Linear Zonal Atmospheric Prediction for Adaptive Optics

Patrick C. McGuire[1], Troy A. Rhoadarmer,
Hanna Coy, J. Roger P. Angel, Michael Lloyd–Hart

Center for Astronomical Adaptive Optics,
Steward Observatory, University of Arizona,
Tucson, AZ 85721, USA

ABSTRACT

We compare linear zonal predictors of atmospheric turbulence for adaptive optics. Zonal prediction has the possible advantage of being able to interpret and utilize wind–velocity information from the wavefront sensor better than modal prediction. For simulated open–loop atmospheric data for a 2–meter 16–subaperture AO telescope with 5 millisecond prediction and a lookback of 4 slope–vectors, we find that Widrow–Hoff Delta–Rule training of linear nets and Back–Propagation training of non–linear multilayer neural networks is quite slow, getting stuck on plateaus or in local minima. Recursive Least Squares training of linear predictors is two orders of magnitude faster and it also converges to the solution with global minimum error. We have successfully implemented Amari's Adaptive Natural Gradient Learning (ANGL) technique for a linear zonal predictor, which premultiplies the Delta–Rule gradients with a matrix that orthogonalizes the parameter space and speeds up the training by two orders of magnitude, like the Recursive Least Squares predictor. This shows that the simple Widrow–Hoff Delta–Rule's slow convergence is not a fluke. In the case of bright guidestars, the ANGL, RLS, and standard matrix–inversion least–squares (MILS) algorithms all converge to the same global minimum linear total phase error (~0.18 rad^2), which is only ~5% higher than the spatial phase error (~0.17 rad^2), and is ~33% lower than the total 'naïve' phase error without prediction (~0.27 rad^2). ANGL can, in principle, also be extended to make non–linear neural network training feasible for these large networks, with the potential to lower the predictor error below the linear predictor error. We will soon scale our linear work to the ~108–subaperture MMT AO system, both with simulations and real wavefront sensor data from prime focus.

Keywords: atmospheric turbulence prediction, linear prediction, neural network prediction, zonal prediction, Widrow–Hoff neural networks, delta–rule training, natural gradient learning, recursive least squares, servo lag error, least–squares predictors

1. Introduction

There are several different contributions to residual wavefront error in adaptive optics systems, including wavefront sensor CCD noise, photon counting noise, wavefront fitting and reconstructor errors, focus and tilt anisoplanatism, and servo lag error. Herein, we concentrate on reducing the servo lag error by developing spatiotemporal predictors of atmospheric turbulence. If we have a quality wavefront predictor in hand, then actuator commands can be sent to the deformable mirror that more accurately represent the conjugate atmospheric wavefront at the 'present' time, as opposed to sending actuator commands to the deformable mirror that actually represent the state of the atmosphere a few milliseconds in the past.

Several groups have tried different atmospheric turbulence predictive techniques for adaptive optics (Aitken, Jorgenson & McGaughey[1]; Wild[2]; Lloyd–Hart, Rhoadarmer & McGuire[3,4]; Dessenne, Madec & Rousset[5]; Rhoadarmer & Ellerbroek[6]; etc.), and all have reported significant improvements in residual wavefront phase error when employing their predictors: for simulations, for post–facto predictive analysis of real wavefront sensor data from a telescope, and/or for in situ on–the–fly prediction. Desenne et al have concentrated on training modal predictors. Rhoadarmer & Ellerbroek have had significant success training closed–loop zonal predictors. Lloyd–Hart, Rhoadarmer & McGuire have concentrated on exploring different advanced open–loop zonal predictive algorithms. Wild concentrated on matrix–algebra based zonal techniques, both open– and closed–loop.

For adaptive optics where the mirror actuator positions are known (such as the MMT adaptive secondary system), the open–loop prediction techniques makes some sense, since the input and output atmospheric wavefronts needed for

---

training the predictor can be determined by adding the known mirror shape to the measured wavefront. Otherwise, if the mirror shape is unknown, the only use of training open−loop predictors is to gain experience before proceeding to the more difficult problem of training closed−loop predictors. The open−loop prediction problem is in principle much easier than the closed−loop prediction problem because the mirror and measurement dynamics are not as intimately coupled into the prediction problem. For open−loop prediction, we are in principle only predicting the statistically moderately−well−understood atmospheric spatiotemporal dynamics.

