Analysis of a physics-inspired GAN approach for super-resolution fluorescence microscopy

par
Hamza Mentagui
Encadrement:
Luca Calatroni, Laure Blanc-Féraud

25 octobre 2023

Contents

1 Introduction **3**
1.1 Motivation . . . 3
1.2 Objectives . . . 3
1.3 Contribution . . . 3

2 Fluorescence microscopy **4**
2.1 Inverse Problems . . . 6

3 Model for stochastic fluctuations **7**
3.1 Physical modelling . . . 7
3.2 Mathematical formalisation . . . 9

4 Learning problem **10**
4.0.1 Comparison of distributions . . . 11
4.0.2 Problem statement and constraints . . . 12

5 Contribution **14**

6 Single-Poisson fluctuation model **15**
6.1 Calculating the gradient . . . 17
6.2 Expansion of the data term . . . 19
6.2.1 Analysis . . . 21

7 Double-Poisson **22**
7.1 Computation of fitting term 23
7.2 Analysis 24
7.3 Gradient Calculation 25
7.4 An approximation of the gradient 26

8 Optimisation **27**
8.1 Optimisation of the discriminator 27
8.2 Optimisation of the generator 27

9 Numerical Results **28**
9.1 Data 30
9.1.1 Ground Truth 30
9.2 Real images 31
9.3 Model setting 32
9.3.1 Generator 32
9.3.2 Discriminator 33
9.4 Numerical results 34
9.5 The performance and comparison criterion 34
9.6 The true and approximate gradient 35
9.7 Evaluation of the complete fluctuation model 37

10 Conclusion **39**

1 Introduction

1.1 Motivation

My internship in the field of fluorescence microscopy research is an exciting and crucial step in my academic and professional journey. Located within the esteemed team Morpheme at the I3S laboratory (Laboratory of Computer Science, Signals, and Systems) and INRIA on the picturesque Côte d'Azur in Sophia Antipolis, this opportunity offored me an exceptional platform for me to dive headfirst into the fascinating field of fluorescence microscopy and the one of inverse problems.

1.2 Objectives

the main objectives of this internship are to gain a comprehensive understanding of fluorescence microscopy and its mathematical modellinge, as well as the challenges it addresses. Subsequently, to familiarize myself with the field of inverse problems and explore their connection to microscopy. Finally, the most significant goal is to continue and expand upon the work [13] of a former intern, which focuses on describing an adversarial optimization method for numerically solving the inverse problem of super-resolution fluorescence microscopy via the analysis of fluorescence stochastic fluctuations..

1.3 Contribution

During my internship, I made various contributions within the scope of the work initiated by the previous intern. Firstly, in the model of stochastic fluctuations considered in previous work (modelling only the fluctuations due to the presence of Poisson noise), which will be presented in detail later, I provided a mathematical analysis of the data fidelity term present in the objective function considered, as well as an optimization method based on random coordinate descent using the true gradient of the functional instead of the approximate gradient. Furthermore, I worked on the general model (describing both the noise fluctuations and the ones more related to the signal itself) by calculating the theoretical gradient of the function to be minimized and analyzing the associated data fidelity term. Lastly, I provided numerical results that put the theoretical part into practice.

2 Fluorescence microscopy

Fluorescence microscopy is a major tool with frequently employed in several disciplines, such as cell physiology. The underlying process of fluorescence involves the absorption of light energy (a photon) by an fluorescent sample followed by the emission of some of this light energy (as another photon) a few nanoseconds later. Because some energy is lost in this process, the emitted photon has less energy than the absorbed photon. Light with a short wavelength (toward the blue) has higher energy than light with a long wavelength (toward the red). Therefore, light emitted from an fluorescent sample usually has a longer wavelength than that of the absorbed (excitation) light. The molecular transitions explaining these processes can be depicted in terms of energy diagrams (1).

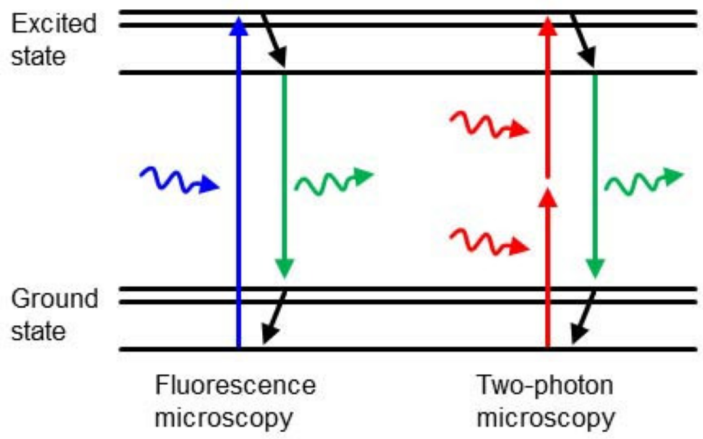


Figure 1: Shematic illustration of single photon and two photon absorption

Most organisms do not naturally contain such fluorescent sample, but artificial fluorescent samples that can bind to the proteins of interest can be introduced. This allows for the precise selection of a specific part of a cell or organism and the simultaneous observation of multiple types of proteins, provided that they are illuminated with the appropriate wavelengths. This is typically achieved using a light source, such as a laser or LEDs, which illuminates the sample at its absorption wavelength, activating the fluorescent markers.

Diffraction: The phenomenon of diffraction plays a crucial role in fluorescence microscopy. Diffraction is an optical phenomenon that constrains the resolution capability of microscopes by preventing the formation of images

of very small objects at very close distances. This means that, without accounting for diffraction, we would not be able to observe details smaller than a certain size.The maximum spatial resolution of a microscope is limited by diffraction. The Abbe diffraction limit is often used to determine the minimum achievable resolution. According to this limit, the maximum resolution is approximately half the wavelength of the light used. Mathematically, this can be expressed as:

$$R \approx \frac{\lambda}{2nsin(\theta)}$$

with:

- R: is the resolution.
- λ is the wavelength of light.
- n is the refractive index of the medium.
- θ is the numerical aperture angle of the microscope.

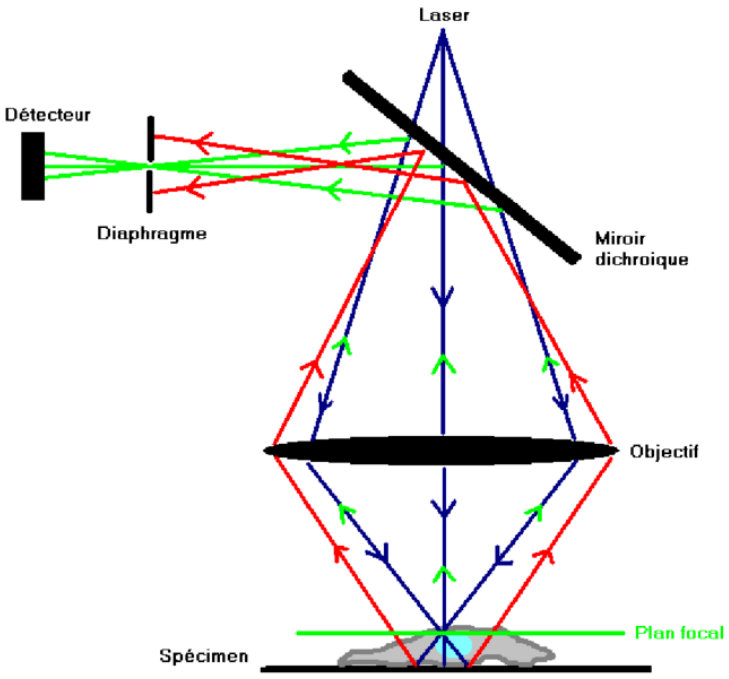
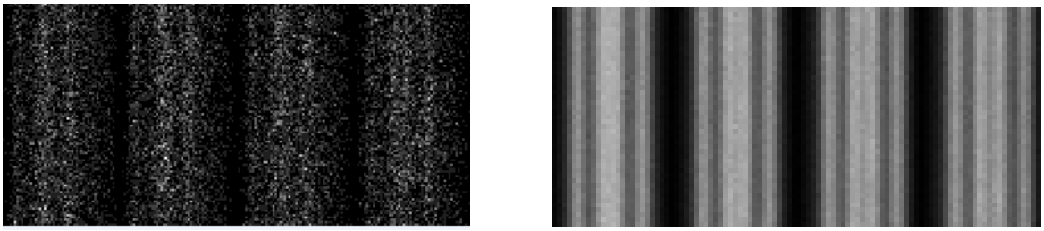


Figure 2: Diffraction phenomena

Diffraction sets a limit on the spatial resolution of a microscope, meaning that

two closely spaced points cannot be distinguished if their separation is below a certain limit. Thus, diffraction is the primary factor that constrains the resolution capability of a microscope. As a result, image blurring is present in images acquired by a microscope due to diffraction. Fine and closely spaced details are not reproduced sharply, making it challenging to visualize sub-cellular structures or individual particles. For standard microscopes, the separation distance is approximately $220nm$.

Stochastic Fluctuations: these phenomena describe to random variations in the detected light signals during fluorescence microscopy image acquisition. In practice, a sequence of blurred, noisy and possibly undersampled images... is observed of noisy images from the same sample (3). These fluctuations can be caused by various factors, including fluctuations in the number of photons emitted by fluorophores, optical imperfections, thermal disturbances, and electronic limitations of detectors.



(a) (b)

Figure 3: (a) and (b) correspond, respectively, to the microscope-acquired image and the sequence average.

These sequences of fluctuations are generated using SOFItools [6], which is a program that simulates acquisition process of the microscope, before to get super-resolution.

2.1 Inverse Problems

Inverse problems are common in science and engineering, including fluorescence microscopy. They involve determining the cause of an observation from measured data. In fluorescence microscopy, a typical inverse problem is to reconstruct the spatial distribution of fluorophores in a sample from acquired images. Inverse problems are common in science and engineering, including fluorescence microscopy. They involve determining the cause of an observation from measured data. Let $\mathbf{y} \in \mathcal{Y}$ be the acquired image, then:

$$\mathbf{y} = \mathcal{H}(\mathbf{x}) + \eta$$

with:

- $\mathbf{x} \in \mathcal{X}$ is the sample.
- $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$ is the operator modeling the microscope.
- η is the noise.

