On Geometric and Analytic Constraints in the $H^\infty$ Fixed-Lag Smoothing

Leonid Mirkin and Gilead Tadmor

Abstract—This paper studies the limiting factors on $H^\infty$ performance achievable in the fixed-lag smoothing problem, in both continuous and discrete time. It shows that geometric performance bounds are inherited from the fixed-interval ($L^\infty$) problem, whereas analytic constraints are affected by the preview interval. Thus, the fixed-lag problem can be viewed as a constrained version of its fixed-interval counterpart. The analysis sheds light on the dependence of achievable performance on the preview length and, in particular, on the generic presence of performance saturation at a finite preview interval.

I. INTRODUCTION

The paper concerns the $H^\infty$ fixed-lag smoothing problem for the setup in Fig. 1. Here $v$ is a signal that we want to reconstruct using the measurements $y$. We assume that both $v$ and $y$ are generated by an LTI finite-dimensional system $G$, driven by an exogenous signal $w$. The estimate $u$ is generated by a stable estimator $K$. The $H^\infty$ norm of the system $G_v$, mapping $w$ to the estimation error $e := v - u$, called the error system, is the indicator of the estimation performance.

The fixed-lag smoothing problem is characterized by a relaxed causality constraint on the estimator $K$. Thus, at the time $t$ the estimator has access to the measurement $y$, up to the time $t + h$, for some fixed preview length, also termed the smoothing lag, $h \geq 0$. The idea is to exploit the preview to improve the estimation performance relative to causal estimation (filtering), where $h = 0$. In practice, the smoothing problem can be viewed as the reconstruction of the delayed version of $v$ by a causal estimator, whereby only causal systems are involved. Here, however, we find it convenient to use the equivalent non-causal formulation, to highlight the advantage of previewed information.

Studies of smoothing and (its dual) preview tracking problems in an $H^2$ (or Kalman filtering) setting trace back to the early ’60s [1], [2]. A renewed interest in preview problems over the past few years includes several revisions of the classical $H^2$ version [3]–[5] and complete solutions to the more challenging $H^\infty$ problem [6]–[10] (a comprehensive bibliography can be found in these papers). The smoothing solutions bear close similarity to their filtering counterparts, especially in their solvability conditions, which are based on similar Riccati equations.

An intriguing property of $H^\infty$ smoothing solutions, noticed in [6], is the saturation of the achievable performance as a function of the smoothing lag. In other words, there generically exists a finite smoothing lag, say $h_s$, such that the performance achievable with $h = h_s$ does not improve with $h > h_s$. The saturation property has no analogy in the $H^2$ case, where the performance index is generically [1] a strictly decreasing function of $h$. Conditions under which the $H^\infty$ performance saturation occurs were quantified in [11] for the special case of a continuous-time system with a strictly proper transfer function from $w$ to $v$. Saturation in discrete-time systems has been demonstrated only by numerical examples [8]. The overall underlying reason for saturation has not been previously clarified.

In this paper we propose an alternative to previously established solution procedures and use the analysis leading to that solution to elucidate the saturation phenomenon. The idea is to address the problem in two steps. The first step is the fixed-interval smoothing ($L^\infty$ estimation), in which the estimator has access to the entire past and future information about $v$; this corresponds to $h = \infty$. Performance limitations arising in this stage are geometric, they are completely determined by the part of $v$ which is invisible through $y$. In the second step, the finite preview of the fixed-lag problem is treated as a constraint imposed on the fixed-interval solution. Performance in this stage is limited by the need to keep the estimator analytic in some region of the complex plane, hence analytic constraints. The second step results in a special Nehari problem with delay, which, in turn, is solvable by the approach proposed in [12].

An attractive property of the proposed two-step solution procedure is the separation of geometric and analytic constraints. Namely, the geometric constraints are independent of the smoothing lag $h$, and remain active even as $h \to \infty$. The analytic constraints do depend on the smoothing lag $h$.

To quantify this dependence, we derive an induced norm interpretation of the Hankel norm associated with the delayed Nehari problem, arising in the second step. This interpretation reveals the fact that limitations imposed by analytic constraints vanish exponentially as $h \to \infty$. Therefore, in the presence of nontrivial geometric constraints, there always exists a finite $h$ at which the analytic constraint become weaker than the constant geometric constraint. The performance reaches its optimal fixed-interval level, and saturates, at that finite $h$. This observation extends the result of [11] for the continuous-time case and constitutes the first analytic proof of saturation in the discrete-time case.

Notation: The nomenclature adopted here is driven by the need for a unified treatment of continuous- and discrete-time systems. We use $\lambda$ for the Laplace transform variable in both the continuous (where $\lambda = s$) and the discrete (where $\lambda = z$) cases. The notation $\mathbb{B}$ stands for the stability boundary in the complex $\lambda$-plane: $\mathbb{B} = \mathbb{R} \cup \infty$ (extended imaginary axis) in the continuous-time case and $\mathbb{B} = \mathbb{D}$ (unit circle) in the discrete-time case; $\mathbb{U}$ denotes the “strictly” (i.e., excluding $\mathbb{B}$) unstable region in the $\lambda$-plane. Given a transfer function $G(\lambda)$, its
II. PROBLEM FORMULATION

Consider the system in Fig. 1 and partition the (rational and proper) signal generator \(G(\lambda)\) compatibly with the signal partition as follows:

\[ G(\lambda) = \begin{bmatrix} G_e(\lambda) & G_s(\lambda) \end{bmatrix}. \]

To guarantee the well-posedness of the estimation problem, assume that

\( \mathcal{A}_1: \) \(G_s(\lambda)\) has full row rank \(\forall \lambda \in \mathbb{B}\)

(in the continuous-time case this assumption also includes the non-singularity of the feedthrough term \(G_s(\infty)\)).

