Optimal Estimate of Monotonic Trend with Sparse Jumps

Dimitry Gorinevsky

***Abstract*— This paper discusses a problem for recovering an underlying trend from noisy data. The key assumption is that the trend is monotonic, e.g., reflects accumulation of irreversible system deterioration. The trend is obtained as a maximum a posteriori probability estimate. The overall problem setup is related to α-β filter and Hodrick–Prescott filter. The main difference is that instead of a Gaussian process noise, a one-sided exponentially distributed noise is assumed. The batch estimate is a solution to a Quadratic Programming problem. The approach works exceptionally well for piece-wise linear trends that have a small number of jumps in the trended variable or its increase rate. Theoretical analysis justifies the sparsity properties for the jumps in the solution.**

I. INTRODUCTION

This paper considers trending of univariate time series. Simple or more sophisticated methods for trending are used in finance, economics, business, decision support, condition-based maintenance, industrial process monitoring, manufacturing automation, enterprize resource planning, medical, geological, weather, and environmental applications.

Optimal estimation of a trend requires a prior knowledge of a trend model. A broadly used model is a second-order random walk where Gaussian increments drive a slope state and an intercept state of a linear trend model. Known as a Holt model (in econometrics) or α-β model, such second-order model is also a mainstay of Kalman filtering in navigation systems. A batch estimation of the trend (smoothing problem) based on the second-order model is commonly used for financial and econometric analysis under the name of Hodrick-Prescott filter.

The problem statement in this paper was motivated by estimation applications where the trend reflects a physical property, such as damage, irreversibly accumulating with time. We consider trending based on second-order monotonic random walk model with monotonicity constraints. Similar to the α-β filter trending, this second order model might have many applications. While there is abundant research and textbook literature on 'usual' linear trending, very little work on monotonic trending seems to exist.

A statistically optimal estimate of a monotonic trend can be obtained by numerical optimization of a log-likelihood index with the monotonicity constraints. The problem can be formulated as Quadratic Programming (QP) and solved efficiently. An interesting property of the solution is that it usually comes out as a piecewise linear function with sparse jumps and inflection points. One can think of the jumps as discrete failure events in the damage accumulation history. We analyze this property in the paper and discuss its relationship to sparsity properties of L_1-optimal solutions of overdetermined linear equation systems, which have recently received significant attention in signal processing literature, see [4], [3], [5], [7], [11] for further references.

The initial motivation for our work came from turbine engine performance trending applications. Some of the recent work in this application area explicitly introduces monotonicity constraints in the estimation of the performance loss. The papers [13], [19] consider a single-step constrained estimation, which is locally optimal. They stop short of formulating an optimal estimation problem.

Optimal statistical estimation with monotonicity constraints has been known in statistics for some time under the name of Isotonic Regression. Much of the early work in this area is summarized in the books [1], [16], see also more recent papers [15], [18]. This prior work leads to QP problem formulations that could be also obtained for monotonic trend estimation based on a first-order random walk model. This is different from the second-order random walk model considered herein.

Some prior work in statistics is aimed at estimating monotonic curves, such as cumulative probability distributions. The monotone spline smoothing considered in [12] leads to a QP problem statement close to one considered in this paper (though it is not motivated by time series estimation). There is also substantial literature on constrained optimization-based estimation, in particular moving horizon estimation, see [8], [14], [20]. Yet this work does not focus on monotonic trending.

This paper extends an earlier conference paper [9], where trending based on the monotonic walk model was introduced. For related work (outside of the scope of this paper) see [17] and the references there. The goals and contribution of this paper are two-fold. First, it provides tutorial material on the monotonic trending. Second, it explains some of the properties of the monotonic trending, in particular, the sparsity of the jumps and inflection points in the trend solution.

II. LEAST SQUARES TRENDING

This section establishes a departure point of the study by briefly reviewing standard linear-quadratic trending methods. The problem with Gaussian noises in this section is well known and considered for tutorial purposes. The main reason for practical success of these methods is in their conceptual clarity. They are easy to implement and have a small number of well-understood parameters to set up and tune. Subsequently we consider other, less standard,

Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, gorin@stanford.edu.
This work was supported by the NSF GOALI Grant ECS-0529426

problem formulation by modifying the baseline problem of this section.

