A Sequential Simplex Algorithm for Automatic Data and Center Selecting Radial Basis Functions

Xiaofeng Ma
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523–1874
Email: markma@rams.colostate.edu

Tomojit Ghosh
Department of Computer Science
Colorado State University
Fort Collins, Colorado 80523
Email: tomojit@rams.colostate.edu

Michael Kirby
Department of Mathematics
Department of Computer Science
Colorado State University
Fort Collins, Colorado 80523–1874
Email: Michael.Kirby@Colostate.edu

***Abstract*—We propose a sequential algorithm for learning sparse radial basis approximations for streaming data. The initial phase of the algorithm formulates the RBF training as a convex optimization problem with an ℓ_1 objective function on the expansion weights while the data fitting problem imposed only as an ℓ_∞-norm constraint. Each new data point observed is tested for feasibility, i.e., whether the data fitting constraint is satisfied. If so, that point is discarded and no model update is required. If it is infeasible, a new basic variable is added to the linear program. The result is a primal infeasible-dual feasible solution. The dual simplex algorithm is applied to determine a new optimal solution. A large fraction of the streaming data points does not require updates to the RBF model since they are similar enough to previously observed data and satisfy the data fitting constraints. The structure of the simplex algorithm makes the update to the solution particularly efficient given the inverse of the new basis matrix is easily computed from the old inverse. The second phase of the algorithm involves a non-convex refinement of the convex problem. Given the sparse nature of the LP solution, the computational expense of the non-convex algorithm is greatly reduced. We have also found that a small subset of the training data that includes the novel data identified by the algorithm can be used to train the non-convex optimization problem with substantial computation savings and comparable errors on the test data. We illustrate the method on the Mackey-Glass chaotic time-series, the monthly sunspot data, and a Fort Collins, Colorado weather data set. In each case we compare the results to artificial neural networks (ANN) and standard skew-RBFs.**

I. Introduction

This paper addresses the problems of *data selection* and *model order determination* that arise when learning nonlinear mappings from observations. The data selection problem has two primary components. Ideally, the algorithm should identify the point at which new observations are not changing the model, i.e., the number of fitting functions and their approximate locations have been determined. Once enough data has been observed, there is no advantage to including additional points in the training phase. Further, only a subset of the total training data may be required to capture the behavior required to construct the model. There is no need to train on new data if it is sufficiently similar to previously processed data. This is particularly important for efficiency in real-time data streaming, or on-line learning, where only novel data being observed should be incorporated into model updates. The approach proposed here addresses both of these data selection issues and in so doing, the model order of the problem is also determined.

In recent work, we proposed a two-phase *batch* algorithm for modeling nonlinear relationships in data using radial basis functions [26], [27]. The first phase required the solution of a linear program for center selection. The Dantzig-Selector, and p-norm convex optimization problems were explored for the center selection problem in the batch, or non-streaming, modeling problem. Given the full training data set at the start of modeling, these sparsity promoting optimization problems provide an estimate of the required complexity for the model, as well as initial spatial locations for the approximating functions. The second training phase involved the nonconvex optimization of the shape and location parameters of skew RBFs proposed in [5], [6]. While this can be an expensive aspect of the training, the fact that a parsimonious model is produced by the first phase greatly reduces the amount of effort required for model refinement. The batch algorithm produced results on standard bench-marking problems comparable to state-of-the-art. However, this batch algorithm doesn't exploit the efficiencies of the sequential simplex algorithm, nor can it identify novel data subsets.

In this paper we demonstrate how the linear nature of the RBF learning problem, first advocated in [19], can be exploited to formulate an adaptive learning algorithm using a sequential simplex algorithm. Streaming data points are identified as feasible and ignored, or infeasible, and used to update the solution to the convex optimization problem. The result is a highly parsimonious model that is then refined by adding skewing terms and non-convex refinement. The algorithm provides a geometric criterion, polyhedron feasibility, to identify *novel* data points which are the only ones used to update the model. Interestingly we see that in the context of time-series, the novel data points tend to occur at extrema, or inflection points, in the target data. As we shall see, these novel data points may also be used to construct minimal training sets for the non-convex refinement phase of the algorithm. We demonstrate this numerically on three data sets including that Mackey-Glass chaotic time-series, Fort Collins weather data, and monthly sunspot data. Our results are shown to be comparable to those of Artificial Neural Networks where substantially more user guidance is required to determine network architecture.

II. Problem Formulation

The Radial Basis Function (RBF) expansion is a widely used tool for data driven modeling of large data sets and has the form

$$f(x) = w_0 + \sum_{k=1}^{m} w_k \phi(\|x - c_k\|). \tag{1}$$

where the RBF functions can be selected from several options, see, e.g., [18] for details. We assume in this paper that the function f and weights w_k are scalars. As we see below, the location, and number of centers $\{c_k\}$ will be determined automatically by the algorithm. In our implementation, the Euclidean inner product is weighted with the metric being determined by the data; see [6] for details.

