IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 14, 2021, accepted November 28, 2021, date of publication November 30, 2021,
date of current version December 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131740

Maximal Associated Regression: A Nonlinear Extension to Least Angle Regression

SANUSH K. ABEYSEKERA[1], (Member, IEEE), YE-CHOW KUANG[1], (Senior Member, IEEE), MELANIE PO-LEEN OOI[1], (Senior Member, IEEE), AND VINEETHA KALAVALLY[2]

[1]Faculty of Science and Engineering, The University of Waikato, Hamilton 3216, New Zealand
[2]Department of Electrical and Computer Systems Engineering, School of Engineering, Monash University Malaysia, Sunway 47500, Malaysia

Corresponding author: Ye-Chow Kuang (yechow.kuang@waikato.ac.nz)

This work was supported by the Ministry of Science, Technology and Innovation Malaysia under Grant 03-02-10-SF0284. The work of Melanie Po-Leen Ooi was supported by the Royal Society of New Zealand Te Apārangi through a Rutherford Discovery Fellowship.

ABSTRACT This paper proposes Maximal Associated Regression (MAR), a novel algorithm that performs forward stage-wise regression by applying nonlinear transformations to fit predictor covariates. For each predictor, MAR selects between a linear or additive fit as determined by the dataset. The proposed algorithm is an adaptation of Least Angle Regression (LARS) and retains its efficiency in building sparse models. Constrained penalized splines are used to generate smooth nonlinear transformations for the additive fits. A monotonically constrained extension of MAR (MARm) is also introduced in this paper to fit isotonic regression problems. The proposed algorithms are validated on both synthetic and real datasets. The performances of MAR and MARm are compared against LARS, Generalized Linear Models (GLM), and Generalized Additive Models (GAM) under the Gaussian assumption with a unity link function. Results indicate that MAR-type algorithms achieve a superior subset selection accuracy, generating sparser models that generalize well to new data. MAR is also able to generate models for sample deficient datasets. Thus, MAR is proposed as a valuable tool for subset selection and data exploration, especially when *a priori* knowledge of the dataset is unavailable.

INDEX TERMS Additive models, least angle regression, nonlinear transformations, subset selection.

I. INTRODUCTION

Linear regression forms the foundation of many modern statistical modelling and data analysis problems. It captures the relationship between the response and predictor covariates of the form $y = \beta_0 + \sum_{j=1}^{p} x_j\beta_j + \epsilon$ for $j = 1, \ldots, p$ by estimating the coefficients $\beta = (\beta_1, \ldots, \beta_p)^T$ and intercept β_0. Here, a normally distributed noise component is represented by ϵ. The predictor covariates $x_j = (x_{1j}, \ldots, x_{nj})^T \in \mathbb{R}^n$ are assumed to be independent and the response vector is represented by $y = (y_1, \ldots, y_n)^T$. Compared to more complex regression techniques, linear models (LM) are preferred when model interpretability is of particular importance. However, the application of LMs to real-world datasets sometimes results in poor model performance; occasionally the model fits are purely artefacts of an LM trying to fit a nonlinear model. A more accurate fit is achieved by introducing nonlinear smoothing terms $\phi_j(x_j)$ to the linear regression formulation (1). These classes of models are known as additive models [1]. Additive models (AM) are less general than its more popular cousin, Generalized Additive Models (GAM) [2] in that it assumes a strictly normal distribution with a unity link function $g(y) = y$.

$$y = \beta_0 + \sum_{j=1}^{p} \phi_j(x_j) + \epsilon \tag{1}$$

The smoothing terms in AMs are also called shape functions. Smoothers used for GAMs can be directly translated to AMs under the appropriate AM constraints. Shape functions are commonly achieved by scatterplot smoothers [2], splines [2]–[5], boosted decision stumps [6], decision tree ensembles [7], and neural networks [8]. In [9], both L2 boosting and P-splines are combined to derive a computationally efficient smoother. Regression splines are the most widely used form of shape functions. There is rich literature [8], [10]–[15] on regression spline-based methods that incorporate various constraints to the spline function while

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Sanchez.

achieving lower model complexity compared to other methods. Classical spine formulations, however, are sensitive to knot placement and control point selection. They require prior assumptions on the underlying distribution of data. Meyer [12] proposes shape restricted regression splines that are robust to knot choice. Constraint penalized splines [10] further extend shape restrictions to incorporate additional constraints on monotonicity and convexity.

Each individual shape function applied to x_j can be considerably more complex than a linear fit. The accuracy of AM fits are frequently higher than LM fits. However, both model accuracy and parsimony are equally important characteristics to reduce overfitting. Therefore, fitting algorithms must be capable of identifying covariates that can be sufficiently explained by lower complexity shape functions. For instance, in [4], [16]–[18] Wood *et al.* proposed several smoothness estimation schemes to control the tradeoff between smoothness and fit by minimizing a generalized cross-validation (GCV) score by a regularized maximum likelihood (REML) framework, or by a generalized Akaike's Information Criterion (AIC). For some covariates, nonlinear shape functions can be avoided entirely when linearity can be reasonably assumed. For such cases, methods such as GAMSEL [19] and SPLAM [20] selects between linear or nonlinear fits for each predictor covariate based on a REML framework.

Parsimony can be further improved by removing predictors from the model that has little-to-no impact on the response. Lasso regularization [21] and subset selection are two such techniques to remove insignificant covariates from the regression model. Originally, lasso regularization was defined for purely linear models. Several attempts have been made to extend lasso to the additive model setting. COSSO [22], SpAM [23], GAMSEL [19], and high-dimensional additive models [24] achieve this using a REML framework. The key difference between these models lie in the type of penalty function used. In [25], Marra and Wood introduces a double penalty approach to the REML framework to penalize both the range space and the null space to induce sparsity.

Subset selection identifies the best subset of predictor covariates that has the highest impact on the response. It iteratively adds covariates to the regression model until a predetermined stopping criterion is satisfied. Classical methods of subset selection include the best subset selection [26], forward selection [27], and backward elimination. These methods however are overly aggressive and may therefore disregard predictors with significant correlation to the response after the first iteration. Forward stage-wise selection [28] is a more prudent algorithm that gradually approaches the final model in small steps. The time complexity of the stage-wise procedure is an order of magnitude higher than the classical methods as correlations are evaluated after each step. The Least Angle Regression (LARS) algorithm [28] offers a geometric interpretation of the stage-wise problem that theoretically finds the optimal step size thus eliminating the need for multiple small steps.

Even though LARS gained wide popularity since its introduction and sits at the heart of rapid development, very few attempts have been made to extend it any further. QuasiLAR introduced by Wu [29] extends the application of LARS to Generalized Linear Models. ConvexLAR introduced by Xiao *et al.* [30] generalizes the square-loss function in classical LARS to an arbitrary convex function. Both QuasiLAR and ConvexLAR require solving ordinary differential equations. Khan *et al.* [31] introduces a robust version of LARS for model estimation in the presence of many candidate predictors with outliers. Group Lasso implemented in LARS by Yuan and Lin [32] accounts for the dependence between variables to improve subset selection. Alfons *et al.* [33] extends group selection to robust LARS. Adaptive lasso proposed by Zou [34] improves the oracle properties of LARS. Finally, the elastic net implementation of LARS proposed by Zou and Hastie [35] stabilizes LARS for problems with $p > n$.

This paper attempts to develop LARS in a different direction. The main objectives achieved by the proposed modifications are,

- Replace correlation with association to perform forward stepwise subset selection to obtain the ranked importance of each individual predictor covariate on the response.
- Introduce a nonlinear shape function to transform predictor covariates to improve regression performance.
- Control the complexity of the shape function to avoid overfitting by selecting linear or nonlinear fits for each covariate based on the data.
- Introduce a monotonic nonlinear extension to the algorithm to solve isotonic regression problems.

The proposed algorithm is named Maximal Associated Regression (MAR). MAR takes advantage of the geometrical insights derived from LARS to achieve computational efficiency while extending its applicability for nonlinear mappings. Like LARS, MAR assumes independence between predictors. The linear correlation used to perform variable sequencing in LARS is replaced by association in MAR. At each regression stage, the subset selection procedure updates the existing model with a new predictor covariate that has the highest association to the current residual. A nonlinear transformation is generated by a constrained penalized spline [10] estimate. AIC is used to select between a linear or spline fit. To preserve the LARS solution path paradigm, the model equation decouples the nonlinear covariate transformations $\phi_j(x_j)$ from the regression coefficient vector β compared to the original AM formulation. Thus, the MAR algorithm calculates them independently (2).

$$y = \beta_0 + \sum_{j=1}^{p} \phi_j(x_j)\beta_j + \epsilon \tag{2}$$

MAR can be extended to solve isotonic regression problems by applying a monotonic constraint to the spline transformations. Theoretical aspects of isotonic regression such as the oracle property, asymptotic distribution of estimators,

and the estimability of the nonlinear components have been covered in [36], [37]. This paper only considers estimable problems and focuses on the algorithmic aspect of regression. Existing work on additive isotonic regression focuses on finding all components simultaneously, or by iterative methods such as back-fitting. A recent development by Bergersen *et al.* [14] achieves a sparse model by combining spline basis functions with the L_1 penalty. However, there is no discussion of stepwise subset selection for additive isotonic regression.

We begin Section II by presenting a geometrical interpretation of LARS and the necessary transformations to transition from LARS to MAR. The monotonic extension is introduced in Section III. Validation of the algorithm is done on both synthetic and real datasets in Sections IV and V respectively. The proposed MAR algorithm is compared against LARS, GLM, and the mgcv implementation of GAM [4]. Concluding remarks are presented in Section VI.

II. FROM LARS TO MAR

Least Angle Regression iteratively converges towards a final regression solution $\beta = (\beta_1, \ldots, \beta_p)^T$ by adding a single coefficient β_j at each iteration. A parsimonious β is achieved by terminating the solution path with a stopping criterion that restrains the complexity of the regression model. Two main assumptions drive the classical LARS algorithm: 1) the predictors are independent and 2) their relationship to the response can be captured by a linear correlation coefficient. The resulting solution path is piece-wise linear. The group Lasso proposed by Yuan and Lin [32] evaluates the effects of predictor dependence on estimation accuracy. Linearity of covariates on the other hand assumes $\phi_j(x_j)$ in (2). The proposed MAR algorithm relaxes the second assumption and attempts to detect non-linearity in the solution path. However, to avoid overfitting, MAR must be equipped to detect and preserve linearity when nonlinear transformations are not required. We start by describing the geometric interpretation behind the formulation of LARS to provide a viable basis to describe the motivation behind MAR. We then provide a high-level overview of MAR followed by a component-wise description of the modifications made to detect nonlinearity.

A. GEOMETRIC INTERPRETATION OF LARS REVISITED

The LARS algorithm is initialized by standardizing the input dataset $X \in \mathbb{R}^{n\times p}$ to have zero mean and unit length, eliminating the β_0 component from the regression coefficients. A candidate vector of regression coefficients $\hat{\beta}$ is achieved by minimizing the total squared error defined by (3) subject to a bound λ. The bound on (3) determines the sparsity of $\hat{\beta}$.

$$S(\hat{\beta}) = \min_{\beta\in\mathbb{R}^p} \|y - X\hat{\beta}\|^2 \text{ subject to } \sum_{j=1}^{p} |\hat{\beta}_j| \leq \lambda \qquad (3)$$

Algorithmically, LARS is a stepwise procedure that adds predictor components to $\hat{\beta}$ from a null vector until the full solution is reached. If the quadratic loss function in (3) has been satisfactorily minimized before all components are fitted to the final model, the algorithm will return a sparse solution. At iteration t, the regression coefficient vector is denoted by $\hat{\beta} = \beta_t$ which will have t non-zero active components that is most correlated to the current residual. The active set $\mathcal{A}_t$ keeps track of active predictors and its complement $\mathcal{I}_t$ holds the inactive set. At any step $t > 0$, the predictor with the highest correlation c_t to the residual $r_t = y - X\beta_t$ is added to the new coefficient vector. The index of the newly activated predictor is moved from the inactive set to the active set. For the current inactive predictor matrix $X_{\mathcal{I}_t}$, LARS defines the linear correlation as,

$$c_t = X_{\mathcal{I}_t}{}^T(y - X\beta_t) \qquad (4)$$

Fig. 1 illustrates the vector space representation of LARS transitioning from iteration $t-1$ to t. Note that the dimensions of this abstract vector space is not p (number of predictors) but n (sample size). A completed LARS solution path will contain p vectors, each vector being of n dimensions. At $t-1$, the estimated response is $\mu_{t-1} = X\beta_{t-1}$ with residual r_{t-1}. The projection of the response y onto $t-1$ subspace $\mathcal{S}_{t-1}$ is indicated by y_{t-1}. In classical forward selection, the solution path will progress until y_{t-1} is reached. This will lead to an intermediate ordinary least squares solution at $t-1$, which essentially mask covariates with potentially high covariance from subset selection. However, LARS terminates path propagation when a new predictor covariate $x_{j\in\mathcal{I}_{t-1}}$ from $\mathcal{I}_{t-1}$ becomes equally correlated to the current residual r_{t-1}. Geometrically, the selected predictor x_t makes the smallest angle with r_{t-1}. The new solution path, which now includes x_t, propagates along the unit vector u_t, thus bisecting the angle between μ_{t-1} and x_t. Propagation terminates at step size γ_t when a new covariate from the current inactive set and $\mu_t = \mu_{t-1} + \gamma_t u_t$ becomes as equally correlated to the new residual r_t. The ingenuity of LARS lies in the use of this geometrical insight to find a theoretical estimation for the optimal value of γ_t within a single step.