Consider a univariate data set Y representing an underlying trend X perturbed by a noise

$$Y = \{y(1), ..., y(N)\}, \quad (1)$$
$$X = \{x(1), ..., x(N)\}, \quad (2)$$

where $y(t)$ are scalars and $x(t) \in \Re^K$ are trend state vectors. As an illustrative example, Figures 1 and 2 show an underlying trend (dashed line) and simulated time series Y (dots) obtained by adding $[-1,\ 1]$ bounded pseudo-random numbers to the trend.

In what follows, we will study the problem of estimating the underlying trend X from the observed data Y. This requires assuming a model for trend X and a relationship between X and Y. Consider a linear model of the form

$$x(t+1) = Ax(t) + \xi(t), \quad (3)$$
$$y(t) = Cx(t) + e(t), \quad (4)$$

where $x(t) \in \Re^K$ is the trend state, $\xi(t) \in \Re^K$ is the state driving noise, $e(t) \in \Re$ is the measurement noise, $A \in \Re^{K,K}$, and $C \in \Re^{1,K}$. In particular, consider a second-order random walk model, where $K = 2$ and

$$A = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad (5)$$

where α is a positive scalar parameter. A free response of (5) describes a linear trend (an affine function of time). The state x_1 is the slope and x_2 is the intercept of the trend.

Assume that $\xi(t)$ and $e(t)$ in (3)–(4) are independent white Gaussian noises. The probability distributions for $x(1)$, $e(t)$, and $\xi(t)$ are

$$p_0(z) \sim N(x_0, \Xi_0), \quad p_e(z) \sim N(0, q), \quad (6)$$
$$p_\xi(z) \sim N(0, \Xi),$$

where the covariance q is a scalar, the initial state x_0 is a $\Re^2$ vector, and the covariances Ξ_0, Ξ are $\Re^{2,2}$ matrices.

A Maximum A posteriori Probability (MAP) estimate of the trend X can be obtained by maximizing the conditional probability $P(X|Y) \rightarrow \max$. Using the Bayes rule, the conditional probability can be expanded as

$$P(X|Y) = P(Y|X) \cdot P(X) \cdot C, \quad (7)$$

where $C = [P(Y)]^{-1}$ is independent of X. Since the noise values $e(t)$ in (4) for different t are independent and identically distributed as $p_e(x)$, we get

$$P(Y|X) = \prod_{t=1}^{N} p_e\left(y(t) - Cx(t)\right) \quad (8)$$

By applying the Bayes rule to equation (3) for each t and using the fact that $\xi(t)$ in (3) are independent with identical distributions $p_\xi(x)$ we get

$$P(X) = p_0\left(x(1)\right) \cdot \prod_{t=2}^{N} p_\xi\left(x(t) - Ax(t-1)\right) \quad (9)$$

The MAP estimate minimizes $L = -\log P(X|Y)$. Obtaining the conditional probabilities in (7)–(9) by using (3)–(4), (6), yields the MAP problem in the form

$$L = \frac{1}{2}\sum_{t=1}^{N} \|y(t) - Cx(t)\|^2 + \frac{1}{2}\|x(1) - x_0\|_{Q_0}^2 \quad (10)$$
$$+\frac{1}{2}\sum_{t=2}^{N} \|x(t) - Ax(t-1)\|_Q^2 \rightarrow \min,$$

where $\|\cdot\|_Q$ is a weighted 2-norm such that $\|x\|_Q^2 = x^T Q x$. The weighting matrices are $Q = q\Xi^{-1}$ and $Q_0 = q\Xi_0^{-1}$.

The solution to the quadratic minimization problem (10) gives an estimate of state history vector X_N that consists of stacked state vectors $x(t)$ and has length of $N \cdot K$ (length of $2N$ for the model, (3)–(4), (5), (6)). The estimated trend output $\hat{y}(t) = Cx(t)$ can be displayed and compared against the original data.

For large N, the solution to (10) does not depend on the initial condition x_0 and its weight Q_0. Since Ξ is diagonal, the second-order model (3)–(4) has two tuning parameters Q_1 and Q_2, where $Q = \mathrm{diag}\{Q_1,\ Q_2\}$.

