On the Importance of Asymmetry and Monotonicity Constraints in Maximal Correlation Analysis

Elad Domanovitz
Dept. of EE-Systems, TAU
Tel Aviv, Israel

Uri Erez
Dept. of EE-Systems, TAU
Tel Aviv, Israel

***Abstract*—The maximal correlation coefficient is a well-established generalization of the Pearson correlation coefficient for measuring non-linear dependence between random variables. It is appealing from a theoretical standpoint, satisfying Rényi's axioms for measures of dependence. It is also attractive from a computational point of view due to the celebrated alternating conditional expectation algorithm, allowing to compute its empirical version directly from observed data. Nevertheless, from the outset, it was recognized that the maximal correlation coefficient suffers from some fundamental deficiencies, limiting its usefulness as an indicator of estimation quality. Another well-known measure of dependence is the correlation ratio but it too suffers from some drawbacks. Specifically, the maximal correlation coefficient equals one too easily whereas the correlation ratio equals zero too easily. The present work recounts some attempts that have been made in the past to alter the definition of the maximal correlation coefficient in order to overcome its weaknesses and then proceeds to suggest a natural variant of the maximal correlation coefficient. The proposed dependence measure at the same time resolves the major weakness of the correlation ratio measure and may be viewed as a bridge between the two classical measures.**

I. Introduction

Pearson's correlation coefficient is a measure indicating how well one can approximate (estimate in an average least squares sense) a (response) random variable Y as a linear (more precisely affine) function of a (predictor/observed) random variable X, i.e., as $Y = aX + b$.[1] The coefficient is given by

$$\rho(X \leftrightarrow Y) = \frac{\mathrm{Cov(X,Y)}}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}. \tag{1}$$

The coefficient is symmetric in X and Y so it just as well measures how well one can approximate X as a linear function of Y.

The correlation ratio of Y on X, also suggested by Pearson (see, e.g., [1]), similarly measures how well one can approximate Y as a general admissible function of X, i.e., as $Y = f(X)$.[2] Specifically, the correlation ratio of Y on X is given by

$$\theta(X \to Y) = \sqrt{\frac{\mathrm{var}(\mathbb{E}[Y|X])}{\mathrm{var}(Y)}} = \sqrt{1 - \frac{\mathbb{E}[\mathrm{var}(Y|X)]}{\mathrm{var}(Y)}}. \tag{2}$$

The correlation ratio can also be expressed as $\theta(X \to Y) = \sup_f \rho(f(X) \leftrightarrow Y)$, where the supremum is taken over all (admissible) functions f (see, e.g., [2]). This measure is naturally nonsymmetric. A drawback of the correlation ratio is that it "equals zero too easily". Specifically, it can vanish even where the variables are dependent.

The Hirschfeld-Gebelein-Rényi maximal correlation coefficient [3]–[5] measures the maximal (Pearson) correlation that can be attained by transforming the pair X, Y into random variables $X' = g(X)$ and $Y' = f(Y)$; that is, how well $X' = aY' + b$ holds in a mean squared error sense for some pair of functions f and g. More precisely, the maximum correlation coefficient is defined as the supremum over all (admissible) functions f, g of the correlation between $f(X)$ and $g(Y)$:

$$\rho^{**}_{\max}(X \leftrightarrow Y) = \sup_{f,g} \rho(f(X) \leftrightarrow g(Y)). \tag{3}$$

This measure is again symmetric by definition. We use the superscript "**" to indicate that both functions (applied to the response and the predictor random variables) need not satisfy any restrictions beyond being admissible.

The maximal correlation coefficient has some very pleasing properties. In particular, in [5], Rényi put forth a set of seven axioms deemed natural to require of a measure for dependence between a pair of random variables. He further established that the maximal correlation coefficient satisfies the full set of axioms. In particular, unlike the correlation ratio, the maximal correlation coefficient "does not equal zero too easily". Further, unlike the correlation ratio, it is symmetric, which was set as one of the axioms. Nonetheless, this comes at the price of "equaling one too easily" as exemplified below.[3]

Rényi's seminal work inspired substantial subsequent work aiming to identify other measures of dependence satisfying the set of axioms. We refer the reader to [6] for a survey of some of these.

