TIKHONOV REGULARIZATION VERSUS SCALE SPACE: A NEW RESULT

Luc Florack, Remco Duits, Joris Bierkens

Eindhoven University of Technology
Department of Biomedical Engineering
PO Box 513
NL-5600 MB Eindhoven, The Netherlands
{L.M.J.Florack,R.Duits}@tue.nl, G.N.J.C.Bierkens@student.tue.nl

ABSTRACT

It is well-known that scale space theory and Tikhonov regularization are close-knit. In previous studies qualitative analogies and formal relations had already been found, but none of these established both an exact as well as operational connection. We establish such a connection for the case of a Gaussian scale space representation and a first order Tikhonov scheme. The free parameter of the latter turns out to be related to a particular attenuation of scales in a procedure whereby one "collapses" the scale space image along the scale axis via the Laplace transform. The result provides a physical interpretation of first order Tikhonov regularization and its associated control parameter.

1. INTRODUCTION

A great deal of attention has been paid in the image literature to circumventing, or solving altogether, the notorious ill-posedness problem of classical differentiation. Several solutions have been provided, and some have become standard tools for image processing. In this paper we will focus on Tikhonov regularization and Gaussian blurring.

Originally ill-posedness was attributed to irregularity of the image data. This led to the desire to smoothen the data by some sort of regularization procedure. In Tikhonov regularization the aim is to find a differentiable function which is, in some sense, close to the original data by means of functional minimization [1]. Although this way of handling ill-posedness hides a misconception (ill-posedness is an artifact of the derivative operators, not their operands!), it does yield sensible results when combined with a discrete difference scheme for classical differentiation when restricted to sufficiently low order. The price for this is a free parameter (or multiple parameters), controlling the degree of visual smoothness (not to be confused with order of differentiability, which does not depend on the value of the parameter).

A rigorous way to dispense with ill-posedness altogether was proposed in the early fifties of the previous century by the French mathematician Laurent Schwartz [2, 3]. Gaussian scale space theory [4, 5, 6, 7, 8] is a constrained instance of this theory of so-called "tempered distributions". If one identifies an image with a tempered distribution (a typically highly irregular "function" with a virtually arbitrary frequency spectrum), and constrains the admissible linear filters ("test functions of rapid decay" or "Schwartz functions") to the Gaussian family [9], one obtains scale space theory. Scale space theory inherits the intrinsic well-posedness of Schwartz' solution, because differentiation has been defined in terms of integration. It likewise introduces a free parameter, viz. scale or resolution.

Scale space differentiation (or distributional differentiation in general) is reminiscent of Tikhonov regularization, as both are rooted on integration. But an exact, operational connection between a Gaussian scale space representation and a Tikhonov regularized image remained hitherto unknown. Qualitative connections have been described in a previous study [10]. There it has been demonstrated that a suitably regularized image can be regarded *approximately* as a scale space image at a particular scale, i.e. as a filtered image, in which the filter is some approximation of a normalized Gaussian. It has also been demonstrated that there exists an exact, *formal* connection, in the sense that a Gaussian scale space image at a fixed scale can be interpreted as a Tikhonov regularized image obtained from a complicated regularization functional in the limiting case of infinitely many regularization terms, involving derivatives up to infinite order. Thus previous studies have mainly revealed analogies.

In this paper we establish an exact, operational relation between scale space theory and Tikhonov regularization. The essential difference with previous results is that (i) we consider only a single, quadratic, first order regularization term, (ii) multiple scales—in the sense of Gaussian scale space theory—are involved in the Tikhonov regularization of a raw image for any fixed value of the regularization parameter, and (iii) the relation between the two paradigms is of an exact and operationally well-defined rather than approximate or formal nature.

