# Hypercontractivity, Maximal Correlation and Non-interactive Simulation

Zi Yin
Department of Electrical Engineering
Stanford University
zyin@stanford.edu

Youngsuk Park
Department of Electrical Engineering
Stanford University
youngsuk@stanford.edu

***Abstract*—In this paper, we investigate the non-interactive simulation proposed by Kamath et. al. [3], and establish outer bounds by using the maximal correlation and hypercontractivity ribbon. For several examples, inner bounds are discussed and explicit schemes are also introduced.**

***Index Terms*—Hypercontractivity, maximal correlation, Non-Interactive Simulation**

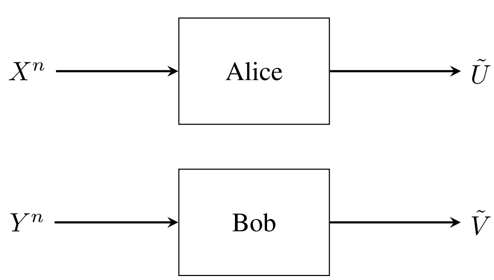
## I. Introduction

In information theory, it is important to find the behavior of a pair of sets $\mathcal{A} \subset \mathcal{X}^n, \mathcal{B} \subset \mathcal{Y}^n$ such that $\mathbb{P}\{Y^n \in \mathcal{B} \mid X^n \in \mathcal{A}\} \geq 1 - \epsilon$. In [1], The authors investigated its behavior and used it in proving strong converses of source coding problem with side information and also the coding problem for degraded broadcast channel (BC). Later in [2], the authors found the connection between the behavior of such sets under $\epsilon \to 0$ and hypercontractivity ribbon, where this naming is inherited from theoretical physics. Also, they formularized the term as the ratio of KL-Divergences. Recently in [3], Anantharam et al. derived another formularization as the ratio of mutual information, and shed light on their operational meaning.

The rest of paper is organized as follow In section 2 of this paper, non-interactive simulation model is revisited. Definitions and properties of maximal correlation and hypercontractivity are briefly discussed in section 3. In section 4, necessary and sufficient condition of several examples are explored.

## II. Problem Setup: Non-Interactive Simulation

Suppose Alice and Bob want to simulate a jointly distribute random variable $(U, V)$ independently. Alice observes a sequence of random variables $X^n$ and Bob observes $Y^n$, where $(X^n, Y^n) \overset{\text{iid}}{\sim} \mathbb{P}_{X,Y}$.




**Non-Interactive Simulation:** If for any $\epsilon > 0$, there exist deterministic functions $\phi, \psi$ on $\mathcal{X}^n, \mathcal{Y}^n$, such that $d_{TV}((\tilde{U}, \tilde{V}), (U, V)) < \epsilon$ where $(\tilde{U}, \tilde{V}) = (\phi(X^n), \psi(Y^n))$. This definition is adopted from Kamath and Anantharam [3]. The total variation distance between two distributions is defined as

$$d_{TV}(U, V) = \sup\{|\mathbb{P}_U(A) - \mathbb{P}_V(A)| : A \in \mathcal{F}\}$$

Note here that the simulation problem is not exact, but to make their distributions as close as possible. Also, different from the problem studied by Gács and Köner [4], this scenario is rateless; we are simulating a single pair of $(U, V)$ using possibly many copies of $(X, Y)$.

### A. Multi-letter Characterization

**Theorem:** Non-interactive simulation is achievable if and only if $(U, V)$ is in the closure of the set

$$\{(\tilde{U}, \tilde{V}) : \tilde{U} - X^n - Y^n - \tilde{V} \text{ forms a Markov chain for some } n\}$$

## III. Backgraound: Hypercontractivity and Maximal Correlation

### A. Motivation

The multi-letter characterization II-A is not tractable, hence we are looking of a outer bound of achievability, which is easy to check. By continuity and the data processing inequality of mutual information, the closure of the set

$$\{(\tilde{U}, \tilde{V}) : I(\tilde{U}; \tilde{V}) \leq I(X^n; Y^n)\}$$

will give an outer bound. However, this outer bound is trivial, as when $n \to \infty$, $I(X^n; Y^n) = nI(X; Y)$ is either 0 or $\infty$. So this motivates us to find other measurements of correlation $T(\cdot, \cdot)$, with the two desired properties:

- Data processing: if $U = f(X), V = g(Y)$,
$$T(U; V) \leq T(X; Y);$$
- Tensorization: if $(X^n, Y^n) \overset{\text{iid}}{\sim} \mathbb{P}_{X,Y}$,
$$T(X^n; Y^n) = T(X; Y).$$

