

Evolution de la résolution de problèmes inverses en imagerie

Nelly PUSTELNIK¹ Caroline CHAUX²

¹ENSL, CNRS, Laboratoire de physique, F-69342 Lyon, France

²CNRS, IPAL, Singapour

Résumé – Les problèmes inverses sont au coeur de nombreuses problématiques en traitement du signal et des images. Cet article a pour objectif de retracer les grandes avancées en terme de résolution de problèmes inverses. Nous partons de la formulation bayésienne, puis décrivons de nombreuses pénalisations rencontrées dans la littérature, ainsi que les grandes classes algorithmiques résolvant les problèmes de minimisation associés. Enfin, nous faisons une ouverture aux approches basées apprentissage profond.

Abstract – Inverse problems are central in many signal and image processing areas. This article aims at reviewing the major advances in terms of solving inverse problems. We take as a starting point the Bayesian formulation, and then report on the numerous penalties encountered in the literature, as well as the main algorithmic classes for solving the associated minimization problems. Finally, we make an opening to deep learning based approaches, which offer very effective reconstruction strategies.

1 Grandes lignes de l'évolution des problèmes inverses en imagerie

La résolution de problèmes inverses permet de remonter à l'information originale d'un objet d'étude à partir d'observations dégradées de ce même objet. La généralité et la cohérence de l'approche inverse permet de traiter de façon globale et avec des méthodologies similaires des données de nature très différentes (images 2D ou 3D, images hyper-spectrales), voire hétérogènes.

Modèle direct – La première étape de résolution d'un problème inverse passe par l'écriture du modèle direct dont la forme la plus simple s'écrit :

$$z = A\bar{u} + \epsilon \quad (1)$$

où \bar{u} désigne l'objet original (inconnu), typiquement une image composée de N pixels, A une transformation linéaire (ou non-linéaire) modélisant le dispositif d'acquisition, ϵ une dégradation stochastique ou en d'autres termes un bruit de mesure, ici supposé additif, et z les observations, typiquement une image composée de M pixels.

Problème inverse – Le problème inverse consiste alors à estimer une image \hat{u} proche de \bar{u} à partir de l'information contenue dans z , d'une information complète ou partielle de A , de la statistique du bruit, et d'*a priori* sur la classe d'images à reconstruire.

Cette unique équation décrit de nombreux problèmes inverses en analyse d'images tels que le débruitage (A est l'opérateur identité et z est une version bruitée de l'objet original) ou la déconvolution (A est un opérateur de convolution et z une version floue et bruitée de l'objet original) communément rencontrée en astronomie ou microscopie. Elle décrit également de nombreux modèles d'acquisition rencontrés en imagerie médicale, notamment en imagerie par résonance magnétique (IRM : échantillonnage dans le domaine de Fourier) ou en tomographie (transformée de Radon).

L'intérêt des approches de type problèmes inverses réside dans leur formulation générale qui les rend applicables à de nombreuses modalités et dans leur capacité à fournir des estimateurs \hat{u} ayant d'excellentes propriétés.

Travaux pionniers – Les premiers travaux de cette discipline remontent à la définition du concept de problème bien posé proposé par J. Hadamard [18] en 1902 basé sur l'existence, l'unicité et la stabilité de la procédure d'inversion. Par construction, la majorité des problèmes inverses considérés en traitement du signal et des images (TSI) sont des problèmes mal posés.

Lorsque A n'est pas inversible, de nombreux travaux ont alors porté sur la résolution par maximum de vraisemblance proposé par R.A. Fisher en 1922 qui se ramène à la résolution des moindres carrés lorsque ϵ est un bruit blanc Gaussien et à la pseudo-inverse de Moore-Penrose si de plus A est linéaire, i.e., $\hat{u} \in \text{Argmin}_u \|Au - z\|_2^2 = (A^*A)^{-1}A^*z$, où A^* désigne l'adjoint de A .