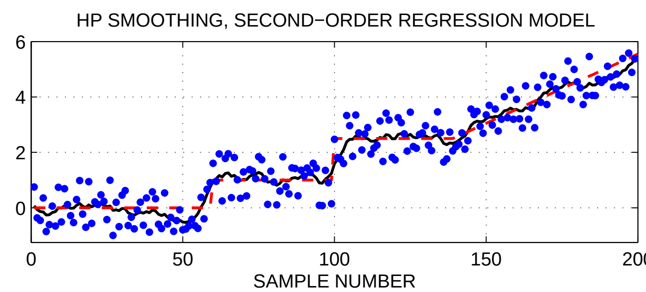

Fig. 1. Smoothing using the second-order trend model - H-P filter

A special case of the second-order smoothing algorithm (10) known as Hodrick–Prescott filter (H-P filter) is broadly used in econometrics. The H-P filter is a tool of choice for financial analysis (for stock valuation trending). For formulating H-P filter assume $\alpha = 1$, $Q_0 = 0$ and $Q = \mathrm{diag}\{Q_1,\ 0\}$. From (3) we get $\|x(t) - Ax(t-1)\|_Q^2 = Q_1 \cdot [x_1(t) - x_1(t-1)]^2$. In accordance with (5) we have $x_1(t) = x_2(t+1) - x_2(t)$. By using that in (10) we get a standard H-P filter formulation

$$L = \frac{1}{2}\sum_{t=1}^{N} \|y(t) - x_2(t)\|^2 + \quad (11)$$
$$\frac{Q_1}{2}\sum_{t=2}^{N-1} \|[x_2(t+1) - x_2(t)] - [x_2(t) - x_2(t-1)]\|^2 \rightarrow \min,$$

The smoothing results for the H-P filter with $Q_1 = 5$ are shown in Figure 1.

III. Monotonic trending

Monotonicity of a trend might be caused by accumulation of irreversible deterioration in a system. The state variable $x_2(t)$ in the monotonic trend model (3)–(4), (5) might have a meaning of the cumulative deterioration accumulated in the system. The state variable $x_1(t)$ describes the nondecreasing rate of the deterioration accumulation. One can say that $x_1(t)$ is a primary deterioration and $x_2(t)$ is a secondary deterioration, which is accumulated because of the presence of the primary, as well as because of other random causes. As an intuitive example, $x_1(t)$ might describe a quality of oil in the car (between oil changes) and $x_2(t)$ overall deterioration of the engine. The engine would deteriorate faster if the oil is bad. Engine deterioration and oil deterioration are considered to be monotonic (irreversible).

We model the monotonicity through probability distribution $p_\xi(x)$ of the process noise $\xi(t)$ assuming that $p_\xi(x) = 0$, for $x < 0$. The assumed $p_\xi(x)$ should be reasonably simple, justifiable, and lead to a solvable formulation. In what follows, we mainly consider a one-sided (positive) exponential distribution $p_\xi(x)$ that yields MAP formulation in the form of a QP (Quadratic Programming) problem. In addition to the exponential distribution, the QP formulation of the MAP problem can be also obtained for a uniform on an interval distribution and a Gaussian distribution with a positivity constraint. All of these assumptions result in similar QP problems.

Consider the second-order model (5). Assume that the process noises $\xi(t) = [\xi_1(t)\ \xi_2(t)]^T$ in (3) are exponentially distributed (always nonnegative). The model states, thus, evolve monotonically. The observation noise $e(t)$ in (4) is assumed to be an independent white Gaussian noise. The probability distributions (6) for $x(1) = [x_1(1) \quad x_2(1)]^T$, $e(t)$, and $\xi(t) = [\xi_1(t)\ \ \xi_2(t)]^T$ are replaced by

$$p_e(x) \sim N(0, q), \qquad p_0(x) \sim N(x_0, \Xi_0), \tag{12}$$
$$p_{\xi_j}(x) \sim \Xi_j e^{-x/\Xi_j} \quad \text{for } x \geq 0, \quad (j = 1, 2),$$

where Ξ_j and Ξ_j have the meaning similar to the process noise covariances in (6).