### B. Maximal Correlation

**Definition:** The *Rényi-Gebelein-Hirschfeld Maximal Correlation* $\rho_m(X;Y)$ is defined to be

$$\rho_m(X;Y) = \sup_{f,g:\mathbb{E}[f(X)]=0,\mathbb{E}[g(Y)]=0} \frac{\mathbb{E}[f(X)g(Y)]}{\sqrt{\mathbb{E}[f(X)^2]\mathbb{E}[g(Y)^2]}}$$

By Cauchy inequality and by setting $f(X) = \mathbb{E}[g(Y)|X]$, we can show the equivalence characterization of $\rho_m$

$$\rho_m(X;Y) = \sup_{g:\mathbb{E}[g(Y)]=0} \frac{||\mathbb{E}[g(Y)|X]||_2}{||g(Y)||_2}$$

**Properties of** $\rho_m(\cdot;\cdot)$**:**

- If $U = f(X), V = g(Y)$, $\rho_m(U;V) \leq \rho_m(X;Y)$. It can be seen immediately from the definition.
- If $(X^n,Y^n) \overset{\text{iid}}{\sim} \mathbb{P}_{X,Y}$, $\rho_m(X^n;Y^n) = \rho_m(X;Y)$. This result is proved by Witsenhausen [3]. When $\mathcal{X},\mathcal{Y}$ are finite, a proof based on the relationship between singular values and tensor product of matrices is obtained by Kumar [7]. A short proof based on [5] is attached in Appendix A.

### C. Hypercontractivity Ribbon

**Definition:** The *Hypercontractivity ribbon* $r_p(X;Y)$ is defined to be

$$r_p(X;Y) = \begin{cases} \inf\{r : \langle f,g\rangle \leq ||f(X)||_q ||g(Y)||_{pr}\}, p \geq 1 \\ \sup\{r : \langle f,g\rangle \geq ||f(X)||_q ||g(Y)||_{pr}\}, p < 1 \end{cases}$$

where $\langle f,g\rangle = \mathbb{E}[f(X)g(Y)]$ and $q$ is the Hölder conjugate of p, i.e. $1/p + 1/q = 1$. In [2] [4] the ribbon is defined as $q_p(X;Y) = pr_p(X;Y)$. By Hölder inequality and by setting $f(X) = \mathbb{E}[g(Y)|X]^{p-1}$, we can show the equivalence characterization of $r_p$

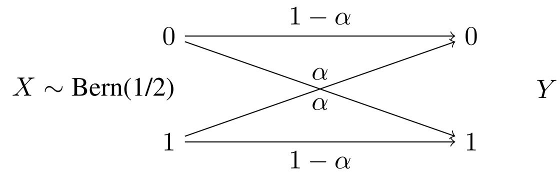
$$r_p(X;Y) = \begin{cases} \inf\{r : ||\mathbb{E}[g(Y)|X]||_p \leq ||g(Y)||_{pr}\}, p \geq 1 \\ \sup\{r : ||\mathbb{E}[g(Y)|X]||_p \geq ||g(Y)||_{pr}\}, p < 1 \end{cases}$$

**Properties of** $s_p(\cdot;\cdot)$**:**

- If $U = f(X), V = g(Y)$,
$$\begin{cases} r_p(U;V) \leq r_p(X;Y), \forall p \geq 1 \\ r_p(U;V) \geq r_p(X;Y), \forall p < 1 \end{cases}$$
Again, this can be seen immediately from the definition.
- If $(X^n,Y^n) \overset{iid}{\sim} (X,Y)$,
$$r_p(X^n;Y^n) = \begin{cases} \max\{r_p(X_i;Y_i)\}, \forall p \geq 1 \\ \min\{r_p(X_i;Y_i)\}, \forall p < 1 \end{cases}$$
A short proof based on [2] is attached in Appendix B.
- For fixed $(X,Y)$, $r_p(X;Y)$ is non-increasing in $p$. This is a direct consequence of Lyapunov's inequality.
- $$\lim_{p\to\infty} \frac{pr_p(X;Y) - 1}{p-1} = \lim_{p\to\infty} r_p(X;Y)$$
exists, and defined to be $s^*(X;Y)$. In [6] they showed
$$\lim_{p\to 1^+} \frac{pr_p(X;Y) - 1}{p-1} = s^*(Y;X).$$

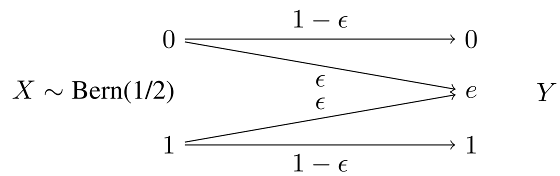
### D. $\rho_m$ and $s^*$ for some special sources

*1) Doubly Symmetric Binary Source (DSBS):* $(X,Y) \sim$ DSBS($\alpha$) if $X \sim Bern(\frac{1}{2})$ and $Y$ is obtained by passing $X$ through a BSC($\alpha$).