B. THE MAR ALGORITHM

LARS updates the solution vector by adding covariates based on the highest un-normalized independent correlation (4). Xiao *et al.* [30] interprets this as a least squares loss function. Prediction scores are assigned to each active predictor and the LARS solution path propagates in a direction that maintains the absolute score of all active predictors at identical values. The common absolute score of the solution decays with solution path propagation. Inactive predictors join the solution when their absolute score becomes equal to the common absolute score of active predictors. This perspective links LARS to the steepest decent algorithm in optimization. Using this interpretation, they generalize the LARS algorithm to ConvexLAR that can handle any convex loss function.

MAR use this insight and relaxes the linearity assumption that forms the basis of LARS solution path propagation. Instead of correlation c, MAR adopts association to perform subset selection. This paper uses distance correlation

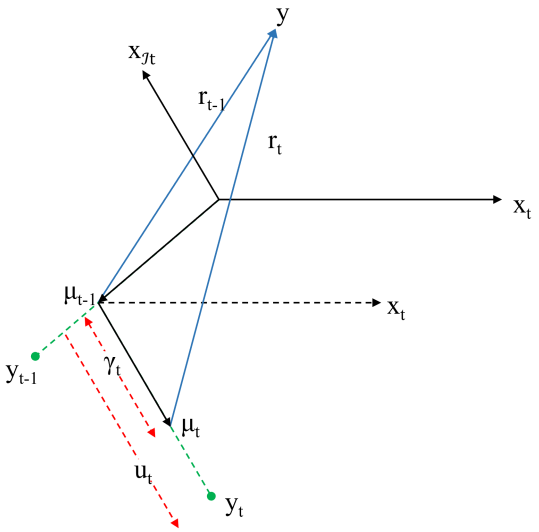

FIGURE 1. Geometric interpretation of the LARS solution path. At iteration $t-1$, the current solution is μ_{t-1}. In the next iteration t, LARS selects the next most correlated covariate and propagates γ_t along the unit vector u_t to achieve the new solution μ_t. If only one more covariate is available, the algorithm terminates by achieving the OLS solutions at iteration $t+1$.

$\mathcal{R}$ [38] to calculate association. For multivariate inference, $\mathcal{R}$ is preferred over other likelihood ratio tests which are inapplicable when the dimensionality of the dataset exceeds the sample size or when assumptions on the distribution do not hold. Furthermore, $\mathcal{R}$ is sensitive to all types of departures from independence, including nonlinear and nonmonotonic dependence. It does not require assumptions on normality for valid inference.

Crucially, MAR does not depend on the type of association used. Thus, $\mathcal{R}$ can be replaced by any other method of association. The performance of the MAR using Kendall's Tau association on synthetic and real datasets are illustrated in the supplementary material. Association is invariant to arbitrary transformations of covariates. Geometrically, this means that the search for the next covariate is not sensitive to any deformation of coordinates in the vector space in Fig. 1. In other words, distortion of predictors will not hide the predictor-response association. This characteristic makes association much more robust against unknown transformations that may be hidden in the data unbeknown to the user. For example, numerical values assigned to categorical labels are often arbitrary. At best, one can expect some association instead of strictly linear correlation between the numerical values and the response y. This deformation of the coordinate system in the vector space will affect the accuracy of the regression model and the reliability of the subset selection operation. Section IV will show some examples where MAR creates better models because it is able to discover hidden nonlinearity in X.

Once a covariate with maximal association is selected, a non-parametric transformation is imposed to linearize the selected covariate with respect to the residual. The choice of

TABLE 1. Maximal associated regression (MAR) algorithm.

Algorithm: Maximal Associated Regression		
Input: Training dataset $X \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$, t_{max}		
Initialize: $\beta = 0$, $\mu = 0$, $\mathcal{A} = (1, \ldots, p)$, $\mathcal{I} = ()$		
while $\mathcal{I}$ not empty $\vee$ $t < t_{max}$		
1	$r_{t-1} \leftarrow y - \mu_{t-1}$	
2	Select x_j from $\mathcal{I}_{t-1}$ with highest association $\mathcal{R}_t$	Section II-D
3	Update $\mathcal{A}_t$ and $\mathcal{I}_t$	
4	$x_j \leftarrow \phi_j(x_j)$	Section II-C
5	Find u_t and γ_t	Section II-E
6	Update β_t and μ_t	
7	$t \leftarrow t+1$	
end		
Output: $\beta \in \mathbb{R}^{p\times t}$ and $\phi \in \mathbb{R}^{n\times p}$		

Three modifications are introduced to transform LARS to MAR. Details can be found on the sub-sections indicated here.

transformation will be elaborated in Section II-C. The complexity of the transformation is controlled by the Akaike's Information Criterion (AIC). When the prediction error does not improve with respect to complexity of the transformation, a linear mapping will be used. The nonlinear transformation will generate a vector space diagram of MAR that is almost identical to Fig. 1. However, unlike LARS, the direction of solution path propagation does not bisect the angle between the residual and the new predictor vector. Section II-E proposes a binary search algorithm to calculate an optimal γ. Table 1 summarizes the MAR procedure in pseudocode. Here, t_{max} is the maximum number of algorithm iterations. The following subsections discusses the modifications in detail.

C. NONLINEAR TRANSFORMATIONS AND MODEL COMPLEXITY

To better illustrate the effects from each modification when converting LARS to MAR, we accompany the remainder of this paper with an example. A synthetic dataset is generated by a normally distributed $X \in \mathbb{R}^{n\times p}$ with mean zero and unit length for $n = 1000$ and $p = 10$. The response distribution, given by (5), is subject to a normally distributed noise component $\epsilon \sim N_n(0, 1^n)$.

$$y = -\frac{1}{8} + \frac{1}{8}e^{x_1} + I(x_2 > 0)^{1/3} - \frac{1}{2}x_3 + \frac{1}{5}\tanh(3x_4) + \epsilon \tag{5}$$

Here, $I(\cdot)$ is an indicator function. Covariate mappings for the dominant predictors estimated by the LARS and the MAR algorithms respectively are illustrated in Fig. 2. It shows that the MAR algorithm creates a better representation of the true nonlinear relationship between the response and the predictors.

Transformation of predictor covariates with respect to r represents the essence of MAR. MAR uses constrained penalized splines [10] to perform nonlinear transformation because the formulation guarantees invariance to knot placement and adherence to global constraints on monotonicity and convexity. Generation of the spline basis functions B_j and its derivatives b_j for the j^{th} predictor covariate is only done once

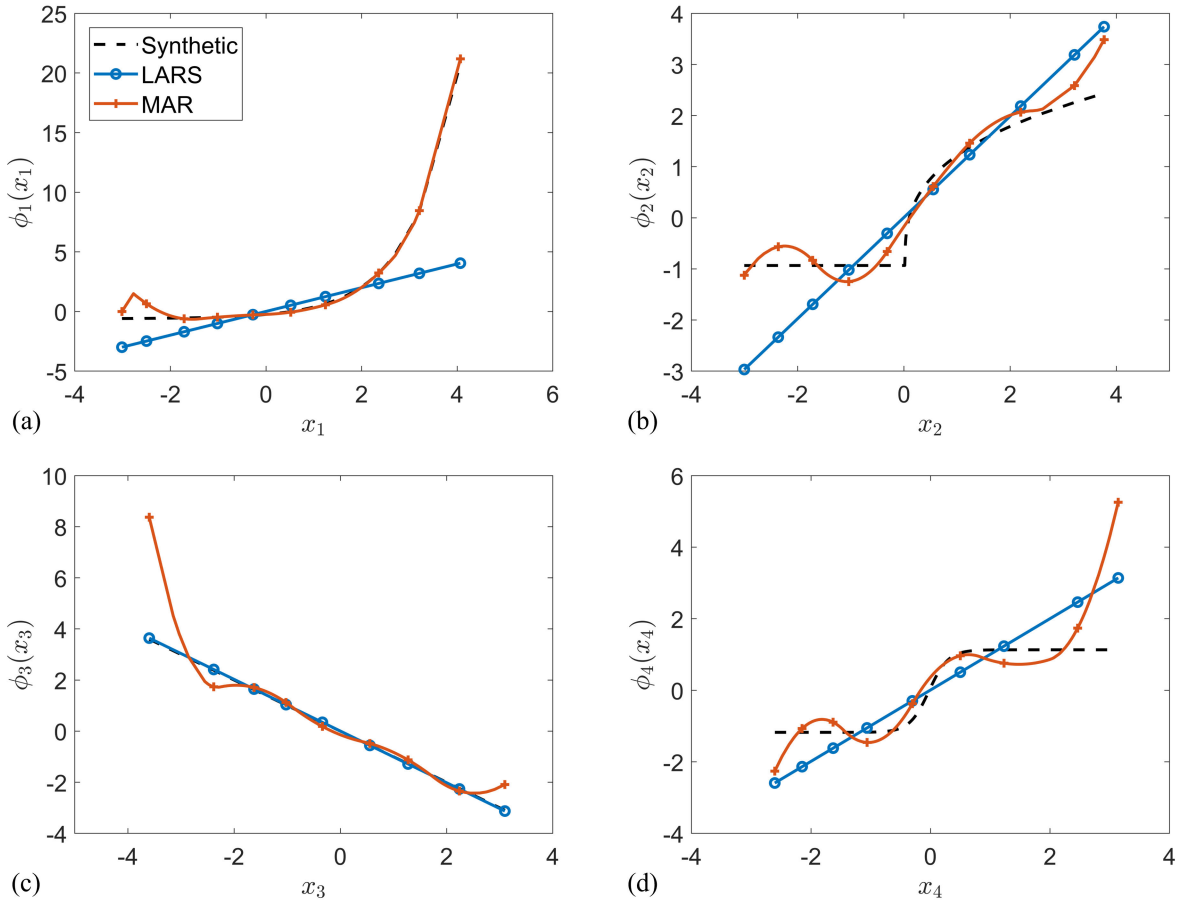

FIGURE 2. The nonlinear mappings of the dominant predictor covariates estimated by the MAR algorithm for the synthetic example (5). The LARS estimate and the actual relationship between the predictors and the response is also given for comparison. MAR achieves a much better fit for the nonlinear predictors in (a), (b), and (c). However, monotonicity constraints are introduced in Section III to reduce the excessive wiggling associated with unconstrained regression.

during initialization of MAR. The degrees of freedom of the spline functions are maintained at $m = 2$ (quadratic spline) to minimize overfitting. The spline complexity is controlled by the number of knots k placed equidistantly along the dataset x_j. Starting from a maximum of k_{max}, the number of knots are reduced for datasets with lower degrees of freedom. Redundant knots are detected by calculating the rank of the Hessian matrix H of the spline basis function. As a minimum of three internal and four external knots are required for a quadratic polynomial, $k_{min} = 7$ is imposed.

MAR achieves monotonic nonlinear mappings for dominant covariates by reducing the RSS between r and the fitted model. To complete the spline function, control points C_j are estimated using quadratic programming (8) to minimize the RSS. It must be noted that this formulation only holds true for quadratic spline basis functions.

$$H = B_j{}^T B_j \tag{6}$$

$$\nabla f = r^T B_j \tag{7}$$

$$C_j = \min_{C_j \in \mathbb{R}^q} \frac{1}{2} C_j{}^T H C_j + \nabla f C_j \tag{8}$$

Table 2 is the predictor covariate transformation procedure. To reduce overfitting, a linear estimate replaces the spline fits when it can sufficiently model variance in the dataset. Once the spline control points are generated from (8), the spline fit is compared against a linear approximation using the Akaike's Information Criterion (AIC). The AIC hyperparameters a_1 and a_2 controls the amount of penalization on model complexity. Lower values favor complex models while higher values are more likely to select simpler models. A lower AIC magnitude is given to models with higher prediction power and a lower complexity.

D. REPLACING PEARSON'S CORRELATION WITH DISTANCE CORRELATION

Solution path propagation in LARS depend on the linear correlation between the predictor covariates and the residual vector r. At each iteration, correlation between the active set covariates and r decreases as the solution path progresses towards an OLS solution. In contrast, the inactive set covariates have a negative correlation with r. The objective of LARS is to estimate the intersection point γ where the minimum correlation of the active set and the maximum

TABLE 2. Predictor covariate transformation algorithm for MAR.