The MAP estimation problem is $L = -\log P(X|Y) \rightarrow \min$. Substituting the conditional probabilities in (7)–(9) from (3)–(4) and (12) we get, similar to how (10) is obtained,

$$L = \frac{1}{2}\sum_{t=1}^{N} \|y(t) - Cx(t)\|^2 + \frac{1}{2}\|x(1) - x_0\|^2_{Q_0} \tag{13}$$
$$+ \sum_{t=2}^{N} \bar{\rho}^T [x(t) - Ax(t-1)] \rightarrow \min,$$
$$\text{subject to } x(t) - Ax(t-1) \geq 0, \tag{14}$$

where A and C are given by (5), $\bar{\rho} = [q\Xi_1^{-1},\ q\Xi_2^{-1}]^T$, and $Q_0 = q\Xi_0^{-1}$. The loss index (13) includes linear and quadratic terms in components $x(t)$ of the decision vector X_N; the constraints (14) are linear inequalities. This QP problem is always feasible: $x(t) = A^{t-1}x_0$ is one feasible solution. Hence, a global optimum can be found using a standard QP solver. The solution depends on two regularization parameters $\rho_1 = q\Xi_1^{-1}$ and $\rho_2 = q\Xi_2^{-1}$.

To illustrate the performance of the algorithm (13)–(14), Figure 2 illustrates the trend estimated for the same data set as in Figure 1. As one can see, the proposed algorithm recovers the underlying trend very well. Compared to the trend estimated with an H-P filter in Figure 1, using the second-order monotonic walk model in Figure 2 yields about 40% better mean square error in recovering the underlying trend.

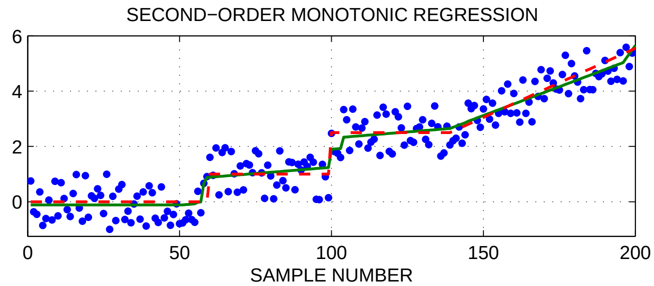

Fig. 2. Second-order monotonic regression for $\rho = [2,\ 5]^T$

Computational experiments show that the formulated monotonic trending algorithm can recover the jumps and inflection points in the underlying trend extremely well. In fact, the trend estimate in most practical cases comes out as having sparse jumps and inflection points. The reasons for the trend estimate having such form are discussed in the next section. When trending the underlying damage, the steps and inflection points can be interpreted as discrete failure events.

The tuning parameters of the algorithm: ρ_1 and ρ_2 are defined by the covariance q of the Gaussian observation noise $e(t)$ and the parameters Ξ_j of the exponential distribution for the process noise $\xi(t)$ in the model (3)–(4), (5). We consider ρ_1 and ρ_2 as tuning knobs of the algorithm and de-emphasize the statistical meaning of the parameters.

IV. Analysis for Monotonic Random Walk

As described in the previous section, a monotonic trend based on the random walk model can be computed by solving a QP problem. This section presents an analysis of the solution properties. The main issue of interest is: why does the solution comes out having a few jumps or inflection points with the straight line segments in-between? We will attempt to answer this question.

Introduce the model deviations $(t = 1,\ \ldots, N-1)$

$$u(t) = x(t+1) - Ax(t) \tag{15}$$

We will further use $u(t)$ as decision variables. The constraint (14) can be expressed as $u(t) \geq 0$. The variables $x(t)$ can be found from $u(t)$ by running a recursive update

$$x(t+1) = Ax(t) + u(t)$$

Let us introduce a decision vector U_N describing the trend (2) a vector Y_N describing the observations (1)

$$U_N = \begin{bmatrix} u(1) \\ \vdots \\ u(N-1) \end{bmatrix}, \qquad Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \tag{16}$$

We will show that the vector $U_N \in \Re^{2(N-1)}$ estimated from $Y_N \in \Re^N$ is sparse - has a relatively small number of nonzero components. Then, the estimated trend consists of straight line segments with sparse jumps (where $u_2(t) > 0$) and sparse inflection points (where $u_1(t) > 0$).

