Increased Disturbance Rejection in Magnetic Disk Drives by Acceleration Feedforward Control and Parameter Adaptation [1]

Matthew T. White and Masayoshi Tomizuka

mtwhite@cmlds6.ME.Berkeley.EDU and tomizuka@euler.Berkeley.EDU

Computer Mechanics Laboratory

Department of Mechanical Engineering

University of California, Berkeley, CA 94720-1740

Abstract

As the density of data on magnetic disk drives increases, so does the need for more precise position control of the read/write head, especially in the presence of internal and external disturbances. This is achieved by measuring the acceleration of the drive and feeding the sensor information forward to the actuator. By matching the electromechanical impedance between the disturbance and the position error, the feedforward controller can cancel the effects of the disturbance. Two techniques are presented for designing the feedforward controller. The first method is an infinite impulse response filter that is designed off-line, and the second is a finite impulse response filter that is adapted on-line using the filtered-x LMS algorithm. Both techniques are tested through shake table experiments, resulting in reductions of the position error signal between 50% and 95%.

[1] submitted to *Control Engineering Practice*

1 Introduction

Despite amazing performance improvements in the recent past, the disk drive industry continues to search for ways to make its products smaller, faster, and able to hold more data. Key elements in this search are increasing radial density, or the number of concentric data tracks stored on the disk surface, and decreasing access time, the time it takes to move from the current track to the desired track. This requires, among other things, more accurate positioning of the read/write head, often in the presence of significant disturbances. These disturbances can be in the form of shock or vibration, while the drive is seeking or following a track. Disk drives are being used in increasingly demanding environments. For example, laptop and handheld computers are becoming even more popular. New applications such as cellular phones are being investigated. With the advancement of automation in industry, even traditional desktop computers are subjected to the rigors of the manufacturing floor. All of these situations elevate the level of external disturbances to which the drives are subjected.

To compound this problem, many of the smaller form factors have abandoned the use of shock mounts that help to isolate the drives from such disturbances. Those drives that have retained the shock mounts are more susceptible to wind-up, the internal reaction force due to actuator motion, which is especially significant during seeks. To decrease seek time, drive engineers are pushing closer to time-optimal, or bang-bang, control. This type of control results in large reaction forces and excites modes of the drive dynamics that have been neglected in the past.

The technique described in this paper uses accelerometers to measure the motion of the drive, and then feeds this information forward to the actuator controller to coordinate the read/write head position with the desired track position.

A number of authors have suggested the use of accelerometers for disturbance rejection in the past. Aruga *et al.* (1990) employed them on a dual, linear actuator system for a 10.5 inch form factor drive. They were concerned with the effects on the track following of one head as the second head performed a seek. Some dynamics of the drive were included, and robustness to modeling parameters was considered analytically. An experimental seek result was shown.

Davies and Sidman (1993) developed a constant, single-parameter, acceleration feedforward controller with low-pass filtering on the accelerometer output. Again, an analysis of the robustness to parameter variations was presented, with experimental results for a seek test and an external impulse disturbance.

Kempf (1994) used an accelerometer on a compact disk player to control the focus length of the reading lens.

He considered the first bending mode of the disk with simple models for the motor and shock mounts. The filtered-x LMS (least mean squares) adaptation algorithm was applied to the controller parameters.

Although drive companies have included low-grade accelerometers as shock sensors for shutdown, until recently the cost of quality-signal accelerometers has prohibited their use for control. However, advancements in sensor technology have lowered prices to a level that makes including these accelerometers in a production drive feasible. Examples of such sensors were used in the experiments. Furthermore, the field of microsensors may soon provide low-cost accelerometers that, due to their small size, can be used in more sophisticated schemes.

Section 2 follows with a description of the disk drive system, the nominal feedback controller, and the experimental platform. Section 3 describes the fixed-parameter feedforward design and the results achieved using this technique. Then the adaptive feedforward control method is presented in Section 4 with the results of its implementation. Conclusions are given in Section 5, and a convergence proof for the filtered-x LMS algorithm is included in the Appendix.

2 Disk Drive System

The disk drive used for experimentation was a 5.25 inch form factor with 2000 tracks per inch and a disk spindle speed of 3600 rpm. (See Fig. 1 for a schematic of the drive.) The drive base was attached to a mounting frame with shock mounts. The drive contained a stack of 8 disks. One disk side was dedicated to servo information to generate the position error signal (PES) between the read/write head and the desired track. The rotary actuator was composed of a voice coil motor (VCM) and the bearing-supported actuator arm with the read/write head and its suspension system.

The drive frame was mounted on a rotary shaker system that provided the external disturbance. The drive was situated so that it rotated about the actuator axis. This configuration transmitted the maximum disturbance to the read/write head. As shown (exaggerated) in Fig. 2, the disturbance to the base caused a relative displacement between the data track and the position of the head. It was this displacement that acceleration feedforward control was designed to minimize.

Two linear accelerometers were mounted on the drive base to measure the tangential components of the acceleration, and were used to calculate the angular acceleration of the drive base. The accelerometer signal was sent through a low-pass filter with a -3 dB point of 1 kHz before it was received by the feedforward compensator.

