Structurally Robust Weak Continuity

N. D. Sidiropoulos, J. S. Baras, and C. A. Berenstein
Institute for Systems Research
University of Maryland
College Park, MD 20742
nikos@glue.umd.edu

Abstract

We pose the following optimization: Given $\mathbf{y} = \{y(n)\}_{n=0}^{N-1} \in \mathbf{R}^N$, *find a finite-alphabet* $\widehat{\mathbf{x}} = \{\widehat{x}(n)\}_{n=0}^{N-1} \in \mathcal{A}^N$, *that minimizes* $d(\mathbf{x}, \mathbf{y}) + g(\mathbf{x})$ *subject to:* $\mathbf{x}$ *satisfies a hard structural (syntactic) constraint, e.g.,* $\mathbf{x}$ *is piecewise constant of plateau run-length* $\geq M$, *or locally monotonic of lomo -degree* α. *Here,* $d(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{N-1} d_n(y(n), x(n))$ *measures fidelity to the data, and is known as the* noise term, *and* $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$ *measures smoothness-complexity of the solution. This optimization represents the unification and outgrowth of several digital nonlinear filtering schemes, including, in particular, digital counterparts of* Weak Continuity (WC) *[6, 7, 2], and* Minimum Description Length (MDL) *[4] on one hand, and nonlinear regression, e.g.,* VORCA *filtering [11], and* Digital Locally Monotonic Regression *[10], on the other. It is shown that the proposed optimization admits efficient Viterbi-type solution, and, in terms of performance, combines the best of both worlds.*

1 Introduction

One of the classic problems in the true spirit of nonlinear filtering is that of detecting and estimating edges in noise. Among the great many approaches proposed so far, a particularly noteworthy one is the (nonconvex) variational *Weak Continuity (WC)* paradigm of Mumford-Shah [6, 7] and Blake-Zisserman [2] (see also the excellent recent book by Morel and Solimini [5]). Weak continuity is a rigorous paradigm for edge detection, which attempts to fit *piecewise*-smooth candidate "interpretations" to the observable data (thus the term *weak* continuity).

In real life we nowadays most often deal with digital data, i.e., sequences of finite-alphabet variables. Following Blake and Zisserman [2], we present a digital version of discrete-time WC. Given a (generally real-valued) sequence of finite extent $\mathbf{y} = \{y(n)\}_{n=0}^{N-1} \in \mathbf{R}^N$, the problem is to find a finite-alphabet sequence, $\widehat{\mathbf{x}} = \{\widehat{x}(n)\}_{n=0}^{N-1} \in \mathcal{A}^N$ (the "reproduction process"), that minimizes

$$\sum_{n=0}^{N-1} (y(n) - x(n))^2 + \sum_{n=1}^{N-1} h_{\alpha,\lambda_{WC}}(x(n) - x(n-1))$$

where

$$h_{\alpha,\lambda_{WC}}(t) = \begin{cases} \lambda_{WC}^2 t^2 & , t^2 < \frac{\alpha}{\lambda_{WC}^2} \\ \alpha & , otherwise \end{cases}$$

There exist essentially two ways to go about solving this problem: Dynamic Programming (DP) [1], and the so-called *Graduated Non Convexity (GNC)* algorithm [2]. For one-dimensional data, DP is probably the best way to go. According to Blake and Zisserman [2], Papoulias [8] was the first to implement a DP WC algorithm. The drawback of DP is that it does not generalize in higher dimensions, for lack of total ordering. The GNC, by comparison, carries over quite effortlessly in higher dimensions.

A related optimization has been advocated by Leclerc [4], based on the *Minimum Description Length (MDL)* principle of Rissanen; the goal is the minimization of:

$$\sum_{n=0}^{N-1} \frac{(y(n) - x(n))^2}{\sigma^2} + \sum_{n=1}^{N-1} \lambda_{MDL} \left[1 - \delta(x(n) - x(n-1))\right]$$

where δ is the Kronecker delta function, and σ^2 is noise variance. Here, $\lambda_{MDL} \geq 0$.

2 Unification and Motivation

Both WC and MDL seek to minimize a nonconvex cost of the following general form

$$\mathcal{V}(\mathbf{y}, \mathbf{x}) = \sum_{n=0}^{N-1} d_n(y(n), x(n)) + \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$$

In the digital world, Leclerc's MDL formulation is a special case of WC. Indeed, if λ_{WC} is sufficiently large (i.e., $\lambda_{WC}^2 > \alpha$), then, t being integer, $h_{\alpha,\lambda_{WC}}(t) = \alpha\left[1 - \delta(t)\right]$. If, in addition, $\alpha = \lambda_{MDL}\sigma^2$, then WC reduces to Leclerc's MDL approach.