Algorithm: Predictor Covariate Transformation
Input: $x_j \in \mathbb{R}^n, r \in \mathbb{R}^n, k, m, q, B_j \in \mathbb{R}^{n\times q}, b_j \in \mathbb{R}^{n\times q}$
Initialize: $a_1 = 2, a_2 = 3 + q$
$\beta \leftarrow {x_j}^T r / {x_j}^T x_j$
$RSS_1 \leftarrow \sum_{i=1}^{n}(r_i - \hat{r}_i)^2$ for $\hat{r} \leftarrow x_j\beta$
$AIC_1 \leftarrow a_1 + n\ln(RSS_1)$
Find H and ∇f
Approximate C_j
$RSS_2 \leftarrow \sum_{i=1}^{n}(r_i - \hat{r}_i)^2$ for $\hat{r} \leftarrow B_jC_j$
$AIC_2 \leftarrow a_2 + n\ln(RSS_2)$
if $AIC_1 \leq AIC_2$ **then**
  $\phi_j(x_j) \leftarrow x_j$
else
  $\phi_j(x_j) \leftarrow B_jC_j$
end
Output: $\phi_j(x_j) \in \mathbb{R}^n$

Akaike's Information Criterion (AIC) performs model inference on the covariate transformations to select between a linear or a nonlinear transformation. A quadratic spline model is used for nonlinear mappings. Control points for the spline transformations are generated by a quadratic approximation.

TABLE 3. Binary search algorithm to estimate solution path termination.

Algorithm: Binary Search
Input: $X \in \mathbb{R}^{n\times p}, y \in \mathbb{R}^n, \mu, u \in \mathbb{R}^n, \mathcal{A} \in \mathbb{R}^{1\times(p-t)}, \mathcal{I} \in \mathbb{R}^{1\times t}, t_{max}, \varphi_{max}$
Initialize: $\gamma_1 = 0, \gamma_2 = 1, r_1 = y - \mu, r_2 = y - (\mu + u), \varphi = \infty, t = 1$
Find $\mathcal{R}_1(\mathcal{A})$ and $\mathcal{R}_1(\mathcal{I})$ at r_1
Find $\mathcal{R}_2(\mathcal{A})$ and $\mathcal{R}_2(\mathcal{I})$ at r_2
while $t < t_{max} \vee \varphi > \varphi_{max}$
  Find γ
  $r_\gamma \leftarrow y - (\mu + u\gamma)$
  Find $\mathcal{R}_\gamma(\mathcal{A})$ and $\mathcal{R}_\gamma(\mathcal{I})$ at r_γ
  if $AIC_1 \leq AIC_2$ **then**
    $\gamma_1 \leftarrow \gamma$
    $\mathcal{R}_1(\mathcal{A}) \leftarrow \mathcal{R}_\gamma(\mathcal{A})$
    $\mathcal{R}_1(\mathcal{I}) \leftarrow \mathcal{R}_\gamma(\mathcal{I})$
  else
    $\gamma_2 \leftarrow \gamma$
    $\mathcal{R}_2(\mathcal{A}) \leftarrow \mathcal{R}_\gamma(\mathcal{A})$
    $\mathcal{R}_2(\mathcal{I}) \leftarrow \mathcal{R}_\gamma(\mathcal{I})$
  end
  $t \leftarrow t + 1$
  $\varphi \leftarrow \gamma_2 - \gamma_1$
end
Output: γ

This algorithm replaces the analytical method in LARS due to the nonlinear relationship between association and the current residual.

correlation of the inactive set covariates becomes equally correlated to r. Thus, LARS uses the un-normalized Pearson's Correlation c_j to accurately estimate the linear correlation between the current residual and the j^{th} predictor covariate.

In MAR, distance correlation $\mathcal{R}$ replaces linear correlation to measure association. Analogous to c_j that estimates linear dependence between random vectors, $\mathcal{R}_j$ is a generalized estimate of correlation that extends to the unconstrained non-monotonic space. Initially proposed by Székely *et al.* [39] and later extended in [40], $\mathcal{R}_j$ provides a non-parametric estimate of association. Unlike classical methods of association, $\mathcal{R}_j$ leads to an accurate estimate even when the predictor dimensions exceed sample size or when prior assumptions on the underlying distribution of X do not hold. Distance correlation satisfies $0 \leq \mathcal{R}_j \leq 1$, and $\mathcal{R}_j = 0$ only if the covariate vectors are independent.

The original formulation of $\mathcal{R}$ has a computational complexity of order $\mathcal{O}(n^2)$ restricting its widespread adaptation for applications with large sample sizes. MAR uses the $\mathcal{R}$ formulation proposed by Huo and Székely [38] that reduces its complexity to $\mathcal{O}(n\log n)$ making it comparable to other computationally efficient algorithms. The optimal γ is found by a binary search algorithm that converge to the optimum solution over multiple iterations.

E. FINDING THE OPTIMAL STEP SIZE

In LARS, the optimal step size γ is estimated based on correlation (4). The correlation is, in turn, related to the angular separation in vector space. The LARS solution path propagates in a direction that maintains the angular symmetry between the active set covariates. MAR does not maintain this symmetry. Association by $\mathcal{R}$ depends on a rank-order system that is not affected by the magnitudes of individual elements. For instance, two predictor covariates, x_1 and x_2 will have the same association to r irrespective of the differences in element-wise magnitude, provided that elements are arranged in the same ascending order.

The non-linearity of $\mathcal{R}$ causes the estimation of γ to be non-trivial. Table 3, proposes an iterative solution for MAR to estimate γ based on a binary search algorithm. The algorithm starts by evaluating the association at the two extremes of the solution path u. The range of possible γ values are represented by γ_1 and γ_2. The minimum association of the active set and the maximum association of the inactive set covariates at γ_1 is represented by $\mathcal{R}_1(\mathcal{A})$ and $\mathcal{R}_1(\mathcal{I})$. Association at γ_2 is represented by $\mathcal{R}_2(\mathcal{A})$ and $\mathcal{R}_2(\mathcal{I})$. The current estimate of γ is updated with (9) at each iteration and replaces either γ_1 or γ_2. The algorithm progresses until convergence to within φ_{max} or until the maximum number of iterations t_{max} is reached.

$$\gamma = \gamma_1 + \frac{\gamma_2 - \gamma_1}{1 - \frac{\mathcal{R}_2(\mathcal{A}) - \mathcal{R}_2(\mathcal{I})}{\mathcal{R}_1(\mathcal{A}) - \mathcal{R}_1(\mathcal{I})}} \tag{9}$$

The unit vector u_t for the t^{th} MAR iteration can be obtained using the original formulation proposed by LARS,

$$u_t = X_{\mathcal{A}_t}\beta_t(\text{ols}) - \mu_t \tag{10}$$

where $X_{\mathcal{A}_t}$, $\beta_t(\text{ols})$, and μ_t represent the active set predictor covariates, the t^{th} step ordinary least squares estimate, and the t^{th} step estimated response respectively.

III. MONOTONIC EXTENSION OF MAR

Shape restricted regression plays an important role in model estimation when *a priori* knowledge of predictor behavior is available. Monotonically constrained shape restrictions are widely imposed for many practical applications in fields such as biology, medicine, and statistics [41], [42]. Monotonicity is implicitly assumed in linear regression, but it is violated

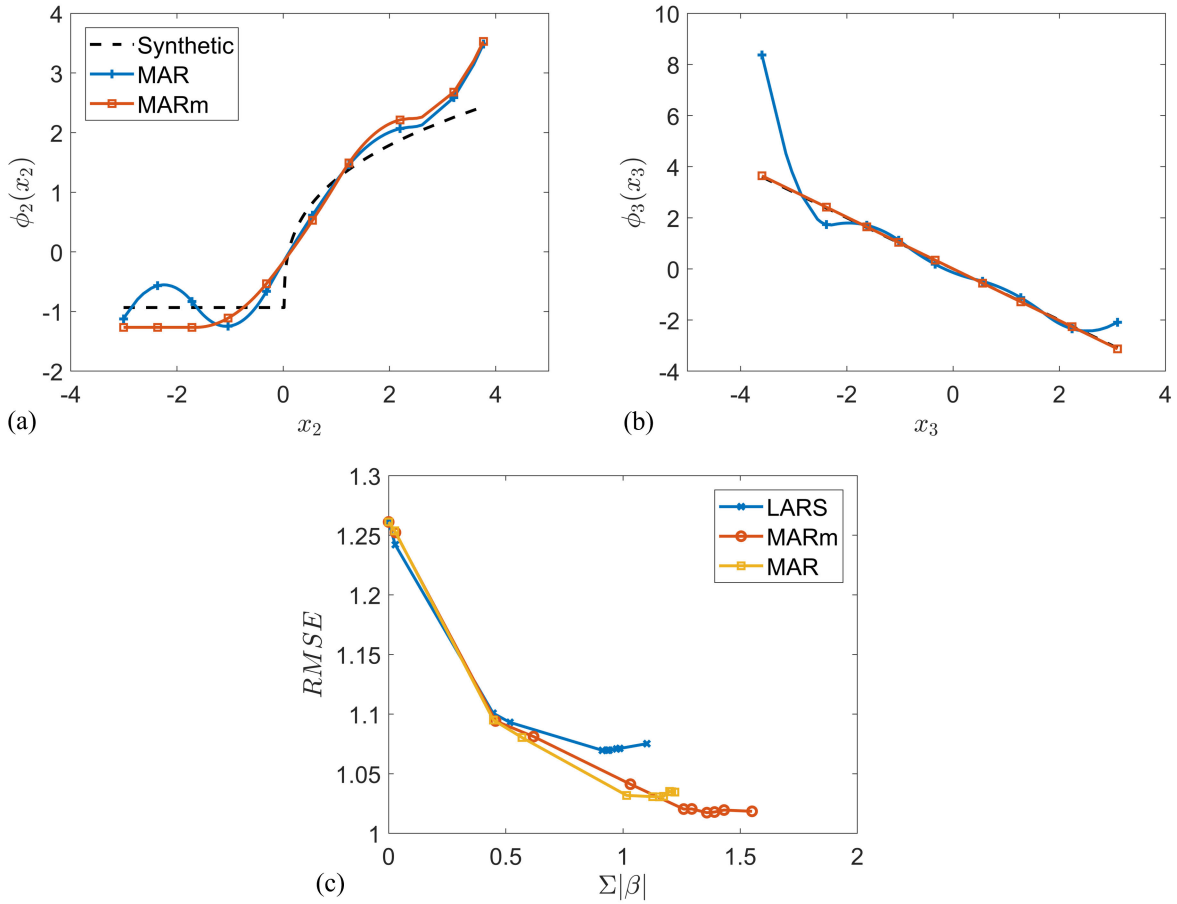

FIGURE 3. Monotonically constrained spline mappings for the synthetic example (5) for predictor covariates (a) x_2 and (b) x_3. The RMSE performance for LARS, MAR, and monotonically constrained MAR (MARm) is given by (c). MARm is able to better minimize the uncontrolled wiggling of the nonlinear fits resulting in better RMSE performance.

by the unconstrained nonlinear transformations introduced by the MAR algorithm. For the synthetic example, Fig. 3 shows that imposing the monotonicity assumption results in a much better fit with lower RMSE.

This section introduces a monotonicity constraint to the MAR algorithm (MARm). In Section II-C, MAR transforms input covariates using unconstrained penalized spline functions. Monotonicity for penalized splines can be achieved by introducing the strictly positive constraint proposed by Meyer [10] to the control point calculation in (8). In practice, monotonicity will only be applied to a specific subset of predictor covariates. This can be achieved by independently applying (11) to the appropriate predictor covariates indexed by j.

$$b_j C_j \geq 0 \tag{11}$$

Compared to alternative spline formulations, constrained penalized splines offers two main advantages: 1) it allows for a high degree of flexibility without the excessive wiggling typically associated with over-fitting in non-parametric regression, and 2) it guarantees global conformity of the spline to monotonicity constraints [10]. This makes constrained penalized splines especially viable for nonlinear model fitting. Although the synthetic example shows a better RMSE under MARm, it must be noted that the monotonicity constraint does not always guarantee a better fit. Some real-world datasets benefit from the added flexibility given by unconstrained spline models. Thus, it is recommended that the monotonicity constraints only be applied when prior knowledge on the behavior of the predictor covariates is available.

