# Learning Variation via Optimal Control

CVBIOUC

Introduction
Numerical Methods and Applications in Total Variation Image Restoration
L-PDE
LVOC
New Insight
LVOC
$\epsilon$-Constrained Method
Other Variation or Functional Method
Active Contour
...
Experimental Results
Pre-analysis
Results

# Contents


The total variation of an image $u : \Omega \to \mathbb{R}$ is defined as:

$$\int_{\Omega} |\nabla u| \, d\boldsymbol{x} \tag{1}$$

Deblurring

Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60:259–268


The total variation of an image $u : \Omega \to \mathbb{R}$ is defined as:

$$\int_{\Omega} |\nabla u| \, d\boldsymbol{x}$$

Inpainting

Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60:259–268


The total variation of an image $u : \Omega \to \mathbb{R}$ is defined as:

$$\int_{\Omega} |\nabla u| d\boldsymbol{x}$$

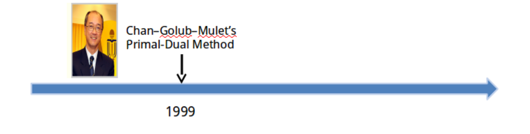
Decomposition

Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60:259–268


The unconstrained total variation based image restoration model reads:

$$\min_{u} \int_{\Omega} \frac{1}{2}(u-f)^2\, d\boldsymbol{x} + \beta \int_{\Omega} |\nabla u|\, d\boldsymbol{x} \tag{2}$$

Acar A, Vogel C (1994) Analysis of bounded variation penalty methods for ill-posed problems. Inverse Probl 10(6):1217–1229



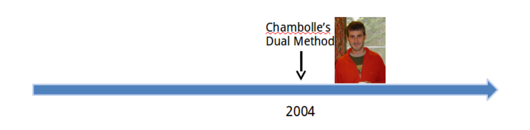

The corresponding Euler-Lagrange equation of (2):

$$K^TKu - K^Tf - \beta div(\frac{\nabla u}{|\nabla u|}) = 0 \tag{3}$$

With the help of an auxiliary variable $\boldsymbol{p} = \frac{\nabla u}{\sqrt{|\nabla u|^2+\epsilon}}$, the singularity will vanish.

$$\begin{aligned} \nabla u - \boldsymbol{p}\sqrt{|\nabla u|^2+\epsilon} &= 0 \\ K^TKu - K^Tf - \beta div\boldsymbol{p} &= 0 \end{aligned} \tag{4}$$

Chan T, Golub G, Mulet P (1999) A nonlinear primal-dual method for total variation-based image restoration. SIAM J Sci Comp 20:1964-1977




A pur dual method which is equivalent to the ROF model avoids the singularity of $|\nabla u| = 0$.

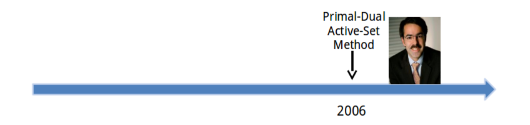
$$\sup_{|\boldsymbol{p}|\leqslant 1}\{-\frac{\lambda^2}{2}\int_\Omega |K^{-T} div\boldsymbol{p} - \frac{f}{\lambda}|^2 d\boldsymbol{x}\} \tag{5}$$

Semi-implicit scheme:

$$\boldsymbol{p}^{n+1} = \frac{\boldsymbol{p}^n + \tau H(\boldsymbol{p}^n)}{\boldsymbol{p}^n + \tau|H(\boldsymbol{p}^n)|} \tag{6}$$

where $H(\boldsymbol{p}) := \nabla[(K^T K)^{-1} div\boldsymbol{p} - \frac{1}{\lambda}K^{-1}f]$

Chambolle A (2004) An algorithm for total variation minimization and applications. J Math Imaging Vis 20:89–97




Consider the general quadratic problem:

$$\min_{y, y \leqslant \phi} < y, Ay > - < f, y > \tag{7}$$

KKT conditions:

$$\begin{aligned} Ay + \lambda &= f \\ \lambda \odot (\phi - y) &= 0 \\ \lambda &\geqslant 0 \\ \phi - y &\geqslant 0 \end{aligned} \tag{8}$$

---

Hintermüller M, Stadler G (2006) A primal-dual algorithm for TV-based inf-convolution-type image restoration. SIAM J Sci Comput 28:1–23


The non-negative constrained TV model:

$$\min_{u, u \geqslant 0} \int_{\Omega} \frac{1}{2}(Ku - f)^2 \, d\boldsymbol{x} + \beta \int_{\Omega} |\nabla u| \, d\boldsymbol{x} \tag{9}$$

