For Piecewise Smooth Signals, l^1 Method Is the Best Among l^p: An Interval-Based Justification of an Empirical Fact

Vladik Kreinovich[1] and Arnold Neumaier[2]
[1]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
email vladik@utep.edu
[2]Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien, Austria
email Arnold.Neumaier@univie.ac.at

Abstract

Traditional engineering techniques use the Last Squares method (i.e., in mathematical terms, the l^2-norm) to process data. It is known that in many practical situations, l^p-methods with $p \neq 2$ lead to better results. In different practical situations, different values of p are optimal. It is known that in several situations when we need to reconstruct a piecewise smooth signal, the empirically optimal value of p is close to 1. In this paper, we provide a new interval-based theoretical explanation for this empirical fact.

Keywords: piecewise smooth, l^1 method, interval uncertainty

1 Formulation of the Problem

l^2-methods: brief reminder. Traditional engineering techniques use the Last Squares Method LSM (i.e., in mathematical terms, the l^2-norm) to process data. For example, if we know that measured values $b_1, \ldots, b_m$ are related to the unknowns $x_1, \ldots, x_n$ by the known dependence $\sum\limits_{i=1}^{n} A_{ij} x_j \approx b_i$, and we know the accuracy σ_i of each measurement, then the LSM means that we find the values x_j for which the function

$$V = \sum_{i=1}^{m} \left(\frac{1}{\sigma_i} \cdot \sum_{i=1}^{n} A_{ij} x_j - b_i \right)^2$$

takes the smallest possible value.

By the Gauss-Markov Theorem [20], this method is provably optimal (the best linear unbiased estimator) under the assumption that the measurement errors $\Delta b_i \stackrel{\text{def}}{=} \sum\limits_{i-1}^{n} A_{ij} x_j - b_i$ are uncorrelated with 0 mean and standard deviation σ_i. In addition, if the Δb_i are independent and normally distributed, the maximum likelihood method [16, 18] $\rho(\Delta b_1, \ldots, \Delta b_n) \to \max$ takes the form

$$\rho(\Delta b_1, \ldots, \Delta b_n) = \rho_1(\Delta b_1) \cdot \ldots \cdot \rho_n(\Delta b_n)$$

and $\rho_i(\Delta b_i) = \dfrac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\dfrac{\Delta b_i^2}{2\sigma_i^2}\right)$.

Since the logarithm is a strictly increasing function, and the logarithm of a product $\rho_1 \cdot \ldots \cdot \rho_n$ is equal to the sum of the logarithms, maximizing the maximal likelihood is equivalent to minimizing the sum of negative logarithms $-\log(\rho_i)$ of ρ_i, i.e., minimizing the sum

$$\psi\left(\frac{\Delta b_1}{\sigma_1}\right) + \ldots + \psi\left(\frac{\Delta b_n}{\sigma_n}\right) \quad (1)$$

with $\psi(x) = x^2$ – i.e., to the Least Squares Method.

Similarly, if we know that the next value x_{i+1} is close to the previous value x_i of the desired signal, and that the average difference between $x_{i+1} - x_i$ is about σ_i, then we can use LSM to find the values x_i which minimize the sum $V = \sum\limits_{i=1}^{n-1} \left(\dfrac{x_{i+1} - x_i}{\sigma_i} \right)^2$.

M-methods: brief reminder. In many practical situations, different measurement errors are independent, but the distribution may be different from normal; see, e.g., [13, 14, 15]. In this case, the maximum likelihood method is still equivalent to minimizing the sum (1), but with a different function $\psi(x) = -\log(\rho(x))$.

In many other practical situations, we know that the distribution is not normal, but we do not know its exact shape. In such situations of *robust statistics*, we can still use a similar method, with an appropriately selected function $\psi(x)$. Such methods are called *M-methods*; see, e.g., [8, 16, 18].

In such situations, if we know that the next value x_{i+1} is close to the previous value x_i of the desired signal, and that the average difference between $x_{i+1} - x_i$

is about σ_i, then we can use LSM to find the values x_i which minimize the sum $V = \sum_{i=1}^{n-1} \psi\left(\frac{x_{i+1} - x_i}{\sigma_i}\right)$.

l^p-methods: brief reminder. Among different M-methods, empirically, l^p-methods – with $\psi(x) = |x|^p$ for some $p \geq 1$ – turn out to be the best for several practical applications; see, e.g., [4]. In this case, we select a signal (= tuple) x_i for which the value $V \stackrel{\text{def}}{=} \sum_{i=1}^{n-1} \left|\frac{x_i - x_{i+1}}{\sigma_i}\right|^p$ is the smallest possible.

