Weight and Time Recursions in Dynamic State Estimation Problem With Mixed-Norm Cost Function

Pavel Akimov and Alexander Matasov, *Member, IEEE*

***Abstract*—The mixed-norm cost functions arise in many applied optimization problems. As an important example, we consider the state estimation problem for a linear dynamic system under a nonclassical assumption that some entries of state vector admit jumps in their trajectories. The estimation problem is solved by means of mixed l_1/l_2-norm approximation. This approach combines the advantages of the well-known quadratic smoothing and the robustness of the least absolute deviations method. For the implementation of the mixed-norm approximation, a dynamic iterative estimation algorithm is proposed. This algorithm is based on weight and time recursions and demonstrates the high efficiency. It well identifies the rare jumps in the state vector and has some advantages over more customary methods in the typical case of a large amount of measurements. Nonoptimality levels for current iterations of the algorithm are constructed. Computation of these levels allows to check the accuracy of iterations.**

***Index Terms*—Estimation, linear systems, optimization, uncertain systems.**

I. Introduction

Some dynamic systems can admit the jumps of stepwise type (discontinuities) in their trajectories at certain time-instants. Such systems arise, e.g., in robotics [1], finance [2], epidemiology [3], [4], inertial navigation [5]. In discrete systems that are numerical models of continuous-time processes, these jumps are manifested in the form of abrupt changes in the average values of trajectories. We study the state estimation problem for this type systems. The traditional least squares estimation method (l_2-norm approximation) gives smoothed in time estimates of the jumps since, figuratively speaking, the l_2-norm approximation "avoids large residuals." To overcome this difficulty one can use mixed l_1/l_2-norm approximation. The point is that the jumps in trajectories can be interpreted as the presence of additional abnormal pulses in the right-hand sides of linear dynamic equations. It is well known that, for static systems with outliers, the least absolute deviations method (l_1-norm approximation) is an effective estimation algorithm [6]–[9]. Moreover, in image processing an l_1-norm total variation regularization term, which corresponds to a simplest dynamic model, allows to preserve the edge location (the analog of the jump location for multidimensional case) [10]–[12]. Therefore, in the cost function of l_2-norm approximation problem, the summands that can contain abnormal pulses should be replaced by l_1-norm terms. Thus we obtain the mixed-norm approximation problem for solving the state estimation problem for uncertain systems with abnormal input pulses. Such optimization problems also arise, e.g., in the Huber approach [13], compressive sensing [14] (LASSO method), l_1 trend filtering [15]; see also a brief overview in [16].

The presence of nonsmooth cost functions leads to computational difficulties, which are especially significant in processing large measurement data for the state estimation in dynamic systems. Conventional numerical methods can be divided into three groups.

The methods of the first group consist in the reduction to equivalent convex programming problems with smooth cost functions and inequality type constraints [8]. The disadvantage of this approach is that it requires large computational resources under processing the long lasting measurements since, in this case, the matrix that specifies the inequality constraints has a large dimension proportional to the amount of nonsmooth summands in the cost function.

An essential step forward in solving the convex optimization problems of high dimension is the so-called alternating direction method of multipliers (ADMM) [17]–[19]. The methods of this type form the second group. They involve an iterative search for a minimum of the Lagrange function for the original problem by parallel solving the problems of smaller dimension. However, one of the ADMM steps still requires the solution of an auxiliary quadratic problem, which is comparable to the original problem in the number of variables [20], [21]. For some relatively simple cases, this difficulty can be overcome [21]. But for a general case of linear dynamic systems, a direct use of ADMM does not allow to decompose the original problem in a required manner.

The third group of methods utilizes a transition to a sequence of approximating quadratic optimization problems [7], [22]. As is shown in the present paper, this approach can be modified for the case of state estimation in dynamic systems. We propose a numerical solving algorithm that consists of two nested iteration procedures. The first one is intended for constructing the auxiliary variational problems; the aim of the second one is to solve these problems by means of Kalman estimation methods [23], [24]. The idea for constructing the auxiliary approximating problems is to replace the absolute values in the cost function by special approximating quadratic functions. Such a replacement for the classical l_1-norm approximation (least absolute deviations method) is presented, e.g., in [7], [22], [25], [26] and is also utilized in [27], [28] for the solution of a geometric problem. The corresponding algorithm is called the method of variational-weighted approximation or the Weiszfeld algorithm. However, in the works mentioned above, this method is used heuristically, without examining the convergence or

Manuscript received November 28, 2013; revised May 29, 2014; accepted October 19, 2014. Date of publication October 29, 2014; date of current version March 20, 2015. This work was supported by the Russian Foundation for Basic Research (RFBR) under Grant 11-08-00004a. Recommended by Associate Editor M. Verhaegen.

The authors are with the Faculty of Mechanics and Mathematics, M.V. Lomonosov Moscow State University, Moscow 119991, Russia (e-mail: akmpavel@rambler.ru; alexander.matasov@gmail.com).

Digital Object Identifier 10.1109/TAC.2014.2365687

accuracy analysis. In the authors' book [29], the nonoptimality levels were constructed for the iterations of the Weiszfeld algorithm, which was applied to static estimation problems. In [5] and [29], we described the application of l_1-norm approximation to a dynamic system that arises in inertial navigation (by a reduction to a static optimization problem).

In the paper, we present new results. They consist in the elaboration of a novel iterative algorithm for solving the dynamic estimation problems with large measurement arrays by means of mixed-norm approximation and in the construction of nonoptimality levels for the iterations of estimation algorithm, taking into account the dynamic structure of the process. In this study, we employ an approach to the construction of approximating problems that is somewhat different from the approach in our previous works; this approach proves to be more efficient.

Let us discuss the link of our paper with some recent works [20], [30]–[38] devoted to nonsmooth optimization for data processing and compressive sensing. These articles in one way or another deal with various l_0/l_2- or l_1/l_2-norm optimization problems. In [38], a hierarchical probabilistic model generates a Bayesian sequential reweighted scheme. In [32], an l_1/l_∞-norm block is used. In the works mentioned above, the sparsity requirement is imposed primarily on the state vector (in [20], [33], [35], for both state vector and system noise). In our paper, the l_1-norm regularization term contains only system disturbances (as for some modifications in [37], where numerical methods were not discussed). The article [20] is most close to our study. It is devoted to robust smoothing of dynamic processes in the presence of rare outliers. A similar filtering problem for a special case is considered in [31]. In [20], [31], other models for system noises are employed by introducing the additional variables for outliers into the state vector and imposing the sparsity constraints on these outliers. The paper [20] differs from our investigation by a more detailed cost function and other numerical methods. In [33], the authors use a Bayesian approach to acquire a sparse signal; but their algorithm is based on a combination of the Kalman filter and a sequential importance sampling with resampling. Thus numerical difficulties are transferred to a complex Monte-Carlo sampling.

An ingenious approach to compressive sensing is presented in [30]. Instead of the l_1-norm (or l_p-norm) summand in the cost function, a fictitious nonlinear (containing absolute values) measurement is introduced. This can be done because the sparsity constraint is imposed on state vector. By means of the equality $|x| = x \cdot \operatorname{sign} x$ these nonlinear measurements are iteratively linearized and the solution of the original problem reduces to the extended Kalman filter type procedures. In our approach, despite a formal resemblance, we do not linearize absolute values in the cost function but replace them with specially weighted quadratic functions; therefore, we can utilize the exact Kalman smoother. Moreover, for our problem statement the approach from [30] is not applicable since we take the l_1-norm for the system disturbances and hence the corresponding sparsity constraint cannot be interpreted as pseudo-measurements of the state vector.

Thus, in contrast to [20], [30]–[38], we reduce the l_1/l_2-norm problem to a sequence of quadratic (l_2-norm) dynamic problems and, what is most important, we explicitly evaluate a nonoptimality level of a current solution at each iteration (see Sections II-C and IV below).

The paper is organized as follows. In Section II, we set the problem, describe the algorithm of weight and time recursions, and introduce the nonoptimality levels. In Section III, the dual problems are considered, which are necessary for constructing the nonoptimality levels. The formulas for the guaranteed nonoptimality level are derived in Section IV. In Section V, numerical examples are discussed. They confirm the efficiency of the proposed algorithm.

II. Approximation Problems

A. Mixed l_1/l_2-Norm Approximation

Consider a linear discrete dynamic uncertain system

$$x(k+1) = Fx(k) + Gq(k) + g(k), \qquad k = 0, \dots, K-1$$

where $x(k) \in \mathbb{R}^n$ is an unknown state vector at an instant k, $q(k) = (q_1(k), \dots, q_l(k))^T \in \mathbb{R}^l$ is an unknown input disturbance vector, $F \in \mathbb{R}^{n\times n}$,$G \in \mathbb{R}^{n\times l}$ are specified matrices, and $g(k) \in \mathbb{R}^n$ is a known vector of system nonhomogeneity.

Suppose $q(k)$, along with 'regular' high-frequency disturbances, can take abnormal values at certain rare time-instants, so that noticeable jumps arise in the trajectories $\{x(k)\}$. Assume for simplicity that the system matrices are constant. All statements and conclusions are valid for time-varying matrices as well.

Suppose the measurements are performed

$$z(k) = Hx(k) + r(k), \qquad k = 0, \dots, K.$$

Here $z(k) \in \mathbb{R}^m$ is the measurement vector, $H \in \mathbb{R}^{m\times n}$ is a given matrix, $r(k) = (r_1(k), \dots, r_m(k))^T \in \mathbb{R}^m$ is a measurement noise.

Also assume we have *a priori* information $\bar{x}(0)$ about the initial state: $\bar{x}(0) = x(0) + \bar{r}$, where $\bar{r} = (\bar{r}_1, \dots, \bar{r}_n)^T \in \mathbb{R}^n$ is an error vector. It is required to estimate the state $\{x(k)\}_{k=0}^K$ by measurements.

