

A STOCHASTIC DIFFUSION APPROACH TO SIGNAL DENOISING

Hamid Krim and Yufang Bao[†]

ECE Dept., NCSU, [†]also BUPT, China, P.R.
Raleigh, NC 27695-7914,
ahk@eos.ncsu.edu, yfbao@eos.ncsu.edu

ABSTRACT

We present a stochastic formulation of a linear diffusion equation (or heat equation), and in light of the potential applications ranging from signal denoising to image enhancement/segmentation of its nonlinear extensions, we propose a more general nonlinear stochastic diffusion. The constructed stochastic framework, in contrast to traditional deterministic approaches, unveils the sources of existing limitations and allows us to further significantly improve the performance by addressing the key problem. Substantiating examples are provided.

1. INTRODUCTION

Research interest in scale-based analysis has significantly grown over the last decade, and scale as an entity has played an increasingly important role in signal and image analysis. Since the ground breaking paper of Witkin [1] who pointed out the equivalence between a heat equation-based evolution of a process and its smoothing with a Gaussian kernel and proposed a linear scale space analysis, several developments have taken place. A systematic multiscale analysis framework has been independently proposed by Mallat[2] using wavelet bases. These wavelet functions with their ability to focus energy on local important features intrinsic to a signal turned out to be well adapted to signal enhancement and denoising [3]. The linear scale space framework originally intended as a continuous scale framework, was hampered by its uniform filtering of signal and noise,

This was in part supported by an AFOSR grant F49620-98-1-0190 and by ONR-MURI grant WUHT-72298-S2 and by NCSU school of Engineering.

and this limitation was first addressed by Perona and Malik (P-M) [4]. Their approach relied on the sharp features in a signal (e.g, singularities) to help improve the performance of signal/image denoising (image segmentation as well). Specifically they proposed a nonlinear Partial Differential Equation (PDE) to operate on a noisy signal and selectively smooth (or diffuse) the regions devoid of large gradients (i.e., where singularities as a step jump or an edge in an image are absent).¹ The novelty of this approach together with the promising results achieved triggered a tremendous research activity in computer vision and applied mathematics (see [5] for a good review of the literature). The performance or the expected performance of all these methods in different noise environments was not well understood, on one part because the inherent nonlinearities introduced technical difficulties, on the other the randomness was never explicitly addressed. In [6], Pollak *et. al.* proposed a new approach which resulted in a remarkable robustness to a wide class of noises, in spite of which the methodology remained, as in all previous techniques, fundamentally deterministic.

Our goal in this paper is to revisit the denoising problem based on nonlinear scale space diffusion and account for the inherent randomness of the noise by developing a stochastic analysis approach readily justified by stochastic differential equations[9]. This will, as will be seen, unfold the advantages of the framework, and unveil its limitations.

In the balance of this paper, we first formulate the problem and provide the necessary background for the subsequent development. In Section 3, we provide a stochas-

¹The gradient size is relative here as the additive white noise, for example, tends to also have large gradients.

tic interpretation of the linear heat PDE and describe its discrete implementation. In Section 4, we provide a stochastic viewpoint of Perona-Malik’s equation, and we in turn use the insight provided by this stochastic interpretation to propose a new nonlinear equation which we term *Nonlinear Stochastic diffusion*. We finally provide substantiating examples in Section 5.

2. BACKGROUND

As noted above, using the fact that a Gaussian function is the Green’s function of a Heat equation[7], Witkin[1] proposed to use the latter to evolve a signal to achieve similar linear filtering of a Gaussian of variance t . Smoothing a noisy signal $f(x)$, where, $x \in \mathbb{R}$ is thus written as

$$\frac{\partial U(t, x)}{\partial t} = \Delta U(t, x), \quad (1)$$

where $U(t, x)$ denotes the signal at scale $t \in \mathbb{R}^+$, $U(0, x) = f(x)$, and Δ is the Laplacian operator (i.e., $\frac{\partial^2}{\partial x^2}$).² $U(\cdot, \cdot)$ at the finest scale is assumed to be comprised of a signal of interest and white Gaussian noise of some variance σ^2 . In this analysis we will focus on the class of piece-wise constant signals (i.e. staircase step functions are of importance in many applications, and the discontinuities represent edges in 2-D signals/images).