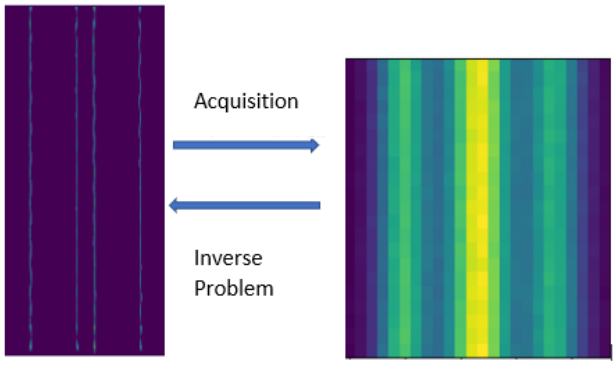


Figure 4: Inverse Problem in Fluorescence microscopy

3 Model for stochastic fluctuations

3.1 Physical modelling

Modeling stochastic fluctuations in fluorescence microscopy is essential for understanding and quantifying the inherent noise and variability in fluorescence signals. Stochastic fluctuations can arise from various sources, including photon statistics, detector noise, and sample variability. the physical approach to model and analyze stochastic fluctuations in fluorescence microscopy is based on four ingredients :

1. **Photon Statistics:**
    - The most fundamental source of noise in fluorescence microscopy is photon shot noise. We can distinguish two types of fluctuations:
        - **Emission noise:** related to the emission of photons by fluorophores.

  - **Detection noise:** related to the detection of photons by the microscope.

2. **Point Spread Function (PSF):**
   - To model the effects of blurring introduced by the diffraction phenomenon.
3. **The space-variant background:**
   - Models here out-of-focus and ambient molecules.
4. **Background Noise:**
   - Detector electronics and other sources can introduce Gaussian noise into the signal. You can model this noise as a Gaussian distribution with a mean of zero and a certain standard deviation.

3.2 Mathematical formalisation

In this section, we elucidate the mathematical tools that are employed to implement the physical modeling discussed in the previous section.

By convention, we work with vectorized images:

$$\mathbf{x} \in (\mathbb{R}^+)^{nn} \iff \mathbf{x} \in (\mathbb{R}^+)^{n^2} \tag{1}$$

We denote $\mathcal{X} = (\mathbb{R}^+)^{n^2}$ and $\mathcal{Y} = (\mathbb{R}^+)^{M^2}$ two spaces of images (they are modeled by vectors) such that $n = L \times M$:

- $\mathbf{x} \in \mathcal{X}$ represents the super-resolved image, which means the spatial distribution of fluorescent molecules discretized on a fine grid.
- $\mathbf{y} \in \mathcal{Y}$ represents the image obtained after microscope acquisition.
- L is the factor resolution.

The microscope acquisition process can be modeled by a stochastic process that depends on the parameters describing the physical model explained previously. Indeed, the stochastic model of fluctuations is described by a Poisson process. Let $\mathbf{x} \in \mathcal{X}$ be the input image, and $\mathbf{y} \in \mathcal{Y}$ the acquired image (by microscope):

$$\mathbf{P}_{\mathbf{y}/\mathbf{x}} \sim \mathcal{P}(\mathbf{U}\hat{\mathbf{H}}\mathcal{P}(\mathbf{x}) + \mathbf{b}) + \eta \tag{2}$$

Such that:

- $\mathcal{P}$: Poisson distribution.
- $\hat{\mathbf{H}} : \mathcal{X} \to \mathcal{X}$: The convolution operator associated with a Gaussian kernel. Its role is to model the blurring due to the diffraction's phenomenon. Here $\hat{\mathbf{H}}$ is a matrix (the matrix form). we denote $\mathbf{C_i} \in \mathbb{R}^{n^2 \times 1}$ its i-th column:

$$\hat{\mathbf{H}} = [\hat{\mathbf{C}}_1, ..., \hat{\mathbf{C}}_{n^2}] \tag{3}$$

- $\mathbf{U} : \mathcal{X} \to \mathcal{Y}$: The undersampling matrix.

- $\mathbf{b} \in \mathbb{R}^{M^2}$: The space-variant background.

- $\eta \sim \mathcal{N}(O_{M^2}, \sigma^2 \mathbf{Id}_{(M^2,M^2)})$: Gaussian distribution which models the electronic noise.

We have to notice that $\mathbf{P}_{\mathbf{y}/\mathbf{x}}$ is the conditional probability of $\mathbf{y}$ giving $\mathbf{x}$.

In an abuse of language, we denote $\mathbf{H} : \mathcal{X} \to \mathcal{Y}$ the matrix that contains the convolution and the undersampling operations:

$$\mathbf{P}_{\mathbf{y}/\mathbf{x}} \sim \mathcal{P}(\mathbf{H}\mathcal{P}(\mathbf{x}) + \mathbf{b}) \tag{4}$$

The columns of $\mathbf{H}$ are positive and normalized:

$$\|\mathbf{C}_j\|_1 = 1 \quad where \quad \mathbf{H} = [\mathbf{C}_1, ..., \mathbf{C}_{n^2}] \tag{5}$$

4 Learning problem

The objective of our work is to reconstruct the super-resolved image with only a finite number of samples acquired from the microscope. Indeed, we do not have guarantees about the legitimacy of our model. In other words, if we denote $\mathbf{x}^*$ the sought-after input and $\mathbf{y}^{real}$ the output image coming from microscope, usualy we have :

$$\mathbf{P}_{\mathbf{y}^{real}/\mathbf{x}^*} \neq \mathcal{P}(\mathbf{H}\mathcal{P}(\mathbf{x}^*) + \mathbf{b}) + \eta \tag{6}$$

In order to avoid confusion between the real output and the simulated one, we denote $\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b})$ (Random variable) the image coming from the Poisson process and $\mathbf{y}^{real}$ the real output :

$$\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) \sim \mathcal{P}(\mathbf{H}\mathcal{P}(\mathbf{x}) + \mathbf{b}) + \eta \tag{7}$$

We are facing a learning problem. In fact, we would like to find $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that approximate $\mathbf{y}^{real}$ by $\mathbf{y}^{sim}(\hat{\mathbf{x}}, \hat{\mathbf{b}})$ in distribtuion. In other words, we want to find (reconstruct) $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that minimize the distance between $\mathbf{P}_{\mathbf{y}^{sim}(\mathbf{x},\mathbf{b})}$ and $\mathbf{P}_{y^{real}}$:

$$\hat{\mathbf{x}}, \hat{\mathbf{b}} = \underset{\mathbf{x},\mathbf{b}}{argmin}\ \mathcal{C}(\mathbf{P}_{\mathbf{y}^{sim}(\mathbf{x},\mathbf{b})}, \mathbf{P}_{y^{real}}) \tag{8}$$

where $\mathcal{C}$ is the distance over the distribution's space.

4.0.1 Comparison of distributions

In this section, we discuss methods for comparing two distributions, $\mathbf{P}$ and $\mathbf{Q}$, particularly divergences and distances:

1. $\mathbf{l}_2$ **norm**: The easiest way to compare $\mathbf{P}$ to $\mathbf{Q}$ is to compute the difference between the expectations of their associated random variables in terms of the $\mathbf{l}_2$ norm:

$$\mathcal{C}(\mathbf{P}, \mathbf{Q}) = \|\mathbf{E}_{X\sim\mathbf{P}}[X] - \mathbf{E}_{Y\sim\mathbf{Q}}[Y]\|^2 \tag{9}$$

   The advantage of this loss is that it can be computed explicitly. However, comparing only two mean images is not very representative of the diversity in the data. A better choice improving the amount of information carried out by (8) consists in comparing distributions using the expectation of $\mathbf{l}_2$ distance between $X \sim \mathbf{P}$ et the expectation of $Y \sim \mathbf{Q}$:

$$\mathcal{C}(\mathbf{P}, \mathbf{Q}) = \mathbf{E}_{X\sim\mathbf{P}}[\|X - \mathbf{E}_{Y\sim\mathbf{Q}}[Y]\|^2] \tag{10}$$

2. **KL divergence** : Another way to compare $\mathbf{P}$ and $\mathbf{Q}$ that are absolutely continuous with respect to a measure on Ω is to compute their Kullback-Leibler divergence:

$$\mathbf{KL}(P\|Q) = \int P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx \tag{11}$$

   Moreover, even in the case of known density functions, such minimisation is generally challenging due, for instance, to the fact that the KL divergence is equal to infinity when one of the two densities vanishes and when they have disjoint supports [8].

3. **Wasserstein distance**: A distance stemming from the optimal transport field [1] :

$$W_1(P, Q) = \inf_{\gamma\in\Gamma(\mathbf{P},\mathbf{Q})} \mathbf{E}_\gamma[\|x - y\|\ ] \tag{12}$$

   where $\Gamma(\mathbf{P}, \mathbf{Q})$ stands for the set of all joint distributions $\gamma(x, y)$ with marginals equal to $\mathbf{P}$ and $\mathbf{Q}$.

   Unfortunetly, this expression has an intractable form.Therefore, it is not useful. One way to circumvent the problem is to use the dual

expression of this distance, called the Kantorovich-Rubinstein dual [2] distance:

$$W_1(P,Q) = \sup_{\|f\|_{Lip}\leqslant 1} \{\mathbf{E}_{X\sim\mathbf{P}}[f(X)] - \mathbf{E}_{Y\sim\mathbf{Q}}[f(Y)]\} \tag{13}$$

where f is 1-Lipschitz function.

Generally, it is not always easy to find f^* such that:

$$f^* = \underset{\|f\|_{Lip}\leqslant 1}{argmax}\ \{\mathbf{E}_{X\sim\mathbf{P}}[f(X)] - \mathbf{E}_{Y\sim\mathbf{Q}}[f(Y)]\} \tag{14}$$

Instead of that, we can make a good choice of a familty of 1-Lipschitz $\{\mathcal{D}_\phi\}$ functions parametrized by ϕ, with $\mathcal{D}_{\phi^*}$ such that :

$$\mathcal{D}_{\phi^*} = \underset{\phi}{argmax}\ \{\mathbf{E}_{X\sim\mathbf{P}}[\mathcal{D}_\phi(X)] - \mathbf{E}_{Y\sim\mathbf{Q}}[\mathcal{D}_\phi(Y)]\} \tag{15}$$

Finaly:

$$W_1(P,Q) \approx \{\mathbf{E}_{X\sim\mathbf{P}}[\mathcal{D}_{\phi^*}(X)] - \mathbf{E}_{Y\sim\mathbf{Q}}[\mathcal{D}_{\phi^*}(Y)]\} \tag{16}$$

4.0.2 Problem statement and constraints

Optimisation problem: We can reformulate our optimization problem in equation (8) as a min-max problem using equation (10) and (15). In other words, for a well-chosen family of 1-Lipschitz functions $\{\mathcal{D}_\phi\}_\phi$ parametrezed by ϕ, we want to find $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that minimize objective function $\mathcal{L}(\mathbf{x},\mathbf{b})$:

$$\hat{\mathbf{x}}, \hat{\mathbf{b}} = \underset{\mathbf{x},\mathbf{b}}{argmin}\ \mathcal{L}(\mathbf{x},\mathbf{b}) \tag{17}$$

with

$$\mathcal{L}(\mathbf{x},\mathbf{b}) = \underset{\|\nabla\mathcal{D}_\phi\|_{Lip}\leqslant 1}{max} \{\mathbf{E}[\mathcal{D}_\phi(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))] - \mathbf{E}[\mathcal{D}_\phi(\mathbf{y}^{real})]\} + \mathbf{E}[\|\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] \tag{18}$$

We added the right-hand term (10) in the equation above to stabilize the convergence of (18) during the optimization process.