These methods have been successfully used to solve inverse problems in geophysics; see, e.g., [6, 17].

In [11], the empirical success of l^p-methods was theoretically explained: it was shown that l^p-methods are the only scale-invariant ones, and that they are the only methods optimal with respect to all reasonable scale-invariant optimality criteria. It is therefore reasonable to use l^p-methods for processing data.

l^p-methods: how to select p. The above-mentioned justification explains that with respect to each optimality criterion, *one* of the l^p-methods is optimal – but does not explain which one. It is known that in different practical situations, different values of p lead to the best signal reconstruction.

For example, in the situation when the errors are normally distributed, $p = 2$ is the best value; for other situations, we may get $p = 1$ or $p \in (1, 2)$.

In each situation, we must therefore empirically select p – e.g., by comparing the result of data processing with the actual (measured) values of the reconstructed quantity.

Empirical fact. In several situations, we know that the reconstructed signal is piecewise smooth. For example, in geophysics, the Earth consists of several layers with abrupt transition between layers; in image processing, an image often consists of several zones with an abrupt boundary between the zones, etc.

It turns out that in many such situations, the empirically optimal value of p is close to 1; see, e.g., [6] for the inverse problem in geophysics, and [1, 7, 12, 19] for image reconstruction.

How this fact is explained now (see, e.g., [1]). In the continuous approximation, the l^p-criterion leads to the minimization of $\int |\dot{x}|^p \, dt$ (in the 1D case; multi-D case is handled similarly). For a transition of magnitude C and width ε, the derivative $\dot{x}$ is $\approx C/\varepsilon$, so the contribution of the transition zone to the integral is of order $\varepsilon/\varepsilon^p = \varepsilon^{-(p-1)}$. For $p > 1$, when $\varepsilon \to 0$, this contribution tends to ∞. Thus, for $p > 1$, the minimum is never attained at the discontinuous transition ("jump") $\varepsilon = 0$, but always at a smoother transition $\varepsilon > 0$.

For $p = 1$, the contribution is finite, so jumps are not automatically excluded – and indeed, they may be correctly reconstructed.

Limitations of this explanation. There are two limitations to this explanation:

- first, it explains why l^p-methods for $p > 1$ do not reconstruct a jump, but it does not explain why l^1 methods reconstruct the jump correctly;
- second, it strongly relies on the continuous case and does not fully explain why a similar phenomenon occurs for real-life (discretized) computations.

What we do in this paper. In this paper, we provide a new interval-based theoretical explanation for the above empirical fact, an explanation that is directly applicable to real-life (discretized) computations.

2 Analysis of the Problem and the Main Results

For simplicity, we will consider 1-D signals $x(t)$. In the interval setting, for several moments of time $t_1 < \ldots < t_n$ (usually, equidistant $t_i = t_1 + (i-1)\Delta t$), we know the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ that contain the actual (unknown) values $x_i = x(t_i)$. Based on this interval information, we would like to select the values $x_i \in \mathbf{x}_i$. According to the l^p-criterion, among all the tuples $(x_1, \ldots, x_n)$ for which $x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n$, we select the one for which the value $V = \sum_{i=1}^{n-1} \left|\frac{x_i - x_{i+1}}{\sigma_i}\right|^p$ is the smallest possible.

To select p, we will consider the case of a "transition zone", i.e., the case when for some values $l < u$, we know two things:

- that the value x_{l-1} right before the zone cannot be equal to the value x_{u+1} right after the zone – i.e., that $\mathbf{x}_{l-1} \cap \mathbf{x}_{u+1} = \emptyset$; and
- that we have (practically) no information about the values of x_i inside the zone – i.e., at least that for all i from l to u, the interval $\mathbf{x}_i$ contains both $\mathbf{x}_{l-1}$ and $\mathbf{x}_{u+1}$.

In this case, the above criterion interpolates the values x_i inside the zone. If we assumed that the signal is smooth, then, no matter how steep the transition, we would have had a smooth interpolation. However, since we consider the situations when the signal is only piecewise smooth, we would rather prefer to have a signal which "jumps" discontinuously from one value to another.

