

# AN OPTIMAL CONTROL PROBLEM IN MEDICAL IMAGE PROCESSING

K. Bredies,<sup>1</sup> D. A. Lorenz,<sup>1</sup> and P. Maass<sup>1</sup>

<sup>1</sup>*University of Bremen, Bibliothekstrasse 2, 28359 Bremen, Germany, {kbredies,dlorenz,pmaass}@math.uni-bremen.de*

**Abstract** As a starting point of this paper we present a problem from mammographic image processing. We show how it can be formulated as an optimal control problem for PDEs and illustrate that it leads to penalty terms which are non-standard in the theory of optimal control of PDEs.

To solve this control problem we use a generalization of the conditional gradient method which is especially suitable for non-convex problems. We apply this method to our control problem and illustrate that this method also covers the recently proposed method of surrogate functionals from the theory of inverse problems.

**Keywords:** generalized conditional gradient method, surrogate functionals, image processing, optimal control of PDEs

## 1. Motivation from medical imaging

For many years medical imaging has aimed at developing fully automatic, software based diagnostic systems. However, the success of those automatic systems is rather limited and the human expert is as much responsible for the final diagnosis as in previous years. Hence, growing effort has been devoted to enhancing the techniques for presenting the medical images as well as additional information.

In Germany a particular effort is made in mammography, i. e. X-ray scans of the female breast for early detection of breast cancer. The process of examination by the medical experts is divided into a very short recognition phase ( $< 1$  sec.) and a second verification phase ( $\approx 1$  min.).

During the recognition phase, the expert first recognizes the coarse features, then more and more fine features. Tests have shown, that the experts usually form their decisions during this very short recognition phase. Nevertheless, the verification phase is the more critical one. The critical and difficult cases, where the recognition phase does not end with a preliminary diagnosis, most often applies to women in the early stages of cancer. During the verification phase

---

*Please use the following format when citing this chapter:*

Bredies, K., Lorenz, D.A., and Maass, P., 2006, in IFIP International Federation for Information Processing, Volume 202, Systems, Control, Modeling and Optimization, eds. Ceragioli, F., Dontchev, A., Furuta, H., Marti, K., Pandolfi, L., (Boston: Springer), pp. 249-259.

the expert shifts forwards and backwards, thereby alternating in examining small details and in catching an overall impression of the location of critical patterns within the organ.

This process can be supported by presenting the expert different versions of the original image during close up and normal subphases. More precisely, the expert sees a version with contrast enhanced small details in a close up phase ('fine scale'), while he sees an image which preserves all major edges but smoothes within regions ('coarse scale') during the normal phase. For enhancing fine details in mammography images a variety of algorithm have been proposed. Many of them are based on the wavelet transform due to its property of dividing an image into different scale representations, see for example [8] and references therein.

In this work we deal with the development of an optimized presentation for one cycle of the verification phase. To put the problem in mathematical terms, we start with a given image  $y_0$  assumed to be a function defined on  $\Omega := [0, 1]^2$ . The fine scale and the coarse scale image are denoted  $y_f$  and  $y_c$  respectively. Under the natural assumption of finite energy images we model them as functions in  $L^2(\Omega)$ . The goal is, to produce a movie (i. e. a time dependent function)  $y : [0, 1] \rightarrow L^2(\Omega)$ , from the given images  $y_0$ ,  $y_f$  and  $y_c$  such that

- the movie starts in  $y_0$ , i. e.  $y(0) = y_0$ ,
- the movie sweeps to the fine scale image and to the coarse scale image, e. g.  $y(t) \approx y_f$  for  $t \in [.2, .4]$  and  $y(t) \approx y_c$  for  $t \in [.6, .8]$ ,
- the movie sweeps in a "natural" way.

An example for a mammography image, a fine scale, and coarse scale image is shown in Figure 1. As a first guess one could try to make a linear interpolation between the fine scale and the coarse scale representation. This method has one serious drawback: It does not take the scale sweep into account, i. e. all fine details are just faded in rather than developing one after another.

Hence, more advanced methods have to be employed. In this article we show a solution of this presentation problem based on optimal control of PDEs. To simplify the presentation we only model a sweep between two images in the following.