Hypotheses:

- $\mathbf{P}_{y^{real}}$ is unknown and we only have a few samples of $\mathbf{y}^{real}$.
- $\mathbf{x}^*$ is a sparse image, that means $\|\mathbf{x}^*\|_0$ is small.
- $\mathbf{b}^*$ is smooth and does not vary abruptly, so $\|\nabla\mathbf{b}^*\|_2^2$ must be small.

Constraints :

Generally, it is not always straightforward to calculate the explicit form of the equation (18) due to the fact that:

- Due to the high-dimensional nature of the problem, we cannot estimate $\mathbf{P}_{y^{real}}$ from a few number of samples of $\mathbf{y}^{real}$.
- Even if $\mathbf{P}_{\mathbf{y}^{sim}}$ is known, it is not always easy to find a closed form of $\mathbf{E}[\mathcal{D}_\phi(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))]$.

To circumvent this problem, we approximate expectations by using empirical quantities. We denote $\{\mathbf{y}_t^{real}\}_{1\leqslant t\leqslant T}$ samples of $\mathbf{y}^{real}$ and $\{\mathbf{y}_t^{sim}(\mathbf{x},\mathbf{b})\}_{1\leqslant t\leqslant T}$ i.i.d samples of $\mathbf{y}^{sim}(\mathbf{x},\mathbf{b})$, then for any function $h:\mathbb{R}^{M^2}\to\mathbb{R}$:

$$\mathbf{E}[h(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))] \approx \frac{1}{T}\sum_{t=1}^{T} h(\mathbf{y}_t^{sim}(\mathbf{x},\mathbf{b})) \tag{19}$$

$$\mathbf{E}[h(\mathbf{y}^{real})] \approx \frac{1}{T}\sum_{t=1}^{T} h(\mathbf{y}_t^{real}) \tag{20}$$

Research questions:

Three questions naturally arise:

1. How can we choose the family of 1-Lipschitz functions $\{\mathcal{D}_\phi\}_\phi$ that approximate well equation (15)?
2. How can we mathematically justify the role of the data-fitting term in equation eq:18:
$$\mathbf{E}[\|\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2]$$
3. Should we add regularization terms to enforce sparsity and smoothness in $\mathbf{x}$ and $\mathbf{b}$, respectively?
4. What can we say about the quality of the reconstruction of $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$:
$$\|\hat{\mathbf{x}} - \mathbf{x}^*\| << \varepsilon \ ; \ \|\hat{\mathbf{b}} - \mathbf{b}^*\| << \varepsilon \tag{21}$$

5 Contribution

The first question was recently addressed by a colleague in the context of an internship [13], in which he considered a more simplified simulation model while assuming that the fluctuations related to photon emission are deterministic, thus neglecting practically the inner Poisson fluctuation model. In other words:

$$\hat{\mathbf{y}}^{sim} \sim \mathcal{P}(\mathbf{H}\mathbf{x} + \mathbf{b}) + \eta \tag{22}$$

Instead of :

$$\mathbf{y}^{sim} \sim \mathcal{P}(\mathbf{H}\mathcal{P}(\mathbf{x}) + \mathbf{b}) + \eta \tag{23}$$

In our work, we study the two models:

1. **Single- Poisson**[13]: $\hat{\mathbf{y}}^{sim}$ comes from equation(22).
2. **Double-Poisson**: $\mathbf{y}^{sim}$ comes from equation (23).

The first model was previously studied. For this model during my internship I have made the following contributions:

- Mathematical justification for the use of the data term.
- In the numerical part, I used the exact gradient instead of the approximation that was used [13]. For this, i used the random coordinate descent (RDC).

For the second model during my internship I provided a mathematical analysis of the cost function:

- Mathematics justification of the fitting term.
- Exact computation of the gradient.
- In numerical part, i used an approximation of the gradient (to reduce the complexity).
- Numerical results.

6 Single-Poisson fluctuation model

We consider the model (22). The idea behind the previous work is to consider a GAN framework [5, 8]. the standard generator is replaced with the model (22) keeping the discriminator, which is a convolutional neural network.

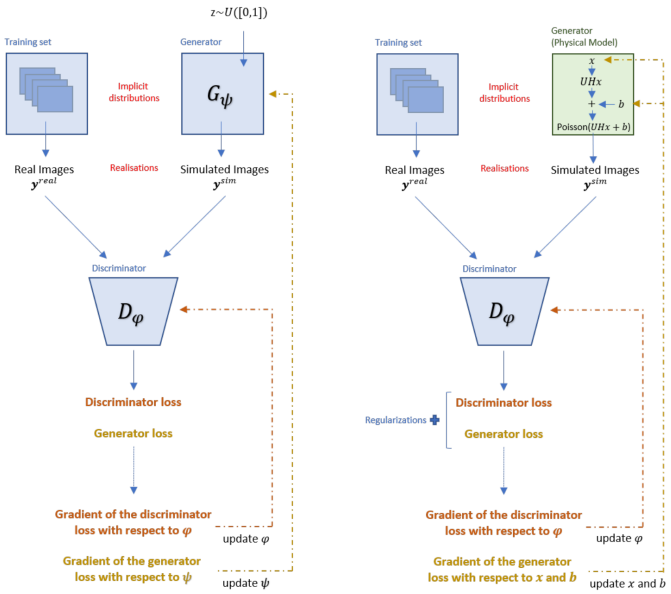


Figure 5: Comparison between a standard GAN architecture (left) and FluoGAN

More precisely, we consider a family of discriminators $\{\mathcal{D}_\phi\} : \mathbb{R}^{M^2} \to [0,1]$ parametrized by $\mathbf{x}$, which are the hyperparameters of the convolutional network that we use in optimization problem (16) following the cost function:

$$\mathcal{L}(\mathbf{x}, \mathbf{b}) = \underset{|\nabla \mathcal{D}_\phi||_{Lip} \leqslant 1}{max} \{\mathbf{E}[\mathcal{D}_\phi(\hat{\mathbf{y}}^{sim}(\mathbf{x}, \mathbf{b}))] - \mathbf{E}[\mathcal{D}_\phi(\mathbf{y}^{real})]\} + \mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x}, \mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2]$$

Optimisation Problem:
The min-max problem (17) is modelled here as an optimisation problem of two cost functions associated with the generator and discriminator, respec-

tively:

$$\mathcal{L}_{\phi}^{gene}(\mathbf{x},\mathbf{b}) = \frac{\gamma}{2}\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] - \delta\mathbf{E}[\mathcal{D}_{\phi}(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] + \lambda_1\|\mathbf{x}\|_1 + \frac{\lambda_2}{2}\|\nabla\mathbf{b}\|^2 \tag{24}$$

$$\mathcal{L}_{\mathbf{x},\mathbf{b}}^{Disc}(\phi) = \mathbf{E}[\mathcal{D}_{\phi}(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] - \mathbf{E}[\mathcal{D}_{\phi}(\mathbf{y}^{real})] + \lambda_D \underbrace{(\|\nabla_y \mathcal{D}_{\phi}(y^{mix})\| - 1)^2}_{\text{The relaxation of the 1-Lipschitz constraint}} \tag{25}$$

where:

- $\delta = 0$: Without the discriminator.
- $\delta = 1$: With the discriminator.
- $\hat{\mathbf{y}}^{mix} = \mu\mathbf{y}^{sim} + (1-\mu)\mathbf{y}^{real}$ with $\mu \sim \mathcal{U}[0,1]$.
- λ_D : Regularization parameter.
- λ_i; i=1,2: $\mathbf{l}_i$ regularization parameter.
- γ : Normalization parameter of data attachment term.

The optimization process takes place in two steps:

1. We fix $\mathbf{x}$ and $\mathbf{b}$, and then we find the parameters $\hat{\phi}$ that minimize (25).
2. We fixe $\hat{\phi}$, and then we find $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that minimize (24).

To use optimization algorithms, we need in particular to calculate the gradient of:

$$\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] - \delta\mathbf{E}[\mathcal{D}_{\phi}(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] \tag{26}$$

with respect to $\mathbf{x}$ and $\mathbf{b}$.

Generally, it is not always straightforward to calculate the gradient of a function in the form:

$$\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))] \tag{27}$$

with $f : \mathbb{R}^{M^2} \to \mathbb{R}$.

For example, for M=1:

$$\begin{aligned}\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] &= \mathbf{E}[f(z)] \;\; where \;\; z \sim \mathcal{P}(hx+b) \in \mathbb{N} \\ &= \sum_{n=0}^{\infty} f(n)\mathbf{P}(z=n) \\ &= \exp^{-(hx+b)} \sum_{n=0}^{\infty} f(n)\frac{(hx+b)^n}{n!}\end{aligned}$$

which is hard to manipulate.

6.1 Calculating the gradient

The calculation is done in two steps:

1. **Without noise**:
   In this first step, we do not take the Gaussian noise into account ($\eta = 0$), as already done in previous work. In other words:

   $$\hat{\hat{\mathbf{y}}}^{sim} \sim \mathcal{P}(\mathbf{z}) \quad where \;\; \mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{b} \tag{28}$$

   The idea is to compute the gradient of $\mathbf{E}_{\mathbf{z}}[f(\hat{\hat{\mathbf{y}}}^{sim})]$ with respect $\mathbf{z}$ and by chains rule with respect $\mathbf{x}$ and $\mathbf{b}$.