For the synthetic example, the growth of the regression coefficients for each algorithm is illustrated in Fig. 4. Unlike LARS, MAR estimates growth in a strictly positive direction as the nonlinear transformations are capable of automatic sign inversion when negative associations are detected. This capability can be switched on and off without any loss of information. In the case of MAR, the direction of association will be encoded in the transformation instead of the regression vector β. Thus, the authors find that it is advantageous to work with $|\beta|$ instead of β when nonlinear transformations are involved. Furthermore, both MAR in Fig. 4 (a) and MARm in Fig. 4 (b) show an identical order of subset selection and similar shapes in $|\beta|$ propagation. This is an expected

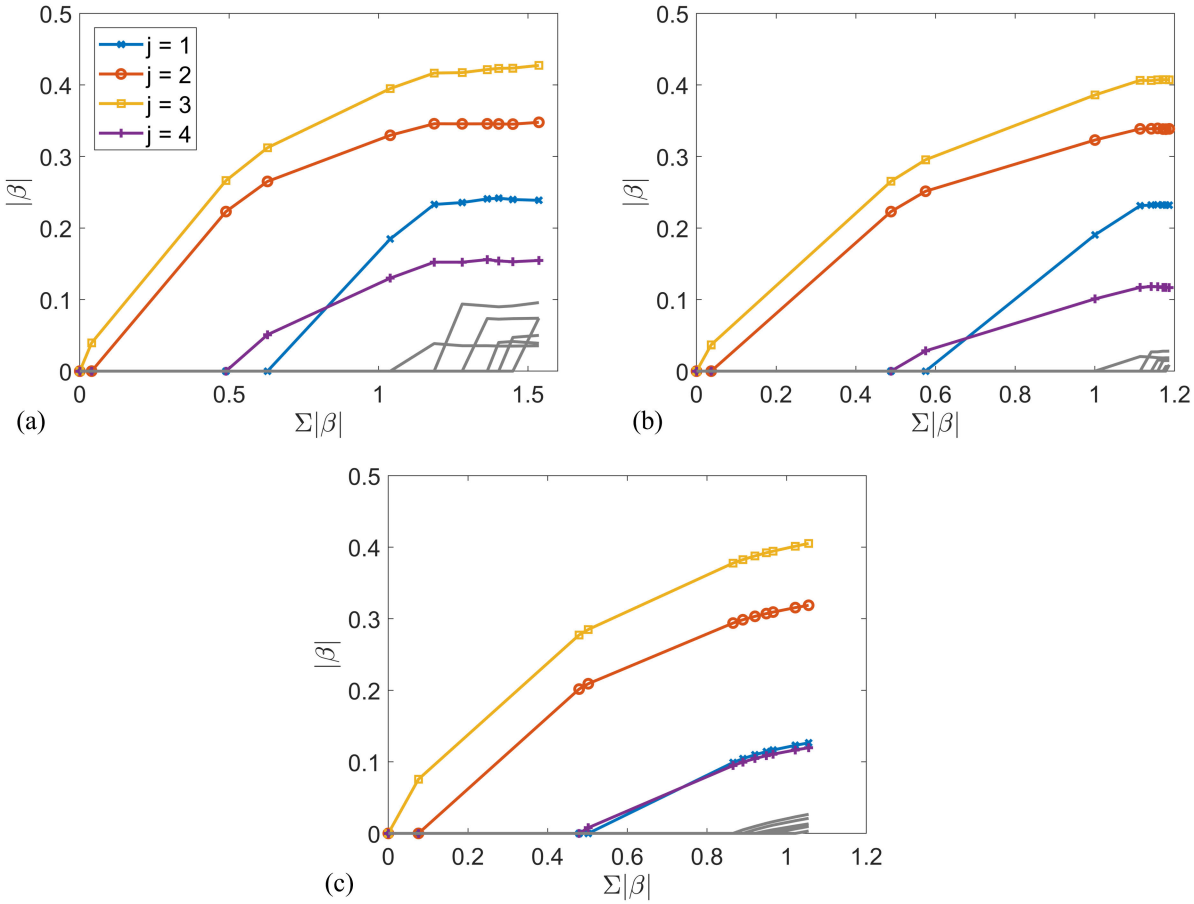

FIGURE 4. The magnitude of the regression coefficients with increasing model complexity $\sum|\beta|$ for (a) MAR, (b) MARm, and (c) LARS algorithms. All algorithms accurately select the dominant feature set. Both MAR and MARm have identical regression coefficients due to monotonicity of the feature set.

phenomenon as both algorithms are based on $\mathcal{R}$ for small datasets with predominantly monotonic covariates.

IV. SIMULATION STUDIES

Synthetic datasets provide a good testing ground to study the behavior of the proposed algorithms. We compare the prediction performance of MAR and monotonically constrained MAR (MARm) against 3 existing regression algorithms: LARS, Generalized Linear Models (GLM), and Generalized Additive Models (GAM). LARS and GLM provide a baseline reference as they are both widely used algorithms to fit linear regression models. GLM is implemented using the inbuilt R function 'glm' while LARS, MAR, and MARm are implemented in MATLAB. GAM is implemented using the R package 'mgcv' [4]. GAM more closely resembles MAR as both algorithms introduce nonlinear transformations to the predictor covariates. For a fair comparison with MAR, both GAM and GLM are implemented under the Gaussian assumption with the unity link function.

Eight synthetic models with five dominant predictors are considered. They are illustrated in Table 4. Covariates on the predictor matrix $X \in \mathbb{R}^{n\times p}$ for $n = 1000$ and $p = 10$ are generated according to a centered multivariate normal distribution $N_p(0, \sigma^2)$. A normally distributed noise component $\epsilon \sim N_n(0, 1^2)$ is added to all datasets. A non-zeros covariance of $q = 0.5^{|j-k|}$ between predictor covariates x_j and x_k or a signal to noise ratio of $snr = 5$ are introduced for selected models. Categorical components are induced by trichotomizing covariates into tertiles such that $x_j \in [0, 1, 2]$. Indicator functions for the categorical components are defined by $I(\cdot)$.

Accuracy of the regression fits generated by each algorithm is assessed by root mean squared error (RMSE). Ten-fold cross-validation is used to ensure stability of the models. Thus, both mean and standard deviations of the performance parameters are obtained. Numerical results are tabulated in the supplementary material (Table S1). For the LARS and MAR type algorithms, parameters N_5 and N_{max} indicate the number of dominant covariates on the active set at algorithm iteration $t = 5$ and at solution path termination t_{max}. Note that for LARS and MAR, t_{max} is not always 10 as subset selection terminate prematurely if the solution path becomes badly conditioned.

TABLE 4. Synthetic models for MAR validation.

Model	Model Equation	SNR	Covariance
1	$y = \frac{1}{2} + x_1^3 + \frac{2}{\sqrt{\pi}} \int_{x_2}^{\infty} e^{-\delta^2}\, d\delta - \frac{1}{3} e^{x_5} - 2x_8^3 + (x_9 I(x_9 > 0))^{1/3} + \epsilon$	×	×
2		×	✓
3	$y = -\frac{1}{5} - 2x_2^2 + \frac{1}{4}\cos(x_3) - \frac{1}{3}\cosh(2x_7) + \tanh(3x_8) + \tanh(3x_9)^2 + \epsilon$	×	×
4		×	✓
5		✓	×
6	$y = \frac{1}{4} + 2x_1 - \frac{1}{8}x_3 - \frac{1}{5}x_5 + 5x_6 - 3x_8 + \epsilon$	×	×
7		✓	×
8	$y = 2 - \frac{1}{3} I(x_2 = 2) - I(x_3 = 0, 2) - 2I(x_5 = 1) + I(x_7 = 0, 1, 2) + \frac{1}{2} I(x_8 = 1, 2) + \epsilon$	×	×

Models 1 and 2 have monotonic and models 3,4, and 5 have nonmonotonic covariates. Models 6 and 7 are linear. Model 8 is categorical.
Models with specifically applied signal to noise ratios or covariances are also indicated.

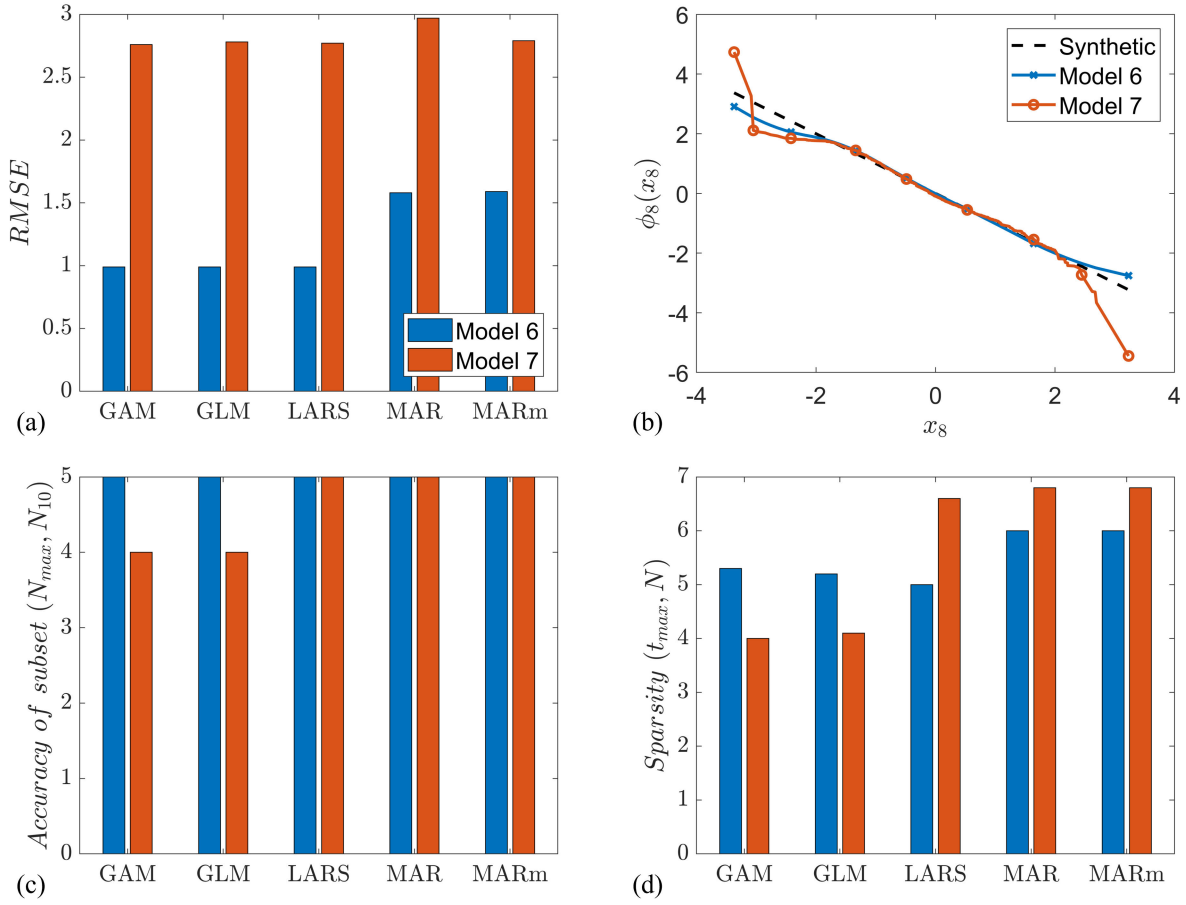

FIGURE 5. Algorithm performance (means from cross-validation) for synthetic models 6 and 7 illustrating (a) RMSE, (c) number of dominant covariates at solution path termination to indicates subset selection accuracy, and (d) the number of covariates in the final model to indicates sparsity. Overfitting of MAR splines due to increased SNR is illustrated in (b) for x_8. The nosier spline fits for model 7 lead to a consistently higher RMSE in (a).

GLM and GAM does not perform subset selection. Instead, it performs a hypothesis test on each predictor covariate. The null hypothesis assumes that the j^{th} predictor covariate is sparse $\beta_j = 0$. Thus, predictors with p-values less than 0.05 are approximated as dominant covariates. Hence, parameters N and N_{10} indicate the total number of covariates (dominant + noise) and the true dominant covariates in the final model respectively.

A t-test is conducted to verify the statistical significance between different quality measures. The performance of MAR and MARm are compared against other algorithms at a 5% significance level. Each dataset is treated independently, their cross-validation results are used to test statistical significance. Both the t-test results and their corresponding p-values are provided in the supplementary material (Tables S2 and S3).

