BERNSTEIN FILTER: A NEW SOLVER FOR MEAN CURVATURE REGULARIZED MODELS

Yuanhao Gong

National University of Singapore

ABSTRACT

The mean curvature has been shown a proper regularization in various ill-posed inverse problems in signal processing. Traditional solvers are based on either gradient descent methods or Euler Lagrange Equation. However, it is not clear if this mean curvature regularization term itself is convex or not. In this paper, we first prove that the mean curvature regularization is convex if the dimension of imaging domain is not larger than seven. With this convexity, all optimization methods lead to the same global optimal solution. Based on this convexity and Bernstein theorem, we propose an efficient filter solver, which can implicitly minimize the mean curvature. Our experiments show that this filter is at least two orders of magnitude faster than traditional solvers.

***Index Terms*—** Mean Curvature, Curvature Filter, Bernstein, convex, half-window regression.

1. INTRODUCTION

Estimating the signal from observed data is a fundamental task in signal processing. There are many theories that have been developed in past decades to achieve this goal, such as variational methods, wavelet theory, dictionary learning, compressed sensing, etc.

Among these theories, Bayesian Theorem can be used to derive variational methods. Given the observed data $I(\vec{x})$, where $\vec{x} \in \Omega \subset R^n$ is the spatial coordinate, Ω is the sampling domain, and n is the dimension, we want to estimate the signal $U(\vec{x})$. This can be carried out by Bayesian Theorem:

$$p(U|I) = \frac{p(I|U)p(U)}{p(I)} \propto p(I|U)p(U)\,, \tag{1}$$

where $p(\cdot)$ denotes the probability. Maximizing the probability $p(U|I)$ is equivalent to minimizing following energy

$$\mathcal{E}(U) = -\log(p(U|I))\,. \tag{2}$$

Plugging in the Eq. 1, we have

$$\mathcal{E}(U) = -\log(p(I|U)) - \log(p(U))\,. \tag{3}$$

Therefore, both the sampling model (generating I from U) and some prior about U have to be assumed in this framework. In general, this links Bayesian Theorem and the variational framework:

$$\begin{array}{llllll} \max \log(p(U|I)) = & \log(p(I|U)) & + & \log(p(U)) & \\ & \Updownarrow & & \Updownarrow & \\ \min \mathcal{E}(U) \quad = & \mathcal{E}_{\Phi_0}(U,I) & + & \mathcal{E}_{\Phi_1}(U)\,, \end{array} \tag{4}$$

where the double headed arrows indicate the counterparts in these two approaches.

The $\mathcal{E}_{\Phi_0}$ is the data fitting energy while the $\mathcal{E}_{\Phi_1}$ is the regularization energy. $\mathcal{E}_{\Phi_0}$ measures how well an estimation of U fits the data I. This generally depends on the sampling process. The ℓ_2 norm is commonly used because the measurement error usually satisfies Gaussian distribution, which leads to $\mathcal{E}_{\Phi_0} = \frac{\|U-I\|_2^2}{\sigma^2}$ (σ is a parameter). Another frequent choice is the ℓ_1 norm which corresponds to Laplacian noise model, because it is robust to outliers. In most of models, $\mathcal{E}_{\Phi_1}$ is a regularization term that imposes the prior knowledge about ground truth U, such as Tikhonov, the ℓ_2 norm of the gradient, symmetry [1], gradient distribution [2, 3, 4, 5], Total Variation (TV) [6], Mean Curvature (MC) [7, 8, 9, 10, 4], or Gaussian Curvature (GC) [11, 12, 4].

Among these regularizations, MC is particularly interesting because it is related to the minimal surfaces that is common in our physical world. Several researchers have shown that imposing MC leads to a better result compared with imposing TV [8, 9]. This, however, has not been explained theoretically before. In this paper, we show that MC imposes a higher order approximation to the signal than TV does.

