

Imagerie radio-interférométrique robuste par dépliement neuronal

Yassine MHIRI¹, Mohammed Nabil EL KORSO³, Arnaud BRELOY², Pascal LARZABAL¹

¹ SATIE, Université Paris-Saclay; ² LEME, Université Paris Nanterre. ³ L2S, Université Paris-Saclay.

{yassine.mhiri}{pascal.larzabal}{m.elkorso}@universite-paris-saclay.fr,
{abreloy}@parisnanterre.fr

Résumé – Les algorithmes d’imagerie radio-interférométrique sont particulièrement sensibles aux bruits non gaussiens affectant les mesures. De telles perturbations, notamment dues à la présence d’interférences de radiofréquences, peuvent être modélisées par des distributions gaussiennes composées. Cette modélisation conduit à la formulation d’un problème d’estimation du maximum de vraisemblance pour la tâche de reconstruction d’image. Cependant, le choix de la distribution et de ses hyperparamètres peut avoir un impact significatif sur l’erreur de reconstruction. Dans cet article, nous généralisons l’algorithme d’espérance-maximisation pour l’imagerie radio-interférométrique récemment introduit par les auteurs à une modélisation du bruit additif par des distributions gaussiennes composées. Ensuite, nous proposons de déplier l’algorithme précédemment introduit, donnant lieu à un algorithme d’apprentissage profond dérivé pour lequel la robustesse au bruit hétérogène est pilotée de manière informée par la base d’apprentissage. Enfin, nous illustrons les performances du réseau proposé sur des données simulées.

Abstract – Radio-interferometric imaging algorithms are particularly sensible to non Gaussian noise in the measurements. Such perturbations, notably due to the presence of radiofrequency interference are modeled by the class of compound Gaussian distributions. This modelization leads to the formulation of a maximum likelihood estimation problem for the image reconstruction task. Yet the choice of the distribution and its hyperparameters can significantly impact the reconstruction error. In this paper, we first derive a general expectation maximization imaging algorithm based on compound Gaussians modelling. Then, we propose a deep unfolding network based on the former algorithm for which a robustness to heterogeneous noise is derived by learning from data. Finally, we illustrate the performances of the proposed network on simulated data.

1 Introduction

Les radiotélescopes interféromètres permettent de mesurer le champ électromagnétique dans le domaine des ondes radio. Cette dernière décennie a vu le développement d’une nouvelle génération de radio-interféromètres, tels que le LOFAR ou le SKA, améliorant la résolution et la sensibilité des mesures. Ces interféromètres fonctionnent en mesurant les corrélations entre plusieurs capteurs, à partir desquels une synthèse d’images est réalisée [1]. La synthèse d’images s’apparente à une tâche de reconstruction à partir de mesures incomplètes du plan de Fourier de l’image du ciel, ce qui conduit à un problème inverse mal posé [2, 3]. Pour ce faire, des algorithmes d’imagerie basés sur de l’optimisation convexe ont été développés [3, 4]. Ces méthodes prennent en compte un a priori parcimonieux sur l’image du ciel et utilisent des algorithmes proximaux tel ISTA (Iterative Soft Thresholding Algorithm) [5]. La synthèse d’image peut également être effectuée par une estimation du maximum de vraisemblance régularisée. Notamment en considérant un bruit blanc gaussien additif sur les mesures, le problème d’optimisation résultant est analogue à un problème d’optimisation convexe [3]. Néanmoins, les données radio interférométriques ne peuvent pas être modélisées de manière appropriée en utilisant un bruit blanc gaussien. En effet, la nouvelle génération de radiotélescopes interféromètres est construite à partir de capteurs de plus en plus sensibles, ce qui augmente la probabilité que les données soient corrompues