2. THEORY

2.1. A Summary of Previously Established Results

Let f_0 be a real-valued raw image. The Tikhonov functional of interest will be the following:

$$\mathcal{E}\left[f_\lambda\right] = \int \frac{1}{2} f_\lambda(x)^2 + \frac{1}{2}\lambda \|\nabla f_\lambda(x)\|^2 - f_\lambda(x) f_0(x)\, dx\,. \quad (1)$$

The bilinear term at the end defines the coupling between the regularized image, f_λ, and the raw image, f_0. Together with the first quadratic term it forces the regularized image to be close to the raw image (in L^2-sense). The second quadratic term is the regularization term, and contains a free control parameter $\lambda > 0$. It forces

This work is part of the DSSCV project supported by the IST Programme of the European Union (IST-2001-35443). Maurice Duits is gratefully acknowledged for his help.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

the regularization to be smooth in a trade-off with the remaining terms. As the notation suggests we have $\lim_{\lambda\to 0} f_\lambda = f_0$ (again, in L^2-sense). The equivalent Fourier formulation reads as follows:

$$\mathcal{E}\left[\hat{f}_\lambda\right] = \int \frac{1}{2}|\hat{f}_\lambda(\omega)|^2 + \frac{1}{2}\lambda\|\omega\|^2|\hat{f}_\lambda(\omega)|^2 - \hat{f}^*_\lambda(\omega)\hat{f}_0(\omega)\,d\omega\,. \tag{2}$$

The corresponding Euler-Lagrange equations are easily solved. We find

$$f_\lambda(x) = (I - \lambda\Delta)^{-1} f_0(x)\,, \tag{3}$$

respectively

$$\hat{f}_\lambda(\omega) = \frac{1}{1+\lambda\|\omega\|^2}\hat{f}_0(\omega)\,. \tag{4}$$

The Fourier route is clearly more explicit. The spatial formula is essentially defined by virtue of its Fourier equivalent [11]. For dimensions $n = 1, 2, 3$ the convolution filters corresponding to the linear operator in Eq. (3) are respectively given by

$$\begin{aligned}
\phi_\lambda(x) &= \frac{1}{\sqrt{4\lambda}}\exp\left[-\frac{|x|}{\sqrt{\lambda}}\right]\,,\\
\phi_\lambda(x,y) &= \frac{1}{2\pi\lambda}K_0\left[\frac{\sqrt{x^2+y^2}}{\sqrt{\lambda}}\right]\,,\\
\phi_\lambda(x,y,z) &= \frac{1}{4\pi\lambda\sqrt{x^2+y^2+z^2}}\exp\left[-\frac{\sqrt{x^2+y^2+z^2}}{\sqrt{\lambda}}\right]\,,
\end{aligned}$$

in which K_0 is the zeroth order modified Bessel function of the second kind.

Before we relate the above results to scale space theory, let us recall that a Gaussian scale space image satisfies the following heat equation with initial condition:

$$\left\{\begin{array}{rcl} \dfrac{\partial u}{\partial s} &=& \Delta u \\ \lim\limits_{s\to 0} u &=& f_0\,. \end{array}\right. \tag{5}$$

In the Fourier domain we have:

$$\left\{\begin{array}{rcl} \dfrac{\partial \hat{u}}{\partial s} &=& -\|\omega\|^2\hat{u} \\ \lim\limits_{s\to 0} \hat{u} &=& \hat{f}_0\,. \end{array}\right. \tag{6}$$

Let us write f_s for the solution at fixed scale $s > 0$. Assuming suitable boundary conditions the solution can be written as [12]

$$f_s(x) = \exp(s\Delta)\, f_0(x)\,, \tag{7}$$

or

$$\hat{f}_s(\omega) = \exp\left(-s\|\omega\|^2\right)\hat{f}_0(\omega)\,. \tag{8}$$

Again, the Fourier route is somewhat easier to grasp.

One observation readily presents itself [10]: If one replaces the inverse scale space generator by its first order Taylor polynomial one obtains

$$\exp(s\Delta) \approx (I - s\Delta)^{-1}\,,$$

respectively

$$\exp\left(-s\|\omega\|^2\right) \approx \frac{1}{1+s\|\omega\|^2}\,.$$

Comparison with Eqs. (3–4) shows that scale s in a Gaussian scale space representation is the analogue of the Tikhonov regularization parameter λ. This is indeed merely an analogy, since the relation holds only by approximation.