The concept of Taylor frozen flow is a good approximation on short time−scales (t $\ll \tau$, where τ is the atmospheric coherence time) for single layers of atmospheric turbulence. With the assumption of frozen flow, wind from some preferred direction will blow a static wavefront of aberration across the telescope aperture. In such a case, it makes much sense to make predictions of the wavefront based on the prior wavefronts over the subaperture of interest and only the neighboring subapertures, and so, zonal prediction seems the preferred choice over modal prediction. In the case of modal prediction, the only way to keep track of wind velocity information for prediction purposes is by keeping track of the dynamics of the relative amplitudes of several different modes simultaneously. Keeping track of the history of a single mode (e.g. spherical aberration) will contain little information as to the wind direction, and hence, utilizing the Taylor frozen flow idea to predict future wavefronts becomes much more complex. However, modal prediction does have the advantage that most of the atmospheric aberration is in the low−order modes. Therefore, if only the time history of a limited number of modes (e.g. less than 16 for the case studied here, and less than 30 for the case of ~100 subapertures) are used to predict future modal amplitudes, then a significant reduction in the number of predictive parameters becomes possible under modal prediction, but high spatial−frequency information is then discarded. If one wants to retain all the measured wavefront information, and if one wants to retain the local nature of the predictor (as opposed to having the prediction of Zernike mode #5 depend with significant amplitude on the past history of Zernike mode # 20, for example) and possibly retain the physics of the Taylor frozen flow concept, then it seems that zonal prediction is preferred over over modal prediction. For anticipated AO systems with 3000 subapertures, a straightforward global predictive analysis (either zonal or modal) will suffer a severe dimensionality curse. However, with the pruning of unimportant predictive parameters afforded by the locality of Taylor flow, it appears to us that zonal prediction will offer significantly better computational properties than non−local modal prediction. With multiple non−interacting layers of Taylor turbulence with different wind velocities and correlation lengths, these arguments become less precise. However, in this work we explore zonal predictors, with the anticipation that one day we can adequately compare the performances of modal predictors with those we obtained here.

We have made forays into both linear and non−linear predictors during our research, and have had success employing both techniques. However, in our experience, for zonal predictors, the explored non−linear techniques (e.g. back−propagation neural networks[3,4]) take a long time (hundreds of thousands of training frames) to converge to their minimum predictive phase error, and often do not have as low a phase error in the end as their linear siblings. Therefore, we decided to explore back−propagation linear networks[4] to see if such linear networks had the same convergence problems as the non−linear networks, and indeed they do, as reported conclusively here for the case of Widrow−Hoff Delta−Rule networks. In the high signal−to−noise case, the Delta−Rule (or back−propagation for multiple layer networks) seems to be plagued for these large networks by running into sub−optimal solutions, which in the parlance of machine−learning are called 'local minima' in the training error surface, as opposed to the true global minimum. Therefore, development of other predictive technologies is paramount, and in this work, we show that three different zonal linear predictor algorithms (matrix inversion, recursive least squares, and adaptive natural gradient descent) all converge to the same minimum error and to the same predictive matrix within 1000−6000 training frames recorded in simulation at 200 Hz. The adaptive natural gradient descent technique has the possibility of being directly extended to multi−layer non−linear networks, and hence the possibility of even further improvements in the residual phase error below the phase error for linear predictors becomes possible, in principle. However, in our initial forays into such a non−linear extension, we have found the memory requirements to be prohibitive for the large networks in question.

2. Techniques

The architecture of a basic zonal predictor is shown in Figure 1. We use all the slope values in the past 4 frames of data each with a linear weight to predict each of the slope values in the next frame of data. For the linear case, the predictive problem can also be represented in matrix notation. The whole problem is to find adequate techniques to determine accurate values of the linear weights between the inputs and the outputs.

The basic goal of deriving a good predictor is to minimize the error when the predictor is utilized. In the case of predicting slopes, the prediction error , E_{slope}, is given by:

$$E_{slope} = \left\langle \sum_{i=1}^{32} (s_i^{Meas}(t) - s_i^{Pred}(t))^2 \right\rangle = \left\langle (\vec{s}^{\ Meas}(t) - \vec{s}^{\ Pred}(t))^2 \right\rangle, \qquad (1)$$

where $\vec{s}(t)$ is the 32 element x & y slope vector at time t; $s_i(t)$ is the ith element of that slope vector; the *Pred* and *Meas* superscripts refer to the predicted slope vector and the actually measured slope vector, respectively; and the angular brackets denote a time average, and a similar expression is defined for the phase error:

$$E \equiv E_{phase} = \left\langle \sum_{i=1}^{6^2 \times 32} (\phi_i^{Recon}(t) - \phi_i^{Pred}(t))^2 \right\rangle = \left\langle (\vec{\phi}^{\ Recon}(t) - \vec{\phi}^{\ Pred}(t))^2 \right\rangle, \qquad (1a)$$

where we sample the phase reconstruction 6^2 times per subaperture, and we use phase−maps reconstructed with a standard least−squares technique from the simulated slope measurements.