The optimal sparse solutions of overdetermined problems are of much interest in signal processing and there is significant past and current research work in this area, see [5], [7], [11] and references thereof. In [3], [4], [5], [7], [11] it is shown that under certain conditions, a sparse solution of overdetermined system of linear equations can be *exactly* reproduced by l_1 optimization. With a proper modification this property holds for noisy data. The problem (13)–(14) is equivalent to l_1-norm optimization with linear constraints and noisy data.

Our analysis of the problem (13)–(14) is inspired by [5], [7], [11]. The formulated results, however, are technically different, use different assumptions. This section considers the second-order random walk model (3)–(5), but the results can be generalized to higher-order random walk models.

For convenience of analysis, this section assumes that $\rho_1 = \rho_2 = \rho$ in (13). This can be assumed without a loss of generality. Consider a variable change $x_1 \to \beta x_1$, where $\beta = \rho_2/\rho_1 = \Xi_2/\Xi_1$. After this variable change the model (3)–(5) keeps the same form with a change of parameters: $\alpha \to \alpha/\beta$, $\Xi_1 \to \Xi_2$; in (13) we get $\rho_1 \to \rho_2$ and $\rho_2 \to \rho_2$.

Using (16), the first term in the loss index (13) can be presented in the form $\|Y_N - A_N U_N - B_N x(1)\|^2$, where

$$A_N = \begin{bmatrix} 0 & 0 & \dots & 0 \\ C & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2} & CA^{N-3} & \dots & C \end{bmatrix},$$

$$B_N = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix}, \quad (17)$$

For further analysis, let us re-write the problem (13)–(14) in the matrix form using notation (16), (17).

$$\frac{1}{2}\|Y_N - A_N U_N - B_N x(1)\|^2 + \frac{1}{2}\|x(1) - x_0\|^2_{Q_0} + \rho \mathbf{1}^T U_N \to \min, \quad (18)$$

$$\text{subject to } U_N \geq 0, \quad (19)$$

where $\mathbf{1}$ is a vector of ones of the length $2(N-1)$ and the inequality (19) is component-wise.

The QP problem (18)–(19) has a positive semidefinite Hessian. About a half of the Hessian eigenvalues are zero. We are interested in showing that solving this problem would yield a sparse solution. Following the analysis approach of [5], [7], [11] we notice that the problem (18)–(19) is equivalent to a problem

$$\mathbf{1}^T U_N \to \min, \quad (20)$$

$$\text{subject to } U_N \geq 0, \quad (21)$$

$$\|Y_N - A_N U_N - B_N x(1)\|^2 + \|x(1) - x_0\|^2_{Q_0} \leq \epsilon^2, \quad (22)$$

where ρ in (18)–(19) is an inverse of a Lagrange multiplier for the convex constraint (22) in (20)–(22). Larger ρ corresponds to smaller ϵ and vice versa.

Sparse solution in the absence of noise

Following the scheme of analysis in [5], [7], [11], consider first a no-noise case. Assume that the data Y_N is exactly generated by the trend model of the form (3)–(5) in the absense of noise such that

$$Y_N = A_N V_N + B_N v_0, \quad (23)$$

where the initial condition of the trend assumes a zero initial slope, $v_0 = [0 \; v_2]^T$ and $x_0 = v_0$.

For $\epsilon = 0$, (20)–(22) becomes an LP (Linear Programming) problem. The second constraint (22) changes into two equalities: $x(1) = v_0$ and

$$Y_N - A_N U_N - B_N v_0 = 0 \quad (24)$$

The overdetermined system of N equations (24) with $2(N-1)$ decision variables allows many solutions. The question is whether a sparse solution of this overdetermined system can be recovered exactly by solving the LP problem.

In [5], [7], [11] it is shown that under certain technical conditions a sparse solution of an overdetermined system of linear equations (which is a combinatorial complexity problem) can be found by solving an LP in a computationally efficient way. Though at the first glance these results seem to be related, they are not applicable to the problem (20), (24). The results of [3] require neighborliness of the system matrix - the property that does not hold for the problem in question. The results of [5], [7], [11] require low mutual coherence. This means cross-correlation between the columns of the system matrix being small, which is not the case for the Toelpitz impulse response matrix A_N (17).