1.1. Traditional Solvers

Although MC is preferred, it is difficult to minimize the mean curvature regularized models. Traditionally, there are two ways to minimize it. One is based on gradient descent method or called diffusion schemes [7, 8]

$$\frac{\partial U(\vec{x},t)}{\partial t} = -\frac{\partial \mathcal{E}(U(\vec{x},t))}{\partial U(\vec{x},t)}\,,\; U(\vec{x},0) = I(\vec{x})\,. \tag{5}$$

This type of methods suffer from the numerical stability requirement (CLF condition). Therefore, the step size in each iteration is limited. They usually need a large number of iterations to converge. The other type is based on Euler Lagrange Equation [9, 10], which is the necessary condition of being a local minimum. In general, these methods are more efficient

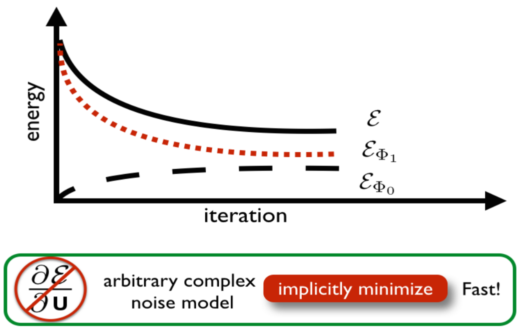


Fig. 1. Regularization is the dominant part in optimization.

than diffusion schemes. But the derived scheme is usually complicated and its physical meaning is missing or difficult to understand [10, 13].

1.2. Motivation

Besides the complicate equation, another drawback of traditional methods is that they are computationally slow [8, 9, 10], which hampers the application of mean curvature. To accelerate the computation, the multigrid strategy is adopted [13]. Even with the multigrid acceleration, these methods are still much slower than our filter as shown in the experiment section. The main reason is that these two types of approaches start from the total energy $\mathcal{E}(U)$ in Eq. 4 without considering the geometric meaning of minimizing mean curvature.

Another problem for previous solvers is that they are not generic. They require the imaging model $\mathcal{E}_{\Phi_0}$ to be differentiable with respect to U. This is not always true, for example, when $\mathcal{E}_{\Phi_0} = \|U - I\|_1$ or $\mathcal{E}_{\Phi_0} = \int |U - I|^{\|\nabla U\|} \mathrm{d}\vec{x}$.

1.3. Contribution

To overcome these issues, we propose to minimize the regularization energy. Our method is inspired by the observation that regularization energy $\mathcal{E}_{\Phi_1}$ is the dominant part during the minimization. As shown in Fig. 1, the regularization energy $\mathcal{E}_{\Phi_1}$ usually decreases while the energy $\mathcal{E}_{\Phi_0}$ usually increases if the initial condition is the original image. Since the total energy has to decrease, $\mathcal{E}_{\Phi_1}$ must be the dominant part. Therefore, as long as the decreased amount in $\mathcal{E}_{\Phi_1}$ is larger than the increased amount in $\mathcal{E}_{\Phi_0}$, the total energy $\mathcal{E}$ decreases.

There are several benefits of doing so. First, we do not require the total energy to be differential. Therefore, our method can handle arbitrary complex noise model. Second, we don't need to compute mean curvature, which means that the estimated image does not need to be smooth. Therefore, the edges are preserved. Third, the resulting filter is simple to compute, and its physical meaning is clear.

Another contribution of this paper is that we prove that the mean curvature regularization is convex when $n \leq 7$ (n is the dimension of imaging domain, for example, $n = 2$ for 2D images). Thanks to this convexity, the results from different solvers converge to the same global unique solution.

2. CONVEXITY OF MEAN CURVATURE REGULARIZATION

First, we show the relationship between mean curvature and minimal surfaces. Then we prove the convexity of mean curvature regularization, based on Bernstein theorem.

2.1. Mean Curvature Regularization

Let's write the mathematical form of mean curvature regularization. We embed the U into a higher dimension space to form a surface $\Psi(\vec{x}) = (\vec{x}, U(\vec{x})) \in R^{n+1}$. And the mean curvature of Ψ is defined as

$$H(\Psi) = \frac{1}{n}\sum_{i=1}^{n} \kappa_i \,, \tag{6}$$

where κ_i is the principle curvature of Ψ [14]. When $n = 2$, this mean curvature becomes the well-known case. The mean curvature regularization of U is defined as

$$\mathcal{E}_{\Phi_1}(U) = \mathcal{E}_H(U) = \mathcal{E}_H(\Psi) = \int_{\vec{x}\in\Omega} |H(\Psi)|^q \mathrm{d}\Psi \,, \tag{7}$$

where $q \geq 1$ is a positive real number. We take $q = 1$ in the rest of this paper.

