



HAL
open science

Une introduction à l'imagerie computationnelle monodétecteur

Nicolas Ducros

► **To cite this version:**

Nicolas Ducros. Une introduction à l'imagerie computationnelle monodétecteur. Imageries optiques non conventionnelles pour la biologie, ISTE Group, pp.247-274, 2023, 10.51926/ISTE.9132.ch8 . hal-03551730

HAL Id: hal-03551730

<https://hal.science/hal-03551730>

Submitted on 1 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1

Une introduction à l'imagerie computationnelle monodétecteur

Nicolas DUCROS¹

¹CREATIS, INSA-Lyon, Lyon

1.1. Introduction

Ce chapitre s'intéresse à une famille de méthodes permettant l'acquisition de l'image d'une scène à partir d'un détecteur ponctuel. En particulier, l'imagerie computationnelle monodétecteur (en anglais, *single-pixel imaging*) consiste à mesurer, au moyen d'un dispositif physique, l'intensité de la scène vue au travers d'un motif de lumière. En répétant cette opération pour différents motifs bien choisis, il est possible de remonter à l'image initiale grâce à une étape de traitement numérique appelée reconstruction d'images.

A l'heure où les capteurs 2D capables de mesurer directement une image ont envahi notre quotidien, en raison de leur coût modéré (quelques dizaines de dollars l'unité) et de leur faible encombrement (de l'ordre du centimètre carré), la pertinence de l'imagerie monodétecteur demeure double. Premièrement, alors que la disponibilité de capteurs 2D s'est considérablement accrue dans la gamme des longueurs d'onde du visible, il n'en va pas de même sur toute la gamme du spectre électromagnétique. En particulier, les capteurs matriciels restent onéreux dans le domaine de l'infrarouge (cf. figure 1.1) où ces derniers sont par ailleurs victimes d'un bruit de mesure important, que les méthodes décrites dans ce chapitre permettent de réduire.

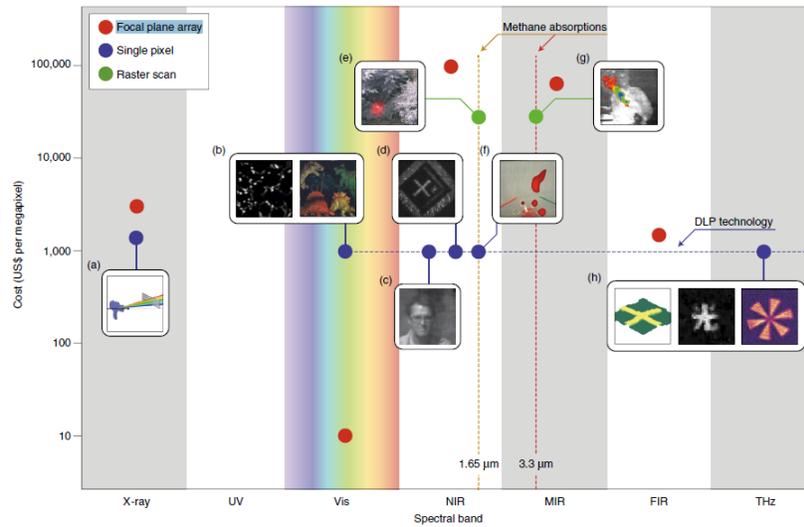


Figure 1.1 – Méthodes d'imagerie disponibles en fonction des longueurs ondes considérées issu de (Edgar et al. 2019). Alors que les systèmes d'imagerie reposant sur des capteurs matriciels ('Focal plane array') sont accessibles à un faible coût dans le domaine du visible ('Vis'), les systèmes monodétecteur ('Single-pixel') sont des alternatives pertinentes dans l'infrarouge proche ('NIR'), moyen ('MIR') ou lointain (FIR).

Deuxièmement, le principe d'acquisition monodétecteur permet d'accéder facilement à une troisième dimension, en plus des deux dimensions spatiales traditionnelles. Cette approche est donc naturellement adaptée à l'imagerie hyperspectrale qui acquiert la longueur d'onde des photons incidents, ou à l'imagerie des temps de vol. L'imagerie computationnelle monodétecteur a un large champ d'applications : imagerie infrarouge (Radwell et al. 2014), imagerie terahertz (Chan et al. 2008), microscopie de fluorescence (Studer et al. 2012), imagerie multispectrale (Rousset et al. 2018), ou encore imagerie de temps de vie de fluorescence (Pian et al. 2017).

L'imagerie computationnelle monodétecteur est associée à l'article fondateur de Duarte et al. (2008), qui a introduit la terminologie de *single-pixel imaging* sous laquelle la technique est aujourd'hui connue. Il est cependant intéressant de remarquer que le concept de spectroscopie de Hadamard, développé quatre décennies plus tôt (Nelson and Fredman 1970), s'appuie sur un principe d'acquisition identique. La spectroscopie de Hadamard peut donc être considérée comme un mouvement précurseur à l'imagerie "*single-pixel*", fournissant ainsi un éclairage théorique complémentaire à cette approche qui semble s'être développée postérieurement, mais de façon indépendante. L'imagerie computationnelle monodétecteur connaît dernièrement une nouvelle

phase d'essor avec les travaux de [Higham et al. \(2018\)](#), qui ont introduit le recours aux méthodes d'apprentissage profond (en anglais *deep learning*) pour la reconstruction.

Nous avons choisi de structurer ce chapitre selon ces trois grandes phases chronologiques qui feront l'objet de trois sections indépendantes : spectroscopie de Hadamard, single-pixel imaging et apprentissage profond. Nous illustrerons les différentes méthodes proposées par de nombreux exemples pour lesquels nous fournirons le code informatique permettant la simulation de données expérimentales et la reconstruction d'images. Ce code s'appuie sur le package Python SPyRit (), spécifiquement dédié à l'imagerie monodétecteur.

Formulation mathématique

Soit $\mathbf{x} \in \mathbb{R}^N$ le signal que l'on cherche à acquérir, généralement l'image d'une scène. Le principe de base de l'imagerie computationnelle consiste à mesurer, au moyen d'un dispositif physique, non pas \mathbf{x} directement mais une version transformée $\mathbf{y} = \mathbf{A}\mathbf{x}$, à partir de laquelle des algorithmes permettront de remonter à \mathbf{x} . La matrice $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M \leq N$, modélise le processus d'acquisition ; ses lignes contiennent des motifs de multiplexage qui seront acquis séquentiellement. Dans le cas où la matrice \mathbf{A} est inversible, l'image \mathbf{x} peut être retrouvée en calculant $\mathbf{A}^{-1}\mathbf{y}$.

Pour illustrer ce principe, considérons le cas simple où l'on cherche à acquérir une image constituée de $N = 2$ pixels à partir de $M = N = 2$ mesures indépendantes. On dispose d'un système expérimental permettant de mesurer des combinaisons linéaires de ces deux pixels. La première mesure y_1 acquiert $-x_1 + x_2$, tandis que la seconde y_2 acquiert $x_1 - 0.5x_2$. Autrement dit, nous sommes dans le cas où $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & -0.5 \end{bmatrix}$. Puisque $\mathbf{A}^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$, l'image d'origine peut être reconstruite simplement à partir de y_1 et y_2 , en calculant $x_1 = y_1 + 2y_2$ et $x_2 = 2y_1 + 2y_2$.

Mise en œuvre expérimentale

Le principe exposé précédemment nécessite de pouvoir moduler l'image selon différents motifs de lumière qui correspondent au ligne de la matrice \mathbf{A} . Un exemple de mise en œuvre expérimentale est représenté à la figure 1.2. Les premières démonstrations reposaient sur des masques perforés, permettant de laisser passer ou non la lumière, mais la mise en œuvre s'est considérablement simplifiée avec l'avènement de modulateurs spatiaux de lumière dont le motif de modulation peut-être contrôlé informatiquement. Différentes technologies de modulateurs sont disponibles. Parmi elles, les DMD (pour *digital micromirror devices*) sont des matrices constituées de plusieurs millions de micro-miroirs, d'une taille typique de $10 \times 10 \mu\text{m}^2$, pouvant être basculés dans deux orientations différentes à des fréquences pouvant dépasser les 20 kHz. Cette rapidité, couplée à un coût modéré de l'ordre de quelques centaines d'euros, font des DMD la technologie ayant la faveur d'une majorité des mises en œuvre récentes (cf. figure 1.2).