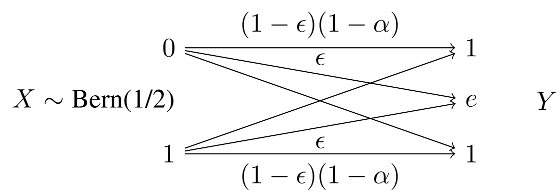

If $(X,Y) \sim$ DSBS($\alpha$), $\rho_m^2(X;Y) = s^*(X;Y) = (1-2\alpha)^2$. The proof is simple calcuclus and can be found in [2].

*2) Symmetric Binary Erasure Source (SEBS):* $(X,Y) \sim$ SEBS($\epsilon$) if $X \sim Bern(\frac{1}{2})$ and $Y$ is obtained by passing $X$ through a BEC($\epsilon$).




If $(X,Y) \sim$ SEBS($\epsilon$), $\rho_m^2(X;Y) = s^*(X;Y) = 1-\epsilon$.

*3) Binary Symmetric Binary Erasure Source (BSBES):* $(X,Y) \sim$ BSBES($\alpha,\epsilon$) if $X \sim Bern(\frac{1}{2})$ and $Y$ is obtained by passing $X$ through a channel indicated below.

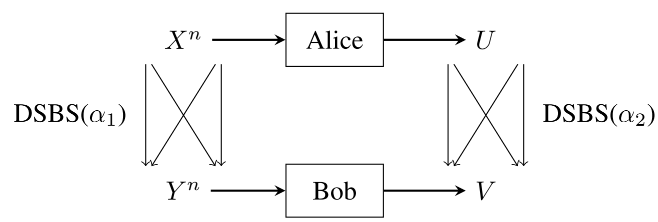



If $(X,Y) \sim$ BSBES($\alpha,\epsilon$), $\rho_m^2(X;Y) = s^*(X;Y) = (1-\epsilon)(1-\alpha)^2$.

## IV. Examples of Non-Interactive Simulation

We investigated several non-interactive simulation scenarios and gave necessary conditions using maximal correlation and hypercontractivity. Sufficient conditions will also be demonstrated using explicit generation schemes.

### A. DSBS($\alpha_1$) Input and DSBS($\alpha_2$) Output



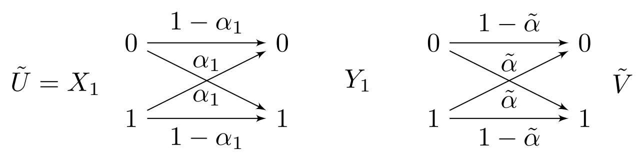

In this example, the source follows DSBS($\alpha_1$) and the output follows DSBS($\alpha_2$), with $\alpha_1, \alpha_2 \leq \frac{1}{2}$.

*1) Necessary Condition:* By tensorization and data processing inequality of maximal correlation,

$$\rho_m^2(X;Y) \geq \rho_m^2(U;V) \Leftrightarrow \alpha_1 \leq \alpha_2$$

is necessary.

*2) sufficient Condition:* We claim that $\alpha_1 \leq \alpha_2$ is also the sufficient condition. Since $\alpha_1 \leq \alpha_2$, there exits $\tilde{\alpha}$, such that $\alpha_1 * \tilde{\alpha} = \alpha_2$. Consider the following scheme



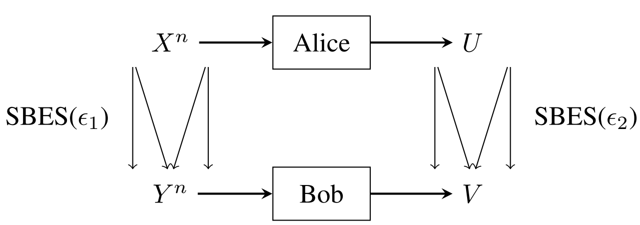

Alice take $\tilde{U} = f(X^n) = X_1$, and Bob passes $Y_1$ through a simulated BSC($\tilde{\alpha}$). Consider

$$S_n = 4\left(\frac{\sum_{i=2}^n Y_i}{\sqrt{n-1}} - \frac{\sqrt{n-1}}{2}\right)$$

By central limit theorem, $S_n \xrightarrow{d} N(0,1)$. Then we set $Z = 0$ wherever $S_n < \Phi^{-1}(\tilde{\alpha})$, and set $Z = 1$ otherwise. Then for such $Z$, let $\tilde{V} = Y_1 \oplus Z$.