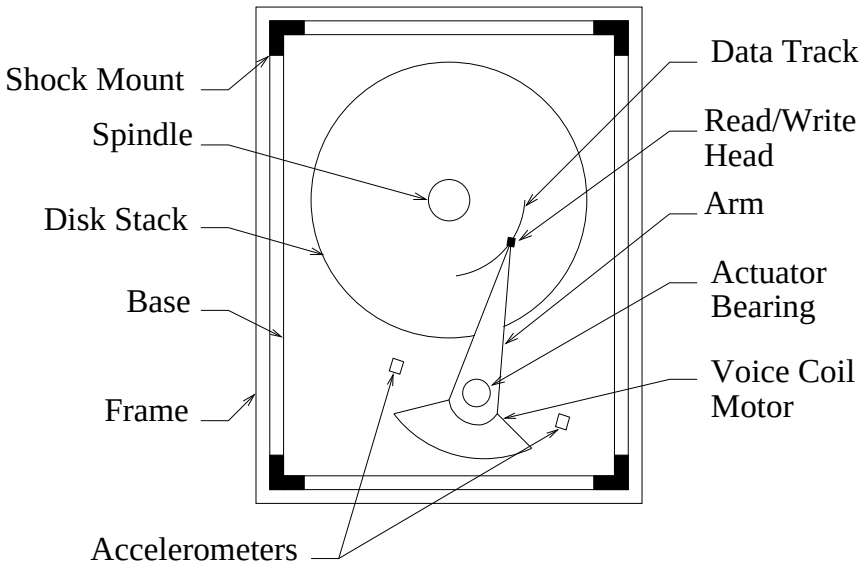

Figure 1: Disk Drive Components

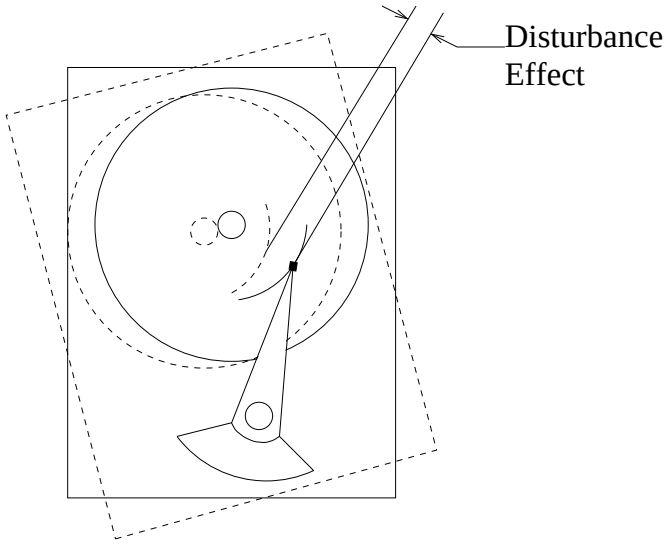

Figure 2: Effect of Disturbance on Track Misregistration

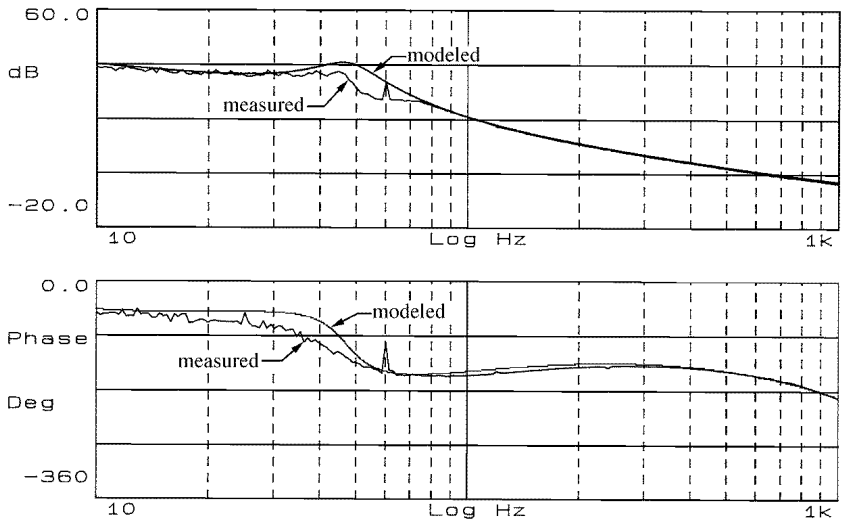

Figure 3: Comparison of the Measured and Modeled Open Loop Frequency Responses

The plant model included the shock mounts, base, and actuator. The damping and stiffness of the actuator bearing had prominent effects. Higher frequency actuator arm resonances were identified but did not significantly affect the experimental results. A time delay of approximately 150 μs contributed additional phase lag to the system. It was not possible to measure the plant frequency response directly, but the plant model was verified through the measurement of the open loop transfer function as described below.

The disk drive was equipped with a factory-installed analog feedback controller. Before experimentation, the analog controller was switched off and replaced with a digital controller. The digital feedback controller was a lead-lag filter that resulted in an open loop gain crossover frequency of 635 Hz. The feedback controller was designed in continuous time and then converted to discrete time using the bilinear transformation.

The open-loop frequency response was calculated from the known digital controller and the modeled plant, and compared to the measured open-loop frequency response as shown in Fig. 3. Frequency response measurements were obtained with a Hewlett-Packard dynamic signal analyzer.

2.1 General Control Structure

A block diagram of the plant and control loops is shown on the next page in Fig. 4, reproduced from (White and Tomizuka, 1995). The variables K_a, K_τ, and K_{pes} represent the gains of the transconductance amplifier, motor torque constant, and position error signal. The dynamics of the actuator between the motor torque τ and the head position θ_{head} are contained in the transfer function $G_{\tau h}$. The symbols $G_{\tau b}$ and G_{bt} represent

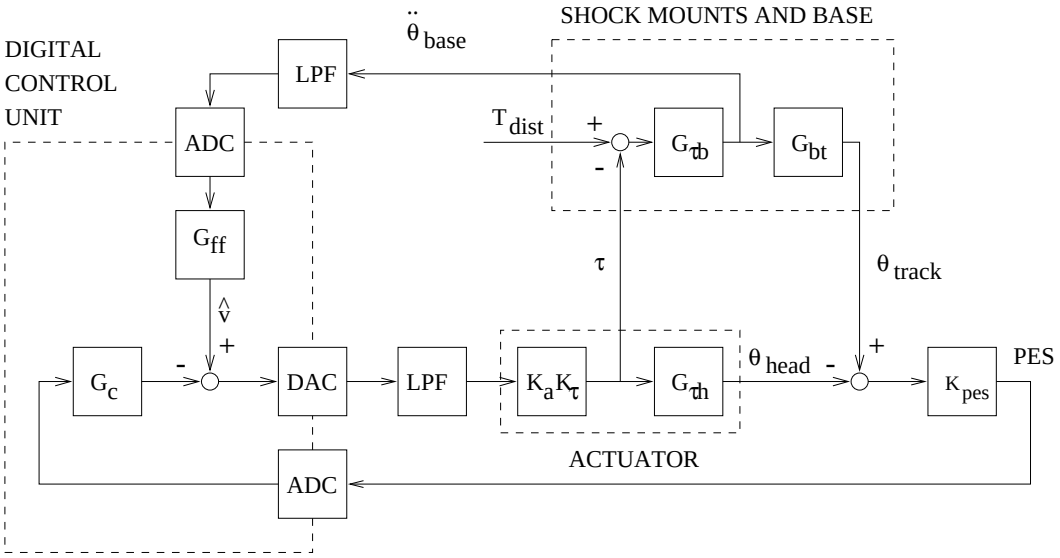

Figure 4: Block Diagram of Plant and Control System

the transfer functions from the disturbance to the base acceleration and from the base acceleration to the desired track position θ_{track}, respectively.

The nominal closed loop system consisted of the drive and the feedback controller, G_c. The drive sent the PES to the feedback controller, which calculated the required input to the actuator.

The measurement of the base acceleration was passed through a low-pass filter (LPF) to the feedforward controller G_{ff}. The feedforward controller calculated the additional control input $\hat{v}$ that was required due to the disturbance to the base T_{dist}, and this signal was added to the actuator input calculated by the feedback controller.

It should be noted that this technique was not pure feedforward. Because of the reaction torque of the actuator, there was a feedback component of the output of the feedforward controller. However, due to the ratio of the masses of the actuator and the base, the feedforward component dominated, and the feedback component did not significantly affect stability.

The feedforward controller was designed to match the electromechanical impedance between the base acceleration and the PES. By ignoring the torque feedback effects, the block diagram of Fig. 4 may be rearranged to appear as Fig. 5. The reference transfer function G_{ref} represents the dynamics to be canceled by G_{ff}, and is given by

$$G_{ref} = \frac{G_{bt}}{K_a K_\tau G_{\tau h}}. \tag{1}$$

The transfer function G_{out} represents additional dynamics before the PES,

$$G_{out} = \frac{K_a K_\tau K_{pes} G_{\tau h}}{1 - K_a K_\tau K_{pes} G_{\tau h} G_c}. \tag{2}$$

From Fig. 5, it can be seen how the proper choice of G_{ff} can cancel the disturbance effects. Note that v is a mathematical artifact, not a physical signal in the system, and therefore could not be measured.

2.2 Experimental Implementation

During testing, the disk drive described previously was connected to a digital control unit. (See Fig. 7 for a diagram of the experimental system.) The digital control unit included a TMS320C30 digital signal processor (DSP) that was used to calculate the feedback and feedforward control inputs for the drive. Also included were analog to digital converters (ADC) for the accelerometer signal and PES, digital to analog converters (DAC) for the control inputs to the drive and signals for analysis, and a LPF for smoothing the VCM command signal.

A personal computer was used to run an interactive system control program that allowed the user to specify the operating mode of the drive. A second computer ran the DSP debugger software with an XDS 510 emulator. An external hardware interrupt signaled the controller sampling period of 50 μs.

3 Fixed-Parameter Feedforward Controller

The first controller design method tested was a simple fixed-parameter, infinite impulse response (IIR) filter. The expected value of G_{ff} was determined through the use of the model information. This result was verified through experimental measurements. Recall that the signal v shown in Fig. 5 is not measurable. Because of this, the frequency response of G_{ff} had to be determined indirectly. To begin, $G_{wo} = G_{ref} G_{out}$ was measured with the feedforward controller turned off. Then the frequency response G_w from the base acceleration $\ddot{\theta}_{base}$ to the PES was measured with a nominal feedforward controller $G_{ff,nom}$ in place. Thus, G_{out} could be calculated as $(G_{wo} - G_w)/G_{ff,nom}$. Finally, the desired G_{ff} was calculated as G_{wo}/G_{out}. The modeled and measured results for G_{ff} are shown in Fig. 6. As with G_c, the transfer function G_{ff} was designed in continuous time and converted to discrete time using the bilinear transformation. A third-order model was found to be adequate for G_{ff}.