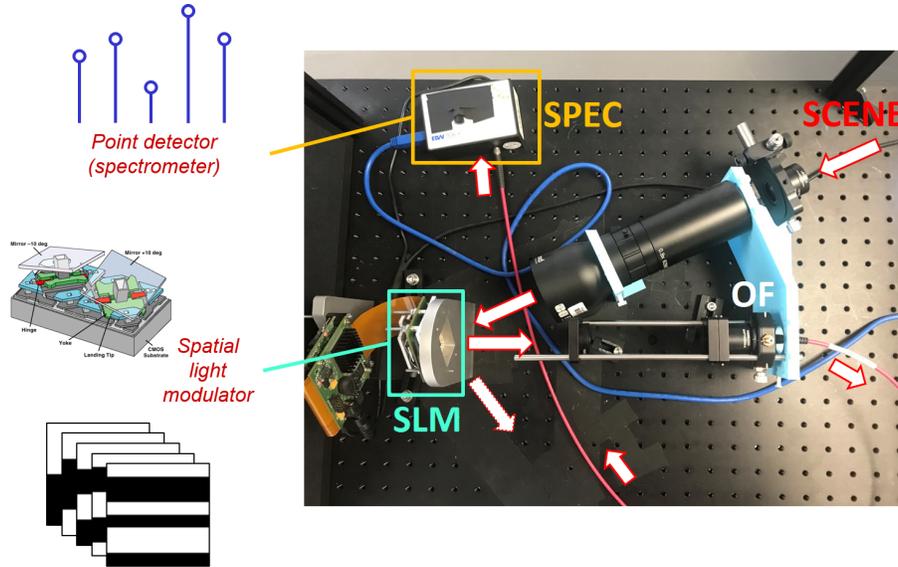


Figure 1.2 – Principe d’acquisition de l’optique par transformée de Hadamard. L’image issue de la scène (‘SCENE’) est projetée sur un modulateur spatial de lumière (‘SLM’) qui renvoie, selon un motif donné, une partie seulement de l’image vers un détecteur ponctuel, ici une fibre optique (‘OF’) conduisant la lumière vers un spectromètre compact (‘SPEC’). Une acquisition complète consiste à afficher différents motifs sur le modulateur spatial de lumière. Ces motifs correspondent aux lignes de la matrice A de l’équation (1.3), alors que les différentes valeurs mesurées par le détecteur ponctuel constitue le vecteur y de cette même équation. Connaissant A et ayant mesuré y , l’objectif est de remonter à l’image de la scène x qui demeure inconnue.

1.2. Optique par transformée de Hadamard : les origines (1970-80)

L’optique par transformée de Hadamard est fondée sur le modèle d’acquisition bruité

$$y_m = a_{m,1} x_1 + \dots + a_{m,N} x_N + \epsilon_m \quad [1.1]$$

$$= \mathbf{a}_m^\top \mathbf{x} + \epsilon_m \quad [1.2]$$

où y_m représente la m^{e} mesure, $\mathbf{a}_m^\top = [a_{m,1} \dots a_{m,N}]$ le m^{e} motif et ϵ_m est une source de bruit additive. Contrairement à une approche classique par balayage où pour laquelle chaque pixel de la scène x_n est acquis successivement, ce qui correspond au cas $y_m = x_m$, $1 \leq m \leq N$, l’optique par transformée de Hadamard mesure un mélange de tous les pixels de la scène, mélange dont les proportions varient d’une

mesure à l'autre. Considérant M mesures, nous avons l'écriture matricielle suivante

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \quad [1.3]$$

où $\mathbf{A} = [\mathbf{a}_1^\top, \dots, \mathbf{a}_M^\top]^\top \in \mathbb{R}^{M \times N}$ est la matrice contenant les M motifs et $\boldsymbol{\epsilon} \in \mathbb{R}^M$ représente le bruit.

L'optique par transformée de Hadamard s'intéresse au cas $M = N$ pour lequel le nombre de mesures est égal au nombre de pixels de l'image la scène. Dans l'hypothèse où les erreurs sont indépendantes entre-elles et indépendantes de l'image, l'image peut-être estimée au sens des moindres carrés par

$$\hat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{y} = \mathbf{x} + \boldsymbol{\eta}, \quad \text{avec } \boldsymbol{\eta} = \mathbf{A}^\dagger \boldsymbol{\epsilon}. \quad [1.4]$$

où \mathbf{A}^\dagger est la pseudo inverse de \mathbf{A} , qui coïncide avec \mathbf{A}^{-1} lorsque \mathbf{A} est inversible, et $\boldsymbol{\eta}$ représente l'erreur de reconstruction.

En supposant les erreurs de mesure à moyennes nulles, soit $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$, les erreurs de reconstruction sont également à valeurs moyennes nulles. En supposant par ailleurs \mathbf{A} inversible, nous avons $\mathbb{E}(\hat{\mathbf{x}}) = \mathbf{A}^{-1} \mathbf{y} = \mathbf{x}$ et l'on dit alors que $\hat{\mathbf{x}}$ est un estimateur non biaisé de \mathbf{x} . En revanche, la variance de l'erreur de reconstruction dépend du choix des motifs \mathbf{A} . En supposant que l'erreur de mesures est de variance connue σ^2 , on peut en effet montrer que

$$\text{Cov}(\hat{\mathbf{x}}) = \sigma^2 \mathbf{A}^\dagger (\mathbf{A}^\dagger)^\top. \quad [1.5]$$

où \mathbf{A}^\top est la transposée de la matrice \mathbf{A} .

REMARQUE.

$$\text{Cov}(\hat{\mathbf{x}}) = \text{Cov}(\boldsymbol{\eta}) = \text{Cov}(\mathbf{A}^\dagger \boldsymbol{\epsilon}) \quad [1.6]$$

$$= \mathbf{A}^\dagger \text{Cov}(\boldsymbol{\epsilon}) (\mathbf{A}^\dagger)^\top \quad [1.7]$$

Les erreur de mesures $\boldsymbol{\epsilon}$ étant indépendantes et de variance connue σ^2 , nous avons $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_N$, où $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ est la matrice identité. Par conséquent,

$$\text{Cov}(\hat{\mathbf{x}}) = \sigma^2 \mathbf{A}^\dagger (\mathbf{A}^\dagger)^\top. \quad [1.8]$$

De nombreux travaux se sont attachés à déterminer de "bonnes" matrices \mathbf{A} en cherchant à réduire l'erreur de reconstruction $\boldsymbol{\eta}$. L'un des résultats principaux, démontré par J. Hadamard, conduit à préconiser, lorsqu'elle existe, la matrice de Hadamard d'ordre N , que l'on note \mathbf{H}_N . Il s'agit d'une matrice binaire, orthogonale, telle que

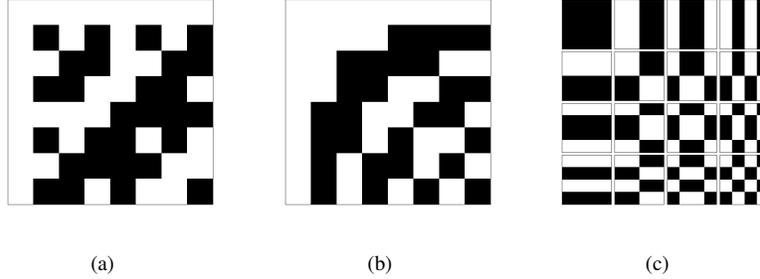


Figure 1.3 – Motifs de Hadamard 1D et 2D. (a) et (b) Matrice de Hadamard dans le cas $N = 8$. Chaque ligne de ces matrices correspond à un profil de modulation 1D ; les lignes de la matrice sont classées selon l'ordre naturel en (a), selon l'ordre de Walsh en (b). Une permutation permet de passer de l'une à l'autre des matrices. (c) Motifs de Hadamard 2D dans le cas $N = 4 \times 4$. Les 16 motifs de Hadamard 2D sont représentés de haut en bas et de gauche à droite, dans l'ordre de Walsh.

$\mathbf{H}_N \mathbf{H}_N^\top = N \mathbf{I}_N$. La figure 1.3 représente différentes matrices de Hadamard, en 1D et en 2D.

On peut caractériser l'erreur de reconstruction dans le cas d'une acquisition avec la matrice de Hadamard en substituant $\mathbf{A}^\dagger = \mathbf{H}^{-1} = \frac{1}{N} \mathbf{H}^\top$ dans l'équation (1.5). On obtient alors

$$\text{Cov}(\boldsymbol{\eta}) = \frac{\sigma^2}{N} \mathbf{I}_N, \quad 1 \leq n \leq N. \quad [1.9]$$

On observe que les erreurs sont indépendantes (les coefficients non diagonaux de $\text{Cov}(\boldsymbol{\eta})$ étant nuls) et que chaque pixel sera reconstruit avec un même niveau de bruit, de variance $\frac{\sigma^2}{N}$. Il a été démontré que l'erreur de reconstruction ainsi obtenue est la plus petite possible et, de ce point de vue, les matrices de Hadamard constituent un choix optimal.

Il est intéressant de comparer l'approche de Hadamard à l'approche intuitive par balayage où l'on acquiert chaque pixel séquentiellement, indépendamment des autres. Ce cas de figure revient à substituer $\mathbf{A} = \mathbf{I}_N$ dans l'équation (1.5) et conduit à obtenir à une variance $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_N$. L'approche de Hadamard offre ainsi un facteur de réduction de l'écart type du bruit de \sqrt{N} . Cette propriété de réduction du bruit est illustrée à la figure 1.4, où l'on retrouve, visuellement, une même qualité de reconstruction pour l'approche par balayage dans le cas $\sigma = 0.25$ et pour l'approche d'Hadamard dans le cas $\sigma = 16$, soit une amélioration d'un facteur $16/0.25 = 64$, qui correspond bien à la racine du nombre de pixels considérés.

