

## Équations aux dérivées partielles

# Théorie des semi-groupes pour l'équation de Perona–Malik

Satyanad Kichenassamy

*Université de Reims Champagne-Ardenne, laboratoire de mathématiques, UMR 6056, Moulin de la Housse,  
B.P. 1039, 51687 Reims cedex 2, France*

Reçu et accepté le 1<sup>er</sup> janvier 2007

Disponible sur Internet le 1<sup>er</sup> février 2007

Présenté par Haïm Brezis

---

### Résumé

On montre que les observations empiriques sur l'équation de Perona–Malik, qui intervient en particulier en traitement d'images et en océanographie, peuvent s'expliquer si l'on considère que cette équation, bien que non parabolique, définit un semi-groupe à croissance exponentielle correspondant à une perturbation lipschitzienne du flot de gradient d'une fonctionnelle convexe, dans un espace à discontinuités fixées. La stratification automatique observée numériquement correspond à l'évolution des sauts engendrée par ce flot non linéaire. **Pour citer cet article :** *S. Kichenassamy, C. R. Acad. Sci. Paris, Ser. I 344 (2007).*

© 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

### Abstract

**Semi-group theory for the Perona–Malik equation.** We show that numerical observations on the Perona–Malik equation, which arises in particular in image processing and in oceanography, may be accounted for by associating to this formally ill-posed PDE the gradient flow with exponential growth generated by a Lipschitz perturbation of a convex functional, in the framework of semi-group theory. The automatic stratification observed numerically is accounted for using an image model with fixed discontinuities, which are selectively attenuated, rather than smoothed, by this nonlinear flow. **To cite this article :** *S. Kichenassamy, C. R. Acad. Sci. Paris, Ser. I 344 (2007).*

© 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

---

### Abridged English version

An equation of the form

$$u_t = \operatorname{div} R(\nabla u) - \alpha(u - f), \quad u(x, y, 0) = f(x, y), \quad (1)$$

where  $\alpha > 0$ ,  $R(\nabla u) = \rho(|\nabla u|^2)\nabla u$ ,  $s \mapsto s\rho(s^2)$  admits a unique positive maximum for  $s = K$ , and tends to zero at infinity, will be called a Perona–Malik equation (PM) [6,3], as modified by Nordström [5], who introduced the 'bias' term  $\alpha(u - f)$ . Two typical examples are:  $\rho = (1 + s^2/K^2)^{-1}$  and  $\rho = \exp(-s^2/2K^2)$ . We consider the initial-value problem with Neumann conditions on a rectangle  $\Omega \subset \mathbb{R}^2$ . This is appropriate for applications in Computer Vision. Although Eq. (1) is not parabolic for gradients greater than  $K$ , straightforward difference schemes derived from it do not exhibit the wild oscillations characteristic of the numerical solution of backward diffusion problems.

---

Adresse e-mail : [satyanad.kichenassamy@univ-reims.fr](mailto:satyanad.kichenassamy@univ-reims.fr).

We account for this paradox by suggesting that the numerical computations imply four implicit assumptions, which may be incorporated into the definition of a functional  $\mathcal{F}$  on a set  $\Omega \setminus \Gamma$ , of which (1) is the gradient flow in the sense of semi-group theory. The set  $\Gamma$  is a rough estimate of the set of edges. Edge location is fixed by design.

The mechanism of the nonlinear evolution, the existence of which follows from semi-group theory, may be described as a pruning procedure, whereby the jumps across the various components of  $\Gamma$  decrease by a given rule, while their location remains fixed. The new elements in this approach, which are motivated by numerical experiments, are as follows:

- (i) We attribute to each edge a natural thickness  $h$ , which prevents the creation of an infinite number of singularities. Difficulties with lower semi-continuity are thereby taken care of. In real images,  $h$  is equal to one pixel.
- (ii) The simplification of the image is not achieved by displacement of the edges, nor by removing components of  $\Gamma$ , but merely by letting the jump across certain components drop below a threshold, here  $hK$ . After this point, numerical schemes cease to identify such components as edges.
- (iii) We take into account not only the arclength of the edge set, but also the size of each jump, which we interpret as an ‘edge strength’. Intuitively, if the gradient remains smaller than  $K$  off  $\Gamma$ , our generalized solutions solve a nonlinear Neumann problem with boundary condition  $\partial u / \partial n = S(j(u)/h)$  across  $\Gamma$ , where  $S(s)$  is the solution of  $S\rho(S^2) = s\rho(s^2)$  with smallest absolute value, and  $j(u)$  is the absolute value of the jump of  $u$ . Thus, for  $\rho(s) = (1 + s^2/K^2)^{-1}$ , we have  $S(s) = \min(s, K^2/s)$  for  $s \geq 0$ .