Introduce the space

\[ D_{-h} H^\infty := \{ F : D_h F \in H^\infty \}. \]

Members of \( D_{-h} H^\infty \) are analytic but not necessarily bounded in \( \mathbb{U} \). Clearly, \( H^\infty \subset D_{-h_1} H^\infty \subset D_{-h_2} H^\infty \subset L^\infty \) for \( 0 < h_1 < h_2 \). The norm on \( D_{-h} H^\infty \) is the standard \( L^\infty \) norm, i.e.,

\[ \| F \|_\infty := \text{ess sup}_{\lambda \in \mathbb{B}} \sigma_{\max}[F(\lambda)]. \]

This space can be thought of as \( H^\infty \) with a relaxed boundedness in \( \mathbb{U} \) and is the stability space for systems with \( h \)-relaxed causality.

The \( H^\infty \) fixed-lag smoothing problem studied in this paper is formulated in these terms, following a standard pattern:

**SP**

Given a signal generator \( G \), satisfying \( \mathcal{A}_1 \), a smoothing lag \( h \geq 0 \) and a performance bound \( \gamma \geq 0 \), determine whether there is an estimator \( K \in D_{-h} H^\infty \) satisfying

\[ G_v - KG_y \in D_{-h} H^\infty \quad \text{and} \quad \| G_v - KG_y \|_\infty \leq \gamma. \]

If the answer is affirmative, find solutions satisfying the performance bound.

The non-strict inequality for the norm of the error system is considered here to treat the optimal case.

Note that we do not require that \( G \) itself be stable. This enables to consider problems arising in stable physical systems, but where typical unstable weights, like the integrator or frequency resonances, are used. The concept of a stable solution relates then to steady-state errors rather than to signal boundedness under all bounded \( w \). For example, if \( G \) has a pole at the origin, the stability of \( G_v \) is merely equivalent to the zero steady-state error \( e \) under a constant \( u \). The stability conditions are elaborated in the next section.

III. STABILIZATION AND NORMALIZATION

Since the system in Fig. 1 is open loop, stability of the error system can only be attained via canceling unstable poles of \( G \) by \( K \). This amounts to imposing interpolation constraints on the estimator. For example, the stability of

\[ G_v(s) = \begin{bmatrix} \frac{1}{s} & 0 \\ 0 & \frac{1}{s} \end{bmatrix} - K(s) \begin{bmatrix} \frac{1}{s} & 1 \end{bmatrix} \]

is equivalent to the condition \( K(0) = 1 \). This constraint can be handled by substituting \( K(s) = s/(s + 1)Q(s) + 1/(s + 1) \) for an unconstrained \( Q \in e^{sh} H^\infty \). This is actually equivalent to considering the system

\[ G_v(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} - Q(s) \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \end{bmatrix} \]

for which stability is guaranteed whenever \( Q \in e^{sh} H^\infty \).

Interpolation constraints on \( K \) become more complicated in the case of multiple poles or matrix-valued \( G_v \). We, therefore, consider an implicit approach of imposing these constraints, which follows the spirit of the Youla parametrization for feedback systems [13]. Namely, we make use of the coprime factorization machinery to reduce the general problem to that with a stable signal generator. With that goal in mind, we establish some preliminary results.

**Proposition 1:** The error system is stabilizable for some \( h > 0 \) iff it is stabilizable for \( h = 0 \).

**Proof:** Since \( D_{-h} K(\lambda) \) is an entire function, singularities arise only from the rational part of the system. The proposition thus follows by standard interpolation arguments, involving only the rational plant [14].

**Proposition 2:** There is a stabilizing \( K \in H^\infty \) iff

\[ G_v = N_v - M_v G_y \]

for some \( N_v, M_v \in H^\infty \).

**Proof:** If (1) holds, then \( K = -M_v \) clearly stabilizes \( G_v \). On the other hand, if \( K \in H^\infty \) is stabilizing, (1) holds for \( N_v = G_e \) and \( M_v = -K \).

Proposition 2 leads to the following result:

**Proposition 3:** There exists a stabilizing \( K \in H^\infty \) iff \( G \) admits a left coprime factorization of the form

\[ G = \begin{bmatrix} I & M_v \\ 0 & M_y \end{bmatrix}^{-1} \begin{bmatrix} N_v \\ N_y \end{bmatrix} \]

for some \( M_v, M_y, N_v, N_y \in H^\infty \).

**Proof:** Since \( G_y \) is rational and proper, it can always be factorized as \( G_y = M_y^{-1} N_y \) for some coprime \( M_y, N_y \in H^\infty \). That is the second row of (2). It then readily follows that the first row of (2) is (1). We thus have to show only that the factorization (2) is coprime. Indeed, let \( X_y \) and \( Y_y \) be Bézout factors for the coprime factors \( M_y \) and \( N_y \). Using these terms one obtains the Bézout identity for the factorization (2):

\[ \begin{bmatrix} I & M_v \\ 0 & M_y \end{bmatrix} \begin{bmatrix} I & -M_v X_y - N_y Y_y \\ 0 & X_y \end{bmatrix} = \begin{bmatrix} N_v \\ N_y \end{bmatrix} \begin{bmatrix} 0 \\ Y_y \end{bmatrix} = I \]

and the proof is complete.

