Sparse prediction of turbulence in time and space

Rufus Fraanje[12,a], Niek Doelman[2], and Michel Verhaegen[1]

[1] Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

[2] TNO Science and Industry, Stieltjesweg 1, 2628 CK, Delft, The Netherlands

Abstract. Using a matching pursuit (MP) approach to solve least squares problems with a sparsity constraint sparse controllers are computed to compensate for atmospheric turbulence. The method has been validated by simulation of von Kármán turbulence with frozen flow for the case of prediction, reconstruction and control.

1 Introduction

The need for numerically efficient controller for large scale adaptive optics (AO) has been widely acknowledged in literature, see, e.g., [5, 6, 10, 12]. The high number of actuators and sensors and the high sampling rate result in high computational load and memory usage. For example, to compensate for a single layer of turbulence with coherence length of $r_0 = 20$cm over a telescope aperture of $D = 40$m the number of actuators and sensors will be in the order of $(D/r_0)^2 = 40.000$. When the sampling rate of the controller is 2kHz even a simple integral controller where the mapping from sensors to actuators is performed by one matrix multiplication takes already a computational load of 6.4Tflops (1flops = 1 floating point operation per second). Current field programmable gate array (FPGA) technology enables peak-performances of 10-100Gflops, such that the matrix multiplication need to be distributed over about 100 FPGAs. Exploiting the temporal correlation in the turbulence by means of optimal prediction, c.f., [8], and control (LQG) as well as multiple layer AO will improve the performance of the AO system but may also further increase the computational complexity. Hence, to prevent extreme system complexity, power dissipation and cost, efficient (predictive) control methods are needed.

To simplify the control problem efficient control algorithm have been derived assuming that the turbulence is constant in time and temporal correlation has been neglected, e.g., [5, 10, 12]. In [7] it has been demonstrated on a laboratory setup that significant performance gain (up to a factor 20) can be obtained by a controller that exploits both the temporal and the spatial correlation that is present in the turbulence by means of a disturbance model. Moreover, [7] proposed a method to determine this controller from measured data rather than on the basis of a priori assumed statistics of the turbulence that may significantly deviate from reality due to variations in wind speed and direction. The computational complexity of the algorithm to determine the controller scales with the cube and the complexity of the controller scales with the square of the number of channels and the degrees of freedom in the temporal correlation. Therefore, in [4] fast adaptive control methods have been used to reduce the computational complexity and to enhance the adaptation to variations in the turbulence statistics. In [11] another predictive control algorithm has been proposed which allows a much more efficient implementation based on the Fast Fourier Transform.

This paper takes a different approach. The control problem is written as a least squares problem which is solved by imposing a bound on the maximum number of non-zero elements. In [2] is proved that this problem is NP-hard, and thus not feasible for the AO purpose. Suboptimal methods have been proposed in literature, e.g., see [1] for a recent overview, among which the class of Matching

[a] e-mail: `p.r.fraanje@tudelft.nl`
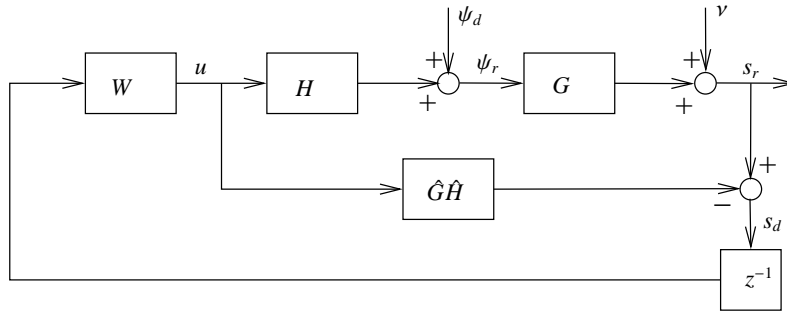
Fig. 1. Block diagram of the AO configuration with internal model control (IMC), where H the DM influence matrix, ψ_d the turbulence disturbance ψ_r the residual disturbance, G the wavefront sensor (WFS) matrix, ν the measurement noise, s the sensor signal and z^{-1} the unit-sample delay. The internal model $\hat{G}\hat{H}$ is used to determine the disturbance signal s_d which is the input to the controller W that computes the control signal u.