Both WC, and Leclerc's MDL approach are powerful and meritorious paradigms; however, both share a nontrivial shortcoming: they are not robust with respect to outliers, in the sense of being susceptible to noise-induced "impulses". Consider a single such outlier, i.e., a Kronecker delta of height Δ. If $(\frac{\Delta}{\sigma})^2 > 2\lambda_{MDL}$, then Leclerc's MDL approach will preserve this "impulse"; similarly, if $\Delta^2 > \frac{\alpha}{\lambda_{WC}^2}$, and $\Delta^2 > 2\alpha$, then WC will also preserve it. Observe that these statements should be interpreted as follows: for each given choice of respective optimization parameter(s), one can find a sufficiently large Δ which forces both "filters" to preserve "impulses" of height $\geq \Delta$. In the context of edge detection in impulsive noise, this behavior is undesirable; these "impulses", no matter how powerful, should not be preserved [12].

"Traditional" nonlinear filters (e.g., the root of the median) are robust with respect to outliers. This robustness stems from the fact that the implicit goal of these filters is to enforce (albeit suboptimally) "hard" structural (syntactic) constraints on the data, e.g., of the type $\mathbf{x}$ is piecewise constant of plateau run-length $\geq M$, or locally monotonic of lomo-degree α. How to optimally enforce such constraints has been the subject of previous work by the first author in so-called VORCA filtering [11] and digital locally monotonic regression [10]. VORCA filtering amounts to solving:

$$\textbf{minimize} \sum_{n=0}^{N-1} d_n(y(n), x(n))$$

$$\textbf{subject to} : \mathbf{x} = \{x(n)\}_{n=0}^{N-1} \in P_M^N$$

where P_M^N is the set of all sequences of N elements of $\mathcal{A}$ which are piecewise constant of plateau (run) length $\geq M$.

A real-valued sequence (string), $\mathbf{x}$, of length N, is *locally monotonic* of degree $\alpha \leq N$ (or *lomo-α*, or simply *lomo* in case α is understood) if each and every one of its substrings of α consecutive symbols is monotonic. Local monotonicity appears in the study of the set of root signals of the median filter [3]; it constraints the roughness of a signal by limiting the rate at which the signal undergoes changes of trend (increasing to decreasing or vice versa). In effect, it *limits the frequency of oscillations, without limiting the magnitude of jump level changes that the signal exhibits.* Local monotonicity implies a different notion of smoothness, as compared to e.g., limiting the support of the Fourier transform; the latter imposes a limit on *both* the frequency of oscillations, *and* the magnitude of jump level changes.

In [9], Restrepo and Bovik developed an elegant mathematical framework in which they studied locally monotonic regressions in $\mathbf{R}^N$. *Digital* locally monotonic regression has been proposed in [10], and it amounts to solving:

$$\textbf{minimize} \sum_{n=0}^{N-1} d_n(y(n), x(n))$$

$$\textbf{subject to} : \mathbf{x} = \{x(n)\}_{n=0}^{N-1} \in \Lambda(\alpha, N, \mathcal{A})$$

where $\Lambda(\alpha, N, \mathcal{A})$ is the set of all sequences of N elements of $\mathcal{A}$ which are locally monotonic of lomo-degree α [10]. Both approaches are robust, in the sense of suppressing impulse-like inputs, while retaining "true" (consistent) edge signals. However, both do not take into account the significance of level changes ("discontinuities") in the solution, i.e., they may declare an edge even when the two resulting levels are very close. This is often undesirable; and it happens exactly because the latter two approaches do not *explicitly* account for smoothness/complexity, i.e., unlike WC, they do not incorporate a "soft" smoothness/complexity penalty into the cost function.

3 Structurally Robust Weak Continuity

It appears quite natural, then, to combine the power of WC with the appeal and demonstrated effectiveness of "hard" structural constraints, and propose the minimization of:

$$\sum_{n=0}^{N-1} d_n(y(n), x(n)) + \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$$

$$\textbf{subject to} : \mathbf{x} \in \mathcal{S}$$

where $\mathcal{S}$ is the set of all sequences of N elements of $\mathcal{A}$ satisfying some local "hard" structural constraint. Here, again, $d(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{N-1} d_n(y(n), x(n))$ is a fidelity measure, and $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$ is a smoothness-complexity measure. We will refer to this optimization as *Structurally Robust Weak Continuity (SR-WC).* When $\mathcal{S} = P_M^N$, *Runlength-Constrained Weak Continuity (RC-WC)* results; similarly, if $\mathcal{S} = \Lambda(\alpha, N, \mathcal{A})$, then *Locally Monotonic Weak Continuity (LM-WC)* results. Both retain the unique merits of WC, are robust with respect to outliers, take complexity into consideration, and admit efficient Viterbi-type solution. In fact, RC-WC, and LM-WC can be solved using exactly the same resources and computational structures as VORCA, and digital locally monotonic regression, respectively [12]. The extension to weak continuity (i.e., the incorporation of the first-order smoothness-complexity measure $g(\mathbf{x}) = \sum_{n=1}^{N-1} g_n(x(n), x(n-1))$ into the cost functional) essentially comes "for free", due to the structure of the Viterbi Algorithm. The resulting complexity of RC-WC, LM-WC is $O((|\mathcal{A}|^2 + |\mathcal{A}|(M-1))N)$, $O(|\mathcal{A}|^2 \alpha N)$, respectively.