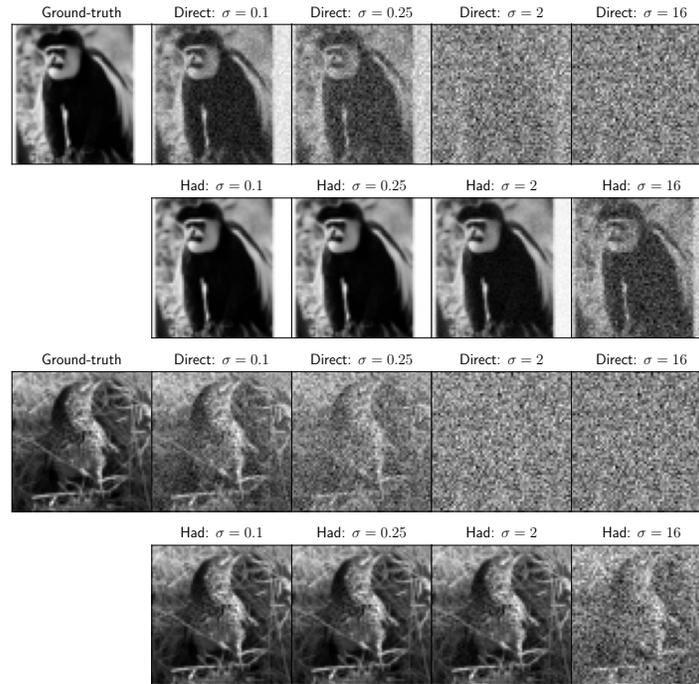


Figure 1.4 – Illustration de l’amélioration de l’approche de Hadamard (‘Had’) par rapport à une acquisition par balayage (‘Direct’) pour différents niveaux de bruit σ . Les mesures ont été simulées pour $N = 64 \times 64 = 4096$ motifs de Hadamard 2D.

Des contraintes supplémentaires sur les motifs, issues de considérations pratiques, ont également fait l’objet d’une attention particulière. Parmi elles, on citera notamment la positivité des motifs ainsi que leur caractère cyclique, nécessaire afin qu’ils puissent être réalisés expérimentalement par simple translation d’un masque physique (Nelson and Fredman 1970 ; Sloane and Harwit 1976). Cette dernière condition était particulièrement critique avant l’avènement de modulateur spatiaux contrôlables informatiquement.

1.3. Compressed sensing : le renouveau (2006-2016)

Acquisition sous échantillonnée

Afin de réduire les temps d'acquisition, il peut être souhaitable de limiter le nombre de mesures. Le problème de reconstruction revient alors à inverser un système linéaire $\mathbf{y} = \mathbf{A}\mathbf{x}$ pour lequel le nombre de mesures est inférieur au nombre d'inconnues, c'est-à-dire pour $\mathbf{A} \in \mathbb{R}^{M \times N}$ telle que $M \leq N$. On parle alors d'acquisition compressée.

En l'absence de bruit, il existe un ensemble de solutions vérifiant $\mathbf{m} = \mathbf{A}\mathbf{x}$. Parmi elles, un choix classique consiste à retenir la solution de norme minimale obtenue par la minimisation sous contrainte

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2^2 \quad \text{tel que } \mathbf{y} = \mathbf{A}\mathbf{x}. \quad [1.10]$$

où $\|\cdot\|_2$ représente la norme ℓ_2 . Lorsque que les lignes de \mathbf{A} sont indépendantes, c'est à dire lorsque \mathbf{A} est une matrice de plein rang, l'équation (1.10) admet la solution analytique suivante

$$\mathbf{x} = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{y}, \quad [1.11]$$

où l'on notera par la suite $\mathbf{A}^\dagger = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1}$ la matrice pseudo-inverse de \mathbf{A} . Dans le cas où les lignes de \mathbf{A} sont orthogonales, c'est-à-dire si $\mathbf{A}\mathbf{A}^\top = \mathbf{I}_M$, on a simplement $\mathbf{A}^\dagger = \mathbf{A}^\top$.

Cependant la solution pseudo-inverse obtenue à l'équation (1.11) est biaisée puisque de façon générale, $\mathbf{A}^\dagger \mathbf{y} = \mathbf{A}^\dagger \mathbf{A} \mathbf{x}^{\text{true}} \neq \mathbf{x}^{\text{true}}$. On dit également que la reconstruction n'est pas parfaite. Dans le cas où les lignes de \mathbf{A} sont indépendantes, une reconstruction parfaite impliquerait la condition $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_N$, qui ne peut être vérifiée si $M < N$.

Principe général du compressed sensing

La théorie du *compressed sensing* offre un cadre théorique pour la reconstruction parfaite de signaux à partir d'un faible nombre d'échantillons. Celle-ci a eu un impact considérable dans le domaine de l'imagerie assistée par ordinateur, durant la décennie ayant suivi son introduction (Donoho 2006).

Cette théorie s'appuie sur la notion de représentation parcimonieuse d'une image, c'est-à-dire sur l'existence d'un opérateur Ψ (typiquement adossé à une transformée en ondelettes ou de Fourier) tel que toute image \mathbf{x} peut-être représentée sous la forme

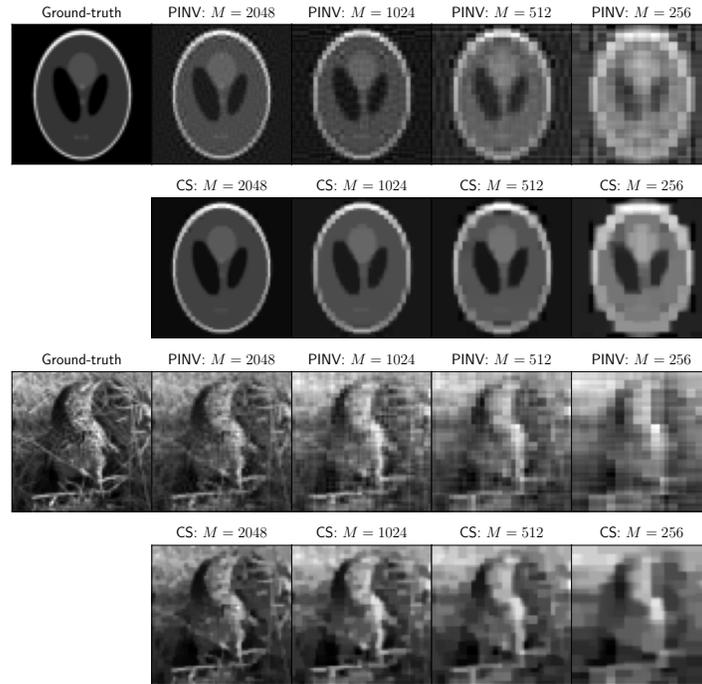


Figure 1.5 – Reconstruction à partir d’acquisitions sous échantillonnées. ’PINV’ : solution au sens des moindres carrés donnée par l’équation (1.11). ’CS’ : *compressed sensing*, solution obtenue en minimisant (1.13) avec $\Phi = [\nabla_x^T, \nabla_y^T]^T$, correspondant à la variation totale $\|\Phi f\|_1 = \|\nabla_x f\|_1 + \|\nabla_y f\|_1$. Les différentes colonnes considèrent différents taux d’échantillonnage $\frac{M}{N}$. Avec $N = 4096$, le taux d’échantillonnage décroît de gauche à droite ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ et enfin $\frac{1}{16}$).

$\mathbf{x} = \Psi\boldsymbol{\alpha}$ où $\boldsymbol{\alpha}$ est un vecteur contenant seulement quelques composantes non nulles (Candes and Wakin 2008). La reconstruction d’une image \mathbf{x} à partir de mesures $\mathbf{y} = \mathbf{A}\mathbf{x}$ peut alors s’effectuer en minimisant le nombre de composantes non nulles du vecteur $\boldsymbol{\alpha}$ sous la contrainte $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\Psi\boldsymbol{\alpha}$.

Ce problème est un problème d’optimisation combinatoire NP-difficile, dont la résolution est notoirement longue et complexe. Cependant, la théorie du *compressed*

sensing montre que, moyennant des hypothèses supplémentaires portant sur le modèle direct \mathbf{A} (notamment son incohérence), le problème de minimisation

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad \text{tel que} \quad \mathbf{y} = \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\alpha} \quad [1.12]$$

où $\|\cdot\|_1$ représente la norme ℓ_1 , permet de remonter au vecteur $\boldsymbol{\alpha}$ qui correspond à l'image \mathbf{x} . Ce résultat est important car le problème de minimisation ℓ_1 défini à l'équation (1.12) est significativement plus simple à résoudre que le problème combinatoire initial.

En présence de bruit, la contrainte $\mathbf{y} = \mathbf{A}\mathbf{x}$ n'est plus vérifiée et il est préférable d'adopter la formulation sans contrainte

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\boldsymbol{\Phi}\mathbf{x}\|_1, \quad [1.13]$$

où λ est un paramètre de régularisation et $\boldsymbol{\Phi}$ un opérateur de parcimonie (que l'on peut voir comme $\boldsymbol{\Psi}^{-1}$ quand il est défini). La norme ℓ_1 n'étant pas différentiable en 0, il est de coutume d'avoir recours à des méthodes proximales pour résoudre (1.13), parmi lesquelles on peut citer les méthodes de Bregman ou l'algorithme des directions alternées.

En dépit de la solidité de ses fondements théoriques, la théorie du *compressed sensing* souffre de limites pratiques (Adcock et al. 2017/ed). Les méthodes d'optimisation proximales sont par nature itératives, ce qui conduit à des temps de calculs pouvant s'avérer prohibitifs, même pour des problèmes de faible dimension. Par ailleurs, l'opérateur de parcimonie $\boldsymbol{\Phi}$ n'étant pas connu a priori, il est de coutume de choisir le gradient discret. On fait alors l'hypothèse que les variations de l'images sont localisées (image à faible variation totale), ce qui conduit à des artefacts typiques en marches d'escalier. Des méthodes plus avancées reposant sur la minimisation de la variation totale isotrope, de la variation totale d'ordre plus élevé ou encore du rang de la solution, permettent de pallier ce problème au prix de temps de calculs significativement plus longs. De plus, la plupart de ces algorithmes dépendent fortement du réglage d'un ou de plusieurs hyparamètres (paramètre de régularisation, pas de descente, nombres d'itération, etc.), dont les valeurs optimales dépendent, entre autre, de l'image en cours d'acquisition et du niveau du bruit de mesure.