Pursuit (MP) algorithms. This paper derives a MP variant to solve for sparse matrices providing a sparse predictor or controller to compensate for atmospheric turbulence.

The following sections are structured as follows. Section 2 introduces the AO problem and derives a least squares problem for the optimal controller coefficients. Section 3 derives the sparse solution to the least squares problem on the basis of Matching Pursuits. Section 4 provides simulation results for a number of various AO configurations.

2 Adaptive Optics optimal prediction and control

Figure 1 illustrates the AO control problem, where $\psi_d(k) \in \mathbb{R}^{n_\psi}$ a vector containing phase-distortions of the observed light-beam at n_ψ locations in the cross-section of the beam at discrete time instant kT_s, where T_s the sampling time. The objective is to compensate this phase-distortion by means of a deformable mirror (DM) resulting in a residual phase distortion

$$\psi_r(k) := \psi_d(k) + Hu(k), \tag{1}$$

where $u(k) \in \mathbb{R}^{n_u}$ the control signal and $H \in \mathbb{R}^{n_\psi \times n_u}$ the influence-matrix of the DM. We assume that ψ_r does not have piston term. Because minimizing the 2-norm of the piston removed residual phase approximately involves maximization of the Strehl ratio, the control objective will be to minimize

$$J(k) := \mathrm{E}[\psi_r(k)^T \psi_r(k)]. \tag{2}$$

A sensor signal proportional to the (piston corrected) residual phase is obtained by

$$s_r(k) := G\psi_r(k) + \nu(k), \tag{3}$$

where $G \in \mathbb{R}^{n_s \times n_\psi}$ the wave-front sensor (WFS) matrix and $\nu(k) \in \mathbb{R}^{n_s}$ measurement noise, which is zero-mean white noise with covariance $\mathrm{E}\left[\nu(k)\nu(k)^T\right] = \sigma_\nu^2 I$ and independent of $\psi_d(\ell)$ for all k, ℓ.

In this paper we assume an AO system in which the sampling time is chosen such that the wavefront sensing, all data-transportation and the computation of the control signal are performed within one sampling time period. This will result in one sample delay in the control loop.

As in [4] a reference signal will be constructed by means of Internal Model Control (IMC)

$$s_d(k) := s_r(k) - \hat{G}\hat{H}u(k) \;=\; s_r(k) - GHu(k) \tag{4}$$

which is assumed to be perfect. It can be shown that $s_d(k) \;=\; G\psi_d(k) + \nu(k)$, which is only determined by the disturbance and the measurement noise. Furthermore, the control signal is determined by

$$u(k) := \sum_{i=1}^{p} W_i s_d(k - i) \tag{5}$$

where $W_i \in \mathbb{R}^{n_u \times n_s}$ and $p \geq 1$ a user chosen parameter determining the complexity of the controller. Note that in (5) $u(k)$ is only determined by delayed values of $s_d(k)$ because of the one sample delay in the closed loop. Also note that the controller (5) is inherently stable if the elements in the matrices W_i and the integer p are all bounded. Hence, the control problem is to determine $\{W_i\}_{i=1}^{p}$ such that (2) is minimized. However, ψ_r cannot be measured directly and thus (2) will be unknown. One approach is to model the correlation coefficients of the turbulence a priori, but the correlation coefficients might be unknown in practice due to changing turbulence statistics.

Therefore, in this paper we will minimize the mean square value of the reconstructed phase $\hat{\psi}_r(k) := G^\dagger s_r(k)$, where $G^\dagger$ the pseudo-inverse of G. Then the cost function to be minimized is

$$J' := \mathrm{tr}\Big(G^\dagger G \mathrm{E}\left[\psi_r(k)\psi_r(k)^T\right](G^\dagger G)^T + \sigma_v^2 G^\dagger G^{\dagger T}\Big) \tag{6}$$

where the white noise property of v has been used. Assuming the turbulence to be wide-sense stationary over at least N samples, (6) is approximated by

$$J'_N = \frac{1}{N}\mathrm{tr}(C + AWB^T)(C + AWB^T)^T \tag{7}$$

where

$$A := G^\dagger GH, \quad B := \begin{bmatrix} s_d(p) & s_d(p+1) & \cdots & s_d(N-1) \\ s_d(p-1) & s_d(p) & \cdots & s_d(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ s_d(1) & s_d(2) & \cdots & s_d(N-p) \end{bmatrix}, \quad C := \begin{bmatrix} \hat{\psi}_d(p+1) & \hat{\psi}_d(p+2) & \cdots & \hat{\psi}_d(N) \end{bmatrix}.$$