2. **With noise** :
   We do the previous calculations in the case where we introduce noise. This is a contribution of my internship.

Step 1 :
For optimising the function we need to compute quantities in the form $\nabla_x \mathbb{E}\left[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b})\right]$ and $\nabla_b \, \mathbb{E}\left[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}))\right]$ where:

$$f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b})) = \frac{\gamma}{2} \, \|\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{y}^{real}\|^2 - \delta\mathcal{D}_{\phi}(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}))$$

and the dependence of $\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b})$ on both $\mathbf{x}$ and $\mathbf{b}$. The gradient of the expected value of a Poisson random variable with respect to its parameter(s) can be computed directly.

[13] gives the exact form of the gradient:

$$\nabla_x \mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim})] = \mathbf{H}^T\{\mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) + 1_i) - \mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}))]\}_{1\leqslant i\leqslant M^2} \quad (29)$$
$$\nabla_b \mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim})] = \{\mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) + 1_i) - \mathbf{E}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}))]\}_{1\leqslant i\leqslant M^2}$$

Step 2(our contribution):
We now consider the case where $\eta \neq 0$ (22) and rewrite it as:

$$\hat{\mathbf{y}}^{sim} = \hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) + \eta \quad (30)$$

with $\{1_i\}_{1\leqslant i\leqslant M^2}$ The canonical basis of $\mathbb{R}^{M^2}$.

There holds:

$$\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] = \mathbf{E}_\eta[\mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]]$$

We have that $\mathbf{E}_{\hat{\hat{\mathbf{y}}}}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]$ depends on $\mathbf{x}$ and $\mathbf{b}$ only and it is $\sigma(\eta)$-measurable function. In previous step, we have shown that $\mathbf{E}_{\hat{\hat{\mathbf{y}}}}[f(\hat{\hat{\mathbf{y}}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]$ is differentiable and its differential is bounded, so we can apply the theorem of differentiation under the integral sign:

$$\begin{aligned}\nabla_x \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] &= \nabla_x \mathbf{E}_\eta[\mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]] \\ &= \mathbf{E}_\eta[\nabla_x \mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]]\end{aligned}$$

According to (29):

$$\begin{aligned}&\nabla_x \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] \\ &= \{\mathbf{E}_\eta[\mathbf{H}^T \mathbf{E}_{\hat{\hat{\mathbf{y}}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta + 1_i + \eta)|\eta] - \mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta + \eta)|\eta]]\}_{1\leqslant i\leqslant M^2} \\ &= \mathbf{H}^T\{\mathbf{E}_\eta[\mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta) + 1_i|\eta]] - \mathbf{E}_\eta[\mathbf{E}_{\hat{\hat{y}}^{sim}}[f(\hat{\hat{y}}^{sim}(\mathbf{x},\mathbf{b}) + \eta)|\eta]]\}_{1\leqslant i\leqslant M^2} \\ &= \mathbf{H}^T\{\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) + 1_i)] - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))]\}_{1\leqslant i\leqslant M^2} \quad (31)\end{aligned}$$

By the same reasoning, we have:

$$\nabla_b \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] = \{\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) + 1_i)] - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))]\}_{1\leqslant i\leqslant M^2} \quad (32)$$

In previous work, to reduce the complexity of calculations, the following approximation was made:

$$\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) + 1_i) - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] \approx \mathbf{E}[\frac{df(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))}{dx_i}] \quad (33)$$

We don't have guarantee that it is a good approximation. To verify this, we explicitly develop the data term, for which we will notice that the L_1 term appears naturally. This insight constitute the main ingredients of the following section.

6.2 Expansion of the data term

In this section, we focus on the calculation and analysis of (10). We consider the expression (30) of the model:

$$\hat{\mathbf{y}}^{sim} = \hat{y}^{sim} + \eta$$

We denote :

$$\hat{\mathbf{y}}^{sim} = \begin{bmatrix} \hat{y}_1^{sim} \sim \mathcal{P}([\mathbf{Hx}]_1 + b_1) + \eta_1 \\ \hat{y}_2^{sim} \sim \mathcal{P}([\mathbf{Hx}]_2 + b_2) + \eta_2 \\ \dots \\ \hat{y}_{M\times M}^{sim} \sim \mathcal{P}([\mathbf{Hx}]_{M^2} + b_{M^2}) + \eta_{M^2} \end{bmatrix}; \quad \mathbf{y}^{real} = \begin{bmatrix} y_1^{real} \\ y_2^{real} \\ \dots \\ y_{M^2}^{real} \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{n^2} \end{bmatrix}; \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_{M^2} \end{bmatrix}$$

with $\{\hat{\mathbf{y}}_i^{sim}\}_i$ independent random variables

and

$$\mathbf{H} = \{h_{i,j}\} \in \mathcal{M}_{M^2,n^2}(\mathbb{R}) \quad \mathbf{Hx} = \begin{bmatrix} [\mathbf{Hx}]_1 \\ [\mathbf{Hx}]_2 \\ \dots \\ [\mathbf{Hx}]_{M^2} \end{bmatrix}$$

we have:

$$\begin{aligned} &\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] = \mathbf{E}[\|(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})]) + (\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}])\|^2] \\ &= \mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})])\|^2] + \|\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\|^2 \\ &+ 2\underbrace{\mathbf{E}[\langle \hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}), \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\rangle]}_{=0} \\ &= \underbrace{\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})])\|^2]}_{\text{Variance}} + \underbrace{\|\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\|^2}_{\text{Bias}} \end{aligned} \tag{34}$$

Remark:
Here, what we call variance is the sum of the variances of each coordinate of $\hat{\mathbf{y}}^{sim}$, unlike the general definition of the variance of $\hat{\mathbf{y}}^{sim}$, which is a

covariance matrix.
We can expand the variance as:

$$\begin{aligned}
\underbrace{\mathbf{E}[\|(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})])\|^2]}_{\text{Variance}} &= \sum_{i=1}^{M^2} (\hat{y}_i^{sim} - \mathbf{E}[\hat{y}_i^{sim}])^2 \qquad (35)\\
&= \sum_{i=1}^{M^2} Var(\hat{y}_i^{sim}) \\
&= \sum_{i=1}^{M^2} Var(\hat{\hat{y}}_i^{sim}) + Var(\eta_i) \\
&= \sum_{i=1}^{M^2} [\mathbf{Hx}]_i + b_i \quad + M^2\sigma^2 \\
&= \sum_{i=1}^{M^2} \sum_{j=1}^{n^2} h_{i,j}x_j \quad + \|\mathbf{b}\|_1 + M^2\sigma^2 \quad ; b_i \geqslant 0 \\
&= \sum_{j=1}^{n^2} x_j \sum_{i=1}^{M^2} h_{ij} \quad + \|\mathbf{b}\|_1 + M^2\sigma^2 \\
&= \sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 + M^2\sigma^2
\end{aligned}$$

$\mathbf{C}_j^i \geqslant 0$ (5) , with:

$$\mathbf{C}_j = \begin{bmatrix} C_j^1 \\ C_j^2 \\ \dots \\ C_j^{M^2} \end{bmatrix}$$

For the bias term:

$$\underbrace{\|\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[y^{rea}]\|^2}_{\text{Bias}} = \|\mathbf{Hx} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2$$

Therefore, we deduce:

$$\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] = \underbrace{\|\mathbf{Hx} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2}_{\text{Bias}} + \underbrace{\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 + M^2\sigma^2}_{\text{Variance}} \qquad (36)$$

6.2.1 Analysis

We observe that the variance term in equation (35) naturally expresses the assumption of sparsity on the Poisson parameter. Indeed, if $\hat{\mathbf{y}}^{sim} \sim \mathcal{P}(\mathbf{z})$ we have:

$$\underbrace{\mathbf{E}[\|(\hat{\mathbf{y}}^{sim} - \mathbf{E}[\hat{\mathbf{y}}^{sim}])\|^2]}_{\text{Variance}} = \|\mathbf{z}\|_1$$

From (35) we make two observations:

- We notice that each coordinate x_i is associated with the L1 norm of the i-th column of $\mathbf{H}$.

- We see that the L1 term of the background $\mathbf{b}$ can lead to a solution $\hat{\mathbf{b}}$ that is sparse which is not always the case for $\mathbf{b}^*$.

In practice, we can numerically show that the variance term is negligible compared to the bias:

$$\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 \ll \underbrace{\|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[y^{real}]\|^2}_{\text{bias}} \tag{37}$$

so we can approximate:

$$\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x}, \mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] \approx \|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[y^{real}]\|^2 + M^2\sigma^2 \tag{38}$$

The minimization of bias does not always lead to an $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that satisfy (21), because H is generally ill-conditioned. To overcome this problem, we add a regularization term that takes into account the sparsity of $\mathbf{x}^*$ and the smoothness of $\mathbf{b}^*$:

$$\|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[y^{real}]\|^2 + \lambda_1\|\mathbf{x}\| + \frac{\lambda_2}{2}\|\nabla\mathbf{b}\|^2 + M^2\sigma^2 \tag{39}$$

Putting everything together in (24)-(25) we get:

$$\mathcal{L}_\phi^{gene}(\mathbf{x}, \mathbf{b}) \approx \frac{\gamma}{2}\|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[y^{real}]\|^2 + \frac{\gamma}{2}M^2\sigma^2 - \delta\mathbf{E}[\mathcal{D}_\phi(\hat{\mathbf{y}}^{sim}(\mathbf{x}, \mathbf{b}))] + \lambda_1\|\mathbf{x}\|_1 + \frac{\lambda_2}{2}\|\nabla\mathbf{b}\|^2 \tag{40}$$

$$\mathcal{L}_{\mathbf{x},\mathbf{b}}^{Disc}(\phi) = \mathbf{E}[\mathcal{D}_\phi(\hat{\mathbf{y}}^{sim}(\mathbf{x}, \mathbf{b}))] - \mathbf{E}[\mathcal{D}_\phi(\mathbf{y}^{real})] + \lambda_D(\|\nabla_y\mathcal{D}_\phi(\mathbf{y}^{mix})\| - 1)^2$$

Remark: The approximation used before (33) works well for the data term (10):

$$\begin{aligned}\nabla_x \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] &= \mathbf{H}^T\{\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})+1_i)] - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))]\}_{1\leqslant i\leqslant M^2} \\ &\approx \mathbf{H}^T\mathbf{E}[\nabla_y f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] \end{aligned} \tag{41}$$

in the specific case where:

$$f(y) = \|y - \mathbf{E}[\mathbf{y}^{real}]\|^2.$$

To justify (41), we develop the right side term:

$$\begin{aligned}\mathbf{H}^T\mathbf{E}[\nabla_y f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] &= 2\mathbf{H}^T\{\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\} \\ &= 2\mathbf{H}^T(\mathbf{H}\mathbf{x}+\mathbf{b}-\mathbf{E}[\mathbf{y}^{real}]) \\ &= \nabla_x\|\mathbf{H}\mathbf{x}+\mathbf{b}-\mathbf{E}[\mathbf{y}^{real}]\|^2\end{aligned}$$

Recalling (38):

$$\nabla_x \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] \approx \nabla_x\|\mathbf{H}\mathbf{x}+\mathbf{b}-\mathbf{E}[\mathbf{y}^{real}]\|^2$$

Finally, the approximation is right (41).