The idea behind matrix−inversion least−squares (MILS) predictors[3] is very similar to the idea behind least−squares reconstructors. The least−squares predictor is determined by first recording the the array of slope values ('slope vector' = $\vec{s}(t)$) determined from the wavefront sensor, for N = a few thousand frames, and then
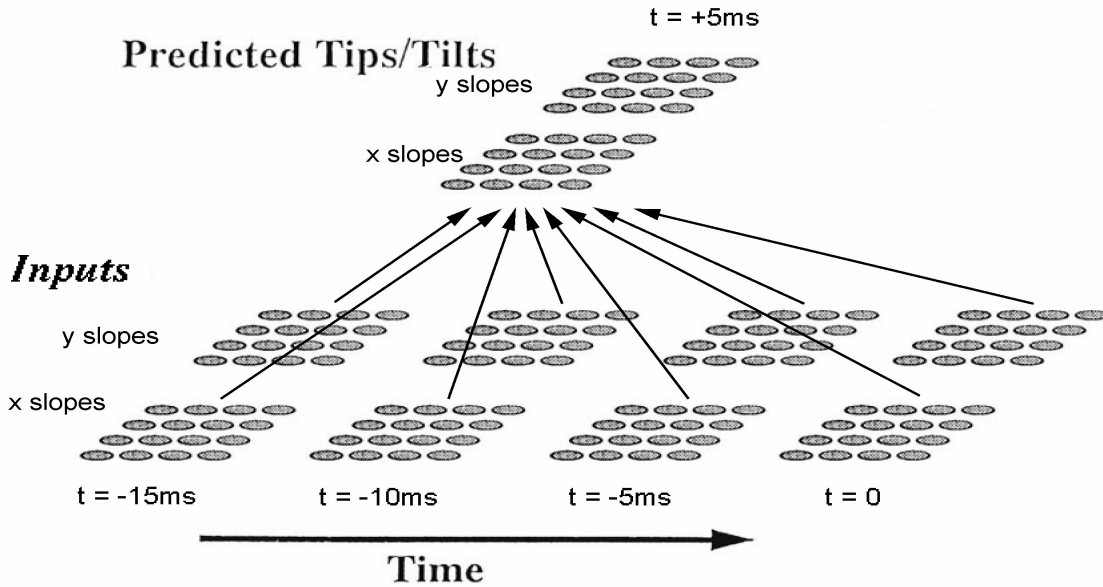

Figure 1: A drawing showing how all the slopes from a 4×4 Shack−Hartmann array for the past four frames are used to predict the future slopes (adapted from Refs. 1 & 4).
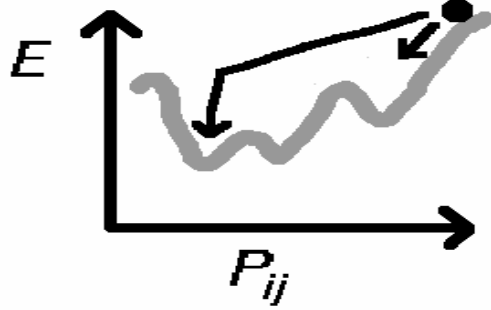
Figure 2: A sketch showing the phase error of a predictor as a function of of a single connection weight, W_{ij} (between past input j and future output i), with all the other weights fixed. We show how unaided gradient descent algorithms (denoted by the ball following the local slope of the short arrow) potentially can get trapped in local minima of the phase error surface, whereas more sophisticated algorithms can actually find the global minimum.

form two matrices **F** (for future) and **H** (for historical) from this data by putting the slope vectors **s**(t) in the columns of these matrices:

$$F \quad = \quad \left(\vec{s}(t) \quad \vec{s}(t-1) \quad \vec{s}(t-2) \quad \cdots \quad \vec{s}(t-(N-4)) \right)$$

$$H \quad = \quad \begin{pmatrix} \vec{s}(t-1) & \vec{s}(t-2) & \vec{s}(t-3) & \cdots & \vec{s}(t-(N-3)) \\ \vec{s}(t-2) & \vec{s}(t-3) & \vec{s}(t-4) & \cdots & \vec{s}(t-(N-2)) \\ \vec{s}(t-3) & \vec{s}(t-4) & \vec{s}(t-5) & \cdots & \vec{s}(t-(N-1)) \\ \vec{s}(t-4) & \vec{s}(t-5) & \vec{s}(t-6) & \cdots & \vec{s}(t-N) \end{pmatrix} \quad (2)$$

Then, with **F** and **H**, we obtain the predictor matrix **P** by the taking the pseudoinverse of **H** and premultiplying the pseudoinverse by **F**:

$$P = F H^{+} = F H^{T} (H H^{T})^{-1} \quad (3)$$

such that:

$$\vec{s}(t) = P \begin{pmatrix} \vec{s}(t-1) \\ \vec{s}(t-2) \\ \vec{s}(t-3) \\ \vec{s}(t-4) \end{pmatrix} \equiv P \vec{s}_{H}(t) \ , \quad (4)$$

where $\vec{s}_{H}(t)$ is a composite vector of the last four slope vectors.