## 2. Modeling as an optimal control problem

### 2.1 PDEs and control problems in image processing

Parabolic partial differential equations are a widely used tool in image processing. Diffusion equations like the heat equation [15], the Perona-Malik

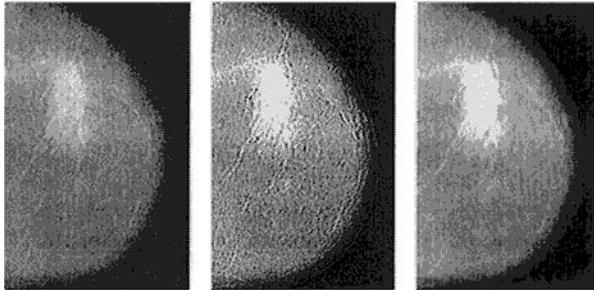


Figure 1. A mammography image. Left: original image  $y_0$ , middle: fine scale image  $y_f$ , right: coarse scale image  $y_c$

equation [11] or anisotropic equations [14] are used for smoothing, denoising and edge enhancing.

The smoothing of a given image  $y_0 \in L^2(\Omega)$  with the heat equation is done by the solution of the equation

$$\begin{aligned} y_t - \Delta y &= 0 \text{ in } [0, 1] \times \Omega \\ y_\nu &= 0 \text{ on } [0, 1] \times \partial\Omega \\ y(0) &= y_0 \end{aligned}$$

where  $y_\nu$  stands for the normal derivative, i. e. we impose homogeneous Neumann boundary conditions.

The solution  $y : [0, 1] \rightarrow L^2(\Omega)$  gives a movie which starts at the image  $y_0$  and becomes smoother with time  $t$ . This evolution is also called scale space and is analyzed by the image processing community in detail since the 1980s. Especially the heat equation does not create new features with increasing time, see e. g. [6] and the references therein.

Thus, the heat equation is well suited to model a sweep from a fine scale image  $y_f$  to a coarse scale image  $y_c$ . Hence, we take the image  $y_f$  as initial value. To make the movie  $y$  end at a certain coarse scale image  $y_c$  instead of its given endpoint  $y(1)$  we propose the following optimal control problem:

$$\begin{aligned} \text{Minimize} \quad & J(y, u) = \frac{1}{2} \int_{\Omega} |y(1) - y_c|^2 dx + \frac{\alpha}{2} \int_0^1 \int_{\Omega} |u|^2 dx dt \\ \text{subject to} \quad & y_t - \Delta y = u \text{ in } [0, 1] \times \Omega \\ & y_\nu = 0 \text{ on } [0, 1] \times \partial\Omega \\ & y(0) = y_f. \end{aligned}$$

In other words, the diffusion process is forced to end in  $y_c$  with the help of a heat source  $u$ .

## 2.2 Adaption to image processing

The above described problem is classical in the theory of optimal control of PDEs, though not well adapted to image processing. The solution of this problem may have several drawbacks: The control  $u$  will be smooth due to the regularization and have a large support. This will result in very smooth changes in the image sequence  $y$  and, more worse, in global changes in the whole image. To overcome these difficulties, different norms can be used for regularization. A widely used choice in image processing is to use Besov norms because they are appropriate to model images. Besov norms can be defined in different ways, e. g. in terms of moduli of smoothness [13] or in terms of Littlewood-Paley decompositions [7]. Here we take another viewpoint and define the Besov spaces via norms of wavelet expansions [3, 10]. For a sufficient smooth wavelet  $\psi$  the Besov semi norm of a function  $f$  on a set  $M \subset \mathbf{R}^d$  is defined as

$$|f|_{B_{p,q}^s(M)}^q = \sum_j \left( 2^{sjp} 2^{j(p-2)d/2} \sum_{i,k} |\langle f, \psi_{i,j,k} \rangle|^p \right)^{q/p}$$

where  $j$  is the scale index,  $k$  indicates translation and  $i$  stands for the directions. The Besov space  $B_{p,q}^s(M)$  is defined as the functions  $f \in L^p(M)$  that has a finite Besov semi norm. See [7, 10] for a more detailed introduction to wavelets and Besov spaces.

