

MATH0461-2  
INTRODUCTION TO NUMERICAL OPTIMIZATION  
Quentin Louveaux

---

# Project: compressed sensing

---

Stephan Champaviller  
Hubar Julien

*Master 1 Data science Engineering*  
*Faculty of Applied Sciences*

Academic years 2020-2021

# Contents

1	Introduction . . . . .	2
2	Modelling . . . . .	2
2.1	$\ell_0$ norm . . . . .	2
2.2	Formulation of the problem using the $\ell_1$ norm . . . . .	3
2.3	Formulation of the problem using the $\ell_2$ norm . . . . .	3
2.4	Closed-form solution to the $\ell_2$ norm problem . . . . .	4
2.5	Robust variants of the $\ell_1$ norm problem . . . . .	4
3	Numerical Experiments . . . . .	5
3.1	Numerical resolution of the problems . . . . .	5
3.2	Interpretation of the dual variables associated with equality constraints in the $\ell_1$ norm . . . . .	6
3.3	Numerical resolution of the $\ell_1$ norm and its robust variants for all noise measurements . . . . .	6
3.4	First variant . . . . .	6
3.5	Second variant . . . . .	7
3.6	Third variant . . . . .	9
3.7	Conclusion . . . . .	10
	<b>Annexes</b>	<b>12</b>
1	Result of non robust with $l_1$ norm . . . . .	13
	<b>Bibliography</b>	<b>16</b>

# 1 Introduction

Compressive sensing exploit the possibility to represent an image with a sparse representation. Either  $x$  an image or a signal, it can be decomposed into a product of a Fourier matrix  $\Psi$  (or universal transform basis) with a sparse  $s$  vector. As soon as this decomposition exists it is possible to express the measure of this signal as the following product.

$$y = C\Psi s = \Theta s \quad (1.1)$$

Where  $C$  is the measurement matrix, it makes it possible to make a link between the measurement and the product of the matrix  $\Psi$  and  $s$ . This system is an under-determined system.

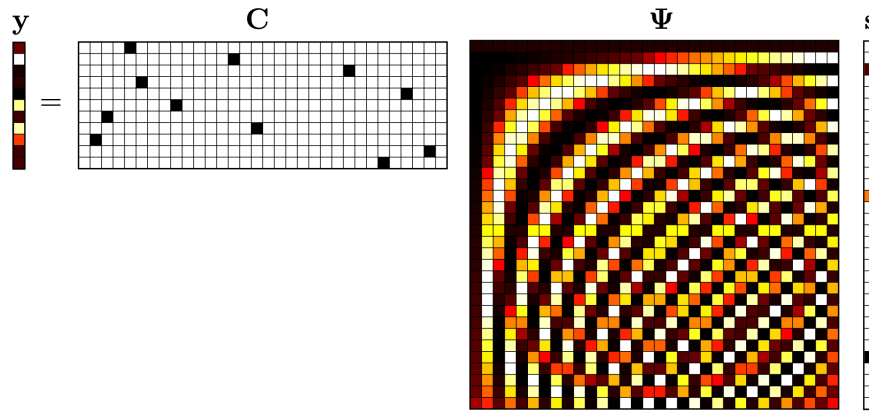


FIGURE 1: Schematic of measurements in the compressed sensing framework(Image credit to Data Driven Science & Engineering [1])

In the datasets we were provided, the transform matrix looks like a Fourier matrix and the measurement matrix seems like a random matrix.

## 2 Modelling

### 2.1 $\ell_0$ norm

The problem of reconstructing the sparse vector can be stated as an  $\ell_0$  problem as follows:

$$\begin{aligned} \min \quad & \|s\|_0 \\ \text{s.t.} \quad & \|\Theta s - y\|_2 = 0 \\ & s \in \mathcal{R}^N \end{aligned}$$

Using the  $\ell_0$  norm amounts to finding the smallest number of ones in the  $s$  vector that satisfies the constraints. This is an NP complete problem (according to literature ).

#### 2.1.1 Proof of non-convexity

For this problem to be convex, we check the conditions described on slide 20 of lecture "Introduction to nonlinear optimization". That is :

- $m = \Phi\Psi x$  is linear. This is true by construction.