Generally for $f(y) = \mathcal{D}_\phi(y)$, this approximation may not hold.

7 Double-Poisson

In this section, we focus on the complete simulation model (23), which is the major contribution of this work, by computing the cost functional and the gradient corresponding to this new model:

$$\mathbf{y}^{sim} \sim \mathcal{P}(\mathbf{H}\mathcal{P}(\mathbf{x}) + \mathbf{b}) + \eta$$

Remark :
In this section, we address three things:

- We expand the expression of the data term.
- we proceed to compute the gradient of the term $\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b})]$ with respect to $\mathbf{x}$ and $\mathbf{b}$.
- we provide approximations to reduce the complexity of the gradient computation.

Similar to the previous section, we analyze the data term within the framework of (22) with a brief analysis. Then, we proceed to calculate the gradient with respect to $\mathbf{x}$ and $\mathbf{b}$.

7.1 Computation of fitting term

Using the same notation as before, we set:

$$\mathbf{y}^{sim} \sim \mathcal{P}(\mathbf{Hz} + \mathbf{b}) + \eta \; ; \quad \mathbf{z} \sim \mathcal{P}(\mathbf{x}) \tag{42}$$

We have as previously (34):

$$\begin{aligned}
&\mathbf{E}[\|\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] \\
&= \mathbf{E}_{\mathbf{z}}[\mathbf{E}[\|(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) - \mathbf{E}[\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b})]) + (\mathbf{E}[\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}])\|^2 | \mathbf{z}]] \\
&= \mathbf{E}_{\mathbf{z}}[\mathbf{E}[\|(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b}) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})]) + (\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}])\|^2]] \\
&= \underbrace{\mathbf{E}_{\mathbf{z}}[\mathbf{E}[\|(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})])\|^2]]}_{\text{Variance}} + \underbrace{\mathbf{E}_{\mathbf{z}}[\|\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\|^2]}_{\text{Bias}}.
\end{aligned}$$

We expand the variance term using (35):

$$\begin{aligned}
\underbrace{\mathbf{E}_{\mathbf{z}}[\mathbf{E}[\|(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})) - \mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})])\|^2]]}_{\text{Variance}} &= \mathbf{E}_{\mathbf{z}}[\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 z_j + \|\mathbf{b}\|_1 + M^2\sigma^2] \\
&= \sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 \underbrace{\mathbf{E}[z_j]}_{=x_i} + \|\mathbf{b}\|_1 + M^2\sigma^2 \\
&= \sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 + M^2\sigma^2
\end{aligned}$$

For the bias term:

$$\begin{aligned}
\underbrace{\mathbf{E}_{\mathbf{z}}[\|\mathbf{E}[\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b})] - \mathbf{E}[\mathbf{y}^{real}]\|^2]}_{\text{Bias}} &= \mathbf{E}_{\mathbf{z}}[\|\mathbf{Hz} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2] \\
&= \mathbf{E}_{\mathbf{z}}[\|(\mathbf{Hz} - \mathbf{Hx}) + (\mathbf{Hx} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}])\|^2] \\
&= \mathbf{E}_{\mathbf{z}}[\|\mathbf{H}(\mathbf{z} - \mathbf{x})\|^2] + \|\mathbf{Hx} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2
\end{aligned}$$

we have:

$$\begin{aligned}
\|\mathbf{H}(\mathbf{z} - \mathbf{x})\|^2 &= (\mathbf{z} - \mathbf{x})^T \mathbf{H}^T \mathbf{H} (\mathbf{z} - \mathbf{x}) \\
&= \sum_{i=1}^{n^2} (z_i - x_i)^2 w_{ii} + 2 \sum_{i<j} (z_i - x_i)(z_j - x_j) w_{ij}
\end{aligned}$$

with the notation:

$$(\mathbf{H}^T\mathbf{H})_{ij} = w_{ij}.$$

Due to the independence of z_i:

$$\begin{aligned}\mathbf{E}_\mathbf{z}[\|\mathbf{H}(\mathbf{z}-\mathbf{x})\|^2] &= \sum_{i=1}^{n^2} Var(z_i)w_{ii} + 2\sum_{i<j} Cov(z_i,z_j)w_{ij} = \sum_{i=1}^{n^2} Var(z_i)w_{ii} + 0 \\ &= \sum_{i=1}^{n^2} x_i w_{ii}\end{aligned}$$

We can easily show that $w_{ii} = \|\mathbf{C_i}\|_2^2$
Finally:

$$\mathbf{E}[\|\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] = \underbrace{\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 + M^2\sigma^2}_{\text{Variance}} \tag{43}$$

$$+ \underbrace{\|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2 + \sum_{i=1}^{n^2} x_i \|\mathbf{C_i}\|_2^2}_{\text{Bias}} \tag{44}$$

7.2 Analysis

The previous computations show that our expansion results in the same data term as (36) with the addition of $\sum_{i=1}^{n^2} x_i\|\mathbf{C_i}\|_2^2$ which also expresses sparsity of $\mathbf{x}$. From (37):

$$\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 \ll \|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2$$

with (5):

$$\|\mathbf{C}_j\|_2^2 < \|\mathbf{C}_j\|_1$$

we have:

$$\sum_{j=1}^{n^2} (\|\mathbf{C}_j\|_1 + \|\mathbf{C}_j\|_2^2)x_j + \|\mathbf{b}\|_1 < 2\sum_{j=1}^{n^2} \|\mathbf{C}_j\|_1 x_j + \|\mathbf{b}\|_1 \tag{45}$$

$$\ll \|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2 \tag{46}$$

finally the following approximation hohlds:

$$\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] \approx \|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2 + M^2\sigma^2 \qquad (47)$$

It thus makes sense to approximate the cost functional as before:

$$\mathcal{L}_{\phi}^{gene}(\mathbf{x},\mathbf{b}) \approx \frac{\gamma}{2}\|\mathbf{H}\mathbf{x} + \mathbf{b} - \mathbf{E}[\mathbf{y}^{real}]\|^2 + \frac{\gamma}{2}M^2\sigma^2 - \delta\mathbf{E}[\mathcal{D}_{\phi}(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] + \lambda_1\|\mathbf{x}\|_1 + \frac{\lambda_2}{2}\|\nabla\mathbf{b}\|^2$$
$$\mathcal{L}_{\mathbf{x},\mathbf{b}}^{Disc}(\phi) = \mathbf{E}[\mathcal{D}_{\phi}(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] - \mathbf{E}[\mathcal{D}_{\phi}(\mathbf{y}^{real})] + \lambda_D(\|\nabla_y\mathcal{D}_{\phi}(\mathbf{y}^{mix})\| - 1)^2$$

We notice that the gradient can be calculated directly.

7.3 Gradient Calculation

Here we calculate the gradient of $\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))]$ with respect to $\mathbf{x}$ and $\mathbf{b}$ in the general model (23), where:

$$f(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b})) = \frac{\gamma}{2}\ \|\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{y}^{real}\|^2 - \delta\mathcal{D}_{\phi}(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))$$

we have:

$$\begin{aligned}\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x},\mathbf{b}))] &= \mathbf{E}_{\mathbf{z}}[\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z},\mathbf{b}))|\mathbf{z}]] \\ &= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[h(\mathbf{z},\mathbf{b})] \qquad (48)\end{aligned}$$

with $h : \mathbf{R}^{n^2} \to \mathbf{R}$ as:

$$h(\mathbf{z},b) = \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z},\mathbf{b}))]$$

To compute the gradient of the right term in (48) with respect to $\mathbf{x}$, we apply directly (29) in the specific case where $\mathbf{H} = Id$ and $\mathbf{b} = 0$:

$$\nabla_{\mathbf{x}}\mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[h(\mathbf{z},\mathbf{b})] = \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\{h(\mathbf{z}+\mathbf{1_i},\mathbf{b}) - h(\mathbf{z},\mathbf{b})\}_{1\leqslant i\leqslant n^2}]$$

Note that:

$$h(\mathbf{z}+1_i,b) = \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z}+1_i,\mathbf{b}))]$$

And we have:

$$\begin{aligned}\hat{\mathbf{y}}^{sim}(\mathbf{z}+1_i,\mathbf{b}) &\sim \mathcal{P}(\mathbf{H}(\mathbf{z}+1_i)+\mathbf{b}) \\ &\sim \mathcal{P}(\mathbf{H}\mathbf{z}+\mathbf{C}_i+\mathbf{b})\end{aligned}$$

For which:

$$\hat{\mathbf{y}}^{sim}(z + 1_i, \mathbf{b}) = \hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b} + \mathbf{C}_i)$$

We can then deduce:

$$\begin{aligned}
\nabla_x \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))] =& \nabla_x \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[h(\mathbf{z}, \mathbf{b})] \\
&= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\{h(\mathbf{z} + \mathbf{1_i}, \mathbf{b}) - h(\mathbf{z}, \mathbf{b})\}_{1\leqslant i\leqslant n^2}] \\
&= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\{\mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[f(\hat{\mathbf{y}}^{sim}(z, \mathbf{b} + \mathbf{C}_i))] - \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[f(\hat{\mathbf{y}}^{sim}(z, \mathbf{b}))]\}_{1\leqslant i\leqslant n^2}] \\
&= \{\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b} + \mathbf{C}_i))] - \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))]\}_{1\leqslant i\leqslant n^2}
\end{aligned}$$

To calculate the gradient of $\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b})]$ with respect to $\mathbf{b}$, we use the differentiation under the integral sign for the term on the right of equation (48), and then we apply the formula (29):

$$\begin{aligned}
\nabla_b \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))] &= \nabla_b \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[h(\mathbf{z}, \mathbf{b})] \\
&= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\nabla_b h(\mathbf{z}, \mathbf{b})] \\
&= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\{\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b}) + 1_i)] - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b}))]\}_{1\leqslant i\leqslant M^2}] \text{ Recalling (33)} \\
&= \{\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) + 1_i)] - \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))]\}_{1\leqslant i\leqslant M^2}
\end{aligned}$$

7.4 An approximation of the gradient

Computing the gradient of $\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b})]$ with respect to $\mathbf{x}$ is of size n^2 is beyond the capacity to store it in memory. Therefore, we use an approximation that reduces such complexity from n^2 to M^2.