Another appealing trait of the maximal correlation coefficient, greatly contributing to its popularity, is its relation to the mean square error and hence to a Euclidean geometric framework. In particular, it is readily computable numerically via the alternating conditional expectation (ACE) algorithm of Breiman and Friedman [7]. Moreover, and as recalled in the sequel, the ACE algorithm naturally extends to cover linear estimation of a (transformed) random variable from a component-wise transformed random vector.

Despite its elegance and it being amenable to computation, the maximal correlation coefficient suffers from some significant deficiencies as was recognized since its inception.

[1]We assume that the random variables X and Y have finite variance.

[2]We define a function $f(\cdot)$ to be admissible w.r.t. the random variable X if it is a Borel-measurable real-valued function such that $\mathbb{E}[f(X)] = 0$ and it has finite and positive variance.

[3]See also footnote 3 in [5].

As noted, for instance, the maximal correlation coefficient "equals one too easily"; see, e.g. [8] and [9]. Specifically, it can equal one even when the two random variables are nearly independent (as also demonstrated below).

Several suggestions have been proposed over the years to alter the measure so as to overcome these drawbacks. One important avenue calls for limiting the functions applied to both random variables to be monotone functions [9], [10]. As we observe next, such a restriction is not quite restrictive enough when it comes to the transformation applied to the response variable. More importantly, it is overly restrictive (in fact, unnecessary) when it comes to the predictor variable. With respect to the latter, it is worth quoting the incisive comments of Hastie and Tibshirani in [11]:

"A monotone restriction makes sense for a response transformation because it is necessary to allow predictions of the response from the estimated model. On the other hand, why restrict predictor transformations... ?"

The goal of the present work is first to reiterate some of the known drawbacks of the correlation ratio and the maximal correlation coefficient. In particular, we demonstrate that requiring that the transformation applied to response variable be monotonic is *almost* sufficient to resolve the drawbacks of the maximal correlation coefficient but that one must strengthen slightly the required "degree" of monotonicity.

Specifically, we introduce the notion of κ-monotonicity and argue in favor of constraining (only) the transformation applied to of the response random variable to be κ-monotonic, leading to a proposed semi-κ-monotone maximal correlation measure. The parameter κ dictates a minimal and maximal slope that the function applied to the response variable must satisfy.

We show that requiring that $0 < \kappa < 1$ yields a measure that does not suffer from the drawbacks of neither the maximal correlation coefficient nor from those of the correlation ratio. The correlation ratio and the measure alluded to by Hastie and Tibshirani can be viewed as extreme cases of the suggested measure, setting κ to be 1 or 0, respectively. The suggested measure satisfies a set of modified Rényi axioms that does not sacrifice any natural requirements in terms of capturing *both* independence and full dependence. This is achieved by doing away with the symmetry requirement so that "full dependence" is directional. Since in the context of prediction/estimation symmetry should not be expected, the desirable properties the new measure possesses do not come at a price.

Finally, as the usefulness of the maximal correlation coefficient is due in part to it being readily computable, we suggest modifications to the ACE algorithm and exemplify the resulting performance via several examples.

II. Shortcomings of the Correlation Ratio and Maximal Correlation Coefficient and a Proposed Resolution

As a simple example, consider two random variables that share only the least significant bit:

$$X = C + \sum_{i=1}^{N} A_i 2^i;\ Y = C + \sum_{i=1}^{N} B_i 2^i \tag{4}$$

where A_i, B_i, C are mutually independent random variables, all taking the value 0 or 1 with equal probability. Clearly, applying modulo 2 to both variables yields a maximal correlation of 1, for any value of N. This seems quite unsatisfactory if our goal is estimation subject to any reasonable distortion metric as the two random variables become virtually independent as N grows.

Remark 1. *It should be noted in this respect that the maximal correlation coefficient is a good measure for a different goal. It quantifies to what extent two random variables share any common "feature/s".*

Disconcerted by this behavior of the maximal correlation coefficient, Kimeldorf and Sampson [9] proposed to alter its definition in such a way that it attains the maximal value of 1 only if the pair of random variables is mutually completely dependent.[4] This motivated them to define a modified measure as

$$\rho_{\max}^{mm}(X \leftrightarrow Y) = \sup_{f,g} \rho(f(X) \leftrightarrow g(Y)). \tag{5}$$

where f and g are not only admissible but also monotone functions.

