2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Efficient Filtering Using Monotonic Walk Model

Dimitry Gorinevsky

Abstract— **This paper proposes a nonlinear filter for estimating monotonic underlying trend from noisy observations. The filter computes Maximum A posteriori Probability (MAP) estimate using a monotonic walk model instead of the random walk model in standard linear filtering. The batch estimate is a solution of Quadratic Programming (QP) problem. This paper shows that the QP has a form of isotonic regression (IR) and has a linear computational complexity.**

The filter is implemented in a Moving Horizon Estimation (MHE) setting. The data beyond the estimation horizon are replaced by the initial condition parameters (arrival cost). The MHE for IR is nonsmooth, so the existing nonlinear MHE theory is not applicable. By exploiting properties of the IR solution, we develop an update of the MHE arrival cost, which is provably close to the full information MAP solution and stable. The analysis is complemented by a Monte Carlo simulation study of the proposed nonlinear filtering algorithm. The simulation results confirm improved performance of the proposed filter compared with a linear filter and the earlier version of the MHE update.

I. Introduction

The filtering approach most broadly used in practice is Exponentially Weighted Moving Average (EWMA) update. The EWMA update could be derived as a steady state Kalman filter for estimating an underlying random walk trend from a noisy time series.

The problem of interest is filtering for a *monotonic walk* model, where the increments of the underlying trend have zero probability of being negative. The monotonic walk model could reflect physical damage accumulation with time. The model is of interest to systems health management, fault diagnostics, and prognostics system development. By analogy with the EWMA filter, a filter based on the monotonic walk model should be broadly useful.

Estimation with the monotonic walk model was considered in our earlier papers [5], [16], [15]. The MAP batch estimation problem was formulated as a Quadratic Programming (QP) with linear monotonicity constraints. The filtering could be implemented by repeatedly solving the QP as the new data points are available. The above cited papers used standard sparse QP solvers.

The best commercially available interior point QP solvers take more than a hundred milliseconds for a horizon of a few hundred time steps on a current PC computer. This works for some, but not for all application. This paper shows that the optimal estimation based on monotonic walk model is reduced to an isotonic regression (IR) problem, which can be solved two orders of magnitude faster. This removes most of the limitations on the use of the proposed monotonic filtering scheme and is the first contribution of this paper.

IR problems are QP problems of special type studied in statistics literature. They are discussed in the books [1], [13] as well as in many papers. A univariate IR problem can be solved with linear complexity algorithms, see [2]. The complexity is the same as for a batch linear-quadratic estimation. A Monte Carlo simulation study in this paper characterizes performance of the fast IR-based monotonic filtering algorithm compared to EWMA filtering. The observed improvement is 7 to 2 dB in root mean square error; the improvement is bigger for larger signal to noise ratio. The simulation study is the second contribution of the paper.

Most of the published IR applications are in statistics, such as estimation of cumulative distribution functions; so far, there has been little work in signal processing and control. IR has been applied to detection of fault conditions in the earlier application papers [3], [6], [7]. A related, but different, signal processing problem is discussed in [17]. The third contribution of the paper is analytical characterization of the filtering properties of the IR solution. The filtering properties were not characterized earlier.

In on-line signal processing applications, the data batch size would eventually overwhelm available computational capacity. A Moving Horizon Estimation (MHE) version of the batch estimation can then be used by disregarding the old data beyond certain time horizon.

Optimization-based estimation in the MHE framework, has been studied earlier, see [4], [11], [18]. These papers do not specifically consider monotonicity constraints. The MHE stability analysis in [4], [11] is based on system linearization, The analysis considers arrival cost, the penalty for the initial condition on the moving horizon. In [11] the arrival cost is approximated by a quadratic functions assuming that the estimator constraints are inactive asymptotically in the delay time. The monotonicity constraints may be active and such assumption is invalid.

Our earlier paper [14] considered optimal estimation using monotonic walk model in the MHE framework. In [14], an arrival cost model was chosen by selecting the initial condition for the current horizon to be on a hard constraint defined by the solution for the previous horizon. This was proven to work if a moving average of the (noisy) data is monotonic – a restrictive condition. The simulation study in this paper shows that the update proposed in [14] might accumulate a large and increasing bias of the noisy data estimate if this condition does not hold. In practice, the noisy data might have non-monotonic moving average, though the underlying trend is monotonic. This makes the update of [14] unsuitable for many applications. This paper introduces an

This work was supported by the NSF GOALI Grant ECS-0529426 and AFOSR grant FA9550-06-1-0514

Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, gorin@stanford.edu

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

MHE formulation, which approximates the full information solution closely and usually matches it exactly. The proposed update of the arrival cost is based on analytical characterization of the monotonic estimation solution. It performs very well in simulations with large noise where the update of [14] fails. This is the third contribution of this paper.