First consider a stochastic problem statement. Suppose for a moment that $\bar{r}_\alpha$, $r_\beta(k)$, and $q_\gamma(j)$ are independent random variables with zero means and known variances. Moreover, let $\bar{r}_\alpha$ and $r_\beta(k)$ be Gaussian variables: $\bar{r}_\alpha \sim \mathcal{N}(0, \sigma^2_{\Pi\alpha})$, $r_\beta(k) \sim \mathcal{N}(0, \sigma^2_{R\beta})$, and $q_\gamma(k)$ has the Laplace distribution density function

$$p_\gamma(y) = \frac{1}{\sqrt{2}\sigma_{Q\gamma}} \exp\left(-\frac{\sqrt{2}|y|}{\sigma_{Q\gamma}}\right)$$

where $\sigma_{\Pi\alpha}$, $\sigma_{R\beta}$, and $\sigma_{Q\gamma}$ are standard deviations. Then the maximum *a posteriori* (under given measurements) approach to estimation [24, pp. 389, 393] leads to the mixed l_1/l_2-norm approximation problem for estimating the state vector $\{x(k)\}_{k=0}^K$:[1]

$$\begin{aligned} \mathcal{I}_0 &= \min_{x,q} \mathcal{I}(x,q), \\ \mathcal{I}(x,q) &= \left\|\Pi^{-1}\left(\bar{x}(0) - x(0)\right)\right\|_2^2 \\ &\quad + \sum_{k=0}^{K} \left\|R^{-1}\left(z(k) - Hx(k)\right)\right\|_2^2 + \sum_{k=0}^{K-1} \left\|Q^{-1}q(k)\right\|_1 \end{aligned} \tag{1}$$

[1]Obviously, the conditional probability density function satisfies the relation $p(x(0), q|z) \propto \exp\{-(1/2)\mathcal{I}(x(x(0), q), q)\}$.

subject to

$$x(k+1) - Fx(k) - Gq(k) - g(k) = 0, \quad k = 0, \dots, K-1. \tag{2}$$

The weight matrices

$$\Pi = \mathrm{diag}(\Pi_1, \dots, \Pi_n), \quad R = \mathrm{diag}(R_1, \dots, R_m), \\ Q = \mathrm{diag}(Q_1, \dots, Q_l)$$

have the form

$$\Pi_\alpha = \sigma_{\Pi\alpha}, \quad R_\beta = \sigma_{R\beta}, \quad Q_\gamma = \frac{\sigma_{Q\gamma}}{2\sqrt{2}}. \tag{3}$$

Hereinafter we use the standard shorthand notation: $(x, q) = \{x(0), \dots, x(K), q(0), \dots, q(K-1)\}$; $\|y\|_1 = \sum_i |y_i|$, $\|y\|_2 = (\sum_i y_i^2)^{1/2}$, $\|y\|_\infty = \max_i\{|y_i|\}$.

The main advantage of stochastic problem setting is that it provides a rigorous justification for the cost function form. However, it should be emphasized that the stochastic approach is based on the implicit assumption that the statistical regularity holds. By this we mean a tendency of repeated experiments to result in the convergence of the averages as more and more trials are made [39, p. 17]. For many real engineering systems, this statistical regularity cannot be reliably verified; often it is a matter of faith. It is hard to verify the independence of noise values or the compliance with Laplace or Gaussian distribution. A detailed probabilistic model usually well motivates a resulting method but narrows its formal applicability.

Therefore, along with stochastic point of view, the noises are often assumed to be unknown deterministic functions (see, e.g., the classical books [8, pp. 292, 331, 334], [24, pp. 101, 387], [40]). The papers [10]–[12], [37] also exploit a deterministic framework. Since our main subject of study is convex optimization problem (1), (2) (which is the same for both approaches), the choice of formalization depends on the preferences of the reseacher. So, in the subsequent discussion we use a less burdensome deterministic framework without losing the essence and treat the weight coefficients Π_α, R_β, Q_γ as typical magnitudes (scales) of the variables $\bar{r}_\alpha$, $r_\beta(k)$, and $q_\gamma(j)$; these weights are chosen by the investigator. In this simplified and more practically oriented context, the noises are classified not according to whether they are Laplacian or Gaussian but according to whether they contain outliers or not (cf. [20, p. 4530]).

In the cost function, we take squared l_2-norm for the measurement residuals, and l_1-norm for the residuals in dynamics. This partition is caused by the assumption that the anomalous disturbances can arise in system dynamics (i.e., in components of $q(k)$) but not in the measurements. We will show in examples below that such approach is very effective for the identification of jumps in system dynamics.

Variational problem (1), (2) is similar to the variational problems with mixed norms that arise in compressive sensing. In fact, the presence of the term $\sum_0^{K-1} \|Q^{-1}q(k)\|_1$ in the cost function makes the q-component of the solution (x, q) sparse. This sparsity results in the fact that, according to (2), the x-component tends to consist of piecewise smooth trajectories with rare jumps, which is consistent with our initial desire to track the jumps precisely. From the mathematical point of view, the problem (1), (2) is a convex optimization problem with a nonsmooth cost function and linear constraints of equality type.

Usually, the methods for solving the nonsmoth problems of type (1), (2) are based on the reduction to smooth convex optimization problems with a large number of equality and inequality constraints. However this involves the operation with matrices of the order of $Kl \times (Kn + 2Kl)$ that, for large K (of the order of several thousands), requires considerable numerical resources. In many applied problems, including the processing of navigation data [41], [42], long lasting measurement series are typical. Therefore it is necessary to develop methods that exploit a dynamic nature of estimation problem and allow to avoid the reduction to static problems with high dimension matrices. One such approach for solving the mixed-norm problems is described in Section II-C.

B. Necessary Results From the Theory of Quadratic Problems

In the subsequent discussion we will essentially exploit the properties of quadratic smoothing problems (l_2-norm approximations) of the following form [23], [24]:

$$J_0 = \min_{x,q} \left(\left\|\Pi^{-1}\left(\bar{x}(0) - x(0)\right)\right\|_2^2 + \sum_{k=0}^{K} \left\|R^{-1}\left(z(k) - Hx(k)\right)\right\|_2^2 + \sum_{k=0}^{K-1} \left\|Q^{-1}q(k)\right\|_2^2 \right) \tag{4}$$

subject to the same constraints (2).

The effective numerical methods for solving the smoothing problem (4), (2) have long been developed; they are based on the Kalman estimation theory [24], [43]. Recall the recurrent relations that determine a solution of (4), (2).

Theorem 1 ([24]): Let $(x^*,\ q^*)$ be a solution for (4), (2). Then this solution is described by the following boundary-value problem:

$$\begin{aligned} x^*(k+1) &= Fx^*(k) + GQ^2G^T\lambda^J(k+1) + g(k), \\ q^*(k) &= Q^2G^T\lambda^J(k+1), \quad k = 0, \dots, K-1, \\ \lambda^J(k) &= F^T\lambda^J(k+1) + H^TR^{-2}\left(z(k) - Hx^*(k)\right), \\ & k = 0, \dots, K \end{aligned} \tag{5}$$

under the boundary conditions

$$\lambda^J(K+1) = 0, \qquad x^*(0) = \bar{x}(0) + \Pi^2\lambda^J(0).$$

Remark 1: Strictly speaking, in the book [24] Theorem 1 was proved for the case when $\bar{x}(0) = 0$, $g(k) = 0$. However, it can be easily verified that this statement directly follows from [24]. One should proceed to the centered process $\tilde{x}(k) = x(k) - \bar{x}(k)$, where the vectors $\bar{x}(k)$ are described by the equations

$$\bar{x}(k+1) = F\bar{x}(k) + g(k), \qquad k = 0, \dots, K-1$$

with specified initial values $\bar{x}(0)$.

Since the boundary values are given at opposite ends, formulas (5) are not suitable for direct solving the problem. To overcome this difficulty we can use the Bryson-Frazier formulas, which are presented in the following statement.

Theorem 2 ([24], [44]): The solution $(x^*,\ q^*)$ of problem (4), (2) is defined by the formulas

$$
\begin{aligned}
x^*(k) &= x^-(k) + P^-(k)\lambda^J(k), \quad k = 0,\dots,K\\
q^*(k) &= Q^2G^T\lambda^J(k+1), \quad k = K-1,\dots,0
\end{aligned}
$$

where $x^-(k), P^-(k)$ are the state vector estimate and the covariance matrix for Kalman filter, respectively

$$
\begin{aligned}
x^-(k+1) &= Fx^-(k) + K_p(k)\left(z(k) - Hx^-(k)\right) + g(k),\\
x^-(0) &= \bar{x}(0),\\
P^-(k+1) &= FP^-(k)F^T + GQ^2G^T\\
&\quad - K_p(k)\left(R^2 + HP^-(k)H^T\right)K_p^T(k),\\
P^-(0) &= \Pi^2,\\
K_p(k) &= FP^-(k)H^T\left(R^2 + HP^-(k)H^T\right)^{-1},\\
&\quad k = 0,\dots,K-1
\end{aligned}
$$

the vectors $\lambda^J(k)$ are defined by the backwards-time recursions

$$
\begin{aligned}
\lambda^J(k) = {} & (F - K_p(k)H)^T\lambda^J(k+1)\\
& + H^T\left(R^2 + HP^-(k)H^T\right)^{-1}\left(z(k) - Hx^-(k)\right),\\
& \lambda^J(K+1) = 0, \quad k = K,\dots,0.
\end{aligned}
$$

The Bryson–Frazier formulas allow to find the solution of smoothing quadratic problem in two passes. One more important aspect of numerical implementation is the application of the square-root method. The detailed description and analysis of the square-root method can be found in the books [24], [43].

C. Weight and Time Recursions for Mixed-Norm Problem

1) Approximation for Absolute Deviations: Here we propose a modification of the Weiszfeld algorithm (the algorithm of variational-weighted approximations). The application of this algorithm for the implementation of least absolute deviations method in static systems was described, e.g., in [7], [22], [25], [29]. The distinctive properties of the present study from these works are as follows: first, the algorithm is applied to the mixed norms; second, we use another quadratic approximation of absolute deviations; third, the auxiliary quadratic problems are dynamic problems and they are solved by the Bryson-Frazier formulas.

Describe the main idea of approximation. Let

$$
\left(x^{(s)}, q^{(s)}\right) = \{x(0,s),\dots,x(K,s),q(0,s),\dots,q(K-1,s)\}
$$

be an approximate solution of (1), (2) at the previous iteration with a number s. Functional (1) contains the l_1-norms of dynamic deviations $q(k)$: $\sum_{k=0}^{K-1}\sum_{i=1}^{l}Q_i^{-1}|q_i(k)|$. We replace the absolute deviations $|q_i(k)|$ by their quadratic approximations

$$
|q_i(k)| \approx \frac{1}{2}\,|q_i(k,s)| + \frac{1}{2}\frac{q_i^2(k)}{|q_i(k,s)|}. \tag{6}
$$

It is implicitly assumed that the closer is $q_i(k,s)$ to the correspondent optimal component, the more accurate is the approximation of absolute values. With (6), the following approximate equality holds: $\mathcal{I}(x,q) \approx \mathcal{J}(x,q,s) + c(s)$, where

$$
\begin{aligned}
\mathcal{J}(x,q,s) = {} & \left\|\Pi^{-1}\left(\bar{x}(0) - x(0)\right)\right\|_2^2\\
& + \sum_{k=0}^{K}\left\|R^{-1}\left(z(k) - Hx(k)\right)\right\|_2^2\\
& + \frac{1}{2}\sum_{k=0}^{K-1}\sum_{i=1}^{l}Q_i^{-1}\frac{q_i^2(k)}{|q_i(k,s)|},\\
c(s) = {} & \frac{1}{2}\sum_{k=0}^{K-1}\sum_{i=1}^{l}Q_i^{-1}\,|q_i(k,s)|\,.
\end{aligned} \tag{7}
$$

Thus, instead of the minimization of $\mathcal{I}(x,q)$, we should search a minimum of the approximate quadratic convex function (7). The last group of summands $c(s)$ does not depend on (x,q) and, under minimization, can be excluded from the cost function.

Remark 2: Our reasoning here are not rigorous and are aimed to clarify the motivation of the proposed algorithm. Besides, with $q_i(k,s) = 0$ our approximation is not valid; however, in this case, the regularization is applied [7] (see below).