One particular Besov space plays a special role in image processing: the space  $B_{1,1}^{d/2}(M)$ . This is because it is close to the space of functions of bounded variation  $BV(M)$  which used to model images [12]. Especially one has  $B_{1,1}^{d/2}(M) \subset BV(M)$ . In the following we use the scale  $B_{p,p}^s([0, 1] \times \Omega)$  of Besov spaces and the wavelet norms

$$|f|_{B_{p,p}^s([0,1] \times \Omega)}^p = \sum_{i,j,k} w_j |\langle f | \psi_{i,j,k} \rangle|^p$$

with the weighting sequence  $w_j = 2^{sjp} 2^{j(p-2)3/2}$ .

## 2.3 The solution of the PDE and the control-to-state mapping

The solution of the heat equation is a classical task. If we assume that the initial value  $y_f$  is in  $L^2(\Omega)$  and the control  $u$  is in  $L^2([0, 1] \times \Omega)$  the solution  $y$  is in  $L^2(0, 1, H^1(\Omega)) \cap C([0, 1], L^2(\Omega))$ . Especially  $y$  is continuous with respect to time and the point evaluation  $y(1)$  makes sense, see e. g. [9].

In our case the solution operator  $u \mapsto y$  is affine linear, due to the non-zero initial value. We make the following modifications to come back to a linear problem: We split the solution into two parts. The non-controlled part  $y^n$  is the

solution of

$$\begin{aligned} y_t^n - \Delta y^n &= 0 \\ y^n(0) &= y_f \end{aligned}$$

and the homogeneous part  $y^h$  is the solution of

$$\begin{aligned} y_t^h - \Delta y^h &= u \\ y^h(0) &= 0 \end{aligned} \tag{1}$$

(both with homogeneous Neumann boundary conditions). Then the solution operator  $G : u \mapsto y^h$  of equation (1) is linear and continuous from  $L^2([0, 1], L^2(\Omega))$  to  $L^2(0, 1, H^1(\Omega)) \cap C([0, 1], L^2(\Omega))$ . With the help of the point evaluation operator we have the control-to-state mapping  $K : u \mapsto y^h(1)$  linear and continuous from  $L^2([0, 1], L^2(\Omega))$  to  $L^2(\Omega)$ . Then the solution is  $y = y^n + y^h$  and we can focus on the control problem for  $y^h$ .

Together with the thoughts of the previous subsection we end up with the following minimization problem:

$$\text{minimize } J(u) = \frac{1}{2} \|Ku - y_c + y^n(1)\|_{L^2(\Omega)}^2 + \alpha |u|_{B_{p,p}^s([0,1] \times \Omega)}^p. \tag{2}$$

### 3. Solution of the optimal control problem

The minimization of the functional (2) is not straightforward. The non-quadratic constraint leads to a nonlinear normal equation which can not be solved explicitly. Here we use a generalization of the conditional gradient method for the minimization.

#### 3.1 The generalized conditional gradient method

The classical conditional gradient method deals with minimization problems of the form

$$\min_{u \in C} F(u), \tag{3}$$

where  $C$  is a bounded convex set and  $F$  is a possible non-linear function. One notices that this constrained problem can actually be written as an “unconstrained” one with the help of the indicator functional

$$I_C(u) = \begin{cases} 0 & u \in C \\ \infty & u \notin C \end{cases}.$$

With  $\Phi = I_C$ , problem (3) thus can be reformulated as

$$\min_{u \in H} F(u) + \Phi(u). \tag{4}$$

To illustrate the proposed generalization, we summarize the key properties of  $F$  and  $\Phi$ :  $F$  is smooth while  $\Phi$  may contain non-differentiable parts. The minimization problem with  $\Phi$  alone is considered to be solved easily while the minimization of  $F$  is comparatively hard. The influence of  $\Phi$  is rather small in comparison to  $F$ .

With these assumptions in mind, the conditional gradient method can also be motivated as follows. Let  $u \in H$  be given such that  $\Phi(u) < \infty$ . We like to find an update direction by a linearized problem. Since  $\Phi$  is not differentiable, we only linearize  $F$ :

$$\min_{v \in H} \langle F'(u)|v \rangle + \Phi(v). \quad (5)$$

The minimizer of this problem serves as an update direction.

So this “generalized conditional gradient method” in the  $(n+1)$ -st step reads as follows: Let  $u_n \in H$  be given such that  $\Phi(u_n) < \infty$ .