In this section, we will show that for $p = 1$, we will indeed get such a jump, while for $p > 1$, we have a smooth transition instead. Let us describe this result in precise terms.

Definition 1 *By an* l^p-problem, *we mean the following problem:*

GIVEN: n intervals $\mathbf{x}_i = [\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$, n real numbers $\sigma_1, \ldots, \sigma_n$, and a real number $p \geq 1$;

AMONG: tuples $x_1 \ldots, x_n$ such that $x_i \in [\underline{x}_i, \overline{x}_i]$ for every i;

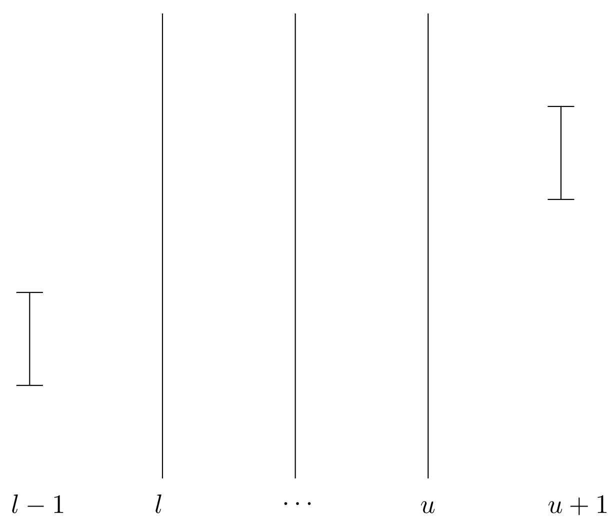
FIND: the tuple for which the value $V = \sum_{i=1}^{n-1} \left| \frac{x_i - x_{i+1}}{\sigma_i} \right|^p$ is the smallest possible.

Definition 2 *An l^p-problem is called* degenerate *if all the values σ_i are different.*

Comment. Almost all combinations $\sigma_1, \ldots, \sigma_n$ are degenerate.

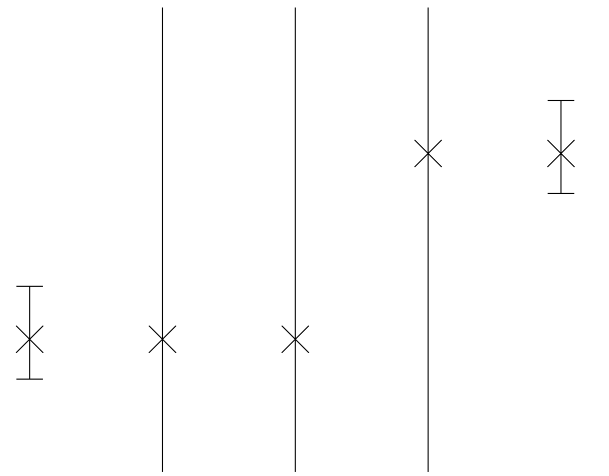
Definition 3 *Let $l < u$ be integers. We say that an l^p-problem contains a* transition zone *between l and u if the following two conditions hold:*

- $\mathbf{x}_{l-1} \cap \mathbf{x}_{u+1} = \emptyset$; *and*
- *for all i from l to u, we have $\mathbf{x}_i \supseteq \mathbf{x}_{l-1}$ and $\mathbf{x}_i \supseteq \mathbf{x}_{u+1}$.*



Proposition 1 *For $p = 1$, for each solution x_i to a non-degenerate l^p-problem, in each transition zone, we have $x_{l-1} = x_l = \ldots = x_t$ and $x_{t+1} = \ldots = x_u = x_{u+1}$ for some t.*

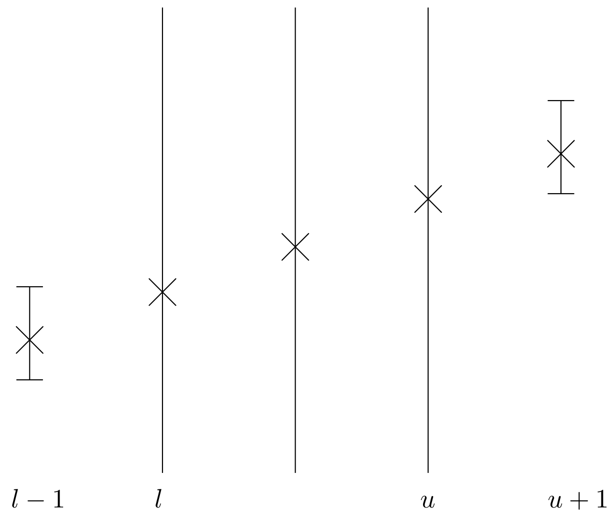
In other words, for $p = 1$, in each transition zone, we have a "jump" from the value x_{l-1} before the transition zone to the value x_{u+1} after the transition zone.