Choix des motifs d'acquisition

Il est de coutume de choisir les motifs d'acquisition \mathbf{A} parmi une base de motifs \mathbf{H} , ce que l'on peut modéliser par

$$\mathbf{A} = \mathbf{S}_M \mathbf{H}, \quad [1.14]$$

où $\mathbf{S}_M \in \{0, 1\}^{M \times N}$ est une matrice de sous-échantillonnage qui sélectionne certaines des lignes de la base complète $\mathbf{H} \in \mathbb{R}^{N \times N}$. Par exemple, la matrice de sous-échantillonnage permettant de retenir respectivement les 2^e, 1^{er} et 6^e motifs d'une base de $\mathbb{R}^{8 \times 8}$ est donnée par

$$\mathbf{S}_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad [1.15]$$

Le choix de la matrice \mathbf{A} revient alors à choisir une base complète \mathbf{H} et une matrice de sous échantillonnage \mathbf{S}_M . Comme détaillé dans la section précédente, la base de Hadamard est un choix naturel en raison de ses propriétés de réduction du bruit, mais les bases de Fourier, de cosinus discret ou d'ondelettes sont des alternatives courantes dans le domaine. Le problème du choix de la matrice de sous-échantillonnage trouve une solution dans la théorie du *compressed sensing* qui prescrit l'emploi de matrices aléatoires (Candes and Wakin 2008). Un second résultat théorique intéressant démontre que l'erreur de reconstruction obtenue par une méthode de reconstruction linéaire est minimisée en choisissant des motifs qui maximisent l'énergie des mesures (Baldassarre et al. 2016). La matrice de sous-échantillonnage est alors obtenue en résolvant

$$\max_{\mathbf{S}_M} \sum_s \|\mathbf{S}_M \mathbf{H} \mathbf{x}_s\|_2^2, \quad [1.16]$$

où les \mathbf{x}_s , $1 \leq s \leq S$, représente les S images constitutives d'une base de données. Par chance, le problème posé par l'équation (1.16) peut être résolu très simplement. Il suffit, dans un premier temps, de calculer l'image $\sum_s (\mathbf{H} \mathbf{x}_s)^2$, où le carré est appliqué pixel à pixel, puis de retenir les M plus larges composantes de cette image d'énergie.

Des choix empiriques distincts, aux performances souvent honorables sont monnaie courante. De nombreuses études ont comparé les résultats obtenus en choisissant différentes bases et méthodes de sous-échantillonnage (Higham et al. 2018 ; Ochoa et al. 2018 ; Vaz et al. 2020). A la figure (1.6), nous comparons trois approches classiques de sous-échantillonnage de la base de Walsh-Hadamard . La première stratégie consiste en un échantillonnage aléatoire, la seconde à l'échantillonnage dit 'naturel' pour lequel les motifs de plus basses fréquences sont conservés, la troisième se base sur la maximisation de l'énergie des mesures tel que décrit par l'équation (1.16). L'échantillonnage aléatoire conduit visuellement aux images de plus basse qualité pour lesquelles on peine à distinguer l'objet principal de la scène. Le PSNR est modeste, typiquement de 15 à 20 dB plus faible qu'avec les deux autres approches. Ces deux dernières sont également nettement meilleures visuellement, avec la reconstruction de détails fins comme les mâts du bateau de la première image. En adéquation avec la théorie, l'approche basée sur l'énergie est en moyenne la meilleure des deux, avec un PSNR plus élevé de l'ordre de 2–3 dB.

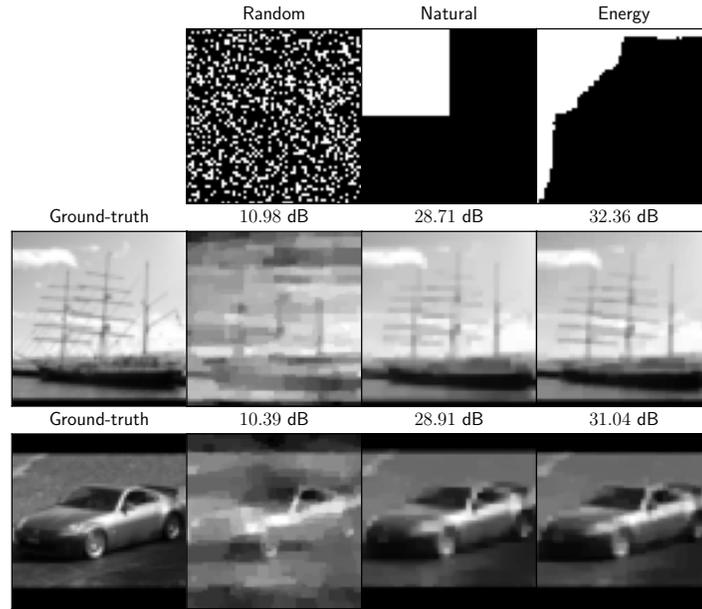


Figure 1.6 – Sous-échantillonnage des motifs d’acquisition, pour différentes méthodes. La première ligne représente un masque qui décrit les mesures retenues par la procédure d’échantillonnage (‘Random’ : échantillonnage aléatoire, ‘Naturel’ : basse fréquence, ‘Energy’ : solution de (1.16)). La base H choisie est celle de Walsh-Hadamard pour $N = 64 \times 64$ et l’on a fixé $M = \frac{1}{4}N$. Les reconstructions sont obtenues en résolvant (1.13).

1.4. Reconstruction par apprentissage profond

Principe général

Les méthodes de reconstruction par apprentissage profond tentent de reconstruire une image directement à partir des mesures, en utilisant une fonction non linéaire qui dépend de paramètres dont les valeurs doivent être judicieusement choisies (Barbasthis et al. 2019). Plus spécifiquement, en notant les mesures $\mathbf{y} \in \mathbb{R}^M$, nous obtenons l’image $\mathbf{x} \in \mathbb{R}^N$ par application de la fonction $\mathcal{G}_\omega : \mathbb{R}^M \rightarrow \mathbb{R}^N$, à savoir

$$\mathbf{x} = \mathcal{G}_\omega(\mathbf{y}), \quad [1.17]$$

où ω représente les paramètres de la fonction \mathcal{G} .

Dans le domaine de l'apprentissage profond, la fonction \mathcal{G}_ω est classiquement appelée modèle et correspond à un réseau de neurones. L'un des points forts de ce type d'approches est que l'évaluation du modèle est généralement rapide (de l'ordre de la milliseconde) ce qui conduit à une diminution significative des temps de reconstruction comparé aux méthodes issues de la théorie du *compressed sensing*.

Architecture du modèle

Les modèles les plus courants en apprentissage profond sont probablement les réseaux de neurones artificiels à propagation vers l'avant (en anglais *feedforward neural network*). Ceux-ci sont constitués de couches de neurones mises en cascade, ce qui peut être modélisé par

$$\mathcal{G}_\omega = \mathcal{G}_{\omega^L} \circ \dots \circ \mathcal{G}_{\omega^1} \quad [1.18]$$

où $\mathcal{G}_{\omega^\ell}$, $1 \leq \ell \leq L$ représente la ℓ^{e} couche du réseau et \circ la composition de fonctions. La sortie d'une couche est obtenue par combinaison linéaire de ses entrées, suivie d'une transformation non-linéaire. En notant le vecteur d'entrée d'une couche $\mathbf{y} \in \mathbb{R}^{M_{\text{in}}}$, sa sortie $\mathbf{x} \in \mathbb{R}^{M_{\text{out}}}$ est ainsi donnée par

$$\mathcal{G}_\omega(\mathbf{y}) = g(\mathbf{W}\mathbf{y} + \mathbf{b}) \quad [1.19]$$

où $\mathbf{W} \in \mathbb{R}^{M_{\text{out}} \times M_{\text{in}}}$ et $\mathbf{b} \in \mathbb{R}^{M_{\text{out}}}$ sont respectivement les poids et les biais de la couche en question, tandis que g est une fonction non-linéaire, classiquement dénommée fonction d'activation. Les valeurs des paramètres du réseau $\omega = (\mathbf{W}, \mathbf{b})$ sont déterminées durant la phase d'entraînement. En revanche, le nombre de couches, le nombre d'entrées et de sorties de chaque couche, la structure des matrices de poids et la nature des fonctions d'activation constituent l'architecture du réseau qui est définie, une fois pour toute, avant entraînement.

Il existe différents types de couches de neurones satisfaisant l'équation (1.19) mais qui diffèrent quant à la structure des connexions qui relient les entrées aux sorties. La figure 1.7 décrit les structures classiques des couches pleinement connectées et des couches de convolution ainsi que la fonction 'ReLU'. Cette dernière est une fonction d'activation typique qui seuille à zéro les valeurs négatives. Dans le cas des couches de neurones pleinement connectées, les matrices de poids \mathbf{W} sont pleines et requièrent donc $M_{\text{out}} \times M_{\text{in}}$ paramètres. En revanche, les couches de convolution considèrent des matrices de poids structurées, ce qui réduit considérablement le nombre de paramètres mis en jeu, typiquement, quelques filtres 3×3 indépendamment de la taille de l'image d'entrée.