Cependant, ces estimateurs sont rarement envisageables ou conduisent à un estimateur irrégulier provenant de l'amplification du bruit de mesure, principalement causé par le mauvais conditionnement de A . On peut chercher à améliorer ce conditionnement, ce qui a conduit à des approches de types SVD tronquée [19] ou alors proposer une régularisation de la solution par approche variationnelle pénalisée.

Approches variationnelles – Depuis ces travaux fondateurs, les avancées majeures concernent la résolution des problèmes inverses par minimisation d'un critère de la forme :

$$\hat{u} \in \underset{u \in C}{\text{Argmin}} \psi(Au, z) + \lambda\varphi(u) \quad (2)$$

où $C \subset \mathbb{R}^N$ représente l'ensemble des solutions admissibles, ψ correspond à une "distance" entre les observations z et leur modèle, φ un terme qui pénalise les solutions trop irrégulières et $\lambda > 0$ un paramètre de régularisation permettant l'ajustement entre adéquation aux données ψ et pénalisation φ . Cette minimisation générique permet de contrer le mauvais conditionnement des matrices (opérateurs) lorsque $\lambda > 0$. De plus, on peut retrouver les solutions irrégulières obtenues par maximum de vraisemblance (en prenant $\lambda = 0$).

Hyperparamètre – Une question sensible dans la résolution des problèmes inverses par approche variationnelle concerne le choix de l'hyperparamètre λ . Les principales techniques proposées pour régler automatiquement ce paramètre sont la minimisation d'estimateurs de l'erreur quadratique moyenne

(estimateur non-biaisé du risque de Stein : SURE) [13], la validation croisée généralisée (GCV) [15], la courbe en L [19], ou les approches bayésiennes par marginalisation de la distribution a posteriori ou méthodes MCMC (ces dernières nécessitant de spécifier la densité de probabilité de λ).

Dimensionnalité – La résolution de problèmes inverses en imagerie fait face à une augmentation toujours croissante du volume de données à traiter mais bénéficie également de l'accroissement de la puissance de calcul des ordinateurs. Les développements méthodologiques algorithmiques combinés aux performances numériques nous ont permis de passer de l'analyse d'images composées de $N = 10^3$ pixels dans les années 80 [14] à l'analyse d'images hyperspectrales de taille 10^7 [17] quarante ans plus tard pour un même temps de calcul et tout cela en améliorant les performances de reconstruction, grâce à l'utilisation de pénalisations φ finement choisies.

Vers le deep learning – Face à ces stratégies variationnelles régularisées constituant l'état-de-l'art, l'analyse d'images et, plus spécifiquement, la résolution de problèmes inverses fait face depuis une dizaine d'années au développement de méthodes reposant sur l'apprentissage profond [16]. Dans la littérature on trouve essentiellement trois classes d'approches : (i) les réseaux d'apprentissage "end-to-end" que l'on peut aussi qualifier de "boîte noire", les plus développés lors des premiers travaux combinant problèmes inverses et apprentissage profond, puis, plus récemment, (ii) les approches Plug & Play (PnP) [25] permettant de conserver le cadre standard des approches variationnelles régularisées où la différence réside dans le choix du terme de pénalisation dont l'opérateur proximal [23] associé prend la forme d'un réseau de neurones fermement contractant, et enfin, (iii) les approches dites d'algorithmes déroulés (unfolded/unrolled) [1] permettant de construire un réseau s'inspirant des algorithmes proximaux tout en convergeant en très peu d'itérations (i.e., de couches de réseaux). Ces deux dernières classes d'approches permettent de proposer un continuum entre les approches variationnelles standards et les approches qualifiées de "boîte noire".

Plan – Après cette introduction permettant de fixer les grandes avancées en résolution de problèmes inverses, nous proposons de revenir sur certains éléments clés de leurs résolution. Nous détaillerons tout d'abord l'interprétation bayésienne en section 2 dont la formulation permet de guider le choix du terme d'attache aux données. Puis, nous présenterons dans la section 3 les grandes classes de pénalisation/régularisation. Les principaux algorithmes mis en oeuvre pour la résolution de (2) seront présentés dans la section 4. Enfin, une ouverture vers l'apprentissage profond sera effectuée dans la section 5 où nous donnerons quelques éléments de compréhension.