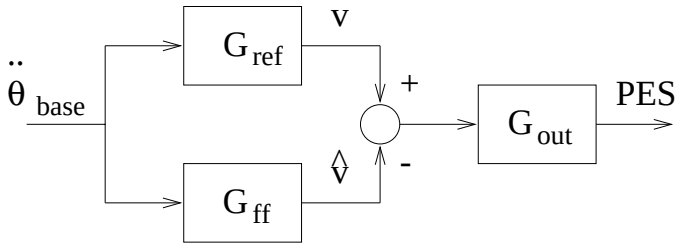

Figure 5: Impedance Matching for Feedforward Controller Design

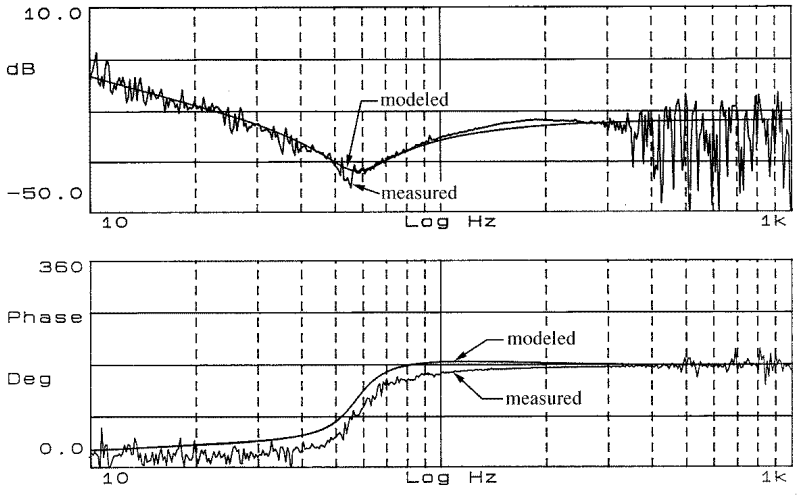

Figure 6: Comparison of the Measured and Modeled Feedforward Controller Frequency Responses

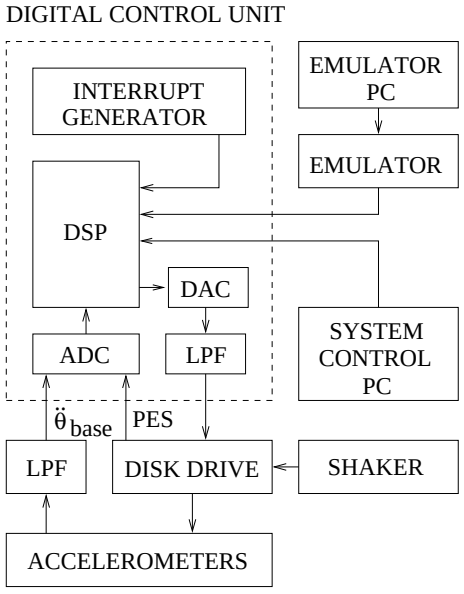

Figure 7: Schematic of Experimental System

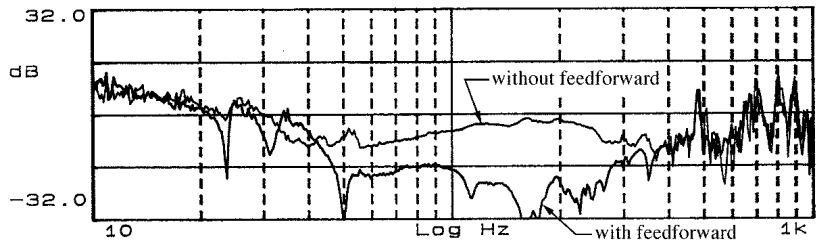

Figure 8: Magnitude of Experimental Frequency Response from Base Acceleration to PES without and with Fixed-Parameter IIR Feedforward Control

3.1 Fixed-Parameter Feedforward Controller Experimental Results

To test the efficacy of the feedforward control technique, comparisons of the position error signal with and without the acceleration information were made. The drive was put into track-following mode. A swept-sine was sent to the shaker system that generated an acceleration of approximately 1 G. This led to position errors on the order of 1 μm for the case with feedback control only, which is approximately the limit for reading and writing on a disk with 2000 tracks per inch. For higher density disks, this limit will obviously decrease.

Fig. 8 shows the frequency response between the acceleration of the drive and the PES. Between 40 Hz and 400 Hz, the PES was reduced from 50% to 90% using the fixed-parameter IIR feedforward controller. The response above and below this range was relatively unaffected. Note that in all cases, the measured PES also contained repeatable runout components due to eccentricities of the drive that were not induced by the disturbance.

Sample time traces are shown in Figs. 9-12 for frequencies of 60 Hz and 200 Hz. At 60 Hz, the PES was 1.36 μm without feedforward control and 0.55 μm with the fixed-parameter IIR feedforward controller. For 200 Hz, the error amplitudes were 1.42 μm and 0.39 μm without and with the feedforward controller, respectively. These correspond to reductions of 59% and 72%, which agree with the transfer function results.

4 Filtered-x LMS Adaptation Algorithm

Experimentation with the fixed-parameter design showed that performance of the feedforward controller was heavily dependent upon accurate modeling of the system. In mass-produced products such as disk drives, variations in system parameters between units are common. Additionally, the dynamics of a single drive are

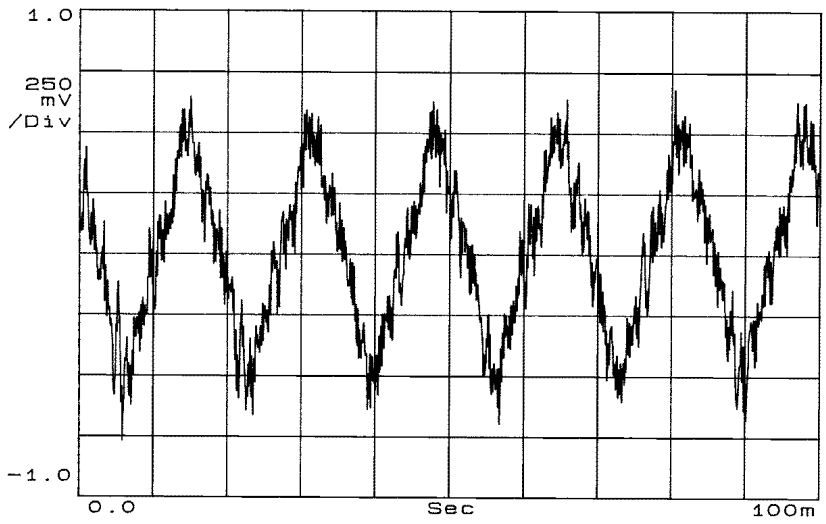

Figure 9: Experimental Time Trace without Feedforward Control at 60 Hz

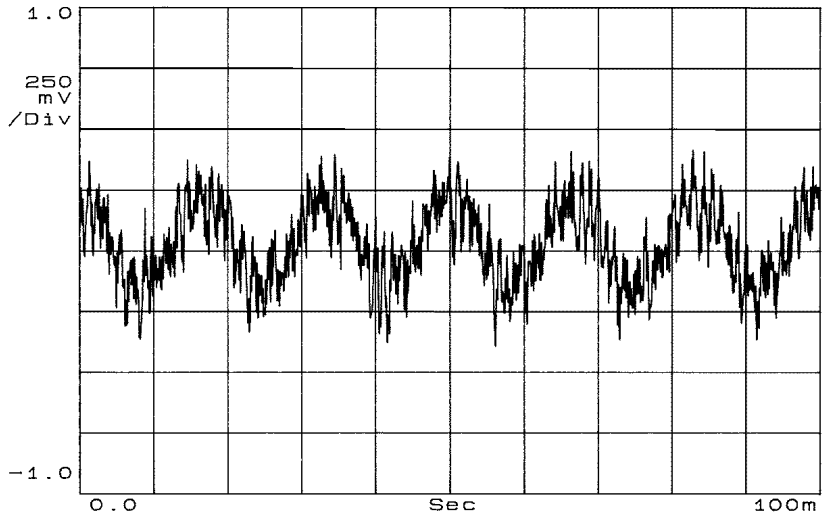

Figure 10: Experimental Time Trace with Fixed-Parameter IIR Feedforward Control at 60 Hz

