylvester's theorem for determinants. It states that for any $\mathbf{P} \in \mathbb{R}^{M\times N}$ and $\mathbf{Q} \in \mathbb{R}^{N\times M}$ it holds that $|\mathbf{I}_M + \mathbf{PQ}| = |\mathbf{I}_N + \mathbf{QP}|$ (see, .e.g, [9]). Now, continue expanding the right hand side

$$\frac{|\Sigma_V|}{|\Sigma_X|} = |\mathbf{I}_N \underbrace{-\Sigma_X^{1/2}\mathbf{A}\Sigma_U^{-1}}_{\mathbf{Q}} \underbrace{\Sigma_U^{-1}(\Sigma_U - \hat{\mathbf{D}}_U)\mathbf{A}^T\Sigma_X^{1/2}}_{\mathbf{P}}| \tag{34}$$

$$= |\mathbf{I}_M - \Sigma_U^{-1}(\Sigma_U - \hat{\mathbf{D}}_U)\underbrace{\mathbf{A}^T\Sigma_X^{1/2}\Sigma_X^{1/2}\mathbf{A}\Sigma_U^{-1}}_{=\mathbf{I}_m}| \tag{35}$$

$$= |\Sigma_U^{-1}| \cdot |\hat{\mathbf{D}}_U|. \tag{36}$$

Since the arguments of the logs in the lemma are equal, so are the logs. □

This Lemma turns out to be the crucial tool to comment on successive refinability. To that end, let us zoom in on a specific region of distortions, namely $D_X < N\lambda_N(\Sigma_X)$. For these D_X, we are ready to prove the absence of rate loss for any $\mathbf{A}$. The case of larger D_X requires some care, which we will explain after the following theorem:

*The authors would like to thank Stefan Apostol for his substantial contribution to Lemma 2.

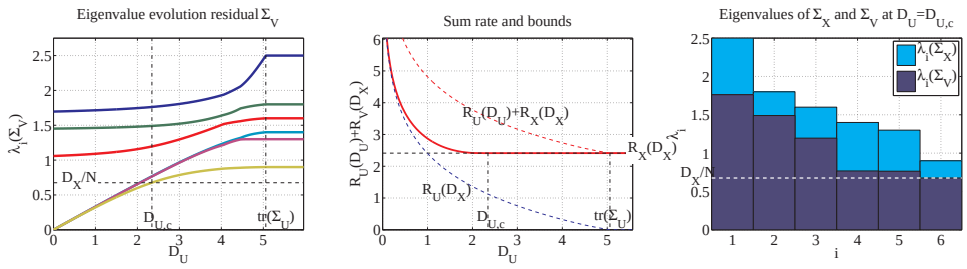

Figure 2: Example for $(N, M) = (6, 3)$ for $\Sigma_X = \text{diag}(2.5, 1.8, 1.6, 1.4, 1.3, 0.9)$ and $\frac{D_X}{N} = 0.675$. Left: $\lambda_i(\Sigma_V)$ are non-decreasing in D_U by Lemma 1. Middle: The sum-rate is bounded as in (30) and is equal to $R_X(D_X)$ for $D_U \geq D_{U,c}$. Right: Eigenvalue drop for $\Sigma_X \to \Sigma_V$ at the critical point $D_U = D_{U,c} \approx 2.36$.

Theorem 2. *For sufficiently small distortion, i.e., $D_X \leq N\lambda_N(\Sigma_X)$, a Gaussian source X is successively refinable after a projection $\mathbf{A}$ if and only if the projection is requested at a $D_U \geq D_{U,c}$ for a critical $D_{U,c}$ that satisfies*

$$\lambda_N(\Sigma_V(D_{U,c})) = \frac{D_X}{N}, \tag{37}$$

where $\Sigma_V(D_{U,c})$ follows (25).

Proof. The theorem says that one must pick a D_U so that no eigenvalue of the residual Σ_V drops below the distortion threshold that is set by reversed water filling for the optimal one-stage coding of $R_X(D_X) = \frac{1}{2}\log\left(\frac{|\Sigma_X|}{(D_X/n)^n}\right)$. Assume for convenience that all the following $\hat{\mathbf{D}}$ are already fixed so that rates are minimized. Then by Lemma 2,

$$R_U(D_U) + R_V(D_X) = \frac{1}{2}\log\left(\frac{|\Sigma_U|}{|\hat{\mathbf{D}}_U|}\right) + \frac{1}{2}\log\left(\frac{|\Sigma_V|}{|\hat{\mathbf{D}}_V|}\right) \tag{38}$$

$$= \frac{1}{2}\log\left(\frac{|\Sigma_X|}{|\Sigma_V|}\right) + \frac{1}{2}\log\left(\frac{|\Sigma_V|}{|\hat{\mathbf{D}}_V|}\right) \tag{39}$$

$$= \frac{1}{2}\log\left(\frac{|\Sigma_X|}{|\hat{\mathbf{D}}_V|}\right). \tag{40}$$