In [5], [7], [11] and related work, sparsity is characterized through the number of nonzero components of the decision vector. If this number is small enough, the sparse solution can be recovered by solving an LP. To illustrate that this does not work for the problem (20), (24) consider a counterexample. Assume that the data is generated by the model of the form (3)–(5) with $u_1(t) = v_1(t)$ and $u_2(t) = v_2(t)$ corresponding to the vector V_N in (23) and $x(1) = 0$. Assume further that the only nonzero components of V_N in (23) are given by

$$v_1(\tau + 1) = \alpha, \quad v_2(\tau + 1) = 1,$$

for some $1 < \tau < N$. A different solution with $u_1(\tau) = 1$ and the rest of U_N components being zero has a smaller 1-norm: $\mathbf{1}^T U_N = 1$ while $\mathbf{1}^T V_N = 1 + \alpha$. The solutions V_N and U_N yield the same output $y(t) = (t - \tau - 1) \cdot \alpha$ for $t > \tau + 1$ and $y(t) = 0$ for $t \leq \tau + 1$. This means in the considered case the sparse solution (25) with only two non-zero components *is not* recovered exactly.

We will show that if V_N in (23) is sparse and satisfies certain additional conditions then solving the LP problem (20), (24) allows to recover the sparse solution exactly, i.e., $U_N = V_N$. The additional sparsity conditions introduced herein is that nonzero component indices in the underlying trend are never adjacent (are separated by gaps). This is quite different from the assumptions in [5], [7], [11].

The following result can be established.

Theorem 1: Consider the problem

$$\mathbf{1}^T U_N \to \min, \tag{25}$$

$$\text{subject to } Y_N = A_N U_N + B_N v_0, \;\; U_N \geq 0, \tag{26}$$

where A_N and B_N are given by (5), (17). Assume that (23) holds, where $v_0 = [0 \;\, v_2]^T$ and the components of vector V_N correspond to the underlying trend $v(t)$. Assume that for $j \neq k$, $v_j(t_1) > 0$ and $v_k(t_2) > 0$ simultaneously only if $|t_2 - t_1| > k \geq 1$. Assume further that in (5) we have $\alpha > 1$.

Then solving the LP problem (25)–(26) yields $U_N = V_N$ (and $u(t) = v(t)$).

Theorem 1 is proved in Appendix A. The proof is specific for the second order trend model (5), but could be possibly generalized for a higher order model (3)–(4).

Recovering sparse solution from noisy data

In practice, the data is always noisy and the question is whether the sparse solution could be recovered from such data. We will show that generally this is the case. Our formulation and analysis approach are related to that of [7]. The difference is that we are considering positivity constraint, while [7] considers a similar QP optimization problem with l_1 norm regularization penalty but without a positivity constraint. In [7] the Lagrange multipliers corresponding to bounded l_1 norm are limited by unity. The Lagrange multipliers for the positivity constraints do not have such bound.

Consider data Y_N generated by the trend model (3)–(5) and distorted by a bounded noise. Instead of (23) assume

$$Y_N = A_N V_N + B_N v_0 + \gamma e, \tag{27}$$

$$\|e\|_\infty \leq 1 \tag{28}$$

where $e \in \Re^N$ is the noise sequence with bounded terms and γ is the noise intensity parameter. We will show that for small enough γ and under some additional technical conditions the sparse solution $U_N = V_N$ could be reasonably reproduced from data (27)–(28) by solving (18)–(19).

In particular, the support of the solution U_N (the index set of nonzero components) is the same as of V_N and the solution error $\|U_N - V_N\|$ is bounded.