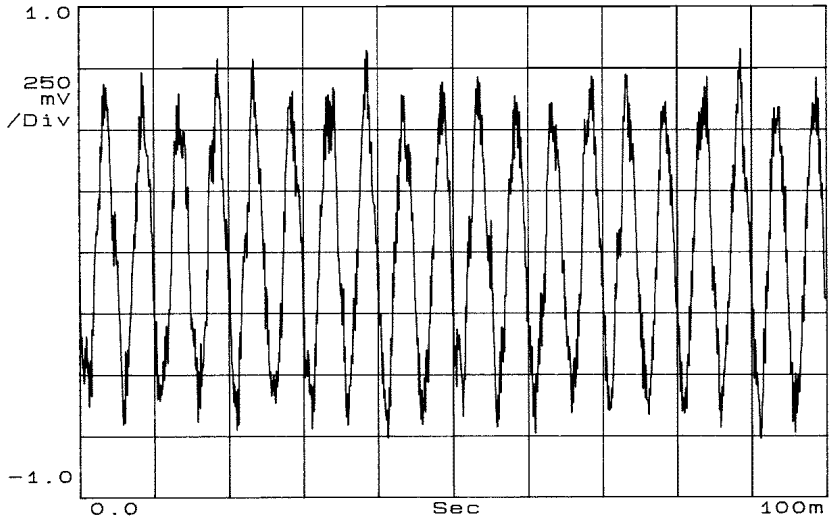

Figure 11: Experimental Time Trace without Feedforward Control at 200 Hz

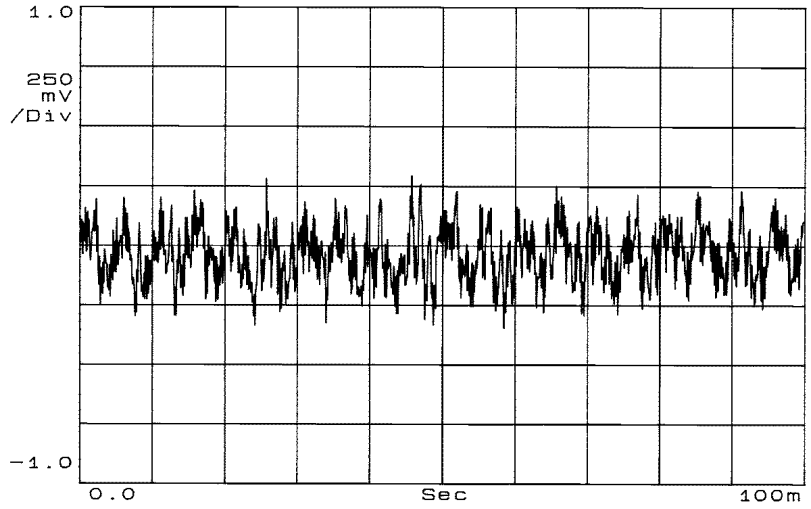

Figure 12: Experimental Time Trace with Fixed-parameter IIR Feedforward Control at 200 Hz

known to vary with age and use. These changes can be long term due to extended wear or short term due to such factors as thermal effects, although time scales are typically long enough that the system may be modeled as time-invariant. To combat the problems associated with parameter variations, adaptation on the feedforward controller was applied using a model reference version of the filtered-x LMS technique described in (Widrow and Stearns, 1985).

The filtered-x LMS algorithm is a modification of the least-mean square (LMS) algorithm, a simple gradient technique. The filtered-x LMS algorithm switches the typical order of the adaptive filter and the plant dynamics in an effort to reduce the effects of the plant noise on the adaptation. The filtered-x LMS algorithm has been used extensively in the field of active noise control (Fuller and von Flotow, 1995; Elliott *et al.*, 1987; Ren and Kumar, 1989).

4.1 Adaptive Feedforward Controller Algorithm Details

The adaptive feedforward control algorithm follows the same basic idea as the fixed-parameter algorithm. The goal is to match the transfer function G_{ref} of the physical system with the feedforward controller G_{ff}. For an FIR filter, G_{ff} takes the form

$$G_{ff}(k, q^{-1}) \quad = \quad w_0(k) + w_1(k)q^{-1} + \ldots + w_{N-1}(k)q^{-(N-1)} \tag{3}$$

where the parameters $w_i(k)$, or tap weights, at time k are adjusted on-line via the filtered-x LMS algorithm, and q^{-1} represents the one-step delay operator. The tap weights are adjusted in the direction that minimizes the expectation of the squared error with constant gain μ, or

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \frac{\partial}{\partial \mathbf{w}(k)} (E[\varepsilon^2(k)]) \tag{4}$$

where $\mathbf{w}(k)$ represents the vector of tap weights, E denotes the expectation, and $\varepsilon(k)$ is the PES. Calculating the gradient $\frac{\partial}{\partial \mathbf{w}}$ results in

$$w_i(k+1) = w_i(k) + 2\mu E[\varepsilon(k) G_{out}(q^{-1}) \ddot{\theta}_{base}(k-i)]. \tag{5}$$

This is the desired update law. However, the expectation is not known, nor is the transfer function G_{out}. Thus, some approximations must be made. The transfer function G_{out} is replaced with an *a priori* estimate $\hat{G}_{out}$, and the current value of $\varepsilon(k)\hat{G}_{out}(q^{-1})\ddot{\theta}_{base}(k)$ is used in place of the expected value. This results in an actual update law given by

$$w_i(k+1) = w_i(k) + 2\mu\varepsilon(k)\hat{G}_{out}(q^{-1})\ddot{\theta}_{base}(k-i). \tag{6}$$

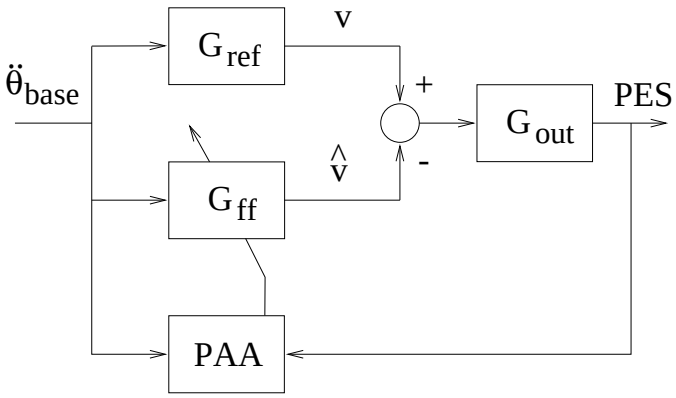

Figure 13: Impedance Matching for Feedforward Controller Design with Adaptive Feedforward Control

From equation 6 it can be seen that the filtered-x LMS algorithm is of the standard recursive adaptation form with $\varepsilon(k)$ as the modeling error and $\hat{G}_{out}(q^{-1})\ddot{\theta}_{base}(k-i)$ as the regressor.

4.2 Discussion of Adaptive Algorithm

Using an *a priori* estimate of G_{out} may seem to defeat the purpose of the adaptation. However, it has been shown experimentally that even a poor estimate of G_{out} is generally acceptable (Elliott *et al.*, 1987; Ren and Kumar, 1989; Widrow and Stearns, 1985). The condition on $\hat{G}_{out}$ for convergence of the filtered-x LMS algorithm is that $\hat{G}_{out}/G_{out}$ is strictly positive real (SPR). The convergence proof is presented in the Appendix.

It is possible to expand the algorithm so that the transfer function $\hat{G}_{out}$ is included in the adaptation, and to include more complicated methods for estimating the expectation. However, this increases the computation time and was not performed in the results presented here.