par des valeurs aberrantes. La présence de valeurs aberrantes, même de faible niveau, dans les mesures peut avoir un impact important sur les performances des algorithmes d’imagerie. Des analyses théoriques et expérimentales ont été menées et ont démontré la nonadéquation d’une distribution gaussienne pour modéliser la présence d’anomalies dans les mesures dans le contexte de la radioastronomie [6, 7]. Nous avons proposé dans [8] un algorithme d’espérance-maximisation (EM) pour l’imagerie radio-interférométrique basé sur une t -distribution complexe qui démontre une meilleure efficacité en présence de données non gaussiennes. À notre connaissance, il n’existe pas d’algorithmes d’imagerie radio-interférométrique se basant sur d’autres distributions non gaussiennes. En effet, la résolution du maximum de vraisemblance dans le cas des distributions de la classe des gaussiennes composées n’est en général pas triviale. La prise en compte de la distribution du bruit peut être mise en oeuvre au travers de méthodes basées sur un apprentissage statistique [9, 10]. Toutefois, dans le contexte de la radio interférométrie, la plupart des architectures neuronales introduites sont complexes et produisent des poids qui ne sont pas directement interprétables, contrairement aux méthodes d’optimisation basées sur une modélisation physique du signal. Cette interprétabilité est cruciale dans le contexte de la reconstruction d’images radio-interférométriques à partir desquelles sont tirées des conclusions scientifiques. En ce sens, nous proposons une méthodologie d’apprentissage informée pour l’imagerie radio-interférométrique. Pour cela, nous considérons le

cadre introduit dans [11] qui a conduit à l'émergence d'architectures neuronales interprétables dérivées à partir d'algorithmes itératifs. L'avantage de ces approches est de produire une architecture explicable à partir d'un algorithme basé sur une modélisation des données. De plus, l'inférence est plus rapide comparée à un algorithme itératif en raison d'un apprentissage supervisé [12]. Cette interprétabilité nous permet d'identifier les poids liés à la statistique du bruit, ce qui conduit à un apprentissage approprié à faible coût du réseau. Dans cet article, nous présentons un algorithme EM qui produit une estimation du maximum de vraisemblance des images du ciel dans le cas de distributions non gaussiennes. Ensuite, nous dérivons une architecture neuronale déroulée à partir de l'algorithme EM proposé. La robustesse de la méthodologie proposée à la présence d'interférences est illustrée sur des données simulées.

2 Algorithme EM robuste pour l'imagerie radio-interférométrique

2.1 Modélisation statistique

Les radio-interféromètres mesurent la cohérence spatiale du champ électromagnétique pour toutes les paires d'antennes qui composent le réseau d'antennes. Ces mesures sont appelées *visibilités* et peuvent être liées à la transformée de Fourier 2D de l'image du ciel que l'on souhaite reconstruire. Plus précisément, les données radio-interférométriques sont des échantillons de la transformée de Fourier 2D de l'image céleste. Ainsi, un modèle discret pour les données radio-interférométriques peut être dérivé,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

où $\mathbf{y} \in \mathbb{C}^{K \times 1}$ représente les visibilités mesurées, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ l'image vectorisée du ciel que nous cherchons à reconstruire, \mathbf{n} un bruit additif et $\mathbf{H} \in \mathbb{C}^{K \times N}$ correspond à l'opérateur de transformée de Fourier discrète (DFT). Les radio-interféromètres sont soumis à diverses perturbations additives qui ne peuvent pas être modélisées efficacement par un bruit gaussien. Notamment, les interférences de radiofréquence (RFI), qui représentent les ondes radios d'origine humaine, peuvent affecter de manière significative les données observées et nécessitent une modélisation spécifique [13]. Ainsi, nous considérons la classe des distributions gaussiennes composées, définie comme le produit d'une composante gaussienne t , appelée *speckle*, et d'une variable aléatoire positive τ , la *texture*,

$$\mathbf{n}[k] = \sqrt{\tau} \mathbf{t}[k], \quad (2)$$

où $\mathbf{t} \sim \mathcal{CN}(0, \mathbf{I})$ et $\tau \sim p(\tau)$. Le modèle direct s'écrit,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \text{diag}(\tau_1, \dots, \tau_K)^{1/2} \mathbf{t}, \quad (3)$$

où $\mathbf{y} \in \mathbb{C}^K$ est le vecteur de mesures. La vraisemblance logarithmique négative des données s'écrit alors,

$$\mathcal{L}(\mathbf{x}; \mathbf{y}) = \sum_{k=1}^K -\log \left(\int_0^\infty (\pi\tau)^{-1} \exp(-\tau^{-1}\delta_k^2) p(\tau) d\tau \right), \quad (4)$$

avec $\delta_k = \|\mathbf{y}_k - \mathbf{h}_k \mathbf{x}\|$, \mathbf{h}_k étant la $k^{\text{ème}}$ ligne de \mathbf{H} et \mathbf{y}_k la $k^{\text{ème}}$ entrée du vecteur de visibilités mesuré. Le problème de maximum de vraisemblance régularisé résultant est formulé en considérant une régularisation L1 afin de prendre en compte la parcimonie des images [3],