The equivalent system:

$$\begin{aligned} \nabla u - \boldsymbol{p}\sqrt{|\nabla u|^2 + \epsilon} &= 0 \\ (K^T K + \alpha I)u - K^T f - \beta \, div p - \lambda &= 0 \\ \lambda - \max\{0, \lambda - cu\} &= 0 \end{aligned} \tag{10}$$

Krishnan D, Lin P, Yip A (2007) A primal-dual active-set method for non-negativity constrained total variation deblurring problems. IEEE Trans Image Process 16(11):2766–2777


Associated with the Fisher-Burmeister function $\phi$, the KKT system is shown as follows.

$$\begin{aligned} \mu \boldsymbol{p} - H(\boldsymbol{p}) &= 0 \\ \phi(\mu, 1 - |\boldsymbol{p}|^2) &= 0 \end{aligned} \tag{11}$$

Ng M, Qi L, Tang Y, Huang Y (2007) On semismooth Newton' s methods for total variation minimization. J Math Imaging Vis 27(3):265–276


The primal-dual formulation:

$$F(u, \boldsymbol{p}) := \frac{1}{2} \int_{\Omega} (Ku - f)^2 \, d\boldsymbol{x} + \beta \int_{\Omega} u \, div \boldsymbol{p} \, d\boldsymbol{x} \tag{12}$$

Two subproblems:

$$\sup_{|\boldsymbol{p}| \leqslant 1} F(u, \boldsymbol{p}) \quad \inf_{u} F(u, \boldsymbol{p}) \tag{13}$$

Zhu M, Chan T (2008) An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report, 08–34

Numerical Methods:

- Dual and Primal-Dual Methods
- Bregman Iteration
- Graph Cut Methods
- Splitting Methods
- Quadratic Programming
- Second-Order Cone Programming
- Majorization-Minimization

---

Scherzer O (2011) Handbook of mathematical methods in imaging. Springer, Vienna

## Optimal Control Governed by Evolutionary PDEs:

minimize $J(f, u)$, where $u \in \mathcal{U}$ controls $f$ via the following PDE:

$$\begin{cases} f_t = L(u, \boldsymbol{\beta}), & (x, t) \in \Omega \\ f = 0, & (x, t) \in \Gamma \\ f|_{t=0} = f_0 & x \in \Omega \end{cases} \tag{14}$$

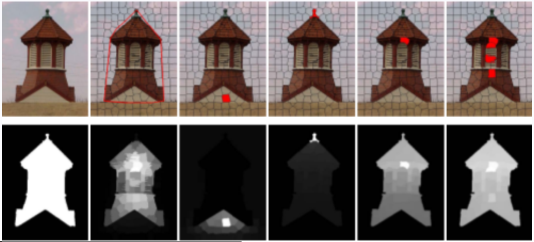
where $L(u, \boldsymbol{\beta}) = \kappa(u) + F(u, \boldsymbol{\beta})$.

Lin Z, Zhang W, Tang X (2008) Learning partial differential equations for computer vision. Technical report, Microsoft Research, MSR-TR-2008-189

# Adaptive partial differential equation learning for visual saliency detection

Saliency diffusion is formulated as an evolutionary PDE:

$$\frac{\partial f(\boldsymbol{p}, t)}{\partial t} = div(K_{\boldsymbol{p}} \nabla f(\boldsymbol{p})) + \lambda(f(\boldsymbol{p}) - g(\boldsymbol{p})), \boldsymbol{p} \in \mathcal{S} \tag{15}$$



Liu R, Cao J, Lin Z, Shan S (2014) Adaptive partial differential equation learning for visual saliency detection. In Proceedings of IEEE conference on computer vision and pattern recognition

The non-negative constrained TV model:

$$\min_{u, u \geqslant 0} \int_\Omega \frac{1}{2}(Ku - f)^2 \, d\boldsymbol{x} + \beta \int_\Omega |\nabla u| \, d\boldsymbol{x}$$

The equivalent system:

$$\begin{aligned} \nabla u - p\sqrt{|\nabla u|^2 + \epsilon} &= 0 \\ K^T K u - K^T f - \beta \, divp - \lambda &= 0 \\ \lambda - \max\{0, \lambda - cu\} &= 0 \end{aligned}$$

**What if beta is adaptive?**

The non-negative constrained TV model:

$$\min_{u, u \geqslant 0} \int_\Omega \frac{1}{2}(Ku - f)^2 \, d\boldsymbol{x} + \beta \int_\Omega |\nabla u| \, d\boldsymbol{x}$$