We define the interpolation matrix as

$$\Phi = \begin{bmatrix} 1 & \phi(\|x_1 - c_1\|) & \dots & \phi(\|x_1 - c_{N_c}\|) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \phi(\|x_m - c_1\|) & \dots & \phi(\|x_m - c_{N_c}\|) \end{bmatrix}$$

and our notation $(\Phi)_i$ is the i'th row of this matrix.

The *Dantzig-Selector* optimization problem was proposed in the context of solving $b = Xw + z$ where X is $n \times p$ with $n << p$ [16], [28]. The Dantzig-selector convex optimization problem is

$$\text{minimize } \|w\|_1 \tag{2}$$

subject to

$$\|X^T(Xw - b)\|_\infty \leq \epsilon \tag{3}$$

This approach was extended to the RBF data fitting problem where $X = \Phi$ as defined above [26], [27]. In this paper, we employ the modified Dantzig-selector where the convex optimization problem has the form

$$\|\Phi w - b\|_\infty \leq \epsilon_c \tag{4}$$

where ϵ_c is estimated as in [27]. The constraint in Equation (4) is preferred to the formulation in Equation (2) since the number of candidate centers may be large and it avoids forming a $p \times p$ matrix for large p. Also, more importantly, Equation (2) does not allow for a sequential application of the linear programming algorithm.

III. Algorithm

The modified (and unmodified) Dantzig-selector problem may be written as a Linear Program (LP) in standard form

$$\text{minimize } c^T x \tag{5}$$

subject to the constraints

$$Ax = b,\ x \geq 0 \tag{6}$$

where $c \in \mathbb{R}^n$ is the cost vector, $A \in \mathbb{R}^{m \times n}$, resources $b \in \mathbb{R}^m$.

A. Formulation of the LP

The conversion of the ℓ_1-norm problem with ℓ_∞ constraints into a linear program in standard form may be accomplished by introducing new variables. First, we transform the objective function introducing t_i such that $-t_i \leq w_i \leq t_i$. Similarly, the ℓ_∞-norm in the constraint $\|\Phi w - b\|_\infty$ can be rewritten $-\epsilon_c \leq \Phi_i^T w - b \leq \epsilon_c$.

At the k'th step the sequential algorithm solves the problem

$$\text{minimize } \sum_{i=1}^{N_c} t_i \tag{7}$$

subject to

$$\Phi_k^T w - b + w_{N_c+1} = \epsilon_c \tag{8}$$

$$-\Phi_k^T w + b + w_{N_c+2} = \epsilon_c \tag{9}$$

$$w_i \leq t_i = 0, i = 1, \dots, N_c \tag{10}$$

$$-w_i \leq t_i = 0, i = 1, \dots, N_c \tag{11}$$

where $w_{N_c+1} \geq 0$, $w_{N_c+2} \geq 0$ are introduced as new basic variables converting the inequality constraints to equality constraints.

B. Feasibility of New LP

At each iteration we add a new data point and the associated LP may be augmented by two basic variables. Given an optimal solution was determined in the previous iteration, the new LP must be dual feasible since the reduced cost of the added basic variables are zero [3]. However, the primal problem may not be feasible since either $w_{N_c+1} < 0$, or $w_{N_c+2} < 0$, but not both. In other words, at least one of the two constraints being added with each point *must* be feasible, and they may both be feasible. This follows since if

$$\Phi_i^T w - b > 0$$

i.e., the first added constraint is violated, then

$$-\Phi_i^T w + b < 0$$

so the second constraint *must be* satisfied. Note that if

$$-\epsilon_c \leq \Phi_i^T w - b \leq \epsilon_c$$

then both added constraints are feasible.

In our algorithm we do not update the LP with constraints that are feasible. We will observe in the numerical experiments that this approach significantly accelerates the learning problem since only a small fraction of the streaming data points lead to infeasibility.

C. Initialization of LP

We start by solving the following initialization problem so that an optimal basic feasible solution can be obtained. The first convex minimization problem to be solved is Equation (12), a linear programming without any data points added.