- $\ell_0(x)$  is convex. This is not true. Let's analyse its epigraph (slide 19). We give a counter example to the convexity. Let's choose  $x_1 = (0, 0, 0, \dots)$  and  $x_2 = (1, 0, 0, \dots)$   $x_1, x_2 \in \mathbb{R}^N$  which both belong to the epigraph and  $\lambda = \frac{1}{2}$ .

$$\begin{aligned}
 |\lambda x_1 + (1 - \lambda)x_2| &\leq \lambda|x_1| + (1 - \lambda)|x_2| \\
 |0 + (1 - \lambda)x_2| &\leq \lambda \times 0 + (1 - \lambda) \times 1 \\
 1 &\leq 1 - \lambda \\
 0 &\leq -\frac{1}{2}
 \end{aligned}$$

## 2.2 Formulation of the problem using the $\ell_1$ norm

When the problem is expressed using the  $\ell_1$  norm. This can be written as follows:

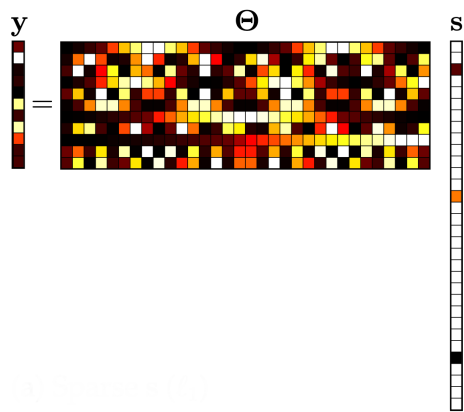
$$\begin{aligned}
 &\min \|s\|_1 \\
 &\text{s.t } \Theta s - y = 0 \\
 &\quad s \in \mathcal{R}^N
 \end{aligned}$$


FIGURE 2:  $\ell_1$  minimum norm solutions to compressed sensing problem. (Image credit to Data Driven Science & Engineering [1])

Thanks to  $\ell_1$  norm this will allow us to have the most sparse  $s$ . This problem is can be represented as a linear program, provided we translate the norm to appropriate constraints inequalities.

## 2.3 Formulation of the problem using the $\ell_2$ norm

The problem is expressed using the  $\ell_2$  norm like so:

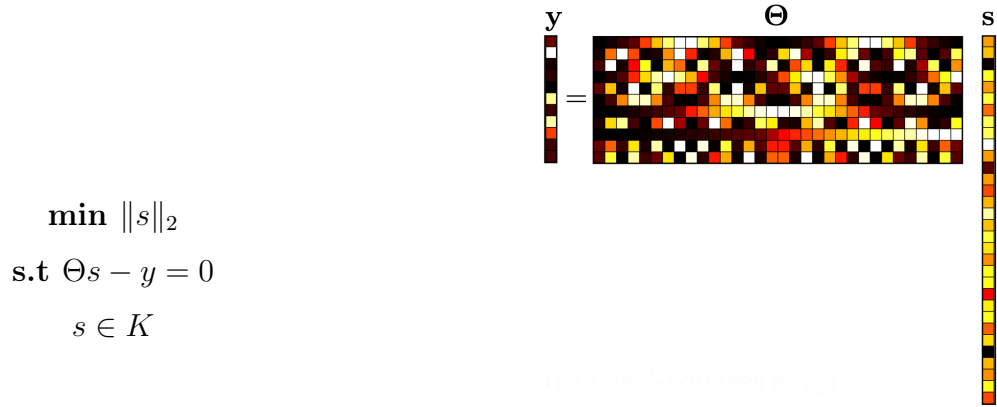


FIGURE 3:  $\ell_2$  minimum norm solutions to compressed sensing problem. (Image credit to Data Driven Science & Engineering [1])

We understand the  $\ell_2$  norm is not the best choice of norm for our problem because it leads to greater probability of finding non sparse solutions (compared to  $\ell_1$  norm).

## 2.4 Closed-form solution to the $\ell_2$ norm problem

We can represent the problem using an  $\ell_2$  norm with a conic program:

$$\begin{aligned} \min \quad & x_0 \\ \text{s.t.} \quad & \Theta s - y = 0 \\ & x \in K \end{aligned}$$

$K$  is a closed, convex, pointed cone :

$$K = \mathbb{L}^n = \{(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1} | x_1^2 + \dots + x_n^2 \leq x_0^2, x_0 \geq 0\}$$

Where  $n$  is the size of our sparse vector. We can say that the dual problem is bounded and strictly feasible.

