Source Coding with Side Information

Presenter: Stark Draper
Date: 15 February 2001
Reading: Cover & Thomas *Elements of Information Theory* [6]
Sections 14.4 (optional), 14.8 & 14.9

1 Introduction

Source coding with side information is one subset of a larger set of distributed source coding problems. These problems consider the encoding a random source signal, $\mathbf{x} = x_1^n = [x_1 x_2 \ldots x_n]^T$, into a codeword where the decoder has access both to the codeword, and to a second source of correlated information.[1] In certain cases this second source is referred to as "side information". In Figure 1 and Table 1 we outline the relation between various distributed source-coding problems and give references. We include the standard single-source lossless and lossy compression problems in the first two rows of Table 1 for reference. After describing the side-information set-up in detail, we will relate it to the other problems indicated in Table 1.

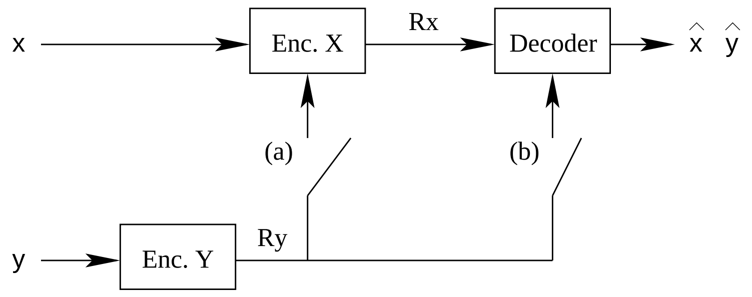

Figure 1: Distributed Source Coding with Two Switches, (a) and (b)

In side-information source coding the goal of the decoder is to produce an estimate of $\mathbf{x}$, namely $\hat{\mathbf{x}}$. The side information provides knowledge about a correlated source $\mathbf{y}$. The decoder is not concerned with estimating $\mathbf{y}$, and so knowledge of $\mathbf{y}$ is useful only to the extent that it allows us to reduce the rate R_x. One possibility is that $R_y \geq H(y)$, in which case the side information is equivalent to a direct observation of $\mathbf{y}$ (this is the case in lossy side information, C&T sec. 14.9)[2]. The other possibility is that $R_y \leq H(y)$, in which case

[1] In this summary we will use the following notation for random variables: x – random variable, $\mathbf{x}$ – random vector, x – sample value, $\mathbf{x}$ – vector of sample values. In general we will suppress the subscripts of densities, i.e. $p_x(x) = p(x)$.

[2] "Lossy" means that at the decoder we want an imperfect reproduction of $\mathbf{x}$, i.e. $E[d(\mathbf{x}, \hat{\mathbf{x}})] < D_x + \epsilon$.

	R_y	(a)	(b)	D_x	D_y	References
Lossless source coding (of $\mathbf{x}$)	n/A	○	○	0	N/A	[6]
Lossy source coding (of $\mathbf{x}$)	N/A	○	○	≥ 0	N/A	[6]
Slepian-Wolf coding	$\leq H(y)$	○	●	0	0	[6, 9]
Lossy multiterminal source coding	$\leq H(y)$	○	●	≥ 0	≥ 0	[15]
Conditional Rate-Distortion	$\geq H(y)$	●	●	≥ 0	∞	[7]
Lossless Side-Info	$\leq H(y)$	○	●	0	∞	[6, 10]
Lossy Side-Info	$\geq H(y)$	○	●	≥ 0	∞	[6, 13]

Table 1: *Relationship between problems and parameters of Figure 1. Column 1 is the problem name. Column 2 indicates whether the y-encoder can losslessly communicate* $\mathbf{y}$ *to the decoder* ($R_y \geq H(y)$)*. This is equivalent to the decoder observing* $\mathbf{y}$ *directly. If* $R_y \leq H(y)$ *the y-encoder cannot losslessly communicate* $\mathbf{y}$ *to the decoder. Columns 3 and 4 indicate whether the respective switches are open (○) or closed (●). Columns 5 and 6 indicate the distortion requirements on the decoded source estimates. A non-zero value indicates a rate-distortion (lossy) problem.* $D_y = \infty$ *means that the decoder is not concerned with estimating this variable. Column 7 indicates the appropriate references.*

the side information contains only indirect knowledge of y (this is the case in lossless side information, C&T sec. 14.8)[3]. In either case, generally $p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} p_{x,y}(x_i, y_i)$; the pairs of random variable are independent and identically distributed (i.i.d).

The basic idea of source coding with side information is only to encode the uncertainty in $\mathbf{x}$ given the knowledge you expect to get about $\mathbf{x}$ from $\mathbf{y}$. The main theoretical question is "What rate R_x must I encode at to achieve some specified distortion D_x?" The main practical question is "How do I do this?" We will touch on both questions here and in class.

The lossless side information problem was introduced by Wyner [10]. The lossy side-information problem was introduced by Wyner and Ziv [13]. Our in-class discussion of technical matters (i.e. proofs) will focus mainly on the lossless scenario. We will discuss some of the intuition and application of the lossy results, but will not delve too deeply into technical matters. Although our discussion in class (and in Cover & Thomas[4]) will focus on discrete sources, the results have been extended to continuous sources. In Section 4, below, we give the achievable rate region for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ which was found by Wyner in [12].