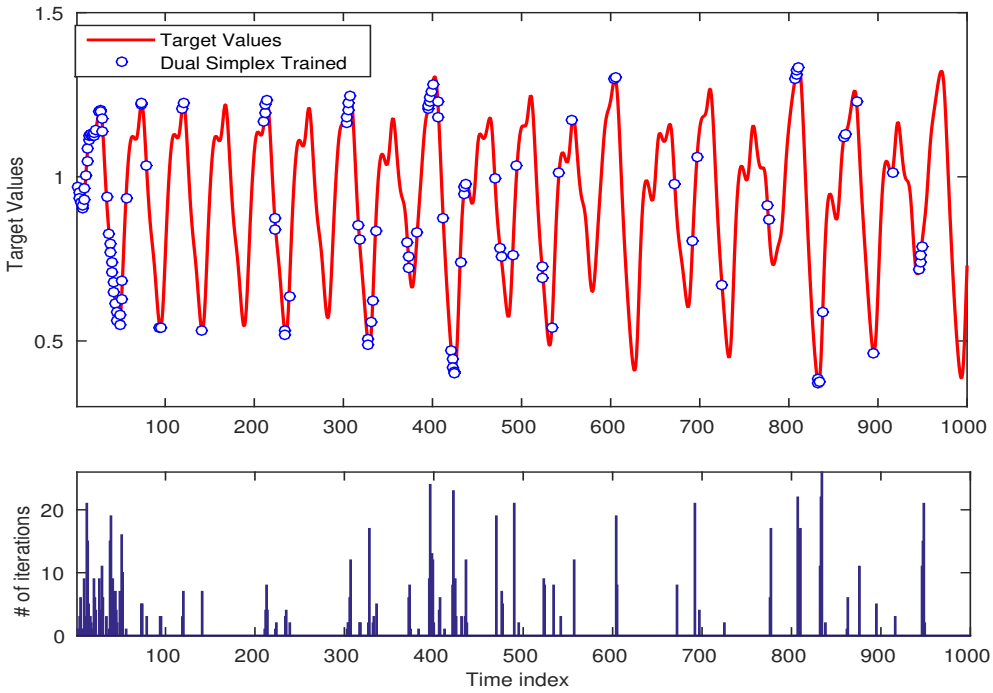

Fig. 1: The result of applying the simplex algorithm to the Mackey-Glass data set using sequential training over 1000 data points. The lower bars denote that an infeasible LP resulted from adding the point. The location of the *novel* point is shown in blue. A total of $N_c = 14$ RBFs were automatically selected from an initial set of 500 during the streaming algorithm.

The optimal solution to this problem is just $w_i = t_i = 0$, $i = 1, \ldots, N_c$.

$$\text{minimize} \sum_{i=1}^{N_c} t_i \tag{12}$$

subject to

$$w_i \leq t_i, i = 1, \ldots, N_c \tag{13}$$

$$-w_i \leq t_i, i = 1, \ldots, N_c \tag{14}$$

With this initialization the LP is now updated by streaming new data points as described in the next section.

D. Infeasibility and Update

In the event that the resulting LP is infeasible there will only be a single variable that will determine the pivoting row in the dual simplex algorithm. The dual simplex algorithm is already feasible so it can be iterated until the primal is feasible while maintaining dual feasibility. We will see that in our application, only a small number of iterations is required in practice to achieve an optimal solution after each new data point is added.

Here we briefly describe the components of the simplex algorithm that make it especially suitable for the current application. A full discussion of the method is outside the scope of this paper and the reader is referred to [3]. The simplex algorithm proceeds by maintaining and updating a basis matrix B consisting of linearly independent columns of the constraint matrix. The solution is improved by identifying a nonbasic column of A which may be used to improve the solution. In this process we require the matrix B^{-1} in order to compute the reduced costs of each non-basic column of A. When the basis B is updated there is a very efficient approach to determine an update for B^{-1} known as the revised simplex method.

In our problem at the m'th update we add a constraint

$$\Phi_m^T w + w_{n+m} = \epsilon$$

The new basis matrix has size $(m+1) \times (m+1)$ and is of the form

$$B_{m+1} = \begin{bmatrix} B_m & 0_m \\ \phi_m^T & 1 \end{bmatrix}$$

where B_m is the basis matrix with m constraints and 0_m is the column vector of m zeros. The vector ϕ_m consists of the components of the vector Φ_i associated with basic columns. The reason this sequential approach is so attractive is the easily verified expression

$$B_{m+1}^{-1} = \begin{bmatrix} B_m^{-1} & 0_m \\ \phi_m^T B_m^{-1} & 1 \end{bmatrix}$$

Hence the update to the LP is extremely efficient.

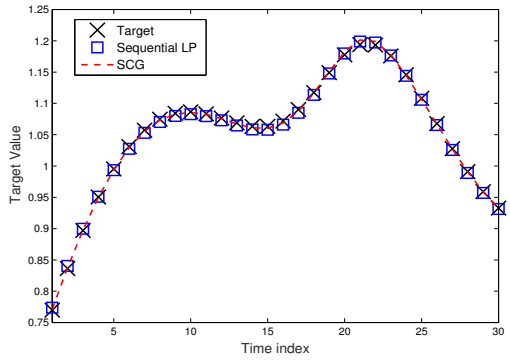

Fig. 2: A comparison of the training phases for the Mackey-Glass problem with the adaptive simplex algorithm. The graph is a blowup of the a segment of the testing data.