2.2. Minimal Surface

Let's see how the mean curvature regularization is related to minimal surfaces. The area of Ψ is defined as

$$\mathcal{A}(\Psi) = \int_{\vec{x}\in\Omega} \sqrt{1 + |\nabla U|^2} \mathrm{d}\vec{x} \,, \tag{8}$$

where ∇ is the gradient operator. If $\mathcal{A}$ is minimal, using Euler Lagrange equation [15], we have

$$\nabla \cdot \left(\frac{\nabla U}{\sqrt{1 + |\nabla U|^2}} \right) = 0 \,, \tag{9}$$

where $\nabla\cdot$ is the divergence operator. This is known as minimal surface equation [16]. If U satisfies this equation, then Ψ is called minimal surface. In 1776, Meusnier proved that $\mathcal{E}_H(\Psi) = 0$ is equivalent to minimal surface equation [17]. Therefore, imposing the mean curvature regularization means assuming that the ground truth is a piecewise minimal surface.

2.3. Linearity and Convexity

Let Ω_i denote an open set region in Ω such that $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$ and $\Omega = \cup_i(\Omega_i \cup \partial\Omega_i)$, where $\partial\Omega_i$ denotes the boundary of Ω_i. On Ω_i, we have following classical result from differential geometry [18]

Theorem 1. **[Bernstein Theorem]** *On $\Omega_i \in R^n$ and $n \leq 7$, if Ψ is a minimal surface, then $U(\vec{x})$ must be a linear function.*

In most of image processing problems, this dimension requirement can be easily satisfied, for example, $n = 2$ for 2D images. This theorem leads to an important conclusion: imposing mean curvature regularization is equivalent to assuming that the signal is piecewise linear. **Therefore, mean curvature regularization is a higher order approximation to the signal, compared with TV regularization, which assumes that the signal is piecewise constant.** This explains why MC is better than TV. This fact has been confirmed numerically in previous works [8, 9]. Since $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$, the linearity of U can be directly obtained on Ω.

Based on this linearity, we can prove following result

Theorem 2. *When $n \leq 7$, $\mathcal{S} = \{U|\mathcal{E}_H(U) = 0\}$ is convex.*

Proof. According Theorem 1, for any $U_1, U_2 \in \mathcal{S}$ and $n \leq 7$, U_1 and U_2 must be a piecewise linear function. Let $U_\lambda = \lambda U_1 + (1-\lambda)U_2$ for $0 \leq \lambda \leq 1$ denote a linear combination of U_1 and U_2. Sine U_1 and U_2 are linear functions, U_λ is also a linear function. We need to prove that $U_\lambda \in \mathcal{S}$.

Let's take arbitrary two imaging regions Ω_1 and Ω_2 for U_1 and U_2, respectively. There are three possible cases:

- ⋆ 1) If $\Omega_1 \cap \Omega_2 = \emptyset$, then U_λ has a linear form on $\Omega_1 \cup \Omega_2$. Clearly, $\mathcal{E}_H(U_\lambda) = 0$, thus $U_\lambda \in \mathcal{S}$.
- ⋆ 2) $\Omega_1 \cap \Omega_2 \neq \emptyset$ is a general case. We have proved the case for $\Omega_1 \cup \Omega_2 \backslash \{\Omega_1 \cap \Omega_2\}$ in case 1. Therefore we only need to prove the theorem on $\Omega_1 \cap \Omega_2$(case 3).
- ⋆ 3) When $\Omega_1 = \Omega_2$, if $U_1(\partial\Omega_1) = U_2(\partial\Omega_2)$, then $U_1 = U_2$. Therefore, $U_\lambda = U_1$ and $U_\lambda \in \mathcal{S}$. Otherwise, U_1 and U_2 do not consist on the boundary. Since $U_\lambda \in \mathcal{S}$ on $\Omega_1 \backslash \partial\Omega_1$, we only need to prove $\int\limits_{\vec{x}\in\partial\Omega_1} |H(U)| \mathrm{d}\vec{x} = 0$ for the boundary. In fact, from Lebesgue integration point of view, when $|H(U)| < +\infty$, we only need to prove $\int\limits_{\vec{x}\in\partial\Omega_1} \mathrm{d}\vec{x} = 0$, which is well-known in mathematics and also a good assumption in signal processing.

In summary, $U_\lambda \in \mathcal{S}$ and thus $\mathcal{S}$ is convex. □

3. BERNSTEIN FILTER

Thanks to this convexity property from theorem 2, the global optimal solution is unique if the data fitting energy $\mathcal{E}_{\Phi_0}$ is also convex. Therefore, the result from different solvers must converge to the same global optimal solution.

