

# CONTRACTION, SMOOTHNESS, AND LOW-PASS FILTERING

Mohammad Ali Khojastepour, Behnaam Aazhang, Richard G. Baraniuk

Department of Electrical and Computer Engineering  
Rice University  
Houston, Texas, USA

## ABSTRACT

We introduce a generalized definition for “low-pass” filters that covers time-varying and nonlinear systems under the same umbrella. We show that the qualitative concept of signal smoothing can be made precise through the concept of contractions in probabilistic metric spaces. For illustration, we consider classical linear time-invariant low-pass filters, the nonlinear median filters, and time-varying guaranteed maximum delay schedulers employed in communication systems.

## 1. INTRODUCTION

Linear filters have played a central role in the history of signal processing and communications. Thanks to their strong theoretical basis and computational efficiency, linear filters have dominated over nonlinear and time-varying filters. However, despite their elegant theory, linear filters do not satisfactorily address all signal processing problems. Unfortunately, it has been difficult to design, optimize, and evaluate the performance of the non-linear and time-varying systems due to the lack of a sound underlying theory.

In this paper, we take a fresh look at *low-pass systems*, introducing a new definition that extends directly to nonlinear and time-varying filters. Interestingly, the definition is based on random signals and contractions in probabilistic metric spaces. This work can be considered as a step toward a unified theory of linear, nonlinear, and time-varying filtering.

Using our definition, we study three classes of low-pass systems: standard linear time invariant (LTI) filters, median filters [1–3], and schedulers that guarantee a maximum delay in a communication system [4]. It has been shown that schedulers act like filters that are not necessarily linear or time-invariant.

This paper is organized as follows: In Section 2, we provide some intuition toward our general definition of low-

pass filters. Then, we present a fundamental property of low-pass filters that leads to our main result. In Section 3, we show how conventional low-pass LTI systems are characterized by our new approach. In Section 4 we apply the definition to the above-mentioned nonlinear time-varying (NLTV) systems. Finally, we conclude in Section 5.

## 2. CONTRACTION AND LOW-PASS SYSTEMS

The superposition principle does not hold in nonlinear systems. Indeed, a non-linear or time-varying system output can even contain frequencies not present in the input signal. Hence, it is not possible to analyze the behavior of the output signal of a non-linear system by means of a signal decomposition. Rather, the behavior in response to *all possible* inputs should be considered in order to characterize a system as low-pass. The approach we will pursue uses a class of stochastic input sequences to characterize systems.

The present definition of a low-pass systems is more qualitative than quantitative. We typically deem a system low-pass when it “reduces noise” and the output is (in some sense) “smoother” than input. Moreover, most definitions of low-pass filters [5] use the notion of the frequency response (decay at high frequencies) or pole-zero representation of the filter and therefore are specific to LTI systems.

The notion of signal smoothness has been defined in different ways. The local monotonicity [6] or differentiability order of a signal [7] are examples of such definitions. A higher differentiability order of a wavelet function corresponds to a larger number of vanishing moments, which can be considered as a measure of smoothness [7]. There is an interesting connection between low-pass systems and the concept of a contraction in a probabilistic metric space. (We will focus on discrete-time signals and systems.) We consider the system responses to all possible stationary and white stochastic input sequences  $X[t]$ , which are fully characterized by their first order statistics (pdf)  $f_X(x)$ . We assume that the output signal  $Y[t]$  of the system to the stationary input is a stationary random process; let  $f_Y(y)$  denote its first order statistics (pdf).

Considering just the first order statistics of the input and

---

This work was supported by NSF and the Texas Instruments Leadership University Program. Thanks to Mike Orchard and Don Johnson for their helpful comments and discussion.