The side-information work of Wyner and Ziv builds upon the theoretical work of Slepian and Wolf in lossless multi-terminal source coding, see [9] and C&T sec. 14.4. (Recent work has generalized Slepian-Wolf coding to lossy multi-terminal source coding [15].) Also relevant

[3] "Lossless" means that at the decoder we want a perfect reproduction of $\mathbf{x}$, i.e. $E[d(\mathbf{x},\hat{\mathbf{x}})] < \epsilon$.

[4] See Blahut [3] for another good introduction to these problems and network information theory.

to Wyner and Ziv's work was that of Gray in conditional rate-distortion theory [7]. For ideas on how to implement Slepian-Wolf and Side-Information encoders see [11] and [1, 8, 14, 16], respectively.

Recently there has been a reawakening of interest in the side-information problem, particularly concerning its connections to watermarking, e.g. see [1, 2, 4, 5]. We will point out the duality of these two problems, which Aaron Cohen will discuss further in his talk on 2/22.

2 Lossless Side-Information

Theorem 1 (Lossless Source Coding with Side-Information) *Given a pair of sources* x *and* y *jointly distributed* $p(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} p(x_i, y_i)$*:*

If (a) x *is encoded at rate* R_{x}*,*
(b) y *is encoded at rate* R_{y}*,*

Then we can recover x *with arbitrarily small* $\Pr(\mathsf{error})$ *iff we can find an auxiliary random variable* u *s.t.:*

(i) $R_{\mathsf{x}} \geq H(\mathsf{x}|\mathsf{u})$,

(ii) $R_{\mathsf{y}} \geq I(\mathsf{y};\mathsf{u})$,

(iii) $\mathsf{x} \rightarrow \mathsf{y} \rightarrow \mathsf{u}$ *(Markov).*

The encoding technique is first to bin randomly all typical x sequences into $2^{nR_{\mathsf{x}}} \leq 2^{nH(\mathsf{x})}$ bins. These are the sequence you want to be able to identify at the decoder. The bin index in which the realized source sequence falls, say bin i, is then transmitted to the decoder at rate R_{x}. This bin contains many typical x sequences (about $2^{n(H(\mathsf{x})-R_{\mathsf{x}})} \simeq 2^{nI(\mathsf{x};\mathsf{u})}$ of them). The side-information (available at rate R_{y}) is used to select which typical source sequence within bin i was the realized sequence. (Note that $R_{\mathsf{y}} \geq I(\mathsf{y};\mathsf{u}) \geq I(\mathsf{x};\mathsf{u})$ by the Markov property and the Data Processing Theorem, so the side-information rate is high enough to do the intra-bin selection.) This selection is done via joint typicality arguments, and relies on one new tool, the Markov Lemma.

An interesting case to note is what happens if we choose the auxiliary random variable $\mathsf{u} = \mathsf{y}$. Then we get $R_{\mathsf{x}} \geq H(\mathsf{x}|\mathsf{y})$, $R_{\mathsf{y}} \geq I(\mathsf{y};\mathsf{y}) = H(\mathsf{y})$. I.e. y can be perfectly transmitted and only the residual randomness in x needs be sent. This is a special case of Slepian-Wolf coding. Because we do not care about decoding y in the side-information problem, we do not generally communicate y to the decoder perfectly. The side-information the decoder has, u, is at best as informative as y, hence this gives us a lower bound to R_{x}, i.e. $R_x \simeq H(x|u) \geq H(x|y)$. This lower bound can also be seen via the Data-Processing Inequality.

3 Lossy Side-Information

Theorem 2 (Lossy Source Coding with Side-Information) *Given a source* $\mathbf{x}$*, side information* $\mathbf{y}$*, and a distortion measure* $\mathsf{d}(\cdot,\cdot)$ *s.t.:*

(a) $p(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} p(x_i, y_i)$,

(b) $d(\mathbf{x},\hat{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i)$,

Then, if $\mathbf{x}$ *is encoded at rate* R_x*, we can recover* x *to within distortion* D_x *with arbitrarily small* $\Pr(\mathsf{failure})$ *iff*

(i) $R_x \geq R^{WZ}_{x|y}(D_x) = \min_{p(w|x)} \min_f (I(x;w) - I(y;w))$

(ii) $y \to x \to w$ *(Markov),*

(iii) $E\left[d(x, f(y,w))\right] \leq D_x$.

We use the notation $R^{WZ}_{x|y}(D_x)$ to refer to the Wyner-Ziv rate-distortion function, to differentiate it from the conditional rate-distortion function of Gray that is often written as $R_{x|y}(D_x)$.

The encoding technique here is similar to the lossless case, except that now instead of working to communicate the actual (typical) $\mathbf{x}$ sequence realized, we communicate a reduced-fidelity ($D_x > 0$) approximation $\mathbf{w}$ instead. This lowers the needed transmission rate R_x.