Fig. 13 shows how the adaptation algorithm may be added to the feedforward technique shown in Fig. 5, where the use of $\hat{G}_{out}$ is included implicitly in the parameter adaptation algorithm (PAA). The filtered-x LMS algorithm is appealing for this application because it is simple and has very few computations. This is due mainly to the fact that it uses a constant adaptation gain. Disk drive control systems typically have little space or time to run complicated processes. In addition, the filtered-x LMS algorithm is designed to decrease the effects of noise. Using an FIR filter is also helpful in decreasing noise sensitivity, and FIR filters are always stable. The stability of IIR filters is dependent upon their parameters, which often means that the algorithm must include extra calculations to check the stability of the IIR filter before it is implemented. This is not necessary with an FIR filter.

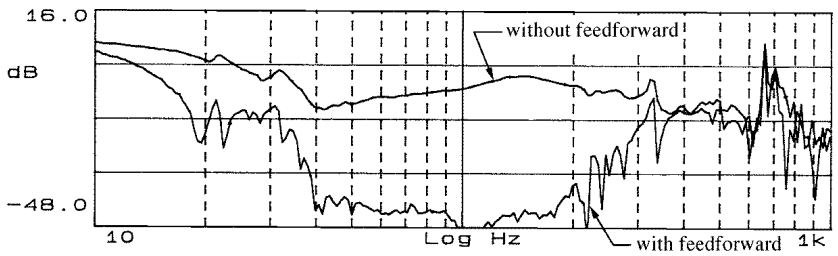

Figure 14: Magnitude of Experimental Frequency Response from Base Acceleration to PES without and with Adaptive FIR Feedforward Control

There are some drawbacks to this algorithm. Because the adaptation gain is constant, this technique converges more slowly than some more complicated algorithms and there is no inherent signal normalization. Also, FIR filters typically require more parameters than IIR filters to describe the same frequency response.

4.3 Adaptive Feedforward Controller Experimental Results

The same experimental tests were run for the filtered-x LMS algorithm that were run for the fixed-parameter design. Fig. 14 shows the magnitude of the frequency response from the disturbance to the PES with the adaptive feedforward controller compared to the case without feedforward. The PES is reduced between 55% and 95% over the range of 15 Hz to 400 Hz. Compare these results to those shown in Fig. 8. Note that the peformance has been improved with the addition of the adaptation, especially in the low frequency region.

Although the FIR filtered-x LMS method is computationally efficient for an adaptation algorithm, it still requires more computations than the fixed-parameter scheme. The computation time for the adaptive algorithm is dependent upon the number of parameters included in the FIR filter. Experimentation showed that the optimal number of tap weights was approximately 40, so the sampling time for the adaptive case was increased to 100 μs. This is still quite reasonable for disk drive applications, as most sample in the 5 kHZ to 10 kHz range. Using fewer parameters at 50 μs sampling, or more parameters at sampling times of 100 μs and greater, did not significantly improve the results. Widrow *et al.* (1976) have shown that increasing the number of parameters does not always improve performance. Sampling the fixed-parameter controller at 100 μs showed little change from the results presented previously.

PES time responses are shown in Figs. 15-18. Fig. 15 and Fig. 17 show the effects of the disturbance without feedforward control to be 3.6 μm at 40 Hz and 6.0 μm at 100 Hz. Figs. 16 and 18 show the PES values

for the same disturbances with the filtered-x LMS feedforward control in place. Initially, the FIR filter parameters are set equal to zero and there is no disturbance. After the disturbance begins, the feedforward parameters are allowed to adapt to their steady state values. After some transient behavior, the PES is reduced to approximately 1.3 μm at both 40 Hz and 100 Hz. This is an improvement of 63% for the 40 Hz disturbance and 78% for the 100 Hz disturbance. The value of the adaptation gain, $\mu = 1.5 \times 10^{-7}$, used in the experiments shown was a fairly moderate value. The rate of convergence can be improved by increasing the value of μ, but at the expense of robustness to the magnitude of the disturbance. There are some conditions on the magnitude of the adaptation gain for convergence of the algorithm, which are discussed in the Appendix.

5 Conclusions

The effects of disturbances on disk drive performance and a control method to reduce these effects were described. This technique used accelerometers to measure the acceleration of the drive and a feedforward filter to send a canceling signal to the drive actuator. Two different implementations of the feedforward filter were used. The first was a fixed-parameter IIR filter. The second was an FIR adaptive filter using the filtered-x LMS algorithm. These techniques were implemented experimentally using a digital signal processor to calculate the control, and a shake table to provide the disturbance. Results demonstrated that acceleration feedforward was effective in reducing the position error of the drive actuator due to disturbances. A comparison between the two feedforward controllers showed that the adaptive controller gave better results, and can adapt to variations in the drive parameters, but at the cost of increased computation.

6 Acknowledgments

This work was supported by the Computer Mechanics Laboratory of the Department of Mechanical Engineering at the University of California at Berkeley. Experimental equipment was provided by Hitachi, Ltd., Texas Instruments Incorporated, and PCB Piezotronics, Inc.

References

Aruga, K., Y. Mizoshita, M. Iwatsubo and T. Hatagami (1990). Acceleration feedforward control for head

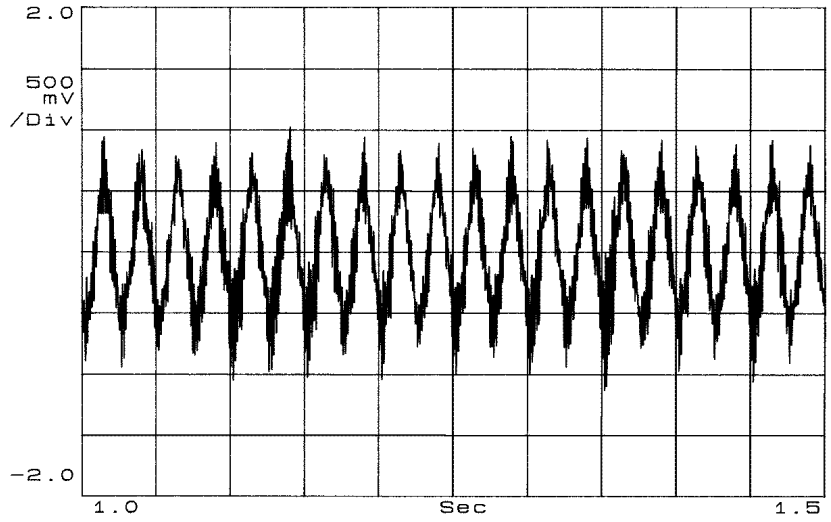

Figure 15: Experimental Time Trace without Feedforward Control at 40 Hz

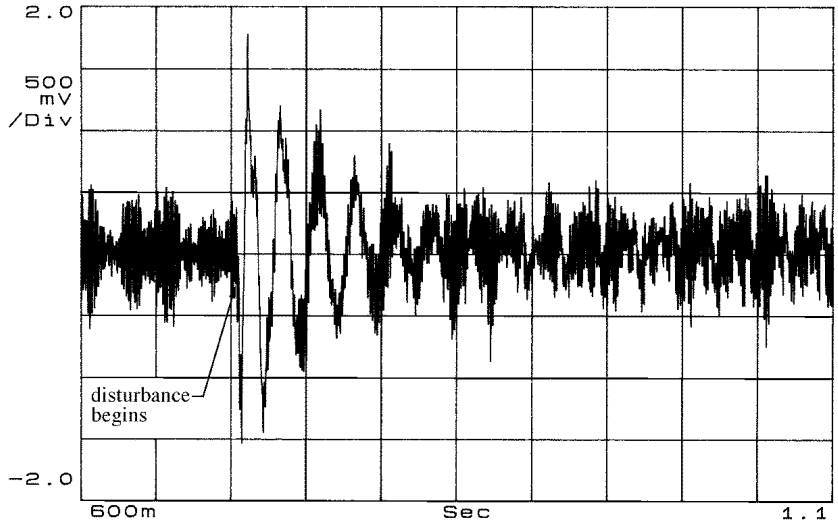

Figure 16: Experimental Time Trace with Adaptive FIR Feedforward Control at 40 Hz