Comment. In the degenerate case, when different values σ_i are equal, the jump is still an optimal solution, but we may also get other solutions, with a smooth transition from x_{l-1} to x_{u+1}. For example, if all the values σ_i are the same, them, as one can easily see, the minimized criterion is proportional to the sum $\sum_{i=1}^{n-1} |x_i - x_{i+1}|$; so, for each solution that monotonically changes from x_{l-1} to x_{u-1}, the corresponding part $\sum_{i=l-1}^{u} |x_i - x_{i+1}|$ of the sum is equal to $|x_{l-1} - x_{u+1}|$. Thus, the value of the minimized criterion is the same for the jump solution and for a different solution in which x_i is the same outside $[l-1, u+1]$ – but strictly monotonically changes between $l-1$ and $u+1$.

Proposition 2 *For $p > 1$, for each solution x_i to an l^p-problem, in each transition zone, we have a strictly monotonic sequence $x_{l-1} < x_l < \ldots < x_u < x_{u+1}$ or $x_{l-1} > x_l > \ldots > x_u > x_{u+1}$.*

Proposition 3 *For $p > 1$, in the limit when all the values σ_i tend to the same value σ, each solution x_i to an l^p-problem, in each transition zone, is linear, i.e., has the form $x_i = a + bi$ for some numbers a and b.*



These results explain why $p \approx 1$ is indeed empirically best for processing piecewise smooth signals: only for $p = 1$, l^p-interpolation leads to a piecewise smooth signal.

Comment. The fact that l^1-methods are the best among l^p-methods does not mean that they are always the best possible interpolation techniques. For example, the above results show that with an l^1-method we always get a jump, both:

- for the steep transition from $\mathbf{x}_{l-1}$ to $\mathbf{x}_{u+1}$, where such a jump is desirable, and
- for a smoother transition from $\mathbf{x}_{l-1}$ to $\mathbf{x}_{u+1}$, where, from the physical viewpoint, we may want to prefer a smooth interpolation.

In other words,

- for small differences $x_i - x_{i+1}$, we would like to have smooth transitions, while
- for large differences $x_i - x_{i+1}$, we would like to have a jump.

Since a jump is reconstructed when $\psi(x) = |x|$ and a smooth transition, when, e.g., $\psi(x) = |x|^2$, a natural idea is to use a *Huber function* $\psi(x)$ which is equal to $|x|^2$ when $|x|$ is below a certain threshold x_0, and which is linear $\psi(x) = C \cdot |x|$ for $|x| > x_0$; from the requirement that the function $\psi(x)$ be continuous, we conclude that $C = |x_0^2| = C \cdot |x_0|$, i.e., that $C = x_0$. Such technique indeed leads to a better reconstruction of piecewise smooth signals; see, e.g., [1] and references therein. Various related choices for $\psi(x)$ have been explored in the context of computer tomography by Kaufman and Neumaier [9, 10].

Huber's function $\psi(x)$, in its turn, has its own limitations; it is worth mentioning that in general, the problem of optimally reconstructing piecewise smooth 2-D signals is NP-hard; see, e.g., [2, 3, 5].

3 Proofs

1°. First, we observe that the solution to an l^p-problem minimizes a continuous function V on a bounded closed set (box) $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$. Thus, this minimum is always attained, i.e., a solution always exists.