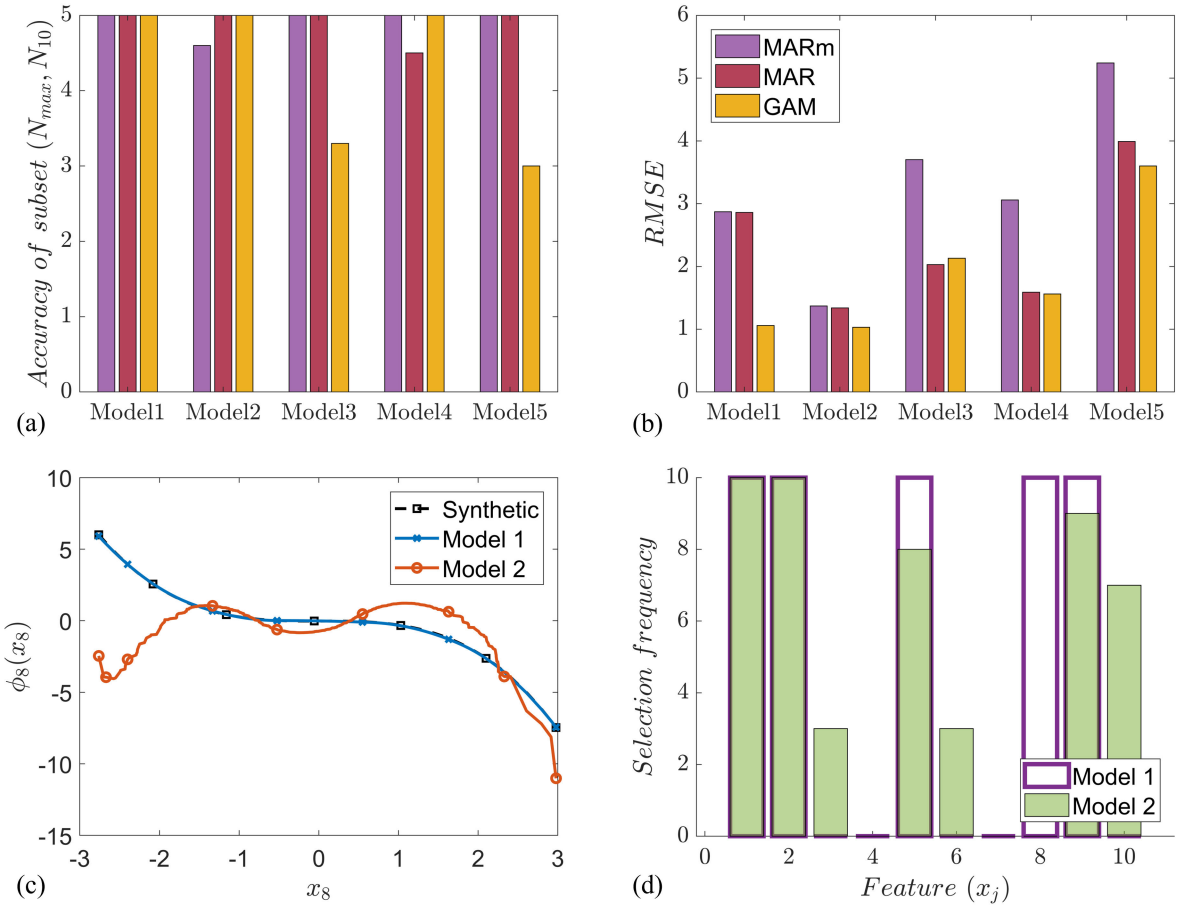

FIGURE 6. Algorithm performance (means from cross-validation) for synthetic models 1 to 5 illustrating the (a) subset selection accuracy and (b) RMSE. Plot (a) shows that MAR selects a more consistent subset compared to GAM. Model fitting performance due to covariance between predictor covariates are assessed by: (c) unconstrained spline fits by MAR for x_8 for models 1 and 2 and (d) the feature selection frequency for models 1 and 2.

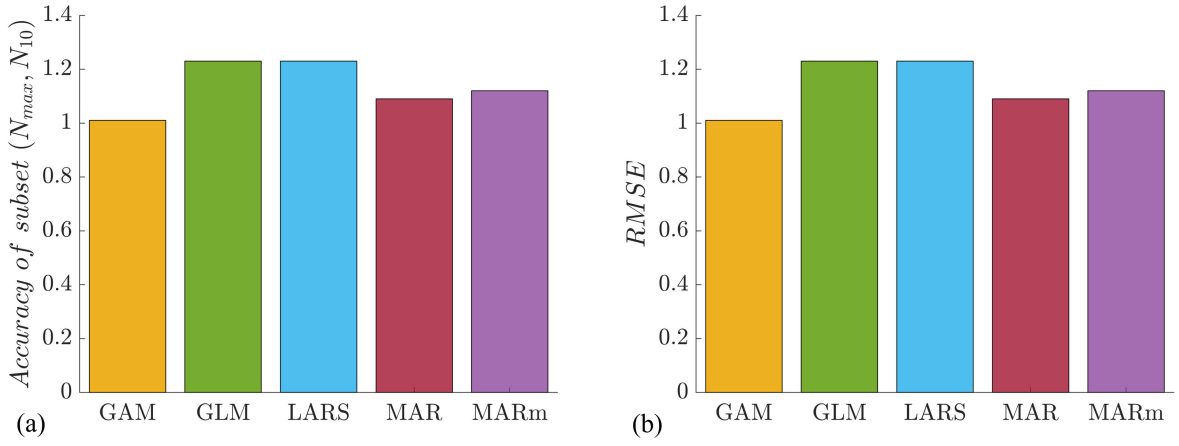

FIGURE 7. Algorithm performance (means from cross-validation) for synthetic model 8 illustrating the (a) subset selection accuracy and (b) RMSE performance.

A. MODEL ESTIMATION WITH LINEAR COVARIATES

For purely linear regression problems depicted by models 6 and 7, LARS and GLM maintain a significant advantage over MAR as covariate transformations in MAR encourage overfitting. For model 6, the RMSE performance in Fig. 5 (a) indicate that despite the noise introduced by the spline fits,

GAM maintains a comparable RMSE to GLM and LARS under low noise conditions. However, the RMSE performance for model 7 shows that MARm achieves an equivalent RMSE when the SNR increases. The performance under higher noise levels is an important factor in model validation as real datasets can contain substantial noise levels. Illustrated by Fig. 5 (b), the slightly poorer RMSE in MAR is due to overfitting of the unconstrained spline model. Thus, MARm is more suitable for high noise datasets.

Depicted by Fig. 5 (c), MAR however has a superior subset selection accuracy compared to GAM with increasing SNR. For model 7, GAM only detects 4 of the 5 dominant covariates while LARS, MAR, and MARm detect all 5. Although GAM leads to a sparser model as seen in Fig. 5 (d), the subset selected by GAM is incomplete. Hence MAR type algorithms are superior when subset selection is of particular importance.

B. MODEL ESTIMATION WITH NONLINEAR COVARIATES

Synthetic models 1 to 5 are used to assess algorithm performance on nonlinear datasets from which 1 and 2 contains strictly monotonic covariates. Nonmonotonic predictors are included in models 3 to 5. Performance of the linear regression algorithms is considerably worse compared to GAM, MAR, and MARm. Thus, in Fig. 6, we only compare performance of additive regression. Numerical performance parameters for all algorithms are available in Table S1 of the supplementary material.

From Fig. 6 (a), it is apparent that the accuracy of subset selection for models 1,3, and 5 by MAR is consistently higher than for GAM. Accuracy of the selected subset in MAR type algorithms become much more evident for models containing nonmonotonic covariates.

From Fig. 6 (b), the RMSE performance of GAM is higher compared to MAR type algorithms. Except for model 1, RMSE of GAM is closely followed by MARm. Models containing high covariance between covariates (models 2, 4) shows a better RMSE compared to models with independent covariates (models 1, 3). This is a false indication of the accuracy of the model fit. As exemplified in Fig. 6 (c), spline fit for x_8 show that MAR is capable of precisely identifying the predictor response relationship for model 1 while being unable to accurately estimate the spline fit for model 2. At $t = 5$, Fig. 6 (d) indicates that the feature selection frequency by cross-validation lead to unstable subsets for model 2 compared to model 1. Thus, GAM and MAR type algorithms are not suitable to fit data with high covariance between covariates. They can be replaced by algorithms specially designed for grouped selection [32]. For real datasets, the Jaccard coefficient can be used to detect instability in subset selection (Section V).

C. MODEL ESTIMATION WITH CATEGORICAL COVARIATES

Model 8 evaluates algorithm performance on purely categorical covariates. Fig. 7 (a) illustrate that compared to GAM, the superiority of subset selection accuracy of MAR type algorithms is maintained for categorical features. In (b), both GAM and MAR type algorithms are shown to achieve comparable RMSE performance which is lower than what is achieved by linear algorithms.

TABLE 5. Summary of validation datasets.

Dataset	Repository	Samples	Predictors
Concrete	UCI	1030	8
Grid Stability	UCI	10000	12
Power Plant	UCI	9568	4
Auto Price	UCI	159	15
CPU Performance	Delve	209	7
Electricity	CMU	55	2
Boston Housing	CMU	506	13
PW Linear	[47]	200	10
Pyrimidines	[48]	74	26
Stock	CMU	950	9
Gas Consumption	UCI	27	4
Bodyfat	CMU	252	14
King County	[46]	21613	18

V. DATA ILLUSTRATIONS

In this section, we evaluate the performance of MAR on 13 public domain datasets obtained from online sources: University of Toronto (Delve) repository [43], University of California (UCI) repository [44], Carnegie Mellon University (CMU) repository [45], the King County house price dataset [46], and from selected publications [47], [48]. A summary of each dataset is depicted in Table 5.

Cross-validated performance of MAR and MARm are compared against LARS, GAM, and GLM algorithms. The statistical significance of the difference for each performance parameter is analyzed in Section V-A. Numerical results of the performance parameters for the first 11 datasets are available in the supplementary material (Tables S4 and S5). Additional analysis of the Pyrimidine dataset and the gas consumption dataset is given in Sections V-B and V-C. The 2 remaining datasets are used for a qualitative analysis of the regression algorithms in Sections V-D and V-E. Accuracy of the model fits are assessed by the normalized RMSE (nRMSE) defined as $RMSE/|\bar{y}|$. Model sparsity is estimated based on the number of active covariates N in the output model. For GAM and GLM, a hypothesis test evaluates the importance of each predictor covariate. Dominant predictors achieve p-values less than 0.05. For LARS, MAR, and MARm, the full solution path β is generated. Subsequently, AIC is used for model inference for all possible sparse models. An adjusted version of AIC (AICc) [49] is used for models with a limited sample size where the ratio between the test set size and the number of free parameters is > 40.

The cross-validated nRMSE accuracy only gives a partial representation of model fitness. Subset selection accuracy is as equally important in determining generalizability of the model to new data. Excessive variation in the active covariates included in the models during cross-validation should give us pause. If consensus cannot be achieved on the dominant features selected during cross-validation, the fitting technique

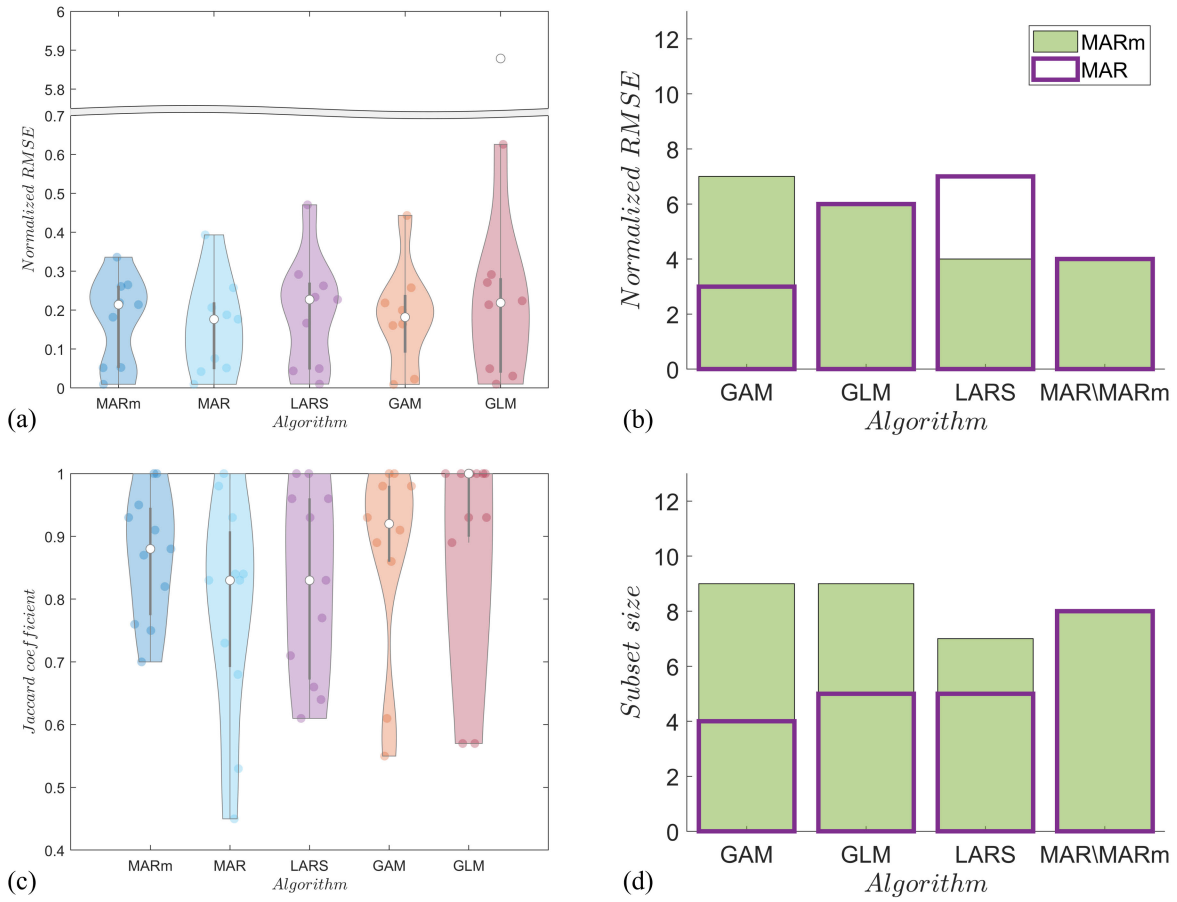

FIGURE 8. The normalized RMSE (nRMSE) performance for each algorithm is illustrated in (a) with the grid stability and PW linear datasets removed due to bad nRMSE by all algorithms. The pyrimidines dataset is shown as an outlier for GLM. The Jaccard coefficient is illustrated in (c) to indicate algorithm stability. A t-test is applied to each dataset independently to benchmark algorithm performance against MAR and MARm. The frequency of the null hypothesis rejection is illustrated as histograms for (b) nRMSE and (d) the subset size N.

must be changed to a more suitable algorithm. Stability of subset selection can be assessed using the Jaccard coefficient (J) [50]. Given multiple models generated by cross-validation, J is the ratio between the pairwise intersection of features and their union. Bounded between 0 and 1, a value of $J = 1$ indicates perfect similarity between feature sets.