### B. Example 2: DSBS Input and SBES Output

The source follows SEBS($\epsilon_1$) and the output follows SEBS($\epsilon_2$).

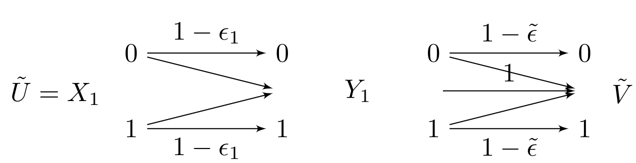



*1) Necessary Condition:*

$$\rho_m^2(X;Y) \geq \rho_m^2(U;V) \Leftrightarrow \epsilon_1 \leq \epsilon_2$$

is necessary for non-interactive simulation.

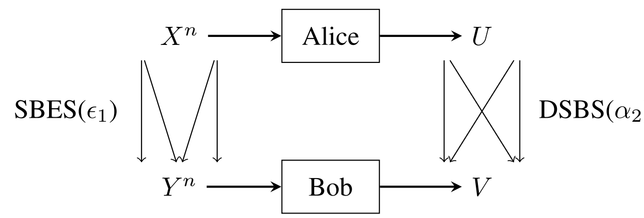
*2) sufficient Condition:* It happens that $\epsilon_1 \leq \epsilon_2$ is sufficient condition too. Note $\tilde{\epsilon} = \epsilon_2 - \epsilon_1 \geq 0$ when $\epsilon_1 \leq \epsilon_2$. Now consider the following scheme




Alice take $\tilde{U} = f(X^n) = X_1$, and Bob passes $Y_1$ through a simulated BEC($\tilde{\epsilon}$). Using $S_n$ defined in Example 1, we set $Z = 0$ wherever $S_n < \Phi^{-1}(\tilde{\epsilon})$. Bob makes an additional erasure whenever $Z = 0$.

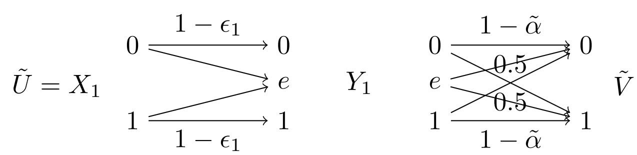
### C. Example 3: SBES Input and DSBS Output

The source follows SEBS($\epsilon_1$) and the output is DSBS($\alpha_2$).




*1) Necessary Condition:* As before, $\rho_m^2(X;Y) \geq \rho_m^2(U;V) \Leftrightarrow \epsilon_1 \leq 4\alpha_2(1-\alpha_2)$ is necessary. Note that this condition matches the channel condition of *less noisy* DM-BC where one channel is BSC($\alpha_2$) and the other channel is BEC($\epsilon_1$).

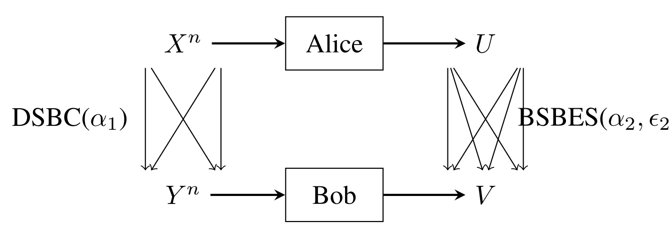
*2) sufficient Condition:* We demonstrate $\epsilon_1 \leq 2\alpha_2$ is sufficient by explicitly construct a simulation scheme. Note that this condition matches the channel condition of *degraded* DM-BC, which is quite interesting. Consider the following scheme




Alice take $\tilde{U} = f(X^n) = X_1$, and Bob passes $Y_1$ through a simulated channel where an erasure is mapped with equal probability to $\{0,1\}$, together with an additional crossover probability $\tilde{\alpha}$ such that $\frac{\epsilon}{2} * \tilde{\alpha} = \alpha_2$. Bob can then simulate the channel, as $S_n$ defined above approaches a Gaussian distribution, any distribution can be simulated using $S_n$.
Notice there is a gap between the inner and outer bounds. One may ask which one is tight. We performed numerical simulation and it turns out that the sufficient condition is tight. The analytical proof is still open.