Approximation (6) differs from that in the papers [22], [25] or in the book [29], where a more simple relation is used

$$
|q_i(k)| \approx \frac{q_i^2(k)}{|q_i(k,s)|}. \tag{8}
$$

A shortcoming of (8) is that, in a neighborhood of $(x^{(s)}, q^{(s)})$, the derivative of the approximating term differs significantly from original value. Indeed, let $q_i(k) \neq 0$, $q_i(k,s) \neq 0$, $q_i(k) \approx q_i(k,s)$; then $\partial|q_i(k)|/\partial q_i(k) = \operatorname{sign} q_i(k)$. It follows from (6) that

$$
\frac{\partial}{\partial q_i(k)}\left(\frac{1}{2}\,|q_i(k,s)| + \frac{1}{2}\frac{q_i^2(k)}{|q_i(k,s)|}\right) = \frac{q_i(k)}{|q_i(k,s)|} \approx \operatorname{sign} q_i(k)
$$

and it follows from (8) that:

$$
\frac{\partial}{\partial q_i(k)}\left(\frac{q_i^2(k)}{|q_i(k,s)|}\right) = 2\frac{q_i(k)}{|q_i(k,s)|} \approx 2\operatorname{sign} q_i(k).
$$

Therefore, formula (6) allows to approximate not only the values of absolute deviations but their derivatives as well. As it will be shown later, this fact substantially influences in the quality of nonoptimality levels for approximate solutions.

2) Iterative Algorithm: Let the pair

$$
\left(x^{(0)}, q^{(0)}\right) = \{x(0,0),\dots,x(K,0),q(0,0),\dots,q(K-1,0)\}
$$

be admissible for problem (1), (2); we consider this pair as an initial approximation of the unknown solution. For example, as the starting point $(x^{(0)},\ q^{(0)})$, one can take the solution of quadratic problem (4), (2).

In accordance with (7), we consider a sequence of l_2-norm approximation problems

$$
\begin{aligned}
\mathcal{J}_0(s+1) = \min_{x,q} \Bigg(& \left\|\Pi^{-1}\left(\bar{x}(0) - x(0)\right)\right\|_2^2\\
& + \sum_{k=0}^{K}\left\|R^{-1}\left(z(k) - Hx(k)\right)\right\|_2^2 + \frac{1}{2}\sum_{k=0}^{K-1}\left\|Q_W^{-1}(k,s)q(k)\right\|_2^2\Bigg)
\end{aligned} \tag{9}
$$

subject to (2), where $s = 0, 1, \dots$.

The matrices $Q_W^{-1}(k,s)$ are defined as follows:

$$Q_W^{-1}(k,s) = \mathrm{diag}\left(Q_{W1}^{-1}(k,s), \ldots, Q_{Wl}^{-1}(k,s)\right),$$
$$Q_{Wi}^{-1}(k,s) = \begin{cases} \left(Q_i^{-1}/\left|q_i(k,s)\right|\right)^{\frac{1}{2}}, & \text{if } |q_i(k,s)| > \alpha Q_i, \\ Q_i^{-1}/\alpha^{\frac{1}{2}}, & \text{if } |q_i(k,s)| \le \alpha Q_i, \end{cases}$$
$$i = 1, \ldots, l, \qquad k = 0, \ldots, K-1 \tag{10}$$

where $\{q_i(k,s)\}$ are the entries of the second vector of pair $(x^{(s)},\ q^{(s)})$. The parameter α in the formulas for $Q_{Wi}^{-1}(k,s)$ characterizes the smallness of residuals (deviations) and is used for regularization: if at any iteration the residuals become too small, i.e., $|q_i(k,s)| \le \alpha Q_i$, then the corresponding weight coefficient is fixed. Since the residuals can have different scales, the smallness of each summand is described by the values αQ_i.

A solution $(x^{(s+1)},\ q^{(s+1)})$ of quadratic problem (9), (2) for next iteration $s+1$ is found by means of the recurrent Bryson-Frazier formulas (see Theorem 2, which certainly is valid for nonstationary weight matrices as well) improved by the square-root method [24]. This recursiveness allows us to deal with vectors and matrices of considerably smaller dimension (of order n and $n \times n$, respectively) that substantially saves computational resources and computer memory compared with the reduction to static case, where we have to operate with an unknown parameter vector of dimension $n + Kl$ and the corresponding matrices of order $(n + (K+1)m + Kl) \times (n + Kl)$.

Hereafter we suppose that the optimal value of the quadratic problem cost function at each iteration $s+1$ is greater than zero: $\mathcal{J}_0(s+1) > 0$. In the opposite case, $\mathcal{J}_0(s+1) = 0$ and each summand in the quadratic cost function equals zero and, therefore, each summand in (1) also equals zero; so, there is no necessity to continue the calculation since the optimal solution has been already found. A solution of the quadratic problem obtained at each iteration is considered as an approximate solution for the original mixed-norm problem.

Thus we have two nested iteration processes: in the inner loop (parameterized by k), the solution of a current quadratic problem is found by means of Theorem 2; in the outer loop (parameterized by s), the auxiliary quadratic problems are formed basing on the previous step of this loop.

There is no rigorous result on the convergence of our algorithm. Hence we should find a way to monitor the accuracy of current iteration. We characterize the accuracy of an approximate solution of (1), (2) by the nonoptimality level [8], [29], [45]

$$\Delta \overset{\text{def}}{=} \mathcal{I}\left(x^{(s+1)}, q^{(s+1)}\right)/\mathcal{I}_0 \tag{11}$$

where $\mathcal{I}_0$ is the unknown optimal value of cost function (1) and $\mathcal{I}(x^{(s+1)},\ q^{(s+1)})$ is the cost function value at the current iteration with number $s+1$. Evidently, $\Delta \ge 1$, and the closer to unity is Δ, the more accurate is the approximate solution.

The nonoptimality level is unknown since the exact solution of (1), (2) is unknown. A constructive calculation of upper bounds for these nonoptimality levels will be our major theoretical issue in the paper (see Section IV below). Namely, if $\Delta \le \Delta_0$, where Δ_0 can be actually calculated and thus can be treated as a guaranteed nonoptimality level, and Δ_0 is fairly close to unity, then the iteration process can be halted and the solution $(x^{(s+1)},\ q^{(s+1)})$ at the final iteration can be considered as the solution of the original mixed l_1/l_2-norm problem. The proximity to unity is determined by a specified threshold Δ_{end}: $\Delta_0 \le \Delta_{\text{end}}$.

The structure of the algorithm is summarized below.

1: *Initialization.*
Set Π, Q, R, α, Δ_{end}.
Then $s := -1$, $Q_W(k,-1) := Q$. Go to 2.
2: *Inner forward and backward time recursions in* k.
Solve (9), (2) by Bryson-Frazier formulas (Theorem 2 from Section II). Get $(x^{(s+1)}, q^{(s+1)})$. Go to 3.
3: *Checking the accuracy.*
Calculate Δ_0 (Theorem 6 from Section IV).
If $\Delta_0 \le \Delta_{\text{end}}$, then STOP; otherwise go to 4.
4: *Updating.*
Set $s := s+1$.
Given $\{q^{(s)}(k)\}_{k=0}^{K-1}$ update $\{Q_W(k,s)\}_{k=0}^{K-1}$ in accordance with (10). Then go to 2 (outer weight recursion in s).

Since all conclusions concerning nonoptimality levels are valid for any iteration, we will not indicate the number of iteration s in the subsequent discussion and will simplify the notation

$$Q_W(i) \overset{\text{def}}{=} Q_W(i,s),$$
$$x^*(k) \overset{\text{def}}{=} x(k,s+1), \quad q^*(i) \overset{\text{def}}{=} q^*(i,s+1),$$
$$\mathcal{J}_0 \overset{\text{def}}{=} \mathcal{J}_0(s+1), \quad i = 0, \ldots, K-1, \quad k = 0, \ldots, K.$$

III. Dual Problems

A. Statement of Dual Problems

In order to construct the nonoptimality levels let us use the facts from the duality theory of convex variational problems [46], [47].

First of all, we represent the variational problems (1), (2) and (4), (2) in another form. Denote the vector of variables for these variational problems by

$$w = \left(x^T(0), x^T(1), \ldots, x^T(K), q^T(0), \ldots, q^T(K-1)\right)^T,$$
$$\dim w = (K+1)n + Kl \overset{\text{def}}{=} N.$$

Let us represent the residuals in functions (1) and (4) in the form of auxiliary vectors u and v

$$u_0 = \bar{x}(0) - x(0),\ u_z(k) = z(k) - Hx(k),\ v_q(j) = -q(j);$$
$$u = \left(u_0^T, u_z^T(0), \ldots, u_z^T(K)\right)^T,$$
$$\dim u = n + (K+1)m \overset{\text{def}}{=} M_u;$$
$$v = \left(v_q^T(0), \ldots, v_q^T(K-1)\right)^T, \quad \dim v = Kl \overset{\text{def}}{=} M_v.$$

Or, in matrix notation

$$\begin{pmatrix} u \\ v \end{pmatrix} = a - \Phi w,$$
$$a = \left(\bar{x}^T(0), z^T(0), z^T(1), \ldots, z^T(K), 0_{M_v\times 1}^T\right)^T \in \mathbb{R}^{M_u+M_v},$$
$$\Phi = \begin{pmatrix} \Phi_1 & 0_{(n+Km)\times M_v} \\ 0_{M_v\times(K+1)n} & E_{M_v\times M_v} \end{pmatrix} \in \mathbb{R}^{(M_u+M_v)\times N},$$
$$\Phi_1 = \begin{pmatrix} E_{n\times n} & 0_{n\times n} & \ldots & 0_{n\times n} \\ H & 0_{m\times n} & \ldots & 0_{m\times n} \\ 0_{m\times n} & H & \ldots & 0_{m\times n} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{m\times n} & 0_{m\times n} & \ldots & H \end{pmatrix}.$$

Constraints (2) also can be represented in matrix form: $b - \Psi w = 0$, where

$$\begin{aligned} b &= \left(g^T(0), g^T(1), \ldots, g^T(K-1)\right)^T \in \mathbb{R}^{Kn}, \\ \Psi &= (\Psi_1 \Psi_2) \in \mathbb{R}^{Kn\times N}, \\ \Psi_1 &= \begin{pmatrix} -F & E_{n\times n} & 0_{n\times n} & \ldots & 0_{n\times n} \\ 0_{n\times n} & -F & E_{n\times n} & \ldots & 0_{n\times n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n\times n} & \ldots & 0_{n\times n} & -F & E_{n\times n} \end{pmatrix}, \\ \Psi_2 &= \begin{pmatrix} -G & 0_{n\times l} & \ldots & 0_{n\times l} \\ 0_{n\times l} & -G & \ldots & 0_{n\times l} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n\times l} & 0_{n\times l} & \ldots & -G \end{pmatrix}. \end{aligned} \tag{13}$$

At last, introduce the vectors of weight coefficients and corresponding matrices

$$\begin{aligned} p_u &= (\tilde{\Pi}, \underbrace{\tilde{R}, \ldots, \tilde{R}}_{(K+1)m})^T, \quad p_v = (\underbrace{\tilde{Q}, \ldots, \tilde{Q}}_{Kl})^T, \\ P_u &= \operatorname{diag}(p_u), \quad P_v = \operatorname{diag}(p_v) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \tilde{\Pi} &= \left(\Pi_1^{-1}, \ldots, \Pi_n^{-1}\right), \quad \tilde{Q} = \left(Q_1^{-1}, \ldots, Q_l^{-1}\right), \\ \tilde{R} &= \left(R_1^{-1}, \ldots, R_m^{-1}\right). \end{aligned}$$

In new notation, problem (1), (2) can be written as

$$\mathcal{I}_0 = \inf_{u,v,w} \left(u^T P_u^2 u + \sum_{i=1}^{M_v} p_{vi}|v_i| \right) \tag{15}$$

subject to

$$a - \Phi w - \begin{pmatrix} u \\ v \end{pmatrix} = 0, \quad b - \Psi w = 0. \tag{16}$$

In its turn, problem (4), (2) is equivalent to

$$J_0 = \inf_{u,v,w} \left(u^T P_u^2 u + v^T P_v^2 v\right) \tag{17}$$

subject to (16).