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} (\mathcal{L}(\mathbf{x}; \mathbf{y}) + \alpha \|\mathbf{x}\|_1). \quad (5)$$

2.2 Algorithme EM pour l'imagerie robuste

Le problème d'optimisation présenté en équation (5) est souvent intractable pour de telles distributions. Nous proposons alors de résoudre le problème de maximum de vraisemblance régularisée par un algorithme EM qui permet d'augmenter de manière itérative la vraisemblance [14] en maximisant itérativement l'espérance d'une vraisemblance complète. Pour ce faire, nous considérons l'espace de données complètes,

$$\boldsymbol{\xi} = (\mathbf{y}, \tau_1, \dots, \tau_N).$$

Ainsi, l'étape E s'exprime comme suit,

$$Q(\mathbf{x}|\mathbf{x}^{(m)}) \propto \sum_{k=1}^K \hat{\omega}_k \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \quad (6)$$

$$\hat{\omega}_k^{(m)} = \int_{\mathbb{R}^+} \frac{1}{\tau} p(\tau|\mathbf{y}_k, \mathbf{x}^{(m)}) d\tau. \quad (7)$$

L'étape M est exprimée en considérant une régularisation L1,

$$\hat{\mathbf{x}}^{(m+1)} = \underset{\mathbf{x}}{\text{argmax}} (Q(\mathbf{x}|\mathbf{x}^{(m)}) + \alpha \|\mathbf{x}\|_1). \quad (8)$$

Le problème d'optimisation convexe suivant peut être résolu avec différents algorithmes. Dans ce qui suit, nous considérons l'algorithme ISTA [5], et dérivons la mise à jour pour une itération,

$$\hat{\mathbf{x}}^{(m+1)} = \mathcal{T}_\alpha (\hat{\mathbf{x}}^{(m)} + \gamma \mathbf{H}^H \hat{\boldsymbol{\Omega}}^{(m)} (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{(m)})). \quad (9)$$

où $\hat{\boldsymbol{\Omega}}^{(m)} = \text{diag}(\hat{\omega}_1^{(m)}, \dots, \hat{\omega}_K^{(m)})$, $\mathcal{T}_\alpha(\cdot) = \max(|\cdot| - \alpha, 0)$ est la fonction de seuillage doux et γ est le pas d'optimisation.

Algorithm 1 Robust EM Imager

input : \mathbf{y}, \mathbf{H}

output : $\hat{\mathbf{x}}$

initialize : $\hat{\mathbf{x}} \leftarrow \mathbf{x}_0$

- 1: **while** critère d'arrêt non satisfait **do**
 - 2: E-Step : $\hat{\omega}_k^{(m)}$ à partir de (7) for $k \in [1, K]$
 - 3: M-Step : $\hat{\mathbf{x}}^{(m+1)}$ à partir de (9)
 - 4: **end while**
-

Dans [8], un algorithme EM est proposé en considérant une t -distribution pour le bruit additif, montrant une amélioration par rapport à une modélisation gaussienne. En effet, dans ce cas, l'espérance calculée en (7) a une forme explicite. Dans la section suivante, nous nous intéressons à une approche basée sur de l'apprentissage statistique pour l'imagerie radio-interférométrique. Plus précisément, nous considérons une version dépliée de l'algorithme EM proposé à partir de laquelle nous proposons un réseau profond déroulé pour l'imagerie radio-interférométrique.

3 Algorithme déroulé pour l'imagerie radio-interférométrique

3.1 Architecture déroulée

Une architecture neuronale est une fonction paramétrique composée de plusieurs couches, chacune prenant comme entrée la sortie de la précédente. Le dépliement neuronal consiste à construire une architecture à partir d'un algorithme itératif déplié [11]. Les poids de l'architecture résultante, pour lesquelles un apprentissage supervisé est conduit, pourront être directement interprétable à partir du modèle dont est issu l'algorithme itératif. Il a été montré que de telles architectures conduisent à un apprentissage à plus bas coût et à une accélération de l'inférence [12]. Ainsi, nous proposons de déplier l'algorithme présenté et d'apprendre l'étape d'espérance de l'algorithme EM de manière supervisée. Pour une itération de l'algorithme 1, la fonction de mise à jour peut se réécrire comme une fonction paramétrique,