The recursive least squares (RLS) technique[7] determines the predictor matrix iteratively so that $P = P(t)$, as opposed to doing it in one big swoop as in the MILS technique. The RLS technique relies on the use of the Kalman filter, and is a rank one update algorithm. The Kalman gain vector $\vec{k}(t)$ is given by $\vec{k}(t) = P(t-1) \vec{s}_{H}(t)$. The RLS predictor is then given by:

$$P(t)=f^{-1}\left(P(t-1)-\frac{\vec{k}(t)\vec{k}^{\,T}(t)}{\alpha(t)}\right) \quad , \quad (5)$$

where f is the forgetting factor (usually chosen to be 0.9), and $\quad \alpha(t)=f+(\vec{s}_H(t))^T\vec{k}(t)\quad$.

        The gradient descent algorithms update each matrix element or weight in an iterative fashion. If the matrix element or weight at row i and column j of P is denoted by P_{ij}, then with stochastic gradient descent (also known as Widrow–Hoff's Delta Rule[7,8]), the weight P_{ij} is updated by an amount ΔP_{ij} given by the analytically–determined slope of (or gradient) of the error surface:

$$\Delta P_{ij}(t) \quad = \quad -\epsilon\frac{dE(t)}{dP_{ij}} \quad , \quad (6)$$

$$= \quad -\epsilon\left(s_i^{Meas}(t)-s_i^{Pred}(t)\right)(s_H)_j(t)$$

or in a vectorized matrix notation:

$$\Delta\vec{P}(t)=-\epsilon\frac{dE(t)}{d\vec{P}} \quad (6a)$$

where we have used equations 1 & 4. The advantage of the Widrow–Hoff (WH) algorithm is its utter simplicity, correcting the weight between input j and output i by a scalar product between the predictor error for output node i and the measured historical slope for input node j.

        The adaptive natural gradient descent learning (ANGL) algorithm[9] has the advantage over the stochastic Widrow–Hoff gradient descent in that it can effectively remove the slowing effect that correlations between the different weight parameters have on training performance. The ANGL algorithm accomplishes this by pre–multiplying the stochastic gradient by an iteratively–determined decorrelating matrix, $G(t)$:

$$\Delta\vec{P}(t)=-\epsilon\,G(t)\frac{dE(t)}{d\vec{P}} \quad , \quad (7)$$

where $G(t)$ is given by:
$$G(t)=(1+\gamma)G(t-1)-\gamma G(t-1)\vec{s}_H(t-1)\vec{s}_H^{\,T}(t-1)G(t-1) \quad , \quad (8)$$

where a value of γ of 0.001 is typically chosen.

        Another way of looking at the ANGL algorithm is that WH algorithm presumes that the weights live in an orthogonal Euclidean space, so that the length of the gradient vector used in equation (6) is just the sum of the squares of its individual components:

$$|dP|^2 = dP\,(dP)^T. \quad (9)$$

The ANGL algorithm presumes a Riemannian space, in which there is a metric much like that in general relativity, which allows non–orthogonality between the different weight coordinate axes, and the length of the gradient vector used in equation (7) is determined by first orthogonalizing with the metric tensor g:

$$|dP|^2 = g\,dP\,(dP)^T. \quad (10)$$

The metric g can be determined on theoretical grounds from a fundamental concept called Fisher information[10]. The matrix G is the inverse of the metric tensor matrix g. The Recursive Least Squares (RLS) algorithm and the ANGL algorithm both use the Kalman filter, and the mathematics is quite similar in the present case of linear systems. However, the ANGL algorithm is directly extendable to the case of non–linear predictors, whereas the RLS algorithm is not extendable to non–linear systems without prior modification.

        One potential problem with the ANGL technique is that it increases the memory requirements since we now need to keep track of $G(t)$, which is a 128×128 matrix for the case of 16 subapertures and a look–back of 4 and no hidden layer. This is not prohibitive for the present case without a hidden layer (in which by luck, the actual (128×32) ×(128×32) G matrix is block diagonal with 128×128 blocks, so only a 128×128 matrix needs to be tracked. If we want to use the full power of non–linear networks with a hidden layer the problem becomes daunting because the matrix is then M×M, where M is the total number of connections and M is typically 128×30 + 30×32 =4800 (with 128 inputs, 30 hidden nodes and 32

outputs).

3. Results

We trained the networks using the same simulated 3–layer atmosphere over a 2 meter diameter telescope with 16 Shack–Hartmann subapertures, as used by Lloyd–Hart & McGuire[3]. The resulting training curves (for zero noise, or a M=0 guidestar) for the 4 different algorithms are shown in figure 3, where we have subtracted the spatial fitting and reconstruction errors (about 0.17 rad^2) to determine the temporal error that is due to the lack of perfect prediction. We used 1000 frames for the matrix inversion (MILS) predictor, and the other predictors are updated every frame. We also compare these sophisticated predictors to the naïve predictor, in which the slopes are presumed to be random walking, so that the best prediction is that the next slope will be identical to the last. The WH algorithm takes over 10^5 frames to converge, even with the highest gain possible (ε~0.01). The other 3 algorithms converge to the same minimum temporal phase error (E~0.01) in less than 10^4 frames, and the RLS algorithm gets very close to its minimum temporal phase error in less than 10^3 frames.