The fourth contribution is analysis of the IR-based filter stability. We show that an 'impulse response' of this nonlinear filter error can be guaranteed to decay with time. The worst-case convergence to zero is as $1/t$.

II. Least squares estimation

This section establishes a departure point of the study by briefly reviewing basic linear-quadratic trending and filtering.

Consider a set of scalar data points $y(t)$ representing an underlying trend $x(t)$ perturbed by a noise

$$
\begin{aligned}
Y_T &= \text{col}\{y(1), ..., y(T)\}, && (1)\\
X_T &= \text{col}\{x(1), ..., x(T)\} && (2)
\end{aligned}
$$

Estimating the underlying trend X_T from the observed data Y_T requires formulating assumptions about trend X_T and the relationship between X_T and Y_T. Perhaps the simplest model, one that is broadly used in many applications, is the random walk model

$$
\begin{aligned}
x(t+1) &= x(t) + \xi(t), && (3)\\
y(t) &= x(t) + e(t), && (4)
\end{aligned}
$$

where $\xi(t)$ is the process noise and $e(t)$ is the observation noise.

Consider Maximum A posteriori Probability (MAP) estimation for of X_T (3)–(4). Assume that $\xi(t)$, $e(t)$ and $x(1)$ in (3)–(4) are independent normally distributed

$$
\begin{aligned}
p_e(x) \sim N(0, q),\ p_0(x) \sim N(x_0, \Xi_0), && (5)\\
p_\xi(x) \sim N(0, \Xi) && (6)
\end{aligned}
$$

The MAP estimation problem can be formulated as $L = -\log P(X_T|Y_T) \to \min$. The model (1)–(6) yields (see [5])

$$
L = \frac{1}{2}\sum_{t=1}^{T}[y(t) - x(t)]^2 + \frac{1}{2}r\sum_{t=2}^{T}[x(t) - x(t-1)]^2 + \frac{1}{2}r_0[x(1) - x_0]^2 \to \min, \quad (7)
$$

where $r = q/\Xi$, $r_0 = q/\Xi_0$. The regularization parameter r in (7) defines the solution smoothness (larger r means smoother trend estimate). The second parameter r_0 defines the initial condition influence (which decays quickly as t increases). The estimate (7) is linear in $y(t)$ and for $r_0 = 0$ can be presented in the form

$$
x(t) = \sum_{\tau=1}^{T} K(t, \tau) y(\tau), \quad (8)
$$

The kernels $K(t, \tau)$ describe the contributions of the data points $y(\tau)$ towards the trended value $x(t)$ and represent noncausal smoothing action of the operator (8). Sufficiently far inside the data interval $[1,\ T]$, the kernels have the same shape, are shift-invariant $K(t, \tau) \sim e^{-c|t-\tau|}$.

Denote by $x(t|T)$ the estimate at time t obtained based on T data points available. As the time series increases from T points to $T+1$, the filtering update relates $x(T+1|T+1)$ and $x(T|T)$. For the problem in hand, the update can be presented in the form of a Kalman Filter. It is well known that for a first-order random walk model (3)–(4), the stationary Kalman Filter has a form of EWMA filter

$$
x(T+1|T+1) = \alpha x(T|T) + (1-\alpha) y(T+1) \quad (9)
$$

The filter (9) corresponds to the kernel $K(T, T-n) = (1-\alpha)\cdot\alpha^n$, where $\alpha = \frac{1}{2}\left(\sqrt{r^2+4r} - r\right)$.

III. Estimation based on monotonic walk model

For random walk model (3), the monotonicity assumption can be expressed as $\xi(t) \geq 0$.

A. Optimal Estimation Problem

For the monotonic walk model, instead of (6), we assume that process noise $\xi(t)$ has one-sided *exponential* distribution

$$
\begin{aligned}
p_\xi(x) &\sim \frac{1}{\Xi}e^{-\frac{x}{\Xi}}, \quad \text{for } x \geq 0 && (10)\\
p_\xi(x) &\sim 0, \quad \text{for } x < 0 && (11)
\end{aligned}
$$

The single parameter Ξ of the distribution $p_\xi(x)$ has a meaning of the average damage accumulation rate.