Theorem 2: Consider the QP problem (18)–(19) where Y_N is given by (27)–(28). Let V_N in (27) be the sparse solution V_N mentioned in the conditions of Theorem 1. Then positive γ_* and ρ_* exist such that for $0 \leq \gamma \leq \gamma_*$ and $0 \leq \rho \leq \rho_*$ solving this QP problem yields U_N with the following properties

1) The support set of the solution vector U_N (the set of nonzero component indexes) is the same as for V_N
2) $\|U_N - V_N\| = O(\gamma, \rho)$

Proof. First, note that for $\gamma = 0$ and sufficiently small ρ, the solution of the QP problem (18)–(19), (27)–(28) is given by $U_N = V_N$. Indeed, V_N achieves the minimum (zero) of the first two quadratic terms in (18). Theorem 1 says that the second term proportional to $\mathbf{1}^T U_N$ is also minimized, conditionally to the first two terms being zero. Since ρ is arbitrarily small, V_N achieves the overall minimum of (18)–(19). (For small enough ρ, an increase in the quadratic terms would always negate any decrease in the linear term).

Second, the problem (18)–(19), (27)–(28) for $\gamma > 0$, $\rho > 0$ can be presented in the form

$$\frac{1}{2} U_N^T Q U_N + (c^T + \rho c_\rho^T + \gamma c_\gamma^T) U_N \to \min \tag{29}$$

$$\text{subject to } U_N \geq 0, \tag{30}$$

where the matrices Q, c, c_ρ, and c_γ are independent of ρ and γ. To obtain (29)–(30), first substitute Y_N from (27)–(28) into (18)–(19); second, find $x(1)$ from unconstrained quadratic optimization of (18); and, third, substitute the optimal $x(1)$ (linearly dependent on U_N) back into (18).

The two perturbation parameters in (29) can be replaced by a single parameter λ such that

$$\rho = \lambda h_\rho, \quad \gamma = \lambda h_\gamma \tag{31}$$

where h_ρ and h_γ are positive constants.

A perturbed quadratic problem of the form (29)–(30), (31), was considered in [10] (and a few other papers cited there). The results of [10] prove that a closed invariancy interval $0 \leq \lambda \leq \lambda_u$ exists where the 'solution support' (a set of nonzero indices of the solution U_N) is the same as for $\lambda = 0$. Moreover, [10] proposes a computational procedure for finding an approximation to this interval.

On the invariancy interval, Property 1 in Theorem 2 holds by definition. Property 2 in Theorem 2 could be proved by computing the norm $\|U_N - V_N\|$ for projections on the solution support subspace (components of U_N and V_N outside of this subspace are zero anyways). The constraints are inactive on the solution support subspace. Thus, U_N can be estimated from linear equations obtained for unconstrained quadratic optimization on the solution support subspace. Since the perturbation parameters enter (29) linearly, Property 2 in Theorem 2 follows immediately. Q.E.D.

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] Boyd, S. and Vandenberge, L., *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004

[3] Donoho, D. L. and Tanner, J., "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci.*, Vol. 102, No. 27, 2005, pp.9446-9451.

[4] Donoho, D. L. and Elad, M., "Maximal sparsity representation via l_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, 2003, pp. 2197-2202.

[5] Donoho, D. L., Elad, M., and Temlyakov, V.N., "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Tran. on Information Theory*, Vol. 52, No. 1, 2006, pp. 6–18.

[6] Feng, G., "Data smoothing by cubic spline filter," *IEEE Tran. on Signal Processing*, Vol. 46, No. 10, 1998, pp. 2790–2796.

[7] Fuchs, J.J., "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Tran. on Information Theory*, Vol. 51, No. 10, 2005, pp. 3601–3608.

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

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

[10] Hadigheh, A.G., Mirnia, K., and Terlaky, T., "Sensitivity analysis in linear and convex quadratic optimization: Invariant active constraint set and invariant set intervals," *INFOR*, Vol. 44, No. 2, 2006, pp.129–155.

[11] Tropp, J.A., "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Tran. on Information Theory*, Vol. 52, No. 3, 2006, pp. 1030–1051.

[12] He, X. M., and Shi, P., "Monotone B-spline smoothing," *Journ. of American Statistical Association*, Vol. 93, No. 442, 1998, pp. 643–650.

[13] Mathioudakis, K., Kamboukos, Ph., and Stamatis, B., "Turbofan performance deterioration tracking using nonlinear models and optimization techniques," *Trans. of ASME, Journ. of Turbomachinery*, Vol. 124, No. 5, 2002, pp. 580–587.

[14] 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.

[15] 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.