## 2.5 Robust variants of the $\ell_1$ norm problem

Robust variants are here to account for noise. When noise is present in the measurements, it makes the goal of finding a sparse solution harder because the solution vectors  $s$  are subject to noise as well (so they are less sparse).

### 2.5.1 First approach

A first robust variant we devised is simply to allow some noise on the  $s$  vector, element wise.

$$\begin{aligned} \min \quad & \|s\|_1 \\ \text{s.t.} \quad & |(\Theta s - y)_i| \leq \epsilon, i \in \{1..N\} \text{ (i denotes the } i^{\text{th}} \text{ element)} \\ & s \in \mathcal{R}^n \end{aligned}$$

### 2.5.2 LASSO

By moving from a system of linear equations to some norm, we allow for a trade off between optimality of the solution of the equations and sparsity of the  $s$  vector (we understand this is known as the problem of "basis pursuit denoising"). So we implemented the LASSO algorithm like this :

$$\begin{aligned} \min \quad & \|s\|_1 \\ \text{s.t.} \quad & \|\Theta s - y\|_2 \leq \epsilon \\ & s \in \mathcal{R}^N \end{aligned}$$

Which translates directly into a conic program.

### 2.5.3 Approximation

We found and implemented another kind of algorithm:

$$\begin{aligned} \min \quad & \|\Theta s - y\|_2 \\ \text{s.t.} \quad & \|s\|_1 \leq \tau \\ & s \in \mathcal{R}^N \end{aligned}$$

Where  $\tau$  is tuning parameter. Contrary to the previous variant, we didn't figure out a intuitive explanation of the  $\tau$  parameter and so we had to figure its values experimentally.

## 3 Numerical Experiments

### 3.1 Numerical resolution of the problems

#### 3.1.1 Reconstruction of images

We used the  $\Psi * s$  product to reconstruct the images.

#### 3.1.2 $\ell_1$ norm for all non-corrupted measurements

As expected using  $\ell_1$  norm and a noiseless measurement matrix we find our 4 images(see [fig4](#)).

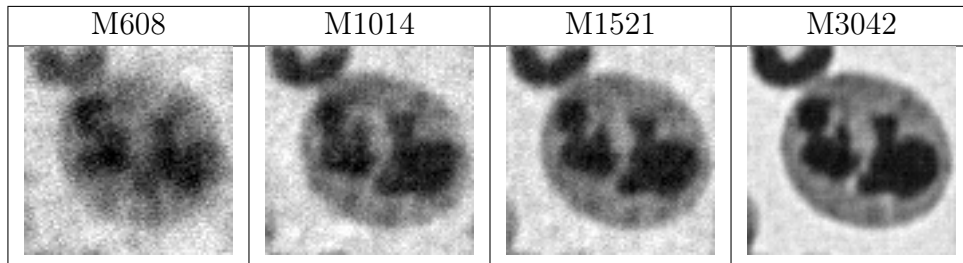


FIGURE 4: Decompression of an image of a neutrophilic polynuclear cell using  $\ell_1$  norm.

It emerges from experience that the larger the matrix, the sharper the image found. The compilation times for the 4 matrices increases enormously with the size of the matrix. Note however that the increase seems to be stagnating from matrix M1521.

### 3.1.3 $\ell_2$ norm for all non-corrupted measurements

Using the  $\ell_2$  norm leads to non sparse vector  $s$ (see [fig3](#)). This has the effect of returning a completely noisy image(see [fig5](#)). So,  $s$  is worse than non sparse, the values of its components triggers higher frequencies waves which lead to noise.

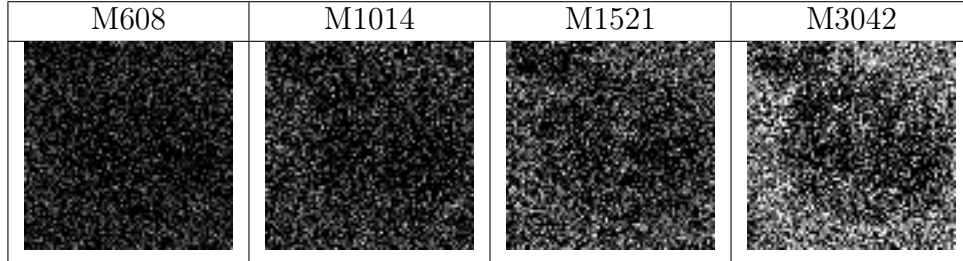


FIGURE 5: Decompression of an image of a neutrophilic polynuclear cell using  $\ell_2$  norm.