The equivalent system:

$$\begin{aligned} \nabla u - p\sqrt{|\nabla u|^2 + \epsilon} &= 0 \\ K^T K u - K^T f - \beta \, div p - \lambda &= 0 \\ \lambda - \max\{0, \lambda - cu\} &= 0 \end{aligned}$$

**How?**

The non-negative constrained TV model:

$$\min_{u, u \geqslant 0} \int_\Omega \frac{1}{2}(Ku - f)^2 \, d\boldsymbol{x} + \boldsymbol{\beta} \int_\Omega |\nabla u| \, d\boldsymbol{x}$$

The semi-smooth Newton's update:

$$\begin{bmatrix} |\nabla u^k|_\epsilon & -(I - \frac{p^k(\nabla u^k)^T}{|\nabla u^k|_\epsilon})\nabla & 0 \\ -\boldsymbol{\beta^k} div & K^T K & -I \\ 0 & \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} F_1^k \\ F_2^k \\ F_3^k \end{bmatrix}$$

$\beta^k$ reflects the difference between $u^k$ and the groudtruth.

Train it!

The non-negative constrained TV model:

$$\min_{u, u \geqslant 0} \int_\Omega \frac{1}{2}(Ku - f)^2 \, d\boldsymbol{x} + \boldsymbol{\beta} \int_\Omega |\nabla u| \, d\boldsymbol{x}$$

The semi-smooth Newton's update:

$$\begin{bmatrix} |\nabla u^k|_\epsilon & -(I - \frac{p^k (\nabla u^k)^T}{|\nabla u^k|_\epsilon})\nabla & 0 \\ -\boldsymbol{\beta^k} div & K^T K & -I \\ 0 & \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} F_1^k \\ F_2^k \\ F_3^k \end{bmatrix}$$

$\beta^k$ reflects the difference between $u^k$ and the groudtruth.

The non-negative constrained variation:

$$\min_{u, u \geqslant 0} \int_{\Omega} [\frac{1}{2}(Ku - f)^2 + V(\boldsymbol{\beta}, u)] d\boldsymbol{x} \tag{16}$$

where $V(\boldsymbol{\beta}, u) = \sum_{i=1}^{3} \beta_i v_i(u)$.

- $v_2(u) = |u|_{TV}$
- $v_1(u) = |u|^2$
- $v_3(u) = |\nabla u|^2$

Optimal control:

$$\min\{\frac{1}{2}\int_{\Omega}(u-\hat{f})^2\,d\boldsymbol{x}+\frac{1}{2}\sum_{i=1}^{3}a_i\beta_i^2\} \tag{17}$$

Learning variation via optimal control:

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\textstyle\int_{\Omega}(u-\hat{f})^2\,d\boldsymbol{x}+\tfrac{1}{2}\sum_{i=1}^{3}a_i\beta_i^2\} \\ \min_{u,u\geqslant 0} \quad & \{\textstyle\int_{\Omega}[\tfrac{1}{2}(Ku-f)^2+V(\boldsymbol{\beta},u)]\,d\boldsymbol{x}\} \end{aligned} \tag{18}$$

*Problem A:*

$$\begin{aligned} \min_{u,\alpha} \quad & [f(x,u,\alpha),\, G(\alpha)] \\ s.t. \quad & g(x,u,\alpha) \leqslant 0 \end{aligned} \tag{19}$$

*Problem B:*

$$\begin{aligned} \min_{u,\alpha} \quad & f(x,u,\alpha) \\ s.t. \quad & G(\alpha) \leqslant \epsilon \\ & g(x,u,\alpha) \leqslant 0 \end{aligned} \tag{20}$$

## Equivalence Theorem

*Let $\epsilon \geqslant \min G(\alpha)$, let $V^*$ solve Problem $B(\epsilon)$, and asume that, if $V^*$ is not unique, then $V^*$ is an optimal solution of Problem $B(\epsilon)$ with minimal $G(\alpha)$ value. Then $V^*$ solves Problem A.*

---

Haimes Y, Lasdon L, Wismer D (1971) On a Bicriterion Formulation of the Problems of Integrated System Identification and System Optimization. IEEE Transactions on Systems, Man, and Cybernetics 1(3) 296–297

Learning variation via optimal control:

$$\min \quad \{\tfrac{1}{2}\int_\Omega (u-\hat{f})^2 d\boldsymbol{x} + \tfrac{1}{2}\sum_{i=1}^3 a_i\beta_i^2\}$$

$$\min \quad \{\int_\Omega [\tfrac{1}{2}(Ku-f)^2 + V(\boldsymbol{\beta}, u)]\, d\boldsymbol{x}\} \qquad s.t. \quad u \geqslant 0$$