2 Interprétation bayésienne

Afin d'intuiter la formulation générique (2), il peut être opportun de faire appel à la formulation bayésienne. Pour cela, nous supposons que u et z sont des réalisations de vecteurs aléatoires U et Z et l'estimation peut alors être atteinte par une stratégie du Maximum A Posteriori (MAP). L'objectif est alors d'estimer \hat{u} qui maximise la distribution a posteriori $\mu_{U|Z=z}$ qui peut être reformulée par le théorème de Bayes puis par monotonie du logarithme comme :

$$\hat{u} \in \underset{u \in \mathbb{R}^N}{\text{Argmin}} \left(\underbrace{-\log \mu_{Z|U=u}(z)}_{\text{vraisemblance}} - \log \underbrace{\mu_U(u)}_{\text{a priori}} \right). \quad (3)$$

Le premier terme de (3) quantifie la fidélité aux observations z et dépend du modèle de formation des données (1). A titre d'exemple, lorsque ϵ modélise un bruit blanc Gaussien de variance σ^2 et $A \in \mathbb{R}^{M \times N}$, la vraisemblance prend la forme

$$\mu_{Z|U=u}(z) = (2\pi\sigma^2)^{-\frac{M}{2}} e^{-\frac{\|Au-z\|^2}{2\sigma^2}}$$

$$\psi(Au, z) = \frac{1}{2\sigma^2} \|Au - z\|^2. \quad (4)$$

Ce même raisonnement permet, par exemple, de motiver une attache aux données de type divergence de Kullback-Leibler lorsque le modèle d'acquisition des données est un processus de Poisson [26].

Le second terme est relié à l'information *a priori* sur l'image originale conduisant à $\varphi(u) = -\log \mu_U(u)$. Si la littérature en restauration d'image s'appuie fortement sur le modèle bayésien pour justifier le choix de l'attache aux données, les modèles de pénalisation ont quant à eux plutôt émergé du côté de la littérature dédiée aux ondelettes, à la parcimonie ou aux modèles variationnels continus.

3 Pénalisation/Régularisation

Une première classe de pénalisations implique des fonctions différentiables et convexes telles que la régularisation de Tikhonov [28], qui fut proposée dans les années 60 aussi appelée régression d'arête (i.e. *ridge regression*) et qui favorise les solutions lisses. La régularisation de Tikhonov peut également être interprétée comme un filtrage de Wiener (1949) en s'appuyant sur la connaissance des densités spectrales du bruit et du signal. On peut également citer la pénalisation de Huber [20] permettant d'approcher la norme ℓ_1 et favorisant des solutions parcimonieuses.

A partir des années 90, la résolution des problèmes inverses a connu la révolution des pénalisations non lisses dont la régularisation la plus connue est sans nul doute la pénalisation par variation totale, proposée par Rudin, Osher, et Fatemi (ROF) [27] et dont la définition originale fut donnée dans une formulation continue. Sous forme discrétisée, plusieurs définitions de la variation totale sont possibles et dépendent essentiellement du choix de l'opérateur de discrétisation. La formulation usuelle de la variation totale isotrope s'exprime :

$$\varphi(u) = \|Du\|_{2,1} = \sum_{n=(n_1, n_2)} \sqrt{((D_1 u)_{n_1, n_2})^2 + ((D_2 u)_{n_1, n_2})^2}$$

avec $(D_1 u)_{n_1, n_2} = u_{n_1+1, n_2} - u_{n_1, n_2}$ et $(D_2 u)_{n_1, n_2} = u_{n_1, n_2+1} - u_{n_1, n_2}$. La pénalisation par variation totale permet d'obtenir de très bonnes performances en débruitage, surtout si l'image contient de larges régions uniformes. Cependant, pour des images naturelles, cette méthode laisse apparaître un effet de *staircasing* que les utilisateurs préfèrent éviter, par exemple en astronomie où une version lissée de la variation totale est souvent utilisée. Il s'agit de la pénalisation par variation totale hyperbolique [6] qui s'exprime $\varphi(u) = \sum_n \sqrt{\|(Du)_n\|_2^2 + \nu}$, faisant apparaître un paramètre supplémentaire $\nu > 0$ à ajuster. Elle possède cependant l'avantage de lisser l'effet constant par morceaux et de fournir une transition plus lisse qui peut apparaître plus réaliste dans certaines modalités d'imagerie. D'autres pénalisations avancées ont été proposées dans la littérature. Par exemple les pénalisations favorisant la parcimonie dans une base ou une trame d'ondelettes [21, 26] et dont la pénalisation prend une forme similaire à la pénalisation par variation totale : $\varphi(u) = \|Du\|_{\bullet}$ où D modélise dans ce cas l'opérateur associé