Difficulties with earlier approaches are accounted for by the properties of  $\mathcal{F}$ : (i) No convergence should be expected as the mesh size tends to zero for fixed  $K$ , because the nonlinear boundary condition depends on  $h$ . (ii) The solution operator is not a contraction: it has exponential growth in time. Therefore, nearby data give nearby solutions for small times only. This is not an indication of ill-posedness. It only suggests to use a small number of iterations in the numerical scheme, in accordance with computations. In practice, it is convenient to combine gradient descent of PM with smoothing steps, see Section 3. Also, the semi-group formalism yields a construction of solutions with data with limited regularity, in any number of space dimensions, which do not seem to be accessible to other methods.

**Remark 1.** Among different, related issues considered in the literature, mostly in one dimension, we may mention: (i) compactness properties of difference schemes along a subsequence as mesh size  $h \rightarrow 0$ ; and (ii) limiting properties as  $h \rightarrow 0$ , assuming  $K$  and  $h$  are related in such a way that  $K$  tends to infinity. We do not dwell on these or other possibilities, since they are not compatible with some or all of hypotheses (i)–(iii) above.

**Remark 2.** An explicit, centered scheme in one dimension decreases total variation [4]. However, this is not sufficient to account for the evolution of jumps discussed here. Note also that the gradient satisfies a maximum principle in one dimension if  $\alpha = 0$ . Examples seem to show that this is not true in general; for this reason, we assume that, if the gradient becomes larger than  $K$  off  $\Gamma$ , evolution continues with a convex nonlinearity at these points. Otherwise, the flow formally solves PM. In this manner, all the cases in which computation is successful are encompassed by the present framework.

## 1. Introduction

On appelle équation de Perona–Malik (PM) [6,3] toute équation de la forme

$$u_t = \operatorname{div} R(\nabla u) - \alpha(u - f), \quad u(x, y, 0) = f(x, y), \quad (2)$$

où  $\alpha > 0$ ,  $R(\nabla u) = \rho(|\nabla u|^2)\nabla u$ ,  $s \mapsto s\rho(s^2)$  est régulière et admet un unique maximum positif pour  $s = K$ , et tend vers zéro à l’infini; le terme  $\alpha(u - f)$  a été introduit par Nordström [5]; on supposera  $\rho(0) = 1$ . Parmi les choix usuels de  $\rho$ , on peut citer:  $\rho_1 = (1 + s^2/K^2)^{-1}$  et  $\rho_2 = \exp(-s^2/2K^2)$ . On considérera le problème de valeur initiale avec données de Neumann sur un rectangle  $\Omega \subset \mathbb{R}^2$ , comme il est d’usage en vision par ordinateur, application qui a suscité le plus grand nombre d’études de cette équation. Par abus de notation, on posera  $R(s) = s\rho(s^2)$ .

**Remarque 1.** Le calcul numérique s’effectue habituellement en discrétisant plutôt

$$u_t = R(u_x)_x + R(u_y)_y. \quad (3)$$

Les considérations ci-dessous s'appliquent à cette équation avec les modifications évidentes. Ce schéma décroît la variation totale en une dimension [4].

**Remarque 2.** Des cas particuliers de cette équation semblent avoir joué un rôle dans le développement de la méthode directe du Calcul des Variations. Voir par exemple [2], où l'on montre qu'une fonctionnelle telle que  $\int_0^1 [1 - (1 + u'^2)^{-1}] dx$  n'est pas s.c.i. sur un espace de fonctions sans discontinuités.

**Remarque 3.** Cette équation apparaît naturellement dans les applications où l'on s'intéresse à la formation spontanée de stratifications—évolution apparemment contradictoire avec l'effet que l'on attendrait d'une diffusion. Pour le problème de la stratification de la salinité en océanographie, voir [7]. On explique ici ce comportement en donnant un cadre fonctionnel dans lequel une telle stratification est encouragée.

