L-Tangent Norm: A Low Computational Cost Criterion for Choosing Regularization Weights and its Use for Range Surface Reconstruction

F. Brunet[1,2,3] A. Bartoli[1] R. Malgouyres[3] N. Navab[2]
[1]LASMEA (CNRS/UBP) [2]CAMPAR (TU München) [3]LAIC (UdA)
Clermont-Ferrand (France) Munich (Germany) Clermont-Ferrand (France)

Abstract

We are interested in fitting a surface model such as a tensor-product spline to range image data. This is commonly done by finding control points which minimize a compound cost including the goodness of fit and a regularizer, balanced by a regularization parameter. Many approaches choose this parameter as the minimizer of, for example, the cross-validation score or the L-curve criterion. Most of these criteria are expensive to compute and difficult to minimize.

We propose a novel criterion, the L-tangent norm, which overcomes these drawbacks. Even though it is empirical, it gives sensible results with a much lower computational cost. This new criterion has been successfully tested with synthetic and real range image data, and shows a behavior similar to cross-validation.

1. Introduction

This paper is concerned with the reconstruction of a surface from range image data (also known as 2.5D data). Such data are obtained by range sensors such as Time-of-Flight cameras or stereo imaging. Reconstructing a surface from scattered data points is important for computing geodesics, texture-mapping, global shape editing, etc.

A surface reconstructed from a range image is usually described as a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ called a surface model. A set of parameters controls the surface shape. Many models have been proposed including the Thin-Plate Spline [2, 6, 14, 15], Radial Basis Functions [8, 12, 13], Bézier surfaces [7] and tensor-product splines over the B-spline basis [3, 4].

The classical approach to reconstruct a surface is by finding the set of parameters which minimizes a cost function. This cost function has two terms: a term reflecting the goodness of fit and a regularization term. These two terms are related by a parameter, the so-called *regularization parameter*, which controls the importance given to the regularization. A small value for the regularization parameter results in a surface passing closely to the data points but prone to overfitting. On the contrary, a large regularization parameter results in a smooth surface which may not approximate the data very well.

One of the challenges in surface reconstruction (and in many other data fitting problems) is selecting the regularization parameter automatically. Many approaches have been proposed. Two of the mostly used methods are the cross-validation score minimization [16] and the L-curve criterion maximization [9, 11]. Cross-validation intends to maximize the ability of the reconstructed surface to generalize. The L-curve criterion selects the regularization parameter by connecting the residual (*i.e.* the closeness of the surface to the data points) and the solution norm (*i.e.* the smoothness of the surface). Unfortunately, these criteria are generally expensive to compute. Moreover, they require to solve optimization problems that are generally difficult. We propose a novel heuristic to select the regularization parameter: the *L-tangent norm*. This new approach is interesting for two main reasons. First, it is cheap to compute. Second, its 'shape' and numerical behavior make the selection of the regularization parameter easy.

Paper organization. Section 2 is dedicated to the problem of reconstructing the surface given a regularization parameter. Section 3 is a short review of existing methods to automatically select the regularization parameter. Then, our new approach, the L-tangent norm, is presented in section 4. Finally, experimental results are shown in section 5.

Notations. Scalars are in italics, *e.g.* x, vectors in bold right fonts, *e.g.* $\mathbf{p}$, and matrices in capitals, *e.g.* M. The vector and matrix transpose is denoted with the symbol $^\mathsf{T}$, *e.g.* M^T. Intervals are denoted with square brackets, *e.g.* $[a, b]$, $]a, b[$ and $]a, b]$ for respectively a closed, an open and a half-open interval.