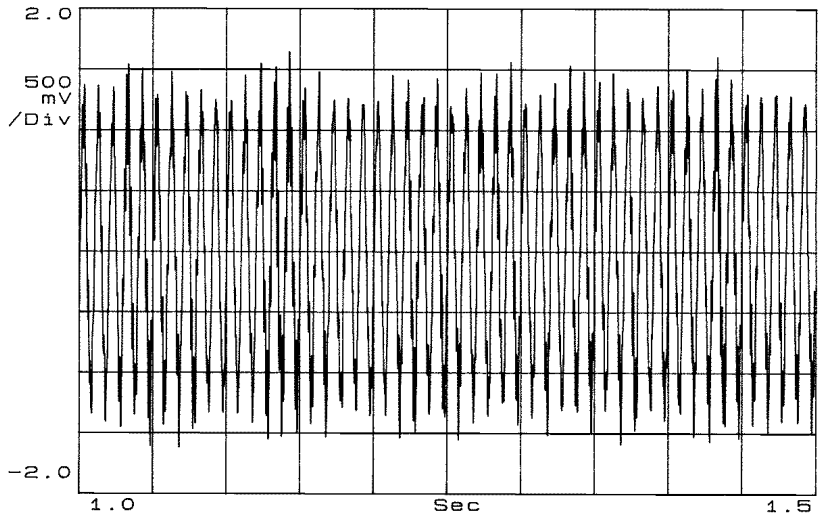

Figure 17: Experimental Time Trace without Feedforward Control at 100 Hz

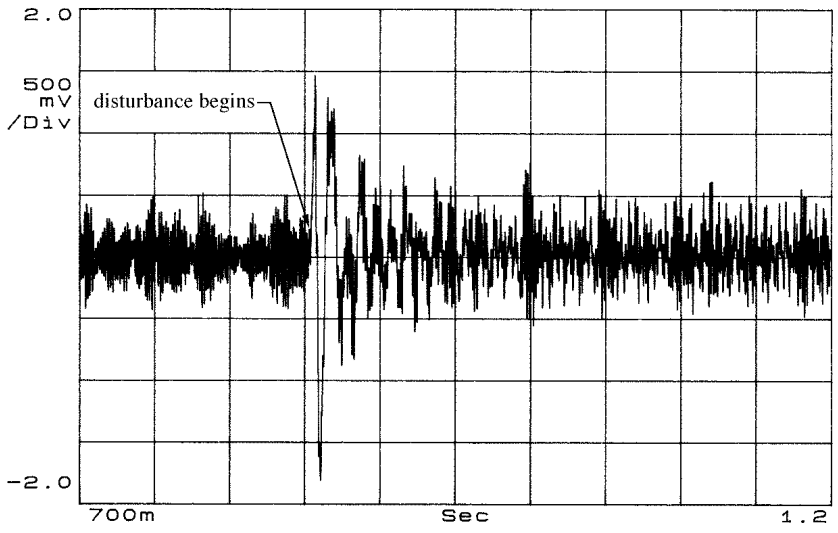

Figure 18: Experimental Time Trace with Adaptive FIR Feedforward Control at 100 Hz

positioning in magnetic disk drives. *JSME International Journal* **33**(1), 35–41. series III.

Benveniste, A. and G. Ruget (1982). A measure of the tracking capability of recursive stochastic algorithms with constant gains. *IEEE Transactions on Automatic Control* **AC-27**(3), 639–649.

Davies, D.B. and M.D. Sidman (1993). Active compensation of shock, vibration, and wind-up in disk drives. In: *Advances in Information Storage Systems*. Vol. 5. pp. 5–20. ASME Press. B. Bhushan, editor.

Elliott, S.J., I.M. Stothers and P.A. Nelson (1987). A multiple error LMS algorithm and its application to the active control of sound and vibration. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-35**(10), 1423–1434.

Fuller, C.R. and A.H. von Flotow (1995). Active control of sounds and vibration. *IEEE Control Systems* pp. 9–19.

Kempf, C.J. (1994). Design of Servo Systems for Disturbance Rejection and Applications to Disk File Storage Systems. PhD thesis. University of California at Berkeley.

Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control* **AC-22**(4), 551–575.

Reason, J. and W. Ren (1993). Estimating the optimal adaptation gain for the LMS algorithm. In: *Proceedings of the CDC*. San Antonio. pp. 1587–1588.

Ren, W. and P.R. Kumar (1989). Adaptive active noise control: Structures, algorithms and convergence analysis. In: *Proceedings of the 1989 International Conference on Noise Control Engineering*. Vol. 1. Newport Beach, CA. pp. 435–440. G.C. Maling, Jr., editor.

White, M.T. and M. Tomizuka (1995). Adaptive acceleration feedforward control of magnetic disk drives. Computer Mechanics Laboratory 7th Annual Sponsor's Meeting. University of California at Berkeley. January 30.

Widrow, B. and S.D. Stearns (1985). *Adaptive Signal Processing*. Prentice Hall. Englewood Cliffs, NJ.

Widrow, B., J.M. McCool, M.G. Larimore and C.R. Johnson, Jr. (1976). Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE* **64**(8), 1151–1162.

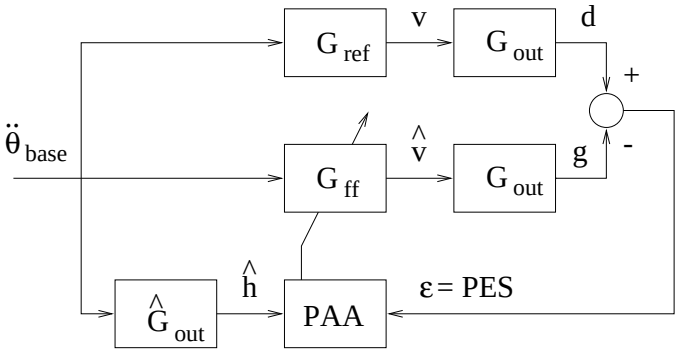

Figure 19: Filtered-x LMS Algorithm

Widrow, B., J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr. and R.C. Goodlin (1975). Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE* **63**(12), 1692–1716.

A Convergence Proof for the Filtered-x LMS Adaptive Algorithm

A.1 Algorithm and Notation

Consider the filtered-x LMS algorithm as shown in Fig. 19, which is essentially Fig. 13 with some slight manipulation for ease of analysis. Then $\hat{v}(k)$, the output of G_{ff} for input $\ddot{\theta}_{base}(k)$, may be written as

$$\hat{v}(k) = [w_0(k), ..., w_{N-1}(k)][\ddot{\theta}_{base}(k), ..., \ddot{\theta}_{base}(k-N+1)]^T \tag{7}$$

$$= \mathbf{w}^T(k)\mathbf{\Theta}(k) \tag{8}$$

and the output of G_{out} for input $\hat{v}(k)$ is $g(k) = G_{out}(q^{-1})\hat{v}(k)$.