For that, we use the following approximation:

$$h(\mathbf{z} + \mathbf{1_i}, \mathbf{b}) - h(\mathbf{z}, \mathbf{b}) \approx \frac{dh(\mathbf{z}, \mathbf{b})}{dz_i} \tag{49}$$

we obtain:

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[h(\mathbf{z}, \mathbf{b})] &= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\{h(\mathbf{z} + \mathbf{1_i}, \mathbf{b}) - h(\mathbf{z}, \mathbf{b})\}_{1\leqslant i\leqslant n^2}] \\
&\approx \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\nabla_z h(\mathbf{z}, \mathbf{b})]
\end{aligned}$$

For which:

$$\begin{aligned}
\mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\nabla_z h(\mathbf{z}, \mathbf{b})] &= \mathbf{E}_{\mathbf{z}\sim\mathcal{P}(\mathbf{x})}[\mathbf{H}^T\{\mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b}) + 1_i)] - \mathbf{E}[f(\hat{\mathbf{y}}^{sim}(\mathbf{z}, \mathbf{b}))]\}_{1\leqslant i\leqslant M^2}] \\
&= \mathbf{H}^T\{\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) + 1_i)] - \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))]\}_{1\leqslant i\leqslant M^2}
\end{aligned} \tag{50}$$

we observed that with this approximation, we can reduce the calculation of the gradient from size n^2 to M^2.

8 Optimisation

In this section, we solve the optimization problem (24)-(25). As explained earlier, the optimization process takes place in two steps:

1. **step 1:** We fix $\mathbf{x}$ and $\mathbf{b}$, and then we find the parameters $\hat{\phi}$ that minimize (25).
2. **step 2:** We fix $\hat{\phi}$, and then we find $\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}$ that minimize (24).

8.1 Optimisation of the discriminator

We use the Adam optimizer [12, 10] to optimize the discriminator(25), which is a convolutional neural network. Adam is a stochastic algorithm known for its excellent performance in deep learning due to its fast convergence and its ability to handle variable learning rates for different parameters [11].

8.2 Optimisation of the generator

To optimize the generator, we use proximal methods, which address the problem of optimizing non-smooth functions of the form $h(\mathbf{z}) = f(\mathbf{z}) + g(\mathbf{z})$. Here, f is differentiable, and g is continuous but not necessarily differentiable, often representing a regularization term. In our case :

$$\begin{aligned} f(\mathbf{x},\mathbf{b}) &= \frac{\gamma}{2}\mathbf{E}[\|\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}) - \mathbf{E}[\mathbf{y}^{real}]\|^2] - \delta\mathbf{E}[\mathcal{D}_\phi(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b}))] + \frac{\lambda_2}{2}\|\nabla\mathbf{b}\|^2 \\ g(\mathbf{x},\mathbf{b}) &= \lambda_1\|\mathbf{x}\|_1 \quad \text{under} \quad \mathbf{x} \geqslant 0. \end{aligned} \tag{51}$$

We minimize $h(\mathbf{x},\mathbf{b})$ with respect to $\mathbf{x}$ and $\mathbf{b}$ using proximal methods associated with the operator $prox_g(.)$.

We have:

$$prox_{\lambda_1\|.\|_1}(x_i) = \begin{cases} x_i + \lambda_1 & \text{if} x_i < \lambda_1 \\ x_i - \lambda_1 & \text{if } x_i \geqslant \lambda_1 \end{cases}$$

Due to the positivity constraint on g:

$$\begin{aligned} prox_g(x_i) &= ReLU(x_i - \lambda_1) \\ ReLU(a) &= max(a, 0). \end{aligned}$$

The simplest proximal algorithm to optimize functions in the form as h is proximal gradient descent, which is the combination of explicit gradient descent for the smooth functional and an implicit one for the non-smooth one with a fixed learning rate τ. We initialize $\mathbf{x}_0 \in \mathcal{X}$, then the iteration reads:

$$\mathbf{x}_{k+1} = prox_{\tau g}(\mathbf{x}_k - \tau \nabla_{\mathbf{x}} f(\mathbf{x}_k)) \tag{52}$$

This algorithm is used to optimize f with respect to $\mathbf{b}$. To optimize f with respect to $\mathbf{x}$, we use FISTA (Fast Iterative Shrinkage-Thresholding Algorithm)[4], which is one of the efficient and fast proximal methods in terms of convergence.

Random coordinate descent: The numerical computation of the gradient for models (22) and (23) is computationally expensive. Indeed, in model (22), the gradient has a size of M^2, and in model (23), it is of size n^2, making them computationally expensive to calculate within an iterative procedure. To overcome this problem, we use the random coordinate gradient (RCD) method[9, 7, 3].

Here, we are employing a simplified version of (RCD), which involves randomly selecting a block of coordinates with respect to which we compute the partial derivatives of the f, rather than computing the entire gradient $\nabla_x f$. In other words, for $\mathbf{x} = (x_{ij}) \in \mathbb{R}_+^{n \times n}$, let $\mathbf{I} \subseteq \{1, ..., n\}$ and $\mathbf{J} \subseteq \{1, ..., n\}$ two randomly selected blocks, each containing consecutive elements. We define $\nabla_x^{\mathbf{I},\mathbf{J}} f$:

$$(\nabla_x^{\mathbf{I},\mathbf{J}} f)_{ij} = \begin{cases} \frac{df(x_{ij})}{dx_{ij}} & \text{if } (i,j) \in \mathbf{I} \times \mathbf{J} \\ 0 & \text{otherwise} \end{cases}$$

In our numerical tests, we test both cases, where $\mathbf{I} \times \mathbf{J}$ is square and rectangular. Initially, we fix the size $|\mathbf{I}|$ and $|\mathbf{J}|$, and in each iteration, we randomly choose $\mathbf{I}$ and $\mathbf{J}$.

9 Numerical Results

In this section we report the numerical results obtained by the methods described previously. We perform two types of comparisons:

1. the results given by the approximate gradient [13] and the theoretical gradient based on RCD applied to model (22).

2. The general model (23) and model(22).

Algorithm 1: FluoGAN optimisation via RCD

Data: Initialization: $\mathbf{x}_0$, $\mathbf{b}_0$, ϕ, $|\mathbf{I}|$ and $|\mathbf{J}|$

Result: the update of $\mathbf{x}_0$, $\mathbf{b}_0$, ϕ

for *epoch: 1...*n_{epoch} **do**
  for *i: 1...*$n_{\mathcal{D}}$ **do**
    for *t: 1...*B **do**
      $\mathbf{y}_t^{real}$ *is randomly chosen from the training set*
      $\mathbf{y}_t^{sim} \leftarrow G(\mathbf{x}, \mathbf{b})$
      $\mathbf{y}_t^{mix} \leftarrow \eta_t \mathbf{y}_t^{real} + (1-\eta_t)\mathbf{y}_t^{sim}$ *where* $\eta_t \sim \mathcal{U}[0,1]$
      $\mathcal{L}_{\mathbf{x}_0,\mathbf{b}_0}^{Disc,t} \leftarrow \mathcal{D}_\phi(\hat{\mathbf{y}}^{sim}(\mathbf{x},\mathbf{b})) - \mathcal{D}_\phi(\mathbf{y}^{real})$
      $+ \lambda_D(\|\nabla_y \mathcal{D}_\phi(y^{mix})\| - 1)^2$
    end
    $\phi \leftarrow Adam(\nabla_\phi \frac{1}{B}\sum_{t=1}^{B} \mathcal{L}_{\mathbf{x}_0,\mathbf{b}_0}^{Disc,t}, \tau_{\mathcal{D}})$
  end
  for *i: 1...*n_G **do**
    Choose $\mathbf{I}$ *and* $\mathbf{J}$
    for *i,j in* $\mathbf{I} \times \mathbf{J}$ **do**
      Initialize : $grad^{\mathbf{I},\mathbf{J}} = 0$
      for *t: 1...*B **do**
        $\mathbf{y}_t^{real}$ *is randomly chosen from the training set*
        $\mathbf{y}_t^{sim} \leftarrow G(\mathbf{x}, \mathbf{b})$
        $f^t + \leftarrow -\delta(\mathcal{D}_\phi(\mathbf{y}_t^{sim} + 1_{ij}) - \mathcal{D}_\phi(\mathbf{y}_t^{sim}))$
        $grad = f^t$
      end
    end
    $\mathbf{x}_0 \leftarrow ReLU(\mathbf{x}_0 - \tau_1(\frac{grad}{B} + \nabla_x \|\mathbf{H}\mathbf{x} + \mathbf{b}\|^2) - \lambda_1)$
    $\mathbf{b}_0 \leftarrow ReLU(\mathbf{b}_0 - \tau_2(\frac{grad}{B} + \nabla_b \|\mathbf{H}\mathbf{x} + \mathbf{b}\|^2) - 2\lambda_2 \nabla^2 b)$
  end
end

9.1 Data

9.1.1 Ground Truth

The input to the microscope, $\mathbf{x}^* \in \mathbb{R}^{n \times n}$ is an image of size $n \times n$ that contains 14 sets of parallel lines of $100nm$ width. The separation distance d (center-to-center distance) between the two middle lines of each set is gradually increasing with a rate of $30nm$. The knowledge of these details allows us to exactly quantity the resolution level reached by each approach.