Meanwhile, theorem 1 is a strong result in the sense that the exact solutions for the mean curvature regularization must have linear forms. Therefore, we can construct a filter solver to minimize mean curvature regularizations based on this theorem. We name this filter as Bernstein Filter (BF).

In this filter, we use all possible linear forms (that are minimal surfaces) to approximate the data and choose the minimal change to update the current estimation. Since the minimal surfaces have been used, Bernstein filter is more efficient in minimizing mean curvature, compared with traditional solvers, as shown in the experiment section.

3.1. All possible planes $\mathcal{P}_k$

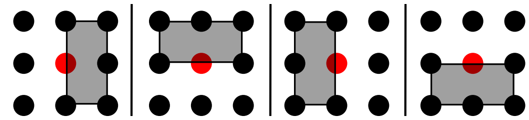


Fig. 2. Four planes $\mathcal{P}_k$ in a 3×3 local window. The plane is determined by the five black points $\{x_i, y_j, U(x_i, y_j)\}$ on the shaded rectangle. The red dot indicates position (x_i, y_j). Thanks to Theorem 1, $\exists k$ such that $(x_i, y_j, U(x_i, y_j)) \in \mathcal{P}_k$.

Let's consider the discrete digital image $I(x_i, y_j)$, where $1 \leq i \leq M$ and $1 \leq j \leq N$ are pixel index. We want to estimate $U(x_i, y_j)$ such that $\{(x_i, y_j, U(x_i, y_j))\}$ lives on a minimal surface. According to Theorem 1, $(x_i, y_j, U(x_i, y_j))$ and its some neighbors must have a linear form. In a local 3×3 window, we can find all possible such linear forms or planes, as illustrated in Fig. 2. We use $\mathcal{P}$ denote these planes and use $\mathcal{P}_k$ to indicate the kth plane in $\mathcal{P}$.

According Theorem 1, we need to update current $U^t(x_i, y_j)$ to a new $U^{t+1}(x_i, y_j)$, such that $\exists k, (x_i, y_j, U^{t+1}(x_i, y_j)) \in \mathcal{P}_k$. First, we compute the signed distance d_k between current $U(x_i, y_j)$ and $\mathcal{P}_k$, for $k = 1, \cdots, 4$. Then, we choose one of d_k to update the current estimation $U(x_i, y_j)$.

3.2. Signed Distance from $U(x_i, y_j)$ to $\mathcal{P}_k$

Let's take the left case in Fig. 2 as an example. First, we take the (x_i, y_j) (the red dot) as origin and use U_0 to denote $U(x_i, y_j)$. In this local coordinate, we have the five points $[1, 1, U_1]$, $[1, 0, U_2]$, $[1, -1, U_3]$, $[0, -1, U_4]$ and $[0, 1, U_5]$, where $U_1 = U(x_{i+1}, y_{j+1})$, $U_2 = U(x_{i+1}, y_j)$, $U_3 = U(x_{i+1}, y_{j-1})$, $U_4 = U(x_i, y_{j-1})$ and $U_5 = U(x_i, y_{j+1})$. Since $\mathcal{P}_k$ has linear form $U = C_2\hat{x} + C_1\hat{y} + C_0$, where $\hat{x}$ and $\hat{y}$ denote the local coordinate, the parameters C_2, C_1 and C_0 can be found by performing a linear regression on this half-window and thus by minimizing

$$T(C_2, C_1, C_0) = \sum_{i=1}^{5} (C_2\hat{x}_i + C_1\hat{y}_i + C_0 - U_i)^2 \,. \quad (10)$$

Letting $\frac{\partial T(C_2, C_1, C_0)}{\partial C_{\{2,1,0\}}} = 0$, we have

$$\begin{bmatrix} 3 & 0 & 3 \\ 0 & 4 & 0 \\ 3 & 0 & 5 \end{bmatrix} \begin{bmatrix} C_2 \\ C_1 \\ C_0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{3} U_i \\ U_1 + U_5 - U_3 - U_4 \\ \sum_{i=1}^{5} U_i \end{bmatrix} . \quad (11)$$

Clearly, it has an analytical solution

$$\begin{bmatrix} C_2 \\ C_1 \\ C_0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{-1}{2} & \frac{-1}{2} \\ \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \end{bmatrix}. \quad (12)$$

Since (x_i, y_j) is the origin (0,0) in the local coordinate, we have the new estimation $U^{t+1}(x_i, y_j) = C_0$, which only needs one plus and one division operations from current estimation ($\frac{U_4^t+U_5^t}{2}$). Then the signed projection distance is

$$d_k = U_0^{t+1} - U_0^t = C_0 - U_0^t = \frac{U_4^t + U_5^t}{2} - U_0^t. \quad (13)$$

3.3. Bernstein Filter

After computing these $\{d_k\}$, we find the d_m that has minimal absolute value in all $\{d_k\}$ and update $U(x_i, y_j)$ such that $(x_i, y_j, U(x_i, y_j)) \in \mathcal{P}_k$. This is summarized in Algorithm 1.