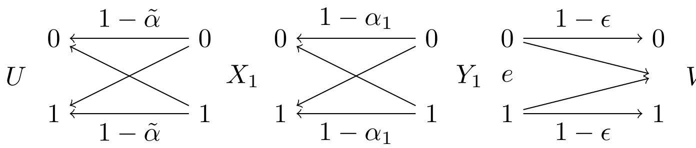
### D. Example 4: DSBS Input and BSBES Output

The source follows DSBS($\alpha_1$) and the output follows BSBES($\alpha_2, \epsilon_2$).




*1) Necessary Condition:* As before, $\rho_m^2(X;Y) \geq \rho_m^2(U;V) \Leftrightarrow (1-\alpha_1)^2 \geq (1-\epsilon_2)(1-\alpha_2)^2$ is necessary.

*2) sufficient Condition:* $\epsilon_1 \leq 2\alpha_2$ is a sufficient condition, by using the following scheme. Note that there exist $\tilde{\alpha}$ such that $\alpha_2 = \tilde{\alpha} * \alpha_1$




Alice and Bob both simulate channels, indicated above. Using non-overlapping time series and previous methodology, the simulation is achievable. Note here a gap between inner and outer bounds also exist, and numerical simulation suggests the inner bound is tight.

### E. Example 5: $s_p$ Gives Sharper Outer Bound than $\rho_m$

In [3], the authors give a example where hypercontractivity gives sharper necessary condition. Consider $(X, Y)$ is DSBS($\alpha$) and $(U, V)$ to be simulated follows

$$\mathbb{P}_{U,V}(0,0) = \mathbb{P}_{U,V}(0,1) = \mathbb{P}_{U,V}(1,0) = \frac{1}{3}$$

*1) Necessary Condition:* Assume that $\alpha \le \frac{1}{2}$. Maximal correlation gives the following outerbound

$$\rho_m(X;Y) \ge \rho_m(U;V) \Leftrightarrow 1 - 2\alpha \ge \frac{1}{2} \Leftrightarrow \frac{1}{4} \le \alpha \le \frac{1}{2}$$

On the other hand, hypercontractivity gives tighter necessary condition. If simulation is possible, there exist functions $f, g$, satisfying $\tilde{U} = f(X^n)$ and $\tilde{V} = g(Y^n)$, and

$$\mathbb{P}\{\tilde{U} = 1\} \approx \mathbb{P}\{\tilde{V} = 1\} \approx \frac{1}{3}, \text{ and } \mathbb{P}\{\tilde{U} = 1, \tilde{V} = 1\} \approx 0$$

Let $A = f^{-1}(\{1\}), B = g^{-1}(\{1\})$. Then, the reversed hypercontractivity inequality under $p < 1$ says,

$$\mathbb{P}((X^n, Y^n) \in A \times B) \ge \mathbb{P}(X^n \in A)^p \mathbb{P}(Y^n \in B)^{pr_p} > 0$$

which means $\mathbb{P}(\tilde{U} = 1, \tilde{V} = 1) = \mathbb{P}((X^n, Y^n) \in A \times B) > 0$, indicating that the simulation is impossible for all $\alpha \le \frac{1}{2}$.

## V. Conclusion

We explored the properties of maximal correlation and hypercontractivity, and illustrated their importance in establishing outer bounds in the non-interactive simulation problem. Further investigations in more general settings are needed to unleash the full power of this technic.

## References


[1] R. Ahlswede and P. Gács, "Bounds on conditional probabilities with applications in multi-user communication", *Z. Wahrscheinlichkeitstheorie und Verw.*, vol. 34, pp. 157-177, 1976.

[2] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontractivity of Markov operators", *Annals of Probability*, vol. 4, pp. 925-936, 1976.

[3] K. Sudeep and V. Anantharam, "Non-interactive simulation of joint distributions: The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon.", *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on.* IEEE, 2012.

[4] P. Gács and J. Körner, "Common information is far less than mutual information", *Problems on Control and Information Theory*, vol. 2, no. 2, pp. 119-162, 1972.

[5] H. Witsenhausen, "On sequences of pairs of dependent random variables." *SIAM Journal on Applied Mathematics* Vol. 28, no. 1, pp. 100-113, 1975

[6] V. Anantharam et. al., "On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover", *arXiv preprint arXiv:1304.6133*, 2013.