The MAP problem based on (3)–(4), (5) and (10)–(11) is similar to (7). The difference is in one term corresponding to (6) (see [5] for detail)

$$
L = \frac{1}{2}\sum_{t=1}^{T}[y(t) - x(t)]^2 + r\sum_{t=2}^{T}[x(t) - x(t-1)] + \frac{1}{2}r_0[x(1) - x_0]^2 \to \min, \quad (12)
$$

$$
\text{subject to: } x(t) \geq x(t-1), \quad (t = 2, \ldots, T), \quad (13)
$$

where with an abuse of notation $r = q/\Xi$, $r_0 = q/\Xi_0$. The second sum in (12) could be collapsed into $r\cdot[x(T) - x(1)]$.

The QP problem (12)–(13) could be solved for $x(t)$ using an off-the-shelf sparse QP solver. The meaning of the regularization (smoothing) parameter r in (12) is similar to r in (7). A specialized interior-point method that solves such QP problems in linear time is discussed in [8]. This special type of QP can be solved even more efficiently.

B. Isotonic Regression

The problem (12)–(13) can be re-written in the form

$$
\sum_{t=1}^{T} w(t)\cdot[a(t) - x(t)]^2 \to \min \quad (14)
$$

$$
\text{subject to: } x(t) \geq x(t-1), \quad (t = 2, \ldots, T), \quad (15)
$$

where the expressions for the coefficients $w(t)$ and $a(t)$ can be obtained by expanding (14)-(15) and matching the linear and quadratic terms with (12)-(13)

$$
\begin{aligned}
w(1) = \frac{1+r_0}{2},\ \ a(1) = \frac{y(1) + r + r_0 x_0}{1+r_0}, && (16)\\
w(t) = 1/2,\ \ a(t) = y(t),\ \ \text{for } \ 2 \leq t \leq T-1, && (17)\\
w(T) = 1/2,\ \ a(T) = y(T) - r. && (18)
\end{aligned}
$$

The problem (14)–(15) is known as Isotonic Regression (IR) problem and was extensively studied in the statistics and operations research literature, see [1], [13]. The problem has an explicit solution (max-min formula in [1]).

$$x(t) = \max_{1\leq s\leq t} \min_{T\geq n\geq t} \frac{\sum_{j=s}^{n} w(j)a(j)}{\sum_{j=s}^{n} w(j)} \tag{19}$$

The expressions (16)–(19) allow computing the solution of the problem (12)-(13) two orders of magnitude faster than with a standard sparse QP solver. For example solving the QP problem of the size $T = 800$ takes about 80 ms on a PC using the state of the art Mosek sparse QP solver called from Matlab and about 1.5 ms using a variant of formula (19) coded in Matlab. (Matlab Optimization Toolbox QP solver takes several seconds). Computing (19) directly is not the fastest algorithm for solving (12)–(13). Several algorithms for solving the IR problem are reviewed in [2]; two of them have linear complexity, including the well-known Pool Adjacent Violators Algorithm (PAVA).

C. Solution Characterization

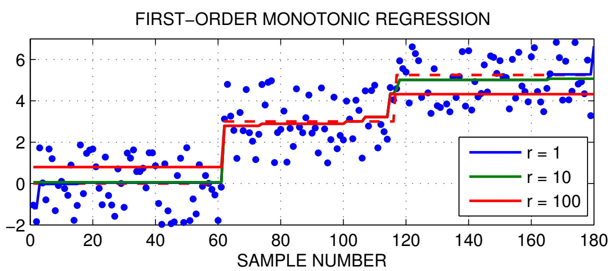

Fig. 1. Batch MAP estimation using monotonic walk model.

Figure 1 shows examples of the monotonic solution (16)—(19) to (12)–(13). The data, shown as dots, are obtained by adding a gaussian noise to a piece wise constant underlying trend (dashed line). The trend estimates for $r = \{1, 10, 100\}$ are shown as solid lines. They recover the underlying trend well. The estimates for larger r are flatter. The trends consists of several constant value segments with jumps between them. These segments are *active set intervals*; the monotonicity constraint (13) is active on each segment.

Theorem 1: Consider a solution to (12)–(13) on an active set interval $[t_0, t_0 + n]$ such that

$$x(t_0 - 1) < x(t_0) = \ldots = x(t_0 + n) < x(t_0 + n + 1), \tag{20}$$

For $1 < t_0$ and $t_0 + n < T$ the solution $x(t)$ satisfies

$$x(t_0) = \ldots = x(t_0 + n) = \frac{1}{n+1} \sum_{t=t_0}^{t=t_0+n} y(t) \tag{21}$$

and for any m such that $0 \leq m < n$

$$\frac{1}{n+1} \sum_{t=t_0}^{t=t_0+n} y(t) < \frac{1}{m+1} \sum_{t=t_0}^{t=t_0+m} y(t) \tag{22}$$

If $t_0 = 1$, then (21), (22) hold with $y(1)$ replaced by $y(1) + r_0 x_0 + r$. If $t_0 + n = T$, then (21), (22) hold with $y(T)$ replaced by $y(T) - r$.