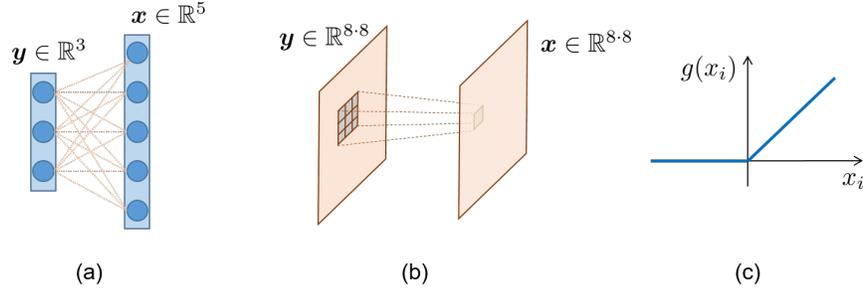


Figure 1.7 – Couches constitutives d'un réseau de neurones où y représente le signal d'entrée et x le signal de sortie. (a) Couche pleinement connectée pour laquelle toutes les sorties (c'est-à-dire toutes les composantes de x) sont connectées à toutes les entrées (c'est-à-dire toutes les composantes de y). Dans le cas représenté, il existe 5×3 connections soit 15 poids ω_i^j , $1 \leq i \leq 3$, $1 \leq j \leq 5$, caractérisant la couche. (b) Couche de convolution pour laquelle la sortie est obtenue en convoluant l'entrée par un filtre de taille fixée, ici 3×3 . Indépendamment de la taille des images d'entrée et de sortie, de taille 8×8 sur cette illustration, le nombre de poids caractérisant la couche sera donc limité à $3 \times 3 = 9$. (c) Fonction d'activation non linéaire, appliquée à la sortie d'une couche avant de la connecter à l'entrée de la couche suivante. La fonction représentée ici est le classique seuillage des valeurs négatives à zéro, connue sous le nom de fonction ReLU (pour *rectified linear unit*).

Comme représentée à la figure 1.8, la première couche d'un réseau de reconstruction profond est généralement une couche pleinement connectée qui associe les mesures $\mathbf{y} \in \mathbb{R}^M$ à une image $\hat{\mathbf{x}} \in \mathbb{R}^N$. Cette première couche est typiquement suivie de couches de convolution qui opèrent dans le domaine image et conduisent à la solution finale $\hat{\mathbf{x}}$.

Entraînement

Les paramètres du modèle sont obtenus avant reconstruction durant la phase d'entraînement. Durant cette phase, on cherche à minimiser la différence entre la sortie du modèle et les images issues d'une base de données. En notant $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_S]$ la base d'images et $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_S]$ les mesures correspondantes, telles que $\mathbf{y}_s = \mathbf{A}\mathbf{x}_s$, la phase d'entraînement minimise une fonction de coût \mathcal{J} (en anglais *loss function*) qui traduit la similarité entre les images originales et les images reconstruites. On considère classiquement l'erreur quadratique moyenne, soit

$$\min_{\omega} \mathcal{J}(\omega) = \min_{\omega} \frac{1}{S} \sum_{s=1}^S \|\mathcal{G}_{\omega}(\mathbf{y}_s) - \mathbf{x}_s\|_2^2. \quad [1.20]$$

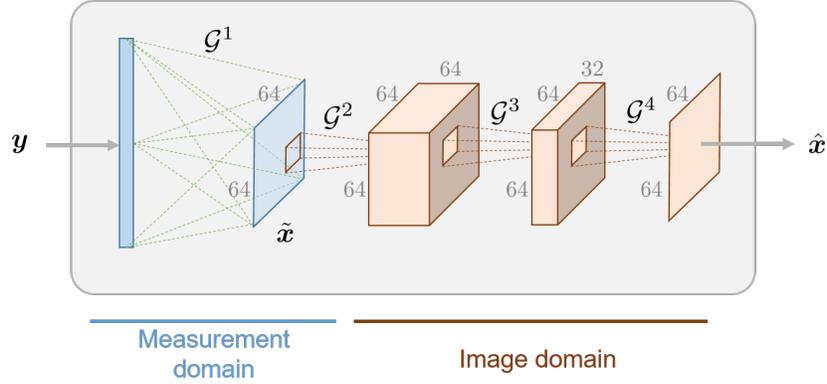


Figure 1.8 – Architecture d’un réseau de reconstruction simple. Ce réseau, composé de 4 couches, est inspiré de Higham et al. (2018). La première couche est une couche pleinement connectée agissant dans le domaine des mesures, suivies de 3 couches de convolution opérant dans le domaine image. Les sorties des couches de convolution sont toutes des images de tailles 64×64 mais leur nombre n^{filt} varie en fonction de la couche considérée. La valeur de n^{filt} correspond au nombre de noyaux de convolution utilisés dans chaque couche, qui vaut 64 pour la première couche, 32 pour la deuxième et 1 pour la troisième. Les noyaux des filtres de ce réseau sont tous de taille 3×3 .

Il est à noter que la fonction de coût est classiquement calculée comme la somme sur la base d’entraînement et, par conséquent, plus cette base est grande, plus les temps d’entraînement sont importants. Cependant, un grand nombre d’images est généralement requis pour obtenir un réseau capable de reconstruire avec une faible erreur des images ne faisant pas partie de la base d’entraînement. La phase d’entraînement peut donc être extrêmement longue (c’est-à-dire nécessiter plusieurs heures voire plusieurs semaines).

La majorité des algorithmes modernes que l’on utilise pour résoudre (1.20) sont des variantes de descentes de gradient stochastique, que l’on peut résumer par la mise à jour suivante :

$$\omega^{(k+1)} = \omega^{(k)} - \tau_k \nabla_{\omega} \mathcal{J}_{\text{batch}}^{(k)}(\omega), \quad [1.21]$$

où ∇_{ω} représente le gradient et $\mathcal{J}_{\text{batch}}^{(k)}$ la fonction de coût partielle, calculée en tenant compte d’un sous-ensemble seulement des paires images-mesures disponibles. En notant \mathbb{B}_k un sous-ensemble de B indices choisis aléatoirement parmi les S indices images possibles, nous avons pour l’erreur quadratique moyenne

$$\mathcal{J}_{\text{batch}}^{(k)}(\omega) = \frac{1}{B} \sum_{s \in \mathbb{B}_k} \|\mathcal{G}_{\omega}(\mathbf{y}_s) - \mathbf{x}_s\|_2^2. \quad [1.22]$$

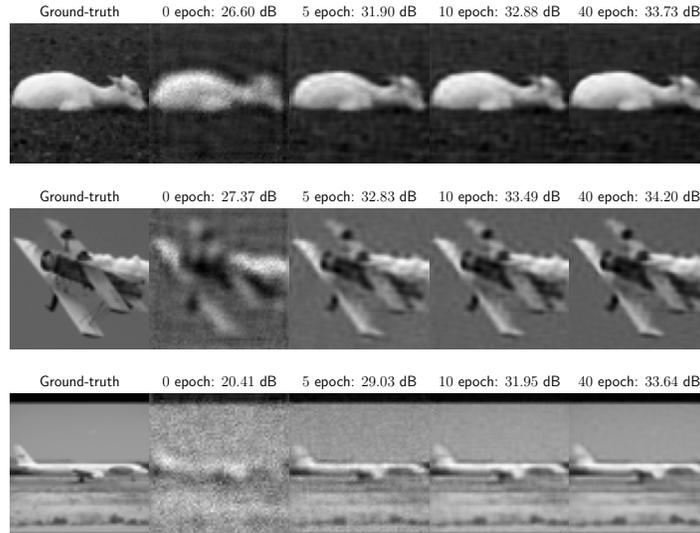


Figure 1.9 – Réseau de reconstruction à différents stades de la phase d’entraînement. On appelle ‘époque’ un cycle d’itérations au bout duquel toutes les images de la base de données ont été considérées une fois, c’est-à-dire prises en compte dans un minibatch tel que présenté à l’équation (1.22). On considère ici le réseau ‘freeNet’ décrit à la figure (1.8). On se place dans le cas $N = 64 \times 64$ et $M = \frac{1}{4}N$ et choisit d’initialiser τ_k à 10^{-3} avant de le multiplier par 0.25 toutes les 20 époques. Les paramètres du réseaux sont initialisés de façon aléatoire avant le début de l’entraînement.

Le paramètre τ_k introduit dans (1.21) est le pas de la descente, connu sous l’appellation *learning rate*, et choisi de façon à décroître au fil des itérations. De nombreuses variantes permettent d’accélérer la vitesse de convergence, notamment par l’introduction d’un moment d’inertie et d’une règle d’adaptation du pas de descente (voir Goodfellow et al. 2016, chap. 8). Il est à noter que les approches basées sur le gradient sont particulièrement populaires en raison de la disponibilité de l’algorithme de rétropropagation du gradient qui permet de calculer $\nabla_{\omega} \mathcal{J}_{\text{batch}}^{(k)}$ efficacement.

A la figure 1.9, nous représentons la sortie d'un réseau de neurones artificiels à différentes époques, c'est-à-dire à différents stades de la phase d'entraînement. On définit une 'époque' comme étant un cycle d'itérations au bout duquel toutes les images de la base de données ont été incluses dans la fonction de coût partielle de l'équation (1.22). On observe que l'objet est reconnaissable dès l'image reconstruite à la première époque ('0 epochs'). Le PSNR décroît ensuite significativement, atteignant des valeurs de l'ordre de 34 dB. Dès la 6^e époque, les PSNR sont déjà de l'ordre de 30 dB et l'image reconstruite avec une bonne fidélité. Les époques successives permettent de réduire les artefacts présents dans l'arrière plan de l'image mais aussi de rehausser les contrastes.