[7] G. Kumar, "On sequences of pairs of dependent random variables: A simpler proof of the main result using SVD", *Personal Homepage*, 2010. `http://web.stanford.edu/~gowthamr/research/Witsenhausen_simpleproof.pdf`


## Appendix

### A. Proof of tensorization of maximal correlation

For any $g$ such that $\mathbb{E}[g(Y_1, Y_2)] = 0$, we have

$$\begin{aligned}
&||\mathbb{E}[g(Y_1, Y_2)|X_1 X_2]||_2 \\
&= ||\mathbb{E}[(g - \mathbb{E}[g|X_1] - \mathbb{E}[g|X_2]) + \mathbb{E}[g|X_1] + \mathbb{E}[g|X_2]|X_1 X_2]||_2 \\
&\overset{(a)}{\le} ||\mathbb{E}[g - \mathbb{E}[g|X_1] - \mathbb{E}[g|X_2]|X_1 X_2]||_2 \\
&\quad + ||\mathbb{E}[\mathbb{E}[g|X_1]|X_1 X_2]||_2 + ||\mathbb{E}[\mathbb{E}[g|X_2|X_1 X_2]]||_2 \\
&\overset{(b)}{\le} \rho_m(X_1;Y_1)\rho_m(X_2,Y_2)||g - \mathbb{E}[g|X_1] - \mathbb{E}[g|X_2]||_2 \\
&\quad + \rho_m(X_2, Y_2)||\mathbb{E}[g|X_1]||_2 + \rho_m(X_1;Y_1)||\mathbb{E}[g|X_2]||_2 \\
&\le \max\{\rho_m(X_1;Y_1), \rho_m(X_2,Y_2)\} \\
&\quad (||g - \mathbb{E}[g|X_1] - \mathbb{E}[g|X_2]||_2 + ||\mathbb{E}[g|X_1]||_2 + ||\mathbb{E}[g|X_2]||_2) \\
&\overset{(c)}{=} \max\{\rho_m(X_1;Y_1), \rho_m(X_2,Y_2)\}||g(Y_1,Y_2)||_2
\end{aligned}$$

where (a) is by Minkowski's inequality, (b) is by definition of $\rho_m$, and (c) is by orthogonality. So we have

$$\rho_m(X_1 X_2; Y_1 Y_2) \le \max\{\rho_m(X_1;Y_1), \rho_m(X_2,Y_2)\}.$$

The other direction is immediate. □

### B. Proof of tensorization of hypercontractivity ribbon

Without loss of generality, assume $p > 1$. For any $g$ and $r > \max\{r_p(X_1;Y_1), r_p(X_2;Y_2)\}$, we have

$$\begin{aligned}
&||\mathbb{E}[g(Y_1,Y_2)|X_1 X_2]||_p \\
&= \{\int\int[\int\int g(y_1 y_2) w_{Y_1|X_1}(dy_1, x_1) w_{Y_2|X_2}(dy_2, x_2)]^p \\
&\quad \mathbb{P}_{X_1}(dx_1)\mathbb{P}_{X_2}(dx_2)\}^{\frac{1}{p}} \\
&\overset{(a)}{\le} \{\int\{\int w_{Y_1|X_1}(dy_1, x_1) \\
&\quad [\int(\int w_{Y_2|X_2}(dy_2,x_2) g(y_1 y_2))^p \mathbb{P}_{X_2}(dx_2)]^{\frac{1}{p}}\}^p \mathbb{P}_{X_1}(dx_1)\}^{\frac{1}{p}} \\
&\overset{(b)}{\le} \{\int[\int w_{Y_1|X_1}(dy_1, x_1)(\int g(y_1 y_2)^{pr}\mathbb{P}_{Y_2}(dy_2))^{\frac{1}{pr}}]^p \mathbb{P}_{X_1}(dx_1)\}^{\frac{1}{p}} \\
&\overset{(c)}{\le} \{\int[\int g(y_1 y_2)^{pr}\mathbb{P}_{Y_2}(dy_2)]^{\frac{1}{pr}pr} P_{Y_1}(dy_1)\}^{\frac{1}{pr}} \\
&= ||g(Y_1 Y_2)||_{pr}
\end{aligned}$$

where (a) is my Minkowski's inequality, (b),(c) are from the definition of $r_p$. As a result, we have

$$r_p(X_1 X_2; Y_1 Y_2) \le \max\{r_p(X_1;Y_1), r_p(X_2,Y_2)\}.$$

The other direction is immediate. □