Theorem 1 is proven in Appendix A. It provides a justification to the IR solution (16)–(19).

Consider now the solution sensitivity to the observed data $y(t)$. Consider small variations in $y(t)$ in (21), such that the active sets do not change. The corresponding change of the solution $x(t)$ has the form reminiscent of (8) with a rectangular window kernel

$$\Delta x(t) = \frac{1}{n} \sum_{k=t_0}^{t_0+n} \Delta y(k), \quad \text{for} \quad t = t_0, \ldots, t_0 + n, \tag{23}$$

where $\Delta x(t)$ is a variation of the solution corresponding to the variation $\Delta y(k)$ of the data point k. The kernel window size n is the length of the largest active constraint interval such that $t \in [t_0, \;\; t_0 + n]$. The rectangular window (23) for the monotonic walk model can be compared to the two-sided exponential kernel (8) for the random walk model.

D. Filtering

Consider filtering based on the monotonic walk estimation. As the time horizon T increases (1)–(2), the IR solution (16)—(19) to (12)–(13) can be re-computed. The filter output is given by the last point of the IR solution and can be computed without finding the entire solution. For $t = T$ the minimization set in (19) is reduced to a single point $n = T$. Computing max in (19) then requires just $2T$ divisions/multiplications and $2T$ additions.

Recall the notation $x(t|T)$ for the estimate of the trend variable $x(t)$ at time t obtained based on the data set of T data points. The linear estimation problem of Section II relates $x(T + 1|T + 1)$ and $x(T|T)$ through the EWMA filtering update (9). The update based on monotonic walk model is more complicated. To understand the filtering update consider a special case of (21) for $t_0 + n = T$ in Theorem 1. Suppose that $[T - n + 1, T]$, where $n < T$, is an active set interval and $x(T - n|T) < x(T - n + 1|T)$. On this interval the solution $x(t|T)$ is constant and

$$x(T|T) = \frac{1}{n} \sum_{t=T-n+1}^{T} y(t) - \frac{r}{n} \tag{24}$$

Some closed-form expressions describing the filtering behavior of the monotonic regression are as follows.

(i) Let $x(T + 1|T + 1)$ be on active constraint interval of length $n_1 \geq 1$ and $x(T|T)$ be on active constraint interval of length $n \geq 1$. Applying (24) for T, n_0 and for $T + 1$, n yields

$$\begin{aligned} x(T+1|T+1) &= \frac{n}{n_1} x(T|T) + \frac{1}{n_1} y(T+1) \\ &+ \frac{1}{n_1} \sum_{t=T-n_1+1}^{T-n} y(t), \quad \text{for } n_1 \geq n+1 \end{aligned} \tag{25}$$

$$\begin{aligned} x(T+1|T+1) &= \frac{n}{n_1} x(T|T) + \frac{1}{n_1} y(T+1) \\ &- \frac{1}{n_1} \sum_{t=T-n+1}^{T-n_1+1} y(t), \quad \text{for } n_1 \leq n \end{aligned} \tag{26}$$

(ii) If $y(T+1)$ is large enough, then the constraint is inactive for $x(T+1|T+1)$. In that case, $n=1$ in (24) and, irrespective to what $x(T|T)$ is, we get

$$x(T+1|T+1) = y(T+1) - r \tag{27}$$

E. Simulation study of filtering

The proposed monotonic regression filter was compared with EWMA filter in an extensive Monte Carlo simulation study. The test data sets of N=250 points were generated. A piece-wise constant underlying trend $x(t)$ was generated containing a given number of jumps. The jump locations were uniformly distributed on $[1,\ N]$ and the jump amplitudes were exponentially distributed with a unit mean. A Gaussian noise was added to $x(t)$ to produce the data $y(t)$. The filter outputs were compared with the underlying trend $x(t)$ to yield a mean square value of the filtering error.

The EWMA filter (9) has a single tuning parameter α. The monotonic regression filter formulation (12)–(13) has a single tuning parameter r (we assumed that $r_0 = 0$). These parameters were chosen to minimize respective root mean square (RMS) errors of recovering the underlying trend for the two filters. To do this, we considered a grids of values for α and a grid for r. For each value on the grid, 1000 simulation runs were completed with differrent random realizations of the additive noise and different realization of the underlying trend (woth a fixed number of jumps). For each run, a RMS error was computed. The chosen parameter value minimizes the average RMS error over the 1000 runs.