The slow convergence of the WH algorithm suggests that it is on a plateau on the error surface in weight space, or that it is stuck in a very broad local minimum. In Figure 4, we check this hypothesis, by directly comparing the weights determined by the various predictors with that of our baseline, the MILS predictor. For both the RLS and ANGL algorithms, the weights are quite similar to the weights determined by the MILS algorithm, differing randomly by ~20%. However, for the WH algorithm, most of the weights are systematically lower in absolute magnitude than the MILS weights by a factor of ~3, though a few of the largest WH weights seem to be adequately close to the MILS weights. So definitely, the WH algorithm is converging to a different solution than the other three algorithms.
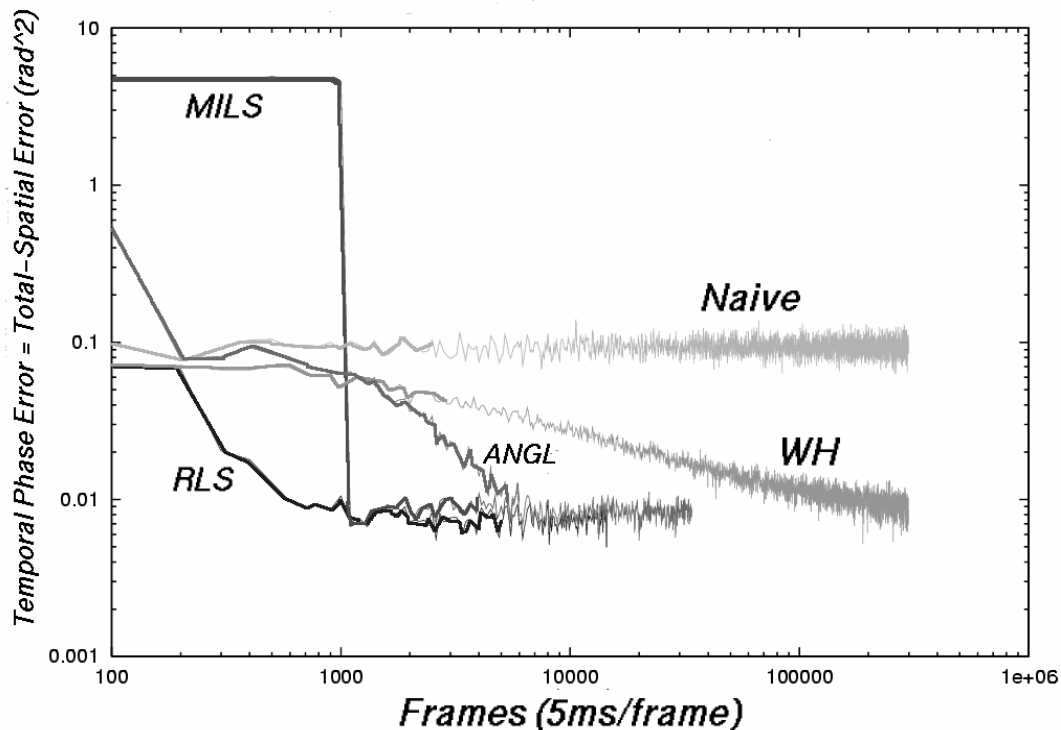

Figure 3: We plot the temporal phase error as a function of training time for 4 different training algorithms. We also show the 'naïve' predictor temporal phase error, in which it is assumed that the atmosphere is random walking, so that the best prediction is that the next slope will be identical to the last. Clearly, the Widrow–Hoff (WH) gradient descent linear network takes over two orders of magnitude longer training time than the other three algorithms. The recursive least squares (RLS) algorithm allows continual updating of the predictor matrix, so it can do better more quickly than the matrix inversion least squares (MILS) solution, and RLS can continue to

update and improve slightly even before enough additional data can be acquired for the next MILS matrix inversion. The adaptive natural gradient descent method (ANGL) also converges to the global minimum rather quickly (relative to WH gradient descent), which suggests that the WH gradient descent algorithm is getting stuck in a local minimum.