To these problems assign dual problems [46], [47].

Theorem 3 ([8]): The dual problem to (17), (16) has the form

$$J^0 = \sup_{\mu_u,\mu_v,\lambda} \left(a^T \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} + b^T\lambda - \frac{1}{4}\left(\mu_u^T P_u^{-2}\mu_u + \mu_v^T P_v^{-2}\mu_v\right)\right) \tag{18}$$

subject to

$$\Phi^T \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} + \Psi^T\lambda = 0. \tag{19}$$

Furthermore, the duality relation holds: $J_0 = J^0$.

Here, $\mu_u \in \mathbb{R}^{M_u}$, $\mu_v \in \mathbb{R}^{M_v}$, $\lambda \in \mathbb{R}^{Kn}$ are the variables of dual problems.

Theorem 4: The dual problem to (15), (16) has the form

$$\mathcal{I}^0 = \sup_{\mu_u,\mu_v,\lambda} \left(a^T \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} + b^T\lambda - \frac{1}{4}\mu_u^T P_u^{-2}\mu_u \right) \tag{20}$$

subject to

$$\Phi^T \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} + \Psi^T\lambda = 0, \qquad \|P_v^{-1}\mu_v\|_\infty \le 1. \tag{21}$$

Furthermore, the duality relation holds: $\mathcal{I}_0 = \mathcal{I}^0$.

The proof is given in Appendix A.

Let us represent the Lagrange multipliers $\lambda \in \mathbb{R}^{Kn}$, $\mu_u \in \mathbb{R}^{M_u}$, and $\mu_v \in \mathbb{R}^{M_v}$ in block form (this form takes into account the dynamic structure of the initial problems)

$$\begin{aligned} \mu_u &= \left(\mu_0^T, \mu_z^T(0), \ldots, \mu_z^T(K)\right)^T, \\ \mu_v &= \left(\mu_q^T(0), \ldots, \mu_q^T(K-1)\right)^T, \\ \lambda &= \left(\lambda^T(1), \ldots, \lambda^T(K)\right)^T, \\ \mu_0 &\in \mathbb{R}^n, \quad \mu_z(k) \in \mathbb{R}^m, \quad \mu_q(j) \in \mathbb{R}^l, \quad \lambda(i) \in \mathbb{R}^n. \end{aligned}$$

Now turn back to initial notation in Theorem 4. Having written in explicit form the expressions (12)–(14) for matrices Φ, Ψ, P_u, P_v and vectors a, b, we get that (20), (21) is equivalent to the problem

$$\begin{aligned} \mathcal{I}^0 = \sup_{\mu_u,\mu_v,\lambda} \Bigg(&\bar{x}^T(0)\mu_0 + \sum_{k=0}^{K} z^T(k)\mu_z(k) + \sum_{k=0}^{K-1} g^T(k)\lambda(k+1) \\ &- \frac{1}{4}\left(\mu_0^T\Pi^2\mu_0 + \sum_{k=0}^{K}\mu_z^T(k)R^2\mu_z(k)\right)\Bigg) \end{aligned} \tag{22}$$

subject to the equality type constraints (here we introduce a fictitious Lagrange multiplier $\lambda(K+1) = 0$)

$$\begin{aligned} -F^T\lambda(1) + \mu_0 + H^T\mu_z(0) &= 0, \\ \lambda(k) - F^T\lambda(k+1) + H^T\mu_z(k) &= 0, \quad \lambda(K+1) = 0, \\ -G^T\lambda(k) + \mu_q(k-1) &= 0, \quad k = 1, \ldots, K, \end{aligned} \tag{23}$$

and, in addition, to the inequality type constraint

$$\max\left\{\|Q\mu_q(k)\|_\infty\right\}_{k=0}^{K-1} \le 1. \tag{24}$$

Similarly, from Theorem 3 we can obtain that the problem (18), (19) [dual to the l_2-norm approximation (17), (16)] is equivalent to the problem

$$\begin{aligned} J^0 = \sup_{\mu_u,\mu_v,\lambda} \Bigg(&\bar{x}^T(0)\mu_0 + \sum_{k=0}^{K} z^T(k)\mu_z(k) + \sum_{k=0}^{K-1} g^T(k)\lambda(k+1) \\ &- \frac{1}{4}\left(\mu_0^T\Pi^2\mu_0 + \sum_{k=0}^{K}\mu_z^T(k)R^2\mu_z(k) + \sum_{k=0}^{K-1}\mu_q^T(k)Q^2\mu_q(k)\right)\Bigg) \end{aligned} \tag{25}$$

subject to (23).

Problem (9), (2) differs from (4), (2) by weight coefficients only (we substitute $(1/2)Q_W^{-2}(k)$ for Q^{-2}). Clearly, all Theorems 1–4 are valid for nonstationary (time-variant) weight matrices $Q(k)$ and $R(k)$ as well. The structure of the proofs remains completely unchanged, only the formulas become even more cumbersome. Therefore, in accordance to Theorem 3 and (25), the dual problem to (9), (2) has the form

$$\begin{aligned} \mathcal{J}^0 = \sup_{\mu_u,\mu_v,\lambda} \Bigg(&\bar{x}^T(0)\mu_0 + \sum_{k=0}^{K} z^T(k)\mu_z(k) + \sum_{k=0}^{K-1} g^T(k)\lambda(k+1) \\ &- \frac{1}{4}\Bigg(\mu_0^T\Pi^2\mu_0 + \sum_{k=0}^{K}\mu_z^T(k)R^2\mu_z(k) \\ &+ 2\sum_{k=0}^{K-1}\mu_q^T(k)Q_W^2(k)\mu_q(k)\Bigg)\Bigg) \end{aligned} \tag{26}$$

subject to (23).

B. Solving the Quadratic Dual Problem

By applying the standard methods of the theory of convex variational problems, we can establish the relationship between the solutions of (26), (23) and (9), (2). In contrast to nonsmooth problem (22)–(24) the solution of (26), (23) can be found in explicit form.

Theorem 5: Let $(x^*,\ q^*)$ be the solution of (9), (2); then the solution of (26), (23) is defined by the following relations:

$$\begin{aligned} \mu_0^* &= 2\Pi^{-2}\left(\bar{x}(0) - x^*(0)\right), \\ \mu_z^*(k) &= 2R^{-2}\left(z(k) - Hx^*(k)\right), \quad k = 0, \ldots, K, \\ \mu_q^*(k) &= -Q_W^{-2}(k) q^*(k), \\ \lambda^*(k+1) &= -2\lambda^J(k+1), \quad k = 0, \ldots, K-1 \end{aligned}$$

where the vectors $\lambda^J(k)$ are found from (5) of Theorem 1

$$\lambda^J(k) = F^T\lambda^J(k+1) + H^T R^{-2}\left(z(k) - Hx^*(k)\right),$$
$$\lambda^J(K+1) = 0, \quad k = 0, \ldots, K.$$

The proof is given in Appendix B.

IV. Guaranteed Nonoptimality Level

In this section, we obtain one of the main results of the paper: a formula for the guaranteed nonoptimality level of a current iteration of the algorithm proposed in Section II-C.

Theorem 6: Let (x^*, q^*) be the solution of the l_2-norm approximation problem (9), (2). Suppose the optimal value of the cost function, $\mathcal{J}_0$, does not equal zero; then $\Delta \le \Delta_0$ and the guaranteed nonoptimality level Δ_0 is determined by the formula

$$\Delta_0 = \mathcal{I}(x^*, q^*) \cdot \left(-\theta_2 \min\left\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\right\}^2 + 2\mathcal{J}_0 \min\left\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\right\}\right)^{-1}$$

where the values $\mathcal{I}(x^*, q^*)$, $\mathcal{J}_0$, θ_2, and θ_∞ are defined by the equalities

$$\begin{aligned} \mathcal{I}(x^*, q^*) &= \left\|\Pi^{-1}\left(\bar{x}(0) - x^*(0)\right)\right\|_2^2 + \sum_{k=0}^{K}\left\|R^{-1}\left(z(k) - Hx^*(k)\right)\right\|_2^2 + \sum_{k=0}^{K-1}\left\|Q^{-1}q^*(k)\right\|_1, \\ \mathcal{J}_0 &= \left\|\Pi^{-1}\left(\bar{x}(0) - x^*(0)\right)\right\|_2^2 + \sum_{k=0}^{K}\left\|R^{-1}\left(z(k) - Hx^*(k)\right)\right\|_2^2 + \frac{1}{2}\sum_{k=0}^{K-1}\left\|Q_W^{-1}(k)q^*(k)\right\|_2^2, \\ \theta_2 &= \left\|\Pi^{-1}\left(\bar{x}(0) - x^*(0)\right)\right\|_2^2 + \sum_{k=0}^{K}\left\|R^{-1}\left(z(k) - Hx^*(k)\right)\right\|_2^2, \\ \theta_\infty &= \max\left\{\left\|QQ_W^{-2}(k)q^*(k)\right\|_\infty\right\}_{k=0}^{K-1}. \end{aligned}$$

Proof of Theorem 6: Consider a current iteration over s of the algorithm of weight and time recursions. Let the solution of (9), (2) at this iteration be $(x^*,\ q^*)$. In order to evaluate the nonoptimality level (see (11) from Section II-C2) from above, let us construct an estimate for the unknown optimal value $\mathcal{I}_0$ from below. By virtue of Theorem 4, $\mathcal{I}_0 = \mathcal{I}^0$, and if $\mathcal{I}^0$ is the maximal value of function (22), then, for any set $\lambda'(k), \mu_0', \mu_z'(k), \mu_q'(k)$ that satisfies constraints (23) and (24), the following inequality holds:

$$\mathcal{I}_0 = \mathcal{I}^0 \ge \bar{x}^T(0)\mu_0' + \sum_{k=0}^{K} z^T(k)\mu_z'(k) + \sum_{k=0}^{K-1} g^T(k)\lambda'(k+1) - \frac{1}{4}\left(\mu_0'^T\Pi^2\mu_0' + \sum_{k=0}^{K}\mu_z'^T(k)R^2\mu_z'(k)\right).$$