Algorithm 1 Bernstein Filter

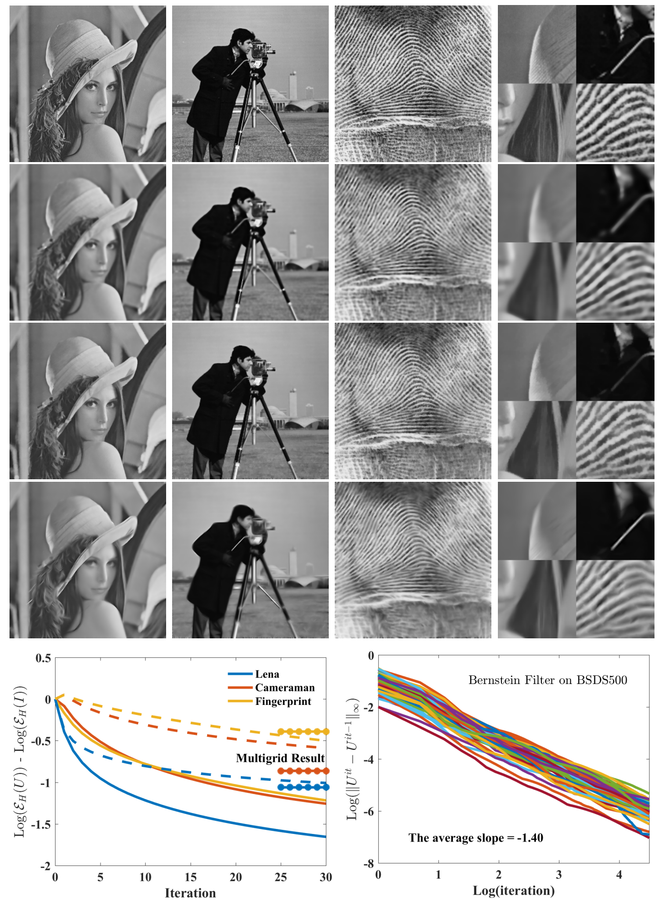
Require: IterationNum, $I(x_i, y_j)$
  $U^0(x_i, y_j) = I(x_i, y_j), t = 0$
  while $t <$ IterationNum **do**
    for i=2:M-1,j=2:N-1 **do**
      $d_1 = \frac{1}{2}[U^t(x_{i-1}, y_j) + U^t(x_{i+1}, y_j)] - U^t(x_i, y_j)$
      $d_2 = \frac{1}{2}[U^t(x_i, y_{j-1}) + U^t(x_i, y_{j+1})] - U^t(x_i, y_j)$
      find d_m such that $|d_m| = \min_{k=1,2}\{|d_k|\}$
      $U^{t+1}(x_i, y_j) = U^t(x_i, y_j) + d_m$
    end for
    $t = t + 1$
  end while
Ensure: $U(x_i, y_j)$

solver (language)	Multigrid (Matlab)	BF (Matlab)	MC filter (C++)	BF (C++)
Lena	183	6.3	0.035	0.025
Cameraman	648	6.5	0.035	0.025
Fingerprint	587	6.4	0.035	0.025

Table 1. running time in seconds (images with 512×512 resolution). The iteration is 30 for MC filter and our filter.