Using the linear heat equation as their paradigm, Perona and Malik’s nonlinear filter behavior is as a Gaussian lowpass filter in regions of low gradients to eliminate any additive noise, and as an almost all-pass filter in regions of high gradients to preserve the sharp features. The corresponding evolution equation is then of the form

$$U(t, x) = \text{div} (g(|\nabla(U(t, x))|) \nabla U(t, x)), \quad (2)$$

where “div” represents the divergence operator, ∇ is the gradient, and $g(\cdot)$ the function which modulates the diffusion according to the above paradigm (i.e., positive and monotonically decreasing with $g(0) = 1$). One possible choice is $g(v) = e^{-\frac{v^2}{K^2}}$ where K determines the rate of decay and thus the extent of smoothing of $U(t, x)$ for a given gradient size. For reasons of space some mathematical technicalities are discussed elsewhere [8].

²The variable t in this context represents scale instead of time.

In light of its direct impact on our proposed framework, the P-M equation will be our primary focus, noting however, that numerous subsequent and important contributions have since appeared in the literature as reported in [5].

3. STOCHASTIC DIFFUSION

The PDE in Eq. 1 models well the diffusion of heat in a homogeneous medium, which fundamentally stems from the motion of particles. The inherent randomness of this motion is well-described by a Brownian motion B_t , and an individual outcome $\omega \in \Omega$ in the prevailing sample space, is associated to a particle. The process B_t can then be interpreted as the distance traveled by particle ω at time t . It is well known that such a transition density for a Brownian motion, for instance, is a Gaussian PDF $p(t, x, y) = \frac{1}{(2\pi t)^{1/2}} e^{-\frac{(y-x)^2}{2t}} \quad \forall x, y \in \mathbb{R}, t > 0$. In light of the above, a stochastic interpretation of a solution to the heat equation, if it exists, and subject to some differentiability conditions, can be given [9] as

$$U(t, x) = E_x\{f(B_t)\} = \int_{\mathbb{R}} p(t, x, y) f(y) dy, \quad (3)$$

where the expectation operator $E(\cdot)$ is taken with respect to x , and $p(t, x, y)$ being the Gaussian PDF.

3.1. A Discrete Formulation

As previously noted, our chief interest here is to achieve a stochastic understanding of scale-space analysis, and hence the importance of reexpressing the above framework in the discrete scale space. Recall that the symmetrical random walk is well known to converge to a Brownian motion as $\tau \rightarrow 0$ and $\delta \rightarrow 0$, with τ, δ respectively denoting scale and distance steps. Hence, a particle following such a trajectory will move on a 1-D line to the right or the left with equal probability of 1/2. Upon discretizing the spatial variable $x_n = x + n\delta$ and the scale $t_n = n\tau$, denoting the probability of a particle to be at position x after n time steps, having departed from x_0 , by $p_n(x_0, x)$, we obtain a standard result from 3, namely the probability of a particle being at x at time $n + 1$ as $\tau \rightarrow 0$ and $\delta \rightarrow 0$,

Proposition 1. *The following discrete equation,*

$$p_{n+1}(x_0, x) = 1/2 p_n(x_0, x - \delta)$$

$$+ 1/2p_n(x_0, x + \delta), \quad (4)$$

converges to

$$\frac{\partial p_t(x_0, x)}{\partial t} = \frac{\partial^2 p_t(x_0, x)}{\partial x^2}, \quad (5)$$