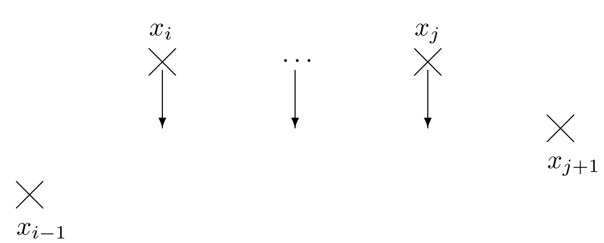
2°. Let us prove that for every p, the solution x_i to the l^p-problem is (non-strictly) monotonic in each transition zone, i.e., that $x_{l-1} \le x_l \le \ldots \le x_u \le x_{u+1}$ or $x_{l-1} \ge x_l \ge \ldots \ge x_u \ge x_{u+1}$.

Let us prove this result by reduction to a contradiction. Namely, let us assume that the solution is attained on some non-monotonic sequence. The fact that x_i is not monotonic on the transition zone means that not all inequalities between the neighboring values are of the same sign, i.e., that we have $x_{i-1} < x_i$ and $x_j > x_{j+1}$ for some indices i and j from this zone. Among such pairs (i,j), let us select a one with the smallest distance $|i-j|$ between i and j.

Without losing generality, we can assume that in this selected pair, $i < j$.

For the selected pair, for indices k between x_i and x_j, we cannot have $x_k < x_{k+1}$ or $x_k > x_{k+1}$ – otherwise we would get a pair with an even smaller difference $|i-j|$. Thus, for all intermediate indices k, we get $x_k = x_{k+1}$. Since $x_i = x_{i+1} = \ldots = x_j$, we thus have $x_i = x_j$. So, we have $x_{i-1} < x_i = \ldots = x_j > x_{j+1}$. Let $\varepsilon = \min(x_i - x_{i-1}, x_j - x_{j+1})$. Let us now keep all the x-values outside (i,j) intact and replace $x_i = \ldots = x_j$ with the values $x_i - \varepsilon = \ldots = x_j - \varepsilon$. The resulting value $x_i - \varepsilon$ is equal to either $x_{i-1} \in \mathbf{x}_{i-1}$ or to $x_{j+1} \in \mathbf{x}_{j+1}$. By the definition of a transition zone, all intermediate intervals $\mathbf{x}_k$ contain both $\mathbf{x}_{i-1}$ and $\mathbf{x}_{j+1}$; hence, the new value of x_k is within the corresponding interval $\mathbf{x}_k$.

By making this change, we decrease the differences $|x_i - x_{i-1}|$ and $|x_{j+1} - x_j|$ and leave all other difference intact – and hence, we decrease the value of the minimized objective function V.



Since the objective function V attains its minimum at the original tuple x_i, the possibility to minimize even further is a contradiction. This contradiction proves that the solution is monotonic in each transition zone.

3°. According to Part 2 of this proof, for the solution, we have $x_{i-1} \le x_i \le \ldots \le x_u \le x_{u+1}$ or $x_{i-1} \ge x_i \ge \ldots \ge x_u \ge x_{u+1}$. To complete the proof of Proposition 1, it is now sufficient to prove that for $p = 1$ and for $k = l, \ldots, u$, we cannot have any strictly intermediate values $x_k \in (x_{l-1}, x_{u+1})$ (or $x_k \in (x_{u+1}, x_{l-1})$).

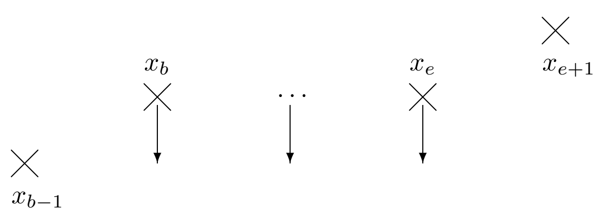
Let us prove this by reduction to a contradiction. Let us assume that an intermediate value x_k does exist. In principle, we may have values equal to x_k. Due to monotonicity, these values form an interval within $[l, u]$. Let x_b be the first value equal to x_k, and let x_e be the last value equal to x_k. Then, we have $\ldots \le x_{b-1} < x_b = \ldots = x_e < x_{e+1} \le \ldots$

Let us now choose a value $\varepsilon \in [x_{b-1} - x_b, x_{e+1} - x_e]$, keep all the x-values from outside $[b, e]$ intact, and replace all the x-values from $[b, e]$ with $x_b + \varepsilon = \ldots = x_e + \varepsilon$. Similarly to Part 2 of this proof, we can show that for every ε, we still have $x_b + \varepsilon \in \mathbf{x}_b$, ..., $x_e + \varepsilon \in \mathbf{x}_e$.