$$\begin{aligned}\hat{\omega}^{(m+1)} &= f_{\Theta}(\mathbf{y}, \hat{\mathbf{x}}^{(m)}, \hat{\omega}^{(m)}), \\ \hat{\mathbf{x}}^{(m+1)} &= \mathcal{T}_{\alpha} \left(\hat{\mathbf{x}}^{(m)} + \gamma \mathbf{H}^H \hat{\Omega}^{(m+1)} (\mathbf{y} - \mathbf{H} \hat{\mathbf{x}}^{(m)}) \right).\end{aligned}\quad (10)$$

où $f_{\Theta}(\mathbf{y}, \hat{\mathbf{x}}^{(m)}, \hat{\omega}^{(m)})$ est une architecture neuronale, que l'on appellera *couche robuste* et Θ contient l'ensemble des paramètres à apprendre. La couche robuste remplace le calcul d'espérance en (7). L'architecture obtenue correspond à une architecture neuronale récurrente. Néanmoins, les architecture récurrentes peuvent conduire à un apprentissage lent et fastidieux dû à une dispersion du gradient [15]. En conséquence, nous proposons l'architecture suivante pour la couche robuste à partir de la méthodologie introduite dans [16], illustrée en figure 1,

$$\mathbf{h}^{(m)} = \text{Sigmoid}(\mathbf{W}_0 \|\mathbf{y} - \mathbf{H} \hat{\mathbf{x}}^{(m)}\|_2^2), \quad (11)$$

$$f_{\Theta}(\mathbf{y}, \hat{\mathbf{x}}^{(m)}, \hat{\omega}^{(m)}) = \text{ReLU}(\mathbf{W}_1 \hat{\omega}^{(m)} + \mathbf{W}_2 \mathbf{h}^{(m)}), \quad (12)$$

où $\mathbf{W}_0 \in \mathbb{R}^{K \times K}$, $\mathbf{W}_1 \in \mathbb{R}^{K \times K}$ et $\mathbf{W}_2 \in \mathbb{R}^{K \times K}$ représentent des couches linéaires, K étant la dimension du vecteur de mesure. Par conséquent, aucune hypothèse sur la distribution du bruit n'est effectuée. L'apprentissage sur des données perturbées par des bruits non gaussiens permet d'obtenir une architecture robuste à de telles perturbations. L'architecture résultante peut être interprétée comme la solution d'un moindres carrés pondérés pour un modèle linéaire gaussien où les pondérations sont obtenues à l'aide de la couche robuste. De plus, contrairement à l'algorithme EM originel, il n'est plus nécessaire de choisir empiriquement les hyperparamètres.

3.2 Résultats numériques

Dans cette section, nous présentons des résultats numériques et comparons les performances du réseau déplié sur 10 itérations avec l'algorithme EM proposé. Nous considérons 2 métriques. Le rapport signal sur bruit (SNR), dérivé de l'erreur quadratique moyenne, mesure la variance de l'estimateur. La

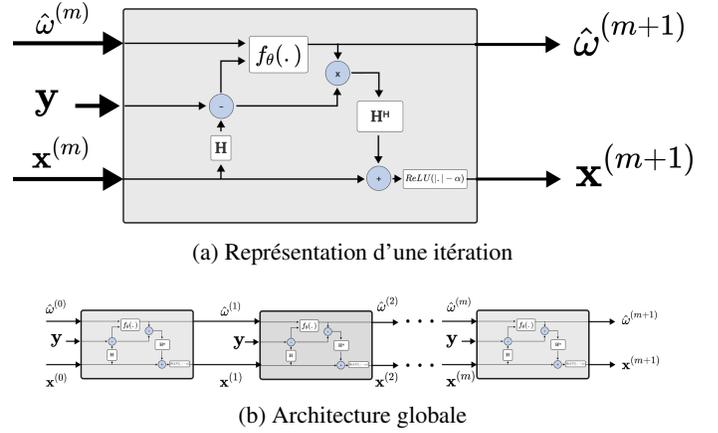


FIGURE 1 – Diagramme de l'architecture neuronale dépliée

similarité structurelle (SSIM) [17] vise à mesurer la similarité perceptive entre deux images.