The problem with adopting this definition in the context of estimation is the symmetric constraints it imposes on the two transformations. As the process of estimation/prediction (and more generally inference) is directional, if the goal of the dependence measure is to characterize how well one can achieve the latter tasks, there is no apparent reason to impose any restriction on the transformation applied to the observed data. Several works aiming to modify Rényi's axioms to reflect this asymmetry include [8], [13] and [14].

Indeed, a natural and quite satisfying directional measure of dependence between random variables is the correlation ratio defined in (2). While Rényi objected to the correlation ratio due to its asymmetric nature, as was noted in [8], when our goal is asymmetric (i.e., estimating Y from X), there is no reason for requiring symmetry from the measure.

Nonetheless, in many cases one does not have strong grounds to assume a particular "parameterization" of the desired (response) random variable Y which is to be estimated. Thus, not allowing to apply any transformation to the response variable, as is the case of the correlation ratio, may in certain cases be too restrictive. In other words, in the absence of a preferred "natural" parametrization of the response variable, one may consider choosing a strictly monotone transformation (change of variables) so as to make it easier to estimate.

Another drawback of the correlation ratio is that it vanishes too easily. Specifically, two dependent random variables can have a correlation ratio of zero. In light of these considerations, we propose the following modification to the definition of the maximal correlation coefficient.

Definition 1. *For $0 \le \kappa \le 1$, a function f is said to be κ-increasing, if for all $x_2 \ge x_1$:*

$$f(x_2) - f(x_1) \ge \kappa(x_2 - x_1),$$

[4]Following the work of Lancaster [12], two random variables are said to be mutually completely dependent if they are almost surely invertible functions of one another.

$$f(x_2) - f(x_1) \le \frac{1}{\kappa}(x_2 - x_1). \tag{6}$$

Definition 2. *For a given $0 < \kappa < 1$, the semi-κ-monotone maximal correlation measure is defined as*

$$\rho^{*m_\kappa}_{\max}(X \to Y) = \sup_{f,g} \rho(f(X) \leftrightarrow g(Y)) \tag{7}$$

where and the supremum is taken over all admissible functions $f(x)$, and over κ-increasing admissible functions $g(y)$.

Remark 2. *Limiting g to be κ-increasing implies that, in particular, it is invertible, which is a natural requirement. The value of κ controls how far we may deviate from the correlation ratio.*

A. The vector observation case

Let $\mathbf{X} = (X_1, \ldots, X_p)$ be a vector of predictor variables. The maximal correlation coefficient becomes

$$\rho^{**}_{\max}(\mathbf{X} \leftrightarrow Y) = \sup_{f,g} \rho(f(\mathbf{X}) \leftrightarrow g(Y)) \tag{8}$$

where the supremum is over all admissible functions.

Following Breiman and Friedman [7], we may also consider a simplified (quasi-additive) relationship between Y and $\mathbf{X}$ where $f(\mathbf{X})$ is restricted to be of the form $f(\mathbf{X}) = \sum_i f_i(X_i)$. In [7], conditions for the existence of optimal transformations $\{f_i\}, g$ such that the supremum is attained are given, and it is shown that under these conditions the ACE algorithm converges to the optimal transformations.

Going back to the rationale for requiring κ-monotonicity, one may object to the example (4) as being artificial and argue that the maximum correlation coefficient merely captures whatever dependence there is between the random variables. However, as exemplified in Section V, the problematic nature of the maximal correlation coefficient becomes more pronounced when considering the multi-variate case and so does the necessity of restricting the transformation of the response variable (only) to be monotone.

III. MODIFIED RÉNYI AXIOMS

We follow the approach of Hall [8] in defining an asymmetric variant of the Rényi axioms; more precisely, we adopt a slight variation on the somewhat stronger version formulated by Li [14]. However, unlike both of these works, when it comes to putting forward a candidate dependence measure satisfying the modified axioms, we adopt the approach suggested in [11] and define a new maximal correlation measure where we restrict *only* the function applied to the response variable.