2. Background

2.1. Reconstruction Given the Regularization Parameter

Assume one is given a set of n range data points. Such a set is composed of n twodimensional points $(x_i, y_i) \in \mathbb{R}^2$, associated to depth information $z_i \in \mathbb{R}$. This set of points is denoted:

$$\{(x_i, y_i) \leftrightarrow z_i \mid i = 1, \ldots, n\}. \tag{1}$$

The surface model is a function of h unknown parameters $\mathbf{p} = \begin{bmatrix} p_1, \ldots, p_h \end{bmatrix}^\mathsf{T} \in \mathbb{R}^h$ that control the shape of the surface:

$$\begin{array}{rccc} f(\cdot;\mathbf{p}): & \Omega \subset \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (x,y) & \longmapsto & f(x,y;\mathbf{p}). \end{array} \tag{2}$$

The surface model we use in this paper is given in section 2.2. The range surface reconstruction problem consists in finding the best set of parameters $\mathbf{p}^*_\lambda \in \mathbb{R}^h$ (which depends on λ) such that:

$$\mathbf{p}^*_\lambda = \arg\min_{\mathbf{p}\in\mathbb{R}^h} \mathcal{E}_d(\mathbf{p}) + \frac{\lambda}{1-\lambda}\mathcal{E}_r(\mathbf{p}) \qquad \lambda \in]0,1[, \tag{3}$$

where $\mathcal{E}_d$ and $\mathcal{E}_r$ are respectively called the *data term* and the *regularization term*. The data term is a function that measures the closeness of the surface to the whole set of data points. The regularization term is a measure of the surface regularity (or smoothness). These two terms are related by the *regularization parameter* λ which controls the trade-off between the goodness of fit and the regularity. In the limit $\lambda \to 0$, the surface is likely to overfit the data. If λ is large, the surface becomes very smooth but may not reflect the data very well ; for instance, with many models, the reconstructed surface is almost a plane when λ is close to 1.

The data term and the regularization term can be chosen as respectively the Mean Squared Residual (MSR) and the bending energy:

$$\mathcal{E}_d(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n} \left(f(x_i, y_i; \mathbf{p}) - z_i\right)^2 \tag{4}$$

$$\mathcal{E}_r(\mathbf{p}) = \iint_\Omega \sum_{d=0}^{2} \binom{2}{d} \left(\frac{\partial^2 f(\cdot;\mathbf{p})}{\partial x^{2-d}\partial y^d}(x,y)\right)^2 \mathrm{d}x\mathrm{d}y. \tag{5}$$

We consider that the surface model is linear with respect to its parameter:

$$\forall (x,y) \in \mathbb{R}^2, \exists\, \mathbf{v}_{x,y} \in \mathbb{R}^h : f(x,y;\mathbf{p}) = \mathbf{v}^\mathsf{T}_{x,y}\mathbf{p},$$

The MSR can be written as:

$$\mathcal{E}_d(\mathbf{p}) = \|M\mathbf{p} - \mathbf{z}\|_2^2 \tag{6}$$

where M is the *collocation matrix* and $\mathbf{z}$ is the vector containing all the depths:

$$M = \begin{bmatrix} \mathbf{v}^\mathsf{T}_{x_1,y_1} & \ldots & \mathbf{v}^\mathsf{T}_{x_n,y_n} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{n\times h},$$
$$\mathbf{z} = \begin{bmatrix} z_1 & \ldots & z_n \end{bmatrix}^\mathsf{T} \in \mathbb{R}^n.$$

The bending energy can be approximated by discretizing the integral sum of equation (5) over a regular grid:

$$\mathcal{E}_r(\mathbf{p}) \approx \frac{1}{ab}\sum_{i=0}^{a-1}\sum_{j=0}^{b-1}\sum_{d=0}^{2} \binom{2}{d} \left(\frac{\partial^2 f\left(\frac{i}{a}, \frac{j}{b}; \mathbf{p}\right)}{\partial x^{2-d}\partial y^d}\right)^2 \tag{7}$$

$$\iff \quad \mathcal{E}_r(\mathbf{p}) \approx \|R\mathbf{p}\|_2^2 \tag{8}$$

where R is the *regularization matrix*. Note that the partial derivatives of a linear model are also linear with respect to $\mathbf{p}$, *i.e.*, for all $d \in \{0,1,2\}$ and for all $(x,y) \in \mathbb{R}^2$, there exists $\mathbf{w}_{x,y,d} \in \mathbb{R}^h$ such that:

$$\frac{\partial^2 f(x,y;\mathbf{p})}{\partial x^{2-d}\partial y^d} = \mathbf{w}^\mathsf{T}_{x,y,d}\mathbf{p}. \tag{9}$$

