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] It is given by

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

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

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

The correlation ratio can also be expressed as

$$\theta(X \to Y) = \sup_f \rho(f(X) \leftrightarrow Y),$$

[1] We assume that both random variables have finite and positive 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.

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", i.e., it can vanish even when 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 $g(Y)$:

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

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 of 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.

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

Despite its elegance and it being amenable to computation, the maximal correlation coefficient suffers from some significant deficiencies as was recognized since its inception. As noted, it "equals one too easily"; see, e.g. [8] and [9]. In fact, it can equal one even for two random variables that are nearly independent (as also demonstrated below).

Disconcerted by this behavior of the maximal correlation coefficient, Kimeldorf and Sampson [9] proposed to alter its definition, introducing monotonicity constraints. They defined a monotone dependence measure as follows.

$$\rho_{\max}^{mm}(X \leftrightarrow Y) = \sup_{f,g} \rho(f(X) \leftrightarrow g(Y)), \tag{4}$$

where f and g are not only admissible but also monotone. Nevertheless, as stated in [9], while the imposed constraints somewhat mitigate the "easiness of attaining the value of one", the measure (4) still can equal one for a pair of random variables that are not completely dependent.

The definition of the monotone dependence measure (4) is unsatisfactory in two respects. The first is that is imposes symmetric constraints 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. In this respect, it is worth quoting the incisive comments (in reference to [10]) of Hastie and Tibshirani [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 second and more subtle deficiency of the monotone dependence measure of Kimeldorf and Sampson (as well as the semi-monotone variant suggested by Hastie and Tibshirani) is that when it comes to the response variable, the requirement that the transformation be monotone is not strong enough.

The goal of the present work is first to reiterate some of the known drawbacks of both the correlation ratio and of the maximal correlation coefficient, and then to suggest a possible resolution. In particular, we demonstrate that while allowing a transformation to be applied to the response variable is important, it is not sufficient to require that it be monotonic. Rather, one must strengthen the required "degree" of monotonicity.

Specifically, we introduce the notion of κ-monotonicity and argue in favor of constraining (only) the transformation applied to 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 maintain.

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 proposed measure satisfies a set of modified Rényi axioms that does not sacrifice the natural requirements of capturing *both* independence and complete dependence.

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 (sequences of) random variables that share only the least significant bit:

$$X^{(N)} = C + \sum_{i=1}^{N} A_i 2^i$$
$$Y^{(N)} = C + \sum_{i=1}^{N} B_i 2^i,$$

where A_i, B_i, C are mutually independent random variables, all taking the values 0 or 1 with equal probability. Clearly, applying modulo 2 to both variables yields a correlation of one. 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. Specifically, the pair $(X^{(N)}/2^N, Y^{(N)}/2^N)$ converges in distribution to a uniform distribution over the unit square.

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

A natural and quite satisfying measure of directional dependence between random variables, that takes the value of one only when the response variable is a function of the predictor variable, 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 that the measure be symmetric.

Nonetheless, in some cases one does not have strong grounds to assume a particular "parameterization" of the desired (response) random variable. Thus, not allowing to apply any transformation to the response variable, as is the case of the correlation ratio, may be too restrictive. In other words, in the absence of a preferred "natural" parameterization of the response variable, one may consider choosing a strictly monotone transformation (change of variables) so as to make it easier to estimate. A more severe drawback of the correlation ratio is that it vanishes too easily, i.e., it can be zero for two dependent random variables.

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 $x_2 \ge x_1$:*

$$f(x_2) - f(x_1) \ge \kappa(x_2 - x_1),$$
$$f(x_2) - f(x_1) \le \frac{1}{\kappa}(x_2 - x_1). \tag{5}$$

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

$$\rho^{*m_\kappa}_{\max}(X \to Y) = \sup_{f,g} \rho(f(X) \leftrightarrow g(Y)) \tag{6}$$

where and the supremum is taken over all admissible functions f*, and over* κ*-increasing admissible functions* g*.*

Remark 2. *Limiting* g *to be* κ*-increasing implies that, in particular, it is invertible, which is a natural requirement. Further, the set of* κ*-increasing admissible functions is closed. We further note that the value of* κ *controls how far the measure can deviate from the correlation ratio.*

A. The vector observation case

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

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

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 in (5) as being artificial and argue that the maximum correlation coefficient merely captures whatever dependence there is between the random variables. In this respect, it is worthwhile quoting Breiman [12] (commenting on [10]):

"I only know of infrequent cases in which I would insist on monotone transformations. Finding non-monotonicity can lead to interesting scientific discoveries. If the appropriate transformation is monotone, then the fitted spline functions (or ACE transformations) will produce close to a monotonic transformation. So it is hard to see what there is to gain in the imposition of monotonicity."