Assume $r(X \to Y)$ is to measure the degree of dependence of Y on X. Then we require that it satisfy the following:

(a) $r(X \to Y)$ is defined for all non-constant random variables X, Y having finite variance.[5]
(b) $r(X \to Y)$ may not be equal to $r(Y \to X)$.
(c) $0 \le r(X \to Y) \le 1$.
(d) $r(X \to Y) = 0$ if and only if X, Y are independent.
(e) $r(X \to Y) = 1$ if and only if $Y = f(X)$ almost surely for some admissible function f.
(f) If f is an admissible bijection on $\mathbb{R}$, then $r(f(X) \to Y) = r(X \to Y)$
(g) If X, Y are jointly normal with correlation coefficient ρ, then $r(X \to Y) = |\rho|$.[6]

We first note that the correlation ratio satisfies all of the modified axioms except for the "only if" part of axiom (d). We next observe that for absolutely continuous (or discrete) distributions, the semi-κ-monotone maximal correlation measure of Definition 2 satisfies the proposed axioms.

It is readily verified that axioms (a), (b) and (c) hold. To show that axiom (d) holds, we note that if X, Y are independent, then obviously $\rho^{*m_\kappa}_{\max}(X \to Y) = 0$, as so is even $\rho^{**}_{\max}(X \leftrightarrow Y)$. As for the other direction, we first note that it suffices to consider the case where the correlation ratio equals 0 and X, Y are dependent. Since the correlation ratio is 0, it follows from (2) that $\mathbb{E}[Y|X] \equiv const$ (in the mean square sense), i.e.,

$$\int p(y|x) y dy = \text{const}. \tag{9}$$

We may break the symmetry of $g(y) = y$ by defining, e.g.,

$$g_{a,\kappa}(y) = \begin{cases} y & y \ge a \\ \kappa y & y < a \end{cases} . \tag{10}$$

Consider two values of x_1 and x_2 for which the functions $p(y|x_i)$ are not identical (as functions of y), as must exist by the assumption of dependence. Let a be a value such that

$$\int^a p(y|x_1) y dy \ne \int^a p(y|x_2) y dy. \tag{11}$$

Without loss of generality, we may assume that the left hand side is smaller than the right hand side (we may rename x_1 and x_2). Recalling that $\kappa < 1$, it follows that

$$\int p(y|x_1) g_a(y) dy > \int p(y|x_2) g_a(y) dy \tag{12}$$

Thus,

$$\mathbb{E}[g_a(Y)|X = x_1] \ne \mathbb{E}[g_a(Y)|X = x_2]$$

and hence the correlation ratio of $Y' = g_a(Y)$ on X is nonzero, giving a lower bound to the semi-κ-monotone maximal correlation measure between Y and X.

To show that axiom (e) holds, we note that by definition, if $Y = f(X)$ (almost surely), then $\rho^{*m_\kappa}_{\max}(X \to Y) = 1$. To show that the opposite direction holds, we note that it can be shown that the supremum in (2) is attained. Recalling that if $\rho^{*m_\kappa}_{\max}(X \to Y) = 1$, then by the properties of Pearson's correlation coefficient, there is a *perfect* linear regression between $g(Y)$ and $f'(X)$ (g, f' being maximizing functions of the measure). Hence, we have $g(Y) = af'(X) + b$ where g is an increasing function with slope greater than κ. Since κ is strictly positive, it follows that not only is g invertible, but also $g^{-1}(Y)$ has finite variance (since the slope of $g^{-1}(Y)$ is at most $\frac{1}{\kappa}$ and Y has finite variance). Therefore, we have $Y = g^{-1}(af'(X) + b)$. Denoting $f(X) = g^{-1}(af'(X) + b)$, we note that if f' is admissible, then so is f.

[5] In [14], the first axiom only requires that $r(X \to Y)$ be defined for *continuous* random variables X, Y.

[6] In [14], the last axiom only requires that if X, Y are jointly normal with correlation coefficient ρ, $r(X \to Y)$ is a strictly increasing function of $|\rho|$.