A complete parametrization of all stabilizing estimators and error systems in Fig. 1, which is an open-loop counterpart of the Youla parametrization, is then formulated as follows:

**Lemma 1:** Let \( G \) admit a coprime factorization over \( H^\infty \) as in (2). Then \( K \in D_{-h} H^\infty \) stabilizes the error equation iff

\[ K = -M_v + Q M_y \]

for some \( Q \in D_{-h} H^\infty \). In this case

\[ G_e = N_v - Q N_y \]

parametrizes the set of all error transfer matrices.
Proof: The “if” part and (4) follow by direct substitutions.
To prove the “only if” part, assume that \( K \in D_{-h}H^\infty \) stabilizes the error equation. Then \( Q := (M_v + K)M_v^{-1} \) should satisfy
\[
[I - Q] \begin{bmatrix} M_v & N_v \\ 0 & M_y & N_y \end{bmatrix} \in D_{-h}H^\infty
\]
Equivalently,
\[
[D_h - D_h Q] \begin{bmatrix} I & M_v & N_v \\ 0 & M_y & N_y \end{bmatrix} \in H^\infty
\]
Because (2) is a left coprime factorization of \( G \), the last equality implies that \( D_h Q \in H^\infty \), so that \( Q \in D_{-h}H^\infty \).

Casting the minimization of \( \|G_e\|_\infty \) in terms of (4), Lemma 1 enables us to consider only estimation problems with stable signal generators. Additional numerator structure is imposed to further simplify the solution in the next section. The following observation is a starting point:

Proposition 4: Let assumption \( \mathcal{A}_1 \) hold. Then the numerator in the factorization (2) can be chosen in the form
\[
\begin{bmatrix} N_v \\ N_y \end{bmatrix} N_v^\sim = \begin{bmatrix} V^\sim \\ I \end{bmatrix}
\]
with a strictly proper \( V \in H^\infty \).

Proof: Let \( N_y \) be selected as a co-inner numerator in the left coprime factorization of \( G \). The existence of such a factorization is a well known consequence of \( \mathcal{A}_1 \) [14]. It is easy to see that when \( N_y \) is co-inner, a parametrization of coprime factorizations of the form (2) is obtained by the following substitutions, with a free selection of \( T \in H^\infty \):
\[
\begin{bmatrix} M_v & N_v \\ M_y & N_y \end{bmatrix} := \begin{bmatrix} M_v - T M_y & N_v - T N_y \\ M_y & N_y \end{bmatrix}
\]
Using this parameterization, we have
\[
\begin{bmatrix} N_v \\ N_y \end{bmatrix} N_v^\sim = \begin{bmatrix} N_v - T N_y \\ N_y \end{bmatrix} N_v^\sim = \begin{bmatrix} N_v N_v^\sim - T \\ I \end{bmatrix}.
\]
The desired form and properties of (5) are then obtained by choosing \( T \) to be the stable part of \( N_v N_v^\sim \) with \( T(\infty) = N_v(\infty)N_v^\sim(\infty) \).

Henceforth, we substitute \( \mathbf{SP}_G \) with the following equivalent \( H^\infty \) smoothing problem:

\( \mathbf{SP}_N \): Given \( N \in H^\infty \) satisfying (5), smoothing lag \( h \geq 0 \) and \( \gamma \geq 0 \), determine whether there is \( Q \in D_{-h}H^\infty \) such that
\[
\|N_v - Q N_y\|_\infty \leq \gamma.
\]
If the answer is affirmative, find solutions satisfying the performance bound.

The equivalence of \( \mathbf{SP}_G \) and \( \mathbf{SP}_N \) means that the optimal performance level, \( \gamma_{opt} \), is identical in the two problem formulations, and that solutions \( K \) of \( \mathbf{SP}_G \) are defined by solutions \( Q \) of \( \mathbf{SP}_N \), via (3). As in similar factorization-based reductions of \( H^\infty \) problems, the reduction of \( \mathbf{SP}_G \) to \( \mathbf{SP}_N \) has two main advantages: interpolation stability constraints are implicitly satisfied and solution formulae, which are yet to be derived, are considerably simplified, owing to the normalization in (5).

IV. Solution of \( \mathbf{SP}_N \)

A. Fixed-interval solution: geometric constraints

We first address the simpler, fixed-interval smoothing version of \( \mathbf{SP}_N \), where no causality constraints are imposed upon \( Q \) (hence, on \( K \)). Thus, the constraint \( Q \in D_{-h}H^\infty \) is relaxed to \( Q \in L^\infty \). Since \( D_{-h}H^\infty \subset L^\infty \), solvability of the \( L^\infty \), fixed-interval problem, is necessary for solvability of \( \mathbf{SP}_N \), and the optimal value of the former is a lower bound on the performance achievable by a fixed-lag solution. Structural aspects of the solution of \( \mathbf{SP}_N \) will be revealed, as a by product of the analysis of the relaxed problem.

