Cell detection by functional inverse diffusion and non-negative group sparsity

Pol del Aguila Pla, Ph.D. Candidate
https://poldap.github.io, https://github.com/poldap
Division of Information Science and Engineering
School of Electrical Engineering and Computer Science

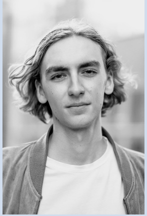
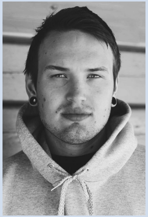
January 28, 2019 at PARIETAL

KTH Royal Institute of Technology

Acknowledgements





J. Jaldén [1]–[5] | V. Saxena [5] | G. Bengtsson | J. Larsson | J. Sörell | E. Ågeby

Royal Swedish Academy of Sciences

Knut and Alice Wallenberg foundation

MabTech AB

Swedish Research Council

KTH Opportunities and EECS school

1. P. del Aguila Pla and J. Jaldén, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse problems," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5407–5421, Oct. 2018
2. P. del Aguila Pla and J. Jaldén, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5422–5437, Oct. 2018
3. P. del Aguila Pla and J. Jaldén, "Cell detection on image-based immunoassays," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2018, pp. 431–435
4. P. del Aguila Pla and J. Jaldén, "Convolutional group-sparse coding and source localization," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 2776–2780
5. P. del Aguila Pla, V. Saxena, and J. Jaldén, "SpotNet — Learned iterations for cell detection in image-based immunoassays,", Accepted in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI), Apr. 2019

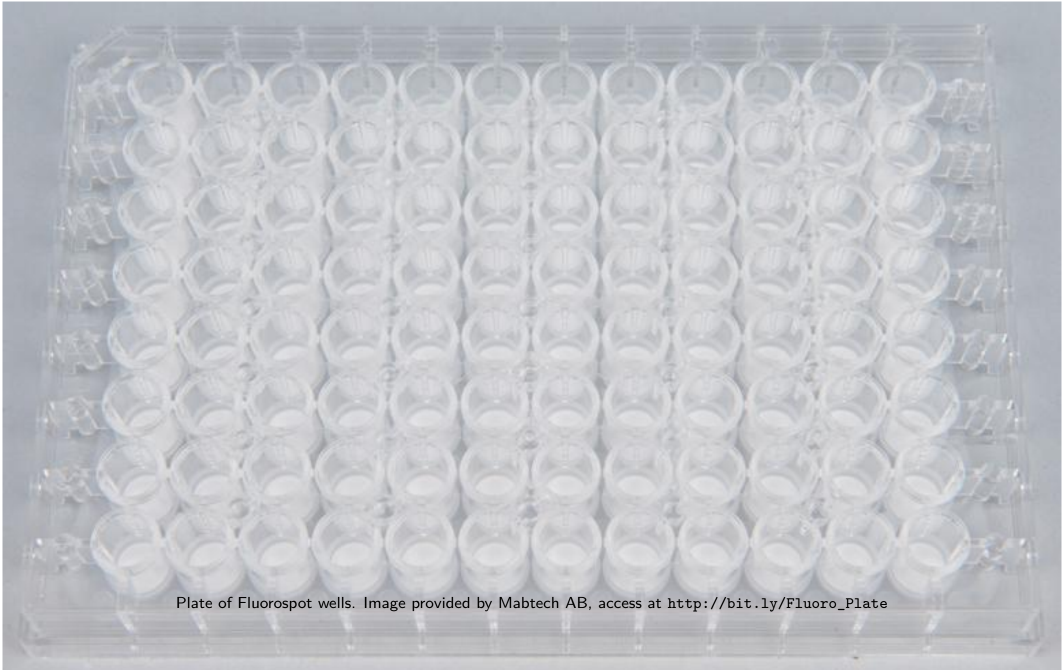


Plate of Fluorospot wells. Image provided by Mabtech AB, access at http://bit.ly/Fluoro_Plate