Note that the original nonsmooth problem (1), (2) and its approximating quadratic problem (9), (2) are related by construction. So, one might expect that the dual problems (22)–(24) and (26), (23) are also related. Indeed, the equality constraints for both dual problems are identical; the cost functions coincide up to the terms with $\mu_q(k)$, which entries are bounded either implicitly [by cost function (26)], or explicitly [by constraint (24)]. Therefore we apply the following approach [8], [29], [45]. We estimate $\mathcal{I}_0$ from below on a lesser [than in problem (22)–(24)] one-dimensional set. Namely, we search a maximal value of the cost function along the direction defined by the solution of (26), (23)

$$\lambda'(i+1) = \sigma\lambda^*(i+1),$$
$$\mu_0' = \sigma\mu_0^*, \quad \mu_q'(i) = \sigma\mu_q^*(i), \quad i = 0, \ldots, K-1,$$
$$\mu_z'(k) = \sigma\mu_z^*(k), \quad k = 0, \ldots, K$$

where a scalar parameter σ can vary and the vectors $\lambda^*(i+1)$, μ_0^*, $\mu_z^*(k)$, $\mu_q^*(i)$ define the solution of (26), (23) (they are presented by Theorem 5). Then we obtain an estimate for $\mathcal{I}^0$ from below

$$\mathcal{I}^0 \ge \max_\sigma\left(\sigma\left(\bar{x}^T(0)\mu_0^* + \sum_{k=0}^{K} z^T(k)\mu_z^*(k) + \sum_{k=0}^{K-1} g^T(k)\lambda^*(k+1)\right) - \frac{\sigma^2}{4}\left(\mu_0^{*T}\Pi^2\mu_0^* + \sum_{k=0}^{K}\mu_z^{*T}(k)R^2\mu_z^*(k)\right)\right) \quad (27)$$

where the σ is restricted by the inequality

$$|\sigma| \cdot \max\left\{\left\|Q\mu_q^*(k)\right\|_\infty\right\}_{k=0}^{K-1} \le 1. \quad (28)$$

(Obviously, constraints (23) hold.)

By virtue of the duality relation, $\mathcal{J}_0 = \mathcal{J}^0$. Since $\{\mu_0^*, \mu_z^*(0), \ldots, \mu_z^*(K),\ \mu_q^*(0), \ldots, \mu_q^*(K-1)\}$ is (a part of) the solution of (26), (23), we have

$$\mathcal{J}_0 = \mathcal{J}^0 = \bar{x}^T(0)\mu_0^* + \sum_{k=0}^{K} z^T(k)\mu_z^*(k) + \sum_{k=0}^{K-1} g^T(k)\lambda^*(k+1) - U$$

where

$$U = \frac{1}{4}\left(\mu_0^{*T}\Pi^2\mu_0^* + \sum_{k=0}^{K}\mu_z^{*T}(k)R^2\mu_z^*(k) + 2\sum_{k=0}^{K-1}\mu_q^{*T}(k)Q_W^2(k)\mu_q^*(k)\right).$$

The substitution of μ_0^*, $\mu_z^*(k)$, $\mu_q^*(i)$ from Theorem 5 into U yields

$$U = (\bar{x}(0) - x^*(0))^T\,\Pi^{-2}\,(\bar{x}(0) - x^*(0)) + \frac{1}{2}\sum_{k=0}^{K-1} q^{*T}(k)Q_W^{-2}(k)q^*(k) + \sum_{k=0}^{K}(z(k) - Hx^*(k))^T\,R^{-2}\,(z(k) - Hx^*(k)) = \mathcal{J}_0.$$

Consequently

$$\bar{x}^T(0)\mu_0^* + \sum_{k=0}^{K} z^T(k)\mu_z^*(k) + \sum_{k=0}^{K-1} g^T(k)\lambda^*(k+1) = 2\mathcal{J}_0.$$

Similarly

$$\frac{1}{4}\left(\mu_0^{*T}\Pi^2\mu_0^* + \sum_{k=0}^{K}\mu_z^{*T}(k)R^2\mu_z^*(k)\right) = (\bar{x}(0) - x^*(0))^T\,\Pi^{-2}\,(\bar{x}(0) - x^*(0)) + \sum_{k=0}^{K}(z(k) - Hx^*(k))^T\,R^{-2}\,(z(k) - Hx^*(k)) \overset{\text{def}}{=} \theta_2.$$

When substituting the expression for $\mu_q^*(k)$ into (28), this inequality takes the form

$$|\sigma| \le \frac{1}{\theta_\infty}, \quad \theta_\infty = \max\left\{\left\|QQ_W^{-2}(k)q^*(k)\right\|_\infty\right\}_{k=0}^{K-1}. \quad (29)$$

Thus inequality (27) can be rewritten in the form

$$\mathcal{I}^0 \ge \max_\sigma\left(-\theta_2\sigma^2 + 2\mathcal{J}_0\sigma\right) \quad (30)$$

subject to (29).

Obviously, the quadratic function $-\theta_2\sigma^2 + 2\mathcal{J}_0\sigma$ under constraint (29) attains its maximal value at $\sigma = \min\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\}$ and consequently, due to the duality relation

$$\mathcal{I}_0 = \mathcal{I}^0 \ge -\theta_2\min\left\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\right\}^2 + 2\mathcal{J}_0\min\left\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\right\}. \quad (31)$$

Note that θ_2 and θ_∞ are not both zero; otherwise all residuals equal zero and, hence, $\mathcal{J}_0 = 0$, which contradicts the second condition of Theorem 6. Therefore, with the convention $1/0 = \infty$ and $\min\{1, \infty\} = 1$, the expression $\min\{\frac{\mathcal{J}_0}{\theta_2}, \frac{1}{\theta_\infty}\}$ is well-defined even for the exotic cases $\theta_2 = 0$ or $\theta_\infty = 0$.

Finally, Theorem 6 follows from the definition of nonoptimality level (11) and inequality (31).

It should be emphasized that the expected similarity of the solutions for (22)–(24) and (26), (23) is not necessary for the validity of Theorem 6. This expectance only motivates the choice of the successful direction in one-dimensional problem. The proposed choice proves to be useful in practice: the nonoptimality levels are estimated quite accurately. □

Theorem 6 extends the range of applicability of nonoptimality levels to dynamic problems with mixed norms. Therein lies an important distinction from the authors' book [29], where the nonoptimality levels were constructed for static least absolute deviations problems. Most of our arguments are valid for a similar situation when the cost function contains the absolute values of measurement residuals and the squares of dynamic disturbances $q(k)$.

V. Numerical Experiments

In this section, we consider the results of numerical experiments. Note that our approach, which was described above, can also be applied to the simpler l_1-norm approximation problem. A brief (and without proofs) sketch of this application is presented in the authors' report [48]. In this compendious report, only the l_1-norm approximation was studied with simple approximation (8). Below we will also compare l_1/l_2-norm approximation and l_1-norm approximation. Besides, we will show that (6) in the mixed-norm problem is much more successful than (8) for the construction of nonoptimality levels.

A. Numerical Examples

Consider a simple dynamic system of the form

$$x(k+1) = \begin{pmatrix} 1 & 0.04 \\ 0 & \delta \end{pmatrix} x(k) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} q(k), \quad \delta = 1, \quad k = 0, \ldots, K-1. \quad (32)$$

Here, $x(k)$, $q(k) \in \mathbb{R}^2$, $g(k) = 0$. We set the state estimation problem for the vectors $x(k)$ by the measurements $z(k)$ and by the prior information $\bar{x}(0)$ about the initial state, where

$$z(k) = x_1(k) + r(k), \qquad k = 0, \ldots, K \quad (33)$$

and

$$0 = \bar{x}(0) = x(0) + \bar{r}.$$

In the numerical experiments, an additional anomalous pulse in $q_2(k)$ (which is equal to -2.5) was modeled. So, a jump in the average values of $x_2(k)$ was generated. The arrays of $x(k)$ and $z(k)$ were formed in accordance with (32) and (33). The amount of discrete time-instances K in this example is equal to 3600. The error $\bar{r}$, the measurement noise $r(k)$, and the high-frequency components of $q(k)$ (without regard for the additional pulse) were simulated by means of Gaussian random number generator. In our experiments

$$\sigma_{\bar{r}1} = \sigma_{\bar{r}2} = 1, \quad \sigma_r = 3, \quad \sigma_{q1} = 0.2, \quad \sigma_{q2} = 0.6.$$

We search the estimates of $x(k)$ as a solution of (1), (2), where the weight coefficients and the regularization parameter are chosen as follows:

$$\Pi_1 = \Pi_2 = 1, \quad R = 3, \quad Q_1 = 0.2,\ Q_2 = 0.6, \quad \alpha = 10^{-3}. \quad (34)$$

Here $n = l = 2$, $m = 1$ and the amount of unknown quantities equals $(K+1)n + Kl = 14402$; the amount of residual entries [summands in the function (1)] equals $n + (K+1)m + Kl = 10803$.

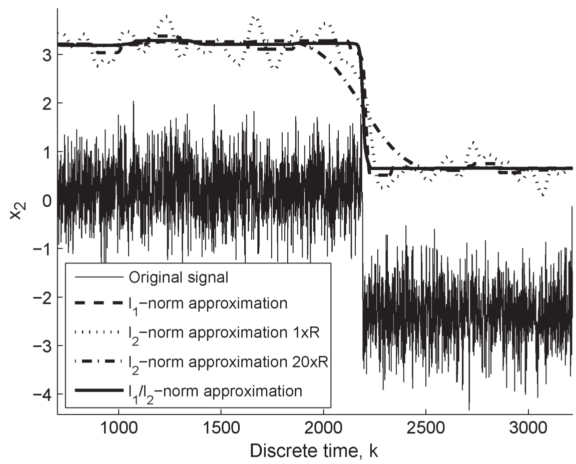

Fig. 1. Estimates of the jump in $x_2(k)$.

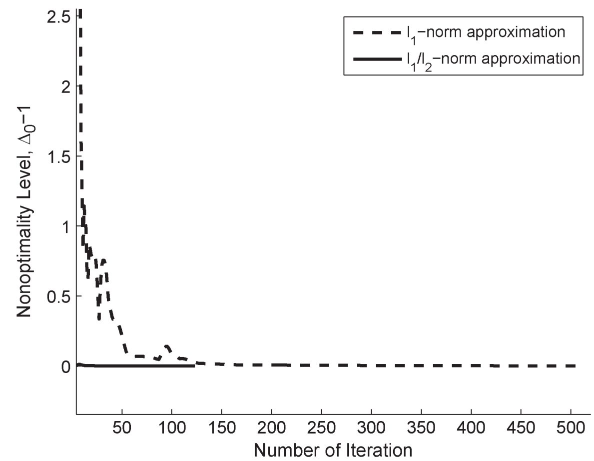

Fig. 2. Normalized guaranteed nonoptimality levels $\Delta_0 - 1$.

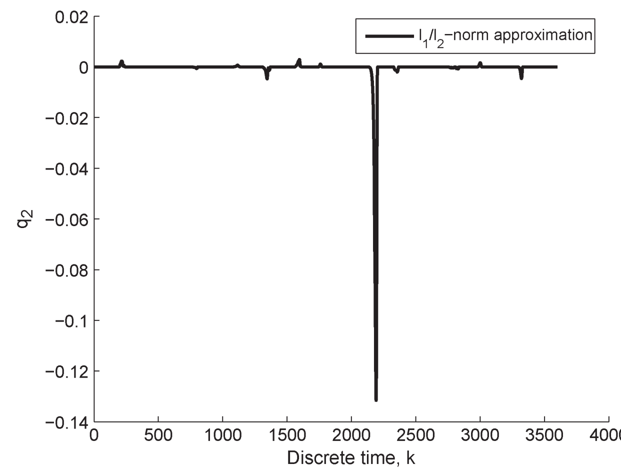

Fig. 3. Estimate of the disturbance $q_2(k)$.