Axiom (f) trivially holds. To show that axiom (g) holds, we recall that it is well known that when X, Y are jointly normal with correlation coefficient ρ, then $\rho^{**}_{\max}(X \leftrightarrow Y) = |\rho|$ (see, e.g., [15] and [16]). Since this implies that the maximal correlation is achieved taking $g(y) = y$ (i.e., a monotone function) and $f(x) = x$ or $f(x) = -x$, it follows that

$$\begin{aligned} \rho^{*m_\kappa}_{\max}(X \rightarrow Y) &= \rho^{**}_{\max}(X \leftrightarrow Y) \\ &= |\rho|. \end{aligned} \tag{13}$$

We note that the restriction $0 < \kappa < 1$ is necessary. Specifically, for $\kappa = 1$, axiom (d) is not satisfied whereas for $\kappa = 0$, axiom (e) is not satisfied.

Finally, we note that one may define other dependence measures satisfying the set of modified Rényi axioms, most notably via the theory of copulas; see [14]. Nonetheless, we believe that the proposed measure has the advantage of being closely tied to linear regression methods and geometric considerations.

IV. Modified ACE Algorithm

We begin by presenting a modification of the ACE algorithm to compute the semi-0-monotone maximal correlation measure $\rho^{*m_0}_{\max}(X \rightarrow Y)$, restricting the function applied to the response variable only to be weakly monotone.

As we do not know of a simple means to enforce the slope constraints, we do not have an algorithm for computing the semi-κ-monotone maximal correlation measure. Instead, we have employed a regularized version of the ACE algorithm as described in Section IV-B.

A. *Evaluating semi-0-monotone maximal correlation*

We now present a modification of the ACE algorithm to compute the semi-0-monotone maximal correlation measure $\rho^{*m_0}_{\max}(X \rightarrow Y)$ for the case of single variate. It is readily seen that the correlation increases in each iteration of the algorithm and thus converges but we do not pursue proving optimality.

Following in the footsteps of [7], recall that the space of all random variables with finite variance is a Hilbert space, which we denote $\mathcal{H}_2$, with the usual definition of the inner product $< X, Y >= \mathbb{E}[XY]$, for $X, Y \in \mathcal{H}_2$. Since applying an admissible function f results in a random variable with finite variance, we may define the subspace $\mathcal{H}_2(X)$ as the set of all random variables that correspond to an admissible function of X.

Similarly, the set of all admissible functions of Y is also a subspace of $\mathcal{H}_2$, which we denote by $\mathcal{H}_2(Y)$. Now, if we limit the functions applied to Y to be monotone increasing, we obtain a closed and convex subset of the Hilbert space $\mathcal{H}_2(Y)$. We denote this set by $\mathcal{M}_0(Y)$.

Denoting by $P_{\mathcal{A}}(Y)$ the orthogonal projection of Y onto the closed convex set $\mathcal{A}$,[7] the modified ACE algorithm is described in Algorithm 1 for the case of single predictor variable.

B. *Regularized ACE algorithm*

Another potential modification for the ACE algorithm is to add the requirement that $g(Y)$ have a minimal slope κ; unfortunately, we do not know of a simple means to enforce an upper limit on the slope.

[7]Note that $\mathcal{P}_{\mathcal{H}_2(X)}(g(Y)) = \mathbb{E}[g(Y) \mid X]$.

Algorithm 1

1: **procedure** Calculate-Semi-0-monotone
2: Set $g(Y) = Y/\|Y\|$;
3: **while** $e^2(g, f)$ decreases **do**
4: $f'(X) = \mathcal{P}_{\mathcal{H}_2(X)}(g(Y))$
5: replace $f(X)$ with $f'(X)$
6: $g'(Y) = \mathcal{P}_{\mathcal{M}_0(Y)}(f(X))$
7: replace $g(Y)$ with $g'(Y)/\|g'(Y)\|$
8: End modified ACE

To enforce that the transformation applied to $g(Y)$ will have a minimal slope κ, we apply the following regularization:[8] That is, step 8 in Algorithm 1 becomes step 9 and the regularization is performed as step 8 (outside the while-loop).

C. *Multi-variate predictor*

The modified ACE algorithms for the case of a multi-variate predictor, is a simple extension of Algorithms 1 and the regularized ACE algorithm and can be found in [17].