On rappelle que, bien que (2) ne soit pas parabolique pour des gradients supérieurs à  $K$ , des schémas discrétisés basés sur cette équation ne présentent pas les grandes oscillations caractéristiques de la solution numérique d'équations paraboliques rétrogrades. Ce paradoxe peut être expliqué [3] en remarquant que, si les estimations de régularité excluent les solutions faibles peu régulières pour lesquelles le gradient est supérieur à  $K$  sur un ouvert, rien n'interdit qu'il existe des solutions pour lesquelles cet ensemble se réduirait à un ensemble de courbes, sur lesquelles PM devrait être remplacée par des conditions aux limites non linéaires. Cela revient à considérer que les calculs numériques contiennent quatre hypothèses implicites :

- (a) il existe un ensemble de courbes lipschitziennes  $\Gamma \subset \Omega$  en dehors duquel le gradient de  $f$  est inférieur à  $K$  ;
- (b)  $u$  admet des discontinuités de première espèce sur  $\Gamma$  ;
- (c) la composante normale du gradient est continue à la traversée de  $\Gamma$  ;
- (d) la composante normale de  $R$  sur  $\Gamma$ , soit  $R(\nabla u) \cdot n$ , est égale à  $R(j(u)/h)$ , où  $j(u)$  désigne la valeur absolue du saut de  $u$ , et  $h$  le pas de discrétisation en espace.

On peut alors construire des solutions généralisées de (2) dans une classe de fonctions possédant les propriétés (a) et (b) ; les propriétés (c) et (d) fournissent alors deux conditions de Neumann non linéaires sur l'interface  $\Gamma$ . Celle-ci représente une estimation grossière de l'ensemble des bords des objets dans l'image. Au cours de l'évolution, on constate que certaines des discontinuités deviennent inférieures à  $hK$  en valeur absolue, ce qui revient en pratique à considérer que ces discontinuités ne doivent pas être retenues comme des bords d'objets dans l'image. C'est ainsi que l'équation (2) permet une simplification automatique de l'image considérée.

On montre dans cette Note que cette solution est donnée par le semi-groupe engendré par le gradient d'une fonctionnelle  $\mathcal{F}$ , où  $\mathcal{F}[u] + \beta \int_{\Omega} u^2$  est convexe, propre et s.c.i. sur  $L^2(\Omega)$ . Tous les résultats d'existence et d'approximation de la théorie des semi-groupes sont alors disponibles ; en particulier, on obtient l'existence de solutions à données peu régulières, en dimension quelconque, qui ne semblent pas pouvoir être fournies par d'autres méthodes.

Les éléments nouveaux de cette approche sont les suivants :

- (i) On attribue à chaque discontinuité une épaisseur positive  $h$  (en pratique, d'un pixel), ce qui exclut d'emblée les difficultés de manque de semi-continuité.
- (ii) La simplification de l'image n'est pas due au déplacement des lignes de discontinuité de la solution, mais à la modification dynamique de la valeur des sauts à la traversée de lignes fixes : les bords des objets sont définis par la condition que ce saut soit supérieur à  $hK$ .
- (iii) La partie non convexe de la fonctionnelle se résume à une intégrale sur  $\Gamma$  qui fournit, par variation, la condition aux limites non linéaire sur  $\Gamma$ . En imposant que la contribution du complémentaire de  $\Gamma$  à  $\mathcal{F}$  soit convexe, on évite l'apparition de nouvelles discontinuités.

On en tire les conséquences pratiques suivantes :

- (a) L'équation (2) est un flot de gradient.
- (b) Son semi-groupe n'étant pas constitué de contractions, des données proches ne conduisent à des solutions proches que pour des temps limités : le problème, bien posé, peut donc paraître « pratiquement instable ».

- (c) La valeur de  $K$  doit être choisie de sorte que le terme  $|\nabla u|/K$  soit comparable à l'unité dans les régions qu'on veut traiter, sans quoi  $R(\nabla u)$  serait assimilable à 0 ou  $\nabla u$ , selon que le rapport serait très grand ou très petit, et l'équation se réduirait alors pratiquement à  $u_t = 0$ , ou à  $u_t = \Delta u$ . Dans les cas les plus courants, ceci force  $R$  à être assez petit. C'est ce qui explique que les solutions constantes par morceaux soient métastables : elles persistent pendant un temps assez long, avant de disparaître.

**Remarque 4.** Bien que l'équation de Perona–Malik soit formellement l'équation d'Euler d'une fonctionnelle non-convexe du premier ordre, ces méthodes de convexification ne s'appliquent pas : l'enveloppe convexe de la fonction  $F(s)$  ci-dessous est en effet identiquement nulle pour les exemples considérés. Cette fonctionnelle non convexe a été proposée par comme contre-exemple au « Principe de Dirichlet » pour des fonctionnelles générales [2].

