Nonlinear Information-theoretic Compressive Measurement Design: Supplementary Material

Liming Wang[1], Abolfazl Razi[1], Migurel Dias Rodrigues[2], Robert Calderbank[1] and Lawrence Carin[1]

[1]Department of Electrical & Computer Engineering, Duke University, Durham, NC 27708, USA
[2]Department of Electronic and Electrical Engineering, University College London, London, UK

1 Proof of Theorem 1

Proof of Theorem 1. The gradient of mutual information $\nabla_A I(X;Y)$ can be expressed as

$$\nabla_A I(X;Y) = \nabla_A[h(Y) - h(Y|X)] \tag{1}$$
$$= \nabla_A h(Y) \tag{2}$$

The equality follows from the fact that $h(Y|X) = \mathbb{E}_X(h(\mathcal{N}(Y; A\psi(X), \Sigma)))$, which is a constant with respect to A. Hence, we have

$$\nabla_A I(X;Y) = \nabla_A h(Y) \tag{3}$$
$$= -\nabla_A \int \log P_Y(y) P_Y(y) dy \tag{4}$$
$$= -\int \nabla_A P_Y(y) dy - \int \log P_Y(y) \nabla_A P_Y(y) dy \tag{5}$$
$$= -\nabla_A \int P_Y(y) dy - \int \log P_Y(y) \nabla_A P_Y(y) dy \tag{6}$$
$$= -\int \log P_Y(y) \nabla_A P_Y(y) dy, \tag{7}$$

where the change of operator orders in (5) and (6) follows from the assumed regularity conditions. We now calculate the term $\nabla_A P_Y(y)$

$$\nabla_A P_Y(y) = \nabla_A \int P_{Y|X}(y|x) P_X(x) dx \tag{8}$$
$$= \int P_X(x) \nabla_A P_{Y|X}(y|x) dx \tag{9}$$
$$= \int P_X(x) \nabla_A \mathcal{N}(y; A\psi(x), \Sigma) dx \tag{10}$$
$$= \int P_X(x) \nabla_y \mathcal{N}(y; A\psi(x), \Sigma) \psi^T(x) dx \tag{11}$$
$$= \int P_X(x) \nabla_y P_{Y|X}(y|x) \psi^T(x) dx, \tag{12}$$

where (11) follows from the chain rule. Plug in (12) back to (7) and by the Fubini's Theorem [1], together with the regularity conditions, we have

$$\nabla_A I(X;Y) = -\int \log P_Y(y) P_X(x) \nabla_y P_{Y|X}(y|x) \psi^T(x) dx dy \tag{13}$$

$$= -\int \nabla_y (\log P_Y(y)) P_X(x) P_{Y|X}(y|x) \psi^T(x) dx dy \tag{14}$$

$$= -\int \frac{P_X(x) P_{Y|X}(y|x)}{P_Y(y)} \nabla_y P_Y(y) \psi^T(x) dx dy \tag{15}$$

$$= -\int \nabla_y P_Y(y) \mathbb{E}[\psi^T(X)|Y = y] dy, \tag{16}$$

where (14) invokes the integration by parts [1]. Again via the assumed regularity conditions, $\nabla_y P_Y(y)$ can be expressed as

$$\nabla_y P_Y(y) = \int \nabla_y P_{Y|X}(y|x) P_X(x) dx \tag{17}$$

$$= \int \nabla_y \log P_{Y|X}(y|x) P_{Y|X}(y|x) P_X(x) dx \tag{18}$$

$$= \int \nabla_y \log \mathcal{N}(y; A\psi(x), \Sigma) P_{X|Y=y}(x|y) P_Y(y) dx \tag{19}$$

$$= \mathbb{E}[(\nabla_Y \log \mathcal{N}(Y; A\psi(X), \Sigma))|Y = y] P_Y(y). \tag{20}$$

It is straightforward to calculate

$$\nabla_y \log \mathcal{N}(y; A\psi(x), \Sigma) = -\Sigma^{-1}(y - A\psi(x)). \tag{21}$$

Hence, we have

$$\nabla_y P_Y(y) = \mathbb{E}[-\Sigma^{-1}(Y - A\psi(X))] P_Y(y) \tag{22}$$

$$= -\Sigma^{-1} \mathbb{E}[W|Y = y] P_Y(y). \tag{23}$$

Combining (23) with (16), we obtain

$$\nabla_A I(X;Y) = \int \Sigma^{-1} \mathbb{E}[W|Y = y] P_Y(y) \mathbb{E}[\psi^T(X)|Y = y] dy \tag{24}$$

$$= \Sigma^{-1} \mathbb{E}[\mathbb{E}[W|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{25}$$

$$= \Sigma^{-1} \mathbb{E}[\mathbb{E}[Y - A\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{26}$$

$$= \Sigma^{-1} \mathbb{E}[\mathbb{E}[Y\psi^T(X)|Y]] - A\mathbb{E}[\mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{27}$$

$$= \Sigma^{-1} \mathbb{E}[\mathbb{E}[(A\psi(X) + W)\psi^T(X)]] - A\mathbb{E}[\mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{28}$$