- 1 Determine the solution of (5) and denote it  $v_n$ .
- 2 Set  $s_n$  as a solution of

$$\min_{s \in [0,1]} F(u_n + s(v_n - u_n)) + \Phi(u_n + s(v_n - u_n)).$$

- 3 Let  $u_{n+1} = u_n + s_n(v_n - u_n)$ .

To ensure existence of a solution in Step 1 we state the following condition:

**ASSUMPTION 1** *Let the functional  $\Phi : H \rightarrow ]-\infty, \infty]$  be proper, convex, lower semi-continuous and coercive with respect to the norm.*

Standard arguments from convex analysis yield the existence of a minimizer in Step 1 of the algorithm [5]. So if  $F$  is Gâteaux-differentiable in  $H$ , the algorithm is well-defined. The convergence of the generalized conditional gradient method is analyzed in detail by the authors in [2]. The main result there is the following theorem.

**THEOREM 2** *Let  $\Phi$  satisfy Assumption 1 and let us assume that every set  $E_t = \{u \in H \mid \Phi(u) \leq t\}$  is compact. Let  $F$  be continuously Fréchet differentiable, let  $F + \Phi$  be coercive and  $u_0$  be given such that  $\Phi(u_0) < \infty$ . Denote  $(u_n)$  the sequence generated by the generalized conditional gradient method.*

*Then every convergent subsequence of  $(u_n)$  converges to a stationary point of  $F + \Phi$ . At least one such subsequence exists.*

Two remarks are in order: First, we notice that the theorem is also valid if the functional  $F$  is not convex. Second, the theorem only gives convergence to a stationary point which may seem unsatisfactory, specially if one wants to minimize non-convex functions. But this does not have to be a drawback, as we will see in the next section:

### 3.2 Application to the control problem

To apply the generalized conditional gradient method to the optimal control problem (2) we split the functional  $J$  into two parts as follows:

$$\begin{aligned}
 F(u) &= \frac{1}{2} \|Ku - y^*\|_{L^2(\Omega)}^2 - \frac{\lambda}{2} \|u\|_{L^2([0,1], L^2(\Omega))}^2 \\
 \Phi(u) &= \frac{\lambda}{2} \|u\|_{L^2([0,1], L^2(\Omega))}^2 + \alpha |u(t)|_{B_{p,p}^s([0,1] \times \Omega)}^p
 \end{aligned}$$

where  $y^* = y_c - y^n(1)$  and  $\lambda > 0$ . The non-differentiable part  $\Phi$  clearly fulfills Condition 1. Moreover, the level sets  $E_t$  are compact, if  $s > 3(2-p)/(2p)$  because then the embedding  $B_{p,p}^s([0,1] \times \Omega) \subset L^2([0,1] \times \Omega)$  is compact.

The derivative of  $F$  is given by

$$F'(u) = K^*(Ku - y^*) - \lambda u,$$

and hence the minimization problem in step one of the generalized gradient method reads as

$$\min_v \langle K^*(Ku - y^*) - \lambda u | v \rangle + \frac{\lambda}{2} \|v\|^2 + \alpha |v(t)|_{B_{p,p}^s([0,1] \times \Omega)}^p.$$

which can be written as in terms of wavelet expansions

$$\begin{aligned}
 \min_v \sum_{i,j,k} \left( \langle K^*(K(u - y^*) - \lambda u | \psi_{i,j,k}) \rangle \langle v | \psi_{i,j,k} \rangle \right. \\
 \left. + \frac{\lambda}{2} \langle v | \psi_{i,j,k} \rangle^2 + \alpha w_j |\langle v | \psi_{i,j,k} \rangle|^p \right).
 \end{aligned}$$

This minimization problem is equivalent to

$$\min_v \sum_{i,j,k} \left( \frac{1}{2} |\langle K^*(K(u - y^*) - \lambda u - \lambda v | \psi_{i,j,k}) \rangle|^2 + \alpha w_j |\langle v | \psi_{i,j,k} \rangle|^p \right).$$

Now the minimization problem reduces to pointwise minimization in the basis coefficients of  $v$ . This can be done analytically by the expression

$$v = \mathbf{S}_{\alpha w_j / \lambda}(u) = \sum_{i,j,k} S_{\alpha w_j / \lambda}(\langle u - \lambda^{-1} K^*(Ku - y^*) | \psi_{i,j,k} \rangle) \psi_{i,j,k}$$

where  $S_\alpha$  is the a shrinkage function given by

$$S_\alpha(x) = \begin{cases} G_{\alpha,p}^{-1}(x) & p > 1 \\ \text{sign}(x) \max(|x| - \alpha, 0) & p = 1 \end{cases}$$

with  $G_{\alpha,p}(x) = x + \alpha p \operatorname{sign}(x)|x|^{p-1}$ .