Proof: subtracting $p_n(x_0, x)$ from both sides of Eq. 4, we obtain

$$\frac{p_{n+1}(x_0, x) - p_n(x_0, x)}{\tau} = \frac{1}{2\tau} [p_n(x_0, x - \delta) - 2p_n(x_0, x) + p_n(x_0, x + \delta)], \quad (6)$$

we conclude the proof by letting $\tau = \delta^2$ and $\delta \rightarrow 0$ ■

3.1.1. Solution via Discrete Expectation

The above results are well known to probabilists and thus not new. The solution to Eq. 5 is a Gaussian transition density function and it characterizes the evolution of a Brownian motion with t and starting at x_0 . Knowing that the limiting process of a random walk is a Brownian motion, we use this convergence to compute the solution at any scale using the continuous solution given in Eq. 3. With the foregoing discretization of x and t we denote by $U(n, x)$ the value of the solution at time step $n\tau$ and state x we can write at step τ ,

$$U(1, x) = \frac{1}{2}f(x - \delta) + \frac{1}{2}f(x + \delta). \quad (7)$$

More generally, we proceed to write the solution to the linear heat equation as a discrete expectation,

$$U(n + 1, x) = \frac{1}{2}U(n, x - \delta) + \frac{1}{2}U(n, x + \delta). \quad (8)$$

Due to the underlying random walker moving to the left and the right with probability $1/2$, it is clear that the the linear evolution will indiscriminately smooth away sharp features along with the noise.

Recall, the P-M equation performs a gradient-based selective smoothing, and using the above discussion as a basis we provide its stochastic formulation in the next section.

4. NONLINEAR STOCHASTIC DIFFUSION

4.1. Stochastic Perona-Malik Diffusion

The nonlinear Perona-Malik equation in Eq. 2 may be stochastically interpreted as well, and this is particularly simplified upon discretizing the scale and space

variables to lead to the following:

$$U((n + 1), x) = p_{1,n+1}(x)U(n, x + \delta) + (1 - p_{1,n+1}(x) - p_{-1,n+1}(x))U(n, x) + p_{-1,n+1}U(n, x - \delta), \quad (9)$$

where

$$p_{1,n+1}(x) = 1/2g(|U(n, x + \delta) - U(n, x)|) = p_{n+1}(x, x + \delta),$$

$$p_{-1,n+1}(x) = 1/2g(|U(n, x) - U(n, x - \delta)|),$$

and $p_{n+1}(x, x + \delta) = P(\xi_{n+1} = x + \delta | \xi_n = x)$ is the transition probability of the underlying Markov chain. This equation is intuitively appealing in that the random walk or the diffusion ξ_n takes place according to the prevailing one sided gradient at position x . At time step $n + 1$, a right moving walk takes place with probability $p_{1,n+1}(x)$, a left moving walk takes place with probability $p_{-1,n+1}(x)$, and the particle remains in place with probability $p_{0,n+1}(x) = 1 - p_{1,n+1} - p_{-1,n+1}$. In contrast to the linear diffusion where the random walk invariably takes place probability $1/2$, the P-M random walk is driven by the intrinsic gradient determined the features we would like to preserve.

4.2. A Bidirectionally Driven Stochastic Diffusion

As we just noted, the P-M diffusion is driven by a one-sided gradient at any position x , which implies that no smoothing takes place in the presence of a relatively high gradient even if caused by noise. On the other hand, if the gradient on both sides of a position x were considered, and knowing that a high gradient due to noise at some given position, tends to be duplicated at the position immediately following, as illustrated on the trailing edge of the noisy signal in Fig. 1. An adapted procedure would recognize that and accordingly diffuse, while the P-M filter which relies on one sided gradient calls for a no transition state in the Markov chain, thus preserving the singularity as displayed. Our proposed technique, precisely addressing such difficulties, is based on the ratio of the two gradients to identify such scenarios and assign a significant probability to diffuse, to eliminate the noise spike. This technical difficulty of the P-M equation may further be compounded in that the transition policy in the Markov chain may eliminate true edges; since at a position x close to the leading edge, and if at position $x + \delta$, $U(\cdot, x)$ is close to

$U(\cdot, x + \delta)$, the probability of transition is finite, and hence smoothing the leading edge of the step signal proceeds. This scenario happens with probability zero when a two sided gradient-based transition probability is used in the policy.