If $\mathbf{r}^\mathsf{T}_k$ is the kth row of R, then we have that:

$$\mathbf{r}^\mathsf{T}_{dab+ib+j+1} = \binom{2}{d}\mathbf{w}^\mathsf{T}_{\frac{i}{a},\frac{j}{b},d}. \tag{10}$$

Finally, equation (3) is equivalent to a Linear Least Squares (LLS) minimization problem [1]:

$$\mathbf{p}^*_\lambda = \arg\min_{\mathbf{p}\in\mathbb{R}^h} \left\| \begin{bmatrix} M \\ \frac{\lambda}{1-\lambda}R \end{bmatrix} \mathbf{p} - \begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} \right\|_2^2. \tag{11}$$

The solution of this problem is given by:

$$\mathbf{p}^*_\lambda = \left(M^\mathsf{T}M + \left(\tfrac{\lambda}{1-\lambda}\right)^2 R^\mathsf{T}R\right)^{-1} M^\mathsf{T}\mathbf{z}. \tag{12}$$

2.2. The Surface Model

We have chosen to use the tensor-product splines over the B-spline basis (TPBS) model. A reason is that the influence of a control point is bounded to its neighborhood due to the local support of the B-splines basis [3, 4, 7]. This property leads to sparse collocation and regularization matrices. This makes the computations fast.

We remind the reader some basic facts about B-splines. An extensive review of splines can be found in [3, 4, 7].

The B-spline of degree $k > 0$ (order $k+1$) having the increasing knot sequence $\mu_0 \leq \ldots \leq \mu_{g+1}$ can be defined recursively by:

$$\begin{cases} N_{i,1}(x) = 1 & \text{if } x \in [\mu_i, \mu_{i+1}[\\ N_{i,1}(x) = 0 & \text{if } x \notin [\mu_i, \mu_{i+1}[\\ N_{i,k+1}(x) = \frac{x-\mu_i}{\mu_{i+k}-\mu_i}N_{i,k}(x) & \\ \qquad\qquad + \frac{\mu_{i+k+1}-x}{\mu_{i+k+1}-\mu_{i+1}}N_{i+1,k}(x). & \end{cases} \tag{13}$$

A TPBS is defined as a linear combination of the B-spline basis functions weighted by the *control points* $C_{i,j}$:

$$s(x,y) = \sum_{i=-k_1}^{g_1} \sum_{j=-k_2}^{g_2} C_{i,j} N_{i,k_1+1}(x) N_{j,k_2+1}(y). \quad (14)$$

Note that equation (14) is linear in the $C_{i,j}$. Note also that it is well-known [4] that the partial derivatives of a TPBS are also TPBS.

In the sequel, we always consider uniform knot sequences. Besides, we take as many knots as possible (regarding the computational complexity) so that the flexibility of the surface model is sufficient to approximate complex shapes.

3. Previous Work

3.1. Cross-Validation

The goal of Ordinary Cross-Validation (OCV) [15, 16] is to choose the regularization parameter so that the reconstructed surface generalizes well. In other words, the reconstructed surface must have a good behavior between the data points. The optimal regularization parameter is the minimizer of the so-called *OCV score*:

$$\lambda^* = \arg \min_{\lambda \in]0,1[} OCV(\lambda). \quad (15)$$

An example of the OCV criterion is given in figure 1a. The OCV score is defined by fitting the model without the ith data point, giving the parameter vector $\mathbf{p}_\lambda^{[i]}$. This is used to predict the ith measurement as $f(x_i, y_i; \mathbf{p}_\lambda^{[i]})$. This prediction is compared against the actual value z_i. This is averaged over the n data points, giving:

$$OCV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(f(x_i, y_i, \mathbf{p}_\lambda^{[i]}) - z_i \right)^2. \quad (16)$$