The average RMS error for the optimal filter parameters depends on the signal to noise ratio. In the simulations, the signal magnitude (the mean jump magnitude) was unity while the noise covariance varied. The results for the two filters are illustrated in the two upper plots in Figure 2. The improvement of the average RMS errors for the monotonic filter over the EWMA is shown in the lower plots in Figure 2. The monotonic filter has a smaller RMS error for all signal to noise ratios. The left plots in Figure 2 show results for 2 jumps in the underlying trend; the right plots show the results for 10 jumps. There are more transients for 10 jumps and the tracking improvement is smaller. The improvement is larger for smaller noise and reaches 7 dB.

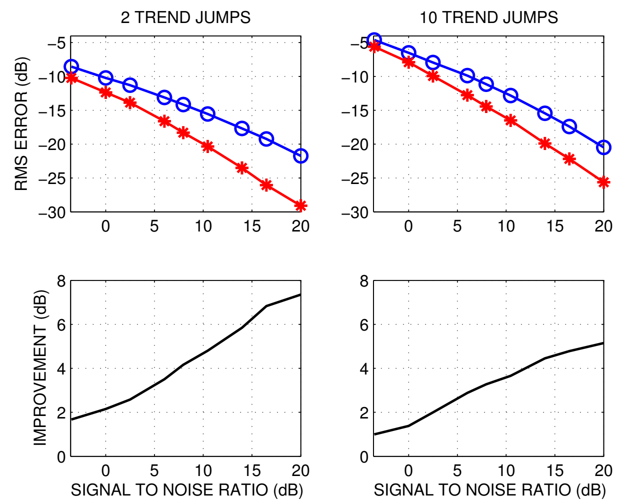

Fig. 2. The upper plot shows RMS error in dB for the monotonic filter (curves marked by *) and EWMA (curves marked by o). The lower plots show improvement in dB.

IV. Moving Horizon Estimation

As the data set size T increases, a practical approach is to keep the most recent data over a fixed horizon N only. The old data is discarded and the computations are performed for the data from time $T-N+1$ to T. This section introduces and analyzes such *Moving Horizon Estimation (MHE)* formulation. We will further refer to the full information monotonic walk estimation problem as the MAP problem. We will denote the solutions of these problems as $x_{MHE}(t|T)$ and $x_{MAP}(t|T)$.

The MAP and MHE designations are introduced for notation convenience. In fact, the MHE problem can also be formulated as optimal Bayesian estimation problem. The problem of optimal estimation on the interval $[T-N+1,\ T]$ is a QP problem of the form (12)–(13)

$$\frac{1}{2}\sum_{t=T-N+1}^{T}[y(t)-x(t)]^2 + r[x(T)-x(T-N+1)] \\ + \frac{r_0(T)}{2}[x(T-N+1)-x_0(T)]^2 \to \min \tag{28}$$

subject to: $x(t) \geq x(t-1), \quad (t = T-N+2,\dots,T)$, (29)

The MHE solution $x_{MHE}(t|T)$ is defined by (28)–(29) for $t \in [T-N+1,\ T]$. As T increases and more data becomes available, the lower end of the interval $[T-N+1,\ T]$ increases as well and the old data is left out. The solution $x_{MHE}(t|T)$ for $t < T-N+1$ can be taken as the solution at the last moving horizon interval still including t, i.e., $x_{MHE}(t|T) \equiv x_{MHE}(t|t+N-1)$. The initial condition penalty parameters $r_0(T)$ and $x_0(T)$ define the MHE arrival cost. The cost depends on the MHE horizon, which is indexed by T. in (28)–(29) depend on T. The update for $r_0(T)$, and $x_0(T)$ is a key part of the MHE algorithm.

A. Proposed MHE update

The last term in the MHE problem (28)–(29) describes *arrival cost*, which characterizes the solution for the discarded points with $t \leq T_N$. In [11], the update based on a linear-quadratic approximation of the *arrival cost* is proposed for a general nonlinear MHE setup. The update is derived by linearizing the estimator around the steady state and computing a Kalman Filter update for the linearized system. The approach of [11] is referenced and used in other publications on the subject, e.g. [4]. Yet, it is not applicable to the problem in question. The monotonicity constraints might be active at the steady state; in that case the approach of [11] does not work and the Kalman Filter update is not valid.