Email: {amir, aaz, richb}@rice.edu. Web: cmc.rice.edu, dsp.rice.edu.

output signals, it can be shown that for any low-pass LTI system the output space is a probabilistic contraction of the input space when the input and the output signal powers are normalized along their respective means. This metric-space contraction is defined as a mapping  $\mathcal{G}(x) : \mathcal{X} \rightarrow \mathcal{Y}$ , such that

$$\begin{aligned} \exists 0 < \gamma < 1 \forall x_1, x_2 \in \mathcal{X} : \\ \|\mathcal{G}(x_1) - \mathcal{G}(x_2)\| < \gamma \|x_1 - x_2\|. \end{aligned} \quad (1)$$

The generalization of a contraction mapping to a probabilistic metric space is considered by Chang et. al. in [8, 9]. We consider a simple measure of probabilistic contraction

$$\forall f_X(x) : \mathbb{E}[\eta^Y] \leq \mathbb{E}[\eta^X] \quad (2)$$

where  $\eta \in \mathcal{R}$ ,  $\eta > 1$ , and  $Y$  is the random variable representing the output of the system at any given time. Note that for a stationary white input process, the output of an LTI “low-pass” filter (defined as in [5]) has smaller variation. Therefore, we see that LTI low-pass filters correspond to contractions. Furthermore, we will show below that any LTI system having this contraction property is a low-pass filter (Theorem 2 specifically considers FIR filters). Thus, we can conclude that an LTI system is low-pass if its response to a stationary white input process is a stationary output process for which the output space is a contraction of the input space.

This same notion can be easily extended to nonlinear or time-varying (NLTV) systems. We will consider two important examples in this class, namely median filters and scheduling systems. In both cases, we show that the contraction principle holds, which leads us to conclude that these non-linear systems are low-pass in nature.

Formally, we define a low-pass system as:

**Definition 1 (Low-pass)** *A system that has stationary output  $Y$  for every stationary white input  $X$  is defined to be low-pass if (i)  $\forall f_X(x) : \mu_X \neq 0 \Rightarrow \mu_Y \neq 0$ , and*

$$(ii) \forall f_X(x) : \mathbb{E}[\eta^{kY}] \leq \mathbb{E}[\eta^X] \quad (3)$$

where  $\eta \in \mathcal{R}$  and  $\eta > 1$ . Also,  $\mu_X = \mathbb{E}[X]$ ,  $\mu_Y = \mathbb{E}[Y]$ , and  $k$  is a scaling constant such that  $k\mu_Y = \mu_X$ .

Let  $|H_X(f)|^2$  denote the ratio of the output Fourier spectrum to the input spectrum for a given input distribution. Then the first condition of the definition can be interpreted as  $|H_X(0)|^2 \neq 0$ . The fact that scaling should not change the low-pass nature of a filter has been captured by the constant  $k$ .

### 3. LTI LOW-PASS FILTERS AND THE DEFINITION

**Proposition 1** *Any weighted moving average filter with positive weights is a low-pass system under Definition 1.*

*Proof:* Consider a weighted linear filter of the form

$$Y[t] = a_1 X[t] + a_2 X[t-1] + \dots + a_n X[t-n+1] \quad (4)$$

with the filter weights  $a_i, i = 1, \dots, n$ . From the first condition of Definition 1 we have  $\sum_{i=1}^n a_i \neq 0$ . Thus, without loss of generality we assume that  $\sum_{i=1}^n a_i = 1$  by choosing an appropriate scaling factor  $k$ . Then, We have

$$\mathbb{E}[\eta^Y[t]] = \mathbb{E}[\eta^{a_1 X[t] + a_2 X[t-1] + \dots + a_n X[t-n+1]}] \quad (5)$$

$$= \mathbb{E}[\eta^{a_1 X[t]}] \mathbb{E}[\eta^{a_2 X[t-1]}] \dots \mathbb{E}[\eta^{a_n X[t-n+1]}] \quad (6)$$

$$= \mathbb{E}[\eta^{a_1 X[t]}] \mathbb{E}[\eta^{a_2 X[t]}] \dots \mathbb{E}[\eta^{a_n X[t]}] \quad (7)$$

$$\leq \mathbb{E}^{a_1}[\eta^{X[t]}] \mathbb{E}^{a_2}[\eta^{X[t]}] \dots \mathbb{E}^{a_n}[\eta^{X[t]}] \quad (8)$$

$$= \mathbb{E}^{a_1 + a_2 + \dots + a_n}[\eta^{X[t]}] = \mathbb{E}[\eta^{X[t]}] \quad (9)$$

where (6) holds because the input process is white and (7) holds thanks to its stationarity. Also, since the coefficients are all positive and sum to one, we have  $0 \leq a_i \leq 1$  for all  $i = 1, 2, \dots, n$ . Thus, (8) is clear from the application of the Jensen’s inequality. ■

On the other hand we can show that any FIR filter satisfying Definition 1 must be a low-pass.