It is almost impossible to directly use this definition of the OCV as its evaluation for a single value of λ requires the reconstruction of n surfaces. It is well-known [15] that there exists a non-iterative formula that approximates closely the OCV score:

$$OCV(\lambda) = \frac{1}{n} \left\| \Delta \left(\frac{1}{\mathbf{1} - \Delta(H_\lambda)} \right) (H_\lambda - I)\mathbf{z} \right\|_2^2 \quad (17)$$

where I is the identity matrix, Δ the diagonal operator (*i.e.* $\Delta(\mathbf{u})$ is a square matrix having $\mathbf{u}$ as its main diagonal and $\Delta(A)$ extracts the diagonal entries of matrix A as a vector) and H_λ the *influence matrix*:

$$H_\lambda = M \left(M^\mathsf{T} M + \left(\frac{\lambda}{1-\lambda} \right)^2 R^\mathsf{T} R \right)^{-1} M^\mathsf{T}. \quad (18)$$

Even with the non-iterative equation (17), two problems remain. First, the amount of computation required to solve the minimization problem (15) is still too heavy for large datasets (say $n > 1000$). Second, minimizing the OCV score can be difficult. Indeed, this criterion is numerically unstable. This has the effect to introduce high frequency oscillations (see figure 1b) . It is thus difficult to estimate the criterion derivative which would be useful in an optimization process such as gradient descent.

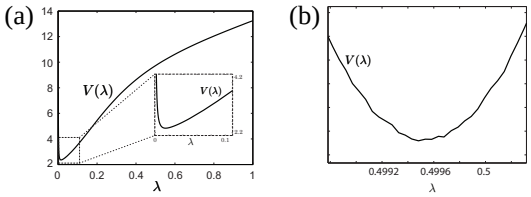

Figure 1. (a) A typical cross-validation score function. (b) High frequency oscillations resulting of numerical instability of the cross-validation score.

3.2. The L-Curve

The L-curve was introduced in [11]. An extensive review of this approach can be found in [9, 10]. The idea of this criterion is to find the best compromise between the goodness of fit and the surface smoothness. To do so, these two quantities are plotted against each other as functions of the regularization parameter.

Let $\rho(\lambda) = \|M\mathbf{p}_\lambda^* - \mathbf{z}\|$ be the *residual norm* and $\eta(\lambda) = \|R\mathbf{p}_\lambda^*\|$ be the *solution norm*. The L-curve is a continuous curve parametrized by the regularization parameter λ and defined by:

$$\left\{ \left(\hat{\rho} = \log \rho(\lambda), \hat{\eta} = \log \eta(\lambda) \right) \in \mathbb{R}_+^2 \mid \lambda \in]0,1[\right\}. \quad (19)$$

The L-curve method chooses one of the maximizers of the L-curve curvature, leading to:

$$\lambda^* = \arg \max_{\lambda \in]0,1[} \kappa(\lambda), \quad (20)$$

where κ is the curvature of the L-curve:

$$\kappa(\lambda) = 2 \frac{\hat{\rho}' \hat{\eta}'' - \hat{\rho}'' \hat{\eta}'}{\left(\hat{\rho}'^2 + \hat{\eta}'^2 \right)^{3/2}}. \quad (21)$$

When the L-curve has the shape of the letter L (see figure 2a), the 'corner' of the curve is well defined: the curvature (figure 2b) has one maximum which corresponds to the regularization parameter λ^* we are searching for. Unfortunately, the curvature often exhibits multiple maxima (see figure 3). In such cases, it is not clear how to choose the regularization parameter.