## 3.2 Interpretation of the dual variables associated with equality constraints in the $\ell_1$ norm

When solving the system, the constraint imposed forces the objective function to return the same values for the primal and the dual(due to the duality theorem). It is interesting to note that as the iterations go on, the values equalize(see in appendices [fig13](#), [fig14](#), [fig15](#) and [fig16](#)).

## 3.3 Numerical resolution of the $\ell_1$ norm and its robust variants for all noise measurements

By relaxing our constraints of optimization it is possible to solve the problem even with noisy measurements. This allows to find the following images(see [fig4](#))

### 3.4 First variant

The first variant offers images similar to what was obtained in [section3.1.2](#). Of course, due to the noise, their quality is not that good. The choice of  $\epsilon$  affects the reconstruction quality. The more we allow for noise, the less the quality. Note that when  $\epsilon = 0.01$ , it means (in our code) that we allow 1% noise. When we look at the noise statistics in the measurements (by comparing uncorrupted and noisy measurements), we see that the actual noise is about 10% of the signal.

Execution times are significantly reduced. Especially for larger matrices where the computation time is divided by 2 (see [fig12](#)). On the other hand, the number of iterations is increased. Please note that the processing time is too long for the last images(M3042). This leads to the search for a better optimization.

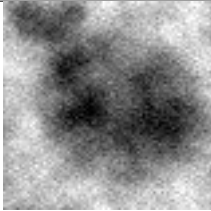
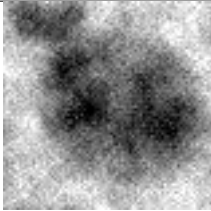
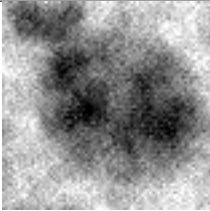
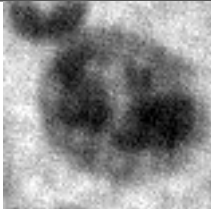
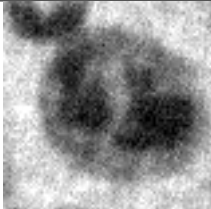
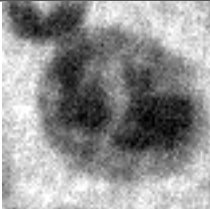
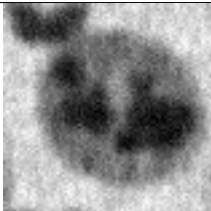
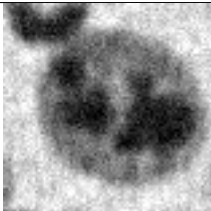
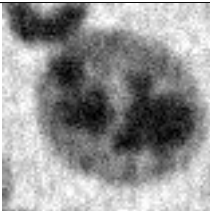
Measurements	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0.001$
M608			
M1014			
M1521			
M3042	Out of time	Out of time	Out of time

FIGURE 6: Results of the first variant experiments with all the provided noisy measurements and associated  $\epsilon$

Looking at the plots of *fig7* we can see that:

- The error histogram (computed on the image obtained by subtracting the reconstructed image from a reference image) is more ore less the same for the two  $\epsilon$ .
- The smaller value of the  $\epsilon$  (0.001) pushes the optimizer towards less sparse solutions.

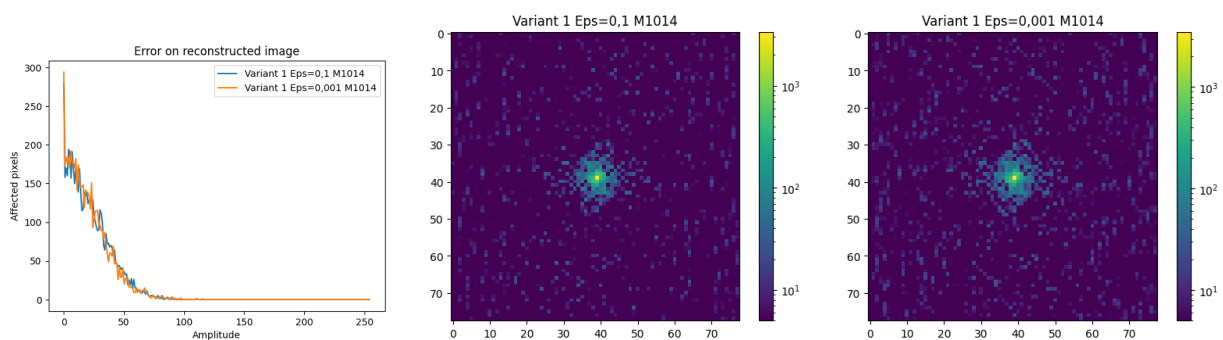


FIGURE 7: Reconstruction error and Fourier transform for two values of  $\epsilon$  for M1014 measurements.