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 applied to 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 [13]. However, unlike both of these works, when it comes to putting forward a candidate dependence measure satisfying the modified axioms, we adhere to a mean square error methodology.

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.[4]
(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 $r(X \to Y) = |\rho|$.

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 6 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 $\mathbb{E}[Y|X] \equiv \mathit{const}$ (in the mean square sense), i.e.,

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

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

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

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

$$\int^a p(y|x_1) y dy \neq \int^a p(y|x_2) y dy. \tag{8}$$

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

$$\int p(y|x_1) g_a(y) dy > \int p(y|x_2) g_a(y) dy \tag{9}$$

Thus,

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

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

To show that axiom (e) holds, we first note that if $Y = f(X)$ (almost surely), then clearly $\rho^{*m_\kappa}_{\max}(X \to Y) = 1$, as this is the case even for the correlation ratio. To show that the opposite direction holds, we note that it can be shown that the supremum in (6) is attained. Recalling that if $\rho^{*m_\kappa}_{\max}(X \to Y) = 1$, then by the properties of the Pearson correlation coefficient, there is a *perfect* linear regression between $g(Y)$ and $f'(X)$

[4] In [13], the first axiom only requires that $r(X \to Y)$ be defined for *continuous* random variables X, Y.

(g, f' being maximizing functions of the measure). Hence, we have $g(Y) = af'(X) + b$ where g is an increasing function with slope no smaller 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.

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 (see, e.g., [14] and [15]) $\rho^{**}_{\max}(X \leftrightarrow Y) = |\rho|$ and is achieved taking $g(y) = y$ (a κ-increasing function) and $f(x) = x$ or $f(x) = -x$. Hence,

$$\rho^{*m_\kappa}_{\max}(X \to Y) = \rho^{**}_{\max}(X \leftrightarrow Y) = |\rho|.$$

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 modified Rényi axioms, most notably via the theory of copulas (which is inherently related to monotone constraints); e.g., a symmetric measure is given in [16] and a directional one is given in [13]. 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 with the goal of computing the semi-0-monotone maximal correlation measure $\rho^{*m_0}_{\max}(X \to Y)$, restricting the function applied to the response variable only to be weakly monotone.

As we do not know of a simple means to maximize correlation subject to slope constraints, we do not have an algorithm for computing the semi-κ-monotone maximal correlation measure. Instead, we apply regularization to the outcome of the modified ACE algorithm as described in Section IV-B.

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

We describe a modification of the ACE algorithm to compute the semi-0-monotone maximal correlation measure $\rho^{*m_0}_{\max}(X \to Y)$ for the case of a 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 by $\mathcal{H}_2$, with the usual definition of the inner product $< X, Y >= \mathbb{E}[XY]$, for $X, Y \in \mathcal{H}_2$. We may further define the subspace $\mathcal{H}_2(X)$ as the set of all random variables that correspond to an admissible function of X. We similarly define the subspace $\mathcal{H}_2(Y)$. Now, if we further limit the functions applied to Y to be non-decreasing, 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 a closed convex set $\mathcal{A}$,[5] the modified ACE algorithm is described in Algorithm 1 for the case of a single predictor variable.

[5]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

B. *Regularized ACE algorithm*

We may enforce that the transformation g satisfies κ-monotonicity by applying the following regularization. Given a monotone transformation g (e.g., the outcome of Algorithm 1), do:[6]

$$g'(Y) = g^{-1}(Y) + \kappa \cdot Y$$
$$g(Y) = g'^{-1}(Y) + \kappa \cdot Y$$

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 are simple extensions of Algorithm 1 and its regularized variant; the details can be found in [17].