Let us compare the solutions to the QP problem (12)–(13) formulated for the entire data set $t \in [1,\ T]$ and the solution to the QP problem (28)–(29) formulated for $t \in [T-N+1,\ T]$. (We assume that $T > N$.)

Theorem 2: Let $[t_0,\ t_1]$ be the largest active set interval for $x_{MAP}(t|T)$ containing $T-N+1$, the beginning of the MHE horizon. Denote $n_0 = T-N+1-t_0$. The solution $x_{MHE}(t|T)$ (28)–(29) on the interval $t \in [T-N+1,\ T]$ coincides with the solution $x_{MAP}(t|T)$ to (12)–(13) on that interval, iff in (28)–(29)

$$r_0(T) = n_0, \quad x_0(T) = \frac{1}{n_0}\sum_{t_0}^{T-N} y(t) - \frac{r_1}{n_0} \tag{30}$$

where $r_1 = 0$ if $t_0 = 1$ and $r_1 = r$ if $t_0 > 1$.

Proof. See Appendix B.

The proposed algorithm for updating the MHE arrival cost is a recursive form of (30). We consider $T \geq N$ only. For $t < N$ the MHE problem is the same as the MAP problem. Consider the MHE interval $[T-N+1,\ T]$ and the MHE solution $x_{MHE}(t|T)$ obtained on this interval. The update logic depends on whether $t = T-N+1$ is the last point of an active set interval and is conditional on

$$\Delta x_T = x_{MHE}(T-N+2|T) - x_{MHE}(T-N+1|T) \tag{31}$$

Algorithm 1: MHE Parameter Update
Initialize.

$$T = N, \quad x_0(T) = 0, \quad r_0(T) = 0 \tag{32}$$

Update.

$$\begin{aligned} x_0(T+1) &= \frac{r_0(T)\cdot x_0(T)}{1+r_0(T)} + \frac{y(T-N+1)}{1+r_0(T)}, \\ r_0(T+1) &= 1 + r_0(T), \text{ if } \Delta x_T = 0 \qquad (33) \\ x_0(T+1) &= y(T-N+1), \\ r_0(T+1) &= 0, \qquad\qquad \text{if } \Delta x_T > 0 \qquad (34) \end{aligned}$$

If $\Delta x_T > 0$ in (31), then $T+N-1$ is a jump point for $x_{MHE}(t|T)$. At next increments of time T this jump point is left outside of the current MHE horizon. We will call the jump points at $t \leq T-N$ historical jump points. The MAP and MHE solutions coincide as long as each MHE jump point is also a MAP jump point. This last assumption is valid most of the time for sufficiently large horizon N.

B. Simulation of the MHE update

The earlier work [14] proposed a different MHE update for the same problem. The update imposed a hard constraint on the initial estimate $x(T-N+2|T+1) = x(T-N+2|T)$. The constraint could be encoded through initial penalty parameters of the form $r_0(T) = +\infty$, and a $x_0(T+1) = x(T-N+2|T)$ In [14] this MHE update was shown to produce accurate results if a moving average of the data is monotonic. This condition is restrictive. If it does not hold, the update might yield inaccurate results.

The performances of the proposed update and of the update of [14] for noisy data are compared in Figure 3. The upper plot shows the data generated by adding a Gaussian noise with a covariance of 0.75 and a harmonic variation with amplitude 0.1 and period of 272 to the underlying trend. The bottom plot shows the ground truth data (underlying trend) as a dashed line. The dash-dotted line, labeled MHE-HARD, shows the MHE filtering results with the hard-constraint update of [14]. The moving window width was $N = 50$ and $r = 10$ The update suffers from irreversible accumulation of large bias. The solid line, labeled MHE-OPTIMIZED shows the results for the proposed update (Algorithm 1). For all practical purposes this MHE filter output coincides with the output of the full information monotonic regression filter, which is labeled as MAP FILTER and plotted by a dotted line (mostly overlaping the solid line).