For the formal, non-approximative connection alluded to in the introduction, consider the following regularization functional in one spatial dimension. Note that it contains infinitely many regularization terms, but only *one* independent regularization parameter, viz. $t > 0$.

$$\mathcal{E}[\,f_t\,] = \int \frac{1}{2}\sum_{i=0}^{\infty}\frac{t^i}{i!}\,f_t^{(i)}(x)\,f_t^{(i)}(x) - f_t(x)\,f_0(x)\,dx\,. \tag{9}$$

Note the similarity with Eq. (1). It is easy to verify that the Euler-Lagrange equation is

$$\exp\left\{-t\frac{d^2}{dx^2}\right\} f_t(x) = f_0(x)\,,$$

in which the function exp is defined in terms of its Taylor expansion in the usual way. The formal solution is

$$f_t(x) = \exp\left\{t\frac{d^2}{dx^2}\right\} f_0(x)\,.$$

To see that this is indeed just linear scale-space theory, differentiate with respect to t, and consider the initial condition at $t = 0$. Writing $u(x,t)$ instead of $f_t(x)$, and using indices to denote partial derivatives, we obtain

$$\left\{\begin{array}{rcl} u_t &=& u_{xx} \\ u(x; t=0) &=& f_0(x)\,. \end{array}\right.$$

The solution is indeed $u(x,t) = f_t(x)$. The higher dimensional case does not bring in additional scale parameters in the isotropic case. For this and further details we refer to the literature [4, 10].

In the next section we show that Gaussian scale space and Tikhonov regularization can also be linked without approximation in the simplest case of a first order scheme. In this case, however, it will turn out that there is no longer a one-to-one relation between scale and regularization parameter.

2.2. A New Result

Recall that the negative exponential distribution,

$$P_\kappa(s) = \kappa\exp(-\kappa s)\,, \tag{10}$$

in which $\kappa > 0$ is a constant, is the only continuous probability distribution which has no memory. That is to say, if $P(S \leq s)$ denotes the unconditional probability that the (positive) stochastic variable S assumes a value less than or equal to s, and $P(S \leq s + t\,|\,S > t)$ the probability that it assumes a value $t < S \leq s + t$ given $S > t$, then the Markov condition

$$P(S \leq s+t\,|\,S>t) = \frac{P(t<S\leq s+t)}{P(S>t)} = P(S\leq s)$$

implies that $P(S\leq s)$ is of the form

$$P(S\leq s) = \int_0^s P_\kappa(t)\,dt$$

for some $\kappa > 0$. The negative exponential distribution satisfies a first order differential equation subject to a normalization condition, viz.

$$\left\{\begin{array}{rcl} \dot{P}_\kappa + \kappa P_\kappa &=& 0 \\ \displaystyle\int_0^\infty P_\kappa(s)\,ds &=& 1\,. \end{array}\right.$$

In nature such "memoryless" distributions govern many natural decay processes, e.g. radioactivity.

Eq. (10) can be used to attenuate coarse scales in a scale space representation. This, in fact, brings us to our main result. It relates the Gaussian scale space paradigm to first order Tikhonov regularization via the Laplace transform.

Result 1 *Let $P_\kappa(s)$ be the negative exponential distribution given by Eq. (10), then*

$$\int_0^\infty \exp(s\Delta)\; P_\kappa(s)\,ds = \left(I - \kappa^{-1}\Delta\right)^{-1} \;,$$

$$\int_0^\infty \exp\left(-s\|\omega\|^2\right)\; P_\kappa(s)\,ds = \frac{1}{1+\kappa^{-1}\|\omega\|^2} \,.$$

Notice that we have reobtained the filters of Eqs. (3–4) for $\kappa = 1/\lambda$. No approximation is involved.

Result 1 shows that first order Tikhonov regularization apparently boils down to "collapsing" a Gaussian scale space image along its scale axis with an attenuation in the form of a negative exponential distribution. Apparently *all* levels of resolution in a Gaussian scale space representation contribute to the first order Tikhonov regularization of an image for any fixed regularization parameter, but not equally. The most dominant levels are those within a boundary layer of scale extent $\mathcal{O}(\kappa^{-1}) = \mathcal{O}(\lambda)$ near zero scale. The result also clarifies the relation between the parameter λ ("regularization scale") and Gaussian scale s, for we have

$$\lambda = \int_0^\infty s\,P_\kappa(s)\,ds \,.$$

In other words, $\lambda = 1/\kappa$ is just the average scale of the distribution. Put differently, the scale "half-life" equals $s_{\frac{1}{2}} = \lambda \ln 2$. Alternatively one can say that regularization scale is the Laplace transform of Gaussian scale. Cf. Figure 1.