A key assumption of the filtered-x LMS algorithm is that G_{out} and G_{ff} commute, at least approximately, which can be justified in the case of small adaptation gain. This implies that

$$g(k) = \left\{G_{out}(q^{-1})\mathbf{\Theta}^T(k)\right\}\mathbf{w}(k) = \mathbf{h}^T(k)\mathbf{w}(k). \tag{9}$$

Now define the error $\varepsilon(k)$ and the mean square error $\xi(k)$ as

$$\varepsilon(k) = d(k) - g(k) \tag{10}$$

$$\xi(k) = E[\varepsilon^2(k)] \tag{11}$$

where $d(k)$ is the desired value. Let $\varepsilon^2(k)$ serve as the estimate of its expectation $\xi(k)$. Thus the estimate of $\nabla(k)$, the gradient of $\xi(k)$, is

$$\hat{\nabla}(k) = \frac{\partial \varepsilon^2(k)}{\partial \mathbf{w}(k)} = -2\mathbf{h}(k)\varepsilon(k). \tag{12}$$

However, the vector $\mathbf{h}(k)$ is not available. Instead, use the modeling approximation $\hat{\mathbf{h}}(k) = \hat{G}_{out}(q^{-1})\mathbf{\Theta}(k)$. This results in

$$\hat{\nabla}(k) = -2\hat{\mathbf{h}}(k)\varepsilon(k). \tag{13}$$

The update law for the parameters of G_{ff} is made in the opposite direction of the approximated gradient

$$\mathbf{w}(k+1) \quad = \quad \mathbf{w}(k) + 2\mu\hat{\mathbf{h}}(k)\varepsilon(k). \tag{14}$$

A.2 Solution for the Optimal Filter Weights

Squaring equation 10 and substituting for $g(k)$ results in

$$\varepsilon^2(k) \quad = \quad d^2(k) + \mathbf{w}^T(k)\mathbf{h}(k)\mathbf{h}^T(k)\mathbf{w}(k) - 2d(k)\mathbf{h}^T(k)\mathbf{w}(k). \tag{15}$$

Assume that $\varepsilon(k)$ is stationary and that $d(k)$ and $\mathbf{\Theta}(k)$ are jointly stationary. Then for constant $\mathbf{w}$, taking the expectation leads to

$$E[\varepsilon^2(k)] \quad = \quad E[d^2(k)] + \mathbf{w}^T E[\mathbf{h}(k)\mathbf{h}^T(k)]\mathbf{w} - 2E[d(k)\mathbf{h}^T(k)]\mathbf{w}. \tag{16}$$

Define $\mathbf{R} = E[\mathbf{h}(k)\mathbf{h}^T(k)]$ and $\mathbf{p} = E[d(k)\mathbf{h}(k)]$. Note that $\mathbf{R}$ is positive semi-definite, so the eigenvalues of $\mathbf{R}$ are real and non-negative. Typically, $\mathbf{R}$ will be positive definite, in which case the eigenvalues are positive. Substituting for $\mathbf{R}$ and $\mathbf{p}$, equation 16 becomes

$$\xi = E[\varepsilon^2(k)] = E[d^2(k)] + \mathbf{w}^T\mathbf{R}\mathbf{w} - 2\mathbf{p}^T\mathbf{w} \tag{17}$$

and the gradient can be calculated as

$$\nabla = \frac{\partial \xi}{\partial \mathbf{w}} = 2\mathbf{R}\mathbf{w} - 2\mathbf{p}. \tag{18}$$

Note that due to the stationarity assumptions, $\mathbf{R}$, $\mathbf{p}$, ξ, and ∇ are all constant with respect to k. Setting $\nabla = \mathbf{0}$ and solving for the optimal weights $\mathbf{w} = \mathbf{w}^*$ results in

$$\mathbf{w}^* \quad = \quad \mathbf{R}^{-1}\mathbf{p} \tag{19}$$

assuming that $\mathbf{R}$ is positive definite, and therefore nonsingular. Substituting these values into equation 17 gives the minimum mean square error

$$\xi_{min} \quad = \quad E[d^2(k)] - \mathbf{p}^T\mathbf{R}^{-1}\mathbf{p}. \tag{20}$$

A.3 Solution for the Converged Filter Weights

Define $\hat{\mathbf{R}} = E[\hat{\mathbf{h}}(k)\mathbf{h}^T(k)]$ and $\hat{\mathbf{p}} = E[d(k)\hat{\mathbf{h}}(k)]$. Assume that

$$F(q^{-1}) = \frac{\hat{G}_{out}(q^{-1})}{G_{out}(q^{-1})} \tag{21}$$

is strictly positive real (SPR), then

$$\mathrm{Re}(F(e^{-j\omega})) > 0 \qquad \forall\, \omega \tag{22}$$

and $\hat{\mathbf{R}}$ is positive semi-definite. To see this, note that

$$\hat{\mathbf{R}} \quad = \quad E[\{F(q^{-1})\mathbf{h}(k)\}\mathbf{h}^T(k)]. \tag{23}$$

Consider any $\mathbf{z} \in \mathcal{R}^N$, the set of real N vectors, then

$$\mathbf{z}^T\hat{\mathbf{R}}\mathbf{z} \quad = \quad \mathbf{z}^T E[\{F(q^{-1})\mathbf{h}(k)\}\mathbf{h}^T(k)]\mathbf{z} \tag{24}$$

$$= \quad E[\tilde{\phi}(k)\phi(k)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(e^{-j\omega})\Phi_{\phi\phi}(\omega)\, d\omega \tag{25}$$

where $\phi(k) = \mathbf{h}^T(k)\mathbf{z}$ and $\tilde{\phi}(k) = \mathbf{z}^T\{F(q^{-1})\mathbf{h}(k)\}$ are scalars, and $\Phi_{\phi\phi}(\omega)$ is the spectral density of $\phi(k)$. Since $\Phi_{\phi\phi}(\omega)$ and $E[\tilde{\phi}(k)\phi(k)]$ are real

$$\mathbf{z}^T\hat{\mathbf{R}}\mathbf{z} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Re}(F(e^{-j\omega}))\Phi_{\phi\phi}(\omega)\, d\omega \tag{26}$$

but $\mathrm{Re}(F(e^{-j\omega})) > 0$, and $\Phi_{\phi\phi}(\omega) \geq 0$, therefore $\mathbf{z}^T\hat{\mathbf{R}}\mathbf{z} \geq 0$. $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ are constant with respect to k, and typically $\hat{\mathbf{R}}$ will be positive definite. Assuming that $\hat{\mathbf{R}}$ is positive definite, define

$$\hat{\mathbf{w}} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}. \tag{27}$$

A.4 Filtered-x LMS Convergence Proof: Constant Adaptation Gain

The convergence proof for the filtered-x LMS algorithm presented in this section follows the proof for the LMS algorithm given in (Widrow and Stearns, 1985). To analyze the convergence of $\mathbf{w}(k)$, take the expectation of equation 14

$$E[\mathbf{w}(k+1)] \quad = \quad E[\mathbf{w}(k)] + 2\mu E[\hat{\mathbf{h}}(k)\varepsilon(k)] \tag{28}$$

$$= \quad E[\mathbf{w}(k)] + 2\mu E[\hat{\mathbf{h}}(k)d(k) - \hat{\mathbf{h}}(k)\mathbf{h}^T(k)\mathbf{w}(k)]. \tag{29}$$

If the common, simplifying assumption that $\mathbf{\Theta}(k)$ and $\mathbf{w}(k)$ are uncorrelated is made (Widrow *et al.*, 1975), then

$$E[\hat{\mathbf{h}}(k)\mathbf{h}^T(k)\mathbf{w}(k)] = \hat{\mathbf{R}}E[\mathbf{w}(k)]. \tag{30}$$

Substituting $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ into equation 29,

$$E[\mathbf{w}(k+1)] \quad = \quad (\mathbf{I} - 2\mu\hat{\mathbf{R}})E[\mathbf{w}(k)] + 2\mu\hat{\mathbf{R}}\hat{\mathbf{w}} \tag{31}$$

where $\mathbf{I}$ is the identity matrix. Let $\hat{\mathbf{Q}}$ be the similarity transformation matrix such that

$$\hat{\mathbf{Q}}^{-1}\hat{\mathbf{R}}\hat{\mathbf{Q}} = \hat{\mathbf{J}} \tag{32}$$

where $\hat{\mathbf{J}}$ is the Jordan form of $\hat{\mathbf{R}}$. Now, transform equation 31 to the principal coordinate system, first by the translation $\mathbf{w}(k) = \mathbf{u}(k) + \hat{\mathbf{w}}$,

$$E[\mathbf{u}(k+1)] = (\mathbf{I} - 2\mu\hat{\mathbf{R}})E[\mathbf{u}(k)] \tag{33}$$

and then by the rotation $\mathbf{u}(k) = \hat{\mathbf{Q}}\mathbf{u}'(k)$,

$$E[\mathbf{u}'(k+1)] = (\mathbf{I} - 2\mu\hat{\mathbf{J}})E[\mathbf{u}'(k)]. \tag{34}$$