### 3.5 Second variant

We observe :

- On the M1521 and M3042 experiments, the reconstruction is better if the  $\epsilon$  is smaller. Which is expected since  $\epsilon$  bounds the linear equations.



- The higher number of measurements makes the reconstruction reliably better.

Looking at the plots of *fig9*, we see that a greater  $\epsilon$  allows to find a sparser solution.

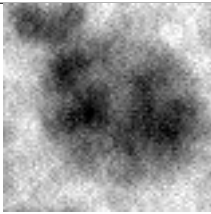
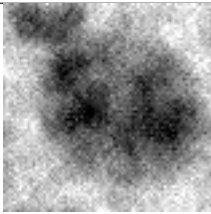
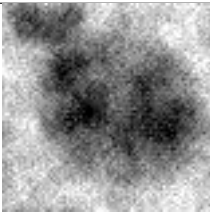
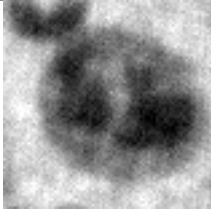
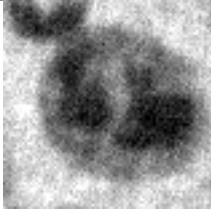
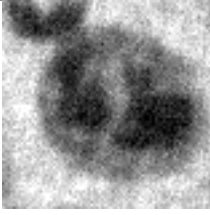
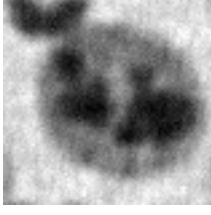
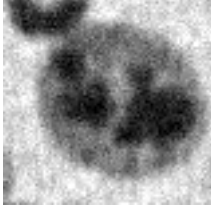
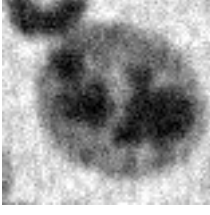



Measurements	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0.001$
M608			
M1014			
M1521			
M3042			

FIGURE 8: Results of the second variant experiments with all the provided noisy measurements and associated  $\epsilon$

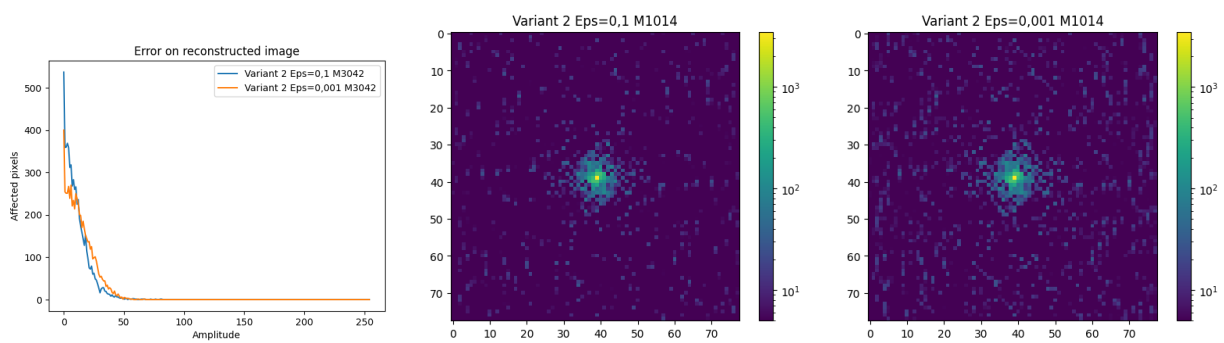


FIGURE 9: Reconstruction error and Fourier transform for two values of  $\epsilon$  for M1014 measurements.