	ANN	RBF	LP	LP lite
train set	1000 pts	1000 pts	1000 pts	200 pts
train RMSE	2.1e-3	3.0e-3	2.8e-3	2.8e-3
val. RMSE	2.4e-3	4.0e-3	3.0e-3	2.9e-3
test RMSE	2.2e-3±5.6e-5	3.0e-3±4e-4	2.7e-3±1e-4	2.7e-3±1e-4
NPE1	9e-3±4e-4	1.3e-2±1.5e-3	1.2e-2±6e-4	1.2e-2±6e-4
NPE2	9e-5±5e-6	2.0e-4±5e-5	1.4e-4±1e-5	1.4e-4±1e-5
mean error	-2e-6	1e-4	8.6e-5	2e-4

TABLE I: Comparison between ANN and Sequential simplex algorithm with non-convex refinement on MackeyGlass data.

E. Linear Optimization Algorithm Summary

For clarity we briefly summarize the main components of the algorithm. We denote the feasible set with m points as P_m. We say that $x_m \in P_m$, i.e., x_m is feasible, if

$$-\epsilon \leq \Phi_i^T w \leq \epsilon,\ i = 1, \ldots, m.$$

- Initialize the LP with the first observation x_1.
- Compute the optimal solution w^* to this LP.
- Add a new data point x_2 and compute B_2^{-1}.
- If $x_2 \in P_2$, then the solution $(w^*, w_{N_c+1}, w_{N_c+2})$ is optimal and a new observation x_3 is made and the process repeated.
- If $x_2 \notin P_2$, then apply the dual simplex algorithm until the problem is primal feasible.
- Add a new observation and repeat.

IV. Non-convex refinement

The linear optimization problem solved via the sequential simplex algorithm provides the number N_c of the RBFs estimated to be required for the data fitting problem, as well as estimated initial locations of the centers $\{c_k\}$.

We use this data to refine the model extending Equation (1) to be

$$f(x) = w_0 + \sum_{k=1}^{m} w_k z(\lambda_k^T(x - c_k))\phi(\|x - c_k\|) \qquad (15)$$

where each Euclidean norm is weighted as described in [6]. The addition of the skewing term $z(\lambda_k^T(x - c_k))$ provides additional shape parameters λ_k to the RBFs making them capable of fitting more complex data with fewer terms. The skewing term $z(r)$ used in this paper is

$$z(r) = \frac{1}{\pi}\arctan(r) + \frac{1}{2}$$

while the RBF function has the usual form

$$\phi(\mu) = \exp(\frac{-\mu^2}{\alpha^2}).$$

The error is now minimized using scaled conjugate gradient (SCG) as described in [26], [27]. See [6] and references therein for an introduction as well as additional details and options for skew functions.

V. Numerical Experiments

We present numerical results comparing Artificial Neural Networks (ANNs), and standard RBFs to the LP (Simplex) RBF on three illustrative data sets. The scaled conjugate gradient method is used in all cases to train the final parameters. LP uses the same training data as ANN and standard RBFs, but with centers selected by the sequential simplex algorithm. LP lite uses only the novel data points augmented with a small number of randomly selected points. The normalized predictive errors NPE1 and NPE2 are defined in [5]. Results are averaged over 25 experiments. In all cases optimal parameters are selected based on performance on a validation set.

A. Mackey-Glass

In our first experiment we apply the sequential simplex algorithm to the Mackey-Glass data set in Figure 1. Blue circles mark the location on the graph where points introduce infeasible constraints, i.e., they are novel when compared to previously streamed data. We see that this happens at peaks and valleys of the function, as well as inflection points. In contrast, data in the stream which is sufficiently similar to previously observed points (determined by the fact that the new constraints are feasible) are discarded and no iterations of the (dual) simplex algorithm are required.

We see in Figure 1 (lower panel) the number of iterations taken by the dual simplex algorithm each time an infeasible data point is added. The median number of iterations, computed over 127 novel points, is 5. In this experiment there were 873 iterations where no update was required, i.e., these data had feasible constraints. Also, note that larger numbers of iterations in the dual simplex algorithm occur in clusters, suggesting that the training data is exploring a new region in state space. In contrast, regions that have been well modeled produce long sequences of where the data is all feasible.

In this example, we observe that the refinement of the parameters using scaled conjugate gradient on the shape parameters and centers only improves the LP solution from MSE 0.0054 to 0.0033; see Figure 2.

The solutions all produce comparable errors as shown in Table I. Fourteen centers selected at random were used in the standard skew-RBF and optimized with the shape parameters using SCG. LP lite was the fastest to train using only an

