

Optimal Periodic Trajectories for Band-Limited Systems

A. J. Fleming and A. G. Wills

Abstract—The speed of an electromechanical scanner is limited by its first resonance frequency. To maximize scan speed, input signals are required that contain negligible frequency components near, or above the first resonance frequency. Such signals are usually obtained by low-pass filtering the desired scan trajectory. However, this introduces curvature and ripple into linear (constant velocity) scan regions. In this work, input signals are designed with guaranteed linear regions and minimal harmonic components above a chosen frequency. The proposed scanning trajectories are proven by simulation and experiment to induce less vibration than existing techniques.

Index Terms—Input signal design, mechatronics, motion control, motion planning, nanopositioning, periodic scanning, trajectory design, vibration control.

I. INTRODUCTION

MANY scientific and industrial machines contain mechanical scanners driven with periodic trajectories. For example, beam steering scanners, manufacturing robots, cam motion generators, and scanning probe microscopes [1]. In this work, without knowledge of system dynamics, periodic input signals are designed to maximize the speed and accuracy of band-limited scanners. We focus on designing inputs for scanning probe microscope nanopositioning stages, as reviewed in [2]–[4].

A. Background

The difficulties associated with high-speed scanners are illustrated in Fig. 1. Here, the system G representing a mechanical scanner is driven with a triangular signal r . In this example, the mechanical system G is a unity-gain second-order low-pass system with resonance frequency ω_n and damping ratio ξ_n , that is

$$G(s) = \frac{\omega_n^2}{s^2 + 2\omega_n\xi_n s + \omega_n^2}. \quad (1)$$

When G is excited by an input with significant frequency content at, or near, the resonance frequency, this content is amplified and appears as output ripple. For systems with settling time

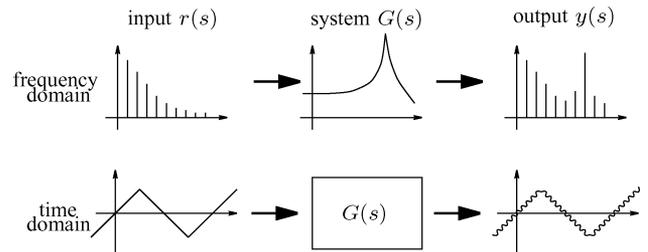


Fig. 1. Triangular scanning signal distorted by a typical mechanical system.

shorter than the scan period, resonance excitation appears after high-frequency events such as the peak of a sharp waveform. In addition to resonance excitation, frequency components of the input above the resonance frequency are attenuated and shifted in phase by 180 degrees.

If we quantify the tracking error e as the difference between input and output, i.e.,

$$e(t) = r(t) - y(t) \quad (2)$$

the error can be expressed in the Laplace domain as

$$e(s) = r(s)(1 - G(s)). \quad (3)$$

Thus, at frequencies where $r(s)$ is significant and $G(s)$ is not close to unity, the error is significant. There are three possible means for reducing error: inverting $G(s)$ (or otherwise filtering $r(s)$); reducing $G(s)$ to unity where $r(s)$ is significant; and reducing $r(s)$ to zero where $G(s)$ is not unity. The characteristics of each approach are discussed in the following.

1) *Inverting G* : Inversion of G is a commonly applied technique that can provide good performance if the plant model or its frequency response is known with high accuracy. When the input is periodic, inversion is easily accomplished by multiplying the Fourier coefficients of the input by the inverse frequency response.

The foremost problem with inversion is the lack of robustness to changes in plant dynamics, especially if the system is resonant [5]. Perfect inversion can also result in large amplitudes if the system response is small or zero at harmonics of the input [6]. Large signal amplitudes can cause actuator saturation and exacerbate amplitude dependent nonlinearity such as hysteresis.

The main attraction of inversion based control is its simplicity and ease of implementation, particularly in high-speed applications. With consideration of plant uncertainty, a significant improvement in imaging speed was achieved in [7] and [8]. Another inversion-based technique [6] avoids large amplitudes by trading off tracking performance for reduced input energy. A

Manuscript received December 03, 2007; revised April 28, 2007. Manuscript received in final form May 29, 2008. First published January 13, 2009; current version published April 24, 2009. Recommended by Associate Editor S. Devasia. This work was supported in part by the Australian Research Council (DP0666620) and by the Centre for Complex Dynamic Systems and Control. Experiments were conducted at the Laboratory for Dynamics and Control of Nanosystems, University of Newcastle.

The authors are with the School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: andrew.fleming@newcastle.edu.au; adrian.wills@newcastle.edu.au).

Digital Object Identifier 10.1109/TCST.2008.2001375

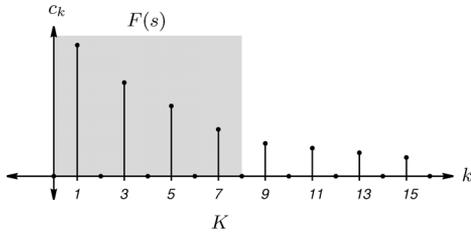


Fig. 2. Fourier coefficients of a triangle wave c_k filtered by $F(j\omega)$. Harmonics k from 9 onwards are removed.

related work [9] generates optimal output trajectories with minimal input energy and was successfully applied to a scanning tunneling microscope scanner.

Iterative inversion is a more elaborate technique that requires a sensor, but overcomes many limitations of linear inversion and can provide excellent performance when no exogenous disturbance is present [10]. Iterative techniques however, require time to converge, can generate large input signals, and require digital signal processing hardware.

Compared to feedback control, it is difficult or impossible to use feedforward compensation for accurate inversion of nonlinearity such as hysteresis. There is also no immunity to exogenous disturbance, offset, and gain drift.

2) *Controlling G* : Controlling G is a popular method for linearizing electromechanical systems at low-frequencies. Proportional-integral (PI) controllers, with and without notch filters for gain-margin improvement are commonly used, for a review see [2]–[4]. If sufficient sensor bandwidth is available, feedback control can also be used to damp mechanical resonance. For this purpose, positive position feedback (PPF) control and variants are straightforward to implement and perform well [11]–[13]. The major disadvantages of feedback are: the addition of sensor-induced noise, limited bandwidth, and tracking lag.

The addition of a feedforward controller can significantly improve the bandwidth and tracking lag of feedback systems without compromising stability or induced noise [14], [15]. However, due to the nature of feedforward control, immunity to hysteresis and disturbance is not improved and performance robustness can be reduced [16].

If only attenuation of mechanical resonance is required, the technique of shunt damping can be employed as an alternative to sensor-based feedback control [17], [18]. Shunt damping can provide attenuation of mechanical resonance without contributing sensor-induced noise.

3) *Reducing the Magnitude of $r(s)$* : Reducing the magnitude of $r(s)$ towards zero at frequencies near and above the resonance is a simple, practical and popular technique for minimizing induced vibration. The most obvious technique for reducing high-frequency content in $r(s)$ is to simply low-pass filter the signal. For periodic signals, this can be performed perfectly in the frequency domain by multiplying the Fourier coefficients of the reference signal with the filter magnitude specification, then applying the inverse Fourier transform.