4. EXPERIMENTS

Bernstein filter is at least **two orders of magnitude faster** than multigrid solver [13], which is the fastest so far. In contrast, one iteration of Augmented Lagrangian Method on 512×512 image usually takes few hundred seconds [10]. The running time comparison is summarized in Tab. 1. The results in Fig 3 along with the $\mathcal{E}_H(U)$ profiles indicate that our filter



(a) Energy rate in Log scale (b) convergence plot

Fig. 3. From top to bottom: original images and zoom in patches; results from Multigrid solver [13]; results from MC filter [4] with 30 iterations; results from our filter with 30 iterations. a)The energy rate $\frac{\mathcal{E}_H(U^{it})}{\mathcal{E}_H(I)}$ in Log scale, BF (solid lines), [4] (dash lines) and [13] (solid line with disk, only shows the converged state). b)The numerical convergence rates of our filter on 500 images.

indeed minimizes mean curvature. We also benchmark our approach on Berkeley Segmentation Data Set (BSDS500) and the numerical convergence rate is shown in Fig. 3(b). Source code is at http://github.com/YuanhaoGong/CurvatureFilter

5. CONCLUSION

In this paper, we have proved the convexity of mean curvature regularization. We have shown that Bernstein filter can efficiently minimize mean curvature regularization. Our experiments show that Bernstein filter is at least two orders of magnitude faster than traditional solvers. We believe that Bernstein filter can be extended into higher dimension and used in a large range of image processing problems in the future.

6. REFERENCES

[1] Yuanhao Gong, Qicong Wang, Chenhui Yang, Yahui Gao, and Cuihua Li, "Symmetry detection for multi-object using local polar coordinate," *Lecture Notes in Computer Science*, vol. 5702, pp. 277, 2009.

[2] Yuanhao Gong and Ivo F. Sbalzarini, "Image enhancement by gradient distribution specification," in *In Proc. Workshop "Emerging Topics in Image Enhancement and Restoration", 12th Asian Conference on Computer Vision*, Singapore, Nov 2014, pp. w7–p3.

[3] Yuanhao Gong and Ivo F. Sbalzarini, "Gradient distribution priors for biomedical image processing," *CoRR*, vol. abs/1408.3300, August 2014.

[4] Yuanhao Gong, *Spectrally regularized surfaces*, Ph.D. thesis, ETH Zurich, Nr. 22616, 2015, http://dx.doi.org/10.3929/ethz-a-010438292.

[5] Y. Gong and I.F. Sbalzarini, "A natural-scene gradient distribution prior for light-microscopy image processing," *Selected Topics in Signal Processing, IEEE Journal of*, vol. PP, no. 99, pp. 1–1, Dec 2015.

[6] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.

[7] Adel I. El-Fallah and Gary E. Ford, "Mean curvature evolution and surface area scaling in image filtering," *IEEE Trans. Image Proc.*, vol. 6, no. 5, pp. 750–753, 1997.

[8] H. Zhu, H. Shu, J. Zhou, X. Bao, and L. Luo, "Bayesian algorithms for PET image reconstruction with mean curvature and Gauss curvature diffusion regularizations," *Computers in Biology and Medicine*, vol. 37, no. 6, pp. 793–804, 2007.

[9] W. Zhu and T. Chan, "Image denoising using mean curvature of image surface," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 1–32, 2012.

[10] Wei Zhu, Xue-Cheng Tai, and Tony Chan, "Augmented Lagrangian method for a mean curvature based image denoising model," *Inverse Probl. Imaging*, vol. 7, no. 4, pp. 1409–1432, November 2013.

[11] D. Firsov and S. H. Lui, "Domain decomposition methods-in image denoising using gaussian curvature," *J. Comput. Appl. Math.*, vol. 193, no. 2, pp. 460–473, September 2006.

[12] Yuanhao Gong and Ivo F. Sbalzarini, "Local weighted Gaussian curvature for image processing," *Intl. Conf. Image Proc. (ICIP)*, pp. 534–538, September 2013.

[13] Carlos Brito-Loeza and Ke Chen, "Multigrid algorithm for high order denoising," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 363–389, 2010.

[14] Michael Spivak, "A comprehensive introduction to differential geometry, vol. 2," *I (Boston, Mass., 1970)*, 1970.

[15] Charles Fox, *An introduction to the calculus of variations*, Courier Corporation, 1950.

[16] Howard Jenkins and James Serrin, "Variational problems of minimal surface type ii. boundary value problems for the minimal surface equation," *Archive for Rational Mechanics and Analysis*, vol. 21, no. 4, pp. 321–342, 1966.

[17] Jean Baptiste Marie Meusnier, *Mémoire sur la courbure des surfaces*, De l'Imprimerie Royale, 1785.

[18] Eugenio Calabi, "Examples of bernstein problems for some nonlinear equations," in *Proc. Symp. Pure Math*, 1970, vol. 15, pp. 223–230.