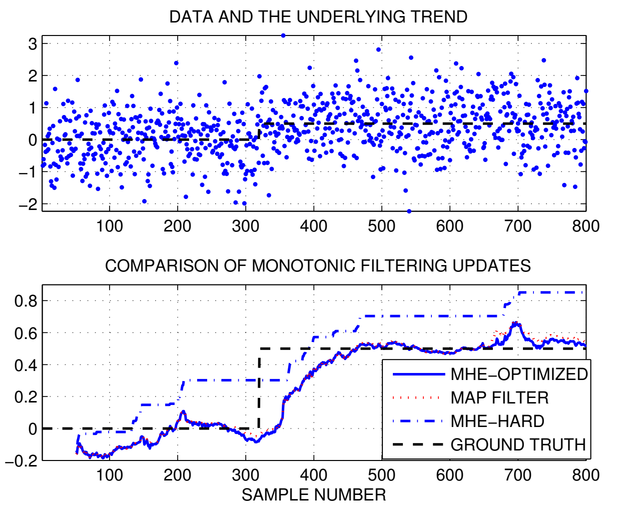

Fig. 3. Comparison of the MHE updates and full-information IR filtering

The MAP filter and proposed MHE filter were compared in the Monte Carlo simulation study mentioned earlier (see Figure 2). The performance loss of the MHE flter was less than 0.016 dB in all cases and less than 0.001 dB in most cases. This is too small to discern in Figure 2.

C. Filter stability

Consider Lyapunov stability of the update, similar to how it is done in [11] for a general nonlinear MHE update. Recall that the results of [11] presume that the system can be linearized around the equilibrium $y(t) = 0$, which is not the case when the monotonicity constraint is present. Both the MAP and MHE solutions converge to a steady state if the initial condition x_0 is off (nonzero). If $x_0 < 0$, the convergence is deadbeat in one step: $x(t) = 0$ for $t > 1$ is the solution. If $x_0 > 0$ and the active set interval is $[1,\ T]$, then the MAP solution can be obtained as

$$x(T|T) = \frac{x_0 r_0}{T + r_0} \tag{35}$$

The MAP filter is stable since the solution (35) asymptotically converges to zero as the time T advances. The solution decays as $1/T$, rather than exponentially. In this case the MHE solution coincides with the MAP solution.

Consider now BIBO (bounded input/bounded output) stability of the filter. This condition is more interesting since it concerns more realistic noisy data input to the filter. We will assume that the input sequence $y(t)$ is bounded in the l_∞ norm such that

$$|y(t)| \leq B_y \tag{36}$$

Let $t_0 \leq T$ be the last jump point of the MAP solution and $n = T - t_0 + 1$. Then (24) holds. By using (36) in (24) we get that $x(T|T)$ is bounded as

$$|x(T|T)| \leq \frac{1}{n} \sum_{t=T-n+1}^{T} |y(t)| + \frac{r}{n} \leq B_y + r \qquad (37)$$

Finally, we will consider an 'impulse response' of the proposed nonlinear filter. Stability of linear filters is often defined as convergence of an impulse response to zero. Consider the following bound on the output signal

$$|x(T|T)| \leq \sum_{t=1}^{T} h(t)|y(t)| + C, \qquad (38)$$

where $h(t) \geq 0$ are the impulse response coefficient bounds for the nonlinear filter and C is a constant. By using (24), a bound of the form (38) can be obtained with $h(t|T) = 1/(T - t + 1)$. The response convergence to zero is slower than exponential.

V. Conclusions

This paper discussed estimation of an underlying trend from noisy data based on monotonic walk model of the trend. The solution is a filter using a moving horizon implementation of an isotonic regression. This nonlinear filter has the the same low computational complexity as a linear filter. Similar to an exponentially weighted moving average filter, the proposed filter has a single tuning parameter. The solution is close to the full-information optimal nonlinear filtering estimate and has well understood properties. The proposed filter provides several dB improvement over a linear filter in recovering monotonic trends from noisy data.

References

[1] Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. *Statistical Inference Under Order Restrictions; The Theory and Application of Isotonic Regression*, New York: Wiley, 1972.

[2] Best, M.J. and Chakravarti N., "Active set algorithms for isotonic regression; a unifying framework," *Mathematical Programming*, Vol. 47, 1990, pp. 425–439.

[3] Chang, J.T., and Fricker, R.D., "Detecting when a monotonically increasing mean has crossed a threshold," *Journal of Quality Technology*, Vol. 31, No. 2, 1999, pp. 217–234.

[4] Goodwin, G.C., Seron, M.M., and De Dona, J.A., *Constrained Control and Estimation*, Springer Verlag, 2004.

[5] Gorinevsky, D., "Monotonic regression filters for trending gradual deterioration faults," *American Control Conference*, pp. 5394–5399, Boston, MA, June 2004.

[6] Hellerstein, J.L. , "A statistical approach to diagnosing intermittent performance-problems using monotone relationships," *SIGMETRICS*, 1989, pp. 20–28

[7] Kang, P. and Birtwhistle, D., "Condition monitoring of power transformer on-load tap-changers. Part 2: Detection of ageing from vibration signatures," *IEE Proceedings - Generation, Transmission and Distribution*, Vol. 148, No. 4, 2001, pp. 307–311.