In figure 5, we train on dim simulated guidestars, and tabulate performance as a function of guidestar magnitude. For bright guidestars ($M_{V+R} \leq 10$), the error ratio graph in Figure 5b shows that the WH algorithm is considerably worse than the other three algorithms. The other three algorithms are admirable in the sense that they allow total phase errors for the brightest guidestars ($M_{V+R} \leq 5$) that is only ~5% higher than the spatial phase error. Actually the WH algorithm trains rather rapidly in the presence of significant noise ($M_{V+R} > 12$) to solutions which are competitive with the other 3 algorithms (as one might expect from the good noise performance[8] found in training Widrow–Hoff reconstructors). However, for all guidestar magnitudes, the RLS algorithm is the best, with the most superiority for dim guidestars ($M_{V+R} > 11$).
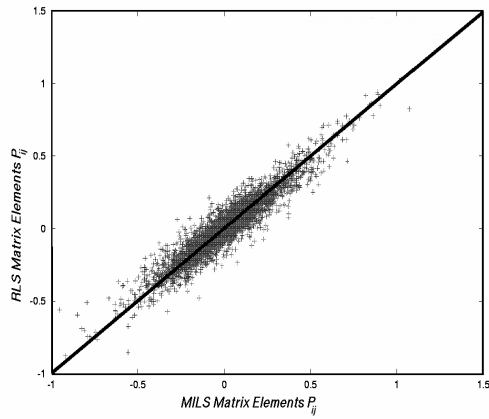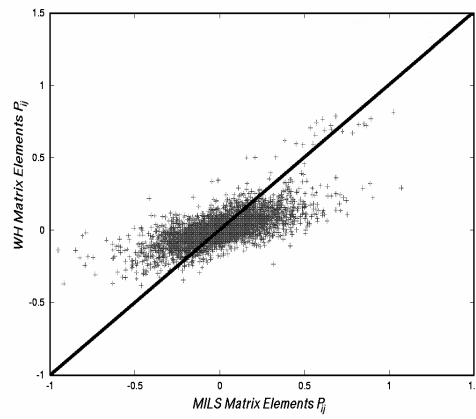
a)   b)
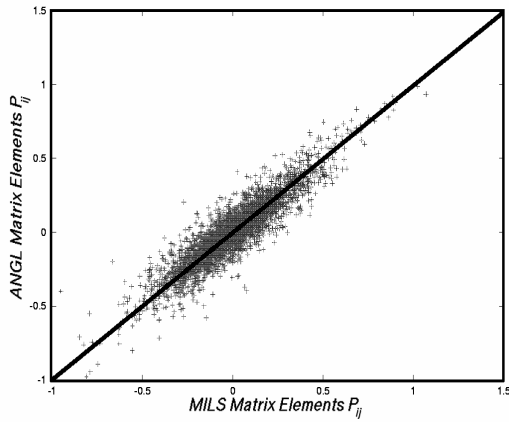
c)

Figure 4: A comparison of the final connection weights for the 3 different training algorithms relative to the matrix inversion least–squares weights. Figure (a) compares the weights of the recursive least squares algorithm (RLS) with the matrix inversion weights (MILS). Figure (b) compares the wights of the Widrow–Hoff gradient descent (WH) algorithm with the MILS weights. Figure (c) compares the weights learned by the adaptive natural gradient algorithm (ANGL) with the MILS weights. Clearly the weights learned by the WH algorithm differ from the weights learned by the other algorithms, whereas the other three algorithms learn very similar weights. This gives further evidence that the WH algorithm is converging to a local minimum rather than the global predictor phase error minimum as the other three algorithms seem to find.
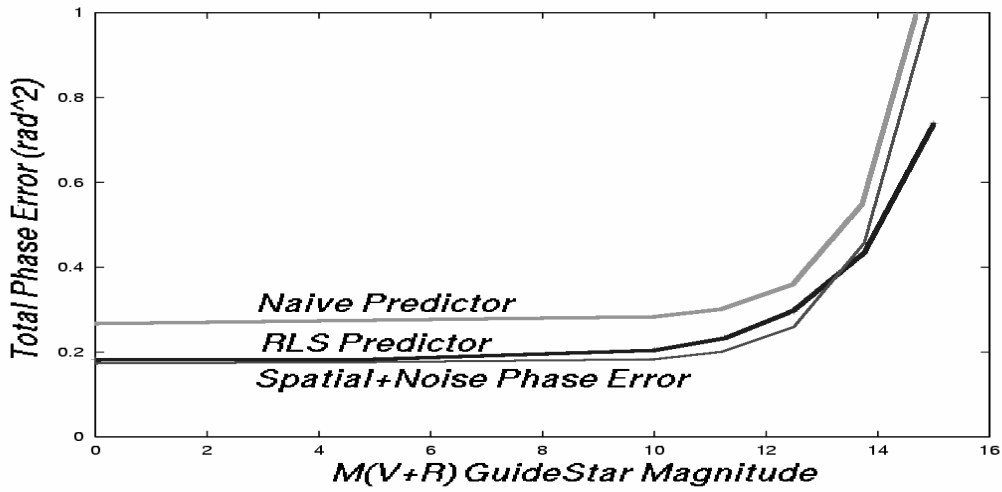
a)
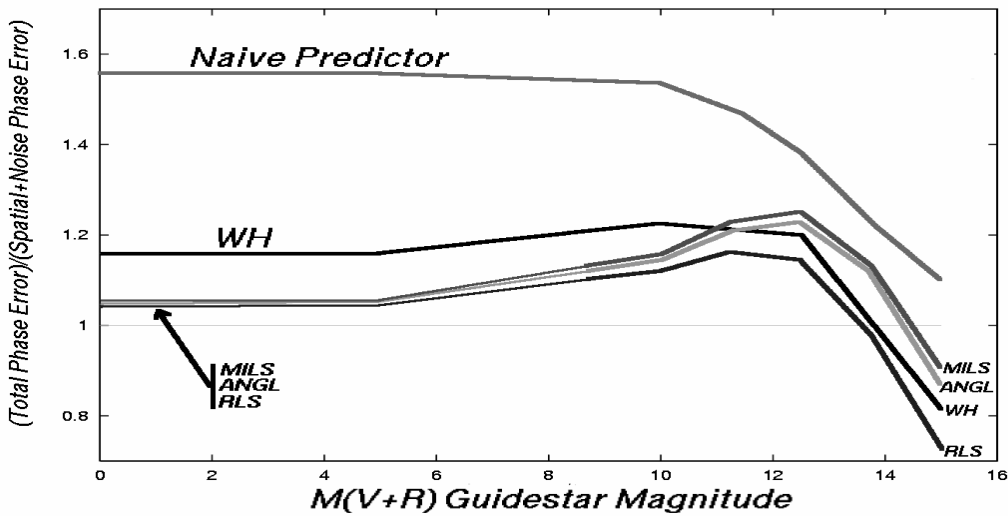
b)