The integral in Result 1 can readily be implemented after discrete approximation, and this is in fact how Figure 1 was obtained. However, there is a fundamental limitation to the accuracy of the result imposed by the lower bound on the physically meaningful scale interval. Since scales smaller than some $\epsilon > 0$ are not represented, one could use the following error measure as a rule of thumb:

$$E(\kappa,\epsilon) = \int_0^\epsilon P_\kappa(s)\,ds = 1 - \exp(-\epsilon\,\kappa) \,,$$

i.e. the discarded part of the distribution's weight. For example, if we set $\epsilon = 1$, which corresponds to a Gaussian width of $\sqrt{2\epsilon} \approx 1.4$ pixels, then the errors for the bottom row images of Figure 1 evaluate to $0.52, 0.095, 0.013, 0.0018$, respectively. Thus at least the first image should be taken with a grain of salt.

Since the Laplace transform has an inverse we may to some extent reverse Result 1.

Result 2 *Definitions as in Result 1. For any $\rho > 0$ we have*

$$\frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \left(I - \kappa^{-1}\Delta\right)^{-1}\; P_\kappa^{-1}(s)\,d\kappa = \exp(s\Delta) \;,$$

or, equivalently,

$$\frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \frac{P_\kappa^{-1}(s)\,d\kappa}{1+\kappa^{-1}\|\omega\|^2} = \exp\left(-s\|\omega\|^2\right) \,.$$

This follows from the residue theorem. To see this, consider the Fourier formula. Close the integration contour by connecting the points at infinity on the vertical line $\mathrm{Re}(z) = \rho$ via a positively oriented semicircle $\partial\Gamma$ of infinite radius in the left part of the $\mathbb{C}$-plane, cf. Figure 2. Let Γ be the interior region, i.e. $\mathrm{Re}(z) < \rho$. Then

$$\frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \frac{P_\kappa^{-1}(s)\,d\kappa}{1+\kappa^{-1}\|\omega\|^2} = \frac{1}{2\pi i}\oint_{\partial\Gamma} \frac{P_\kappa^{-1}(s)\,d\kappa}{1+\kappa^{-1}\|\omega\|^2} \,.$$

Using the residue theorem and the fact that the integrand has a single pole at $\kappa = -\|\omega\|^2$ the right hand side can be written as

$$\mathrm{Res}_{\kappa=-\|\omega\|^2} \frac{P_\kappa^{-1}(s)}{1+\kappa^{-1}\|\omega\|^2} = \exp\left(-s\|\omega\|^2\right) \,.$$

The argument also shows why the result does not depend on the value of ρ, as long as it remains positive.

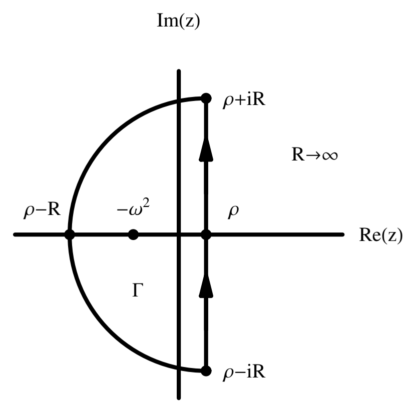

Fig. 2. Integration contour used in the residue theorem.

Unfortunately, the inversion formula of Result 2 has no operational significance, since the pole lies in the nonphysical halfplane $\kappa \le 0$. It thus remains an open question whether one can actually obtain a Gaussian scale space representation given the first order Tikhonov regularization of an image for all regularization parameters $\kappa > 0$.

It should be obvious, at least in the Fourier representation, that one can generalize Results 1 and 2 by replacing Δ and $\|\omega\|^2$ by $-(-\Delta)^\alpha$ and $\|\omega\|^{2\alpha}$, respectively, with $0 < \alpha \le 1$. For details on the resulting, so-called α scale spaces the interested reader is referred to the literature [11, 13]. Of particular interest is the so-called Poisson scale space, which corresponds to $\alpha = \frac{1}{2}$.