Learning variation via optimal control:

$$
\begin{aligned}
&\min \quad \left\{\frac{1}{2}\int_\Omega (u-\hat{f})^2 d\boldsymbol{x} + \frac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2\right\} \\
&\min \quad \left\{\int_\Omega \left[\frac{1}{2}(Ku-f)^2 + V(\boldsymbol{\beta}, u)\right] d\boldsymbol{x}\right\} \quad s.t. \quad u \geqslant 0
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
&\min \quad \left\{\frac{1}{2}\int_\Omega (u-\hat{f})^2 d\boldsymbol{x} + \frac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2\right\} \\
&s.t. \quad \int_\Omega \left[\frac{1}{2}(Ku-f)^2 + V(\boldsymbol{\beta}, u)\right] d\boldsymbol{x} \leqslant \epsilon, u \geqslant 0
\end{aligned}
$$

Learning variation via optimal control:

$$
\begin{array}{ll}
\min & \{\frac{1}{2}\int_{\Omega}(u-\hat{f})^2 d\boldsymbol{x} + \frac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2\} \\
\min & \{\int_{\Omega}[\frac{1}{2}(Ku-f)^2 + V(\boldsymbol{\beta}, u)]\, d\boldsymbol{x}\} \quad s.t. \quad u \geqslant 0
\end{array}
$$

$$\downarrow$$

$$
\begin{array}{ll}
\min & \{\frac{1}{2}\int_{\Omega}(u-\hat{f})^2 d\boldsymbol{x} + \frac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2\} \\
s.t. & \int_{\Omega}[\frac{1}{2}(Ku-f)^2 + V(\boldsymbol{\beta}, u)]\, d\boldsymbol{x} \leqslant \epsilon, u \geqslant 0
\end{array}
$$

$$\downarrow$$

$$
\begin{array}{ll}
\min & \{\frac{1}{2}\int_{\Omega}(u-\hat{f})^2 d\boldsymbol{x} + \frac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2\} \\
s.t. & K^T Ku + Lag(V(\boldsymbol{\beta}, u, p)) = 0 \\
& |\nabla u|_{\epsilon} p - |\nabla u| = 0 \\
& u, |p|^2 - 1 \geqslant 0
\end{array}
\tag{21}
$$

The corresponding minimization:

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\int_\Omega (u-\hat{f})^2 d\boldsymbol{x} + \tfrac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2 \\ & + \int_\Omega \phi_1(K^T K u + Lag(V(\boldsymbol{\beta}, u, p)))\, d\boldsymbol{x} \\ & + \int_\Omega \phi_2(|\nabla u|_\epsilon p - \nabla u)\, d\boldsymbol{x} \\ & + < u, \lambda_1 > + < |p|^2 - 1, \lambda > \end{aligned} \tag{22}$$

The KKT conditions for the problem(21):

$$\begin{aligned} (u-\hat{f}) + Lag(Lag(V)) + \nabla(\phi_2 p^2) - \nabla\phi_2 - \lambda + \lambda_2 \odot p &= 0 \\ K^T K u + Lag(V(\boldsymbol{\beta}, u, p)) &= 0 \\ |\nabla u|_\epsilon p - \nabla u &= 0 \\ (|p|^2 - 1) \odot \lambda_2 &= 0 \\ u \odot \lambda_1 &= 0 \\ \lambda_1, \lambda_2 &\geqslant 0 \end{aligned} \tag{23}$$

The equivalent minimization:

$$\begin{aligned} \min \quad & \{\tfrac{1}{2}\int_\Omega (u-\hat{f})^2\,d\boldsymbol{x} + \tfrac{1}{2}\sum_{i=1}^{3} a_i\beta_i^2 \\ & + \int_\Omega \phi_1(K^TKu + Lag(\tilde{V}(\boldsymbol{\beta},u,p)))\,d\boldsymbol{x} \end{aligned} \tag{24}$$

The corresponding system:

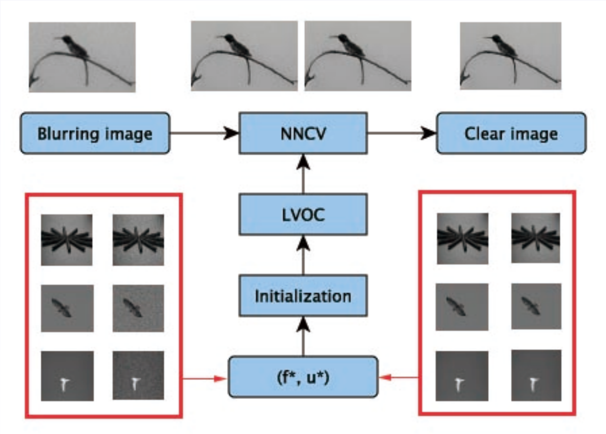
$$(u-\hat{f}) + Lag(Lag(\tilde{V})) = 0 \tag{25}$$

$$K^TKu + Lag(\tilde{V}(\boldsymbol{\beta},u,p)) = 0 \tag{26}$$

$$|\nabla u|_\epsilon p - \nabla u = 0 \tag{27}$$

$$\lambda - \max\{0, \lambda - cu\} = 0 \tag{28}$$

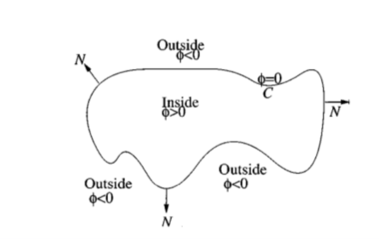
where $\tilde{V} = \beta_1 u - \beta_3 \Delta u - \beta_2\, divp - \lambda - K^Tf$. The left hand sides are denoted by $F_1, F_2, F_3, F_4$ respectively.


Blurring image
NNCV
Clear image
LVOC
Initialization
(f*, u*)

# Active contours model

Snake model:

$$E_{snake} = \int_1^2 \frac{1}{2}[\alpha(s)|\frac{\partial v}{\partial s}|^2 + \beta(s)|\frac{\partial^2 v}{\partial s^2}|^2]ds + \int_1^2 E_{ext}(v(s))ds \tag{29}$$

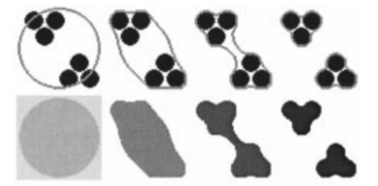



Kass M, Witkin A, Terzopoulos D (1988) Snake: Active contours models. Int. J. comput. Vis., vol. 1, pp. 321-331

# Active contours model

Snake model:

$$E_{snake} = \int_1^2 \frac{1}{2}[\alpha(s)|\frac{\partial v}{\partial s}|^2 + \beta(s)|\frac{\partial^2 v}{\partial s^2}|^2]ds + \int_1^2 E_{ext}(v(s))ds$$



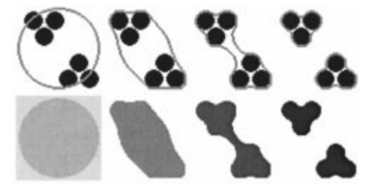
Kass M, Witkin A, Terzopoulos D (1988) Snake: Active contours models. Int. J. comput. Vis., vol. 1, pp. 321-331

# Active contours model

**Why not associate optimal control with Snake model?**

Snake model:

$$E_{snake} = \int_1^2 \frac{1}{2}[\alpha(s)|\frac{\partial v}{\partial s}|^2 + \beta(s)|\frac{\partial^2 v}{\partial s^2}|^2]ds + \int_1^2 E_{ext}(v(s))ds$$



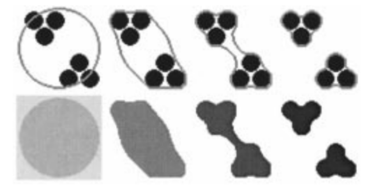
Kass M, Witkin A, Terzopoulos D (1988) Snake: Active contours models. Int. J. comput. Vis., vol. 1, pp. 321-331

# Active contours model

**Future Work!**

Snake model:

$$E_{snake} = \int_1^2 \frac{1}{2}[\alpha(s)|\frac{\partial v}{\partial s}|^2 + \beta(s)|\frac{\partial^2 v}{\partial s^2}|^2]ds + \int_1^2 E_{ext}(v(s))ds$$



Kass M, Witkin A, Terzopoulos D (1988) Snake: Active contours models. Int. J. comput. Vis., vol. 1, pp. 321-331

Thanks!

此处将列举其他能量泛函方法, 敬请期待!

此处将列举 $V(u, \boldsymbol{\beta})$ 的处理效果, 敬请期待!

此处将列举 LVOC 的处理效果 (加大训练集), 并进一步分析, 包括线性拟合最优参数和评价指标的内容, 敬请期待!