A. EVALUATION OF REGRESSION ALGORITHMS

A t-test is performed on the cross-validation results for each dataset independently to verify the statistical significance between different quality parameters. Performance of the benchmark algorithms are compared against MAR and MARm at a 5% significance level. Histograms in Fig. 8 (b) and (d) illustrate the frequency of rejection of the null hypothesis for each quality parameter. Numerical results are available in the supplementary material (Table S6 and S7). Numerical results from cross validation for nRMSE and the Jaccard coefficient (J) are illustrated by the violin plots [51] in Fig. 8 (a) and (c).

Accuracy of each algorithm is analyzed in Fig. 8 (a). It shows that GLM has a significantly poorer accuracy compared to all other algorithms. This is more evident in the Pyrimidines dataset, depicted as a outlier in (a). Comparing nonlinear algorithms, GAM maintains the highest accuracy for most datasets followed by MAR. The statistical significance of the difference in nRMSE is analyzed in (b). Comparing GAM and MAR shows that only 3 out of the 13 datasets has a significantly higher nRMSE in GAM. Thus, in terms of accuracy, MAR is a close followup to GAM. Both accuracy and parsimony in MAR can be further improved by only applying the monotonicity constraints to selected variables. However this entails additional pre-processing for variable selection that is beyond the scope of our paper. It must also be noted that a further disadvantage of using GAM is that it is unable to generate models for sample deficient datasets. For instance, it did not return a result for the gas consumption dataset.

Model sparsity by each algorithms is illustrated in Fig. 8 (d). It shows that MARm generates models with higher sparsity compared to all other algorithms (numerical results in Tables S4 and S5). This is a clear indication that the monotonicity constraint imposes a form of regularization on MAR that prevents weaker predictors from entering the model. The stability of the algorithm is analyzed in Fig. 8 (c) which

TABLE 6. Algorithm performance on the pyrimidines dataset.

Algorithm	RMSE	nRMSE	N	J
LARS	**0.11** (*0.05*)	**0.17** (*0.08*)	**1.3** (*0.48*)	**0.77**
MARm	**0.12** (*0.04*)	**0.18** (*0.06*)	**1** (*0*)	**0.87**
MAR	**0.05** (*0.02*)	**0.08** (*0.03*)	**16.1** (*6.97*)	**0.53**
GAM	**0.17** (*0.1*)	**0.26** (*0.02*)	**8.9** (*2.33*)	**0.55**
GLM	**3.88** (*11.96*)	**5.88** (*18.12*)	**6.7** (*1.89*)	**0.57**

Evaluation parameters for accuracy (RMSE), normalized RMSE (nRMSE), number of non-zero covariates (N), and the Jaccard coefficient (J) are considered.

shows that GAM and MARm are the most stable. However, stability of GAM is distributed across a larger spectrum while MARm consistently lies closer to the upper quadrant of J for all datasets.

B. INTERPRETABILITY OF REGRESSION MODELS: PYRIMIDINES DATASET

The pyrimidines dataset originally analyzed by Hirst *et al.* [52] models the quantitative structure-activity relationship (QSAR) of the inhibition of dihydrofolate reductase (DHFR) by pyrimidines. QSAR is a process in which physicochemical properties of a series of chemical compounds are linked to biological or chemical activity by an empirical equation. The dataset contains structural information of 74 pyrimidines, with 27 predictor covariates produced by 3 positions of chemical activity with 9 attributes per position. The response is $1/\log(K_i)$, where K_i is the inhibition constant that is experimentally measured.

QSAR based modelling has 2 main objectives, to identify an empirical equation that links chemical properties of a compound to its response, and to create a set of easily interpretable decision rules that lead to understanding the dominant attributes affecting the response. Hirst *et al.* [52] points out that while stepwise linear regression meets both these criterion, it is incapable of automatically detecting nonlinear dependence which has been reported in QSAR literature. As such, artificial neural networks (ANN), random forest, support vector machines, k-nearest neighbors, and naïve bias classifiers are used to replace simple linear regression [53]. However, interpretation of decision rules derived from these methods require significantly more work. Furthermore, Hirst *et al.* show that application of ANN to the dataset does not show a statistically significant improvement in performance compared to stepwise linear regression.

MAR provides an ideal solution to these problems as it maintains the interpretability of the solution path while detecting nonlinear dependence. Performance of the regression algorithms are validated in Table 6. The validation agrees with [52] where GLM shows a significantly poor regression performance compared to other algorithms. Interestingly however, LARS maintains nRMSE comparable to the nonlinear regression models. This is because LARS type algorithms are more suitable to handle sample deficient datasets (Section V-C). However, MARm maintains a clear superiority in-terms of accuracy, stability, and sparsity compared to all other algorithms.

TABLE 7. Algorithm performance on the bodyfat dataset.

Algorithm	RMSE	nRMSE	N	J
Bodyfat (x_1)				
Siri	1.26	6.58	-	-
LARS	**0.98** (*0.99*)	**5.12** (*5.17*)	**3.1** (*1.2*)	**0.67**
MARm	**1.38** (*1.78*)	**7.21** (*9.30*)	**1** (*0*)	**1**
MAR	**1.46** (*1.84*)	**7.62** (*9.61*)	**3.3** (*3.5*)	**0.61**
GAM	**1.29** (*1.36*)	**6.74** (*7.10*)	**3.9** (*1.45*)	**0.58**
GLM	**1.02** (*1*)	**5.33** (*5.22*)	**1** (*0*)	**1**
Bodyfat ($\sim x_1$)				
Jackson & Pollock	14.60	76.24	-	-
Tran & Weltman	19.90	103.92	-	-
LARS	**4.97** (*0.56*)	**25.95** (*2.92*)	**1.2** (*0.63*)	**0.87**
MARm	**6** (*0.84*)	**31.33** (*4.39*)	**1.8** (*0.42*)	**0.88**
MAR	**4.91** (*1.32*)	**25.64** (*6.89*)	**8.4** (*2.99*)	**0.71**
GAM	**4.46** (*0.42*)	**23.29** (*2.19*)	**5.5** (*0.71*)	**0.56**
GLM	**4.41** (*0.54*)	**23.03** (*2.82*)	**3.8** (*1.03*)	**0.65**

The regression algorithms are compared against existing bodyfat equations by Siri, Jackson & Pollock, and Tran & Weltman.

C. MANAGING A LIMITED NUMBER OF OBSERVATIONS: GAS CONSUMPTION DATASET

A limiting factor commonly encountered in most medical and pharmaceutical research applications is where the predictor covariates greatly outnumber the available sample size. In genetic research for instance, microarray datasets contain thousands of gene interactions (predictors) and only a few hundred samples. Thus, regression and subset selection algorithms used to identify dominant genes must be conditioned to manage deficiencies in sample size. The elastic net algorithm introduced by Zou and Hastie [35] that extends LARS for $p > n$ datasets is one such example.

A more common scenario is when the number of observations slightly exceed the number of predictors. The gasoline consumption dataset containing 27 observations and 4 predictors is such a sample deficient dataset. GAM is not capable of generating a regression model for this dataset. However, models generated by MAR and MARm is able to account for 86% and 87% of the variance in response achieving good model fits. Thus, MAR is preferred over GAM when the datasets are sample deficient. Quantitative validation of this dataset can be found in the supplementary material (Tables S4, S5).

D. LINEAR MODEL ESTIMATION: BODYFAT DATASET

This dataset contains 252 observations and 14 predictor covariates of body circumference measurements to estimate bodyfat percentage of individuals obtained from the CMU repository. Many research articles from the health and biology community advocates the need for an accurate model to measure bodyfat percentage in individuals due to a strong correlation between excess fatty tissue and chronic diseases [54], [55]. A suitable formulation is given by Siri's equation (12) which predicts the amount of bodyfat using the body density feature (x_1).

$$BF = \frac{495}{x_1} - 450 \tag{12}$$

The main drawback of (12) however is that density measurements require underwater weighing equipment which

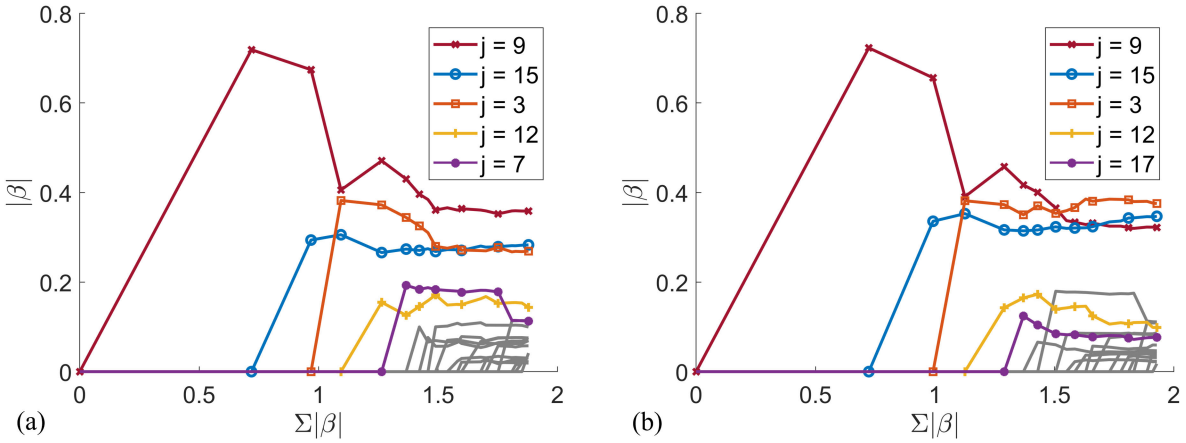

FIGURE 9. Solution path propagation for the 4 most dominant predictor covariates for the (a) MARm and (b) MAR algorithms.

are not readily accessible for rapid measurement. Instead, Kanellakis *et al.* [55] argues that anthropometry-based measurements are more suitable for real world applications. A comprehensive list of 31 anthropometric equations is presented in [55], from which most assume a linear relationship between predictors.

Initially, we fit regression models for the full dataset including x_1. The performance is depicted in Table 7. All regression algorithms correctly identify x_1 as a dominant predictor. Both LARS and MAR type algorithms select x_1 as the start of the solution path. LARS and GLM achieves better nRMSE compared to Siri as they include additional features in the model. Although nRMSE of the GAM and MAR type algorithms closely follow LARS, the nonlinearity encourage overfitting.

A second set of regression models are generated for the dataset without x_1. They are validated against 2 pre-existing models from [55] which are compatible with the predictors on our dataset. However, these existing models show very poor nRMSE performance. This may be caused by the inter-observer variation in measurements between the training and validation datasets. Thus, the results underscore the importance of setting adequate guidelines when acquiring medical data. Except for MARm, all other algorithms achieve similar nRMSE performance. However, compared to other algorithms, LARS achieves good accuracy for a much lower cost on model complexity (N). A lower J for GAM and MAR also indicates overfitting due to the spline transformations. During subset selection, both MAR and LAR type algorithms rank the same predictor: abdomen circumference (x_7) as the primary predictor covariate. Thus, the results indicate that MAR can achieve subset selection and model fitting accuracies comparable to LARS for linear datasets.

E. EXPLORING HIDDEN NON-LINEARITY: KING COUNTY DATASET

The King County dataset contains 21613 observations and 18 predictor covariates for modelling real-estate sales prices in King County, Washington between May 2014, and May 2015. This dataset contains a mix of continuous and categorical features akin to most real-world regression problems. The 18 attributes associated with this dataset are listed in Table 8.

TABLE 8. Description of predictor covariates in King County dataset.