The greatest disadvantage of low-pass filtering is the ripple introduced into linear (constant velocity) regions of the scan. As an explanation, consider the Fourier coefficients c_k of a periodic triangle wave shown in Fig. 2. If the filter $F(s)$ is designed to pass the first K harmonics and attenuate the remainder, the

filtered triangle wave $y(t)$ can be viewed as the original ideal trajectory $r(t)$, minus an error signal $e(t)$, i.e.,

$$y(t) = r(t) * F(t) = r(t) - e(t). \quad (4)$$

Conceptually, the error signal $e(t)$ is the *rippled part* of $y(t)$. In the frequency domain, $e(s)$ is comprised of the frequency components removed from $r(s)$ by $F(s)$, i.e.,

$$e(s) = r(s) - y(s) = r(s) - F(s)r(s) = r(s)(1 - F(s)). \quad (5)$$

More exactly, the Fourier coefficients of $e(t)$ are those of the original triangle above $k = K$. That is, if e_k is the Fourier coefficients of $e(t)$

$$e_k = \begin{cases} 0, & \text{when } -K < k < K \\ c_k. & \text{otherwise.} \end{cases} \quad (6)$$

The power P_e in the error signal $e(t)$ can be quantified using Parseval's equality

$$P_e = \sum_{k=-\infty}^{\infty} |c_k|^2. \quad (7)$$

As a consequence of this equality, the error becomes larger as signal bandwidth is reduced. This contradicts the original goal of low-pass filtering, to reduce scan error. Furthermore, as the filter $F(s)$ becomes more efficient, i.e., provides faster roll-off and better attenuation, the error also increases.

To eliminate the ripple and curvature introduced by frequency domain filtering, time-domain signal shaping was developed. This allows critical parts of the trajectory to be retained while corners and turnaround points are smoothed to reduce high-frequency content. The most straightforward signal shaping method is the minimum acceleration technique. This involves replacing the turning points of a trajectory with a smooth quadratic curve. Although this minimizes inertial force, it does not lead to optimal tracking performance. Minimum acceleration signals were used by Rost and colleagues to achieve SPM imaging rates of up to 200 frames/s [19].

Better performance than the minimum acceleration signal can be achieved by convolving the desired trajectory with a signal that minimizes induced vibration [20]–[22]. The foremost reported disadvantages of convolution techniques are: the significant filter length, sensitivity to parameter variation, and increased control signal magnitude [20].

A more recent technique that was introduced specifically for high-speed scanning is the shaped-triangle technique [23]. This signal is a triangle wave with a flat section, persisting for half the resonance period, at each signal apex. The shaped-triangle technique can provide excellent performance if the resonance frequency is exactly known and the mechanical system is second order. Unfortunately, the performance degrades if the resonance frequency is not exactly known or if the system order is greater than two.

B. Contribution of This Work

In this work, a new technique for designing periodic input trajectories is proposed. The method optimizes a desired trajectory based on frequency domain and/or time domain cost-functions. A key feature is that certain parts of the trajectory can be fixed.

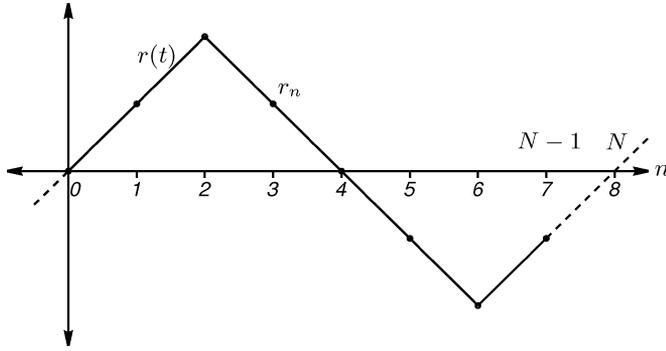


Fig. 3. Periodic signal $r(t)$ and its samples r_n .

For scanning applications, the proposed technique can be used to design input signals with perfectly linear (constant velocity) regions and minimal signal power above a chosen frequency. Comparison with other techniques shows a significant reduction in tracking error.

The proposed technique is most closely related to signal shaping and convolution techniques discussed in the previous subsection. The resulting optimal signals are similar in appearance to minimum acceleration signals but provide much improved performance. Unlike feedforward and feedback techniques, a parametric model or sensor is not required and the implementation is straightforward.

In the following section, a signal optimization scheme is proposed that allows parts of the trajectory to be fixed. In Section III, a range of cost functions are described that minimize properties such as acceleration and signal power. These can be used to generate signals with fixed and free regions that are optimal with respect to the chosen cost function. The frequency-weighted-power cost function is discussed in Section V as a technique for generating input signals for low-bandwidth positioning stages. The performance with respect to other techniques is evaluated by simulation in Section VI and experiment in Section VII. A summary of results and conclusions follow in Section VIII.

II. SIGNAL OPTIMIZATION

In this section, the signal optimization problem is defined and solved. The method begins with an ideal scanning trajectory, this is split into regions that are fixed, and regions that can be modified. The variable parts are then redesigned to minimize a quadratic cost function. In Section III, cost functions are described for various time and frequency domain objectives.

Consider the triangular waveform $r(t)$ plotted in Fig. 3. The samples of $r(t)$ are denoted $r_n = r(\Delta n)$, where Δ is the sampling interval, $n \in \{0, 1, 2, \dots, N-1\}$ and N is the number of samples per period. In the illustration, the sampling frequency $F_s = 1/\Delta$ is equal to eight times the triangle frequency F_T .

The samples of $r(t)$ over one period can be written in vector notation

$$r = \begin{bmatrix} r_1 \\ r_2 \\ r_2 \\ \vdots \\ r_{N-1} \end{bmatrix} = \begin{bmatrix} r(0) \\ r(\Delta) \\ r(2\Delta) \\ \vdots \\ r((N-1)\Delta) \end{bmatrix}. \quad (8)$$

This notation will be used throughout the remainder of this paper. That is, the vector of samples of one period of a waveform $x(t)$ will be denoted x , where $x \in \mathcal{R}^{N \times 1}$.

In this work, we seek a new signal y that is equal to r at an arbitrary set of sample indices S and free to vary elsewhere. The free part of the signal is varied to minimize the quadratic cost $y^T H y$. That is, we seek y that is the solution to

$$y = \arg \min_x x^T H x \quad (9)$$

subject to $x_k = r_k \quad k \in S$

where $x \in \mathcal{R}^{N \times 1}$ and $H \in \mathcal{R}^{N \times N}$. Problem (9) is equivalent to the linearly constrained convex quadratic optimization problem [24]

$$y = \arg \min_x x^T H x + 2f^T x \quad (10)$$

subject to $Ax = r(S)$

where A is the selection matrix representing S and $r(S)$ is a row vector containing the samples of r_n indexed by the values of S .

The solution to problem (10) can be stated in matrix form as [24]

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} -f \\ r(S) \end{bmatrix} \quad (11)$$

where λ are the Lagrange multipliers [24].

A solution to (11) may be obtained by

$$\begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} H & A^T \\ A & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -f \\ r(S) \end{bmatrix} \quad (12)$$

provided the above matrix inverse exists.

To this end, we notice that A has full row rank (since it is constructed as rows of the identity matrix) and $AA^T = I$, so that A^T forms a basis for the row space of A . Let Z be defined as the matrix formed from the rows of the identity matrix that are not present in A , e.g., if $N = 5$ and $S = \{2, 3\}$ then

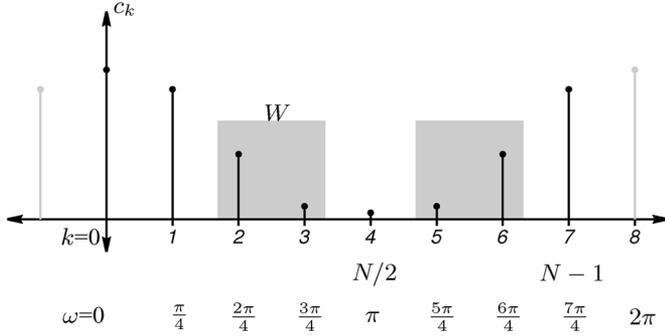
$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that $AZ^T = 0$ and that the rows of Z form a basis for the null space of A . Therefore, according to [24, pp. 231–237], the inverse in (12) exists and problem (10) has a unique minimizer if ZHZ^T is positive definite. While this condition may be difficult to prescribe, it is easily checked. Indeed, for all the examples presented here, this condition was satisfied.