Figure 5: We show how the predictor error for the four different algorithms changes with guidestar magnitude. Figure (a) shows the total phase error (for only the RLS algorithm), and figure (b) shows the ratio of the total phase error to the spatial phase error (for all four algorithms). The spatial+noise phase error is a combination of the fitting and reconstructor errors and the photon noise error. The total phase error encroaches below the spatial phase error for dim guidestars ($M_{V+R}>13.5$) due to the allowance the predictor provides for temporal averaging of several frames together. Clearly, the Widrow−Hoff (WH) algorithm performs worse than the other three algorithms for bright guidestars, though WH perfoms admirably for the high noise case of dim guidestars ($M_{V+R}>11$). The other three algorithms converge to the same total phase error (~0.18 rad^2) for bright guidestars ($M_{V+R}<=5$) , which is only ~5% higher than the spatial phase error (~0.17 rad^2), and ~33% lower than the naïve predictor total phase error (~0.27 rad^2).

Presumably the MILS predictor is the worst of the 3 'good' predictors for all guidestar magnitudes since we have done nothing to optimize the number of training frames used in the MILS matrix inversion. This lack of optimization allows the RLS algorithm to continue to improve sligthly over the MILS algorithm after the nominal 1000 MILS frames for bright guidestars. For dimmer guidestars, the 1000 MILS frames seems to be inadequate, as the improvement over MILS offered by RLS is enhanced over the case of bright guidestars. The mediocre performance of the ANGL algorithm for dim guidestars likely also is due to the lack of optimization of the learning rate parameter, ε , and the smoothing parameter, γ, used in determining *G.* But it is possible that these three algorithms have inherently different characteristics when training in the presence of noise. Intuitively, the poorer performance of the MILS algorithm as compared to the RLS algorithm, in the case of bright guidestars as RLS continues to improve after 1000 training frames, and also in the case of dim guidestars, is primarily due to the fact that the MILS algorithm is being tested on data that it has not been trained upon adequately, whereas the RLS algorithm dynamically adjusts the effective amount of training time so as to offer superior performance to MILS.

One might note that for the dimmest guidestars (M_{V+R}> 14) all the predictors have total phase errors that are less than the spatial+noise phase error. This is due to the 4–frame temporal averaging of the slope vectors afforded by a predictor as opposed to the baseline case of a reconstructor that only has 1 frame to base its estimates of the phase upon, and hence the possibility of a four–fold gain in signal–to–noise of the slope measurements.

As the guidestars become quite dim (M_{V+R} > 14), the spatial+noise phase error increases considerably, to more than 1 rad^2, and the naïve predictor error increases considerably as well, tracking the spatial+noise phase error, which suggests that most of the information that can be used for prediction is getting washed out by the noise.

In our eyes, since we made the ANGL algorithm work as well as MILS and the RLS algorithms in the high signal–to noise case, and since the ANGL algorithm is a more complex version of the WH algorithm, this means that the simple WH algorithm for linear networks is not powerful enough for this problem. This also suggests more generally that the back–propagation training technique for non–linear neural networks is also not an appropriate technique for this problem. In the low signal–to–noise case, the WH algorithm may offer some marginal advantages in simplicity and speed for predictive hardware at the telescope, but hopefully most AO systems will not normally need to work with such dim guidestars, so it behooves us to focus our energies on techniques like the RLS predictor.

Some of the Future of Prediction for Adaptive Optics

We have had success[11] in predicting real open–loop atmospheric turbulence data from the 1.5 meter telescope at Starfire Optical Range with the RLS predictor.