**Remarque 5.** Supposons que  $F(s) = F_1(s/K)$ , où  $F_1(s) \rightarrow 1$  lorsque  $s \rightarrow \infty$ . Le terme en  $j(u)$  dans  $\mathcal{F}$  devient  $F_1(j(u)/hK)$  ; par suite, lorsque  $h$  et  $K$  varient de telle sorte que  $hK \rightarrow 0$ ,  $\mathcal{F}$  tend formellement vers une fonctionnelle de type Mumford–Shah. Cependant, cela n'est plus vrai si  $F$  tend vers  $+\infty$  à l'infini. Les nonlinéarités  $\rho_2$  et  $\rho_1$  du début de cette Note illustrent respectivement ces deux situations.

On donne maintenant le résultat principal et une idée de sa démonstration.

## 2. Relation avec la théorie des semi-groupes

Soit  $\tilde{F}(s)$  une fonction convexe à croissance quadratique à l'infini, qui coïncide avec  $F(s) := \int_0^s R(\sigma) d\sigma$  pour  $|s| \leq K$ . On se donne un ensemble  $\Gamma \subset \Omega$ , tel que les conditions suivantes soient valides :

- (i) Il existe deux constantes  $a$  et  $b$  telles que  $G(s) := \tilde{F}(s) - as^2$  et  $H(s) = F(s) + bs^2$  soient convexes sur  $\mathbb{R}$ .  
(ii) La valeur  $j(u)$  du saut de  $u$  à la traversée de  $\Gamma$  définit une application compacte de  $H^1(\Omega \setminus \Gamma)$  dans  $L^2(\Gamma)$ , et

$$(\forall \varepsilon > 0)(\exists c_\varepsilon > 0)(\forall u \in H^1(\Omega \setminus \Gamma)) \quad \int_{\Gamma} (j(u))^2 \leq \varepsilon \int_{\Omega \setminus \Gamma} |\nabla u|^2 + c_\varepsilon \int_{\Omega \setminus \Gamma} u^2.$$

On peut prendre par exemple pour  $\Gamma$  un ensemble fini de courbes lipschitziennes dans  $\Omega$ . La condition (i) peut être quelque peu généralisée, au prix de complications techniques que l'on omettra ici. On considère alors, pour  $\alpha \geq 0$ , la fonctionnelle convexe de domaine  $H^1(\Omega \setminus \Gamma)$ , définie, pour  $u$  dans ce domaine, par

$$\mathcal{F}[u] = \int_{\Omega \setminus \Gamma} \tilde{F}(|\nabla u|) + \alpha(u - f)^2 + \int_{\Gamma} hF(j(u)/h).$$

Pour  $\beta \geq 0$ , on pose

$$\mathcal{F}_\beta[u] = \mathcal{F}[u] + \beta \int_{\Omega \setminus \Gamma} u^2.$$

**Remarque 6.** L'introduction du paramètre  $\alpha$  permet de pénaliser la déviation de la solution par rapport à l'image initiale. L'effet de ce terme sur la convexité de  $\mathcal{F}$  est comparable à celui de  $\beta$ .

**Théorème 2.1.** Si  $0 < \varepsilon \leq ah/b$  et  $\alpha + \beta \geq bc_\varepsilon/h$ , alors  $\mathcal{F}_\beta$  est convexe, propre et s.c.i. sur  $L^2(\Omega \setminus \Gamma)$ .

On en déduit le résultat suivant :

**Théorème 2.2.** L'équation de PM sous la forme

$$u_t + \partial \mathcal{F}(u) \ni 0; \quad u = f \quad \text{for } t = 0$$

admet une unique solution faible globale, dont la norme  $L^2$  croît au plus exponentiellement.

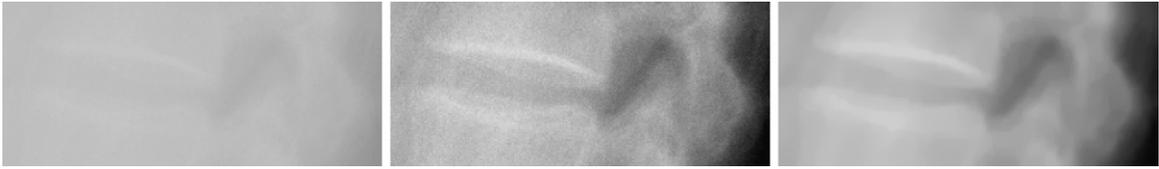


Fig. 1. Image avant traitement (gauche), après ajustement d'histogramme (centre), image traitée (droite).