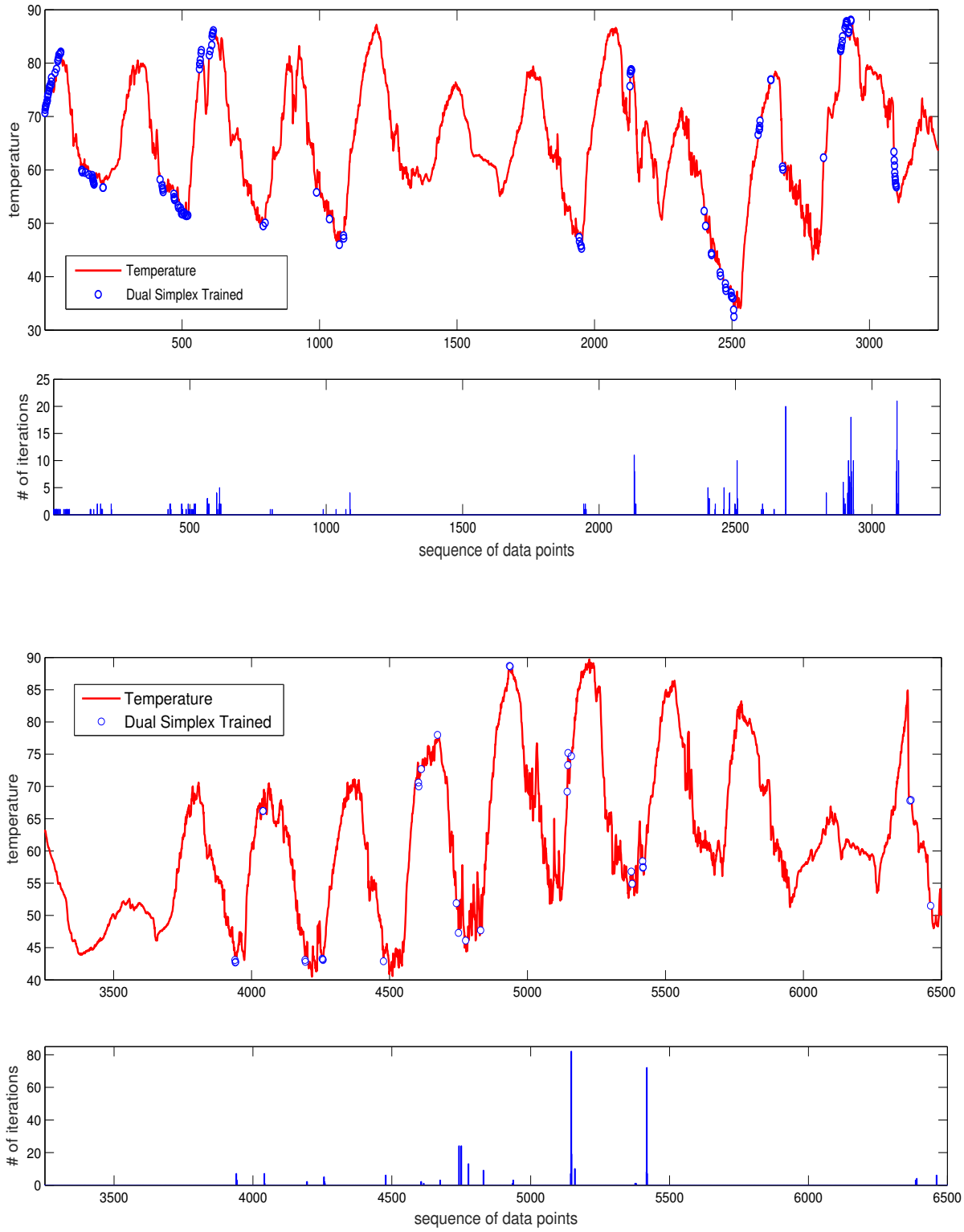

Fig. 3: Fort Collins weather data training results (split into two panels of 3250 data points each). A 3-dimensional embedding with time delay $T = 6$ (30 minutes) is used and we are predicting 12 steps (1 hour) ahead. A total of 6500 input-output pairs are generated for training. A total of 2000 initial centers are sampled from the domain of training data. The RBFs are selected to be Gaussians all with width 10. The sequential simplex algorithm identifies 180 out of 6500 data points as infeasible requiring dual simplex updates to solve for primal feasibility. This plot shows the location of the novel points with number of iterations used in dual simplex algorithm directly in the panel below. A total of $N_c = 13$ centers were selected for this model.

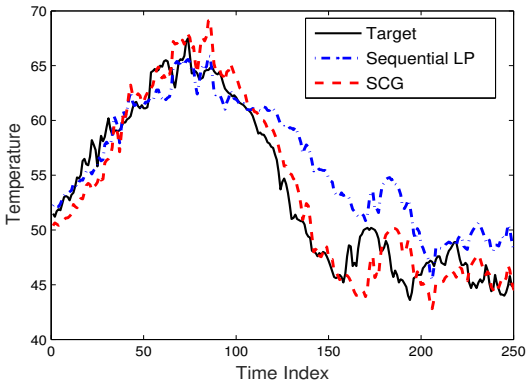

Fig. 4: A comparison of the RBF simplex and non-convex refinement models for the Fort Collins weather data on the testing data. Only 180 data points were used to update the simplex model out of 6500 training data points. An additional 650 were randomly added to these to refine the model using SGC.