If $\lambda_N(\Sigma_V) \geq \frac{D_X}{N}$ then $\hat{D}_{V,i} = \frac{D_X}{N}$ for all $i = 1, \cdots, N$ by the reversed water filling procedure. That means that $\hat{\mathbf{D}}_V = \hat{\mathbf{D}}_X$ and thus $R_U(D_U) + R_V(D_X) = R_X(D)$. For the converse assume that D_U is picked so small that $\lambda_N(\Sigma_V) < \frac{D_X}{N}$. Then the reversed water filling procedure will assign individual distortion levels,

$$|\hat{\mathbf{D}}_V| = \prod_{i=1}^{N} \min(\theta_V, \lambda_i(\Sigma_V)), \tag{41}$$

where θ_V is picked such that $\sum_{i=1}^{N} \min(\theta_V, \lambda_i(\Sigma_V)) = D_X$. This product is maximized by $\left(\frac{D_X}{n}\right)^N$, which can only be a solution if all $\lambda_i(\Sigma_V) \geq \frac{D_X}{N}$, contradicting the assumption. We have $|\hat{\mathbf{D}}_V| < |\hat{\mathbf{D}}_X|$ otherwise, resulting in a sum-rate strictly higher than $R_X(D_X)$. Since $\lambda_i(\Sigma_V)$ are continuous and non-decreasing in D_U, successive refinability is guaranteed for all D_U larger than the D_U at which $\lambda_N(\Sigma_V(D_U)) = \frac{D_X}{N}$. □

The right plot of Figure 2 shows an example of the eigenvalues at this critical point. Note that everything above the dotted line $\frac{D_X}{N}$ is what $\hat{X}$ would have been by a one-stage optimal coding via $R_X(D_X)$. A too small D_U so that any $\lambda_i(\Sigma_V)$ drops below $\frac{D_X}{N}$ means that $\mathbb{E}[X|\hat{U}]$ contains information on X not present in the $\hat{X}$ that would have been the one-stage optimal solution. One can show that if $\mathbf{A}$ consists of any set of M eigenvectors of Σ_X, then $D_{U,c} = \frac{M \cdot D_X}{N}$. It appears that $D_{U,c} > \frac{M \cdot D_X}{N}$ for any other projection matrix, but we are yet to formally show this.

Low distortion, $D_X < N\lambda_N(\Sigma_X)$, ensures that all principal components of X require coding in the first place. Under this condition there always exists a $D_{U,c} < \mathrm{tr}(\Sigma_U)$ for *any* $\mathbf{A}$. Were the condition not true, then $\mathbf{A}$ could have been aligned in the direction of components that are not part of the one-stage optimal description of X. One could thus, for example, exclude these directions from $\mathbf{A}$ to also achieve successive refinability for $D_X > N\lambda_N(\Sigma_X)$. In general, the critical point $D_{U,c}$ in Theorem 2 is more formally found by comparing the partaking principal components $\lambda_i(\Sigma_V(D_{U,c}))$ to θ_X, the coding threshold the reversed water filling procedure uses to compute $R_X(D_X)$. All in all, the successive refinability of X is not so much restricted by the dimensionality reduction imposed by $\mathbf{A}$, it is directionality to which it is most sensitive.

References

[1] W. Equitz and T. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 269–275, 1991.

[2] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *Information Theory, IEEE Transactions on*, vol. 40, no. 1, pp. 253–259, 1994.

[3] J. Nayak and E. Tuncel, "Successive coding of correlated sources," *Information Theory, IEEE Transactions on*, vol. 55, no. 9, pp. 4286–4298, 2009.

[4] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *Information Theory, IEEE Transactions on*, vol. 46, no. 1, pp. 236–246, Jan 2000.

[5] J. Nayak, E. Tuncel, D. Gunduz, and E. Erkip, "Successive refinement of vector sources under individual distortion criteria," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1769–1781, 2010.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory (2. ed.)*. Wiley, 2006, pp. 301–346.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, p. 184.

[8] M. Zhu and N. Deng, "On the symmetric rank-m correction of matrices," *CALCOLO*, vol. 26, no. 1, pp. 33–39, 1989.

[9] A. G. Akritas, E. K. Akritas, and G. I. Malaschonok, "Various proofs of sylvester's (determinant) identity," *Mathematics and Computers in Simulation*, vol. 42, no. 46, pp. 585 – 593, 1996.