The key property we exploit in addressing the \( L^\infty \) problem is that it can be solved in the frequency domain as an amalgamation of independent static problems, at each frequency \( \lambda \in \mathbb{B} \), see [15, Ch. 11]. Specifically, for \( \gamma \geq \gamma_{opt} \) and a \( \gamma \)-(sub)optimal \( Q \in L^\infty \), the following inequality must be satisfied, for all \( \lambda \in \mathbb{B} \):
\[
\|G_e\|_\infty \leq \gamma \iff G_e(\lambda)G_e^\sim(\lambda) \leq \gamma^2 I.
\]
Invoking (5) and suppressing \( \lambda \) dependence, we have:
\[
G_e G_e^\sim = (N_v - Q N_v)(N_v - Q N_v)^\sim
= N_v N_v^\sim - V^\sim V^\sim + Q Q^\sim \pm V^\sim V
= N_v(I - N_v N_v) N_v^\sim + (V^\sim - Q)(V^\sim - Q)^\sim
\]
The first term on the right-hand side,
\[
\Phi_{\sim \gamma} := N_v(I - N_v N_v) N_v^\sim
\]
is nonnegative because \( I - N_v N_v \) is the orthogonal projection onto the null-space of \( N_v \), and therefore defines a nonnegative matrix for all \( \lambda \in \mathbb{B} \). Thus, the problem reduces to
\[
(V^\sim - Q)(V^\sim - Q)^\sim \leq \gamma^2 I - \Phi_{\sim \gamma}.
\]
Since, here, the left hand side is nonnegative, a solution exists (e.g., the central solution \( Q = V^\sim \) iff \( \gamma \geq \Phi_{\sim \gamma} \)). In other words, the optimal achievable performance for the fixed-interval (\( L^\infty \)) smoothing problem is
\[
\gamma \sim := \|N_v(I - N_v N_v)\|_\infty = \sqrt{\|\Phi_{\sim \gamma}\|_\infty}.
\]
This quantity is the \( L^\infty \)-norm of the restriction of \( N_v \) to the null-space of \( N_v \). It therefore reflects a geometric constraint on the achievable smoothing performance, imposed by the problem structure. This constraint is independent of the causality of \( Q \) and is present in any smoothing problem.

B. Analytic constraints and reduction to Nehari problem

The incorporation of the causality condition \( Q \in D_{-h}H^\infty \) brings in the requirement that \( Q(\lambda) \) be analytic in \( \lambda \in \mathbb{U} \). In particular, \( Q \) can no longer be treated independently, at each frequency. For instance, the central \( L^\infty \) solution, \( Q = V^\sim \), is not analytic in \( \mathbb{U} \) unless \( V = 0 \). If \( V \neq 0 \), the analyticity condition means that \( \|V^\sim - Q\| \) cannot be made arbitrarily small. Constraints imposed by a nonzero \( V \) shall be referred to as analytic constraints and are caused by the required causality. As such, one may expect that these constraints are closely associated with the preview length, \( h \), which
determines the precise level of causality relaxation in $\text{SP}_G$ and $\text{SP}_N$. Our purpose here is to understand how the preview length, $h$, affects analytic constraints.

The first step towards this end is to factorize the right-hand side of (8). The geometric constraint
\begin{equation}
\gamma \geq \gamma_{\infty} \quad \text{(for } \gamma_{\infty} \text{ defined by (9)),}
\end{equation}
guarantees that $\gamma^2 I - \Phi_{\psi,\lambda}(\lambda) \geq 0$ for all $\lambda \in \mathbb{B}$. Hence, there exists a spectral factorization [16, Ch. 7]
\begin{equation}
W_r W_r^* = I - \gamma^{-2} \Phi_{\psi,\lambda}.
\end{equation}
where $W_r \in H^\infty$. To simplify the reasonings below we defer the treatment of singular cases (including $\gamma_{\infty} = 0$) to §IV-D, and assume here that
\[
A_2: \exists \lambda_0 \in \mathbb{B} \text{ such that } \gamma_{\infty}^2 I - \Phi_{\psi,\lambda}() \text{ is nonsingular.}
\]
This technical assumption ensures that the spectral factor $W_r$ in (11) is square and analytically invertible over $\mathbb{U}$. Moreover, if $\gamma > \gamma_{\infty}$, we can choose $W_r$ so that $W_r^{-1} \in H^\infty$ as well [16], while if $\gamma = \gamma_{\infty}$, $W_r^{-1}$ will have at least one pole in $\mathbb{B}$ (including the possibility that $W_r^{-1}$ is not proper).

Using the spectral factorization (11) we rewrite (8) as
\[
W_r^{-1}(V^\infty - Q)(V^\infty - Q)^*(W_r^{-1})^{-1} \leq \gamma^2 I
\]
or, equivalently, as
\begin{equation}
\|W_r^{-1}(V^\infty - Q)\|_2 \leq \gamma, \quad Q \in D_{-h} H^\infty.
\end{equation}
This inequality can be thought of as a weighted distance problem from the function $V^\infty$ to the space $D_{-h} H^\infty$. Our next step is to get rid of the weighting function $W_r^{-1}$ by restructuring $Q$. This is possible even if $\gamma = \gamma_{\infty}$, in which case the weighting function imposes additional interpolation constraints on $\tilde{Q}$ over $\mathbb{B}$.

**Lemma 2:** Let a rational $W_r \in H^\infty$ be such that $W_r^{-1}$ has no poles in $\mathbb{U}$. Then there exist $V_{\gamma^-}, V_{\gamma^+} \in H^\infty$ where $V_{\gamma^+}$ is strictly proper and such that:

1. There holds
\begin{equation}
W_r^{-1}V^\infty = V_{\gamma^+}^\infty + W_r^{-1}V_{\gamma^-}.
\end{equation}
2. $W_r^{-1}(V^\infty - Q) \in L^\infty$ for $Q \in D_{-h} H^\infty$ iff
\begin{equation}
Q = W_r \tilde{Q} + V_{\gamma^-}.
\end{equation}
for some $\tilde{Q} \in D_{-h} H^\infty$.