$$= \Sigma^{-1} \mathbb{E}[\mathbb{E}[A\psi(X)\psi^T(X)]] - A\mathbb{E}[\mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{29}$$

$$= \Sigma^{-1} A\mathbb{E}[\mathbb{E}[\psi(X)\psi^T(X)]] - \mathbb{E}[\mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y]] \tag{30}$$

$$\begin{aligned} &= \Sigma^{-1} A\mathbb{E}[\mathbb{E}[\psi(X)\psi^T(X)]] - \mathbb{E}[\psi(X) \mathbb{E}[\psi^T(X)|Y] \\ &- \mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)]] + \mathbb{E}[\psi(X)|Y] \mathbb{E}[\psi^T(X)|Y] \end{aligned} \tag{31}$$

$$= \Sigma^{-1} A\mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y]][\psi(X) - \mathbb{E}[\psi(X)|Y]]^T]. \tag{32}$$

Note that we do not assume specific form of $\psi(x)$ in previous proof, thus it applies to arbitrary basis vector $\psi(x)$, provided that it satisfies the regularity conditions. □

2 Proof of Theorem 2

Proof of Theorem 2. First we notice that

$$\begin{aligned} I(C;Y) &= h(Y) - h(Y|C) && (33)\\ &= h(Y) - h(Y|X) + h(Y|X,C) - h(Y|C) && (34)\\ &= I(X;Y) - I(X;Y|C), && (35) \end{aligned}$$

where the second equality is due to the fact that $C \to X \to Y$ forms a Markov chain and $P_{Y|X,C} = P_{Y|X}$. Following by the similar steps in the proof of Theorem 1, we have

$$\nabla_A I(X;Y|C) = \Sigma^{-1} A\mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y,C]][\psi(X) - \mathbb{E}[\psi(X)|Y,C]]^T]. \quad (36)$$

Using the following facts that

$$\begin{aligned} \mathbb{E}[\mathbb{E}[\psi(X)|Y,C]\mathbb{E}[\psi(X)^T|Y,C]] &= \mathbb{E}[\mathbb{E}[\psi(X)|Y,C]\psi(X)^T] = \mathbb{E}[\psi(X)\mathbb{E}[\psi(X)^T|Y,C]] && (37)\\ \mathbb{E}[\mathbb{E}[\psi(X)|Y,C]\mathbb{E}[\psi(X)^T|Y]] &= \mathbb{E}[\mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y,C]] = \mathbb{E}[\mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y]] && (38)\\ \mathbb{E}[\psi(X)\mathbb{E}[\psi(X)^T|Y]] &= \mathbb{E}[\mathbb{E}[\psi(X)|Y]\psi(X)^T] = \mathbb{E}[\mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y]], && (39) \end{aligned}$$

we have the following expressions

$$\begin{aligned} &\mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y]][\psi(X) - \mathbb{E}[\psi(X)|Y]]^T]\\ &= \mathbb{E}[\psi(X)\psi(X)^T - \psi(X)\mathbb{E}[\psi(X)^T|Y] - \mathbb{E}[\psi(X)|Y]\psi(X)^T + \mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y]] && (40)\\ &= \mathbb{E}[\psi(X)\psi(X)^T - \psi(X)\mathbb{E}[\psi(X)^T|Y,C] - \mathbb{E}[\psi(X)|Y,C]\psi(X)^T + \mathbb{E}[\psi(X)|Y,C]\mathbb{E}[\psi(X)^T|Y,C]\\ &+ \mathbb{E}[\psi(X)|Y,C]\mathbb{E}[\psi(X)^T|Y,C] - \mathbb{E}[\psi(X)|Y,C]\mathbb{E}[\psi(X)^T|Y] - \mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y,C]\\ &+ \mathbb{E}[\psi(X)|Y]\mathbb{E}[\psi(X)^T|Y]] && (41)\\ &= \mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y,C]][\psi(X) - \mathbb{E}[\psi(X)|Y,C]]^T] && (42)\\ &+ \mathbb{E}[[\mathbb{E}[\psi(X)|Y] - \mathbb{E}[\psi(X)|Y,C]][\mathbb{E}[\psi(X)|Y] - \mathbb{E}[\psi(X)|Y,C]]^T]. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \nabla_A I(C;Y) &= \nabla_A I(X;Y) - \nabla_A I(X;Y|C) && (43)\\ &= \mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y]][\psi(X) - \mathbb{E}[\psi(X)|Y]]^T]\\ &- \mathbb{E}[[\psi(X) - \mathbb{E}[\psi(X)|Y,C]][\psi(X) - \mathbb{E}[\psi(X)|Y,C]]^T] && (44)\\ &= \mathbb{E}[[\mathbb{E}[\psi(X)|Y] - \mathbb{E}[\psi(X)|Y,C]][\mathbb{E}[\psi(X)|Y] - \mathbb{E}[\psi(X)|Y,C]]^T]. && (45) \end{aligned}$$

□

References

[1] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications.* Wiley New York, 1999.