First we generate a codebook of about $2^{nI(x;w)}$ $\mathbf{w}$-sequences that well approximate the typical $\mathbf{x}$ sequences to distortion D_x. These are the $\mathbf{w}$-sequences we want to be able to identify at the decoder. Next, the codebook sequences are binned randomly into $2^{nR_x} \leq 2^{nI(x;w)}$ bins. The bin index in which the $\mathbf{w}$ sequence that well-approximates the realized $\mathbf{x}$ source sequence falls, say bin i, is then transmitted to the receiver at rate R_x. This bin contains many $\mathbf{w}$ sequences (about $2^{n(I(x;w)-R_x)} \simeq 2^{nI(y;w)}$ of them). The side-information is used to select which $\mathbf{w}$-sequence within bin i is the one that well approximates the realized sequence $\mathbf{x}$. This selection is done via joint typicality arguments, and the Markov Lemma.

Putting the Markov relationship ($y \to x \to w$) together with Fano's Inequality we can lower-bound $R_{x|y}(D_x)$ as: $I(x;w) - I(y;w) = H(w|y) - H(w|x) = H(w|y) - H(w|x,y) = I(x;w|y) \geq H(x|y) - \epsilon(D_x) \to H(x|y)$ as $D_x \to 0$, which brings us back to the Slepian-Wolf results.

4 The Gaussian Source for MSE Distortion, and Practical Constructions

Wyner determined the rate-region for the Gaussian case with mean square error as the distortion metric [12]. He showed that for a jointly Gaussian zero-mean source $p(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} p(x_i, y_i)$

$$R_{\mathsf{x}|\mathsf{y}}^{WZ}(D_{\mathsf{x}}) = \begin{cases} \frac{1}{2} \log \frac{\lambda_{\mathsf{x}|\mathsf{y}}}{D_{\mathsf{x}}}, & 0 \leq D_{\mathsf{x}} \leq \lambda_{\mathsf{x}|\mathsf{y}} \\ 0, & \lambda_{\mathsf{x}|\mathsf{y}} < D_{\mathsf{x}} \end{cases}$$

where $\lambda_{\mathsf{x}|\mathsf{y}}$ is the Bayesian estimation error:

$$\begin{aligned} \lambda_{\mathsf{x}|\mathsf{y}} &= E\left[(\hat{\mathsf{x}}(\mathsf{y}) - \mathsf{x})^2\right] \\ &= E\left[\mathsf{x}^2\right] - \frac{E\left[\mathsf{x}\mathsf{y}\right]^2}{E\left[\mathsf{y}^2\right]}. \end{aligned}$$

Interestingly, in this case $R_{\mathsf{x}|\mathsf{y}}^{WZ}(D_{\mathsf{x}}) = R_{\mathsf{x}|\mathsf{y}}(D_{\mathsf{x}})$, the conditional rate-distortion function where $\mathbf{y}$ is known at both encoder and decoder. The rate-distortion region is also known for discrete binary-symmetric sources (DSBS). In the Gaussian case the rate-distortion limit can be achieved via nested lattice quantizers [1, 16]. In the DSBS case, the rate-distortion limit can be achieved via nested linear codes [8].

References

[1] R. J. Barron. *Systematic Hybrid Analog/Digital Signal Coding.* PhD thesis, Mass. Instit. of Tech., 2000.

[2] R. J. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and its implications and applications. *Submitted to IEEE Trans. Info. Theory*, January 2000.

[3] R. E. Blahut. *Principles and Practice of Information Theory.* Addison-Wesley, 1987.

[4] B. Chen. *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems.* PhD thesis, Mass. Instit. of Tech., 2000.

[5] M. H. Costa. Writing on dirty paper. *IEEE Trans. Info. Theory*, 29:439–441, May 1983.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, 1991.

[7] R. M. Gray. Conditional rate-distortion theory. Technical report, Stanford Electronics Laboratories, No. 6502-2, 1972.

[8] S. Shamai, S. Verdu, and R. Zamir. Systematic lossy source/channel coding. *IEEE Trans. Info. Theory*, 44:564–579, March 1998.

[9] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Info. Theory*, 19:471–480, July 1973.

[10] A. D. Wyner. A theorem on the entropy of certain binary sequences and applications ii. *IEEE Trans. Info. Theory*, 19:772–777, November 1973.

[11] A. D. Wyner. Recent results in the shannon theory. *IEEE Trans. Info. Theory*, 20:2–10, January 1974.

[12] A. D. Wyner. The rate-distortion function for source coding with side information at the decoder–ii: General sources. *Information and Control*, 38:60–80, 1978.

[13] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory*, 22:1–10, January 1976.

[14] R. Zamir. The rate loss in the wyner-ziv problem. *IEEE Trans. Info. Theory*, 42:2073–2084, November 1996.

[15] R. Zamir and T. Berger. Multiterminal source coding with high resolution. *IEEE Trans. Info. Theory*, 45:106–117, January 1999.

[16] R. Zamir and S. Shamai. Nested linear/lattice codes for wyner-ziv encoding. In *1998 Information Theory Workshop, Kilarney, Ireland*, pages 92–93, June 1998.