After this replacement, we change only two differences $|x_{i+1} - x_i|$:

- the difference $|x_b - x_{b-1}| = x_b - x_{b-1}$ is replaced with $x_b - x_{b-1} + \varepsilon$, and
- the difference $|x_{e+1} - x_e| = x_{e+1} - x_e$ is replaced with $x_{e+1} - x_e - \varepsilon$.

Thus, after this replacement, the original value V of the minimized objective function is replaces with $V + \Delta V$, where $\Delta V \stackrel{\text{def}}{=} \varepsilon \cdot \left(\frac{1}{\sigma_{b-1}} - \frac{1}{\sigma_e} \right)$.



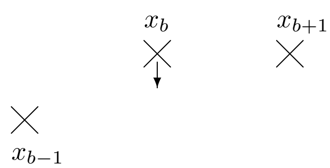
Since the problem is non-degenerate, i.e., all the values σ_i are different, the coefficient at ε in ΔV is non-zero. If this coefficient is positive, we can take negative ε and decrease V; it is is negative, we can decrease V by taking $\varepsilon > 0$. In both cases, we get a contradiction with the fact that the original tuple x_i minimizes V. This contradiction proves that intermediate values are impossible. Proposition 1 is proven.

4°. Let us now prove that for $p > 1$, the solution is strictly monotonic.

We will prove this by reduction to a contradiction. Let us assume that it is not strictly monotonic. We know that the solution is monotonic (Part 2 of the proof). Since it is monotonic, the only way for the solution to be not strictly monotonic is to have $x_i = x_{i+1}$ for some index i. We may have several indices with an x-value equal to this x_i; let b be the first such index, and let e the last such index. Then, $x_b = x_{b+1} = \ldots = x_e$.

Since the intervals $\mathbf{x}_{l-1}$ and $\mathbf{x}_{u+1}$ have no common points, we cannot have $x_{l-1} = x_{u+1}$. Thus, either $b \neq l-1$ or $e \neq u+1$. Without losing generality, we can assume that $b \neq l-1$. Also, without losing generality, we can assume that the solution x_i is increasing. Thus, we have $x_{b-1} < x_b = x_{b+1}$.

Let us now pick a small value $\varepsilon > 0$ and replace x_b with $x_b - \varepsilon$ – while leaving all other x-valued intact.



This replacement changed the original value V of the minimized function with a new value $V + \Delta V$, where

$$\Delta V = \frac{(x_b - x_{b-1} - \varepsilon)^p}{\sigma_{b-1}^p} + \frac{\varepsilon^p}{\sigma_b^p} - \frac{(x_b - x_{b-1})^p}{\sigma_{b-1}^p}.$$

By applying the first term of Taylor expansion to the first ratio in the expression for ΔV, we conclude that

$$\Delta V = -\frac{p \cdot (x_b - x_{b-1})^{p-1}}{\sigma_{b-1}^p} \cdot \varepsilon + O(\varepsilon^2) + \frac{\varepsilon^p}{\sigma_b^p}.$$

We consider the case when $p > 1$. In this case, for sufficiently small ε, the first term dominates, so the difference ΔV is negative – which means that we can further decrease V.

This possibility contradicts to the fact that the tuple x_i minimizes V. Thus, the solution is indeed strictly monotonic. Proposition 2 is proven.

5°. Let us now prove Proposition 3.

By definition of the transition zone, for each index i from this zone, we have $\mathbf{x}_{l-1} \subseteq \mathbf{x}_i$, hence $x_{l-1} \in \mathbf{x}_{l-1} \subseteq \mathbf{x}_i$ and $x_{l-1} \in [\underline{x}_i, \overline{x}_i]$ – thence $\underline{x}_i \le x_{l-1}$. Similarly, from $\mathbf{x}_{u+1} \subseteq \mathbf{x}_i$, we conclude that $x_{u+1} \le \overline{x}_i$.

Due to strict monotonicity (Part 4 of this proof), we have $x_{l-1} < x_i < x_{u+1}$. Thus, $\underline{x}_i \le x_{l-1} < x_i$ and $\underline{x}_i < x_i$ and similarly, $x_i < \overline{x}_i$.