**Proof:** Decompose the rational $W_r^{-1}V^\infty$ as
\[
W_r^{-1}V^\infty = V_{\gamma^+}^\infty + U
\]
for some $U$, analytic over $\mathbb{U}$ and a strictly proper $V_{\gamma^+} \in H^\infty$.
Similarly, select $V_{\gamma^-} \in H^\infty$ and a strictly proper $\tilde{U} \in H^\infty$, such that
\[
W_r V_{\gamma^+}^\infty = \tilde{U}^\infty - V_{\gamma^-}.
\]
Substituting $V_{\gamma^+}^\infty$ from the first decomposition above to the left-hand side of the second one, we get
\[
(V - \tilde{U})^\infty = W_r U - V_{\gamma^-} \in H^\infty.
\]
It thus follows that $\tilde{U} = V$. This, in turn, proves (13).

Assume now that $Q$ is chosen according to (14), with $\tilde{Q} \in D_{-h} H^\infty$. The inclusion $H^\infty \subset D_{-h} H^\infty$ and the fact that $W_r \in H^\infty$ imply that $Q \in D_{-h} H^\infty$. Moreover,
\begin{equation}
W_r^{-1}(V^\infty - Q) = V_{\gamma^+}^\infty - \tilde{Q} \in L^\infty,
\end{equation}
which completes the proof of the “if” part of statement 2).

To prove the “only if” part, let $Q \in D_{-h} H^\infty$ be such that $W_r^{-1}(V^\infty - Q) \in L^\infty$ and define $\tilde{Q} := W_r^{-1}(Q - V_{\gamma^-})$, according to (14). Then all the components of
\[
D_h \tilde{Q} = W_r^{-1}(D_h Q - D_h V_{\gamma^-})
\]
are analytic functions over $\mathbb{U}$. Moreover, by (15), there also holds
\[
D_h \tilde{Q} = D_h W_r^{-1}(Q - V^\infty) + D_h V_{\gamma^+}^\infty \in L^\infty.
\]
Therefore, $D_h \tilde{Q} \in H^\infty$.

Using Lemma 2 we substitute $Q$ from (14) into (12) and state $\text{SP}_N$ in terms of the inequality
\begin{equation}
\|V_{\gamma^+}^\infty - \tilde{Q}\|_2 \leq \gamma, \quad \tilde{Q} \in D_{-h} H^\infty,
\end{equation}
or, equivalently,
\begin{equation}
\|D_h V_{\gamma^+}^\infty - Q_v\|_2 \leq \gamma, \quad Q_v \in H^\infty,
\end{equation}
This equivalent formulation concerns the unweighted distance from the function $D_h V_{\gamma^+}^\infty$ to the space $H^\infty$. It is worth emphasizing that $V_{\gamma^+}$, which is the stable part of $V(W_r^{-1})$, is uniformly bounded for all $\gamma$ satisfying (10).

Calculating the minimal $\gamma$ for which (16) is attainable is complicated by the fact that $\gamma$ is involved both as a norm bound and via the dependence of $V_{\gamma^+}$ on $\gamma$. A standard approach to resolve this double dependence is to implement a binary search procedure: having fixed a candidate $\gamma$, one solves the minimization problem
\begin{equation}
\alpha_{\gamma,h} := \min_{Q_v \in H^\infty} \|D_h V_{\gamma^+}^\infty - Q_v\|_2.
\end{equation}
The inequality (16) is then satisfied if
\begin{equation}
\gamma \geq \alpha_{\gamma,h} \quad \text{(for } \alpha_{\gamma,h}, \text{ as defined by (17)),}
\end{equation}
which is the quantification of analytic constraints. Thus, $\text{SP}_G$ is solvable iff $\gamma$ satisfies both (10) and (18).

The problem of finding $\alpha_{\gamma,h}$ is a classical Nehari extension problem [13]. In particular, Nehari’s theorem says that $\alpha_{\gamma,h}$ equals the induced $L^2$ norm of the Hankel operator with symbol $D_{-h} V_{\gamma^+}$, i.e.,
\[
\alpha_{\gamma,h} = \|\Pi_X D_{-h} V_{\gamma^+} \|_{H^2}.
\]
where $\Pi_X$ stands for the orthogonal projection from $L^2$ onto its subspace $X$. It is readily verified that
\[
\Pi_X D_{-h} = D_{-h} \Pi_{D_h} H^2.
\]
This, together with the fact that $D_{-h}$ preserves the $L^2$ norm, implies that
\begin{equation}
\alpha_{\gamma,h} = \|D_h D_{-h} V_{\gamma^+} \|_{H^2}.
\end{equation}

The last formula says that, given $\gamma$, the optimal performance for every $h$ is\(^1\) the $L^2(-\infty, 0] \mapsto L^2[0, \infty)$ induced norm of $d_k$.

\(^1\)We discuss here only the continuous-time case. Discrete interpretation is the same modulo obvious notational changes.
the very same system, $V_{y\rightarrow}$. In other words, for all values of $h$ we consider inputs from the same space, $L^2(-\infty, 0]$, whereas the truncation of outputs in $L^2[h, \infty]$ becomes increasingly restrictive, as $h$ increases. This implies that $\alpha_{\gamma, h}$ is a monotonically decreasing function of $h$ for each $\gamma$. Moreover, because the impulse response of $V_{y\rightarrow}$ decays exponentially (as $V_{y\rightarrow}$ is finite dimensional), $\alpha_{\gamma, h}$ is an exponentially decreasing function, vanishing as $h \rightarrow \infty$.

The discussion above can be summarized as follows:

**Lemma 3:** Let $A_2$ hold. Then the function $\alpha_{\gamma, 0}$ is bounded for $\gamma$ satisfying (10), and for each such $\gamma$, $\alpha_{\gamma, h}$ is an exponentially decreasing function of $h$, satisfying $\lim_{h \rightarrow \infty} \alpha_{\gamma, h} = 0$.