In what follows, we focus on such qualitative features of $x(k)$ as jumps.[2] The high-frequency components of $x(k)$ represent less interest. Moreover, it is clear that any reasonable estimation algorithm will average the high-frequency components of $x(k)$; so, our attention will be centred on estimating the low-frequency components of $x(k)$.

Consider four methods for estimating the state vector (see Fig. 1): the l_1-norm approximation, i.e., the least absolute deviations method [48] (dashed line in the graphs below); the l_2-norm approximation, i.e., the standard least squares method (4), (2) (dotted line); the mixed-norm approximation from Section II-C (bold solid line); the standard l_2-norm approximation (4), (2) with a twentyfold increase of weight coefficient R (dash-dot line). For the graph of the original signal not to overlap with the other curves in Fig. 1, the graphs for estimates of $x_2(k)$ are artificially shifted up by 3.0 (the same is true for other figures with the original signal). All quadratic problems are solved by the Bryson-Frazier formulas; the mixed-norm problem and the l_1-norm problem are solved by the weight and time recursions algorithm described above (with the corresponding modification for l_1-norm presented in [48]). At each iteration over s the guaranteed nonoptimality level was calculated; the threshold Δ_{end} for stopping criterion $\Delta_0 \leq \Delta_{\mathrm{end}}$ was chosen equal to $1 + 10^{-3}$.

First let us discuss the estimates obtained by means of the traditional quadratic smoothing problem (dotted and dash-dot lines). We see from Fig. 1 that, with the initial value of R [cf. (34)], the estimate of the jump is slightly blurred (in time) and contains noticeable inadequate fluctuations, which correspond neither to the average values of $x_2(k)$ nor to the high-frequency component of $x_2(k)$. With a twentyfold increase of R (which is practically equivalent to a twentyfold decrease of Q), the estimate of the jump is more stable but it becomes too distorted and, therefore, the instant of jump is determined quite inaccurately. Besides, the estimates are very sensitive to the choice of weight coefficients. So, the traditional least squares algorithm cannot be well adjusted for estimating the jumps.

It follows from Fig. 1 that the l_1-norm approximation and the l_1/l_2-norm approximation are much more preferable for the identification of jumps. Moreover, they are quite close. Additional numerical experiments indicate that the l_1-norm and l_1/l_2-norm approximations are considerably less sensitive to the variation of weight coefficients than l_2-norm approximation. The difference between l_1-norm and l_1/l_2-norm approximations manifests itself in evaluating the efficiency of computational procedures. Let us examine this question in detail.

The guaranteed nonoptimality level as a function of s is presented in Fig. 2. With l_1-norm and l_1/l_2-norm approximations, the solution is found in 508 and in 123 iterations over s, respectively. The computational process takes 358 s in the first case, and 74 s in the second case (for the classical l_2-norm approximation, 0.5 s). It should be emphasized that, for the amount of variables mentioned above, the reduction to static problems with inequality/equality constraints does not solve the nonsmooth problems at all because of the lack of computational resources. At the same time, the algorithm of weight and time recursions solves the mixed-norm problem in case of large measurement arrays.

The estimate for the disturbance $q_2(k)$ is shown in Fig. 3. As expected, the sparsity of this estimate is distinctly visible. In enlarged scale, the graph in Fig. 3 in fact contains a series of unidirectional pulses; the total sum of these pulses is equal to -2.5, which is the true value of the pulse in $q_2(k)$.

[2]The discrete systems of equations are supposed to be discrete models of continuous-time processes; so, it is clear what is meant by "jump".

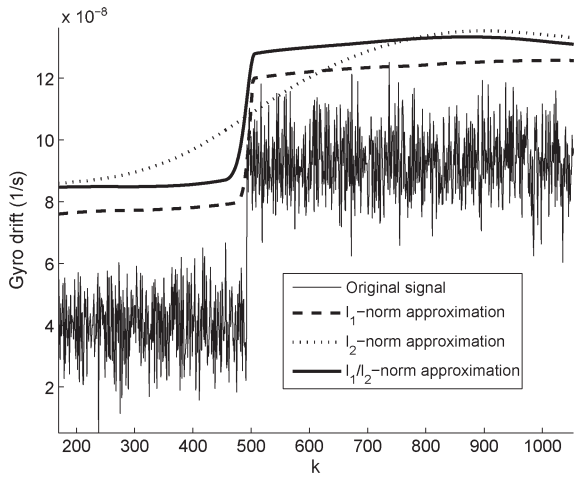

Fig. 4. Estimates of gyro drift.

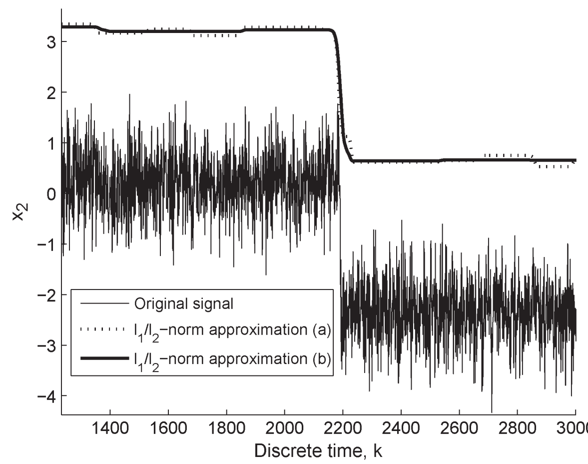

Fig. 5. Estimates of $x_2(k)$ for two approximation methods.

We could also consider other examples with different values of δ in (32) when the average values of $x_2(k)$ in system (32) are not exactly stepwise as before. All mentioned above conclusions hold true.

The regularities that are observed for these simplest cases are valid for more nontrivial issues with several jumps. In particular, our approach also was applied for the identification of rare stepwise jumps in the sensors (accelerometers and gyros) of a strapdown inertial navigation system (SDINS) under bench testing. This problem was set to us by engineers [5], [29], [48]. Here, the system consists of the error equations for SDINS supplemented with the measurements of velocity. The dimensions of state vector, measurements, and disturbances are equal to $n = 8, m = 2$, and $l = 4$, respectively. The amount of unknown variables is of order 20 000, and the number of summands in the cost function is of order 10 000. In contrast to the previous example, the jumps in state vector less obviously affect the output signal $z(k)$. Nevertheless, the use of mixed norms in this case is also noteworthy: the estimates of the jumps are crisp and the accuracy of estimates for the jumps is the same as for l_1-norm approximation [48]; moreover, the duration of computation is 3–6 times less. This is a typical feature for other examples: the computational burden for problem (1), (2) is several (3–6) times smaller than for the pure l_1-norm approximation. In navigation problem, the advantage of l_1/l_2-norm approximation over l_2-norm approximation in the accuracy of identification of jumps is much more remarkable: the l_2-norm estimates are very smoothed and it is impossible to identify the instants of the jumps (see Fig. 4). Furthermore, in multi-variable systems, the l_2-norm approximation can give a nearly constant estimate for the original signal with a significant jump (as in [48]).

B. Comparison of Two Approximation Methods for Absolute Values

Recall that, in this paper, we have modified the method for smooth approximation of absolute values, viz., instead of (8) from earlier works [22], [25], we use more complicated approximation (6). Obviously, with (8), the structure of all formulas for the solutions of primal and dual problems remains the same. The only difference is that we substitute $Q_W^{-2}(k)$ for $(1/2)Q_W^{-2}(k)$. In particular, the solution of the dual problem $(\lambda^a, \mu_0^a, \mu_z^a, \mu_q^a)$ is defined by Theorem 5 with the replacements $Q_W^{-2}(k)$ by $2Q_W^{-2}(k)$ and $(x^*(k), q^*(k))$ by $(x^a(k), q^a(k))$, where $(x^a(k), q^a(k))$ is the solution of the approximating quadratic problem with the trivial approximation (8). However, this difference has a significant impact on the calculation of the guaranteed nonoptimality level.

Indeed, the main idea of the proof of Theorem 6 is that, in order to estimate $\mathcal{I}_0$ from below, a maximum of (22) is searched along the direction that is given by the solution of another problem (26), (23). It is supposed that the estimate will be successful if the solutions of (22)–(24) and (26), (23) are close. This is usually achieved in later iterations when the entries $q_i(k, s)$ vary insignificantly from iteration to iteration

$$|q_i(k, s+1)| \,/\, |q_i(k, s)| \approx 1. \tag{35}$$

In accordance with the rule for setting the Q_W^{-2} from Section II-C2, formula (35) gives that the components μ_{qi}^a and μ_{qi}^* are estimated (in order of magnitude) as follows:

$$\left|\mu_{qi}^a(k)\right| = \left|2Q_W^{-2}(k)q^a(k)\right| \approx 2Q_i^{-1},$$
$$\left|\mu_{qi}^*(k)\right| = \left|Q_W^{-2}(k)q^*(k)\right| \approx Q_i^{-1}$$

i.e., at later iterations, $|\mu_{qi}^a(k)|$ is approximately two times greater than $|\mu_{qi}^*(k)|$. Hence, $\mu_{qi}^*(k)$ is more consistent with constraint (24): $|Q_i\mu_{qi}(k)| \leq 1$. In addition, numerical experiments show that the solution components $\mu_{0i}^a(k)$, $\mu_{0i}^*(k)$ and $\mu_{zi}^a(k)$, $\mu_{zi}^*(k)$ are sufficiently close. Hence, in the compound vector $(\mu_0^a, \mu_z^a, \mu_q^a)$, the first two subvectors are close to "correct" subvectors but the third subvector is twice greater (in absolute value) than it is required: $|Q_i\mu_{qi}^a(k)| \sim 2$. As a result, we get that the direction given by (λ^*, μ^*) is more preferable than the direction given by (λ^a, μ^a).

The calculation of the guaranteed nonoptimality levels confirms this claim: our approximation (6) allows to estimate nonoptimality levels much more accurately than the simplest approximation (8). Consider the same system (32), (33) and make an additional numerical experiment. The estimates of $x_2(k)$ obtained by two different methods are depicted in Fig. 5: the dotted line corresponds to the simplest approximation (8), the bold solid line (as earlier) corresponds to our basic case (6). It follows from Fig. 5 that both estimates are close, however the approximation (6) gives slightly better result. But for the guaranteed nonoptimality levels the results are very different, see Fig. 6. With our choice (6), the required threshold value

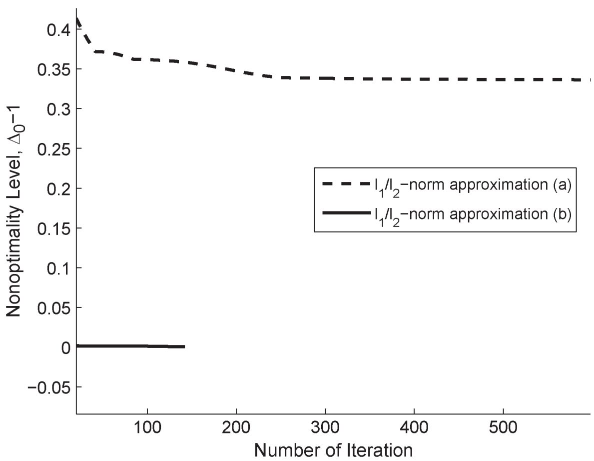

Fig. 6. Normalized guaranteed levels $\Delta_0 - 1$ for two approximation methods.