Lien simple avec les méthodes classiques

On peut appréhender le lien entre les méthodes de reconstruction classiques des sections 1.2 et 1.3 et les reconstructions obtenues par des réseaux de neurones en fixant les paramètres de certaines couches au lieu de les apprendre. Notamment, la sortie de la première couche, que l'on note $\tilde{\mathbf{x}} \in \mathbb{R}^N$, s'écrit

$$\tilde{\mathbf{x}} = \mathbf{W}^1 \mathbf{y} + \mathbf{b}^1, \quad [1.23]$$

en retirant la fonction d'activation, ce qui revient à choisir g^1 comme la fonction identité. A partir de cette remarque, il est possible d'obtenir, par identification avec des estimateurs linéaires classiques, des expressions analytiques pour les paramètres $\omega^1 = (\mathbf{W}^1, \mathbf{b}^1)$ de cette couche.

Tout d'abord, il est possible d'identifier les paramètres de l'équation (1.23) avec la solution pseudo-inverse donnée à l'équation (1.11). Dans le cas où les motifs de lumière sont orthogonaux, la première couche du réseau génère la solution pseudo inverse en choisissant

$$\mathbf{W}_{\text{pinv}}^1 = \mathbf{A}^\top \quad \text{et} \quad \mathbf{b}_{\text{pinv}}^1 = \mathbf{0} \quad [1.24]$$

Avant d'aller plus loin, notons $\tilde{\mathbf{y}} \in \mathbb{R}^N$ le vecteur mesures complet, défini tel que $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{x}$, où \mathbf{H} est la base d'acquisition complète, c'est-à-dire qui n'a pas été sous-échantillonnée. Ce vecteur est complet dans le sens où il contient à la fois les mesures $\mathbf{y} \in \mathbb{R}^M$ acquises par le système expérimental et celles $\mathbf{y}_2 \in \mathbb{R}^{N-M}$ qui ne l'ont pas été. La base complète \mathbf{H} peut être vue comme $\mathbf{H} = [\mathbf{A}, \mathbf{A}_2]$, où $\mathbf{A}_2 \in \mathbb{R}^{(N-M) \times N}$ représente les motifs de lumière qui n'ont pas été acquis.

Dans ce cadre, nous pouvons remarquer (cf. encadré ci-dessous) que la solution pseudo inverse $\tilde{\mathbf{x}} = \mathbf{A}^\top \mathbf{y}$ de l'équation (1.11) est également donnée par $\tilde{\mathbf{x}} = \mathbf{H}^\top \tilde{\mathbf{y}}_0$,

où $\tilde{\mathbf{y}}_0 = [\mathbf{y}, \mathbf{0}]$ est le vecteur d'acquisition où les données manquantes sont remplacées par des zéros.

REMARQUE.

$$\tilde{\mathbf{x}} = \mathbf{H}^\top \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}_2 \end{bmatrix}^\top \tilde{\mathbf{y}} = [\mathbf{A}^\top \ \mathbf{A}_2^\top] \tilde{\mathbf{y}} \quad [1.25]$$

$$= [\mathbf{A}^\top \ \mathbf{A}_2^\top] \begin{bmatrix} \mathbf{y} \\ \mathbf{y}_2 \end{bmatrix} = \mathbf{A}^\top \mathbf{y} + \mathbf{A}_2^\top \mathbf{y}_2 \quad [1.26]$$

Dès lors, on retrouve bien $\tilde{\mathbf{x}} = \mathbf{A}^\top \mathbf{y}$ en imposant $\mathbf{y}_2 = \mathbf{0}$. ■

Meilleur estimateur linéaire : complétion Bayésienne

Cette approche peut être généralisée en envisageant de compléter les mesures manquantes par des valeurs non nulles pertinentes $\tilde{\mathbf{y}}_2$ avant multiplication par la transposée de la base complète. Soit

$$\tilde{\mathbf{x}} = \mathbf{H}^\top \tilde{\mathbf{y}}, \quad \text{où} \quad \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}}_2 \end{bmatrix}. \quad [1.27]$$

Dans un cadre Bayésien, on peut estimer les mesures manquantes à partir de leur corrélation avec les mesures effectuées (Ducros [et al. 2020](#)), ce qui conduit à l'expression analytique suivante

$$\tilde{\mathbf{y}}_2 = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_1^{-1} (\mathbf{y} - \boldsymbol{\mu}_1), \quad [1.28]$$

où $\boldsymbol{\mu}_1$ et $\boldsymbol{\mu}_2$ représentent les moyennes des mesures, $\boldsymbol{\Sigma}_1$ et $\boldsymbol{\Sigma}_{21}$ leur covariance. Plus spécifiquement, $\boldsymbol{\mu}_1 \in \mathbb{R}^M$ est la moyenne des mesures effectuées, $\boldsymbol{\mu}_2 \in \mathbb{R}^{N-M}$ la moyenne des mesures manquantes, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{M \times M}$ la variance des mesures effectuées et $\boldsymbol{\Sigma}_{21} \in \mathbb{R}^{N-M \times M}$ la covariance entre les mesures manquantes et celles acquises. Une illustration géométrique de l'équation (1.28) est donnée à la figure 1.10 qui illustre la corrélation entre la 2^e la 7^e composante du vecteur de Hadamard complet $\tilde{\mathbf{y}}$. Dans cette exemple, si l'on mesure $\tilde{y}_2 = 20$, on fixera $\tilde{y}_7 = 9$ plutôt que $\tilde{y}_7 = 0$ si cette composante n'a pu être mesurée.

Les différentes quantités impliquées dans l'équation (1.27) peuvent être estimées à partir d'une base d'images $\{\mathbf{x}_s\}$, $1 \leq s \leq S$, où S est le nombre d'images contenues

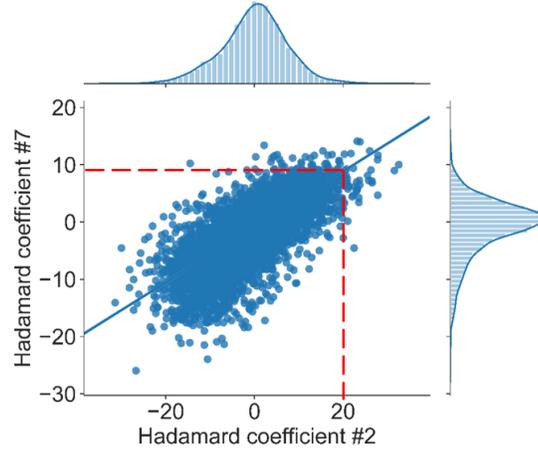


Figure 1.10 – Interprétation géométrique de la complétion Bayésienne. Chaque point représente une image de la base STL-10, dont la position selon l'axe des abscisses (respectivement des ordonnées) dépend de la valeur de la 2^e (respectivement 7^e) composante du vecteur de Hadamard complet $\tilde{y}_s = Hx_s$. La pente de la droite bleue en trait plein traduit la corrélation entre ces deux composantes.

dans la base, en utilisant les estimateurs classiques de la moyenne

$$\mu_1 = \frac{1}{S} \sum_s A x_s \quad [1.29]$$

$$\mu_2 = \frac{1}{S} \sum_s A_2 x_s \quad [1.30]$$

et de la covariance

$$\Sigma_1 = \frac{1}{S-1} \sum_s (A x_s - \mu_1)(A x_s - \mu_1)^\top \quad [1.31]$$

$$\Sigma_{21} = \frac{1}{S-1} \sum_s (A_2 x_s - \mu_2)(A x_s - \mu_1)^\top \quad [1.32]$$

Les équations (1.27) et (1.28) étant linéaires, il est possible d'obtenir une reconstruction par complétion Bayésienne au moyen d'une couche de neurones dont les paramètres sont bien choisis. En substituant (1.28) dans (1.27), puis par identification avec (1.23), nous obtenons

$$W_{\text{comp}}^1 = H^\top \begin{bmatrix} I_M \\ \Sigma_{21} \Sigma_1^{-1} \end{bmatrix} \quad \text{et} \quad b_{\text{comp}}^1 = H^\top \begin{bmatrix} \mathbf{0}_M \\ \mu_2 - \Sigma_{21} \Sigma_1^{-1} \mu_1 \end{bmatrix}. \quad [1.33]$$

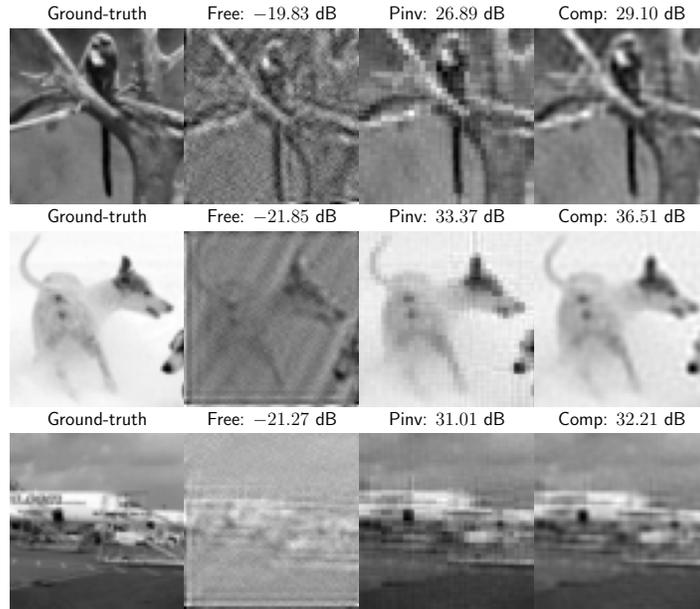


Figure 1.11 – Sortie de la première couche pleinement connectée d'un réseau de reconstruction profond. 'Free' couche libre, c'est-à-dire dont les paramètres sont appris lors de l'entraînement ; 'Pinv' pseudo-inverse donnée par (1.24), 'Comp' complétion des mesures manquantes donnée par (1.33). Lors de l'entraînement de 'Free', les couches en aval de la première sont celles décrites à la figure 1.8. Nous reportons au dessus de chaque image reconstruite son PSNR exprimé en décibels.