L'ensemble de données d'entraînement est construit à partir d'un modèle de sources célestes. En effet, les émissions d'ondes radio provenant de galaxies correspondent à des ellipsoïdes gaussiennes et ont été efficacement utilisées comme données d'entraînement dans des approches supervisées [9]. Nous dérivons alors un ensemble de données d'entraînement à partir d'un ensemble de modèles de ciel composés de sources gaussiennes. Plus précisément, nous considérons un ensemble de données avec 1000 images de taille 64×64 pixels composées d'un certain nombre de sources gaussiennes aléatoires issues d'une distribution uniforme, $N_{sources} \sim \mathcal{U}(2, 10)$. De même, pour chaque source, la puissance et l'échelle sont issues d'une distribution uniforme, $P \sim \mathcal{U}(0.5, 10)$ et $\sigma \sim \mathcal{U}(2, 3)$. Pour chaque image simulée, les visibilité sans bruit correspondantes sont générées à partir de l'opérateur direct tel que défini dans (1), en considérant un radiotélescope interféromètre composé de 25 antennes placées aléatoirement dans le plan. Un bruit issu d'une mixture de gaussiennes composées est ensuite ajouté aux données d'entraînement et permet de simuler la présence d'interférences dans les mesures. Plus précisément, on considère une mixture de gaussiennes composées dont la texture est issue d'une distribution gamma, d'une inverse gamma ou d'une gaussienne inversée [18]. Dans ce cas, l'étape E de l'algorithme EM proposé ne peut être calculée de manière explicite. Nous utilisons l'erreur quadratique moyenne comme fonction coût pour l'entraînement, $l(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$, effectué sur 1000 *epochs*. Nous utilisons l'optimiseur Adam [19] avec un pas d'apprentissage de 0.0001. La différenciation automatique est effectuée avec Pytorch (1.13.1) et l'ensemble du code sources ayant permis d'obtenir les résultats numériques présentés sont mis à disposition dans un dépôt github public¹.

La figure 2 permet de comparer les métriques sur un jeu de données de test perturbées par une mixture de gaussiennes composées pour l'architecture dépliée et l'algorithme EM considérant un bruit t -distribué.

Pour l'ensemble des métriques considérées, nous affichons

1. <https://github.com/y-mhiri/robii/tree/GRETSI>

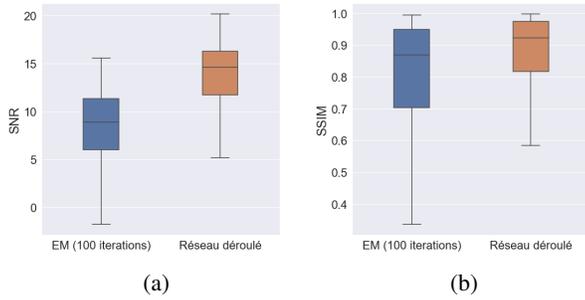


FIGURE 2 – Comparaison des performances de l’algorithme EM pour 100 itérations avec l’architecture neuronale proposée dépliée sur 10 itérations sur le SNR (a) et la SSIM (b).

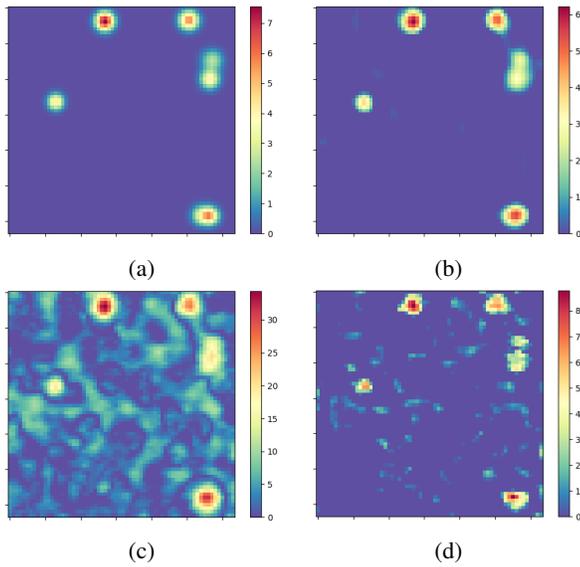


FIGURE 3 – (a) Image à reconstruire, (b) Image obtenue en utilisant l’architecture dépliée sur 10 itérations, (c) Image estimée par 10 itérations de l’algorithme EM considérant une t -distribution, (d) Image estimée par 100 itérations de l’algorithme EM considérant une t -distribution

les performances de l’algorithme EM pour 100 itérations ainsi que celles de l’architecture dépliée sur 10 itérations. L’algorithme déplié apporte de meilleures performances pour l’ensemble des métriques à un coût computationnel moindre.