III. FREQUENCY DOMAIN COST FUNCTIONS

The weighting matrix H can be chosen so that the quadratic cost $x^T H x$ represents a wide variety of frequency domain cost functions, for example, frequency-weighted-power. Techniques for selecting H follow.


 Fig. 4. Fourier components c_k of r .

A. Background: Discrete Fourier Series

The discrete Fourier series c_k of a periodic signal r_n is described by the analysis function [25]

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} r_n e^{-jn \frac{2\pi k}{N}}. \quad (13)$$

The synthesis function is [25]

$$r_n = \sum_{k=0}^{N-1} c_k e^{jn \frac{2\pi k}{N}} \quad (14)$$

where $\hat{\omega} = (2\pi k/N)$ is the normalized frequency, and $(2\pi/N)$ is the normalized fundamental frequency. The real frequency in Hertz is related to $\hat{\omega}$ by $f = (\hat{\omega}/2\pi\Delta)$. As an example, the discrete Fourier components of an eight sample signal are shown in Fig. 4.

The discrete Fourier coefficients of r can be written in matrix notation

$$c = \frac{1}{N} E r$$

where

$$c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{N-1} \end{bmatrix} \quad r = \begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_{N-1} \end{bmatrix}$$

and

$$E = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-j \frac{2\pi 1}{N}} & e^{-j 2 \frac{2\pi 1}{N}} & \dots & e^{-j(N-1) \frac{2\pi 1}{N}} \\ 1 & e^{-j \frac{2\pi 2}{N}} & e^{-j 2 \frac{2\pi 2}{N}} & \dots & e^{-j(N-1) \frac{2\pi 2}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j \frac{2\pi(N-1)}{N}} & e^{-j 2 \frac{2\pi(N-1)}{N}} & \dots & e^{-j(N-1) \frac{2\pi(N-1)}{N}} \end{bmatrix}. \quad (15)$$

B. Minimizing Signal Power

By Parseval's equality, the average power P_r of a discrete time signal r is

$$P_r = \frac{1}{N} \sum_{n=0}^{N-1} |r_n|^2 = \sum_{k=0}^{N-1} |c_k|^2 = \|c_k\|_2^2 \quad (16)$$

where the sequence $|c_k|^2$ for $k \in \{0, 1, 2, \dots, N-1\}$ is the distribution of power as a function of frequency, or the power spectral density. This can be written in vector form

$$\begin{aligned} P_r &= c^* c \\ &= \frac{1}{N^2} r^T E^* E r \end{aligned} \quad (17)$$

thus, referring to (9), minimum power is achieved when

$$H = \frac{1}{N^2} E^* E. \quad (18)$$

C. Minimizing Frequency Weighted Power

In Fig. 4, a frequency dependent weighting W is shown. The power resident in the shaded bandwidth can be calculated by summing only these components. W must be symmetric around π .

We wish to specify a cost function in (9) that represents power above a certain frequency or harmonic. This allows complete freedom in signal power up to the K th harmonic while imposing a power penalty at higher frequencies. The frequency weighted power P_r^W of r is

$$P_r^W = \frac{1}{N^2} r^T E^* W E r \quad (19)$$

where $W = \text{diag}(Q)$ and

$$Q = \begin{cases} 0, & k \in [0 \dots K] \\ 1, & k \in [K+1 \dots N-K-1] \\ 0, & k \in [N-K \dots N-1] \end{cases}$$

thus, referring to (9), minimum frequency-weighted-power is achieved when

$$H = \frac{1}{N^2} E^* W E. \quad (20)$$

It is worth mentioning that frequency-weighted-power signals are not band-limited. Rather, a frequency-weighted-power signal contains the least possible power above a certain frequency with the imposed time-domain constraints. If perfect band-limiting is desired, the Fourier coefficients above $k = K$ can be removed via the discrete Fourier transform and its inverse. The consequences of such filtering, namely the addition of ripple and curvature, are discussed in Section I-A3. The root-mean-square error as a result of filtering is also quantified in (7). As the frequency-weighted-power signal contains the least power above the K th harmonic, if the signal is then band-limited, the resulting signal has the least possible root-mean-square error (7). In applications where band-limiting is required, this is an important result. Restated, frequency-weighted-power signals suffer the least possible distortion when perfectly band-limited.

D. Minimizing Velocity and Acceleration

The use of frequency dependent weighting matrices in Section III-C can also be extended for weighting velocity or acceleration. The Fourier transform of the i th order derivative or integral of $x(t)$ is $(j\omega)^i X(j\omega)$, where i is positive for differentiation and negative for integration.

Rather than calculating the Fourier series of r in (15), we can calculate the Fourier series of its derivatives and integrals. The Fourier coefficients of the differentiated or integrated signal are

$$c = \frac{1}{N} D E r \quad (21)$$

where E and r are defined in (15), $D = \text{diag}(Q)$ and

$$Q = \begin{cases} (jk \frac{F_s}{N})^i, & k \in [0 \cdots N/2] \\ (j(N-k) \frac{F_s}{N})^i, & k \in [N/2 + 1 \cdots N - 1]. \end{cases} \quad (22)$$

This can be simplified to

$$c = \left(\frac{F_s}{N} \right)^i \frac{1}{N} \tilde{D} E r \quad (23)$$

where $\tilde{D} = \text{diag}(\tilde{Q})$ and

$$\tilde{Q} = \begin{cases} (jk)^i, & k \in [0 \cdots N/2] \\ (j(N-k))^i, & k \in [N/2 + 1 \cdots N - 1]. \end{cases} \quad (24)$$

The average power P_r^i in the chosen i th derivative or integral of r is

$$P_r^i = \left(\frac{F_s}{N} \right)^{2i} \frac{1}{N^2} r^T E^* \tilde{D}^* \tilde{D} E r. \quad (25)$$

Thus, referring to (9), minimum velocity or acceleration is achieved when $i = 1$ or 2 respectively, and

$$H = \left(\frac{F_s}{N} \right)^{2i} \frac{1}{N^2} E^* \tilde{D}^* \tilde{D} E. \quad (26)$$

Analogous to Section III-C, we can also consider a frequency weighted version of P_r^i

$$P_r^{i,W} = \left(\frac{F_s}{N} \right)^{2i} \frac{1}{N^2} r^T E^* \tilde{D}^* W \tilde{D} E r. \quad (27)$$

Referring to (9), minimum frequency-weighted velocity or acceleration is achieved when

$$H = \left(\frac{F_s}{N} \right)^{2i} \frac{1}{N^2} E^* \tilde{D}^* W \tilde{D} E. \quad (28)$$

E. Single-Sided Frequency Domain Calculations

Real valued signals with an even number of samples have a symmetric spectrum about the Nyquist frequency. The problem size of (9) can be significantly reduced by considering only one half the spectrum. The signal power is simply twice the sum contained in each half spectrum. That is

$$P = 2P_{0:N/2} - P_{N/2} \quad (29)$$

where the additional $P_{N/2}$ term is due to the power at the Nyquist rate only occurring once. The error in neglecting this additional term becomes smaller as the number of samples increases. For large N it is sufficient to approximate

$$P = 2P_{0:N/2}. \quad (30)$$

Using this simplification, the E , D , \tilde{D} and W need only be computed for $k = 0$ to $N/2$.