Variable	Description	Datatype
x_1	Number of bedrooms	Continuous
x_2	Number of bathrooms	Continuous
x_3	Living area (sqft)	Continuous
x_4	Lot area (sqft)	Continuous
x_5	Floors	Continuous
x_6	Waterfront (0/1)	Categorical
x_7	View rating (0-4)	Continuous
x_8	Condition (1-5)	Continuous
x_9	Rating by agent (1-13)	Continuous
x_{10}	Above area (sqft)	Continuous
x_{11}	Basement area (sqft)	Continuous
x_{12}	Year built	Continuous
x_{13}	Year renovated	Continuous
x_{14}	Zip code	Continuous
x_{15}	Latitude	Continuous
x_{16}	Longitude	Continuous
x_{17}	Mean living area of 15 neighbors (sqft)	Continuous
x_{18}	Mean lot area of 15 neighbors (sqft)	Continuous

The regression algorithms are first applied to the original King County dataset. Numerical results are depicted in Table 9. LARS and GLM algorithms achieve the poorest nRMSE performance indicating a nonlinear relationship between the predictor and the response for this dataset. GAM achieves the best nRMSE. However, the GAM algorithm does not compromise on model complexity N leading to no meaningful subset selection. MAR type algorithms closely follow the accuracy of the GAM models albeit with much better parsimony. Thus, MAR is better suited for subset selection. The solution paths for the MAR type algorithms are illustrated in Fig. 9. Both the order and the identity of the covariates are common for both algorithms. Together, the 4 top ranking covariates capture 76.96% and 75.55% of the observed variance for the MAR and MARm algorithms, respectively.

TABLE 9. Algorithm performance on the King County dataset.

Algorithm	RMSE	nRMSE	N	J
King county				
LARS	**201452** (*6920*)	**37.29** (*1.28*)	**16.2** (*0.79*)	**0.95**
MARm	**180045** (*9307*)	**33.33** (*1.72*)	**16.2** (*1.55*)	**0.92**
MAR	**176026** (*21467*)	**32.59** (*3.97*)	**14.7** (*5.36*)	**0.70**
GAM	**165074** (*11052*)	**30.56** (*2.05*)	**18** (*0*)	**1**
GLM	**201430** (*10240*)	**37.29** (*1.90*)	**16.2** (*0.63*)	**0.96**
King county ($x_{15} \otimes x_{16}$)				
MARm	**154086** (*11081*)	**28.52** (*2.05*)	**13.4** (*2.55*)	**0.81**
MAR	**153248** (*18712*)	**28.37** (*3.46*)	**14.3** (*4.72*)	**0.74**
GAM	**163518** (*9585*)	**30.27** (*1.77*)	**17** (*0*)	**1**

Evaluation parameters for accuracy (RMSE), normalized RMSE (nRMSE), number of non-zero covariates (N), and the Jaccard coefficient (J) are considered.

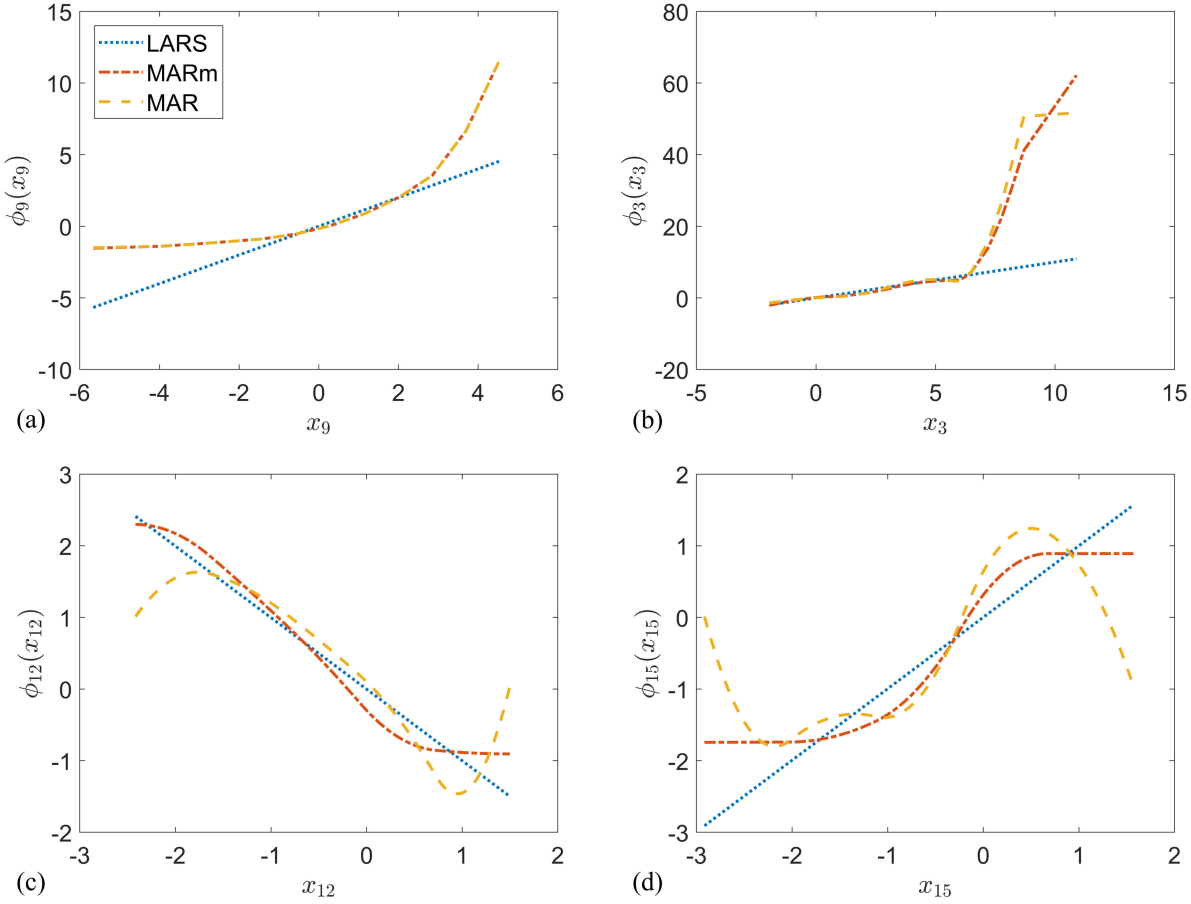

FIGURE 10. Nonlinear transformation of the 4 dominant predictor covariates (a) agent rating, (b) living area, (c) year built, and (d) latitude by the LARS, MAR, and MARm algorithms. MARm and MAR show a similar nonlinear mapping for all 4 dominant predictor covariates indicating a highly significant predictor-response relationship.

Nonlinear transformations of the dominant covariates are illustrated in Fig. 10 with LARS mapping indicated as a reference. In Fig. 10 (a), the MAR and MARm transformations perfectly overlap for the x_9 covariate. The nonlinear transformation $\phi_9(x_9)$ is almost a straight-line on the semi-log plot. This provides evidence that the house price increases exponentially with respect to the rating. Thus, having a good rating for the house can greatly increase its value. In Fig. 10 (b), feature $\phi_3(x_3)$ show that houses with larger living areas are more expensive. Unlike linear regression models however, MAR shows that the price saturates at extremely large x_3. In Fig. 10 (c), all algorithms indicate that older houses built in the early 1900s are more expensive, likely caused by its high sentimental value. However, houses built post 2000 with modern amenities will also likely postulate high selling prices. MARm is unable to adequately capture this relationship due to the monotonicity constraint. Thus, MAR can be used to better represent this relationship.

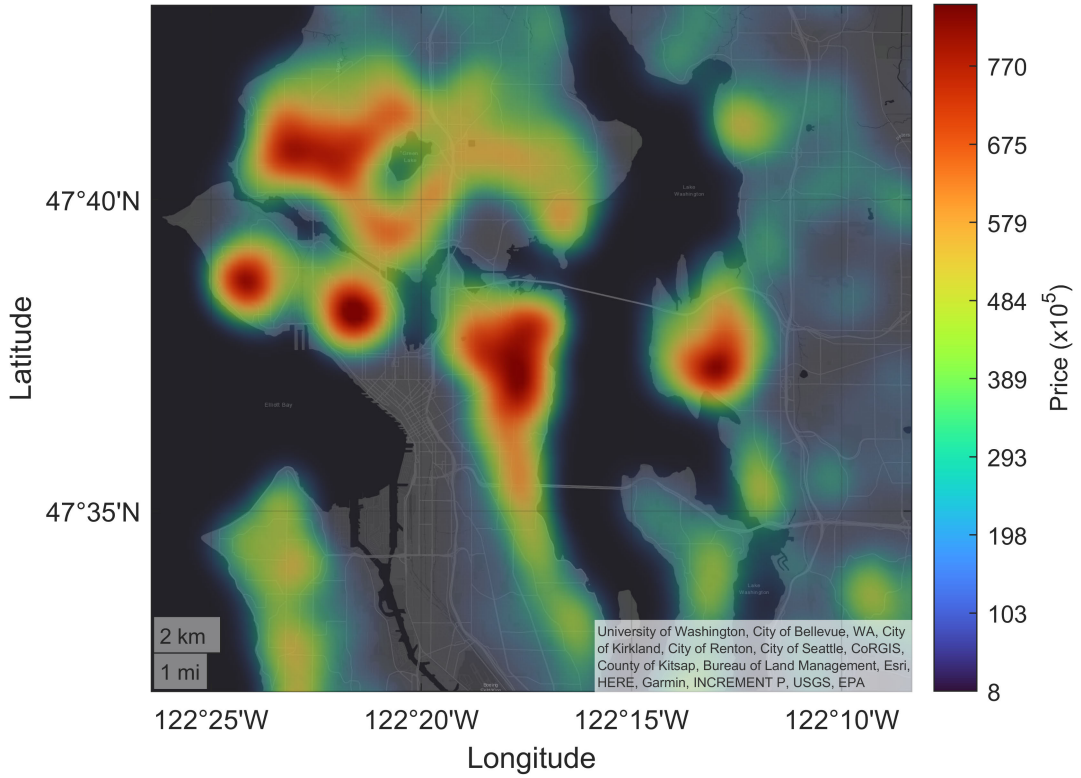

FIGURE 11. A heatmap of house prices superimposed on the Seattle city map.

The latitude feature (x_{15}) in Fig. 10 (d) initially included in the dataset to support visualization turns out to hold the second most significant predictive power, greater than other commonsense predictors such as the living room area and the year built. This surprising outcome highlights one very important advantage of algorithmic regression over procedures that heavily rely on human judgement: sometimes algorithms will discover patterns hidden in the data that even experts miss. Superimposing the spline function of latitude on the Seattle city map lead us to suspect that x_{15} has acted as the proxy for geographic location in the regression model. Geographic location holds a high predictive power as it encapsulates many hidden factors that are not explicitly available in the dataset.

The meaning of x_{15} becomes clear when it is projected onto the King County map. Seattle is the center of the cosmopolitan area where most of the transactions take place. In Fig. 11, the metropolitan area developed predominantly along a north-south axis due to the unique geographical constraints of the area. The initial asymptote of the constrained spline transformation in Fig. 10 (d) corresponds to the small city of Kent 20 miles south of Seattle. The price steadily increases towards the north (higher altitude) and saturates near downtown Seattle. Once again, the prices reduce further north as x_{15} moves away from the local population center. The unconstrained spline transformation almost perfectly captures this effect and mirrors the relationship. Hence latitude has acted as a proxy to represent urbanity of the area. This observation agrees with the common belief about the relationship between house price and the proximity to urban centers.

We further investigate the impact of geographical location on house price by merging the latitude and longitude features by a tensor product [4]. Performance of the regression models on the modified dataset are illustrated in Table. 9. The new model shows that real-estate in the areas corresponding to urban city center and the affluent northern suburbs of the city holds the highest value. The first 4 dominant covariates in the MAR and MARm effectively capture 82.68% and 81.60% of the observable covariance indicating a significant increase in predictor performance. Both MAR type algorithms achieves the best nRMSE performance.

VI. CONCLUSION

This paper introduces an association-based regression and subset selection algorithm. MAR demonstrates robustness to unknown transformation of covariates which is often encountered in practice. The nonlinear transformation of covariates offer a data-driven insight into the relationship between the predictors and the response. Compared to ad-hoc pre-processing such as logarithmic transformations, these data-driven relationships are easier to use and is more robust. This aspect of MAR makes it a valuable tool for data exploration as priori knowledge of the relationships between covariates are replaced by high level constraints such as monotonicity and smoothness. MAR behaves like LARS

when the underlying problem can be solved satisfactorily though a linear model. Compared to GAM, MAR performs well even when the dataset is sample deficient compared to the number of covariates. Analysis on real datasets show that MAR is very flexible and capable of producing parsimonious regression models by selecting the appropriate predictors.

A. RELATIONSHIP WITH OTHER REGRESSION ALGORITHMS

The use of association in subset selection is rarely used to develop regression algorithms. MAR differs from LARS in the sense of subset selection criterion. Therefore, MAR can be viewed as a version of LARS that is invariant with respect to nonlinear transformation of predictor covariates. Technically, MAR is a subset of Generalized Additive Models (GAM) introduced by Hastie and Tibshirani [2]. However, it should be noted that MAR is not designed to tackle the problems that are generally associated with GAM. It has been designed to address the problem of unreliable subset selection when one applies linear regression to nonlinear problems. Therefore, it is more useful to view MAR as a linear regression algorithm that is capable of subset selection and for identification of suitable transformations to linearize the problem. The boosting methodology proposed by Tutz and Leitenstorfer [56] for subset selection is closely related to the stage-wise selection strategy of MAR but differs substantially in the way variables are selected.