	ANN	RBF	LP	LP lite
training set size	6500 pts	6500pts	6500 pts	800 pts
training RMSE	3.19	3.19	3.43	3.51
val. RMSE	2.94	2.94	2.99	2.93
test RMSE	2.97±0.06	2.91±0.11	2.71±0.24	2.75±0.29
NPE1	0.30±0.008	0.3±0.01	0.28±0.03	0.28±0.03
NPE2	0.12±0.005	0.12±0.009	0.10±0.02	0.11±0.03
mean residual	1.14	-0.87	-0.46	-0.29

TABLE II: Comparison between ANN and Sequential simplex algorithm with non-convex refinement on Fort Collins weather data.

average of 131 ± 4 novel data points augmented to a total of 200 points.

B. Streaming Weather Data

The data for this experiment were collected at the Christman Field Weather Station at Colorado State University [30]. The data set consists of temperature measurements collected every five minutes over the month of September, 2016.

In this example the sequential LP is used on the first 6500 data points. It is observed that only 180 data points are identified as infeasible during the streaming data update process; see Figure 3. The blue circles identify the location of novel (infeasible) data which we again see occur at extrema and inflection points. We see that the number of iterations of the dual simplex is generally small, e.g., 5, and rarely exceeds 20.

In the nonconvex refinement of the simplex algorithm model we use the last 10% (650) data points for validation. For LP lite 174 ± 7 novel data points were found and augmented to 800 for the SCG optimization. In Figure 4 we display the results of the prediction of the weather on 250 test data points. A detailed comparison of the methods is shown in Table II. LP and LP lite are seen to have the best overall error rates. The RBFs all have 14 centers.

	ANN	RBF	LP	LP lite
training set size	360 pts	360pts	360 pts	120 pts
training RMSE	12.39	14.85	14.54	16.20
val. RMSE	8.91	19.86	9.28	9.35
test RMSE	15.86±1.63	18.94±0.76	13.64±1.33	11.95±2.14
NPE1	0.20±0.02	0.23±0.06	0.17±0.02	0.15±0.03
NPE2	0.05±0.01	0.08±0.06	0.04±0.007	0.03±0.01
mean residual	8.66	11.70	9.28	9.30

TABLE III: Comparison between ANN and Sequential simplex algorithm with non-convex refinement on monthly sunspot data.

C. Monthly Sunspot Data

Lastly, we model the sunspot data obtained from the World DATA Center SILSO, Royal Observatory of Belgium, Brussels [29]. This data set consists of smoothed monthly total sunspot numbers from 1749 to 1789.[1]

In this example the sequential LP is applied on the first 360 input output pairs (30 years) and the subsequent 120 points (10 years) are used for testing. An average of 63 ± 3 data points are identified as infeasible during the sequential dual simplex update process. The results of this training process can be seen from Figure 5. We observe that the downslope portions of the curve, e.g., months 100-190, apparently contain little novelty in the data. The number of iterations of the dual simplex algorithm is also generally small and never exceeds 12.

In Figure 6 we show plots of the prediction on the test data. See also Table III for numerical details. The test RMSE for LP lite is approximately four sunspots more accurate when compared to the ANN. Eight RBFs were used for all the RBF sunspot examples.

VI. Background and Related Work

We present a new algorithm for sequentially learning an RBF model using the dual simplex algorithm. This approach is distinct from prior work in several fundamental ways. The compressed sensing problem has been solved using sequential observations in [1]. They update their LP after each new observation and solve an augmented primal, or Big-M problem [3]. In another compressed sensing study, the sequential problem was formulated as a quadratic program [2]. Our approach uses the dual simplex algorithm and formulates the linear program using a different treatment of the absolute value constraints.

In other recent work RBF networks have been proposed as a tool for sparse signal recovery [23]. LASSO and LARS techniques have also been incorporated with RBF networks [24]. Sparse RBF kernel methods with ℓ_1-norm penalty have been proposed [25]. The paper [26] proposed the ℓ_1-norm minimization of the weight vector only in the objective functions while imposing the RBF fitting problem as a constraint on the feasible set. The sequel [27] expanded the numerical results of [26] providing further evidence that the convex optimization

[1]The sunspot number data can be freely downloaded from [29].