V. Numerical examples

As we do not know of an efficient means to evaluate the semi-κ-monotone maximal correlation measure for $\kappa \neq 0$, we for the most part demonstrate the advantage of the semi-0-monotone maximal correlation measure over the standard maximal correlation measure, in the context of estimation of a random variable Y from a random vector $\mathbf{X}$.

We begin with a multi-variate example where one of the two observed random variables "masks" the other while the latter is more significant for estimation purposes. We then demonstrate why taking $\kappa = 0$ is not sufficient in general, and show heuristically that the regularized ACE algorithm yields more satisfying results.

For simulating ACE, we used the ACE Matlab code provided by the authors of [18]. To limit g to be a monotonic function we used isotonic regression.

A. *Example 1 - Multi-variate predictor*

Assume that the response variable Y is distributed uniformly over the interval $[0, 1]$. Assume we have two predictor variables

$$X_1 = \mathrm{mod}(\mathrm{Y}, 0.2) + \mathrm{N}_1; \;\; \mathrm{X}_2 = \mathrm{Y}^3 + \mathrm{N}_2 \tag{15}$$

where N_1, N_2 are independent zero-mean Gaussian variables with $\sigma^2_{N_1} = 0.01$ and $\sigma^2_{N_2} = 0.2$.

As can be seen from Figure 1, the ACE algorithm, in order to maximize the correlation, chooses similar functions as in case of running only on Y and X_1, practically choosing to ignore X_2 (the outcome of running ACE between Y and X_1 and between Y and X_2 can be found in [17]). While,

[8]Note that this method of regularization actually forces the minimal slope to be slightly larger than κ.

$$g(Y) = g(Y) + \kappa \cdot Y. \tag{14}$$

indeed, this maximizes the correlation coefficient, it is far from satisfying from an estimation viewpoint.

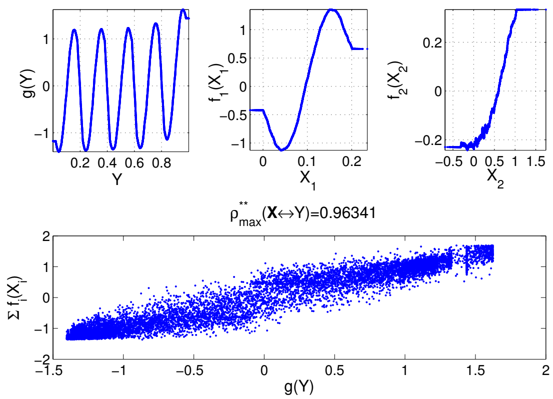

Fig. 1. Example 1: Running ACE on Y, X_1 and X_2.

As can be seen from Figure 2, the resulting value of the semi-0-monotone maximal correlation measure is very close to the maximal correlation value between Y and X_2. Thus, the algorithm "chooses to ignore" X_1 (even though it suffers from a lower noise level) and bases the estimation on X_2.

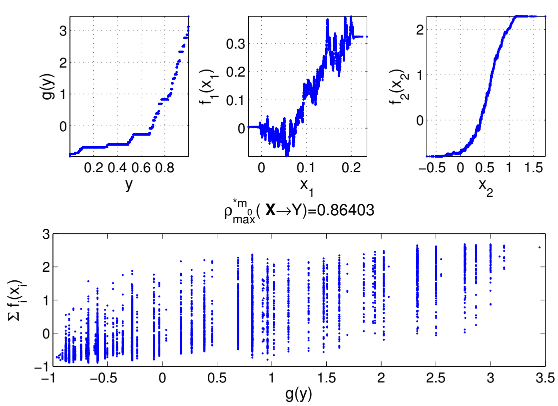

Fig. 2. Example 1: Running modified ACE (Algorithm 1) on Y, X_1 and X_2 with $\kappa = 0$.

B. *Example 2 - Semi-0-monotonicity is insufficient*

To illustrate why it does not suffice to limit g to be merely monotone, consider the following example. Assume that the response Y is distributed uniformly over the interval $[-10, 10]$ and that

$$X = \begin{cases} X = Y & Y > 9 \\ X = N_1 & \text{otherwise} \end{cases} \tag{16}$$

where $N_1 \sim \text{Unif}([-1, 1])$ and is independent of Y.