### 3.6 Third variant

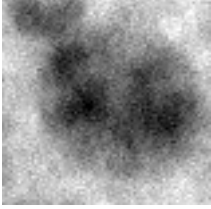
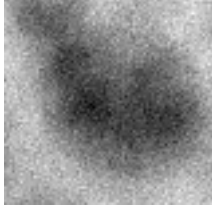

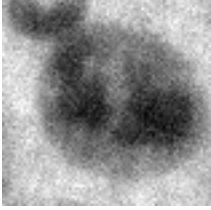
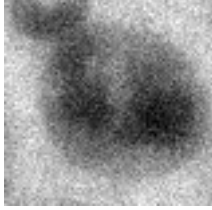

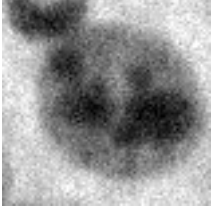
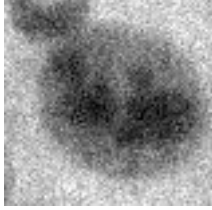
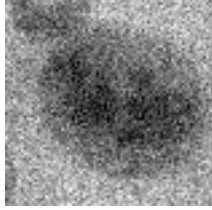
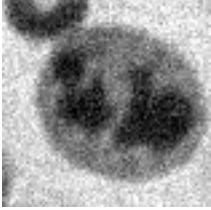
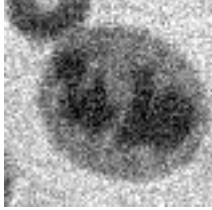
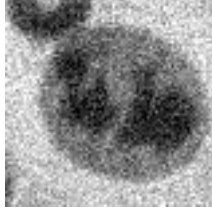
Measurements	$\tau = R/2$	$\tau = R$	$\tau = 2R$
M608			
M1014			
M1521			
M3042			

FIGURE 10: Results of the third variant experiments with all the provided noisy measurements and associated  $\tau$

Looking at the plots of [fig11](#) we can see that:

- The error profile is different for different values of  $\tau$ . A small  $\tau$  leads to less noise in the reconstructed image.
- The smaller value of the  $\tau$  (507) pushes the optimizer towards more sparse solutions.

In conclusion, a smaller  $\tau$  makes sparsity better and error better.

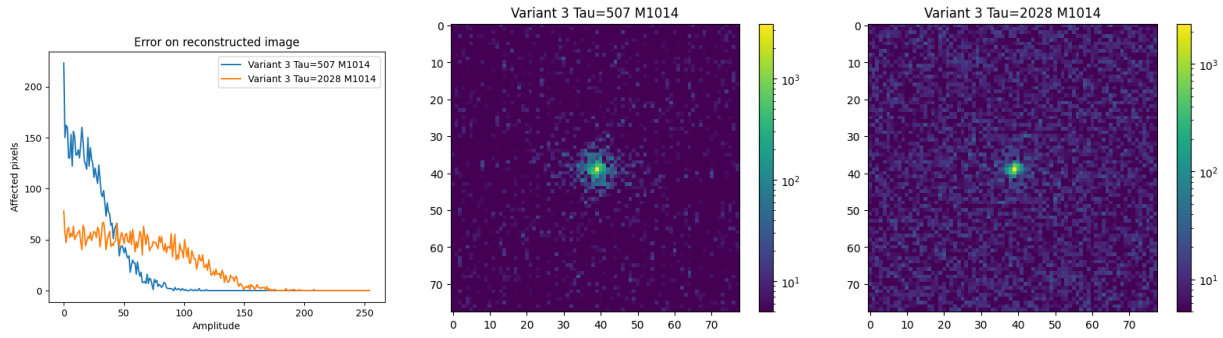


FIGURE 11: Reconstruction error and Fourier transform for two values of  $\tau$  for M1014 measurements.

### 3.7 Conclusion

When decompressing the image, the neutrophilic polynuclear cell can be recognized in almost all  $\ell_1$  norm decompressions. Moreover, the larger the size of the measurement matrix, the more easily the cell is recognized. However, as shown in the result table, the compilation times for the largest matrices are longer. Therefore, a choice, a tradeoff between computation time and result must be made. Indeed, in the case of a neutrophilic polynuclear cell is recognizable from matrices of 1014 with version 2 and  $\epsilon = 0.01$ . Another choice could be made for other cell. This choice can be made by evaluating the results of the table below [12](#).