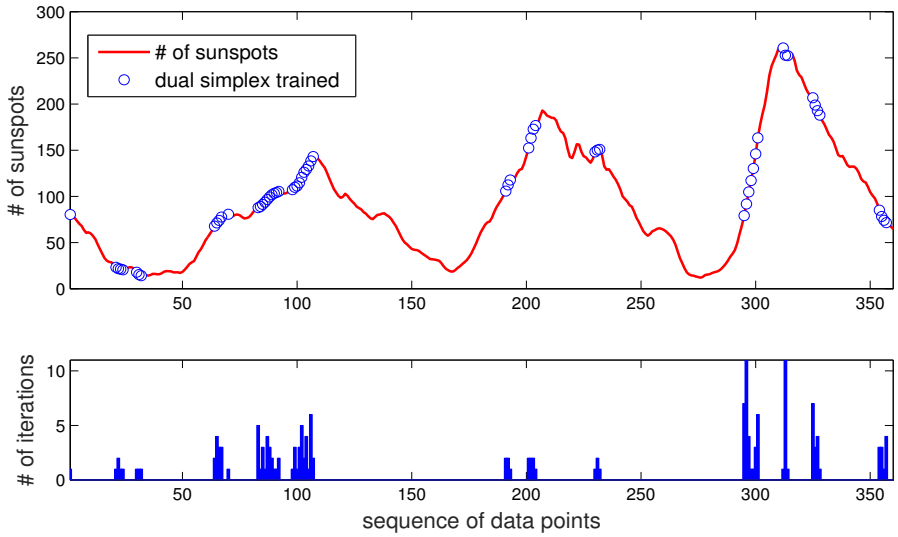

Fig. 5: Smoothed monthly sunspot data training results. A 3-dimensional time delayed embedding with delay $T = 15$ (months) is used and the predictions are 6 months ahead. During training the dual simplex algorithm 61 out of 360 data points were infeasible and were used to update the model. The maximum number of iterations performed in dual simplex is 12. A total of $N_c = 7$ centers were selected from an initial set of 500 randomly selected from the domain of the data (but not actual data points).

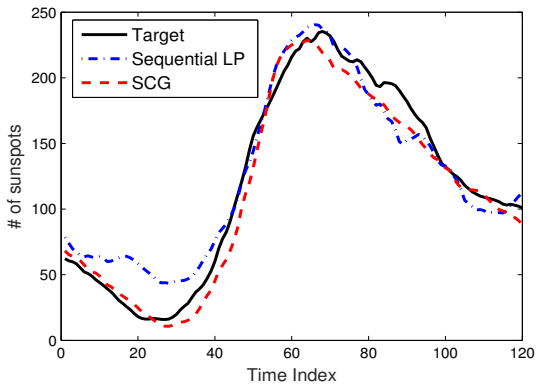

Fig. 6: A comparison of the predictions for the monthly sunspot data with the adaptive simplex algorithm. First 360 input output pairs (30 years) are generated for training. A 3-dimensional time delayed embedding is used with time delay $T = 15$ (months) and we are predicting 6 months ahead. The graph is the test result on the next 120 data points (10 years).

formulation leads to very competitive models. In these papers, in contrast to the current paper, the training was done in batch mode and there was no identification of *novel* training data.

More generally, Radial Basis Functions were proposed as a tool for arbitrary function approximation [19], [21]. A universal approximation theorem was provided for continuous functions over compact domains in [20]. There have been a series of papers whose goal has been to develop fast learning algorithms starting with [21], but this still required the tuning of a variety of *ad hoc* parameters. A comprehensive theoretical introduction to RBFs may be found in [4].

Previous efforts have been made to develop black box algorithms that require no user input for the model development, such as model order selection [10], [18]. This work proposed the use of compact skew-radial basis functions expansions to approximate time-series data. It has been observed computationally that skew-radial basis functions have impressive data fitting characteristics and can even fit abrupt jumps in data [7], [5]. More recent modifications to this algorithm with theoretical foundations have identified even more parsimonious models for, e.g., the Mackey-Glass problem [6].

VII. Conclusion

We have proposed a sequential RBF learning algorithm for streaming data that exploits the mathematical framework of the primal and dual simplex algorithms. Each new observation is incorporated into the linear program as two inequality constraints that result in the addition of two new basic variables that are required to put the problem into standard form. If the solution to the LP is still optimal, then no additional computations are required for that point and the data is discarded. If the solution is primal infeasible, the dual simplex algorithm is applied until a new optimal solution is determined.

The sequential algorithm is particularly suitable for streaming data since new observations often lead to feasible constraints. Thus, the algorithm has built into it a measure of the novelty of the training data. If a new observation resides in the polyhedron of the data constraints that have already been

observed processed, then no update is required. This aspect of the sequential update makes the algorithm very efficient in practice. We have seen that it is very effective to use the novel data, suitably augmented with randomly selected data, to train the nonconvex optimization problem providing additional acceleration of the algorithm. As the numerical examples illustrate, this LP lite approach results in substantial computational savings over using the entire training data set.

We presented results on the Mackey-Glass chaotic time-series, temperature from the Colorado State University weather station and the monthly sunspot data set. In the extreme case, we were able to process 6500 streaming data points with updates to the model needed less than 3% of the time averaged over 25 experiments. This is due to the fact that the weather data, being sampled every five minutes, is highly correlated over short times. The sunspot data was the most variable and had the highest novelty per observation at 17%. The Mackey-Glass showed an average of 13% novel data. All comparisons with the ANNs suggested the methods had comparable errors on the final models across several measures of error.

The main advantage of our proposed approach is the lack of design input required from the user and information gained about the novelty of the exemplars. The number and location of the RBFs are automatically determined solving the sequential convex optimization problems. We also note that this sequential simplex approach can provide a criterion for terminating training on a streaming data set. Namely, after a sufficiently long sequence with no infeasible constraints detected suggests training could be stopped.