Limiting g only to be monotone (with no slope limitations) results in a correlation value of 1 since the optimal solution is to set $g(y) = 0$ in the region it cannot be estimated and $g(y) = y$ otherwise (and then apply normalization). Clearly, the function g is non-invertible (and can be found in [17]).

Next, we ran the regularized ACE algorithm, enforcing a minimal slope of $\kappa = 0.1$. The results are depicted in Figure 3.

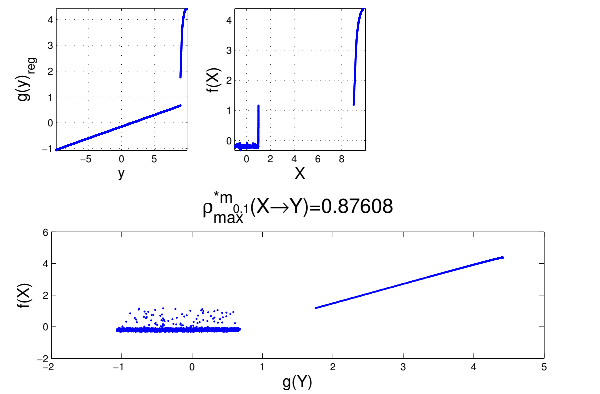

Fig. 3. Example 2: Running regularized ACE on Y, X with $\kappa = 0.1$

This example sheds light on the trade-off that exists when setting the value of κ. Setting κ to be large limits the possible gain over the correlation ratio whereas setting it too low risks overemphasizing regions where the noise is weaker.

REFERENCES

[1] H. Cramér, *Mathematical methods of statistics (PMS-9)*. Princeton University Press, 2016, vol. 9.
[2] A. Rényi, "New version of the probabilistic generalization of the large sieve," *Acta Mathematica Hungarica*, vol. 10, no. 1-2, pp. 217–226, 1959.
[3] H. O. Hirschfeld, "A connection between correlation and contingency," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, no. 4. Cambridge University Press, 1935, pp. 520–524.
[4] H. Gebelein, "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung," *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364–379, 1941.
[5] A. Rényi, "On measures of dependence," *Acta Mathematica Hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.
[6] D. Drouet-Mari and S. Kotz, *Correlation and dependence*. World Scientific, 2001.
[7] L. Breiman and J. H. Friedman, "Estimating optimal transformations for multiple regression and correlation," *Journal of the American Statistical Association*, vol. 80, no. 391, pp. 580–598, 1985.
[8] W. Hall, *On characterizing dependence in joint distributions*. University of North Carolina, Department of Statistics, 1967.
[9] G. Kimeldorf and A. R. Sampson, "Monotone dependence," *The Annals of Statistics*, pp. 895–903, 1978.
[10] J. O. Ramsay, "Monotone regression splines in action," *Statistical Science*, vol. 3, no. 4, pp. 425–441, 1988.
[11] T. Hastie and R. Tibshirani, "[monotone regression splines in action]: Comment," *Statistical Science*, vol. 3, no. 4, pp. 450–456, 1988.
[12] H. Lancaster, "Correlation and complete dependence of random variables," *The Annals of Mathematical Statistics*, vol. 34, no. 4, pp. 1315–1321, 1963.
[13] K. Joag-Dev, "Measures of dependence," *Handbook of Statistics*, vol. 4, pp. 79–88, 1984.
[14] H. Li, "A true measure of dependence," University Library of Munich, Germany, Tech. Rep., 2016.
[15] H. O. Lancaster, "Some properties of the bivariate normal distribution considered in the form of a contingency table," *Biometrika*, vol. 44, no. 1/2, pp. 289–292, 1957.
[16] Y. Yu, "On the maximal correlation coefficient," *Statistics & Probability Letters*, vol. 78, no. 9, pp. 1072–1075, 2008.
[17] E. Domanovitz and U. Erez, "The importance of asymmetry and monotonicity constraints in maximal correlation analysis." [Online]. Available: https://arxiv.org/abs/1901.03590
[18] H. Voss and J. Kurths, "Reconstruction of non-linear time delay models from data by the use of optimal transformations," *Physics Letters A*, vol. 234, no. 5, pp. 336–344, 1997.