**Proposition 2** *Any FIR filter satisfying Definition 1 is a low-pass filter in the conventional sense (see [5], for example).*

*Proof:* Consider an FIR filter defined as

$$Y[t] = a_1 X[t] + a_2 X[t-1] + \dots + a_n X[t-n+1] \quad (10)$$

where  $a_i, i = 1, \dots, n$  are the filter coefficients. As discussed in the proof of Proposition 1, without loss of generality we assume that  $\sum_{i=1}^n a_i = 1$ . We now show that if the filter (10) satisfies the Definition 1, then  $0 \leq a_i \leq 1$  for all  $i = 1, 2, \dots, n$ . We prove this by contradiction. Assume that the filter satisfies Definition 1 and that the maximum value of its coefficients is  $a_j$  with  $a_j > 1$ . Consider a discrete input distribution where  $x_{\max}$  is the maximum value of the random variable  $X$  that has nonzero probability. Consider both  $\mathbb{E}[\eta^{Y[t]}]$  and  $\mathbb{E}[\eta^{X[t]}]$  as functions of  $x_{\max}$ . We have

$$\mathbb{E}[\eta^{Y[t]}] = \mathbb{E}[\eta^{a_1 X[t] + a_2 X[t-1] + \dots + a_n X[t-n+1]}] \quad (11)$$

$$= \mathbb{E}[\eta^{a_1 X[t]}] \mathbb{E}[\eta^{a_2 X[t-1]}] \dots \mathbb{E}[\eta^{a_n X[t-n+1]}] \quad (12)$$

$$= O(\eta^{a_j x_{\max}}). \quad (13)$$

On the other hand,  $\mathbb{E}[\eta^{X[t]}] = O(\eta^{x_{\max}})$ . Since we assumed that  $a_j > 1$ , for large enough  $x_{\max}$  we have that  $\mathbb{E}[\eta^{Y[t]}] > \mathbb{E}[\eta^{X[t]}]$ . Therefore, there exists a distribution for the input process  $X[t]$  such that  $x_{\max}$  is large enough such that it leads to a contradiction.

Now, let the minimum value of the filter coefficients be  $a_j$ . We can follow the same line of argument to show that if  $a_j < 0$  then filter does not satisfy Definition 1. Consider a discrete input distribution where  $x_{\min} < 0$  is the minimum value of the random variable  $X$  which has nonzero probability. Consider both  $\mathbb{E}[\eta^Y[t]]$  and  $\mathbb{E}[\eta^X[t]]$  as a function of  $x_{\min}$ . As in (11-13), we can show that  $\mathbb{E}[\eta^Y[t]] = O(\eta^{a_j x_{\min}})$ . On the other hand,  $\mathbb{E}[\eta^X[t]] = O(\eta^{x_{\min}})$ . Since we assumed that  $a_j < 0$  and also that  $x_{\min} < 0$  for large enough  $|x_{\min}|$  we have  $\mathbb{E}[\eta^Y[t]] > \mathbb{E}[\eta^X[t]]$ . Therefore, there exists a distribution for input process  $X[t]$  such that  $|x_{\min}|$  is large enough such that it leads to a contradiction.

Thus, we have proved that  $0 \leq a_i \leq 1$  for all  $i = 1, 2, \dots, n$ , which means that the filter is a weighted moving average filter that is low-pass in the conventional sense [5]. ■

It should be noted that Definition 1 does not subsume all known linear low-pass filters. A better choice of convex function  $c(x)$  instead of  $c(x) = \eta^x$  might lead to a better characterization of low-pass filters. However, could make the analysis more complicated. An example of such a convex function is

$$c(x) = \begin{cases} e^x - 1 & x > 0 \\ -\ln(1 - x) & x \leq 0. \end{cases} \quad (14)$$

It is easily verified that the second derivative of this function exists and is always positive; that is, the function is strictly convex. Furthermore, it has attractive property of being infinitely differentiable.