IV. TIME DOMAIN COST FUNCTION

In addition to the frequency domain objectives discussed in the previous section, the quadratic cost in (9) can also represent a function of time. This is useful for incorporating finite-impulse response (FIR) weighting functions used in previous trajectory design techniques. The time domain approach is also numerically robust when specifying optimizations that include a weighting on signal derivatives, for example velocity and acceleration.

The time domain cost function is defined as the output power of an FIR filter whose input is y . That is, we seek to minimize

$$z_n = \frac{1}{N} \sum_{n=0}^{N-1} |B(q^{-1})y_n|^2 \quad (31)$$

where $B(q^{-1})$ is an FIR filter of order N_B and length $N_B + 1$.

In matrix form, $z_n = B(q^{-1})y_n$ can be written

$z = \mathbf{B}y$, where

$$z = \begin{bmatrix} z_{0+N_b} \\ z_{1+N_b} \\ z_{2+N_b} \\ \vdots \\ z_{N-1} \end{bmatrix}, y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & b_{N_B} & \cdots & b_1 & b_0 \end{bmatrix}. \quad (32)$$

The power in z is

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} |B(q^{-1})y_n|^2 &= \frac{1}{N} \|B(q^{-1})y_n\|_2^2 \\ &= \frac{1}{N} z^T z \\ &= \frac{1}{N} y^T \mathbf{B}^T \mathbf{B} y. \end{aligned} \quad (33)$$

Thus, referring to (9), the power in z is minimized when

$$H = \frac{1}{N} \mathbf{B}^T \mathbf{B} \quad (34)$$

where \mathbf{B} is the matrix of FIR filter coefficients described in (32).

A. Minimum Velocity

The discrete velocity of y_n is the first-order time derivative

$$\frac{dy_n}{dt} = \frac{y_n - y_{n-1}}{\Delta}. \quad (35)$$

Thus, the FIR filter that represents differentiation is

$$B(q^{-1}) = \frac{1}{\Delta} (1 - 1q^{-1}). \quad (36)$$

This filter can be used in the time-domain cost function (34) to penalize velocity. The filter coefficients are $b_0 = 1$ and $b_1 = -1$.

B. Minimum Acceleration

The discrete acceleration of y_n is the second-order time derivative

$$\begin{aligned} \frac{d^2 y_n}{dt^2} &= \frac{1}{\Delta} \left(\frac{dy_n}{dt} - \frac{dy_{n-1}}{dt} \right) \\ &= \frac{(y_n - y_{n-1}) - (y_{n-1} - y_{n-2})}{\Delta^2} \\ &= \frac{y_n - 2y_{n-1} + y_{n-2}}{\Delta^2}. \end{aligned} \quad (37)$$

Thus, the FIR filter that represents double differentiation is

$$B(q^{-1}) = \frac{1}{\Delta^2} (1 - 2q^{-1} + 1q^{-2}). \quad (38)$$

This filter can be used in the time-domain cost function (34) to penalize acceleration. The filter coefficients are $b_0 = 1$, $b_1 = -2$, and $b_2 = 1$.

C. Frequency Weighted Objectives

Analogous to the frequency weighted cost functions in Section III, time domain cost functions can also be subjected to frequency domain weightings, however, the process is less direct.

Frequency weighted power can be achieved by using the filter $B(q^{-1})$ to implement the desired frequency weighting. In this case, the quadratic cost H representing power at the output of the filter is described in (33) and (34). If the filter $B(q^{-1})$ has already been utilized, for example to specify velocity or acceleration, a frequency weighting can still be applied by generating a second filter $B_2(q^{-1})$, whose frequency response represents the desired weighting, and convolving the two, i.e.,

$$B(q^{-1}) = B_1(q^{-1}) \otimes B_2(q^{-1}) \quad (39)$$

where $B(q^{-1})$ is the filter used in the cost function (34), $B_1(q^{-1})$ is the filter used for example to specify velocity and $B_2(q^{-1})$ is the frequency weighting filter.

V. APPLICATION TO SCAN GENERATION

In periodic scanning applications, it is desirable to scan as quickly as possible without exciting mechanical resonance. In other words, an input signal is required that contains the least possible power at frequencies near and above the first mechanical resonance. This objective is satisfied by the frequency-weighted-power cost function described in Section III-C. The resulting trajectory contains the least possible power above a certain frequency while maintaining perfect scanning over a portion of the range.

For triangular and sawtooth scanning waveforms, the linear range is easily specified by a single parameter β . Referring to Fig. 5, the optimal trajectory y_k is equal to r_k when $r_k < |\beta|$, otherwise there is no restriction. Using the notation in Section II, the previous statement can be rewritten as $y(S) = r(S)$, where S is the set of sample indices for which $r_k < |\beta|$.

To specify the frequency weighting, it is convenient to stipulate the number of unrestricted low-frequency harmonics that may appear in the optimal signal. The spectrum of a triangular scanning signal is shown in Fig. 6. The frequency components of the optimal signal are unrestricted between dc and the K th harmonic. All harmonics greater than K are penalized equally.

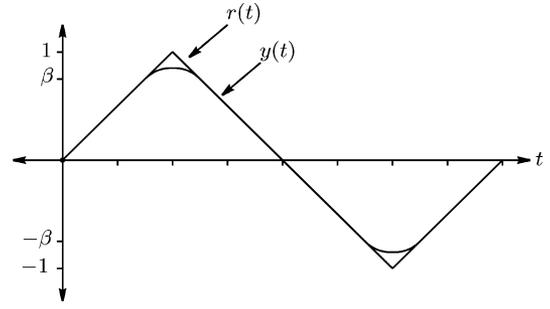


Fig. 5. Reference and optimal trajectory $r(t)$ and $y(t)$. The optimal signal is equal to $r(t)$ when $r(t) < |\beta|$, otherwise there is no restriction.

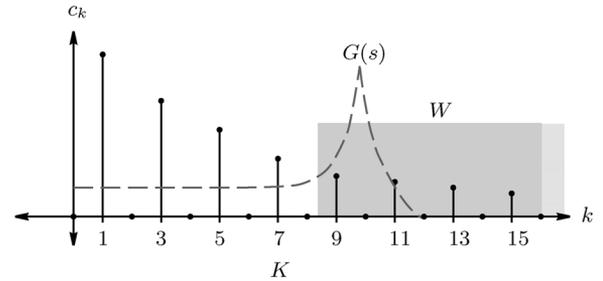


Fig. 6. Fourier components of a triangular scanning signal plotted against harmonic number k . The optimal signal is unrestricted in spectral content between dc and the K th harmonic. All harmonics greater than K are penalized to avoid excitation of the system G .

A MATLAB function that generates and simulates optimal scanning signals, named `generateTriangle`, is available by contacting the first author.

A. Choosing β and K

When using the frequency weighted power objective, frequency content above the cutoff is minimized by decreasing β and increasing K . If either parameter is fixed, the other can be varied to reduce scan error to an arbitrary value.

Assuming the allowable bandwidth is known, two possible scenarios arise when considering the choice of β and K , these are as follows.