à la base ou trame d'ondelettes. On parle alors de formulation à l'analyse, devenant formulation à la synthèse lorsque D est déplacé dans le terme d'attache aux données par utilisation de son adjoint. Par ailleurs, $\|\cdot\|_\bullet$ est une pénalisation favorisant la parcimonie pouvant prendre la forme d'une norme ℓ_1 ou d'une norme mixte $\ell_{1,2}$ comme définie dans (3) offrant une flexibilité dans le choix de D et des groupes de coefficients considérés. On peut se référer à [2] pour des choix de groupes judicieux. Il existe également des formes avancées de pénalisation par variation totale comme la variation totale généralisée (TGV) [4], la variation totale non-locale ou des pénalisations par tenseurs de structure [8].

Une troisième classe de pénalisation repose sur l'utilisation de pénalisations non-convexes de type ℓ_1 -pondérée ou quadratique tronquée [24], qui cependant introduisent des difficultés algorithmiques associées à la minimisation.

Une dernière classe de pénalisation, basée apprentissage, est apparue récemment. Elle sera évoquée dans la section 5.

4 Solutions algorithmiques

Les avancées algorithmiques majeures pour la résolution de problèmes inverses émanent de 2 sources : 1) les adaptations entre propriétés des fonctions φ et ψ et 2) les avancées en optimisation numérique.

Les travaux fondateurs en problèmes inverses sont indissociables de l'algorithme de gradient explicite et des versions accélérées (e.g., Levenberg-Marquardt ou L-BFGS) [30].

Il faudra attendre les années 2000 et l'essor des méthodes proximales [3, 10, 2] (méthodes de sous-gradient implicite) pour être en mesure de traiter de grands volumes de données combinés avec des pénalisations avancées (comme celles mentionnées dans la section précédente) et ainsi obtenir des gains significatifs en qualité de reconstruction. Ces algorithmes reposent sur la notion clé d'opérateur proximal [23] associé à une fonction f , défini pour tout $\tau > 0$ par

$$(\forall x \in \mathbb{R}^N) \quad \text{prox}_{\tau f} x = \arg \min_{y \in \mathbb{R}^N} \frac{1}{2} \|y - x\|_2^2 + \tau f(y). \quad (5)$$

L'opérateur proximal est une généralisation de la projection sur un ensemble convexe fermé non vide C de \mathbb{R}^N , notée P_C , en remarquant que $\text{prox}_{\iota_C} = P_C$ avec ι_C la fonction indicatrice associée à l'ensemble C telle que, pour tout $x \in \mathbb{R}^N$, $\iota_C(x) = 0$ si $x \in C$ et $+\infty$ sinon. L'opérateur proximal associé à la norme ℓ_1 (très utilisé en analyse d'images) se réduit à une opération de seuillage doux de paramètre τ .

Les méthodes proximales, en plus de permettre la gestion de fonctions non-lisses, ont permis d'offrir un cadre unificateur entre optimisation lisse sous-contrainte (e.g. contraintes de dynamiques ou épigraphiques) et optimisation non-lisse.