#### 4. NLTV LOW-PASS FILTERS AND THE DEFINITION

Definition 1 provides a framework to evaluate the low-pass characteristics of an extensive class of nonlinear and time-varying filters. In this section we present two examples of systems that should be rated in the class of low-pass filters. We discuss how Definition 1 leads to such characterizations.

##### 4.1. Median filters

Median filters are used widely in signal processing, especially in image processing for noise reduction as opposed to linear smoothing filters. The reason is that median filters tend to preserve the sharpness of image edges while removing noise [1, 3]. Median filters can be extended to weighted median filters, center weighted median filters, and max-median filters [2, 3].

**Proposition 3** *Median filters are low-pass in the sense of Definition 1.*

*Proof:* Because of space limitations we present only a sketch of the proof. Consider a median filter of length  $n$

defined as

$$Y[t] = \text{median}(X[t], X[t-1], \dots, X[t-n+1]). \quad (15)$$

Since the input process is white with distribution  $f_X(x)$ , the output at time  $t$  is simply the median of  $n$  independent samples of  $X$ , say  $X_1, X_2, \dots, X_n$ , defined as

$$\tilde{X} = \begin{cases} \frac{1}{2}(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}) & \text{if } n \text{ is even} \\ X_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \end{cases} \quad (16)$$

where  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  denote the order statistics of the random sample  $X_1, X_2, \dots, X_n$ . The distribution of the order statistics for continuous random variable  $X$  with cdf  $F_X(x)$  and pdf  $f_X(x)$  is given by

$$f_{X_{(j)}} = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1-F_X(x)]^{n-j} \quad (17)$$

Thus, the distribution of the output of a median filter of length  $n$  can be written in the form  $f_Y(x) = f_X(x)g(x)$ . It is not difficult to show that  $g(x)$  is an increasing function for  $x \in (-\infty, x_{\text{median}}]$  and a decreasing function for  $x \in [x_{\text{median}}, \infty)$ . Thus, the distribution of the output  $Y[t] = \tilde{X}$  is more concentrated toward the median of the input distribution,  $x_{\text{median}}$ . Considering any strip  $[y, y + dy)$  of the output pdf there are two strips of  $[x_1, x_1 + dx)$  and  $[x_2, x_2 + dx)$  of the input pdf such that their probabilities sum up to the probability of the strip  $[y, y + dy)$  and also  $x_1 < y < x_2$ . Because of the convexity of the function  $\eta^X$  in Definition 1, the expected value of  $\eta^y$  over the interval  $[y, y + dy)$  is smaller than the expected value of  $\eta^x$  over the two intervals  $[x_1, x_1 + dx)$  and  $[x_2, x_2 + dx)$ . Therefore, the summation over the domain of the output pdf function  $f_Y(y)$  gives the desired result. ■

The same analysis can be performed for weighted median filters [2] and center weighted median filters [3] to show that both are low-pass.

##### 4.2. Schedulers Guaranteeing a Maximum Delay

Most multimedia sources are bursty in nature, a property that can be used to trade queuing delay for reduced average transmission power [10]. Scheduling deals with the problem of how to choose the output service rate of the queues based on the input arrivals in order to optimize the use of system resources, usually average power, while maintaining some quality of service (QoS requirements), usually packet delays. In [4], a connection between filtering and scheduling with a guaranteed maximum delay  $D_{\max}$  has been established. It has been shown that the necessary and sufficient conditions for any scheduler that guarantees a maximum delay  $D_{\max}$  are given by these inequalities

$$\sum_{i=t}^{t+k} X[i] \leq \sum_{i=t}^{t+k+D_{\max}-1} Y[i], \quad (18)$$

$$\sum_{i=t}^{t+k} Y[i] \leq B[t] + \sum_{i=t}^{t+k} X[i], \quad (19)$$

for any  $t$  and  $k$  where  $X[t]$  is the rate of the input arrival to the queue,  $Y[t]$  is the output service rate of the queue, and  $B[t]$  is the queue backlog at time  $t$ . Based on these conditions, it has been shown that if the scheduler is time invariant, then the scheduler is always low-pass [4]. This property has an attractive interpretation in terms of power-efficiency. It is intuitive that additional delay helps to reduce the required average power for the scheduler by smoothing the input arrival process via queuing.