- 1) The error and scan range are fixed. What is the maximum scan frequency? This is characteristic of most practical circumstances where scan range and precision are more highly valued than frequency. The scan frequency is simply reduced to a point where the number of in-bandwidth harmonics are sufficient to satisfy the error criterion.
- 2) The error and scan frequency are fixed. What is the maximum scan range β ? This case arises in high speed applications where scan range is sacrificed for increased frequency. Given the number of allowable harmonics, e.g., 3, β is reduced until the error is satisfactory. If the resulting scan range is impractically small, the scan frequency must be revised.

Both of these scenarios are easily resolved by plotting the free parameter versus error.

In general purpose applications where no fixed limit on frequency or scan range exists, some insight can be gained by plotting the high-frequency signal content versus the scan range β and number of harmonics K as shown in Figs. 7 and 8. Here,

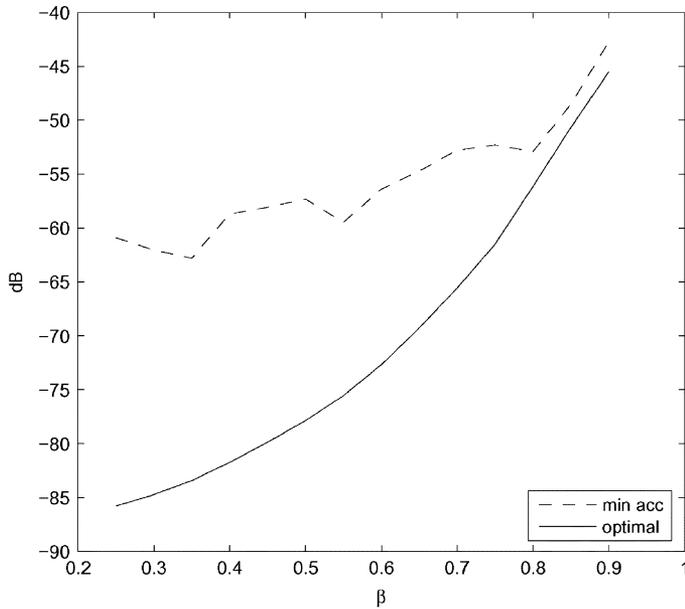


Fig. 7. Absolute sum of the first 50 out of bandwidth Fourier coefficients versus scan range β ($K = 9$).

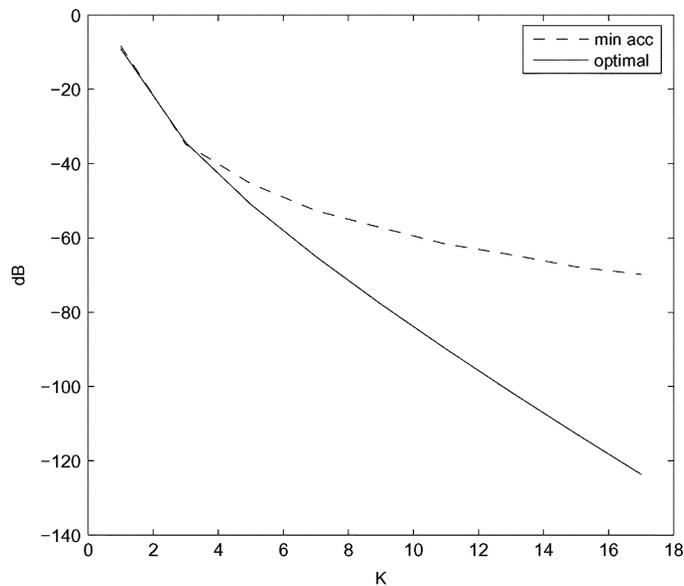


Fig. 8. Absolute sum of the first 50 out of bandwidth Fourier coefficients versus the number of included harmonics K ($\beta = 0.5$).

the high-frequency signal power is defined as the absolute sum of the first 50 harmonic components above K .

If scan range is valued highly, a good choice for β is 0.7, this provides approximately the maximum scan range before high-frequency content significantly increases. Beyond $\beta = 0.8$ there is little difference between the optimal, and minimum acceleration signals. If β is chosen fairly large ($\beta \geq 0.7$), the scan error must be minimized by including a large number of harmonics. For example, if the scan frequency is one twentieth the mechanical resonance frequency, K can be chosen up to 19. In Fig. 8 it is clear that $K = 19$ will provide a very high degree of performance.

If scan speed is highly valued, K must be small. If the scan speed is 10% the resonance frequency, K must be 9 or less.

The smallest reasonable value for K is 5, which allows only 3 sine waves in the optimal signal and scan speeds up to 20% the resonance frequency. In such cases, β must be severely reduced to minimize induced vibration. In Fig. 7, reducing β to 0.3 can provide excellent performance at ultra high speed.

The authors recommend the following two general purpose choices for β and K .

- $\beta = 0.7$ and $K = 9$. This provides good scan range, operation up to 10% of the mechanical resonance frequency and a reasonable minimum of induced vibration. Slower scan speeds with higher K improve performance.
- $\beta = 0.5$ and $K = 5$ or 7. This is more suitable for high performance scanning where scan frequency approaches 20% the resonance frequency. Vibration can be reduced by further reducing β to 0.4 or less.

B. Improving Feedback and Feedforward Controllers

1) *Feedback*: In addition to improving the performance of open-loop scanners, optimized input signals are also useful as reference commands for feedback control loops. As tracking control loops are typically limited in bandwidth to around one-tenth that of the open-loop system, the frequency content of reference commands must be strictly conserved if tracking error is to be kept low.

Further limitations arise in many electromechanical systems that exhibit nonlinearity such as hysteresis. In these systems, high controller loop-gain is required to attenuate tracking error. In integral control loops, significant loop-gain is only available one-decade below the closed-loop bandwidth. Thus, the system should only be driven by reference commands that contain frequency components significantly lower than the closed-loop bandwidth, which is typically only a fraction of the first resonance frequency. In such cases, an optimized reference trajectory can provide the best utilization of the small bandwidth available.

In more general circumstances, reference commands with lower high-frequency content relax the close-loop bandwidth requirement. This, in turn, requires less controller gain, resulting in greater robustness and less feed-through of sensor noise to the regulated variable.

2) *Feedforward*: In systems using inversion based feedforward control, the choice of reference signal is critical. Wide bandwidth input signals have spectral components at frequencies where the inversion filter can be highly sensitive to modeling error [16]. Sensitivity to modeling error can be reduced if the reference signal has minimal harmonic content in the bandwidth where inversion is required [16]. The frequency-weighted-power signal, discussed in Section V, is such an input that contains minimum high-frequency power and can provide the greatest immunity to modeling error.

Frequency-weighted-power signals also minimize control signal magnitude by avoiding frequencies where the plant response is small. This is highly advantageous in iterative systems that achieve near perfect inversion [10]. If the internal reference signal contains frequency components at, or near, plant zeros, extremely large inputs are generated in compensation. Frequency-weighted-power signals that contain minimal high-frequency harmonics can greatly reduce this problem.