Nous affichons sur la figure 3 l’image du ciel estimée par l’algorithme EM proposé pour 10 et 100 itérations, ainsi que les résultats obtenus à l’aide du réseau déplié.

4 Conclusion

Dans cet article, nous dérivons un algorithme d’espérance-maximisation pour la reconstruction d’images dans le contexte de radio-interféromètres en présence d’interférences modélisées par une gaussienne composée. Nous proposons ensuite une architecture neuronale dépliée à partir de cet algorithme. L’architecture neuronale résultante a des poids interprétables. Cela

permet alors de développer un algorithme robuste à des bruits non gaussiens grâce un apprentissage informé. Nous illustrons les performances de l’architecture apprise sur des images de ciel simulées. Comparativement aux méthodes de maximum de vraisemblance, l’algorithme appris montre de meilleurs résultats sur plusieurs métriques. Cet apprentissage informé apparaît prometteur comparativement à une approche deep learning classique [9].

Références

- [1] P. Zarka *et al.*, “Nenufar : Instrument description and science case,” in *2015 International Conference on Antenna Theory and Techniques (ICATT)*. IEEE, 2015, pp. 1–6.
- [2] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and synthesis in radio astronomy*. Springer Nature, 2017.
- [3] Y. Wiaux, L. Jacques, G. Puy, A. M. Scaife, and P. Vanderghenst, “Compressed sensing imaging techniques for radio interferometry,” *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.
- [4] L. Pratley *et al.*, “Robust sparse image reconstruction of radio interferometric observations with purify,” *Monthly Notices of the Royal Astronomical Society*, vol. 473, no. 1, pp. 1038–1058, 2018.
- [5] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [6] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, “Robust calibration of radio interferometers in non-gaussian environment,” *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5649–5660, 2017.
- [7] S. Kazemi and S. Yatawatta, “Robust radio interferometric calibration using the t -distribution,” *Monthly Notices of the Royal Astronomical Society*, vol. 435, no. 1, pp. 597–605, 2013.
- [8] Y. Mhiri, M. N. El Korso, A. Breloy, and P. Larzabal, “A robust em algorithm for radio interferometric imaging in the presence of outliers,” in *2022 IEEE Workshop on Signal Processing Systems (SiPS)*. IEEE, 2022, pp. 1–5.
- [9] L. Connor, K. L. Bouman, V. Ravi, and G. Hallinan, “Deep radio-interferometric imaging with polish : Dsa-2000 and weak lensing,” *Monthly Notices of the Royal Astronomical Society*, vol. 514, no. 2, pp. 2614–2626, 2022.
- [10] M. Terris, A. Dabbech, C. Tang, and Y. Wiaux, “Image reconstruction algorithms in radio interferometry : From handcrafted to learned regularization denoisers,” *Monthly Notices of the Royal Astronomical Society*, vol. 518, no. 1, pp. 604–622, 2023.
- [11] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proceedings of the 27th international conference on international conference on machine learning*, 2010, pp. 399–406.
- [12] V. Monga, Y. Li, and Y. C. Eldar, “Algorithm unrolling : Interpretable, efficient deep learning for signal and image processing,” *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.
- [13] Y. Mhiri, M. N. El Korso, A. Breloy, and P. Larzabal, “Multifrequency array calibration in presence of radio frequency interferences,” *Signal Processing*, p. 108613, 2022.
- [14] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [15] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *International conference on machine learning*. Pmlr, 2013, pp. 1310–1318.
- [16] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, “Advances in optimizing recurrent networks,” in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 8624–8628.
- [17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment : from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [18] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, “Bayesian calibration using different prior distributions : an iterative maximum a posteriori approach for radio interferometers,” in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2673–2677.
- [19] D. P. Kingma and J. Ba, “Adam : A method for stochastic optimization,” *arXiv preprint arXiv :1412.6980*, 2014.