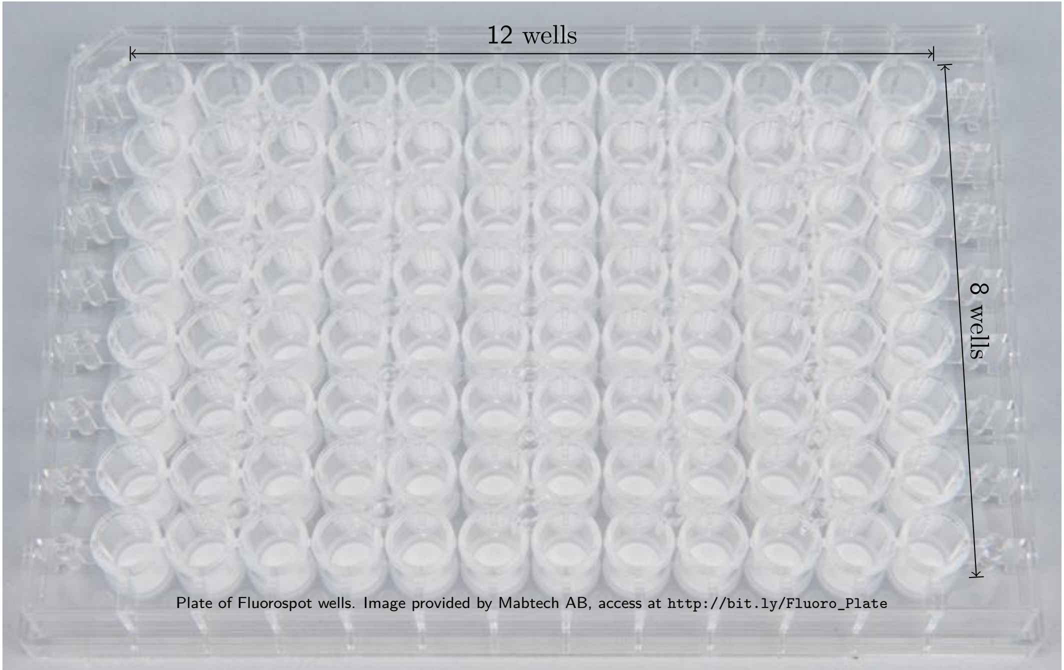


Plate of Fluorospot wells. Image provided by Mabtech AB, access at http://bit.ly/Fluoro_Plate