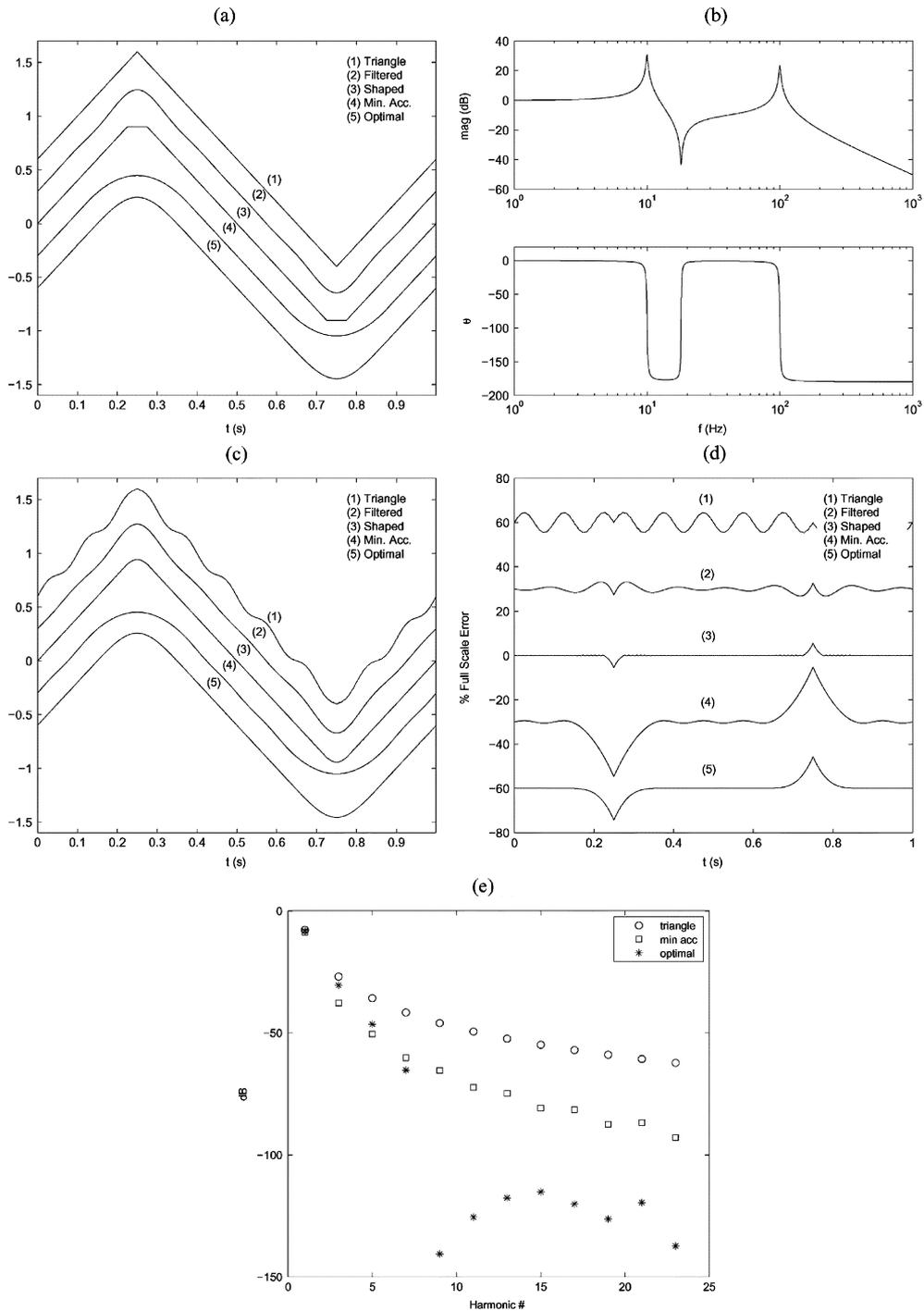


Fig. 9. Input signals under comparison are plotted in (a). Each signal is plotted with an offset for clarity. The example scanner model $G(s)$ (40), whose frequency response is plotted in (b), was excited with each signal. The resulting displacement and difference to an ideal triangle signal is plotted in (c) and (d). Plot (e) contains the first 25 Fourier coefficients of the triangle, minimum-acceleration and frequency-weighted power waveforms. (a) Input signals. (b) Frequency response. (c) Resulting displacement. (d) Error. (e) Fourier coefficients.

VI. COMPARISON TO OTHER TECHNIQUES

As discussed in the introduction, a number of techniques have been proposed for minimizing induced vibration in mechanical scanners. In this section, these techniques are compared to the frequency-weighted-power signal discussed in Section V.

A simple scanner model is considered with two resonances, one at 10 Hz and another at 100 Hz. The transfer function is

$$G(s) = \frac{0.7\omega_1^2}{s^2 + 2\omega_1\xi_1s + \omega_1^2} + \frac{0.3\omega_2^2}{s^2 + 2\omega_2\xi_2s + \omega_2^2} \quad (40)$$

where $\omega_1 = 2\pi 10$, $\omega_2 = 2\pi 100$, and $\xi_1 = \xi_2 = 0.01$. The frequency response of $G(s)$ is plotted in Fig. 9(b).

TABLE I

SIMULATED ROOT-MEAN-SQUARE ERROR BETWEEN THE OUTPUTS AND A TRIANGLE WAVE (CALCULATED IN THE TIME RANGE WHERE THE OPTIMAL SIGNAL IS LINEAR). TWO CASES ARE CONSIDERED, ONE WHERE THE RESONANCE FREQUENCY IS 10 HZ, AND ANOTHER WHERE THE RESONANCE FREQUENCY IS REDUCED BY 10% TO 9 HZ

Resonance Freq.	10 Hz	9 Hz
Triangle	3.1%	1.2%
Filtered-Triangle	0.53%	0.77%
Shaped-Triangle	0.062%	0.46%
Min. Acc.	0.39%	0.13%
Optimal	0.0075%	0.011%

It is desired to operate the scanner at one-tenth the resonance frequency, i.e., 1 Hz. The five input signals under consideration are as follows.

- 1) *Triangle Signal*: 1 Hz, unity amplitude triangle wave with a linear range of ± 1 .
- 2) *Filtered-Triangle*: Triangle signal, non-causally filtered by the minimum order Butterworth frequency response that achieves less than 3 dB ripple below 7 Hz and more than 80 dB attenuation at 9 Hz. The linear range is ± 0.75 .
- 3) *Shaped-Triangle*: Triangle signal with 0.05 s flat area at each apex as described in [23]. This signal provides excellent performance if the resonance frequency is known and the mechanical system is second order. The performance degrades if the resonance frequency is not exactly known or the system order is greater than 2. The linear range is ± 0.9 .
- 4) *Minimum Acceleration (Min. Acc.)*: Minimum acceleration trajectory with a linear scan range of $\pm 0.5(\beta = 0.5)$.
- 5) *Optimal*: Frequency-weighted-power signal with a linear scan range of $\pm 0.5(\beta = 0.5)$ and $K = 7$ as described in Section V.

The five input signals under consideration are plotted in Fig. 9(a). When applied to the example system $G(s)$, the resulting output and corresponding error are shown in Fig. 9(c) and (d). To summarize the results, root-mean-square errors are presented in Table I. The frequency-weighted-power signal is observed to outperform other techniques by between 8 and 400 times. In Table I, results from a second simulation where the resonance frequency is reduced by 10% are also reported. While the shaped-triangle signal performs well in the nominal simulation, it is not robust to changes in the resonance frequency. This is due to its dependency on the scanner resonance. In contrast, the frequency-weighted-power signal is not model-based and performs well when the resonance frequency is not known or prone to variation. Further insight can be gained by considering Fig. 9(e) where the Fourier coefficients of the triangle, minimum-acceleration and frequency-weighted-power signal are plotted. Clearly, after the seventh harmonic, the frequency-weighted-power coefficients drop to extremely small magnitudes. Hence, variations in system dynamics after the seventh harmonic have little effect on the tracking error.