Lemma 3 implies that for any $\gamma \geq \gamma_\infty$ there exists a finite preview, say $h_\gamma$, such that inequality (18) holds for all $h \geq h_\gamma$. Thus, the analytic constraints (18) become inactive and the optimal $L^\infty$ (fixed-interval) performance level is achievable at the finite preview length $h_\gamma$. Larger values of $h$ will maintain the smoother performance unchanged. This phenomenon is known as performance saturation. It was analyzed by use of state-space machinery in [11], where the continuous-time case was considered under the assumption that $G_x(\infty) = 0$. The technique of [11], however, is not readily extendible to the discrete-time case, where the existence of performance saturation was so far known only through numerical examples. The present result is therefore the first proof of the existence of performance saturation in discrete time.

**Remark 4.1:** It might happen that $\alpha_{\gamma, 0} \leq \gamma_\infty$. This implies that the analytic constraints are inactive already in the filtering, $h = 0$, formulation. It is worth emphasizing that this situation does not necessarily require that $V = 0$. An example of such a problem is equalization with a minimum-phase communication channel, see [15, §15.4.1].

Having chosen $\gamma$ for which (16) holds, the next step is to find an admissible $Q_\gamma$, followed by $Q$, as defined by (14) (with $\hat{Q} = D_{-h}Q_x$), and finally $K$, via (3). The approach of [12] to construct $Q_x$ is to use the the fact that a partition

$$D_h V_{y\rightarrow}^\gamma = V_{y\rightarrow}^\gamma + \Delta$$

always exists, with a strictly proper $V_h \in H^\infty$ and an FIR (finite impulse response) $\Delta \in H^\infty$. Absorbing $\Delta$ into $Q_x$, (16) is reduced to an equivalent problem without delay

$$\|V_{y\rightarrow}^\gamma - Q_h\|_2 \leq \gamma, \quad Q_h := Q - \Delta \in H^\infty.$$

This is a finite dimensional problem, involving rational functions, and is solved by standard methods. The solution of $SP_G$ is thus complete. Continuous-time state-space formulæ for a suboptimal (i.e., $\gamma > \gamma_\infty$) solution, derived by this approach, can be found in [17].

**C. Discussion and comparison with $H^2$ case**

As mentioned above, our approach extends existing results on the $H^\infty$ smoothing performance saturation in the continuous-time case and leads to the first discrete-time proof. Not less important is that the proposed approach reveals underlying mechanisms behind the saturation phenomenon: it occurs when analytic constraints on achievable performance become inactive.

The source of geometric constraints is quite clear: they exist if the estimated signal $v$ is not co-directed with the measured signal $y$. Sources of analytic constraints are less evident. If $G_y$ is square, then it is readily seen that analytic constraints are caused by its nonminimum-phase (transmission) zeros. Indeed, if $G_y$ is minimum phase, $N_v = I$ and it follows from Proposition 4 that $V = N_v^\rightarrow = 0$, so that no analytic constraints are present. If $G_y$ does have unstable zeros, so does $N_v$, and these zeros are the poles of $V$ unless they are canceled by zeros of $N_v$. As follows from (1), common unstable zeros of $N_v$ and $G_y$ are also zeros of $G_v$. We thus may conclude that analytic constraints are caused by unstable zeros of $G_y$, which are not zeros of $G_v$. In this sense, analytic constraints play a generic interpolation condition role. The situation in the case where $G_y$ is not square (i.e., when $\dim y < \dim u$) is less transparent and analytic constraints might be present even if $G_y$ has no unstable zeros, as illustrated by a simple example in Section V. The poles of $V$, however, are the unstable poles of the Moore-Penrose pseudo-inverse of $G_y$ which are not canceled by unstable zeros of $G_v$, so that we may still think of analytic constraints as caused by unstable zero dynamics.

It is interesting to compare the $H^\infty$ result with $H^2$ optimization. In the latter, the problem is to minimize the $L^2$-norm of the error system, i.e.,

$$J = \|G_e\|^2_2 := \frac{1}{2\pi} \int_B \text{tr}(G_v(\lambda)G_v^\rightarrow(\lambda))|d\lambda|.$$

Taking into account (6) and the linearity of the matrix trace, the following equality can be obtained (we follow the steps in [18] here):

$$J = \|V^\rightarrow - Q\|^2_2 + \|\Phi_{\psi|y}\|^2_2.$$

The second term on the right-hand side, above, does not depend on $Q$. Hence, provided $\Phi_{\psi|y} \in L^2(S)$, the minimization of $J$ reduces to that of $\|V^\rightarrow - Q\|^2_2$ over $Q \in D_{-h}H^2$. This is a Hilbert space distance problem and, as such, is solvable by the orthogonal projection $Q_{opt} = \Pi_{D_{-h}H^2}V^\rightarrow$. The solution is an FIR system whose impulse response is the truncation of the impulse response of $V^\rightarrow$ to $[-h, 0]$ (if $h = 0$, $Q_{opt} = 0$ and (3) yields the Kalman filter $K = -M_v$).

With this choice, the optimal performance is

$$J_{opt} = \|\Pi_{D_{-h}H^2}V\|^2_2 + \|\Phi_{\psi|y}\|^2_2,$$

where the fact that $\Pi_{L^2(D_{-h})}V^\rightarrow = (\Pi_{D_{-h}}H^2 V^\rightarrow)$ was exploited. Clearly, $J_{opt}$ approaches $\|\Phi_{\psi|y}\|^2_2$ as $h \rightarrow \infty$ and the performance saturates iff $\Pi_{D_{-h}}H^2 V = 0$, for some finite $h$. The latter is obviously possible iff $V$ is FIR, which, in turn, might happen only in the discrete-time case (unless $V = 0$).