Since the eigenvalues λ_i of $\hat{\mathbf{J}}$ are the diagonal elements, the eigenvalues of $(\mathbf{I} - 2\mu\hat{\mathbf{J}})$ also lie on the diagonal. The eigenvalues of $\hat{\mathbf{R}}$, and therefore $\hat{\mathbf{J}}$, are positive. Thus, the system described by equation 34 is stable if

$$|1 - 2\mu\lambda_i| < 1 \tag{35}$$

for $i = 1, ..., N$ or if

$$0 < \mu < \frac{1}{\lambda_{max}} \tag{36}$$

where λ_{max} is the maximum eigenvalue of $\hat{\mathbf{J}}$ and $\hat{\mathbf{R}}$. Note that

$$\lambda_{max} \leq \mathrm{tr}[\hat{\mathbf{J}}] = \mathrm{tr}[\hat{\mathbf{R}}] \tag{37}$$

thus, convergence of the mean is assured if

$$0 < \mu < \frac{1}{\mathrm{tr}[\hat{\mathbf{R}}]} \tag{38}$$

since this implies

$$\lim_{k \to \infty} E[\mathbf{u}'(k)] = \lim_{k \to \infty} E[\mathbf{w}(k) - \hat{\mathbf{w}}] = \mathbf{0}. \tag{39}$$

Since $\ddot{\theta}_{base}(k)$ is assumed to be stationary,

$$\mathrm{tr}[\hat{\mathbf{R}}] = N \times E[\{\hat{G}_{out}(q^{-1})\ddot{\theta}_{base}(k)\}\{G_{out}(q^{-1})\ddot{\theta}_{base}(k)\}] \tag{40}$$

$$= N \times E[\hat{h}(k)h(k)] \tag{41}$$

or N times the filtered input signal power. This is a convenient, if possibly conservative, estimate of λ_{max}. Since the signal $h(k)$ is not available, use the average value of $\hat{h}^2(k)$ for the estimate of $E[\hat{h}(k)h(k)]$. Widrow and Stearns (1985) suggest a value of μ approximately equal to one-tenth of the maximum value given by equation 38. Note that due to the trade-offs between fast convergence and sensitivity to noise, the maximum value of μ is rarely the optimal value (Benveniste and Ruget, 1982; Reason and Ren, 1993).

Using $\hat{G}_{out}$ in place of G_{out} for calculating the bound on μ limits the magnitude of the allowable modeling error in $\hat{G}_{out}$. Considering equations 38 and 40, it is evident that there is an inverse relationship between the accuracy of the estimate of the bound on μ and the size of the $\hat{G}_p$ magnitude error. To set μ close to its maximum value, an accurate estimate of the magnitude of G_p is required. However, as stated above, it is typical to use a value of μ significantly lower than the maximum. This relaxes the requirement on the estimate of the magnitude of G_p.

The other requirement on the estimate $\hat{G}_p$ is the SPR condition stated previously. The SPR condition limits the phase error in $\hat{G}_p$, since it is equivalent to the condition

$$-90^\circ < \angle(G_p(e^{-j\omega})) - \angle(\hat{G}_p(e^{-j\omega})) < 90^\circ \qquad \forall\, \omega \tag{42}$$

where $\angle(G(e^{-j\omega}))$ denotes the phase angle of $G(e^{-j\omega})$. Thus, there is a requirement for the magnitude estimate of G_p and a requirement for the phase estimate of G_p to achieve convergence of the mean. The SPR condition of equation 22 is not terribly restrictive, and if μ is chosen to be near one-tenth of its maximum, neither is the requirement on the $\hat{G}_p$ magnitude error.

Convergence of the filtered-x LMS algorithm may also be shown using the ODE method introduced by Ljung in (Ljung, 1977). This proof was omitted due to space limitations. Both the ODE method and the technique of Widrow and Stearns show convergence of the filtered-x LMS algorithm tap weights to $\hat{\mathbf{w}}$. The ODE result is stronger, with convergence with probability one compared to convergence of the mean for the Widrow and Stearns technique. However, the ODE method requires, among other conditions, that the adaptation gain $\gamma(k) \to 0$ as $k \to \infty$, which the Widrow and Stearns technique does not.

Ljung suggests that it is reasonable to assume that the behavior of the adaptive algorithm for small, nonzero, adaptation gains is similar to the case with vanishing adaptation gain (Ljung, 1977). In comparison, recall that the Widrow and Stearns technique places a limit on the magnitude of the adaptation gain, and that the justification for the commutivity of G_{out} and G_{ff} is that the controller parameters are slowly varying. Since the assumption of a small adaptation gain is implicit in the use of the filtered-x LMS scheme, the results of the ODE analysis can be expected to predict the behavior of the filtered-x LMS algorithm for nonvanishing adaptation gains as well.

A.5 Analysis of the Mean Square Error

The mean square error with the converged values of the filtered-x LMS algorithm is equation 17 with

$$\mathbf{w} = \hat{\mathbf{w}} = \mathbf{w}^* + (\hat{\mathbf{w}} - \mathbf{w}^*) = \mathbf{w}^* + \mathbf{\Delta w} \tag{43}$$

or

$$\xi \;\; = \;\; \xi_{min} + \mathbf{\Delta w}^T \mathbf{R} \mathbf{\Delta w} + 2\left(\mathbf{w}^{*T}\mathbf{R} - \mathbf{p}^T\right)\mathbf{\Delta w}. \tag{44}$$

If $\hat{G}_{out} = G_{out}$, then $\mathbf{\Delta w} = \mathbf{0}$ and $\xi = \xi_{min}$. Ren and Kumar (1989) suggest a second situation where $\xi = \xi_{min}$. Let the transfer function F be defined as before and let f_i be such that

$$F(q^{-1}) = \sum_{i=0}^{\infty} f_i q^{-i}. \tag{45}$$

Then

$$\mathbf{R}\hat{\mathbf{w}} - \mathbf{p} = \sum_{i=1}^{\infty} \frac{f_i}{f_0} E[\mathbf{h}(k-i)\varepsilon(k, \hat{\mathbf{w}})] \tag{46}$$

since from equation 27

$$E[\hat{\mathbf{h}}(k)\mathbf{h}^T(k)\hat{\mathbf{w}}] = E[\hat{\mathbf{h}}(k)d(k)] \tag{47}$$

$$E[\hat{\mathbf{h}}(k)\varepsilon(k,\hat{\mathbf{w}})] = \mathbf{0} \tag{48}$$

$$F(q^{-1})E[\mathbf{h}(k)\varepsilon(k,\hat{\mathbf{w}})] = \mathbf{0}. \tag{49}$$

Substituting for F and splitting the summation results in

$$-f_0 E[\mathbf{h}(k)\varepsilon(k,\hat{\mathbf{w}})] = \sum_{i=1}^{\infty} f_i q^{-i} E[\mathbf{h}(k)\varepsilon(k,\hat{\mathbf{w}})] \tag{50}$$

$$E[\mathbf{h}(k)\mathbf{h}^T(k)]\hat{\mathbf{w}} - E[\mathbf{h}(k)d(k)] = \sum_{i=1}^{\infty} \frac{f_i}{f_0} E[\mathbf{h}(k-i)\varepsilon(k,\hat{\mathbf{w}})] \tag{51}$$

which is equation 46. When $\mathbf{w}^*$ achieves complete cancellation of the error component that is correlated with the input $\ddot{\theta}_{base}(k)$

$$E[\ddot{\theta}_{base}(s)\varepsilon(k,\mathbf{w}^*)] = 0 \tag{52}$$

for all $k \neq s$, which implies

$$E[\mathbf{h}(k-i)\varepsilon(k,\mathbf{w}^*)] = \mathbf{0} \tag{53}$$

for all $i \neq 0$, so $\mathbf{w}^*$ is a solution of equation 46 even if $\hat{G}_{out} \neq G_{out}$. Since $\mathbf{R}$ is assumed to be positive definite, equation 46 has a unique solution, and therefore $\hat{\mathbf{w}}$ must equal $\mathbf{w}^*$.