En l'absence de corrélation entre les mesures acquises et manquantes, c'est-à-dire sous l'hypothèse $\Sigma_{21} = \mathbf{0}$, et lorsque les mesures sont à moyennes nulles, c'est-à-dire pour $\mu_2 = \mathbf{0}$, l'équation (1.33) revient alors à l'expression simple de la pseudo inverse donnée par l'équation (1.24).

À la figure 1.11, nous comparons, pour quelques images, la solution pseudo-inverse, celle obtenue par complétion Bayésienne, ainsi que celle obtenue en apprenant la première couche $\omega^1 = (\mathbf{W}^1, \mathbf{b}^1)$. Dans ce dernier cas, la couche pleinement connectée ω^1 est apprise conjointement avec les couches de convolution en aval ω^ℓ ,

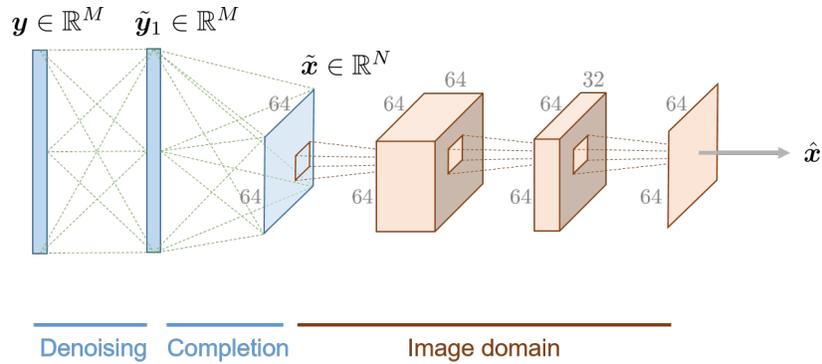


Figure 1.12 – Réseau de complétion débruité (DC-Net pour *denoised completion network*). La première couche est une couche pleinement connectée qui calcule l’image de plus faible erreur quadratique moyenne, au moyen de la formule analytique indiquée à l’équation (1.37). Comme pour le réseau introduit à la figure 1.8, la solution approchée $\tilde{\mathbf{x}}$ est corrigée par des couches de convolution mises en cascade, ce qui a pour effet de corriger les artefacts résiduels dans le domaine image.

$2 \leq \ell \leq 4$. Il est intéressant de noter que la première couche apprise fournit effectivement une solution approchée dans le domaine image. Le PSNR est particulièrement faible en raison d’une mauvaise mise à l’échelle ; cependant, on reconnaît aisément pour les trois images reconstruites les objets présents sur l’image originale. La solution pseudo inverse améliore très significativement la qualité de reconstruction, de même que la solution complétée qui offre, visuellement et en termes de PSNR, la meilleure qualité de reconstruction.

Prise en compte du bruit

Nous nous intéressons désormais à l’influence du bruit de mesure sur la qualité de reconstruction. En particulier, nous considérons le bruit de Poisson intrinsèque à toute mesure de signal lumineux. Les mesures bruitées \mathbf{y}^α sont données par

$$\mathbf{y}^\alpha \sim \frac{1}{\alpha} \mathcal{P}(\alpha \mathbf{y}), \quad [1.34]$$

où α est un facteur multiplicatif qui pilote le niveau de bruit et \mathcal{P} représente la loi de Poisson. Plus α est grand, plus le nombre de photons collectés est grand et meilleur est le rapport signal sur bruit des mesures. La division par α vise simplement à ramener les données dans une même plage de variation, quel que soit le niveau de bruit. En approchant la loi de Poisson par une loi normale, nous avons

$$\mathbf{y}^\alpha \sim \mathbf{y} + \mathcal{N}(\mathbf{0}, \Sigma_\alpha), \quad [1.35]$$

où $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ représente la loi normale de moyenne $\boldsymbol{\mu}$ et de variance $\boldsymbol{\Sigma}$ et $\boldsymbol{\Sigma}_\alpha \in \mathbb{R}^{M \times M}$ est une matrice de covariance qui dépend de la mesure. Le bruit considéré ici n'est donc pas additif.

Il est possible de généraliser l'idée de complétion Bayésienne en présence de bruit. Pour cela, il convient de remplacer les mesures brutes \mathbf{y} à l'équation (1.27) par une version débruitée $\tilde{\mathbf{y}}_1$ de ces mesures, conduisant ainsi à

$$\tilde{\mathbf{x}} = \mathbf{H}^\top \tilde{\mathbf{y}} \quad \text{avec} \quad \tilde{\mathbf{y}} = \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \end{bmatrix} \quad [1.36]$$

Les mesures débruitées et complétées sont données par (Lorente Mur et al. 2021)

$$\tilde{\mathbf{y}}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 [\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_\alpha]^{-1} (\mathbf{y}^\alpha - \boldsymbol{\mu}_1), \quad [1.37a]$$

$$\tilde{\mathbf{y}}_2 = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_1^{-1} [\tilde{\mathbf{y}}_1 - \boldsymbol{\mu}_1], \quad [1.37b]$$

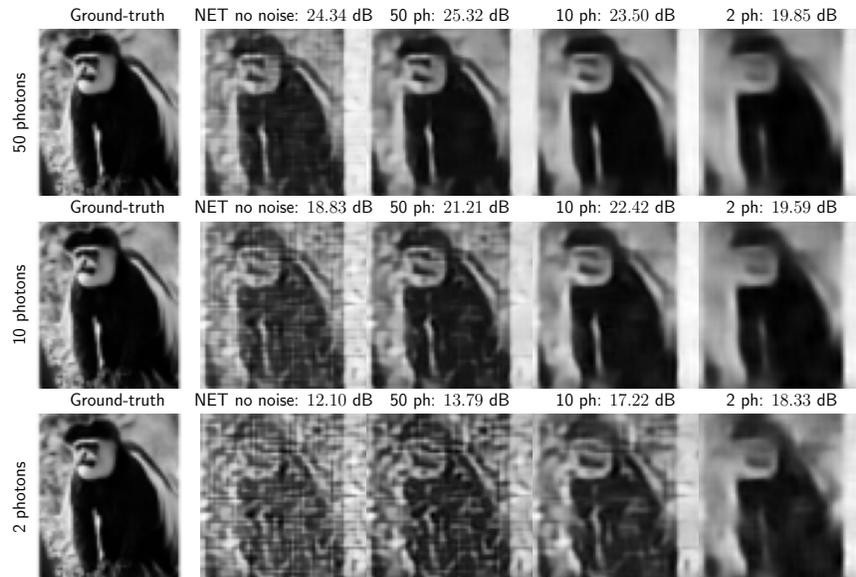
où l'équation (1.37a) correspond au débruitage des mesures acquises alors que l'équation (1.37b) correspond à la complétion des mesures manquantes. En comparant les équations (1.37b) et (1.28), on peut remarquer que les étapes de complétion en présence et en l'absence de bruit sont identiques (addition et soustraction des moyennes et multiplication par $\boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_1^{-1}$). Seules changent les données auxquelles s'appliquent ces opérations : mesures brutes \mathbf{y} dans un cas, mesures débruitées $\tilde{\mathbf{y}}$ dans l'autre. Cette solution est celle de plus faible erreur quadratique moyenne (en anglais MMSE pour *minimum mean squared error*) pouvant être obtenue par un estimateur linéaire.

Par identification entre (1.23), (1.36) et (1.37), nous pouvons choisir les paramètres de la première couche du réseau permettant de calculer la solution de moindre erreur quadratique moyenne. Ceux-ci sont donnés par

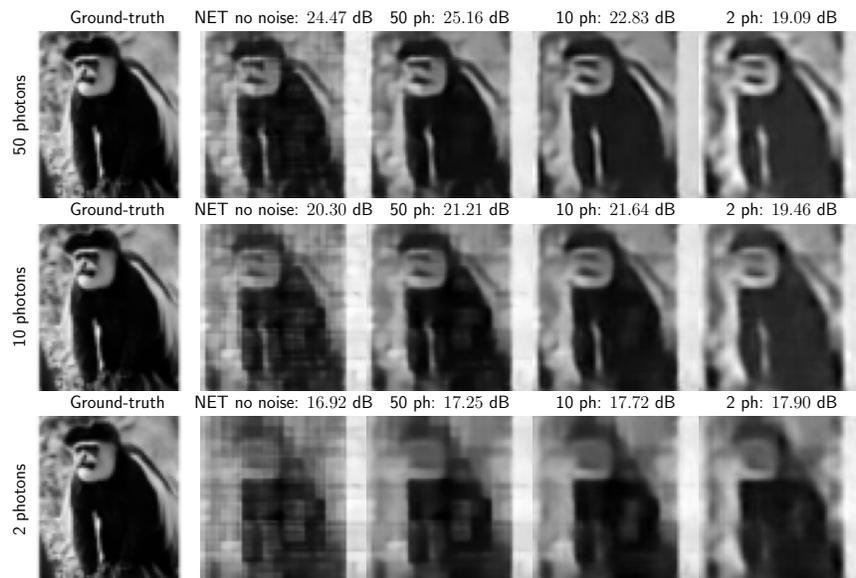
$$\mathbf{W}_{\text{mmse}}^1 = \mathbf{H}^\top \begin{bmatrix} \boldsymbol{\Sigma}_1 [\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_\alpha]^{-1} \\ \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_1^{-1} \end{bmatrix} \quad \text{et} \quad [1.38a]$$

$$\mathbf{b}_{\text{mmse}}^1 = \mathbf{H}^\top \begin{bmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_1 [\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_\alpha]^{-1} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}_1 \end{bmatrix}. \quad [1.38b]$$