To sum up, the generalized conditional gradient method for the control problem reads as:

- Initialize with  $u_0$  and choose parameter  $\lambda > 0$ .
- Calculate direction  $v_n = \mathbf{S}_{\alpha w_j/\lambda}(u_n)$
- Calculate step size  $s_n$  such that  $J(u_n + s_n(v_n + u_n))$  is minimal.
- Update  $u_{n+1} = u_n + s_n(v_n - u_n)$  and set  $n \leftarrow n + 1$ .

**REMARK 3** *The choice of the step size can be done analytically for  $p = 2$  and  $p = 1$  easily with just one evaluation of the operator  $K$  (again we refer to [2] for details). Furthermore, one can choose an approximate step size which decreases the functional  $J$  nevertheless. This is always the case is the step size is small enough.*

### 3.3 Equivalence with the method of surrogate functionals

In this subsection we will point out, that the above proposed algorithm for minimization problems of the form

$$J(u) = \frac{1}{2} \|Ku - f\|^2 + \alpha |u|_{B_{p,p}^s}^p$$

is an extension of the method of surrogate functionals as proposed recently by Daubechies, Defries, De Mol [4]. This method is based on the so called surrogate functional

$$\tilde{J}(u, a) = J(u) - \frac{1}{2} \|Ku - Ka\|^2 + \frac{1}{2} \|u - a\|^2.$$

One notices that  $\tilde{J}(u, u) = J(u)$ . The minimization of  $J$  is done by alternate minimization of  $\tilde{J}$  with respect to the variables  $a$  and  $u$ . Since minimization with respect to  $a$  just gives  $u = a$  one can rewrite this iteration as

$$u_{n+1} = \operatorname{argmin} \tilde{J}(u, u_n).$$

In [4] it is proved, that this minimization leads to the iteration

$$u_{n+1} = \mathbf{S}_{\alpha w_j}(u_n).$$

Hence, the method of surrogate functionals produces the same sequence as the generalized conditional gradient method with  $\lambda = 1$  and step size equals to one. Furthermore, in [4] it is proved, that the produced sequence converges strongly to a minimized of  $J$  if the sequence of weights  $w_j$  is strictly bounded away from zero.

Our result shows, that the method of surrogate functionals can be interpreted as a gradient descend method. Again, we refer to [2] for a more detailed description.

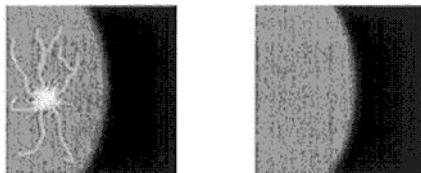


Figure 2. Images used for illustration. Left: fine scale image, right: coarse scale image.

## 4. Application

Here we show the application of the above described methodology. Since the effects can be seen more clearly in artificial images, we will not use original images. The artificial images we used are shown in Figure 2.

For illustration we use the values  $p = 1$  and  $s = 3/2 + \varepsilon > 3/2$  in the minimization problem (2), since this is close to the  $BV$ -norm and we have  $B_{1,1}^{3/2+\varepsilon}([0, 1] \times \Omega) \subset L^2([0, 1] \times \Omega)$  compactly.

The results are presented in Figure 3. The figure shows a comparison of the linear interpolation, the pure result of the application of the heat equation and the result of the optimal control problem. One sees that the linear interpolation is only fading out the details. In the uncontrolled result (middle column) the details are vanishing one after another but the process does not end in the desired endpoint. The result of the optimal control problem (right column) exhibits both a nice vanishing of the details and end in the given endpoint.

## 5. Conclusion

We have seen that the application of the theory of optimal control of PDEs to image processing problems is a fruitful field of research. Besides promising result, even for easy models like the linear heat equation, new interesting mathematical problems arise, like the treatment of non-quadratic penalty terms. For future research, better adapted PDEs (like the anisotropic diffusion equations) could be investigated.