Since the value x_i is strictly inside the interval $\mathbf{x}_i$, the derivative of the minimized function V is equal to 0. Differentiating V relative to x_i (and taking monotonicity into account), we conclude that

$$p \cdot (x_i - x_{i-1})^{p-1} \sigma_{i-1}^p - p \cdot (x_{i+1} - x_i)^{p-1} \sigma_i^p = 0.$$

When $\sigma_i \to \sigma$, we get $x_i - x_{i-1} = x_{i+1} - x_i$. So, the difference $x_i - x_{i-1}$ is indeed the same for all i within the transition zone. Thus, we get the desired linear dependence of x_i on i. The proposition is proven.

Acknowledgments

This work was supported in part by NSF grants EAR-0225670 and DMS-0532645 and by Texas Department of Transportation grant No. 0-5453. The authors are thankful to the participants of the the International Conference on Finite Element Methods in Engineering and Science FEMTEC'2006 (El Paso, Texas, December 11–15, 2006) for valuable discussions.

References

[1] U. M. Ascher, E. Haber, and H. Huang, "On effective methods for implicit piecewise smooth surface recovery", *SIAM Journal on Scientific Computing*, 1006, Vol. 28, No. 1, pp. 339–358.

[2] M. G. Averill, K. C. Miller, G. R. Keller, V. Kreinovich, R. Araiza, and S. A. Starks, "Using Expert Knowledge in Solving the Seismic Inverse Problem", *Proceedings of the 24th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2005*, Ann Arbor, Michigan, June 22–25, 2005, pp. 310–314.

[3] M. G. Averill, K. C. Miller, G. R. Keller, V. Kreinovich, R. Araiza, and S. A. Starks, "Using Expert Knowledge in Solving the Seismic Inverse Problem", *International Journal of Approximate Reasoning* (to appear).

[4] P. J. Bickel and E. L. Lehmann, "Descriptive Statistics for Nonparametric Models. 1. Introduction", *Ann. Statist.*, 1975, Vol. 3, pp. 1045–1069.

[5] S. D. Cabrera, K. Iyer, G. Xiang, and V. Kreinovich, "On Inverse Halftoning: Computational Complexity and Interval Computations", *Proceedings of the 39th Conference on Information Sciences and Systems CISS'2005*, John Hopkins University, March 16–18, 2005, Paper 164.

[6] D. I. Doser, K. D. Crain, M. R. Baker, V. Kreinovich, and M. C. Gerstenberger "Estimating uncertainties for geophysical tomography", *Reliable Computing*, 1998, Vol. 4, No. 3, pp. 241–268.

[7] H. Feng, D. A. Castañon, and W. C. Karl, "Underground Imaging Based on Edge-Preserving Regularization", *Porc. of the IEEE International Conference on Information Intelligence and Systems*, October 31 – November 3, 1999, Bethesda, Maryland, IEEE Computer Society, Los Alamistos, California, pp. 460–464.

[8] P. J. Huber, *Robust Statistics*, Wiley, N.Y., 1981.

[9] L. C. Kaufman and A. Neumaier, "PET regularization by envelope guided conjugate gradients", *IEEE Trans. Medical Imag.*, 1996, Vol. 15, pp. 385–389.

[10] L. C. Kaufman and A. Neumaier, "Regularization of ill-posed problems by envelope guided conjugate gradients", *J. Comput. Graph. Stat.*, 1997, Vol. 6, pp. 451–463.

[11] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[12] M. Nikolova, "Minimizations of cost-functions involving nonsmooth data fidelity terms", *SIAM J. Numer. Anal.*, 2002, Vol. 40, pp. 965-994.

[13] P. V. Novitskii and I. A. Zograph, *Estimating the measurement errors*, Energoatomizdat, Leningrad, 1991 (in Russian).

[14] A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

[15] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.

[16] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[17] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, 1987.

[18] H. M. Wadsworth, Jr. (editor), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., N.Y., 1990.

[19] J. Zhou, H. Shu, T. Xia, and L. Luo, "PET Image Reconstruction Using Mumford-Shah Regularization Coupled with L^1 Data Fitting", *Proceedings of the 27th Annual IEEE Conference on Engineering in Medicine and Biology*, China, September 1–4, 2005, pp. 1905–1908.

[20] G. Zyskind and F. B. Martin, "On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure", *SIAM J. Appl. Math,*, 1969, Vol. 17, pp. 1190–1202.