The same story is not true for the NLTV schedulers. In fact, it is possible to imagine an intuitively high-pass NLTV filter that guarantees the same maximum delay. However, with the general Definition 1, we can show that the optimal (not necessarily LTI) scheduler is also low-pass. In fact, the required average power of the scheduler is proportional to

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N 2^{Y[t]} \quad (20)$$

which can be written in the form  $\mathbb{E}[2^Y]$  under stationary conditions. Considering the fact that the optimal scheduler has the minimum average power, we have

$$\mathbb{E}[2^Y] \leq \mathbb{E}^{D_{\max}}[2^{X/D_{\max}}], \quad (21)$$

where  $\mathbb{E}^{D_{\max}}[2^{X/D_{\max}}]$  is the required average power of a simple moving average filter of size  $D_{\max}$ . Thus, based on Jensen's inequality, we have

$$\mathbb{E}[2^Y] \leq \mathbb{E}^{D_{\max}}[2^{X/D_{\max}}] \leq \mathbb{E}[2^X] \quad (22)$$

which means that the optimal scheduler satisfies the second condition of Definition 1. Also, since we assume that all packets will be delivered with maximum delay of  $D_{\max}$  and that no packets will be dropped, the average input arrivals is equal to the average output service rates. In other words, every packet that arrives at the queue eventually exits the queue. Thus, the first condition is also satisfied.

Furthermore, if a scheduler requires power  $P_1$  to guarantee a maximum delay  $D_{\max}$  and is low-pass under Definition 1, then any scheduler that guarantees the same maximum delay  $D_{\max}$  with lower power  $P_2 < P_1$  is also low-pass. In other words, it is impossible to beat the performance of a low-pass scheduler with a scheduler that is not low-pass. This property again confirms the intuition behind the fact that the output of the optimal scheduler is smoother than the input.

## 5. CONCLUSIONS

We have developed a general framework to help understand and unify LTI and NLTV low-pass systems. This work is

only the first step toward a complete characterization; there is still much more to be done. A good next step would be to consider the second-order temporal dependencies of the input and output signals, which we have not considered in this paper.

## 6. REFERENCES

- [1] J.W. Tukey, "Nonlinear (nonsuperposable) methods for smoothing data," in *Congr. Rec. EASCON*, 1974, p. 673, (Abstract only).
- [2] D.R.K. Brownrigg, "The weighted median filter," *Commun. Assoc. Comput. Machin.*, vol. 27, no. 8, pp. 807–818, 1984.
- [3] S.-J. Ko and Y.H. Lee, "Center weighted median filters and their applications to image enhancement," *IEEE Trans. Circ. Syst.*, vol. 38, no. 9, pp. 984–993, 1991.
- [4] M. A. Khojastepour and A. Sabharwal, "Power optimal scheduling with maximum delay constraints," in *41<sup>st</sup> Annual Allerton Conf.*, 2003.
- [5] J. H. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, and Applications*, Prentice-Hall: New Jersey, 1996.
- [6] A. Restrepo and A. C. Bovik, "Locally monotonic regression," in *IEEE Trans. on Signal Processing*, 1993.
- [7] J. E. Odegard and C. S. Burrus, "Toward a new measure of smoothness for the design of wavelet basis," in *Proc. ICASSP*, 1996.
- [8] S. S. Chang et. al., "Generalized contraction mapping principle and differential equations in probabilistic metric spaces," *Proc. AMS*, vol. 124, no. 8, 1996.
- [9] S. S. Chang, Y. J. Cho, and S. M. Kang, *Probabilistic Metric Spaces and Nonlinear Operator Theory*, Sichuan Univ. Press, Chengdu, 1994.
- [10] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.