[16] Robertson, T., Wright, F.T., and Dykstra, R.L., *Order restricted statistical inference*, New York: Wiley, 1988.

[17] 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.

[18] 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.

[19] Simon, D. and Simon, D. L., "Aircraft turbofan engine health estimation using constrained Kalman filtering," *ASME Journ. of Engineering for Gas Turbines and Power*, Vol. 127, No. 2, 2005, pp. 323-328.

[20] 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

Consider the differentiated output signal $Dy(t) = y(t) - y(t-1)$. Based on (3)–(5), the equality constraint (26) can be re-written as

$$Dy(t) = u_2(t) + \alpha \sum_{k=1}^{t-1} u_1(k) \tag{32}$$

Note that if the components $u(t)$ of the vector U_N in (26) satisfy (32), then $x_1(1)$ can be always chosen such that (26) is satisfied; $x(1)$ does not influence the loss index (25). Also $u_2(N)$, $u_1(N)$, and $u_1(N-1)$ do not influence the output y in (3)–(4) on the time interval $[1, N]$. The minimization in (25), together with nonnegativity constraint in (26) yield $u_2(N) = u_1(N) = u_1(N-1) = 0$. By summing up the equalities (32) we obtain

$$y(N) - y(1) = \mathbf{1}^T U_N + \sum_{k=1}^{N-1} w(k) u_1(k) \tag{33}$$

$$w(k) = \alpha(N-k+1) - 1 \tag{34}$$

Since by Theorem assumption $\alpha > 1$, $w(k) > 0$ is a decreasing sequence of positive weights.

Similar to (32) and (33), the sparse trend input $v(t)$ that generates the data satisfies

$$Dy(t) = v_2(t) + \alpha \sum_{k=1}^{t-1} v_1(k) \tag{35}$$

$$y(N) - y(1) = \mathbf{1}^T V_N + \sum_{k=1}^{N-1} w(k) v_1(k) \tag{36}$$

For each t, when $v_2(t) = 0$, comparing (32) and (35) yields

$$u_2(t) + \alpha \sum_{k=1}^{t-1} u_1(k) = \alpha \sum_{k=1}^{t-1} v_1(k), \tag{37}$$

where $u_2(t) \geq 0$ (recall the positivity constraints (14), (15)).

Let t_j be a jump time when $v_1(t_j - 1) > 0$ or $v_2(t_j) > 0$. In accordance with the sparsity conditions of the Theorem, $v_2(t_j + 1) = 0$ and $v_1(t_j) = 0$. The following chain of inequalities holds

$$\sum_{k=1}^{t_j-1} u_1(k) \leq \sum_{k=1}^{t_j} u_1(k) \leq \sum_{k=1}^{t_j-1} v_1(k) = \sum_{k=1}^{t_j} v_1(k) \tag{38}$$

where the first inequality holds because $u_1(t_j) \geq 0$, the last equality holds since $v_1(t_j) = 0$, and the middle inequality follows from $u_2(t_j + 1) \geq 0$ and (37). Together, (37) and (38) mean that the inequality

$$\sum_{k=1}^{t-1} u_1(k) \leq \sum_{k=1}^{t-1} v_1(k) \tag{39}$$

holds for any $t \in [1,\ N]$, whether t is a jump time or not.

Add the inequalities (39) with the weights $w(t+1) - w(t)$, for $t < N$ and the inequality for $t = N$ with the weight $w(N)$. In accordance with (34), these weights are positive. The addition yields

$$\sum_{k=1}^{N-1} w(k) u_1(k) \leq \sum_{k=1}^{N-1} w(k) v_1(k) \tag{40}$$

By comparing (33), (36), and (40) we get

$$\mathbf{1}^T U_N \geq \mathbf{1}^T V_N \tag{41}$$

Since U_N is an optimal solution to the LP problem (25)–(26) and V_N is a feasible solution, the opposite sign inequality must hold and we get $\mathbf{1}^T U_N = \mathbf{1}^T V_N$. This means the inequalities in the derivation chain must be equalities. In particular, (39) is equality, which means $u_1(k) = v_1(k)$. From that $u_2(k) = v_2(k)$ follows easily. Hence $U_N = V_N$. QED.