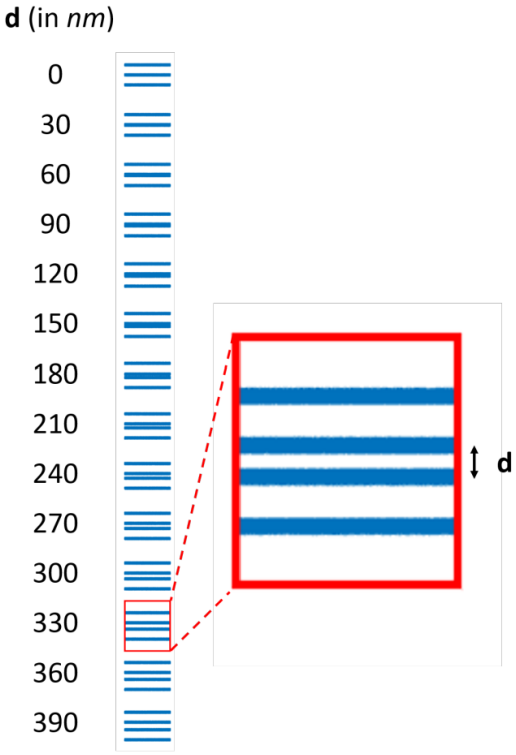


Figure 6: Ground truth pattern

In our case we use the filament block for which d=120nm:

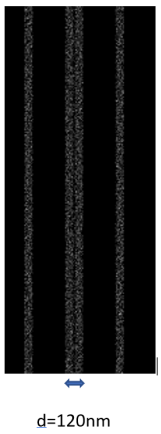


Figure 7: Ground truth of the size 120x120

The pixel size is equal to 103nm.

The space-variant background image, $\mathbf{b}^* \in \mathbb{R}^{M \times M}$, is of size $M \times M$ with $M = 20$. It models the out of focus and ambient molecules and it is add after the convolution and the undersampling operation of the truth groud with a factor resolution L=6.

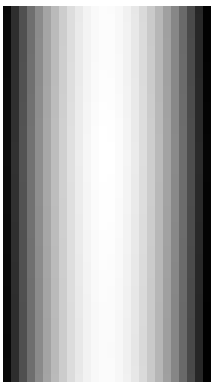


Figure 8: space-variant background $\mathbf{b}^*$

9.2 Real images

The $\{\mathbf{y}_t^{real}\}_t$ are generated from **SOFI**-tools [6]. It models the emitters as alternating sources that operate in cycles. Each emitter switches from an active state to an inactive state once per cycle according to a fixed distribution.

We have a video of $\{\mathbf{y}_t^{real}\}_t$ with 500 frames at a rate of 100 frames per second. It models the sequence of images of size 20x20 taken by the microscope after the acquisition process.

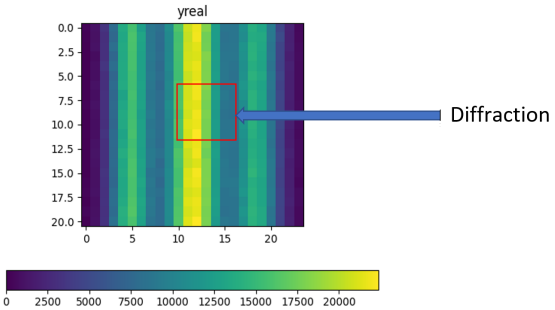


Figure 9: output of the miscroscope

9.3 Model setting

9.3.1 Generator

The generator is characterized by various factors:

- FWHM: Full Width at Half Maximum which characterizes the standard deviation of the Gaussian kernel.
- Undersampling coefficient: makes the image coarser to an image on a less fine grid.
- σ: variance of the Gaussian noise.
- kwidth: The size of the Gaussian kernel support is determined by its spread or standard deviation (sigma). A larger standard deviation will result in a larger support for the kernel. The support essentially defines how many neighboring pixels will be considered when applying the kernel for convolution.
- ksigma: Standard deviation of the Gaussian kernel which is linked to the value FWHM:

$$ksigma = \frac{FHWM}{2.35 \times px} \times undersampling$$

with px is the size of the pixel which $103nm$.

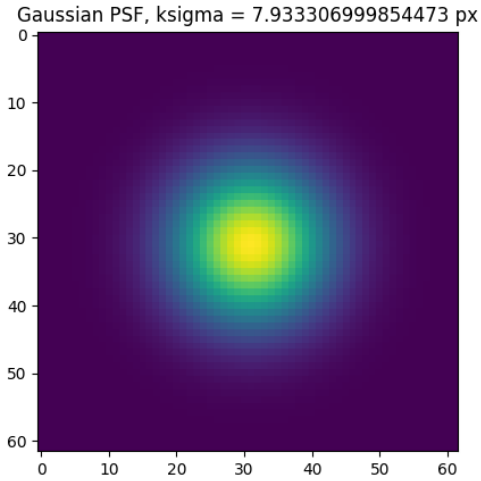


Figure 10: Gaussian Kernel

Parametres	value
FWHM	318nm
Undersampling	6
σ	0.5
kwidth	61
ksigma	7.93

9.3.2 Discriminator

The discriminator is a convolutionel neural network. It helps to extract features from the image $\mathbf{y}^{real}$ for efficient comparison with $\mathbf{y}^{sim}$.

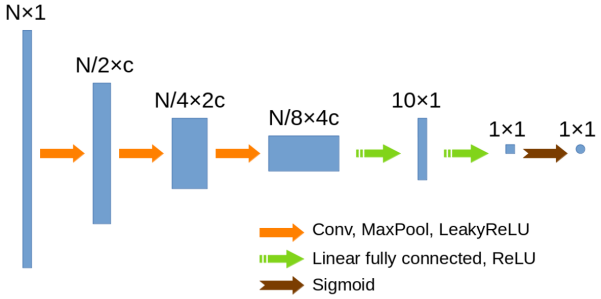


Figure 11: discriminator's architecture

Parametres	value
number of convolutional layers	3
number of initial channels	32
the size of convolutions	3
max pooling	1

Parameters of training:

Parametres	value
Number of iterations for the discriminator per epoch	1
Number of iterations for the generator per epoch	1
Number of epochs	5000
Discriminator GP regularization parameter	10^{-6}
Fidelity term parameter of $\mathcal{D}(\mathbf{y}^{sim})$	0.25
L^2 fidelity term parameter	1
L_1 regularization parameter	1
learning rate for $\mathbf{x}$	1.5
learning rate for $\mathbf{b}$	2.5
learning rate for of the discriminator	2×10^{-6}
batch size for mono poiss	32
batch size for double poisson	200

9.4 Numerical results

In this section, we discuss different results:

- Comparison of the model (22) with approximate gradient and true one (31).
- RDC performance in (22) in both square and rectangle cases.
- Testing the general fluctuation model (23) and compare it to (22).

9.5 The performance and comparison criterion

To compare the reconstructed images in the single Poisson and double Poisson models under both true gradient and random coordinate descent, with the image reconstructed in the first model using the approximate gradient, we will utilize the Root Mean Square Error (RMSE) criterion:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_1^i - \mathbf{y}_2^i)^2}.$$

The RMSE value quantifies the average magnitude of errors made by the model. In our case, it indicates how much the reconstructed image is close from the one reconstructed by the approximate gradient, while larger RMSE values suggest greater discrepancies.

9.6 The true and approximate gradient

In this section, image reconstruction is performed using the single Poisson model under both the true gradient and the approximated gradient. Subsequently, we will apply the Random Coordinate Descent (RCD) algorithm to the same model, followed by an analysis of the results.
First comparison:

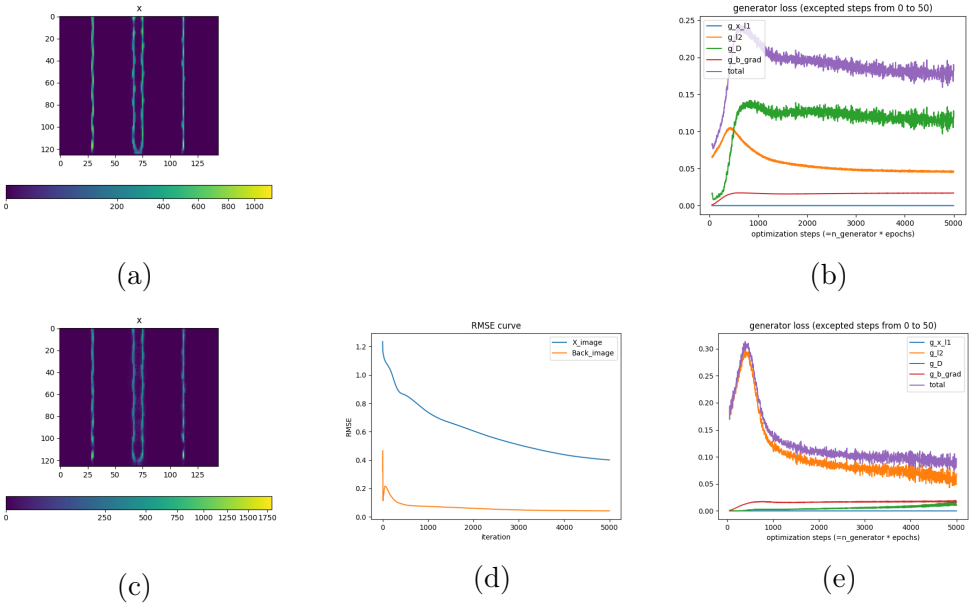


Figure 12: (a) and (b) correspond, respectively, to the image reconstructed by the single Poisson model with the approximated gradient and the associated convergence plot. (c), (e), (d) correspond to the same model with the true gradient.

In the convergence plot, we plot the evolution of five functions corresponding to the terms that appear in the optimization objective (25) (Generator function), which is that of the generator. Notably, in the case of the approximated gradient, the minimized function (in purple) gradually decreases. However, this decrease is relatively slow compared to the true gradient case, where the objective function rapidly converges to its minimum. It is observed that in

the case of the approximated gradient, the discriminator's role in achieving convergence is relatively slower. This could be attributed to the instability of GANs when separating closely related distributions. Initially, the L2 fidelity term in our objective stabilizes the optimization process, converging to an image with a distribution close to the real image. Subsequently, the discriminator's role comes into play, distinguishing the two based on finer details. We also observe that the RMSE with respect to $\mathbf{x}$ decreases over iterations. This implies that the image reconstructed with the true gradient is close to the one reconstructed with the approximated gradient. However, this doesn't provide much insight into the performance of the case with the approximated gradient, as we are essentially working at a microscopic scale where even small differences can have a significant impact. As for the background, we see that the RMSE converges to 0; hence, the background reconstructed with the true gradient remains the same as with the approximated gradient.

RDC algorithm:
Here, we will apply Random Coordinate Descent (RCD) in the optimization process. Indeed, reconstruction with the true gradient is computationally expensive and time-consuming (the previous reconstruction took 1 hour to execute). With this approach, we aim to reduce the computation time to approximately 4 to 8 times less than the time required for the previous reconstruction. However, it's important to note that this method does not guarantee achieving similar or better results.