Formellement, si le problème considéré est de la forme $\min_{x \in \mathbb{R}^N} f(x)$ avec f une fonction convexe non-lisse, une itération de l'algorithme du point proximal (ou sous-gradient implicite) s'écrit :

$$\begin{aligned} x^{[k+1]} &= x^{[k]} - \tau_k g^{[k]} \quad \text{avec} \quad g^{[k]} \in \partial f(x^{[k+1]}) \quad (6) \\ &= \text{prox}_{\tau_k f} x^{[k]} \quad (7) \end{aligned}$$

où $\tau_k > 0$ et $\partial f(x) = \{w \in \mathbb{R}^N \mid (\forall y \in \mathbb{R}^N) \langle y - x, w \rangle + f(x) \leq f(y)\}$ désigne la sous-différentielle de Moreau [23] qui se réduit au gradient lorsque la fonction est différentiable (i.e., $\partial f = \{\nabla f\}$). Dans le cas où la sous-différentielle est calculée en $x^{[k]}$ et non en $x^{[k+1]}$, on obtient une itération de l'algorithme de sous-gradient explicite.

Contrairement à la méthode de sous-gradient explicite, la convergence de l'algorithme proximal repose sur un choix de pas τ_k moins contraignant. L'efficacité des itérations proximales est conditionnée à la connaissance sous forme explicite de l'opérateur proximal. De nombreuses formes explicites ont été proposées dans la littérature et sont répertoriées pour la plupart sur le site web Prox-Repository.

Une difficulté supplémentaire intervient lorsqu'il s'agit de calculer l'opérateur proximal associé à une somme de deux fonctions (i.e., $f = \psi + \lambda\varphi$) comme c'est le cas dans (2). Dans ce cas, très peu de formes explicites existent et il est alors nécessaire d'avoir recours à des algorithmes proximaux d'éclatement (*splitting*) qui sont des variations autour de ces itérations de façon à minimiser une somme de fonctions convexes pouvant être non-lisses. L'approche de référence est l'algorithme explicite-implicite (FB : *forward-backward*) [7] qui suppose ψ de gradient β -Lipschitz et dont les itérations sont de la forme :

$$x^{[k+1]} = \text{prox}_{\tau_k \lambda \varphi} \left(x^{[k]} - \tau_k \nabla \psi(x^{[k]}) \right) \quad (8)$$

où $\tau_k < 2\beta^{-1}$. L'étape explicite est une étape qui permet de gérer l'inversion de A et l'étape implicite est une étape de débruitage activant les différentes pénalisations présentées dans la section précédente. Il peut cependant parfois être coûteux numériquement d'avoir recours à des sous-itérations pour calculer l'opérateur proximal et dans ce cas, il peut être opportun d'avoir recours à des algorithmes primaux-duaux [5, 11].

Pour la gestion de fonctions non-convexes, une littérature reposant sur l'inégalité de Kurdyka-Łojasiewicz a permis de proposer des algorithmes ayant une convergence vers un point critique, utile dans le cas de pénalisations non-convexes ou encore en déconvolution aveugle [9].

5 Vers l'apprentissage profond

Une alternative récente à la minimisation de (2) combinant un choix judicieux d'attache aux données, de pénalisation et de stratégies d'optimisation, consiste à recourir à un apprentissage supervisé reposant sur des réseaux de neurones profonds.

On rappelle que les réseaux profonds sont un empilement de couches. Chaque couche est composée d'une transformation linéaire (par exemple, des convolutions) et une fonction d'activation non linéaire. Formellement, un réseau composé de K couches peut être écrit comme

$$u^{[K]} = d_\Theta(u^{[1]}) = \eta^{[K]}(T^{[K]} \dots \eta^{[1]}(T^{[1]}u^{[1]} + b^{[1]}) \dots + b^{[K]})$$

où, pour chaque $k \in \{1, \dots, K\}$, $T^{[k]}$ est la matrice de poids, $b^{[k]}$ est le biais, et $\eta^{[k]}$ est la fonction d'activation non linéaire.

Pour la résolution de problèmes inverses, l'approche la plus standard, qualifiée de "end-to-end", consiste à considérer $\hat{u} = d_\Theta(z)$ où d_Θ modélise un réseau de type CNN ou U-net.