Pour démontrer le Théorème 2.1, on estime d'abord  $\mathcal{A}[u, v] = \mathcal{F}_\beta[tu + (1-t)v] - t\mathcal{F}_\beta[u] - (1-t)\mathcal{F}_\beta[v]$  pour  $t \in [0, 1]$  en utilisant l'égalité  $(tu + (1-t)v)^2 = tu^2 + (1-t)v^2 - t(1-t)(u-v)^2$ . Tenant compte de la convexité des fonctions  $G$  et  $H$ , on obtient  $\mathcal{A}[u, v] \leq -t(1-t)Q$ , où

$$Q = \int_{\Omega \setminus \Gamma} a |\nabla(u-v)|^2 + (\alpha + \beta)(u-v)^2 - bh^{-1} \int_{\Gamma} (j(u-v))^2.$$

On estime le second terme grâce à la seconde hypothèse. Le choix de  $\varepsilon$  et de  $\beta$  assure que  $Q$  est manifestement positive ou nulle. La convexité de  $\mathcal{F}_\beta$  en résulte. Sa sous-différentielle engendre donc un semi-groupe de contractions.

La sous-différentielle de  $\mathcal{F}$  en est une perturbation lipschitzienne. En appliquant [1, Proposition 3.12, pp. 106–107], on obtient alors le Théorème 2.2.

**Remarque 7.** Une autre idée naturelle consisterait à travailler sur un convexe formé de fonctions de gradient de norme inférieure ou égale à  $K$  sur  $\Omega \setminus \Gamma$ . Mais ceci semble autoriser la formations de zones « plastiques » à l'intérieur de  $\Omega \setminus \Gamma$ , qui ne sont pas observées numériquement. Néanmoins, une telle procédure rentrerait également dans le cadre général de la théorie des semi-groupes [1].

**Remarque 8.** Parmi d'autres tentatives d'interprétation du comportement des schémas numériques pour PM, généralement en une dimension, on peut mentionner l'étude (i) de la compacité des solutions lorsque le paramètre de discrétisation  $\Delta x \rightarrow 0$  et (ii) du comportement des solutions numériques lorsque  $\Delta x \rightarrow 0$  et  $K \rightarrow \infty$ , en supposant une relation entre ces deux paramètres. En particulier, contrairement au présent travail, ces tentatives n'attribuent pas aux lignes de discontinuité une épaisseur finie.

### 3. Exemple

L'équation (2) ayant tendance à préserver des structures de petite taille, il est classique de régulariser au préalable l'image par convolution. Il est à noter que cette opération n'a pas pour objet de rendre le problème bien posé. Dans la Fig. 1, on a appliqué cette méthode à une image radiologique, à 256 niveaux de gris. On prend  $\rho = \rho_1$ . Après ajustement de l'histogramme (au centre), on a appliqué 28 itérations de la discrétisation explicite centrée de (3) avec  $K = 0, 8, 3$  étapes de régularisation ( $K = 1000$ ), puis 33 itérations avec  $K = 0, 8$ . Le résultat final est l'image de droite.

### Références

- [1] H. Brezis, Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert, North-Holland, Amsterdam, 1973.
- [2] R. Courant, Über direkte Methoden in der Variationsrechnung und über verwandte Fragen, Math. Ann. 97 (1) (1927) 711–736.
- [3] S. Kichenassamy, The Perona–Malik paradox, SIAM J. Appl. Math. 57 (5) (1997) 1328–1342.
- [4] S. Kichenassamy, Nonlinear diffusion and hyperbolic smoothing for edge enhancement, in: M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.), Proc. ICAOS'96, in: Lecture Notes in Control and Inform. Sci., vol. 219, Springer-Verlag, Berlin, 1996, pp. 119–124.
- [5] K. Nordström, Biased anisotropic diffusion—a unified regularization and diffusion approach to edge detection, Image and Vision Computing 8 (11) (1990) 318–327.
- [6] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Machine Intell. 12 (1990) 629–639. See also Proc. of IEEE Comp. Soc. Workshop on Comp. Vision, Miami, 1987, pp. 16–22.
- [7] E.S. Posmentier, The generation of salinity finestructure by vertical diffusion, J. Phys. Oceanography 7 (1977) 298–300.