$\Delta_{\rm end} = 1 + 10^{-3}$ is obtained in 149 steps; with the simplest version (8), the desired value is not reached at all within the foreseeable time.

The analogous comparison of two approximation methods also can be made for l_1-norm approximation; recall that in [48] the choice (8) was used. However, since the l_1-cost function is homogeneous, the modified version (6) results in the appearance of the common factor 1/2 in all summands of the approximating l_2-cost functions; so, the use of (6) in this case is to no avail.

VI. Conclusion

In the paper, the algorithm of weight and time recursions is proposed for solving the l_1/l_2-norm approximation problem in linear dynamic systems. The application of mixed-norm approximation allows us to effectively estimate the state vector with possible rare jumps. The use of variational-weighted approximations reduces the solution of the original nonsmooth problem to the solution of a sequence of quadratic problems. The recursive methods in dynamic quadratic smoothing problems enable to avoid the operations with vectors and matrices of high dimension. The simple structure of the algorithm results in savings in computation and allows to process the large amounts of measurement data. The guaranteed nonoptimality levels of current iterations characterize the accuracy of approximate solutions and give a criterion for stopping the calculation process. The numerical experiments show that by means of l_1/l_2-norm approximation the jumps can be identified more clearly than by the traditional l_2-norm approximation. Moreover, the mixed-norm approximation outperforms l_1-norm approximation in computation time.

Appendix A

Proof of Theorem 4: Let us follow a standard approach for constructing the dual problems [46], [47]. The Lagrange function for (15), (16) is defined by the relation

$$\mathcal{L}(u,v,w,\mu_u,\mu_v,\lambda) = u^T P_u^2 u + \sum_{i=1}^{M_v} p_{vi}|v_i| + \lambda^T(b - \Psi w) + \left(\mu_u^T, \mu_v^T\right)\left(a - \Phi w - \begin{pmatrix} u \\ v \end{pmatrix}\right).$$

Then the dual problem to the convex optimization problem (15), (16) has the form

$$\mathcal{I}^0 = \sup_{\mu_u,\mu_v,\lambda}\left(\inf_{u,v,w} \mathcal{L}(u,v,w,\mu_u,\mu_v,\lambda)\right). \tag{36}$$

For the calculation of $\inf_{u,v,w}\mathcal{L}$ with fixed Lagrange multipliers μ_u, μ_v, λ we consider two cases.

1) Let $\lambda^T\Psi + (\mu_u^T, \mu_v^T)\Phi \neq 0$. Then $\mathcal{L}$ is linear in w and $\inf_{u,v,w}\mathcal{L}(u,v,w,\mu_u,\mu_v,\lambda) = -\infty$.
2) Let $\lambda^T\Psi + (\mu_u^T, \mu_v^T)\Phi = 0$. Then $\mathcal{L}$ does not depend on w and

$$\inf_{u,v,w}\mathcal{L}(u,v,w,\mu_u,\mu_v,\lambda) = \inf_u\left(u^T P_u^2 u - \mu_u^T u\right) + \inf_v\left(\sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}\mu_{vi}v_i\right) + \lambda^T b + \left(\mu_u^T, \mu_v^T\right)a.$$

Obviously

$$\inf_u\left(u^T P_u^2 u - \mu_u^T u\right) = -\frac{1}{4}\mu_u^T P_u^{-2}\mu_u.$$

Now let us find the second infimum. The following relations hold:

$$\begin{aligned}\sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}\mu_{vi}v_i &\geq \sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}|\mu_{vi}||v_i| \\ &= \sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v} p_{vi}^{-1}|\mu_{vi}|p_{vi}|v_i| \\ &\geq \sum_{i=1}^{M_v} p_{vi}|v_i| - \max_{i=1,\ldots,M_v}\left\{\left|p_{vi}^{-1}\mu_{vi}\right|\right\}\cdot\left(\sum_{i=1}^{M_v} p_{vi}|v_i|\right) \\ &= \left(1 - \left\|P_v^{-1}\mu_v\right\|_\infty\right)\cdot\left(\sum_{i=1}^{M_v} p_{vi}|v_i|\right)\end{aligned}$$

where, with notation (14), $\max_{i=1,\ldots,M_v}\{|p_{vi}^{-1}\mu_{vi}|\} = \|P_v^{-1}\mu_v\|_\infty$. Consequently, if $\|P_v^{-1}\mu_v\|_\infty \leq 1$, then

$$\inf_v\left(\sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}\mu_{vi}v_i\right) = 0.$$

Consider the opposite case: $\|P_v^{-1}\mu_v\|_\infty > 1$. Hence there exists a j such that $p_{vj}^{-1}|\mu_{vj}| > 1$. Take v in the form: $v_i = 0,\ i \neq j$ and $v_j \neq 0$; then

$$\begin{aligned}\sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}\mu_{vi}v_i &= p_{vj}|v_j| - \mu_{vj}v_j \\ &= v_j p_{vj}\left(\operatorname{sign} v_j - p_{vj}^{-1}\mu_{vj}\right).\end{aligned}$$

Clearly, if $\mu_{vj} > 0$, then $v_j p_{vj}(\operatorname{sign} v_j - p_{vj}^{-1}\mu_{vj}) \to -\infty$ as $v_j \to +\infty$. Similarly, if $\mu_{vj} < 0$, then $v_j\ p_{vj}\ (\operatorname{sign} v_j - p_{vj}^{-1}\mu_{vj}) \to -\infty$ as $v_j \to -\infty$. Thus, with $\|P_v^{-1}\mu_v\|_\infty > 1$, our function is unbounded from below

$$\inf_v\left(\sum_{i=1}^{M_v} p_{vi}|v_i| - \sum_{i=1}^{M_v}\mu_{vi}v_i\right) = -\infty.$$

By combining the obtained results, we finally get

$$\inf_{u,v,w} \mathcal{L}(u,v,w,\mu_u,\mu_v,\lambda) = \begin{cases} -\frac{1}{4}\mu_u^T P_u^{-2}\mu_u + \lambda^T b + \left(\mu_u^T,\mu_v^T\right) a \\ \qquad \text{if } \lambda^T\Psi + \left(\mu_u^T,\mu_v^T\right)\Phi = 0, \ \left\|P_v^{-1}\mu_v\right\|_\infty \le 1, \\ -\infty \qquad \text{otherwise.} \end{cases}$$

Then it follows from (36) that the problem dual to (15), (16) has the form (20), (21). The equality $\mathcal{I}_0 = \mathcal{I}^0$ follows from the duality theorem for convex variational problems [46], [47]. □

Appendix B

Proof of Theorem 5: First let us find a solution of (25), (23) (the dual problem to the classical l_2-norm approximation problem). Write out the Lagrange function for (25), (23)

$$\begin{aligned}\tilde{L}&(\mu_u,\mu_v,\lambda,\beta) \\ &= \beta_0 \left(\bar{x}^T(0)\mu_0 + \sum_{k=0}^{K} z^T(k)\mu_z(k) + \sum_{k=0}^{K-1} g^T(k)\lambda(k+1) \right. \\ &\qquad \left. - \frac{1}{4}\left(\mu_0^T\Pi^2\mu_0 + \sum_{k=0}^{K}\mu_z^T(k)R^2\mu_z(k) + \sum_{k=0}^{K-1}\mu_q^T(k)Q^2\mu_q(k) \right)\right) \\ &\quad + \beta_z^T(0)\left(-F^T\lambda(1) + \mu_0 + H^T\mu_z(0)\right) \\ &\quad + \sum_{k=1}^{K}\beta_z^T(k)\left(\lambda(k) - F^T\lambda(k+1) + H^T\mu_z(k)\right) \\ &\quad + \sum_{k=1}^{K}\beta_q^T(k-1)\left(-G^T\lambda(k) + \mu_q(k-1)\right)\end{aligned}$$

where

$$\beta = \left(\beta_0, \beta_z^T(0),\ldots,\beta_z^T(K),\beta_q^T(0),\ldots,\beta_q^T(K-1)\right)^T$$

is the vector of Lagrange multipliers, $\beta_0 \in \mathbb{R}_+$, $\beta_z(k) \in \mathbb{R}^n$, $\beta_q(j) \in \mathbb{R}^l$, $k = 0,\ldots,K$, $j = 0,\ldots,K-1$.

Apply the Lagrange principle for smooth convex problems with equality type constraints [46], [47]. Suppose $\{\mu_0^*, \mu_z^*(0),\ldots,\mu_z^*(K), \ \mu_q^*(0),\ldots,\mu_q^*(K-1),\lambda^*(1),\ldots,\lambda^*(K)\}$ is a solution of (25), (23); then there exist Lagrange multipliers $\beta_0, \ \beta_z(0),\ldots,\beta_z(K), \ \beta_q(0),\ldots,\beta_q(K-1)$ (not all zero) such that the stationarity condition holds

$$\frac{\partial\tilde{L}}{\partial\mu_u}(\mu_u^*,\mu_v^*,\lambda^*,\beta) = 0, \quad \frac{\partial\tilde{L}}{\partial\mu_v}(\mu_u^*,\mu_v^*,\lambda^*,\beta) = 0,$$
$$\frac{\partial\tilde{L}}{\partial\lambda}(\mu_u^*,\mu_v^*,\lambda^*,\beta) = 0.$$

Consider two cases.

1) Suppose $\beta_0 = 0$; then $\tilde{L}(\mu_u,\mu_v,\lambda,\beta)$ is a linear function in λ,μ_u,μ_v and we get from the stationarity condition that

$$\beta_z(0) = 0, \quad H\beta_z(k) = 0, \quad k = 0,\ldots,K,$$
$$\beta_q(j) = 0, \quad j = 0,\ldots,K-1,$$
$$\beta_z(k) - F\beta_z(k-1) - G\beta_q(k-1) = 0, \quad k = 1,\ldots,K.$$

Since $\beta_z(0) = 0$, $\beta_q(k) = 0$, we have

$$\beta_z(k) = F\beta_z(k-1) = F^k\beta_z(0) = 0.$$

Hence all $\beta_0, \ \beta_z(0),\ldots,\beta_z(K), \ \beta_q(0),\ldots,\beta_q(K-1)$ are equal to zero, which comes into conflict with the Lagrange principle.