Finally, MAR is capable of generating lasso-like solution paths after imposing sign restrictions as described by Efron *et al.* [28]. However, there is no known optimization formulation that leads to LARS hence it is difficult to ascertain the meaning of this lasso-like modification.

B. COMPUTATIONAL MATTERS

The computation of association is much slower than the computation of correlation. Linear correlation has a computational complexity of $\mathcal{O}(n)$ while the association measures have a complexity of $\mathcal{O}(n \log(n))$. Nevertheless, an $\mathcal{O}(n \log(n))$ algorithm is generally considered fast and is widely used in diverse real-world applications. Like LARS, finding the association between the inactive set and the current residual can be broken down into many independent computational tasks. This makes MAR compatible with parallel computing technologies making it easily scalable to tackle large problems.

The computation of spline approximations are small optimization problems. It is not sensitive to the sample size because the number of knots has been chosen *a priori*. More refined approximation of the nonlinear transformation can be achieved through tuning of spline parameters. However, this will come at an expense of higher computational costs. The fine-tuning of spline functions can be done during a follow-up study after suitable subsets have been selected.

ACKNOWLEDGMENT

The authors would like to acknowledge the work done at the Intelligent Lighting Laboratory, Monash University Malaysia.

REFERENCES

[1] L. Breiman and J. H. Friedman, "Estimating optimal transformations for multiple regression and correlation," *J. Amer. Statist. Assoc.*, vol. 80, no. 391, pp. 580–598, 1985.
[2] T. Hastie and R. Tibshirani, "Generalized additive models," *Stat. Sci.*, vol. 1, no. 3, pp. 297–310, 1986.
[3] S. N. Wood, Y. Goude, and S. Shaw, "Generalized additive models for large data sets," *J. Roy. Stat. Soc., C, Appl. Statist.*, vol. 64, no. 1, pp. 139–155, 2015.
[4] S. N. Wood, *Generalized Additive Models: An Introduction With R*. Portland, OR, USA: CRC Press, 2017.
[5] S. N. Wood, "Thin plate regression splines," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 65, no. 1, pp. 95–114, 2003.
[6] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[7] Y. Lou, R. Caruana, and J. Gehrke, "Intelligible models for classification and regression," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2012, pp. 150–158.
[8] X. He and P. Shi, "Monotone B-spline smoothing," *J. Amer. Stat. Assoc.*, vol. 93, no. 442, pp. 643–650, 1998.
[9] M. Schmid and T. Hothorn, "Boosting additive models using component-wise P-splines," *Comput. Statist. Data Anal.*, vol. 53, no. 2, pp. 298–311, Dec. 2008.
[10] M. C. Meyer, "Constrained penalized splines," *Can. J. Statist.*, vol. 40, no. 1, pp. 190–206, Mar. 2012.
[11] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Stat. Sci.*, vol. 11, no. 2, pp. 89–121, May 1996.
[12] M. C. Meyer, "Inference using shape-restricted regression splines," *Ann. Appl. Statist.*, vol. 2, no. 3, pp. 1013–1033, Sep. 2008.
[13] J. O. Ramsay, "Monotone regression splines in action," *Stat. Sci.*, vol. 3, no. 4, pp. 425–441, Nov. 1988.
[14] L. C. Bergersen, K. Tharmaratnam, and I. K. Glad, "Monotone splines lasso," *Comput. Statist. Data Anal.*, vol. 77, pp. 336–351, Sep. 2014.
[15] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Statist.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.
[16] S. N. Wood, "Fast stable direct fitting and smoothness selection for generalized additive models," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 70, no. 3, pp. 495–518, Jul. 2008.
[17] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 73, no. 1, pp. 3–36, 2011.
[18] S. N. Wood, N. Pya, and B. Säfken, "Smoothing parameter and model selection for general smooth models," *J. Amer. Stat. Assoc.*, vol. 111, no. 516, pp. 1548–1563, Oct. 2016.
[19] A. Chouldechova and T. Hastie, "Generalized additive model selection," 2015, *arXiv:1506.03850*.
[20] Y. Lou, J. Bien, R. Caruana, and J. Gehrke, "Sparse partially linear additive models," *J. Comput. Graph. Statist.*, vol. 25, no. 4, pp. 1126–1140, 2016.
[21] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
[22] Y. Lin and H. H. Zhang, "Component selection and smoothing in multivariate nonparametric regression," *Ann. Statist.*, vol. 34, no. 5, pp. 2272–2297, Oct. 2006.
[23] P. Ravikumar, J. Lafferty, H. Liu, and L. Wasserman, "Sparse additive models," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 71, no. 5, pp. 1009–1030, 2009.
[24] L. Meier, S. van de Geer, and P. Bühlmann, "High-dimensional additive modeling," *Ann. Statist.*, vol. 37, no. 6B, pp. 3779–3821, 2009.
[25] G. Marra and S. N. Wood, "Practical variable selection for generalized additive models," *Comput. Statist. Data Anal.*, vol. 55, no. 7, pp. 2372–2387, Jul. 2011.
[26] R. R. Hocking and R. N. Leslie, "Selection of the best subset in regression analysis," *Technometrics*, vol. 9, no. 4, pp. 531–540, Nov. 1967.
[27] N. R. Draper and H. Smith, *Applied Regression Analysis*, vol. 326. New York, NY, USA: Wiley, 1998.
[28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.
[29] Y. Wu, "An ordinary differential equation-based solution path algorithm," *J. Nonparametric Statist.*, vol. 23, no. 1, pp. 185–199, Mar. 2011.

[30] W. Xiao, Y. Wu, and H. Zhou, "ConvexLAR: An extension of least angle regression," *J. Comput. Graph. Statist.*, vol. 24, no. 3, pp. 603–626, Jul. 2015.
[31] J. A. Khan, S. Van Aelst, and R. H. Zamar, "Robust linear model selection based on least angle regression," *J. Amer. Stat. Assoc.*, vol. 102, no. 480, pp. 1289–1299, Dec. 2007.
[32] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 68, no. 1, pp. 49–67, 2006.
[33] A. Alfons, C. Croux, and S. Gelper, "Robust groupwise least angle regression," *Comput. Statist. Data Anal.*, vol. 93, pp. 421–435, Jan. 2016.
[34] H. Zou, "The adaptive lasso and its Oracle properties," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.
[35] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.
[36] E. Mammen and K. Yu, "Additive isotone regression," in *Asymptotics: Particles, Processes and Inverse Problems*. Beachwood, OH, USA: Institute of Mathematical Statistics, 2007, pp. 179–195.
[37] M. C. Meyer, "Semi-parametric additive constrained regression," *J. Nonparametric Statist.*, vol. 25, no. 3, pp. 715–730, Sep. 2013.
[38] X. Huo and G. J. Székely, "Fast computing for distance covariance," *Technometrics*, vol. 58, no. 4, pp. 435–447, Oct. 2016.
[39] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, "Measuring and testing dependence by correlation of distances," *Ann. Statist.*, vol. 35, no. 6, pp. 2769–2794, 2007.
[40] G. J. Székely and M. L. Rizzo, "Partial distance correlation with methods for dissimilarities," *Ann. Statist.*, vol. 42, no. 6, pp. 2382–2412, Dec. 2014.
[41] R. Luss and S. Rosset, "Bounded isotonic regression," *Electron. J. Statist.*, vol. 11, no. 2, pp. 4488–4514, Jan. 2017.
[42] Q. Han, T. Wang, S. Chatterjee, and R. J. Samworth, "Isotonic regression in general dimensions," *Ann. Statist.*, vol. 47, no. 5, pp. 2440–2471, Oct. 2019.
[43] (1996). *Delve Dataset Repository*. Accessed: Oct. 1, 2021. [Online]. Available: https://www.cs.toronto.edu/~delve/data/datasets.html
[44] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. Accessed: Oct. 1, 2021. [Online]. Available: http://archive.ics.uci.edu/ml
[45] P. Vlachos. (2005). *Statlib Datasets Archive*. Accessed: Oct. 1, 2021. [Online]. Available: http://lib.stat.cmu.edu/datasets/
[46] (2016). *King County House Price*. Accessed: Oct. 1, 2021. [Online]. Available: https://www.kaggle.com/harlfoxem/housesalesprediction
[47] D. Kilpatrick and R. Cameron-Jones, "Numeric prediction using instance-based learning with encoding length selection," in *Proc. Int. Conf. Neural Inf. Process. Intell. Inf. Syst.*, vol. 1, 1998, pp. 984–987.
[48] R. D. King, S. Muggleton, R. A. Lewis, and M. Sternberg, "Drug design by machine learning: The use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase," *Proc. Nat. Acad. Sci. USA*, vol. 89, no. 23, pp. 11322–11326, 1992.
[49] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, Nov. 2004.
[50] S. Y. Ho, L. Wong, and W. W. B. Goh, "Avoid oversimplifications in machine learning: Going beyond the class-prediction accuracy," *Patterns*, vol. 1, no. 2, May 2020, Art. no. 100025.
[51] B. Bechtold. (2016). *Violin Plots for MATLAB*. Accessed: Oct. 1, 2021. [Online]. Available: https://github.com/bastibe/Violinplot-Matlab
[52] J. D. Hirst, R. D. King, and M. J. E. Sternberg, "Quantitative structure-activity relationships by neural networks and inductive logic programming. I. the inhibition of dihydrofolate reductase by pyrimidines," *J. Comput.-Aided Mol. Des.*, vol. 8, no. 4, pp. 405–420, Aug. 1994.
[53] J. B. Mitchell, "Machine learning methods in chemoinformatics," *Wiley Interdiscipl. Rev., Comput. Mol. Sci.*, vol. 4, no. 5, pp. 468–481, 2014.
[54] A. J. Chambers, E. Parise, J. L. Mccrory, and R. Cham, "A comparison of prediction equations for the estimation of body fat percentage in non-obese and obese older Caucasian adults in the United States," *J. Nutrition, Health Aging*, vol. 18, no. 6, pp. 586–590, Jun. 2014.
[55] S. Kanellakis, E. Skoufas, V. Khudokonenko, E. Apostolidou, L. Gerakiti, M.-C. Andrioti, E. Bountouvi, and Y. Manios, "Development and validation of two equations based on anthropometry, estimating body fat for the Greek adult population," *Obesity*, vol. 25, no. 2, pp. 408–416, Feb. 2017.
[56] G. Tutz and F. Leitenstorfer, "Generalized smooth monotonic regression in additive modeling," *J. Comput. Graph. Statist.*, vol. 16, no. 1, pp. 165–188, Mar. 2007.

SANUSH K. ABEYSEKERA (Member, IEEE) received the B.Eng. degree (Hons.) in mechatronics engineering and the Ph.D. degree in electrical and computer systems engineering from Monash University, Malaysia.

He currently works as a Postdoctoral Fellow at the Faculty of Science and Engineering, The University of Waikato, New Zealand. He is affiliated with the Waikato Robotics, Automation and Sensing Group. His research interests include statistical machine learning, image processing, computer vision, and robotics.

YE-CHOW KUANG (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electromechanical engineering and the Ph.D. degree in non-invasive diagnostic techniques from the University of Southampton.

He worked in the areas of machine intelligence, uncertainty modeling in engineering design, and automation. He is currently affiliated with the Waikato Robotics, Automation and Sensing Group as well as the Artificial Intelligence Institute, The University of Waikato. He is leading the IEEE Instrumentation and Measurement Society Fault Tolerant Measurement Systems Technical Committee. He was a recipient of the 2019 IEEE Instrumentation and Measurement Society Best Technical Committee Award and the International Education Association of Australia Award for Best Innovation in International Education.

MELANIE PO-LEEN OOI (Senior Member, IEEE) is currently an Associate Professor and Assistant Dean (Research) at The University of Waikato. She has developed new testing techniques and test data processing methodologies that have been adopted by multinational companies, such as Texas Instruments, Freescale Semiconductor (now NXP Semiconductors), and Western Digital. Her work in measurement uncertainty propagation has been adopted by the South African National Accreditation System (national body responsible for conformity assessment) in their guidelines document TG 50-02, since October 2017. She is an Administrative Committee Member of the IEEE Instrumentation & Measurement Society (I&MS) and a Secretary and a member of the Technical Committee on Fault Tolerant Measurement Systems (TC-32) of the IEEE I&MS. She is the youngest female fellow appointed to the Institution of Engineering and Technology. She has been acknowledged by several awards, including the 2019 Rutherford Discovery Fellowship from New Zealand's Royal Society Te Apārangi, the 2017 Mike Sargeant Career Achievement Award from the Institution of Engineering and Technology, U.K., and the Outstanding Young Engineer Award of the Year 2014 from the IEEE Instrumentation and Measurement Society. She is a Guest Editor for the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. She is a U.K. Chartered Engineer.

VINEETHA KALAVALLY received the Ph.D. degree from Monash University, Australia, in 2012. She is currently an Associate Professor with the Department of Electrical and Computer Systems Engineering, School of Engineering, Monash University Malaysia, where she leads the Intelligent Lighting Laboratory. Her research interests include diverse applications of solid-state lighting, visual and non-visual quality of light, wearable light sensors, and visible light communication.