Plus récemment, la conception de d_Θ s'est appuyée sur la connaissance que l'on a du problème physique [29]. Par exemple dans le cas des approches PnP, il s'agit d'avoir recours à une approche de type FB où $\text{prox}_{\tau_k \lambda \varphi}$, qui s'apparente à une action de débruitage, peut être remplacé par des débruiteurs de type BM3D [12] ou un réseau de neurones [25]. On peut également faire référence aux approches déroulées qui consistent à déployer des méthodes itératives telles que l'algorithme FB. Dans le cas où ψ est définie par (4) [16] cela conduit à $T^{[k]} = \text{Id} - \tau_k A^* A$, $\eta^{[k]} = \text{prox}_{\tau_k \lambda \varphi}$, et $b^{[k]} = \tau_k A^* z$. Cette approche permet alors de prendre en charge l'apprentissage

du paramètre de régularisation, le pas de l’algorithme et on peut même envisager l’apprentissage de l’opérateur linéaire via des formulations primale-duales ou sous-itérations lorsque $\psi = \|D \cdot\|_1$ [22] au prix d’un coût d’apprentissage plus élevé.

6 Conclusions et perspectives

Cet article retrace l’évolution des méthodes de résolution de problèmes inverses en imagerie. La figure 1 permet de retracer cette évolution sur un exemple de restauration d’image. Les performances obtenues avec les méthodes récentes reposant sur la combinaison d’outils standards et de réseaux de neurones permettent d’atteindre des résultats de restauration de très grande qualité. Cependant, de nombreuses pistes restent à explorer, allant d’une compréhension approfondie et moins énergivore des approches reposant sur les réseaux de neurones, à l’estimation conjointe des paramètres sous-jacents (déconvolution aveugle, hyperparamètres de régularisation, . . .), en passant par l’accélération des procédures.

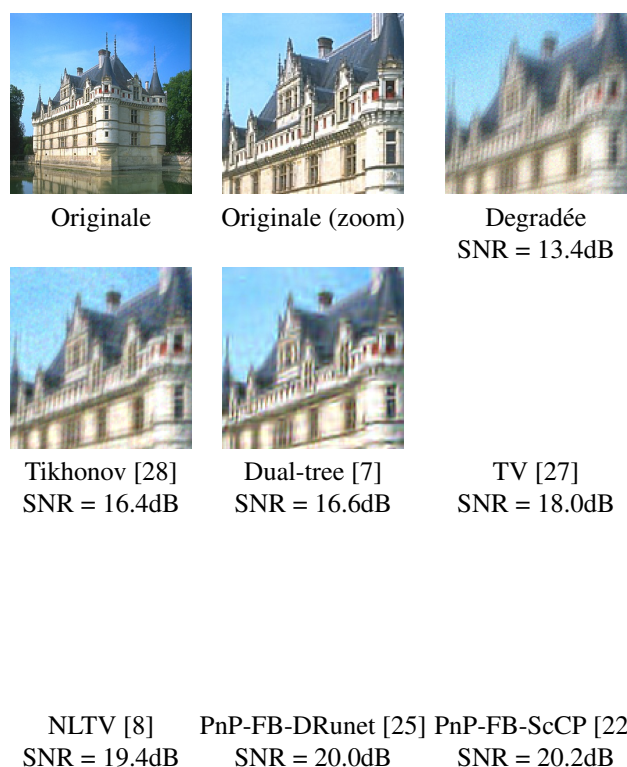


FIGURE 1 : Résultats visuels de restauration pour différents choix de régularisation ou par méthodes PnP.

Références

[1] J. ADLER et O. ÖKTEM : Unrolled iterative algorithms for inverse problems. *Inverse Problems*, 34(1):014004, 2018.

[2] F. BACH, R. JENATTON, J. MAIRAL et G. OBOZINSKI : Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.*, 4(1):1–106, 2012.

[3] H. H. BAUSCHKE et P. L. COMBETTES : *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2017.

[4] K. BREDIES, K. KUNISCH et T. POCK : Total generalized variation. *SIAM J. Imaging Sci.*, 3(3):492–526, Sep. 2010.

[5] A. CHAMBOLLE et T. POCK : An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[6] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT et M. BARLAUD : Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6(2):298–311, 1997.