The qualitative difference between the effects of the smoothing lag $h$ on the $H^\infty$ and $H^2$ performance measures becomes apparent when conditions (10) and (18) are compared with (21). In both cases the achievable performance is restricted by both geometric (via $\Phi_{\psi|y}$) and analytic (via $V$) constraints and in both cases the effect of analytic constraints decreases exponentially as $h$ increases. Yet in the $H^2$ case the overall performance is the sum of these two components, whereas in the $H^\infty$ case, it is the maximum of the two. Thus, if present, analytic constraints always contribute to the $H^2$ performance
but they do not contribute to the $H_\infty$ performance when their potential contribution is dominated by the geometric constraint.

**D. Singular cases**

It follows from (21) that $\Phi_{\psi,Y}$ is the spectral density of the optimal estimation error in the least mean squares (Kalman) fixed-interval smoothing problem. The non-generic condition where $A_2$ fails is that where $\Phi_{\psi,Y}$ has all-pass maximal singular values. We shall now briefly discuss this situation. Throughout this section we assume that $V \neq 0$, which excludes the trivial case when the $H_\infty$ optimal solution for all $h$ is the Kalman filter $K = -M_v$ (corresponding to $Q = 0$).

1) $\Phi_{\psi,Y}(\lambda) = \gamma_\infty I$ (including $\gamma_\infty = 0$): In this case the solvability condition (8) reads $\|V^\infty - Q\|_\infty \leq \gamma_\infty^2$ and thus the optimal fixed-lag smoothing performance satisfies

$$y_{\infty}^2 = \min_{\gamma, \epsilon \in H_\infty} \|D_h V^\infty - Q\|_\infty^2 + \gamma_\infty^2.$$ 

Using the reasonings of $\S$IV-B, here $\gamma_\infty$ is an exponentially decaying function of $h$ approaching $\gamma_\infty$. Hence, performance saturates in this case only if $V$ is an FIR system. This, in turn, is only possible in the discrete-time case and corresponds to the (rare) situation when the Kalman filter yields an FIR error system. Otherwise, the smoothing performance approaches the optimal $L_\infty$ level of $\gamma = \gamma_\infty$ only as $h \to \infty$.

2) $\Phi_{\psi,Y}(\lambda) \neq \gamma_\infty I$: The situation here is more complicated. Let the normal rank of $1 - \gamma^2 \Phi_{\psi,Y}$ drops from $n_v := \dim v$ (the input and output dimension of $\Phi_{\psi,Y}$) for $\gamma > \gamma_\infty$, to $r$ satisfying $0 < r < n_v$, at $\gamma = \gamma_\infty$. In this case, as $\gamma \searrow \gamma_\infty$, the spectral factor $W_\gamma$ in (11) converges [16] to a singular transfer matrix, having the normal rank $r$. Without loss of generality we can then rewrite the spectral factorization in the limiting case as

$$W_\gamma = \begin{bmatrix} L & 0 \end{bmatrix} V^\infty = I_{n_v - r} - \gamma_\infty^2 \Phi_{\psi,Y},$$

where $W_\gamma \in H_{n_v \times n_v}$ is analytically invertible in $\mathbb{U}$, like in the nonsingular case. This leads to the solvability condition

$$W^{-1}_\gamma (V^\infty - Q)(V^\infty - Q)^\ast (W^{-1}_\gamma)^{\ast} \leq \gamma_\infty^2 \begin{bmatrix} L & 0 \end{bmatrix}.$$ 

As $W_\gamma$ might have zeros in $\mathbb{B}$, we have to limit our attention to $Q$’s guaranteeing that $W^{-1}_\gamma (V^\infty - Q) \in L_\infty$. At this point Lemma 2 applies and we end up with the inequality

$$(V^\infty - \tilde{Q})(V^\infty - \tilde{Q})^\ast \leq \gamma_\infty^2 \begin{bmatrix} L & 0\end{bmatrix}$$

for some strictly proper $V^\infty + H_\infty$. This inequality clearly requires that

$$\begin{bmatrix} 0 & I_{n_v - r} \end{bmatrix} (V^\infty + \tilde{Q}) = 0, \quad Q \in D_{-h} H_\infty. \quad (22)$$

Two situations are possible here:

- If $V^\infty + \tilde{Q}$ is FIR (or zero), then (22) can be satisfied for some $h$. The problem then reduces to the Nehari problem of finding $\min\| \begin{bmatrix} L & 0 \end{bmatrix} (V^\infty + \tilde{Q})\|_\infty$ and the arguments of $\S$IV-B apply. Thus, here we still have performance saturation.

- If $V^\infty + \tilde{Q}$ is IIR, then (22) cannot be satisfied with a finite $h$. This implies that saturation does not occur.

**V. AN ILLUSTRATIVE EXAMPLE**

To illustrate the ideas of the previous section, consider the following very simple continuous-time example:

$$G_v(s) = \begin{bmatrix} \frac{s^2 - \xi^2}{s^2 + \xi^2} & 0 \end{bmatrix} \quad \text{and} \quad G_\beta(s) = \begin{bmatrix} \frac{\beta^2 + 1}{\beta} & 1 \end{bmatrix}.$$ 

parametrized by the constants $\xi$ and $\eta$. If $\xi \geq 0$, this problem corresponds to a simulation problem, in which a (low-pass) signal, modeled as the output of the weighted function $\frac{s^2}{s^2 + \xi^2}$, is to be reconstructed after passing through the communication channel $\frac{s^2}{s^2 + \xi^2}$ and being corrupted by a measurement noise.