3. SUMMARY AND DISCUSSION

We have established a new result that relates Gaussian scale space theory to first order Tikhonov regularization in a non-approximative and operationally well-defined way. The result shows that the choice of regularization parameter in the Tikhonov scheme does not correspond to the selection of a particular scale. Rather, regularization is shown to be equivalent to "collapsing" a Gaussian scale space along the scale axis with an attenuation in the form of a negative exponential distribution. The regularization parameter can be identified—up to an irrelevant proportionality constant—with the average scale or the "half-life" that characterizes the negative exponential scale distribution.

Fig. 1. Top row: Four scale levels from a Gaussian scale space image, $s = 1.36, 10.0, 74.2, 548$, or, in terms of the physical width parameter $\sigma = \sqrt{2s}$: $\sigma = 1.65, 4.48, 12.2, 33.1$ pixels. Bottom row: Corresponding levels of Tikhonov regularization with synchronization $\lambda = s$, i.e. Gaussian scale s is taken to be equal to the average scale λ of the negative exponential distribution, Eq. (10). All images are of size 640×480. Slightly more detail can be discerned in the bottom row images as compared to corresponding ones above due to the high resolution bias intrinsic to Tikhonov regularization, but apart from that scales and regularization parameters appear well synchronized.

In mathematical terms, first order Tikhonov regularization scale space is the forward Laplace transform of Gaussian scale space (the same holds for the corresponding parameters). This implies that the latter is the inverse Laplace transform of the former. However, only the forward transform is operationally well-defined and well-posed. It remains hitherto unknown whether Gaussian scale space can be retrieved from first order Tikhonov regularization scale space in a stable manner by some alternative procedure.

Tikhonov regularization schemes of finite order provide only a limited degree of regularity. For instance, the first order scheme proposed in this paper admits only second order differentiation, regardless of the value of the regularization parameter. It does not solve ill-posedness of differential operators, but circumvents problems for the admissible low orders of differentiation by processing the input image. For this reason it is conceptually more natural to resort to scale space theory (or distribution theory in general), in which well-posed differentiation is manifest regardless of the value of the scale parameter.

4. REFERENCES

[1] A. Tikhonov and V. Y. Arseninn, *Solution of Ill-Posed Problems*, John Wiley & Sons, New York, 1977.

[2] L. Schwartz, *Théorie des Distributions*, Publications de l'Institut Mathématique de l'Université de Strasbourg. Hermann, Paris, second edition, 1966.

[3] F. G. Friedlander, *Introduction to the Theory of Distributions*, Cambridge University Press, Cambridge, 1982.

[4] L. M. J. Florack, *Image Structure*, vol. 10 of *Computational Imaging and Vision Series*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[5] T. Iijima, "Basic theory on normalization of a pattern (in case of typical one-dimensional pattern)," *Bulletin of Electrical Laboratory*, vol. 26, pp. 368–388, 1962, (in Japanese).

[6] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

[7] T. Lindeberg, *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[8] A. P. Witkin, "Scale-space filtering," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, 1983, pp. 1019–1022.

[9] J. J. Koenderink and A. J. van Doorn, "Receptive field families," *Biological Cybernetics*, vol. 63, pp. 291–298, 1990.

[10] M. Nielsen, L. M. J. Florack, and R. Deriche, "Regularization, scale-space and edge detection filters," *Journal of Mathematical Imaging and Vision*, vol. 7, no. 4, pp. 291–307, October 1997.

[11] R. Duits, L. Florack, J. de Graaf, and B. ter Haar Romeny, "On the axioms of scale space theory," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 3, pp. 267–298, 2004.

[12] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "The Gaussian scale-space paradigm and the multiscale local jet," *International Journal of Computer Vision*, vol. 18, no. 1, pp. 61–75, April 1996.

[13] M. Felsberg and G. Sommer, "The monogenic scale-space: A unifying approach to phase-based image processing in scale-space," Accepted for publication in Journal of Mathematical Vision and Imaging.