Using Markov chains ξ_n , we can simply model these dynamics via transition probabilities which, as we mentioned, may be specified in terms of the ratio of the bidirectional gradient. We denote the transition probability by $p_{n+1}(x) = P\{\xi_{n+1} = x + \delta | \xi_n = x\}$ and defined as

$$p_{n+1}(x) = \frac{\mathcal{N}}{\mathcal{D} + \mathcal{N}}, \quad (10)$$

where $\mathcal{N} = |U(n, x - \delta) - U(n, x)|^2$ and $\mathcal{D} = |U(n, x + \delta) - U(n, x)|^2$. Using the above transition probability, our newly proposed filter is thus given by

$$U(n+1, x) = U(n, x - \delta) \times (1 - p_{n+1}(x)) + U(n, x + \delta) \times p_{n+1}(x). \quad (11)$$

We should note that in the cases where the two-side gradients are very small, and thus little significance can be attached to their ratio, the motion of the particle is based on a symmetrical random walk, i.e., with probability $1/2$ moving to the right or to the left.

We can in fact prove that in the limit, this Markov chain tends to a stochastic continuous Markov process which in turn leads to further generalizations which are currently under investigation[8].

5. EXPERIMENTAL RESULTS

To illustrate the performance of our nonlinear filter, we add white Gaussian noise to a staircase function whose discontinuities represent a major challenge to a linear filter. The SNR is less than $3dB$. A performance comparison between the newly proposed approach and the P-M filter is shown in Figure 1. The better performance evident from the figure is in fact consistent in a variety of different scenarios and different signals.

6. REFERENCES

[1] A. P. Witkin, "Scale space filtering," in *Proc. Int. Joint Conf. on Artificial Intelligence*, (Karlsruhe, Germany), pp. 1019–1023, 1983.

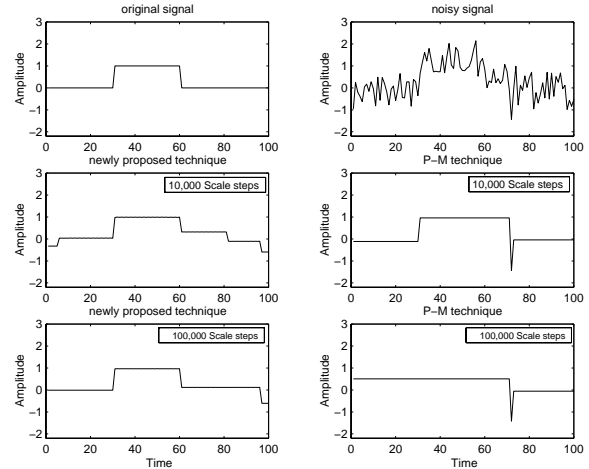


Figure 1: A denoising comparison of staircase signals.

[2] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PAMI-11, pp. 674–693, Jul. 1989.

[3] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Informat. Theory*, vol. IT-38, pp. 617–643, Mar. 1992.

[4] P. Perona and J. Malik, "A network for multiscale image segmentation," in *IEEE Int. Symp. on Circuits and Systems*, (Helsinki), pp. 2565–2568, June 1988.

[5] E. Book, *Geometry Driven Diffusion*. Dordrecht: Kluwer Academic Publishers, 1994.

[6] I. Pollak, A. Willsky, and H. Krim, "Image segmentation and edge enhancement with stabilized inverse diffusion equations," *To be submitted to IEEE Trans. on Image Processing*.

[7] V. Arnold, *Mathematical Methods in Classical Mechanics*. Applied Mathematics, New-York, Paris: Springer-Verlag, 2nd ed., 1989.

[8] H. Krim and Y. Bao, "Stochastic diffusion in signal analysis," *in preparation*.

[9] R. Durrett, *Stochastic Calculus: A Practical Introduction*. Florida: CRC Press, first ed.