[8] Kim, S.-J., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D., "A method for large-scale ℓ_1-regularized least squares problems with applications in signal processing and statistics," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 1, No. 4, 2008, pp. 606–617.

[9] Oppenheim, A.V., Schafer, R.W., and Buck, J.R., *Discrete-Time Signal Processing*, Prentice Hall, 1999.

[10] Pollock, D.S.G., *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*, Academic Press, London, 1999.

[11] Rao, C.V., Rawlings, J.B., and Mayne, D.Q., "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations," *IEEE Trans. on Automatic Control*, Vol. 48, No 2, 2003, pp. 246–258.

[12] Restrepo, A. and Bovik, A.C., "On the statistical optimality of locally monotonic regression," *IEEE Trans. on Signal Processing*, Vol. 42, No. 6, 1994, pp. 1548–1550.

[13] Robertson, T., Wright, F.T., and Dykstra, R.L., *Order Restricted Statistical Inference*, New York: Wiley, 1988.

[14] Samar, S., Gorinevsky, D., and Boyd, S., "Moving horizon filter for monotonic trends," *IEEE Conf. on Decision and Control*, Paradise Island, Bahamas, December 2004.

[15] Samar, S., Gorinevsky, D., and Boyd, S., "Embedded estimation of fault parameters in an unmanned aerial vehicle," *IEEE Conf. on Control Applications*, Munich, Germany, October 2006.

[16] Samar, S. and Gorinevsky, D., "Model predictive estimation of evolving faults. *American Control Conf.*, Minneapolis, MN, June 2006.

[17] Sidiropoulos, N.D. and Bro, R., "Mathematical programming algorithms for regression-based nonlinear filtering in R^N," *IEEE Tran. on Signal Processing*, Vol. 47, No. 3, 1999, pp. 771–782.

[18] Tyler, M. L. and Morari, M., "Stability of constrained moving horizon estimation schemes," Zurich, Switzerland: Swiss Federal Inst. Technol. (ETH), 1996, Available: http://control.ethz.ch.

Appendix

A. Proof of Theorem 1

The KKT (Karush–Kuhn–Tacker) optimality conditions for the problem (12)–(13) can be expressed in the form

$$x(t) - y(t) + [\lambda(t) - \lambda(t-1)] = 0, \qquad (39)$$

where $\lambda(t) \geq 0$ are the Lagrange multipliers corresponding to the monotonicity constraints and $1 < t < T$. The KKT condition for $t = 1$ has the form (39), where $\lambda(T) = r$. The condition for $t = T$ has the form (39), where $\lambda(0) = 0$ and $y(1)$ is replaced with $y(1) + r_0 x_0 + r$.

On each of the active set intervals, the constraint (13) is active. The Lagrange multipliers in the active set interval are positive, $\lambda(t) > 0$. For each of the internal points that do not belong to an active set (jump points), $\lambda(t) = 0$. Summing up (39) from t_0 to $t_0 + m$ yields

$$x(t_0) = \frac{1}{m+1} \sum_{t=t_0}^{t=t_0+m} y(t) + \frac{\lambda(t_0+m)}{m+1} - \frac{\lambda(t_0-1)}{m+1} \qquad (40)$$

The two strict inequalities in (20) mean that $\lambda(t_0 - 1) = 0$ and $\lambda(t_0 + n) = 0$. For $m = n$, we have $\lambda(t_0) = 0$ and $\lambda(t_0 + n + 1) = 0$ in (40), which yields (21). For $m < n$, we have $\lambda(t_0 - 1) = 0$ and $\lambda(t_0 + m) > 0$; the inequality (22) follows from (21) and (40). Q.E.D.

B. Proof of Theorem 2

The QP problem (28)–(29) is a convex problem with positive definite Hessian and has a unique solution. We will show that $x_{MAP}(t|T)$ satisfies the KKT equations and, thus, is the solution for (28)–(29).

Consider the solution $x_{MAP}(t|T) = const$ on the interval $[T-N+1,\ t_1]$. We assume that $1 < t_0 \leq t_1 < T$. Then the solution can be obtained from (21). The case where $t_0 = 1$ and/or $t_1 = T$ is similar. By summing up the KKT equations for the MAP solution over the $[t_0,\ T-N+1]$ interval we obtain the first KKT equation (for $t = T - N + 1$) for the MHE solution, which is satisfied by $x_{MHE}(T-N+1|T) = x_{MAP}(T - N + 1|T)$.

On the interval $[T - N + 2,\ T]$ the KKT equations for both MAP and MHE cases are identical, and are satisfied by the identical solutions. Q.E.D.