In the near future, our group plans efforts on two fronts:
1) Upgrading the present code for the 3 predictors for small AO systems to larger AO systems with more than 100 subapertures. For the case of RLS, this is feasible even in the full zonal case, as it avoids the large matrix inversions necessary for MILS.
2) Extending our present predictive analysis efforts to real atmospheric turbulence data, first with the 1.5 meter data we have from Starfire, and then with dynamical wavefront sensor data that we should acquire soon from the 6.5 meter MMT, either at Prime Focus, or at Cassegrain focus.

In the more distant future, we hope to:
1) Find out if significant improvement can be afforded by extending our predictive techniques back into the non–linear 'neural network' arena.
2) Directly compare our open–loop zonal predictive techniques to other predictive techniques, such as modal prediction, Wild's matrix–algebra predictors, closed loop prediction, and phase (instead of slope) prediction.
3) Follow up on Schoeck's work[12], in which he reconstructs the wind velocity and direction for each of the turbulence layers with a single Shack–Hartmann wavefront sensor, and also estimates the Taylor and non–Taylor component of the flows in each of the turbulent layers. Such information should prove invaluable when training better predictors.

REFERENCES

1. M.B. Jorgenson & G.J.M. Aitken, "Wavefront Prediction for Adaptive Optics", *European Southern Observatory Conf. on Active and Adaptive Optic*, Ed. F. Merkle, Garching, Germany, p. 143, 1994;

   G.J.M. Aitken & D.R. McGaughey, "Predictability of Atmospherically–Distorted Stellar Wavefronts", *Proc. European Southern Observatory Conf. on Adaptive Optics* **54**, Ed. M. Cullum, Garching, Germany, p. 89, 1996;

   G.J.M. Aitken, D. Rossille, & D.R. McGaughey, "Predictability of Fractional–Brownian–Motion Wavefront Distortions and Some Implications for Closed–Loop Adaptive Optics Control", *Proc. SPIE Conf. on Adaptive Optical System Technologies* **3353**, pp. 1060–1069, 1998.

2. W.J. Wild, "Predictive Optimal Estimators for Adaptive Optics Systems", *Optics Letters* **21**, pp. 1433–1435, 1996.

3. M. Lloyd–Hart & P.C. McGuire, "Spatio–Temporal Prediction for Adaptive Optics Wavefront Reconstructors," Adaptive Optics: Topical Mtg. & Tabletop Exhibit, Technical University of Munich, Garching, Germany, 1995.

4. P. C. McGuire, D. G. Sandler, M. Lloyd–Hart, & T. A. Rhoadarmer. *Scientific Applications of Neural Nets, Lecture Notes in Physics*, chapter 2, "Adaptive Optics: Neural Network Wavefront Sensing, Reconstruction, and Prediction", Springer, Heidelberg, pp. 97–138, 1999.

5. C. Dessenne, P.Y. Madec, & G. Rousset:

   "Modal Prediction for Closed–Loop Adaptive Optics", *Optics Letters* **22**, pp. 1535–1537, 1997;

   "Optimization of a predictive controller for closed–loop adaptive optics", *Applied Optics* **37**, pp.4623, 1998;

   "Sky Implementation of Modal Predictive Control in Adaptive Optics", *Optics Letters* **24**, pp. 339–341, 1999.

6. T.A. Rhoadarmer & B.L. Ellerbroek, "Real–time Adaptive Optimization of Wave–front Reconstruction Algorithms for Closed–loop Adaptive–Optical Systems", *Proc. SPIE Conf. on Adaptive Optical System Technologies* **3353**, pp. 1174–1185, 1998.

7. P. Strobach, *Linear Prediction Theory: A Mathematical Basis for Adaptive Systems*, Springer–Verlag, Berlin, 1990.

8. J.R.P. Angel, "Wavefront Reconstruction by Machine Learning Using the Delta Rule", *Proc. SPIE Conf. On Adaptive Optics for Astronomy* **2201**, Kona, Hawaii, p. 629, 1994.

9. S. Amari, "Natural gradient works efficiently in learning ", *Neural Computation* **10**, pp. 251–276, 1998;

   S. Amari, H.Park and K.Fukumizu, "Adaptive method of realizing natural gradient learning for multilayer perceptrons", RIKEN Brain Science Institute preprint, 1999.

10. B.R. Frieden, *Physics from Fisher Information: a Unification*, Cambridge University Press, New York, 1998.

11. T.A. Rhoadarmer & M. Lloyd–Hart, unpublished, 1996.

12. M. Schöck & E. Spillar:

    "Measuring Wind Speeds and Turbulence with a Wave–front Sensor ", *Optics Letters* **23**, pp. 150–152, 1998;

    "Turbulence Analysis with the Starfire Optical Range 3.5–meter Telescope", *Astron.Soc.Pac.***174**, pp. 119, 1999;

    "An Analysis of Turbulent Atmospheric Layers with a Wave Front Sensor: Testing the Frozen Flow Hypothesis", *Proc. SPIE Conf. on Adaptive Optics Systems and Technology* **3762**, pp. 225–236, 1999.