Lastly, we note that this structure proposed here can be adapted to the center selection problem. One can also incrementally add centers to the training problem as non-basic variables. This leaves the primal problem feasible, and may or may not leave the dual feasible. If the problem remains optimal, one could conclude that the center is not necessary. We will explore this direction in future work.

ACKNOWLEDGMENT

This paper is based on research partially supported by the National Science Foundation under Grants Nos. DMS-1322508 and IIS-1633830. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Malioutov, Dmitry M., Sujay R. Sanghavi, and Alan S. Willsky. Sequential compressed sensing. *IEEE Journal of Selected Topics in Signal Processing*, 4.2 (2010): 435-444.

[2] S. Sra and J. A. Tropp, Row-action methods for compressed sensing, in ICASSP, vol. 3, 2006, pp. 868871.

[3] D. Bertsimas and J. Tsitsiklis, Introduction to Linear Optimization, Athena Scientific 1997.

[4] Martin D. Buhmann and M. D. Buhmann. Radial Basis Functions. Cambridge University Press, New York, NY, USA 2003

[5] A. Jamshidi and M. Kirby. Skew-radial basis function expansions for empirical modeling. *SIAM J. Sci. Comput.*, 31(6):4715–4743, 2010.

[6] Arta A Jamshidi and Michael J Kirby. A radial basis function algorithm with automatic model order determination. *SIAM Journal on Scientific Computing*, 37(3):A1319–A1341, 2015.

[7] A. Jamshidi and M. Kirby. Skew-radial basis functions for modeling edges and jumps. In *Mathematics in Signal Processing Conference Digest*, Royal Agricultural College, Cirencester, U.K., December 2008. The Insititute for Mathematics and its Applications.

[8] D.S. Broomhead and M. Kirby. A new approach for dimensionality reduction: Theory and algorithms. *SIAM J. of Applied Mathematics*, 60(6):2114–2142, 2000.

[9] D.S. Broomhead and M. Kirby. The Whitney reduction network: a method for computing autoassociative graphs. *Neural Computation*, 13:2595–2616, 2001.

[10] A.A. Jamshidi and M. Kirby. Towards a black box algorithm for nonlinear function approximation over high-dimensional domains. *SIAM J. Sci. Comput.*, 29:941, 2007.

[11] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[12] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[13] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[14] M. Kirby. *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. Wiley, 2001.

[15] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[16] Emmanuel Candes and Justin Romberg. ℓ_1-magic: Recovery of sparse signals via convex programming. *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 4:46, 2005.

[17] Michael C Mackey and Leon Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, 1977.

[18] Arta Jamshidi. *Modeling Spatio-Temporal Systems with Skew Radial Basis Functions: Theory, Algorithms and Applications*. Ph.D. dissertation, Colorado State University, Department of Mathematics, 2008.

[19] David S Broomhead and David Lowe. Radial basis functions, multivariable functional interpolation and adaptive networks. Technical report, DTIC Document, 1988.

[20] Jooyoung Park and Irwin W Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.

[21] John Moody and Christian J Darken. Fast learning in networks of locally-tuned processing units. *Neural computation*, 1(2):281–294, 1989.

[22] A. A. Jamshidi and M. J. Kirby. Examples of Compactly Supported Functions for Radial Basis Approximations. In E. Kozerenko H. R. Arabnia and S. Shaumyan, editors, *Proceedings of The 2006 International Conference on Machine learning; Models, Technologies and Applications*, pages 155–160, Las Vegas, June 2006. CSREA Press.

[23] Vidya L., Vivekanand V., Shyamkumar U., and Deepak Mishra. RBF-network based sparse signal recovery algorithm for compressed sensing reconstruction. *Neural Networks*, 63:66 – 78, 2015.

[24] Quan Zhou, Shiji Song, Cheng Wu, and Gao Huang. Kernelized lars–lasso for constructing radial basis function neural networks. *Neural Computing and Applications*, 23(7):1969–1976, 2013.

[25] Lihua Fu, Meng Zhang, and Hongwei Li. Sparse RBF networks with multi-kernels. *Neural Processing Letters*, 32(3):235–247, 2010.

[26] T. Ghosh, M. Kirby, and X. Ma, Sparse skew radial basis functions for time-series prediction, *Proceedings International Work Conference on Time Series Analysis* , pp. 296-307, Granada, Spain, June 2016.

[27] T. Ghosh, M. Kirby, and X. Ma, Dantzig-Selector Radial Basis Function Learning with Nonconvex Refinement, *submitted ITISE Springer Book*.

[28] Candes, Emmanuel, and Terence Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 2313-2351, 2007.

[29] SILSO, World Data Center - Sunspot Number and Long-term Solar Observations, Royal Observatory of Belgium, on-line Sunspot Number Catalog: http://www.sidc.be/SILSO/, 1749–1789

[30] Christman Field Weather Station, Colorado State University, www.atmos.colostate.edu/wx/fcc/.