M608	Time computation	Iteration	$\epsilon$ or $\tau$
$v_1$	155	13	0.1
$v_2$	26.85	15	0.1
$v_3$	147.46(s)	17	304.0
$v_1$	149	13	0.01
$v_2$	22.55(s)	17	0.01
$v_3$	130.55	15	608
$v_1$	132	13	0.001
$v_2$	23.08(s)	18	0.001
$v_3$	99.80(s)	11	1016
M1521	Time computation	Iteration	$\epsilon$ or $\tau$
$v_1$	1603	11	0.1
$v_2$	153.11(s)	16	0.1
$v_3$	170.88 (s)	15	760.5
$v_1$	1930	13	0.01
$v_2$	87.65(s)	19	0.01
$v_3$	453.33(s)	12	1521
$v_1$	1827	11	0.001
$v_2$	86.66	18	0.001
$v_3$	202.18	13	3042
M1014	Time computation	Iteration	$\epsilon$ or $\tau$
$v_1$	581	12	0.1
$v_2$	89.77(s)	16	0.1
$v_3$	207.77(s)	16	507.0
$v_1$	426	13	0.01
$v_2$	60.37	17	0.01
$v_3$	693.90	15	1014
$v_1$	582	13	0.001
$v_2$	57.49	17	0.001
$v_3$	468.56	14	2028
M3042	Time computation	Iteration	$\epsilon$ or $\tau$
$v_1$	too long	n/a	0.1
$v_2$	397.32	13	0.1
$v_3$	567.85(s)	13	1521
$v_1$	too long	n/a	0.01
$v_2$	454.05	19	0.01
$v_3$	423.26 (s)	10	3042
$v_1$	too long	n/a	0.001
$v_2$	427.05(s)	21	0.001
$v_3$	415.91(s)	10	6084

FIGURE 12: Results of all variants experiments with all the provided noisy measurements and associated  $\epsilon$  or  $\tau$

# Annexes

# 1 Result of non robust with $l_1$ norm

	Objective		Residual			
Iter	Primal	Dual	Primal	Dual	Compl	Time
0	1.12935753e+05	0.00000000e+00	2.84e-13	0.00e+00	1.12e+01	37s
1	2.38562838e+04	9.94777225e+00	6.88e-09	1.58e-07	1.31e+00	42s
2	2.22911632e+03	5.43189723e+01	3.51e-08	6.29e-07	1.19e-01	48s
3	8.05005508e+02	9.78901726e+01	8.41e-09	3.17e-07	3.87e-02	55s
4	4.02891949e+02	1.61961299e+02	1.64e-09	1.09e-07	1.32e-02	62s
5	2.66974175e+02	2.12019228e+02	1.39e-10	2.36e-08	3.01e-03	70s
6	2.37033230e+02	2.24502933e+02	1.24e-11	7.35e-09	6.87e-04	78s
7	2.31912635e+02	2.28639777e+02	1.53e-12	2.25e-09	1.79e-04	86s
8	2.30953686e+02	2.29946911e+02	2.66e-13	7.55e-10	5.52e-05	93s
9	2.30737759e+02	2.30420360e+02	2.34e-13	2.07e-10	1.74e-05	99s
10	2.30650549e+02	2.30547400e+02	2.45e-13	6.01e-11	5.65e-06	106s
11	2.30614232e+02	2.30584653e+02	2.56e-13	1.79e-11	1.62e-06	112s
12	2.30603840e+02	2.30597520e+02	6.55e-10	3.11e-10	3.46e-07	119s
13	2.30601323e+02	2.30600191e+02	1.82e-09	7.01e-10	6.20e-08	127s
14	2.30601095e+02	2.30600798e+02	5.64e-10	1.83e-10	1.63e-08	133s
15	2.30601020e+02	2.30601011e+02	1.74e-10	3.51e-12	4.62e-10	139s
16	2.30601013e+02	2.30601013e+02	2.90e-10	1.05e-14	6.33e-13	146s