By tedious albeit straightforward computations it can be shown that in this case

$$V(s) = \frac{(1 - \eta \mu)(\mu + \xi)}{s + \mu} \quad \text{and} \quad \Phi_{\psi,Y}(s) = \frac{\mu^2(s^2 - \xi^2)}{s^2 - \mu^2},$$

where

$$\mu := \frac{1}{\sqrt{1 + \eta^2}} \in (0, 1].$$

Thus, $\gamma_\infty = \max(\{\xi, \mu\})$ and $A_2$ is satisfied iff $|\xi| \neq \mu$.

**A. Non-singular case: $|\xi| \neq \mu$**

We start with addressing the non-singular case. Straightforward calculations yield that

$$W^{-1}_\gamma(s) = \frac{\gamma(s + \mu)}{\sqrt{\gamma^2 - \mu^2}s + \mu} > \frac{\sqrt{\gamma^2 - \mu^2}s + \mu}{\sqrt{\gamma^2 - \mu^2}s + \mu},$$

which belongs to $H_\infty$ whenever $\gamma > \gamma_\infty$. If $\gamma = \gamma_\infty$, $W^{-1}_\gamma$ has either a pole at the origin (if $\gamma_\infty = |\xi| > \mu$) or is non-proper (if $\gamma_\infty > |\xi|$). In any event, the stable part of $V(W^{-1}_\gamma)^{-1}$,

$$V_{\gamma+}(s) = \frac{\gamma(1 - \eta \mu)(\mu + \xi)}{\sqrt{\gamma^2 - \mu^2}s + \mu},$$

is well defined. The next step is to calculate $\alpha_{\gamma,h}$ by (19). The key observation here is that, as a first-order system, the $L_2[h, \infty)$-norm of the response to an $L_2[(-\infty, 0)$ input (equivalently, the response to a non-zero initial condition) is always proportional to the $L_2[0, \infty)$-norm of the same response by the factor of $e^{-\mu h}$. An immediate consequence of this fact is that the norm in (19) is the Hankel norm of $V_{\gamma+}$ multiplied by $e^{-\mu h}$, i.e.,

$$\alpha_{\gamma,h} = \frac{\gamma(1 - \eta \mu)(\mu + \xi)}{\sqrt{\gamma^2 - \mu^2}s + \mu}.$$ 

Because $\alpha_{\gamma,h} \neq 0$, analytic constraints are always present, even though $G_v$ has no zeros.

Analytic constraints, however, are not necessarily active, even for $h = 0$. Indeed,

$$\alpha_{\gamma,0} = \frac{1 - \eta \mu}{\mu} \sqrt{\frac{|\mu + \xi|}{|\mu - \xi|}},$$

might be contractive, in which case (18) holds. The unshaded area in Fig. 2 represents this region in the $(\eta, \gamma)$-plane. For $\eta$ and $\gamma$ in that region the $H_\infty$ filter achieves the $L_\infty$ performance without any need in previewed information.

If $\alpha_{\gamma,0} > \gamma_\infty$ (the gray areas in Fig. 2), the fixed-interval performance level, $\gamma_\infty$, can no longer be achieved in the
filtering setting, and for optimal performance, the necessary preview length satisfies

\[ e^\eta h > \frac{1 - \eta \mu}{\mu} \sqrt{\frac{\mu + \zeta}{\mu - \zeta}}. \]

The minimal smoothing lag for which the optimal fixed-interval performance is achieved is then given by

\[ h_{\gamma_{\infty}} = \max \left\{ 0, \frac{1}{\mu} \ln \left( \frac{1 - \eta \mu}{\mu} \sqrt{\frac{\mu + \zeta}{\mu - \zeta}} \right) \right\}. \]

The (solid) contour lines in Fig. 2 represent the parameter sets for several given values of this \( h_{\gamma_{\infty}} \).

**B. Singular cases:** \( |\zeta| = \mu \)

If \( |\zeta| = \mu \), we have that \( \Phi_{V,Y}(s) = \mu^2 \). If \( \zeta \) is negative, i.e., if \( \zeta = -\mu \) and \( V(s) = 0 \) and there are no analytic constraints, in this case the fixed-interval performance is obviously reachable at \( h = 0 \). Indeed, the entire line \( \zeta = -\mu \) (the lower gray dashed line) is located within the unshaded region of Fig. 2.

If \( \zeta \) is positive, \( V(s) \neq 0 \) and the situation is as described in §IV-D1. It can be shown that in this case

\[ \gamma_{\text{opt}}^2 = (1 - \eta \mu)^2 e^{-2\eta h} + \mu^2, \]

which approaches the \( L^\infty \) performance level, \( \gamma_{\infty} = \mu \), only exponentially. The curve to which the contour lines in Fig. 2 converge is actually the curve satisfying \( \zeta \sqrt{1 + \eta^2} = 1 \) (the upper gray dashed line).

**ACKNOWLEDGMENT**

The authors would like to thank the anonymous referees for useful comments, leading to the expansion of the discussion in §IV-C and to the inclusion of this example.

**VI. CONCLUDING REMARKS**

The solution of the fixed-lag smoothing problem has been derived as a constrained version of the fixed-interval smoothing problem. This approach leads to a partition of the factors determining smoother performance. Geometric constraints are independent of the smoothing lag, and are inherited from the fixed-interval problem. Analytic constraints, arising due to the presence of nonminimum-phase zeros, decay exponentially as the smoothing lag grows. The generic situation is therefore that where the optimal fixed-lag smoothing performance reaches its minimum—the optimal fixed-interval performance—and saturates, with a finite-lag. The paper also analyzes the singular exceptions to this general description.

**REFERENCES**