VII. EXPERIMENTAL APPLICATION

Two-axis micro- and nano-positioning stages are used extensively in many forms of scanning probe microscope. They typically comprise a pair of piezoelectric actuators, mechanical dis-

placement amplifiers, and a flexure guided sample platform. Although these configurations can achieve high precision with millimeter range motion, the internal displacement amplifiers, large piezoelectric stacks, and platform mass contribute to a low mechanical resonance frequency. An example of such a stage is the Physik Intrumente P-734. This stage has a range of $100 \mu\text{m}$ but a resonance frequency of only 420 Hz. The frequency response of a single axis is plotted in Fig. 10(b). The unity gain bandwidth extends from dc to around 140 Hz where a phase and magnitude shift of 5 degrees and 1 dB exists. Above this frequency, the phase and magnitude response degrade rapidly. To achieve accurate scanning in open-loop, the input signal spectrum should be retained to within 140 Hz.

Without using model-based inversion, the fastest practical scan speed for the platform under consideration is around 20 Hz. In this case, the 3rd, 5th, and 7th harmonics occur at 60, 100, and 140 Hz. An optimal signal can be designed to achieve high scan range with minimal harmonic content above 140 Hz, this implies $\beta = 0.5$ and $K = 7$. With a sampling rate of 20 kHz (1000 points per period), the 20 Hz optimal input signal can be generated with the command: `generate-Triangle(20000, 20, 0.5, 7)`. This signal and the other signals discussed in Section VI were applied to develop a scan with $13 \mu\text{m}$ linear range. As the choice of β and K is identical to that in Section VI, Fig. 9(e) also pertains to the signals here.

The resulting displacement and difference to an ideal triangle wave is plotted in Fig. 10(c) and (d). The performance is summarized in Table II. Although the frequency-weighted-power signal outperforms other techniques, the magnitude of the error is significantly greater than expected from the spectra plotted Fig. 9(e). The difference is due to the presence of measurement noise and piezoelectric hysteresis that set a minimum bound on the achievable error.

In general, piezoelectric hysteresis will be worsened if the optimization increases peak signal amplitude and *vice versa*.

VIII. CONCLUSION

In this work, scanning trajectories for band-limited systems are designed that minimize a frequency or time domain cost function while enforcing linearity over a certain range. Specific cost functions include minimum velocity, acceleration or power. These are easily combined to achieve multiple objectives, and/or subjected to frequency domain weighting.

The frequency-weighted-power objective was introduced to maximize the scanning performance of band-limited systems. It enforces linearity over a certain range ($\pm\beta$) while minimizing signal power above a chosen frequency. The key advantages of the frequency-weighted-power signal are as follows:

- perfect linearity over a certain range ($\pm\beta$);
- minimum frequency content above the chosen K th harmonic;
- β and K can be varied to achieve arbitrarily low oscillation;
- simplifies and improves the performance of feedforward and feedback control systems.

The frequency-weighted-power signal outperforms present techniques in simulation and experiment on a standard nano-positioning platform. Even with conservative values of β and K ,

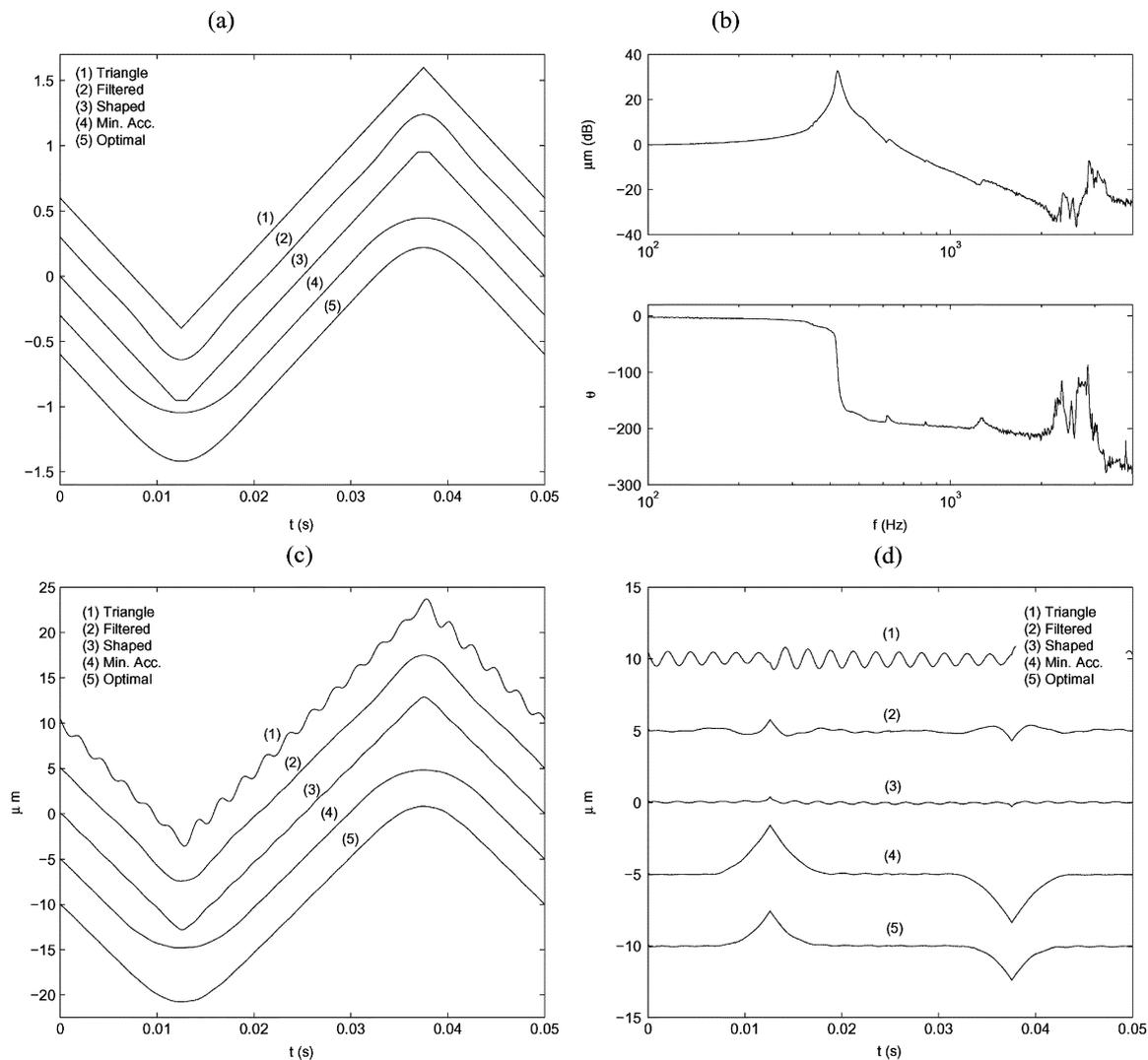


Fig. 10. 20 Hz input signals in (a) were experimentally applied to the scanner with frequency response shown in (b). The resulting displacement and corresponding error is plotted in (c) and (d). (a) Input signals. (b) Frequency response (in micrometers per Volt). (c) Resulting displacement. (d) Error.