FIGURE 13: M608 non robust computation

Iter	Objective		Residual		Compl	Time
	Primal	Dual	Primal	Dual		
0	1.43381958e+05	-1.96311539e-15	3.13e-13	8.07e-17	1.74e+01	88s
1	4.41703758e+04	2.25330250e+01	1.72e-08	3.15e-07	2.42e+00	108s
2	2.37746737e+03	6.00215036e+01	4.92e-10	9.72e-08	1.27e-01	127s
3	7.28705412e+02	1.28885502e+02	8.23e-11	3.79e-08	3.29e-02	149s
4	3.96322980e+02	2.08568402e+02	1.54e-11	1.24e-08	1.03e-02	172s
5	2.91738714e+02	2.58064380e+02	8.21e-13	1.72e-09	1.84e-03	197s
6	2.74741887e+02	2.68731737e+02	3.98e-13	3.15e-10	3.29e-04	222s
7	2.72124674e+02	2.71235156e+02	2.84e-13	3.44e-11	4.87e-05	246s
8	2.71727659e+02	2.71556577e+02	3.69e-13	7.03e-12	9.37e-06	269s
9	2.71669860e+02	2.71624652e+02	2.95e-13	2.07e-12	2.48e-06	291s
10	2.71660241e+02	2.71644618e+02	3.84e-13	6.68e-13	8.56e-07	314s
11	2.71655937e+02	2.71650061e+02	3.13e-13	2.62e-13	3.22e-07	334s
12	2.71654319e+02	2.71651988e+02	3.73e-13	1.10e-13	1.28e-07	355s
13	2.71653742e+02	2.71653331e+02	2.24e-13	1.39e-14	2.26e-08	376s
14	2.71653527e+02	2.71653500e+02	9.87e-10	3.07e-12	1.50e-09	399s
15	2.71653508e+02	2.71653507e+02	4.87e-10	4.64e-14	1.46e-11	422s

FIGURE 14: M1014 non robust computation

Iter	Objective		Residual		Compl	Time
	Primal	Dual	Primal	Dual		
0	1.81951026e+05	0.00000000e+00	1.81e-13	0.00e+00	1.81e+01	223s
1	5.66832987e+04	2.36269863e+01	2.61e-09	1.88e-07	3.11e+00	282s
2	1.66655171e+03	6.29061734e+01	8.98e-09	2.29e-05	8.79e-02	346s
3	5.37982230e+02	2.00518157e+02	1.80e-08	1.01e-06	1.85e-02	413s
4	3.33381908e+02	2.85645799e+02	1.22e-08	8.56e-08	2.62e-03	491s
5	3.07814451e+02	3.01208611e+02	4.74e-09	2.96e-08	3.62e-04	570s
6	3.03910099e+02	3.03152330e+02	1.01e-09	2.87e-09	4.15e-05	633s
7	3.03507459e+02	3.03376699e+02	8.12e-10	2.74e-10	7.16e-06	697s
8	3.03429649e+02	3.03412895e+02	1.53e-09	2.13e-09	9.18e-07	760s
9	3.03417910e+02	3.03416103e+02	1.54e-09	7.27e-10	9.90e-08	824s
10	3.03416559e+02	3.03416437e+02	1.90e-09	5.04e-11	6.68e-09	891s
11	3.03416466e+02	3.03416459e+02	5.78e-09	1.34e-12	3.58e-10	966s
12	3.03416461e+02	3.03416460e+02	8.09e-10	2.85e-14	8.26e-12	1043s

FIGURE 15: M1521 non robust computation

	Objective					
Iter	Primal	Dual	Primal	Dual	Compl	Time
0	3.12155369e+05	0.00000000e+00	2.70e-13	0.00e+00	3.11e+01	648s
1	1.05004183e+05	2.60837192e+01	2.28e-09	2.56e-07	5.76e+00	735s
2	3.18761740e+03	6.35727354e+01	4.93e-09	1.41e-05	1.71e-01	798s
3	6.16450925e+02	1.83987264e+02	7.65e-09	8.72e-07	2.37e-02	852s
4	3.96403093e+02	3.07688850e+02	1.57e-08	4.97e-08	4.86e-03	947s
5	3.63342741e+02	3.49681126e+02	6.82e-09	3.81e-09	7.48e-04	1045s
6	3.57630415e+02	3.55437546e+02	1.64e-09	1.44e-09	1.20e-04	1136s
7	3.56428356e+02	3.56108417e+02	1.82e-09	2.24e-09	1.75e-05	1220s
8	3.56238156e+02	3.56204920e+02	1.97e-09	6.99e-10	1.82e-06	1288s
9	3.56215272e+02	3.56213223e+02	2.10e-09	5.72e-10	1.12e-07	1347s
10	3.56214237e+02	3.56214077e+02	7.27e-09	2.51e-11	8.82e-09	1419s
11	3.56214125e+02	3.56214114e+02	3.38e-10	9.86e-13	5.80e-10	1482s
12	3.56214118e+02	3.56214117e+02	1.99e-09	1.15e-14	3.64e-11	1552s

FIGURE 16: M3042 non robust computation



# Bibliography

- [1] Steven L. Brunton and J. Nathan Kutz. Data driven science + engineering.