The minimization of (7) is a least-squares problem, and a minimizing solution is given by

$$W = -A^\dagger C (B^\dagger)^T \tag{8}$$

which is unique if and only if A has full column rank and B has full row rank. This solution appears not to be sparse, even if G, H are sparse. When a sparsity structure of W is known a priori, then (7) can be minimized subject to this structure, which still is a least squares problem. However, usually the sparsity structure is not known, and one may impose a condition that W only has a constrained number of nonzero elements. It turns out that this problem is NP-hard [2].

3 Sparse Prediction by Matching Pursuits

Matching pursuits (MP's) are typically used to determine (suboptimal) sparse solutions of linear equations with more unknowns than equations. Various type of versions have been proposed, c.f. [1] for an overview. In this paper the basic MP algorithm as proposed in [9] will be used because of its simplicity. The MP methods in literature are solving for sparse vectors, rather than matrices. Indeed, it is possible to vectorize W by stacking its column vectors on top of eachother resulting in the vector $w = \mathrm{vec}(W) \in \mathbb{R}^{pn_s}$, then (7) can be rewritten as

$$J'_N = \frac{1}{N}\Big(\mathrm{vec}(C) + (B \otimes A)w\Big)^T \Big(\mathrm{vec}(C) + (B \otimes A)w\Big) \tag{9}$$

where $A \otimes B$ denotes the Kronecker product of A and B. Then the basic MP algorithm is given by:

Algorithm 1: Basic Matching Pursuit (BMP) [9]

1. Initialize $n = 0$, $r_n = \mathrm{vec}(C)$, $w'_n = 0$.
2. Let $A' := (B \otimes A)D^{-1}$ where D a diagonal matrix which elements are the norms of the corresponding columns of $B \otimes A$, hence A' as unit-norm column vectors.

3. Until stopping criterion is met, $n = n + 1$ and

$$\delta_n = -A'^T r_{n-1} \tag{10}$$

$$i_n = \arg_i \max |\delta_{n,i}| \tag{11}$$

$$w'_{n,i_n} = w'_{n-1,i_n} + \delta_{n,i_n} \tag{12}$$

$$r_n = r_{n-1} + A'_{i_n} \delta_{n,i_n} \tag{13}$$

4. Output r_n and $w_n = D^{-1} w'_n$.

where $\delta_{n,i}$ denotes the ith element of δ_n and A_{i_n} the i_n^{th} column of A.

The stopping criterion can be a maximum on the number of nonzero elements in w'_n, the norm or the change in the norm of r_n has a below some threshold parameter or the number of iterations has exceeded a maximum.

In [3] estimates are given for the rate of convergence. Note, that Algorithm 1 can be implemented more efficiently by computing δ_n recursively, which will be explained in more detail in the next subsection. Despite the simplicity of the MP algorithm, construction of $C \otimes B \in \mathbb{R}^{pn_s n_\psi \times N n_\psi}$ will result in a huge matrix resulting in high computational complexity and memory usage. This drawback is solved by the MP algorithm given in the next subsection for solving a sparse matrix rather than a vector

The extension for matrix unknowns can be made by noting that in each iteration of the MP algorithm a scalar coefficient is determined which minimizes the residual cost most. For matrix unknowns the search for the most minimizing scalar now ranges not only over the rows, but also over the columns. Then, the matrix MP algorithm is given by

Algorithm 2: Matrix Matching Pursuit (MMP)

1. Initialize $n = 0$, $R_n = C$, $W'_n = 0$.
2. Let $A' := A D_A^{-1}$ and $B' := B D_B^{-1}$, where D_A and D_B the diagonal matrices which diagonal elements are the norms of the corresponding columns of A and B respectively.
3. Until stopping criterion is met, $n = n + 1$ and

$$\Delta_n = -A'^T R_{n-1} B' \tag{14}$$

$$(i_n, j_n) = \arg_{i,j} \max |\Delta_{n,i,j}| \tag{15}$$

$$W'_{n,i_n,j_n} = W'_{n-1,i_n,j_n} + \Delta_{n,i_n,j_n} \tag{16}$$

$$R_n = R_{n-1} + A'_{i_n} \Delta_{n,i_n,j_n} B'^T_{j_n} \tag{17}$$

4. Output R_n and $W_n = D_A^{-1} W'_n D_B^{-1}$.

where $\Delta_{n,i,j}$ denotes the i, j^{th} element of Δ_n.