TABLE II
EXPERIMENTAL ROOT-MEAN-SQUARE ERROR BETWEEN AN IDEAL TRIANGLE WAVE AND THE MEASURED OUTPUT (CALCULATED IN THE TIME RANGE WHERE THE OPTIMAL SIGNAL IS LINEAR). TWO CASES ARE CONSIDERED, ONE WHERE THE SCANNER IS UNLOADED, AND ANOTHER WHERE A SAMPLE PLATE IS ADDED THAT REDUCES THE RESONANCE FREQUENCY BY 16% TO 350 HZ

Resonance Freq.	420 Hz	350 Hz
Triangle	2.9%	3.87%
Filtered-Triangle	0.61%	0.72%
Shaped-Triangle	0.56%	1.1%
Min. Acc.	0.21%	0.60%
Optimal	0.18%	0.22%

an order of magnitude improvement in induced oscillation can be achieved. This improvement increases dramatically as scan range is sacrificed, or more harmonics are allowed.

REFERENCES

- [1] E. Meyer, H. J. Hug, and R. Bennewitz, *Scanning Probe Microscopy. The Lab on a Tip*. Heidelberg, Germany: Springer-Verlag, 2004.
- [2] Q. Zou, K. K. Leang, E. Sadoun, M. J. Reed, and S. Devasia, "Control issues in high-speed AFM for biological applications: Collagen imaging example," *Asian J. Control*, vol. 6, no. 2, pp. 164–176, June 2004.
- [3] D. Y. Abramovitch, S. B. Andersson, L. Y. Pao, and G. Schitter, "A tutorial on the mechanisms, dynamics, and control of atomic force microscopes," in *Proc. Amer. Control Conf.*, NY, Jul. 2007, pp. 3488–3502.
- [4] S. Devasia, E. Eleftheriou, and S. O. R. Moheimani, "A survey of control issues in nanopositioning," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 5, pp. 802–823, Sep. 2007.
- [5] Y. Zhao and S. Jayasuriya, "Feedforward controllers and tracking accuracy in the presence of plant uncertainties," in *Proc. Amer. Control Conf.*, Baltimore, MD, Jun. 1994, pp. 360–364.
- [6] J. S. Dewey, K. K. Leang, and S. Devasia, "Experimental and theoretical results in output-trajectory redesign for flexible structures," *ASME J. Dyn. Syst., Meas. Control*, vol. 120, pp. 456–461, 1998.
- [7] G. Schitter and A. Stemmer, "Identification and open-loop tracking control of a piezoelectric tube scanner for high-speed scanning-probe microscopy," *IEEE Trans. Control Syst. Technol.*, vol. 12, no. 3, pp. 449–454, May 2004.
- [8] D. Croft, G. Shed, and S. Devasia, "Creep, hysteresis, and vibration compensation for piezoactuators: Atomic force microscopy application," *Trans. ASME, J. Dyn. Syst., Meas., Control*, vol. 123, pp. 35–43, Mar. 2001.
- [9] H. Perez, Q. Zou, and S. Devasia, "Design and control of optimal scan trajectories: Scanning tunneling microscope example," *J. Dyn. Syst., Meas., Control*, vol. 126, pp. 187–197, Mar. 2004.

- [10] Y. Wu and Q. Zou, "Iterative control approach to compensate for both the hysteresis and the dynamics effects of piezo actuators," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 5, pp. 936–944, Sep. 2007.
- [11] J. L. Fanson and T. K. Caughey, "Positive position feedback control for large space structures," *AIAA J.*, vol. 28, no. 4, pp. 717–724, 1990.
- [12] S. S. Aphale, B. Bhikkaji, and S. O. R. Moheimani, "Minimizing scanning errors in piezoelectric stack-actuated nanopositioning platforms," *IEEE Trans. Nanotechnol.*, vol. 7, no. 9, pp. 79–90, Jan. 2008.
- [13] S. S. Aphale, A. J. Fleming, and S. O. R. Moheimani, "Integral control of resonant systems with collocated sensor-actuator pairs," *IOP Smart Mater. Structures*, vol. 16, pp. 439–446, Apr. 2007.
- [14] K. K. Leang and S. Devasia, "Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in AFM piezoactuators," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 5, pp. 927–935, Sep. 2007.
- [15] L. Y. Pao, J. A. Butterworth, and D. Y. Abramovitch, "Combined feedforward/feedback control of atomic force microscopes," in *Proc. Amer. Control Conf.*, New York, Jul. 2007, pp. 3509–3515.
- [16] S. Devasia, "Should model-based inverse inputs be used as feedforward under plant uncertainty?," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1865–1871, Nov. 2002.
- [17] A. J. Fleming and S. O. R. Moheimani, "Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 1, pp. 33–44, Jan. 2006.
- [18] S. S. Aphale, A. J. Fleming, and S. O. R. Moheimani, "High speed nano-scale positioning using a piezoelectric tube actuator with active shunt control," *IET Micro Nano Lett.*, vol. 2, no. 1, pp. 9–12, Jan. 2007.
- [19] M. J. Rost, L. Crama, P. Schakel, E. van Tol, G. B. E. M. van Velzen Williams, C. F. Overgaw, H. ter Horst, H. Dekker, B. Okhuijsen, M. Seynen, A. Vijftigschild, P. Han, A. J. Katan, K. Schoots, R. Schumm, W. van Loo, T. H. Oosterkamp, and J. W. M. Frenken, "Scanning probe microscopes go video rate and beyond," *Rev. Scientific Instrum.*, vol. 76, no. 5, pp. 053710-1–053710-9, Apr. 2005.
- [20] R. A. Masterson, W. E. Singhose, and W. P. Seering, "Setpoint generation for constant-velocity motion of space-based scanners," *AIAA J. Guid., Control, Dyn.*, vol. 23, no. 5, pp. 892–895, 2000.
- [21] W. Singhose, N. Singer, and W. Seering, "Comparison of command shaping methods for reducing residual vibration," presented at the Eur. Control Conf., Rome, Italy, Sep. 1995.
- [22] N. C. Singer and W. P. Seering, "Preshaping command inputs to reduce system vibration," *ASME J. Dyn. Syst., Meas., Control*, vol. 112, no. 2, pp. 76–82, Mar. 1990.
- [23] G. Schitter, G. E. Fantner, P. J. Thurner, and J. Adams, "Design and characterisation of a novel scanner for high-speed atomic force microscopy," in *Proc. IFAC Symp. Mechatron. Syst.*, Heidelberg, Germany, Sep. 2006, pp. 819–824.
- [24] R. Fletcher, *Practical Methods of Optimization*. Chichester, U.K.: Wiley, 1987.
- [25] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, NJ: Pearson Education Inc., 2007.



Andrew J. Fleming was born in Dingwall, Scotland, in 1977. He received the Bachelor degree in electrical engineering and the Ph.D. degree from The University of Newcastle, Australia (Callaghan campus), in 2000 and 2004, respectively.

In 2005, he was a Research Academic with the Centre for Complex Dynamics and Control, The University of Newcastle, Australia. He is currently an Australian Research Council APD Fellow stationed with the School of Electrical Engineering and Computer Science, The University of Newcastle.

His research interests include micro-cantilever sensors, nano-positioning, and sensor-less control of sound and vibration.



Adrian G. Wills was born in Orange, N.S.W. Australia. He received the B.E. degree in electronics and the Ph.D. degree from The University of Newcastle, Australia (Callaghan Campus), in 1999 and 2003, respectively.

Since then, he has held a Postdoctoral Research position with Newcastle, where the focus of his research has been in the area of system identification and model predictive control.