V. Numerical examples

We first 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 then demonstrate why taking $\kappa = 0$ is not sufficient in general, and show heuristically that the regularized variant of the modified 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

$$X_1 = \mathrm{mod}(\mathrm{Y}, 0.2) + \mathrm{N}_1$$
$$X_2 = Y^3 + N_2,$$

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 (computing maximal correlation) essentially chooses to ignore X_2 and applies similar functions as in the case of running ACE only on Y and X_1 (the outcome of running ACE on Y and X_1 and on Y and X_2 can be found in [17]). While, indeed, this maximizes the correlation coefficient, it is far from satisfying from an estimation viewpoint.

[6]Note that this method of regularization actually forces the slope to be between κ and $1/\kappa + \kappa$.

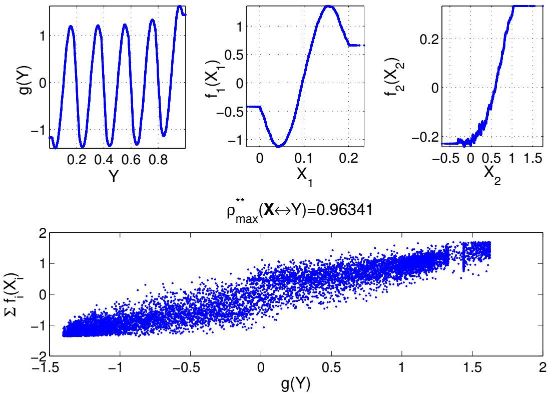

Fig. 1. Example 1: Running ACE on Y, X_1 and X_2.

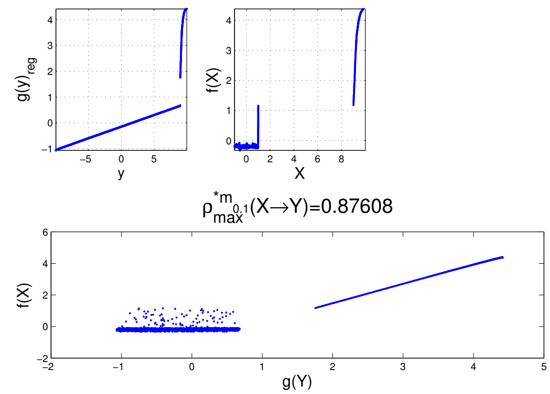

Fig. 3. Example 2: Running regularized modified ACE on Y, X with $\kappa = 0.1$

In contrast, as can be seen from Figure 2, running Algorithm 1 (for the semi-0-monotone maximal correlation measure) results in essentially "choosing to ignore" X_1 (even though it suffers from a lower noise level) and basing the estimation on X_2. This yields similar results to running the ACE algorithm on Y and X_2 only.

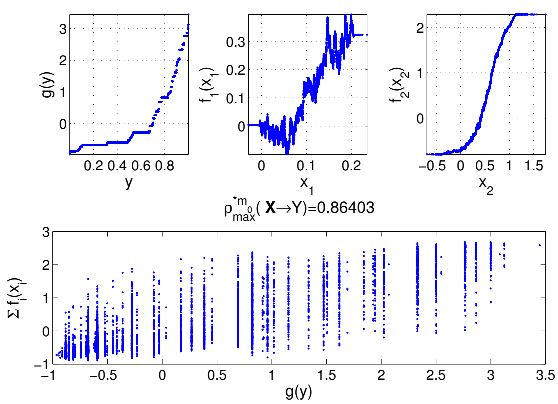

Fig. 2. Example 1: Running modified ACE (Algorithm 1) on Y, X_1 and X_2 with $\kappa = 0$.

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 $Y \sim \mathrm{Unif}([-10, 10])$, $N_1 \sim \mathrm{Unif}([-1, 1])$ and is independent of Y, and that

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

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.

Next, we ran the regularized ACE algorithm, enforcing a minimal slope of $\kappa = 0.1$. The results are depicted in Figure 3. 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] L. Breiman, "[monotone regression splines in action]: Comment," *Statistical Science*, vol. 3, no. 4, pp. 442–445, 1988.

[13] H. Li, "A true measure of dependence," University Library of Munich, Germany, Tech. Rep., 2016.

[14] 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.

[15] Y. Yu, "On the maximal correlation coefficient," *Statistics & Probability Letters*, vol. 78, no. 9, pp. 1072–1075, 2008.

[16] B. Schweizer, E. F. Wolff *et al.*, "On nonparametric measures of dependence for random variables," *The annals of statistics*, vol. 9, no. 4, pp. 879–885, 1981.

[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.