## References

- [1]
- [2] K. Bredies, D. A. Lorenz, and P. Maass. Equivalence of a generalized conditional gradient method and the method of surrogate functionals. Preprint of the DFG Priority Program 1114 No. 135, University of Bremen, 2005.
- [3] A. Cohen. *Numerical Analysis of Wavelet Methods*. Elsevier Science B. V., 2003.
- [4] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

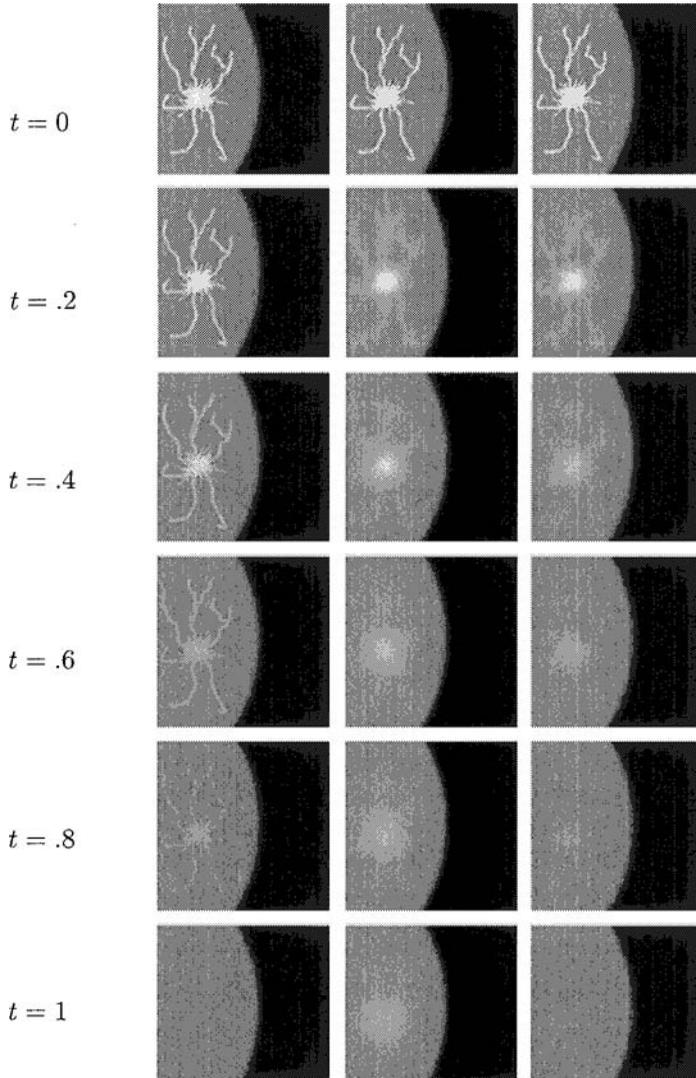


Figure 3. Comparison of different movies. Left column: linear interpolation between  $y_f$  and  $y_c$  from Figure 2. Middle column: Solution of the heat equation with initial value  $y_f$ . Right column: Solution of the optimal control problem (2).

- [5] I. Ekeland and R. Temam. *Convex analysis and variational problems*. North-Holland, Amsterdam, 1976.
- [6] L. Florack and A. Kuijper. The topological structure of Scale-Space images. *Journal of Mathematical Imaging and Vision*, 12:65–79, 2000.
- [7] M. Frazier, B. Jawerth, and G. Weiss. *Littlewood-Paley theory and the study of function spaces*. Number 79 in Regional Conference Series in Mathematics. American Mathematical Society, 1991.
- [8] P. Heinlein, J. Drexl, and W. Schneider. Integrated wavelets for enhancement of microcalcifications in digital mammography. *IEEE Transactions on Medical Imaging*, 22(3):402–413, March 2003.
- [9] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer, 1971.
- [10] Y. Meyer. *Wavelets and Operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1992.
- [11] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [12] L. I. Rudin, S. J. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [13] H. Triebel. *Theory of Function Spaces II*. Monographs in Mathematics. Birkhäuser, 1992.
- [14] J. Weickert. *Anisotropic diffusion in image processing*. Teubner, Stuttgart, 1998.
- [15] A. P. Witkin. Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983.