As in the vector based MP algorithm, the inner product (14) can be computed recursively using the update rule for the residual (17), which yields

$$\Delta_{n+1} = \Delta_n - E_{i_n} \Delta_{n,i_n,j_n} F_{j_n}^T \tag{18}$$

where $E := A^T A$, $F := B^T B$ and further E_{i_n} and F_{j_n} denote the i_n^{th} and the j_n^{th} column of E and F respectively.

Because the number of samples N may be large, such that the number of rows of B will be large, a QR data compression step is usually recommended. Let the QR-decomposition of $[B \ C^T]$ be given by

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} = \begin{bmatrix} B \ C^T \end{bmatrix} \tag{19}$$

where R_{11} is upper triangular, then (7) can be written as

$$J'_N = \frac{1}{N} \text{tr} \left((R_{12}^T + A W R_{11}^T)(R_{12}^T + A W R_{11}^T)^T + R_{22}^T R_{22} \right) \tag{20}$$

Table 1. Relative performance and complexity of static inversion, least squares and the matrix MP algorithm.

A :	P		P	/R
	R .	# C	R . E	# C
SI:	0.23	1296	0.23	2592
LS1:	0.22	1296	0.22	2592
LS30:	0.097	38880	0.097	77760
MP1:	0.22	256	0.46	256
MP30:	0.10	1024	0.36	1024
SI+MP1:	-	-	0.22	2848
SI+MP30:	-	-	0.17	3291
	P	/C	P	/R /C
	R . E	# C	R . E	# C
SI:	0.23	2592	0.23	2592
LS1:	0.22	2592	0.22	2592
LS30:	0.097	38880	0.11	77760
MP1:	0.22	256	0.48	256
MP30:	0.12	1024	0.51	1024
SI+MP1:	-	-	0.22	2848
SI+MP30:	-	-	0.16	3291

and thus B and C can be replaced by R_{11} and R_{12}^T respectively. Note, that this QR decomposition can also be evaluated recursively using new measurements $s_d(N + 1)$ and $\hat{\psi}_d(N + 1)$.

4 Simulation results

Simulations of the Matrix Matching Pursuit algorithm have been performed with von Kármán frozen-flow turbulence with Fried parameter $r_0 = 0.16$m and an outer diameter of $L_0 = 1.00$m. The wind velocity was 10m/s in a constant direction. The telescope aperture is square of dimension 1.00×1.00m and the piston mode has been removed from the turbulence. The sampling rate is $f_s = 660$Hz. To generate a time sequence of the turbulence (60 sec.'s, the first 20 sec.'s for identification of W and the remaining 40 sec.'s for performance evaluation), an AR model of 100 taps has been computed on the basis of the correlation coefficients resulting from the von Kármán frozen-flow model.

The DM consists of 36 actuators in a 6×6 square grid. Each actuator has a Gaussian shaped influence function given by e^{-d^2/σ^2} where d the distance from the actuator position and $\sigma = 0.15$. The WFS measures spatial gradients in two directions at the same positions as the actuators.

To analyze the influence of the DM and the WFS on the results experiments with different choices of the DM influence matrix H and the WFS matrix G are performed:

1. Prediction: $H = I$ and $G = I$;
2. Prediction and reconstruction: $H = I$ and G the WFS matrix mapping the phases into slopes;
3. Prediction and control: H the influence matrix of the DM, $G = I$;
4. Prediction, reconstruction and control: H the influence matrix of the DM and G the WFS matrix.

In the experiments it was further assumed that there is no measurement noise, i.e., $\sigma_v = 0$. For each experiment, various methods are used to determine W:

1. SI: Static inversion, i.e., $W = -H^\dagger G^\dagger$;
2. LS1: The least squares solution (8) with $p = 1$ taps;
3. LS30: The least squares solution (8) with $p = 30$ taps;
4. MP1: Matching pursuit algorithm (Algorithm 2) with maximum number of $p = 1$ taps;
5. MP30: Matching pursuit algorithm (Algorithm 2) with maximum number of $p = 30$ taps.