By virtue of the above, efficient computation of SR-WC can be taken for granted. What is intriguing and unexplored is how to go about choosing fidelity and smoothness/complexity measures. We know that, at least for some specific choices, e.g., "classic" WC, MDL, or VORCA, we may expect very good nonlinear filtering results. The question is, can we make even better choices, and in what sense. This is partially explored in the following.

4 Example

This particular example demonstrates the effectiveness of simple RC-WC. Figure 1 depicts a typical input sequence. This particular input has been generated by adding i.i.d. noise on some artificial "true" noise-free test data. The noise has been generated according to a mixture of a uniform distribution and an "outlier" distribution, the mixture being heavily weighted in favor of the uniform distribution, and most of the data points are contaminated. It should be stressed that we do not utilize our exact knowledge of the noise model to fully match the optimization to the noise characteristics, which is certainly a possibility [11, 10, 9]. Instead, as it will be explained shortly, we only use some crude noise measurements to help us pick reasonable values for two optimization parameters. The noise-free test data has not been reproduced on its own, due to space limitations; instead, it has been overlaid on the restoration plots, using a dashed line. This is meant to help the reader judge filtering "quality".

Here we take $d_n(y(n), x(n)) = |y(n) - x(n)|$, $\forall n \in \{0, 1, \cdots, N-1\}$, and $g_n(x(n), x(n-1)) = \lambda_{WC}^2 \left[1 - \delta(x(n) - x(n-1))\right]$, $\forall n \in \{0, 1, \cdots, N-1\}$, $\mathcal{A} = \{0, \cdots, 99\}$, $N = 512$, and $\mathcal{S} = P_M^N$.

For $M = 1$, we obtain "plain" WC, and the result for $\lambda_{WC}^2 = 25$ is depicted in Figure 2. This is excellent filtering, yet powerful outliers are preserved. We could, in principle, further increase λ_{WC}^2, thereby eventually eliminating outliers, but, at the same time, also "mending" true edges. Clearly, this is not the way to go about ameliorating this problem, for, no matter what our choice of λ_{WC}^2 is, one can always find a sufficiently powerful outlier that will fool WC.

For $\lambda_{WC}^2 = 0$, we obtain "plain" VORCA, and the result for $M = 15$ is depicted in Figure 3. This too is excellent filtering, the outliers have been effectively eliminated, yet some undesirable "weak" edges still remain. For $\lambda_{WC}^2 = 25$, and $M = 15$ we have "true" hybrid RC-WC, and the result is depicted in Figure 4. It is obvious that RC-WC combines the power of both methods: this is, indeed, almost perfect filtering.

One obvious objection may be anticipated: one may wonder about how we came up with the particular choices of M, λ that led to these results. In the following, we address this subject.

4.1 Selection of Optimization Parameters

We will use the following definitions. A *feature (outlying burst) of width* $w < M$ is a "short" arbitrary deviation from a plateau, consisting of a total of w perturbed samples. A *constant segment of saliency (width-strength product)* $\mu = w \cdot H$ is a (potentially long) equidistant deviation from a plateau, i.e., a string of w equal samples which differ by H from the plateau level.

The following two claims refer to *this particular instance of RC-WC*, i.e., $d_n(y(n), x(n)) = |y(n) - x(n)|$, $\forall n \in \{0, 1, \cdots, N-1\}$, and $g_n(x(n), x(n-1)) = \lambda_{WC}^2 \left[1 - \delta(x(n) - x(n-1))\right]$, $\forall n \in \{0, 1, \cdots, N-1\}$. Proofs can be found in [12].