2) Suppose $\beta_0 > 0$; then without loss of generality we may put $\beta_0 = 1$. Then we obtain from the stationarity condition

$$\begin{aligned}\mu_0^* &= 2\Pi^{-2}\left(\bar{x}(0) + \beta_z(0)\right), \\ \mu_z^*(k) &= 2R^{-2}\left(z(k) + H\beta_z(k)\right), \quad k = 0,\ldots,K, \\ \mu_q^*(j) &= 2Q^{-2}\beta_q(j), \quad j = 0,\ldots,K-1. \\ -\beta_z(k) &= -F\beta_z(k-1) - G\beta_q(k-1) + g(k-1), \\ &\qquad k = 1,\ldots,K. \end{aligned} \tag{37}$$

Since $\mu_0^*, \mu_z^*(k), \mu_q^*(k)$ is a solution of (25), (23), the equalities (23) also hold. If we introduce a fictitious Lagrange multiplier $\lambda^*(0) \overset{\text{def}}{=} F^T\lambda^*(1) - H^T\mu_z^*(0)$, then from (37) and (23) we get the following recurrent relations for $\beta_z(k)$ and $\lambda^*(k)$:

$$\begin{aligned}-\beta_z(k+1) &= -F\beta_z(k) - \frac{1}{2}GQ^2G^T\lambda^*(k+1) + g(k), \\ \lambda^*(k) &= F^T\lambda^*(k+1) - 2H^TR^{-2}\left(z(k) + H\beta_z(k)\right), \\ &\qquad k = 0,\ldots,K \end{aligned} \tag{38}$$

under the boundary conditions

$$\lambda^*(K+1) = 0, \quad \beta_z(0) = -\bar{x}(0) + \frac{1}{2}\Pi^2\lambda^*(0).$$

The boundary-value problem from Theorem 1 (except for the second equation of (5)) and the boundary-value problem (38) are identical up to the replacement $\beta_z(k)$ by $-x^*(k)$ and $\lambda^*(k)$ by $-2\lambda^J(k)$. It is well known that the boundary-value problem defined by (5) has a unique solution (see, e.g., [24]). Therefore

$$\beta_z(k) = -x^*(k), \quad \lambda^*(k) = -2\lambda^J(k), \quad k = 0,\ldots,K,$$
$$\beta_q(i) = -q^*(i), \quad i = 0,\ldots,K-1.$$

By virtue of (37), this fact establishes a link between the solutions of primal and dual problems

$$\begin{aligned}\mu_0^* &= 2\Pi^{-2}\left(\bar{x}(0) - x^*(0)\right), \\ \mu_z^*(k) &= 2R^{-2}\left(z(k) - Hx^*(k)\right), \quad k = 0,\ldots,K, \\ \mu_q^*(i) &= -2Q^{-2}q^*(i), \quad i = 0,\ldots,K-1. \end{aligned} \tag{39}$$

Problem (9), (2) differs from (4), (2) by weight coefficients only. Recall that all considerations are valid for time-varying weight matrices as well. Hence, the substitution of $\frac{1}{2}Q_W^{-2}(k)$ for Q^{-2} into (39), where now (x^*, q^*) is the solution of (9), (2), completes the proof. □

Acknowledgment

The authors would like to thank B.T. Polyak for his valuable advice on convex optimization and the reviewers for helpful comments.

References

[1] J. Grizzle, G. Abba, and F. Plestan, "Asymptotically stable walking for biped robots: Analysis via systems with impulse effects," *IEEE Trans. Autom. Control*, vol. 46, no. 1, pp. 51–64, 2001.
[2] R. Korn, "Some applications of impulse control in mathematical finance," *Math. Methods Oper. Res.*, vol. 50, no. 3, pp. 493–518, 1999.
[3] C. Briat and A. Seuret, "Convex dwell-time characterizations for uncertain linear impulsive systems," *IEEE Trans. Autom. Control*, vol. 57, no. 12, pp. 3241–3246, 2012.
[4] C. Briat and E. Verriest, "A new delay-SIR model for pulse vaccination," *Biomed. Signal Processing Control*, vol. 4, no. 4, pp. 272–277, 2009.
[5] P. A. Akimov and A. I. Matasov, "State estimation via l_1-norm approximation: Application to inertial navigation," in *Proc. 18th World Congress Int. Fed. Autom. Control*, Milan, Italy, Aug. 2011, vol. 18, pp. 14 386–14 391.
[6] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations: Theory, Applications, Algorithms*. Boston, MA: Birkhauser, 1983.
[7] V. I. Mudrov and V. L. Kushko, (in Russian), *Methods for Processing the Observations: Quasi-Likelihood Estimates*. Moscow, Russia: Radio and Communication, 1983.
[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[9] M. Dasgupta and S. K. Mishra, Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review, Munich Personal RePEc Archive, Feb. 2007. [Online]. Available: http://mpra.ub.uni-muenchen.de/1781/
[10] D. Strong and T. Chan, Exact Solutions to Total Variation Regularization Problems, UCLA CAM Report, 1996.
[11] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems*, vol. 19, pp. S165–S187, 2003.
[12] T. Chan and S. Esedoglu, "Aspects of total variation regularized L_1 function approximation," *SIAM J. Appl. Math.*, vol. 65, no. 5, pp. 1817–1837, 2005.
[13] P. J. Huber, "Robust regression: Asymptotics, conjectures and Monte Carlo," *Annals Stat.*, vol. 1, no. 5, pp. 799–821, 1973.
[14] R. Tabshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.
[15] S. Kim, K. Koh, S. Boyd, and D. Gorinevski, " l_1 trend filtering," *SIAM Rev.*, vol. 51, no. 2, pp. 339–360, 2009.
[16] B. T. Polyak, "Methods of l_1 optimization in control and filtering," in *Proc. 3rd Multi-Conf. Control Problems*, St. Petersburg, Russia, Oct. 2010, [CD-ROM].
[17] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Comp. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.
[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2010.
[19] L. Vandenberghe, "Convex optimization algorithms for sparse and low-rank representations," in *Proc. Eur. Control Conf.*, Zurich, Switzerland, Jul. 2013, [CD-ROM].
[20] S. Farahmand, G. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4529–4543, 2011.
[21] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," in *Proc. 16th IFAC Symp. Syst. Ident.*, Brussels, Belgium, Jul. 2012, [CD ROM].
[22] E. J. Schlossmacher, "An iterative technique for absolute deviations curve fitting," *J. Amer. Stat. Assoc.*, vol. 68, no. 344, pp. 857–859, 1973.
[23] A. E. Bryson and H. Yu-Chi, *Applied Optimal Control*. London, U.K.: Blaisdell, 1969.
[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Princeton, NJ: Prentice-Hall, 2000.
[25] R. C. Fair, "On the robust estimation of econometric models," *Annals Econ. Soc. Meas.*, vol. 3, no. 4, pp. 667–677, 1974.
[26] R. F. Philips, "Least absolute deviations estimation via the EM algorithm," *Stat. Comp.*, vol. 12, no. 3, pp. 281–285, 2002.
[27] E. Weiszfeld, "Sur le point par lequel la somme des distances de n points donnés est minimum," *Tohoku Math.*, no. 43, pp. 355–386, 1937.
[28] E. Weiszfeld and F. Plastria, "On the point for which the sum of the distances to n given points is minimum," *Annals Oper. Res.*, vol. 167, no. 1, pp. 7–41, 2008.
[29] P. A. Akimov, A. V. Derevyankin, and A. I. Matasov, (in Russian), *Guaranteed Approach and l_1-norm Approximation in Parameter Estimation Problems for SDINS under Bench Testing*. Moscow, Russia: Moscow Univ. Press, 2012.
[30] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, 2010.
[31] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 50–61, 2010.
[32] N. Vaswani, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4108–4120, 2010.
[33] D. Sejdinovic, C. Andrieu, and R. Piechocki, "Bayesian sequential compressed sensing in sparse dynamical systems," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comp.*, Allerton, IL, Sep. 2010, pp. 1730–1736.
[34] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.
[35] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," in *Proc. 45th Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2011, pp. 1–6.
[36] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4967–4972, 2012.
[37] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization (MASTeR) for accelerated dynamic MRI," *Magn. Resonance Med.*, vol. 70, no. 3, pp. 800–821, 2013.
[38] A. Charles and C. Rozell, "Dynamic filtering of sparse signals using reweighted l_1," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Vancouver, Canada, May 2013, pp. 6451–6455.
[39] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, Random Processes, Statistical Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[40] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
[41] A. A. Golovan and N. A. Parusnikov, (in Russian), *Mathematical Foundations of Navigation Systems, Part I. Mathematical Models of Inertial Navigation*. Moscow, Russia: Moscow Univ. Press, 2010.
[42] J. A. Farrell, *Aided Navigation: GPS With High Rate Sensors*. New York, NY, USA: McGraw Hill, 2008.
[43] P. S. Maybeck, *Stochastic Models, Estimation, Control*. New York, NY, USA: Academic Press, 1979.
[44] A. E. Bryson and M. Frazier, "Smoothing for linear and nonlinear dynamic systems," in *Proc. Optim. Syst. Synth. Conf.*, Ohio, OH, USA, Sep. 1963, pp. 353–364.
[45] A. I. Matasov, *Estimators for Uncertain Dynamic Systems*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.
[46] V. M. Alekseev, V. M. Tikhomirov, and S. V. Fomin, *Optimal Control*, (in Russian), Moscow, Russia: Fizmatlit, 2005.
[47] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*. Amsterdam, The Netherlands: North Holland Pub. Co., American Elseveir Pub. Co., 1976.
[48] P. A. Akimov and A. I. Matasov, "Recursive estimation algorithm for l_1-norm approximation in dynamic systems with nonoptimality levels," in *Proc. Eur. Control Conf.*, Zurich, Switzerland, Jul. 2013, pp. 2427–2432.

Pavel Akimov received the Ph.D. degree from M.V. Lomonosov Moscow State University, Moscow, Russia, in 2011.

He held a part-time position at the Faculty of Mechanics and Mathematics, M.V. Lomonosov Moscow State University (2013) and participated in two projects of the Russian Foundation for Basic Research (2009–2013). Since 2008, he has been with the Credit Risk Management Department, Commercial Russian Bank. He co-authored ten papers and one book. His main research interests lie in the field of estimation, optimization, numerical methods, and data mining.

Alexander Matasov (M'97) received the Ph.D. and D.Sc. degrees from M.V. Lomonosov Moscow State University, Moscow, Russia, in 1982 and 1993, respectively.

From 1975 to 1979 he held engineering positions in aviation industry. Since 1979, he has been with the Faculty of Mechanics and Mathematics, M.V. Lomonosov Moscow State University, where he is currently a Leading Researcher at the Laboratory of Control and Navigation; he is a Full Professor (2000) and Soros Professor (2001). He also held part-time positions at Moscow Aviation Institute (Technical University) (1998–2003) and at the Space Research Institute (IKI), Russian Academy of Sciences (1999–2003). Prof. Matasov is the author of about 150 research publications and three monographs: *Estimators for Uncertain Dynamic Systems* (Kluwer, 1999); *Guaranteed Estimation Method* (Moscow University Press, 2009, in Russian); *Guaranteed Approach and l_1-norm Approximation in the Problems of SDINS Parameter Estimation under Bench Testing* (Moscow University Press, 2012, in Russian). His current research interests include navigation, estimation and control theory, theory of functional differential equations.

Dr. Matasov is a member of the Moscow Mathematical Society, the Academy of Nonlinear Sciences, the Academy of Navigation and Motion Control (Russia), the American Mathematical Society and the IEEE Control Systems Society (USA).