A la figure 1.13, nous étudions l'influence de bruit de mesure sur la qualité de reconstruction. Nous observons qu'un réseau de reconstruction entraîné sans bruit verra ses performances décroître rapidement en présence de mesures bruitées (cf. deuxième colonne de figure 1.13). Pour $\alpha = 2$ photons, l'objet de la scène est difficilement discernable, ce qui illustre la nécessité d'entraîner le réseau en présence de bruit. Pour un niveau de bruit de mesure donné, le réseau le plus performant est celui entraîné pour le même niveau de bruit. A titre d'exemple, pour des mesures à $\alpha = 10$ photons, le meilleur réseau est celui entraîné à 10 photons (PSNR de 22.42 dB) alors que ceux entraînés à 50 et 2 photons conduisent à des PSNR réduits de 21.21 et 19.59 dB,



(a) C-Net



(a) DC-Net

Figure 1.13 – Influence du niveau de bruit pour deux architecture de réseaux : (a) C-Net défini par l'équation (1.33) (b) DC-Net défini par l'équation (1.38). Chacune des figures (a) et (b) rassemble des reconstructions obtenues à partir de différents niveaux de bruit de mesure et d'entraînement. Chaque ligne correspond à un niveau de bruit de mesure. Le rapport signal sur bruit augmente de haut en bas : $\alpha = 50$ photons pour la première ligne, $\alpha = 10$ photons pour la deuxième et $\alpha = 2$ photons pour la troisième. Chaque colonne présente les reconstructions obtenues par des réseaux entraînés à différents niveaux de bruit : 'no noise' sans bruit, '50 ph' pour $\alpha = 50$ photons, '10 ph' pour $\alpha = 10$ photons et '2 ph' pour $\alpha = 2$ photons.

respectivement. Visuellement, un réseau entraîné avec un niveau de bruit trop faible conduit à de nombreux artefacts, alors qu'un réseau entraîné avec un niveau de bruit trop élevé lisse excessivement la reconstruction.

1.5. Conclusion

Ce chapitre dresse un tour d'horizon succinct de l'imagerie computationnelle monodétecteur, depuis l'optique par transformée de Hadamard jusqu'aux approches récentes fondées sur l'apprentissage profond. L'optique par transformée de Hadamard est la première à avoir proposé d'acquérir une image en combinant les signaux issus de différents pixels de la scène sur un détecteur ponctuel. Sous l'hypothèse d'un bruit additif Gaussien, cette approche conduit à améliorer significativement le rapport signal sur bruit de l'image acquise, en comparaison d'une approche naïve par balayage. La théorie du *compressed sensing* a ensuite offert une solution élégante pour reconstruire une image à partir d'un faible nombre de mesures, ce qui permet de réduire fortement le nombre de mesure et donc les temps d'acquisition. Plus récemment, les approches par apprentissage ont permis de réduire considérablement les temps de reconstruction tout en améliorant la qualité des images obtenues.

Comme nous avons tâché de le montrer, il existe des liens étroits entre les méthodes de reconstruction classiques celle plus récentes qui sont fondées sur l'apprentissage. En particulier, la première couche d'un réseau de reconstruction peut être choisie de façon à implémenter des solutions classiques en problèmes inverses, telles la solution pseudo inverse ou celle de moindre erreur quadratique moyenne. Les couches en aval peuvent alors être apprises de façon à corriger les artefacts résiduels.

Les réseaux de reconstruction offrent actuellement les meilleures performances, que ce soit visuellement ou en termes de PSNR. Lorsque le bruit photonique devient significatif, typiquement en deçà de 50 photons par pixel, il convient d'entraîner le réseau en présence de bruit. Le niveau de bruit d'entraînement doit être comparable à celui des mesures, sous peine d'une dégradation sensible des performances. La robustesse d'un réseau vis-à-vis d'un niveau de bruit inconnu fait actuellement partie des thématiques de recherche de différents groupes. Il s'agit d'un vaste champ d'étude, nécessitant un chapitre à lui seul. Parmi ces méthodes on pourra notamment citer celles qui interprètent les algorithmes itératifs classiques comme les couches successives d'un réseau. La phase d'apprentissage permet alors de régler les hyperparamètres, de déterminer un terme de régularisation ou son gradient, d'approcher un opérateur proximal. Pour approfondir le sujet, le lecteur pourra s'orienter vers la revue d'[Arridge et al. \(2019\)](#).

1.6. Remerciements

Je tiens à remercier Corinne Fournier, Antonio Lorente Mur, Laurent Mahieu Williame ainsi que Françoise Peyrin, qui ont eu la gentillesse de se prêter à une relecture attentive.

1.7. Bibliographie

- Adcock, B., Hansen, A. C., Poon, C., Roman, B. (2017/ed), BREAKING THE COHERENCE BARRIER : A NEW THEORY FOR COMPRESSED SENSING, Forum of Mathematics, Sigma, 5.
- Arridge, S., Maass, P., Öktem, O., Schönlieb, C.-B. (2019), Solving inverse problems using data-driven models, Acta Numerica, 28, 1–174.
- Baldassarre, L., Li, Y.-H., Scarlett, J., Gözcü, B., Bogunovic, I., Cevher, V. (2016), Learning-Based Compressive Subsampling, IEEE Journal of Selected Topics in Signal Processing, 10(4), 809–822.
- Barbastathis, G., Ozcan, A., Situ, G. (2019), On the use of deep learning for computational imaging, Optica, 6(8), 921–943.
- Candes, E. J., Wakin, M. B. (2008), An Introduction To Compressive Sampling, IEEE Signal Processing Magazine, 25(2), 21–30.
- Chan, W. L., Charan, K., Takhar, D., Kelly, K. F., Baraniuk, R. G., Mittleman, D. M. (2008), A single-pixel terahertz imaging system based on compressed sensing, Applied Physics Letters, 93(12), 121105.
- Donoho, D. L. (2006), Compressed sensing, IEEE Trans. Inform. Theory, 52, 1289–1306.
- Duarte, M., Davenport, M., Takhar, D., Laska, J., Sun, T., Kelly, K., Baraniuk, R. (2008), Single-Pixel Imaging via Compressive Sampling, Signal Processing Magazine, IEEE, 25(2), 83–91.
- Ducros, N., Lorente Mur, A., Peyrin, F. (2020), A completion network for reconstruction from compressed acquisition, in IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, Iowa City, United States, pp. 619–623.
- Edgar, M. P., Gibson, G. M., Padgett, M. J. (2019), Principles and prospects for single-pixel imaging, Nature Photonics, 13(1), 13–20.
- Goodfellow, I., Bengio, Y., Courville, A. (2016), Deep Learning, MIT Press.
- Higham, C. F., Murray-Smith, R., Padgett, M. J., Edgar, M. P. (2018), Deep learning for real-time single-pixel video, Scientific Reports, 8(1), 2369.
- Lorente Mur, A., Lorente Mur, A., Leclerc, P., Peyrin, F., Ducros, N., Ducros, N. (2021), Single-pixel image reconstruction from experimental data using neural networks, Optics Express, 29(11), 17097–17110.

- Nelson, E. D., Fredman, M. L. (1970), Hadamard Spectroscopy, *JOSA*, 60(12), 1664–1669.
- Ochoa, M., Pian, Q., Yao, R., Ducros, N., Intes, X. (2018), Assessing patterns for compressive fluorescence lifetime imaging, *Opt. Lett.*, 43(18), 4370–4373.
- Pian, Q., Yao, R., Sinsuebphon, N., Intes, X. (2017), Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging., *Nature photonics*, 11, 411–414.
- Radwell, N., Mitchell, K. J., Gibson, G. M., Edgar, M. P., Bowman, R., Padgett, M. J. (2014), Single-pixel infrared and visible microscope, *Optica*, 1(5), 285–289.
- Rousset, F., Ducros, N., Peyrin, F., Valentini, G., D'Andrea, C., Farina, A. (2018), Time-resolved multispectral imaging based on an adaptive single-pixel camera, *Opt. Express*, 26(8), 10550–10558.
- Sloane, N. J. A., Harwit, M. (1976), Masks for Hadamard transform optics, and weighing designs, *Applied Optics*, 15(1), 107–114.
- Studer, V., Bobin, J., Cahid, M., Mousavi, H. S., Candes, E., Dahan, M. (2012), Compressive fluorescence microscopy for biological and hyperspectral imaging., *Proceedings of the National Academy of Sciences of the United States of America*, 109(26), E1679–E1687.
- Vaz, P. G., Amaral, D., Ferreira, L. F. R., Morgado, M., Morgado, M., Cardoso, J. (2020), Image quality of compressive single-pixel imaging using different Hadamard orderings, *Optics Express*, 28(8), 11666–11681.