[7] C. CHAUX, P.L. COMBETTES, J.-C. PESQUET et V.R. WAJS : A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518, Jun. 2007.

[8] G. CHIERCHIA, N. PUSTELNIK, B. PESQUET-POPESCU et J.-C. PESQUET : A nonlocal structure tensor-based approach for multicomponent image recovery problems. *IEEE Trans. Image Process.*, 23(12):5531–5544, Dec. 2014.

[9] E. CHOUZENOUX, J.-C. PESQUET et A. REPETTI : Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.*, 162(1):107–132, 2014.

[10] P. L. COMBETTES et J.-C. PESQUET : Proximal splitting methods in signal processing. In H. H. BAUSCHKE, R. S. BURACHIK, P. L. COMBETTES, V. ELSER, D. R. LUKE et H. WOLKOWICZ, éditeurs : *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2011.

[11] L. CONDAT, D. KITAHARA, A. CONTRERAS et A. HIRABAYASHI : Proximal splitting algorithms : A tour of recent advances, with new twists. *SIAM Rev.*, Apr. 2019.

[12] K. DABOV, A. FOI, V. KATKOVNIK et K. EGIAZARIAN : Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, Aug. 2007.

[13] C.-A. DELEDALLE, S. VAITER, J. FADILI et G. PEYRÉ : Stein unbiased gradient estimator of the risk (sugar) for multiple parameter selection. *SIAM J. Imaging Sci.*, 7(4):2448–2487, 2014.

[14] S. GEMAN et D. GEMAN : Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, 1984.

[15] G. H. GOLUB, M. HEATH et G. WAHBA : Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[16] K. GREGOR et Y. LECUN : Learning fast approximations of sparse coding. In *Proc. Int. Conf. Mach. Learn.*, ICML’10, page 399–406, Madison, WI, USA, 2010. Omnipress.

[17] C. GUILLOTEAU, T. OBERLIN, O. BERNÉ et N. DOBIGEON : Hyperspectral and multispectral image fusion under spectrally varying spatial blurs—application to high dimensional infrared astronomical imaging. *IEEE Trans. Comput. Imaging*, 6:1362–1374, 2020.

[18] J. HADAMARD : Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bulletin*, pages 49–52, 1902.

[19] P. C. HANSEN : *Rank-Deficient and Discrete Ill-Posed Problems*. Society for Industrial and Applied Mathematics, 1998.

[20] P. J. HUBER : Robust estimation of a location parameter. *Ann. Statist.*, 53(1):73–101, 1963.

[21] L. JACQUES, L. DUVAL, C. CHAUX et G. PEYRÉ : A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12):2699–2730, 2011.

[22] H. T. V. LE, N. PUSTELNIK et M. FOARE : The faster proximal algorithm, the better unfolded deep learning architecture ? the study case of image denoising. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 947–951, 2022.

[23] J. J. MOREAU : Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[24] M. NIKOLOVA : Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Model. Simul.*, 405(3):960–991, 2005.

[25] J.-C. PESQUET, A. REPETTI, M. TERRIS et Y. WIAUX : Learning maximally monotone operators for image recovery. *SIAM J. Imaging Sci.*, 14(3):1206–1237, 2021.

[26] N. PUSTELNIK, A. BENAZZA-BENHAYIA, Y. ZHENG et J.-C. PESQUET : Wavelet-based image deconvolution and reconstruction. *Wiley Encyclopedia of EEE*, 2016.

[27] L. I. RUDIN, S. OSHER et E. FATEMI : Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259–268, 1992.

[28] A. N. TIKHONOV : Solution of incorrectly formulated problems and the regularization method. *Sov. Math. Dokl.*, 4:1035–1038, 1963. English translation of Doklady Akademii Nauk SSSR, vol. 151, pp. 501–504.

[29] B. WEN, S. RAVISHANKAR, Z. ZHAO, R. GIRYES et J. C YE : Learning for computational imaging. *IEEE Signal Process. Mag.*, 40, 2023.

[30] S. WRIGHT et J. NOCEDAL : Numerical optimization. *Springer Science*, 35(67-68):7, 1999.