Theorem 1 *Assume that M is odd. RC-WC eliminates all features (outlying bursts) of width $w \leq \frac{M-1}{2}$, regardless of λ_{WC}^2, and the same is true for $\lambda_{WC}^2 = 0$, i.e., "plain" VORCA filtering with respect to the above choice of $d_n(\cdot, \cdot)$.*

Theorem 2 *RC-WC suppresses all constant segments of saliency (width-strength product) $\mu = w \cdot H < 2\lambda_{WC}^2$, i.e., "mends" the "weak" edges at the endpoints of such segments, and the same holds for $M = 1$, i.e., "plain" WC with respect to the above choice of $d_n(\cdot, \cdot)$, $g_n(\cdot, \cdot)$.*

The overall conclusion is that this particular instance of RC-WC suppresses features of either (i) width $w \leq \frac{M-1}{2}$ (M: odd), *regardless of strength*, or (ii) saliency (width-strength product) $\mu = w \cdot H < 2\lambda_{WC}^2$. This allows us to essentially separately fine-tune two important aspects of filter behavior. In a nutshell, M controls outlier rejection, whereas λ_{WC}^2 controls residual ripple.

5 Conclusions

Motivated by the power of WC-based methods [6, 7, 2, 4], "complementary" previous work by the first author in optimal nonlinear filtering under "hard" structural (so-called syntactic) constraints [11, 10], and realizing that a potential shortcoming of WC could be ameliorated by introducing "hard" structural constraints, whereas a drawback of the methods of [11, 10] could be rectified by introducing "soft" weak continuity constraints, we have posed, solved, and demonstrated the effectiveness of a novel hybrid optimization, dubbed Structurally Robust Weak Continuity, combining the advantages while avoiding the shortcomings of its constituent elements. SR-WC includes its constituent elements as special cases, and inherits efficient Viterbi implementation from [11, 10]. What is most intriguing is how to go about choosing fidelity and smoothness/complexity measures. This deserves further investigation, and long-term research in this direction is currently underway.

References

[1] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.

[2] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, Mass., 1987.

[3] N. Gallagher. Median filters: a tutorial. In *Proc. IEEE Int. Symp. Circ., Syst., ISCAS-88*, pages 1737–1744, 1988.

[4] Y. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *Int. J. Computer Vision*, 3(1):73–102, 1989.

[5] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhauser, Boston-Basel-Berlin, 1994.

[6] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, 1985.

[7] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Math.*, 42:577–685, 1989.

[8] A. Papoulias. Curve Segmentations Using Weak Continuity Constraints. M.Sc. thesis, Univ. of Edinburgh, 1985.

[9] A. Restrepo and A. C. Bovik. Locally Monotonic Regression. *IEEE Trans. Signal Processing*, 41(9):2796–2810, Sep. 1993.

[10] N. Sidiropoulos. Fast Digital Locally Monotonic Regression. Submitted for publication, IEEE Trans. Signal Processing. Sum mary to appear in Proc. 1996 IEEE Int. Symp. on Circuits and Systems, May 12-15, Atlanta, GA.

[11] N. Sidiropoulos. The Viterbi Optimal Runlength-Constrained Approximation Nonlinear Filter. *IEEE Trans. Signal Processing*, 44(3), March 1996.

[12] N. Sidiropoulos, J. Baras, and C. Berenstein. Structurally Robust Weak Continuity. Submitted for publication, IEEE Trans. Signal Processing.

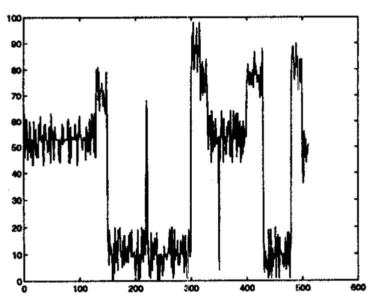

Figure 1. Input sequence, $\{y(n)\}_{n=0}^{511}$

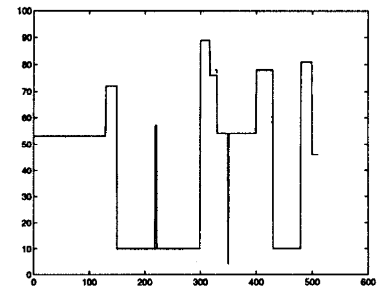

Figure 2. Output of digital WC, $\lambda_{WC}^2 = 25$

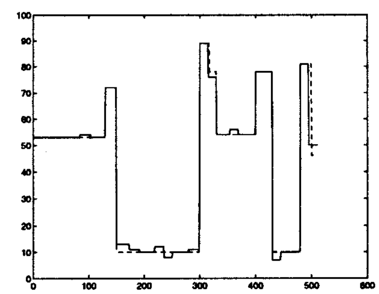

Figure 3. Output of VORCA, $M = 15$

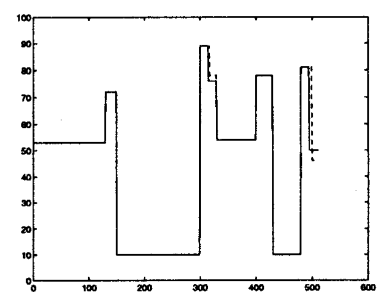

Figure 4. Output of RC-WC, $M = 15$, $\lambda_{WC}^2 = 25$