For the cases where $G \neq I$, i.e., for reconstruction, also the following algorithms have been performed:

6. SI+MP1: Same as MP1 but the input to the filter W is the reconstructed phase $\hat{\psi}_d(k) = G^\dagger s_d(k)$;
7. SI+MP30: Same as SI+MP1 but with a maximum number of $p = 30$ taps.

Table 1 gives the relative error, i.e., the mean squared error divided by the mean squared value of the turbulence for the various methods as well as the number of filter coefficients. The stopping criterion for the MP algorithms was a bound on the number of coefficients as given in the #Coefficients column.

It is observed that in the prediction problem the phase can be predicted with approximately the same accuracy with the matching pursuit algorithms but with much fewer number of coefficients. Also increasing the number of taps from 1 to about 30 improves the prediction, i.e., the error is reduced by a factor 2. The same conclusions can be drawn for the case of prediction and control, which means that the inversion of the influence matrix H, for Gaussian shaped influence functions with limited spatial support, does not significantly reduce the performance by imposing sparsity constraints. This is different when gradients are measured, and the phase need to be reconstructed by the feedforward controller W, which show a significant loss of performance for MP1 and MP30. Part of the performance, was restored by feeding the controller with the reconstructed phase as is done in SI+MP1 and SI+MP30.

5 Conclusions

The paper has shown how the optimal AO control problem can be reformulated in terms of a least squares problem and how sparse solutions can be determined by means of matching pursuit (MP) algorithms. For the matrix least squares problem considered in this paper, a matrix version of the basic MP algorithm has been derived, which is validated on by simulation with von Kármán turbulence. Simulations show that prediction of the wavefront phase can be done very efficiently but reconstruction, i.e., minimizing the wavefront phase on the basis of gradients significantly lowers the performance. Further extensions of the algorithm may be in the direction of adaptive and faster implementations, other filter structures and using orthogonal or conjugate gradient iterations for each MP iteration.

References

1. T. Blumensath and M.E. Davis. Gradient pursuits. *IEEE Trans. Signal Process.*, 56(6):2370–2382, 2008.
2. G. Davis. *Adaptive nonlinear approximations*. PhD thesis, New York Univ., New York, 1994.
3. R.A. DeVore and V.N. Temlyakov. Some remarks on greedy algorithms. *Advances in Computational Mathematics*, 5:173–187, 1996.
4. N. Doelman, R. Fraanje, I. Houtzager, and M. Verhaegen. Real-time optimal control for adaptive optics systems. *European Journal of Control*, 15(3-4):480–488, 2009.
5. B.L. Ellerbroek. Efficient computation of minimum variance wavefront reconstructors using sparse matrix techniques. *J. Opt. Soc. Am. A*, 19(9):1903–1816, 2002.
6. L. Gilles, C.R. Vogel, and B.L. Ellerbroek. Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction. *J. Opt. Soc. Am. A*, 19(9):1817–1822, 2002.
7. K. Hinnen, M. Verhaegen, and N. Doelman. Exploiting the spatiotemporal correlation in adaptive optics using data-driven H_2-optimal control. *J. Opt. Soc. Am. A*, 24(6):1714–1725, 2007.
8. M.B. Jorgensen and G.J.M. Aitken. Prediction of atmospherically induced wave-front degradations. *Optics Letters*, 17(7):466–468, 1992.
9. S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.
10. L.A. Poyneer, D.T. Gavel, and J.M. Brase. Fast wave-front reconstruction in large adaptive optics systems with use of the fourier transform. *J. Opt. Soc. Am. A*, 19(10):2100–2111, 2002.
11. L.A. Poyneer, B.A. Macintosh, and J.-P. Véran. Fourier transform wavefront control with adaptive prediction of the atmosphere. *J. Opt. Soc. Am. A*, 24(9):2645–2660, 2007.
12. M. Tallon, E. Thiébaut, and C. Béchet. A fractal iterative method for fast wavefront reconstruction for extremely large telescopes. In *Proc. Adaptive Optics: Analysis and Methods*, Vancouver, Canada, Jun. 2007.