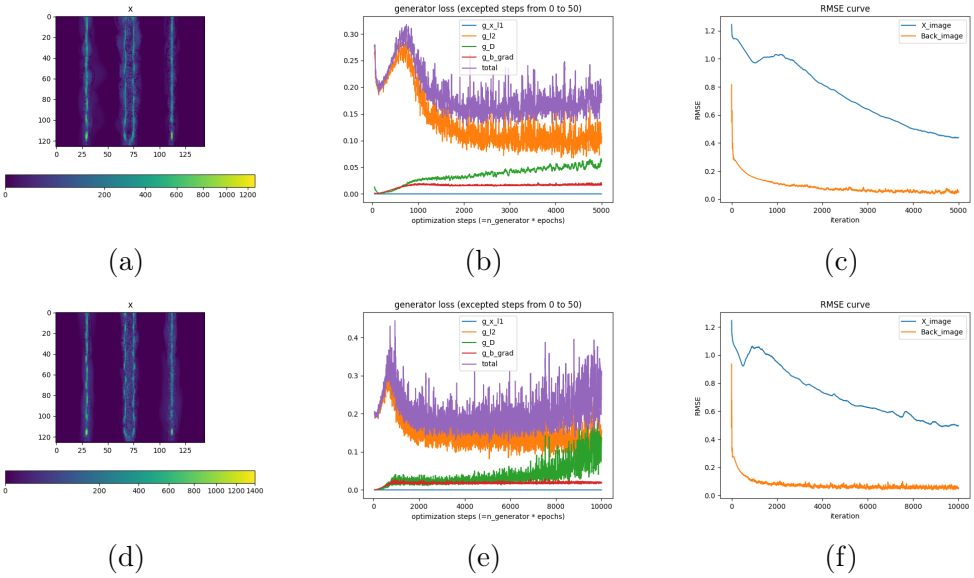


Figure 13: The first line corresponds to the choice where $\mathbf{I} \times \mathbf{J}$ is a rectangle (same notation as the optimisation section) which $\|\mathbf{I}\| = 20$ and $\|\mathbf{j}\| = 5$. In this case, if we view the gradient as an image of the same size as the reconstructed image, at each iteration, we randomly select a rectangle for which we calculate the associated partial derivatives, and the rest are set to zero then we use it instead of the true gradient.The second line corresponds to the choice of squares where $\|\mathbf{I}\| = 5$ and $\|\mathbf{j}\| = 5$.

Here, we observe that the convergence curves of the objective function associated with each case exhibit significant oscillations. This is attributed to the random nature of the block selection. Furthermore, we note that in the case of rectangles, the objective function and the RMSE decreases to approximately 0.1 and 0.4, respectively, which corresponds to the case with the true gradient. In the case of squares, we observe that full convergence is not achieved due to the increasing role of the discriminator term. This comparison leads to the conclusion that selecting blocks compatible with the geometric structure of the true image yields better optimization results and objective function convergence.

9.7 Evaluation of the complete fluctuation model

In this section, we perform image reconstruction using the double Poisson model with the gradient calculated using the previous approximation (50). Firstly, we compute the right-hand side term of (50) which is $\{\mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}) + 1_i)] - \mathbf{E}[f(\mathbf{y}^{sim}(\mathbf{x}, \mathbf{b}))]\}_{1 \leqslant i \leqslant M^2}$, which has a similar computational cost to the

true gradient in the single Poisson model. Subsequently, we will apply Random Coordinate Descent (RDC) to reduce the computation time.

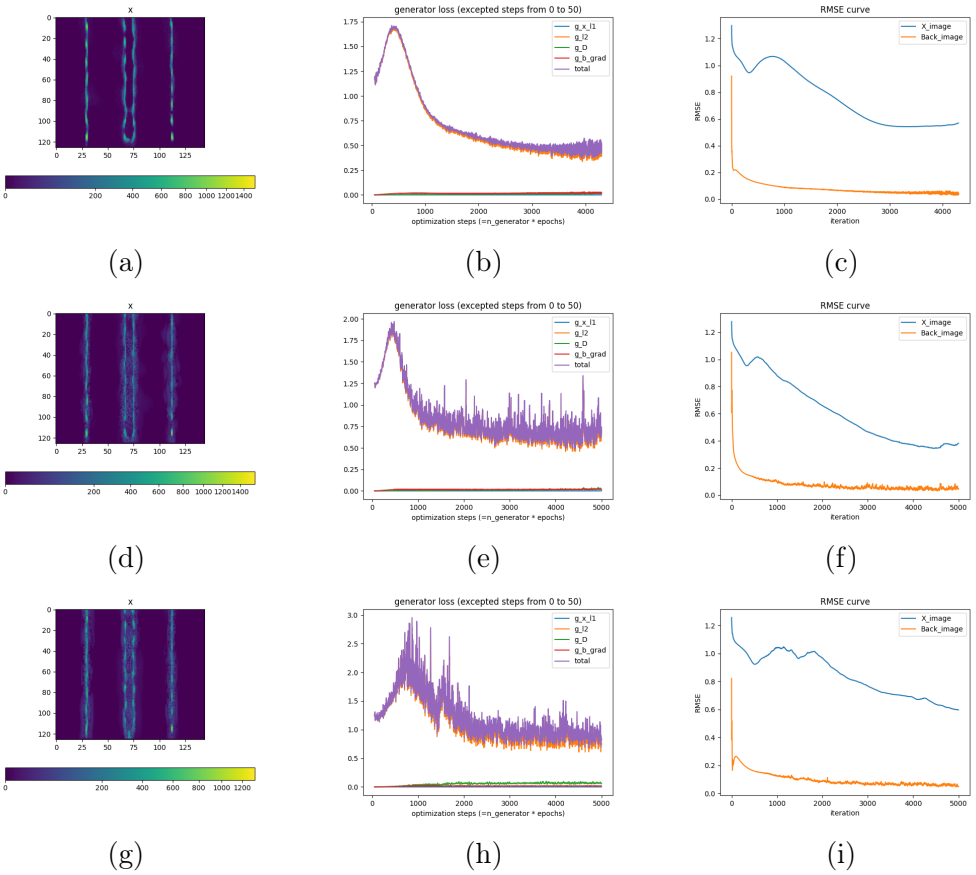


Figure 14: The first row of images corresponds to the double Poisson model with the complete computation of the approximate gradient (50) (without RDC). The two subsequent rows correspond to the application of RDC in the cases of rectangular and square blocks.

We observe that in the case without RDC, the objective function converges rapidly after a few iterations, and towards the end, a slight oscillation occurs due to discriminator instability. This could be interpreted by the fact that the double Poisson model provides a better description of the real data. We also notice that the RMSE decreases at a slower rate than before, which is attributed to the differences between the two models, but it also suggests that the images reconstructed by each model are not significantly different. Furthermore, in the RDC case, the curve corresponding to rectangles converges, on average, more quickly than in the case of squares, confirming the hypoth-

esis that block selection should be compatible with the geometric structure of the true image.

10 Conclusion

This internship focused on the development of the work already initiated from a broader perspective. We were able to deploy a better model describing fluorescence fluctuations (double Poisson model), both theoretically and in practical applications. Notably, the analysis and development of the data fidelity term and gradient computation constituted a significant portion of the internship which was helpful to justify the existing and new approaches from a theoretical perspective. Moreover, the encouraging numerical results pave the way for further development and application of this work to real and more complex data. The choice of regularization parameters and learning rates in the numerical part was crucial for the successful convergence of the algorithms and demanded a substantial amount of time. In the end, this internship has provided me with a solid foundation in the field of inverse problems in the field of fluorescence microscopy.

References

[1] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser Basel, 2005. ISBN: 978-3-7643-8721-1. DOI: 10.1007/978-3-7643-8722-8.

[2] Cédric Villani. *Optimal Transport*. Springer Berlin, Heidelberg, 2009. ISBN: 978-3-540-71050-9 Published: 26 October 2008. DOI: https://doi.org/10.1007/978-3-540-71050-9.

[3] A. K. Bakhtin, G. P. Bakhtina, and I. V. Denega. *Estimates of product of inner radii of mutually non-overlapping domains in multidimensional complex spaces*. 2012. arXiv: 1207.4893 [math.CV].

[4] Jian Li, Jiufeng J. Tu, and Joseph L. Birman. *Optical selection rules in topological insulators Bi_2Sb_3, Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3*. 2012. arXiv: 1204.3068 [cond-mat.mes-hall].

[5] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[6] Arik Girsault et al. "SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging". In: *PLOS ONE* 11.9 (Sept. 2016), pp. 1–13. DOI: 10.1371/journal.pone.0161602. URL: https://doi.org/10.1371/journal.pone.0161602.

[7] Fulvio Melia. "Definitive test of theiR/isubh/sub = ict/iuniverse using redshift drift". In: *Monthly Notices of the Royal Astronomical Society: Letters* 463.1 (Aug. 2016), pp. L61–L63. DOI: 10.1093/mnrasl/slw157. URL: https://doi.org/10.1093%2Fmnrasl%2Fslw157.

[8] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein Generative Adversarial Networks". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, June 2017, pp. 214–223. URL: https://proceedings.mlr.press/v70/arjovsky17a.html.

[9] Marco Boggi and Eduard Looijenga. *Deforming a canonical curve inside a quadric*. 2017. arXiv: 1702.00770 [math.AG].

[10] Nitish Shirish Keskar and Richard Socher. "Improving Generalization Performance by Switching from Adam to SGD". In: *CoRR* abs/1712.07628 (2017). arXiv: 1712.07628. URL: http://arxiv.org/abs/1712.07628.

[11] Sebastian Bock, Josef Goppold, and Martin Georg Weiß. "An improvement of the convergence proof of the ADAM-Optimizer". In: *CoRR* abs/1804.10587 (2018). arXiv: 1804.10587. URL: http://arxiv.org/abs/1804.10587.

[12] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. "On the Convergence of Adam and Beyond". In: *CoRR* abs/1904.09237 (2019). arXiv: 1904.09237. URL: http://arxiv.org/abs/1904.09237.

[13] Mayeul Cachia et al. "Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)". In: *Inverse Problems* 39.5 (Apr. 2023), p. 054006. DOI: 10.1088/1361-6420/acc889. URL: https://dx.doi.org/10.1088/1361-6420/acc889.