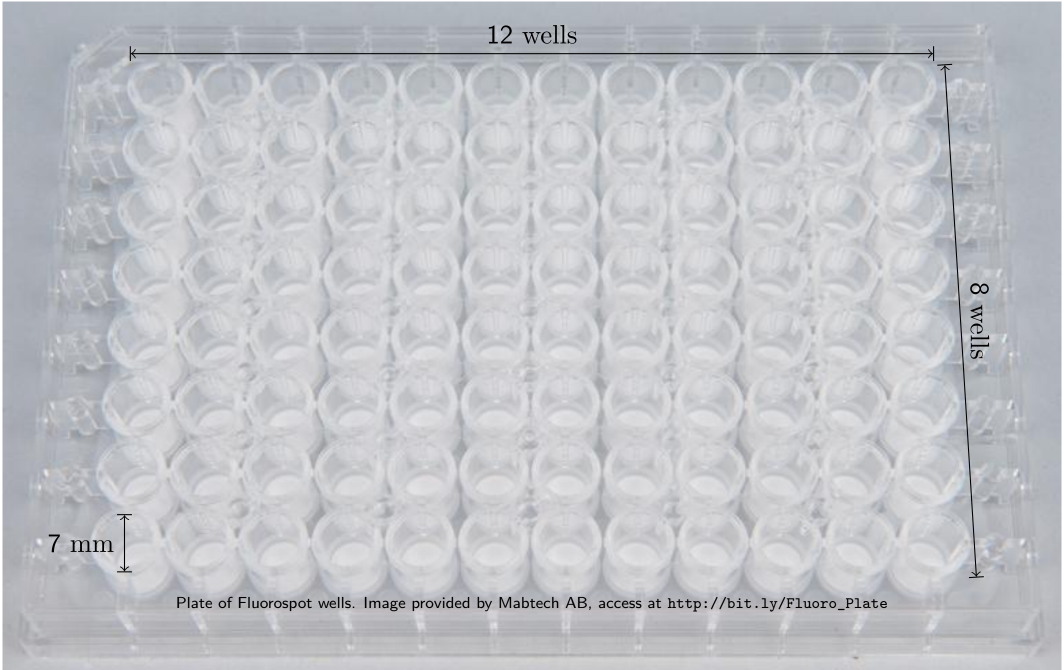
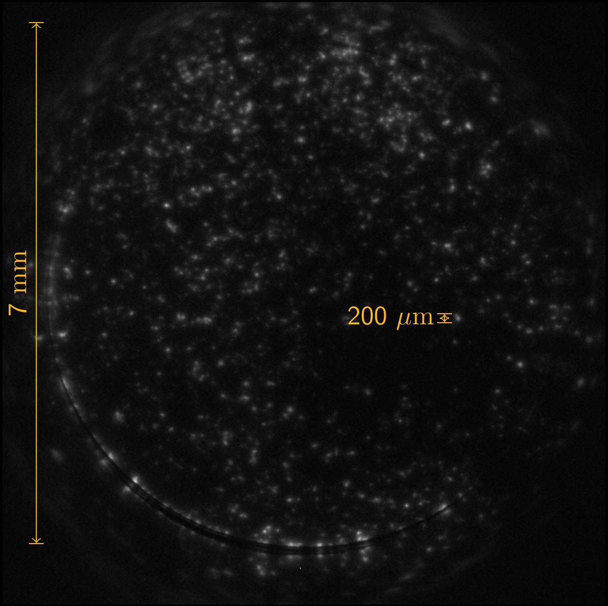
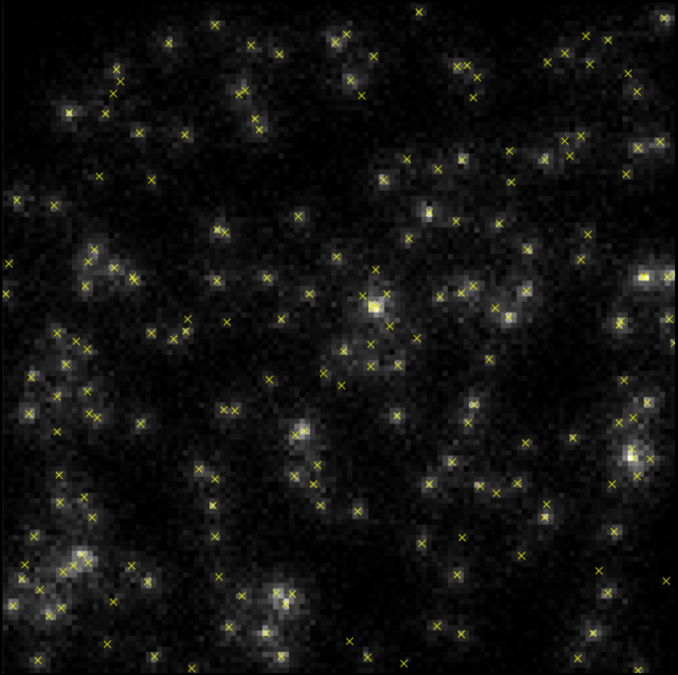
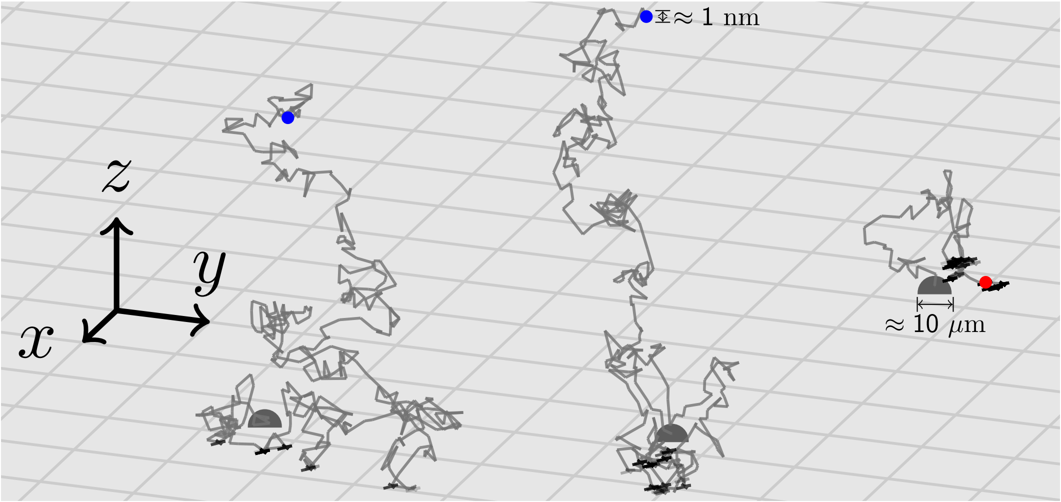


Plate of Fluorospot wells. Image provided by Mabtech AB, access at http://bit.ly/Fluoro_Plate



Fluorospot image, provided by Mabtech AB





≈ 1 nm
z
y
x
≈ 10 μm

A Physical Model for Biomedical Assays (Modeling I)

Relevant quantities for the assay are

- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\mathrm{obs}}(x, y) = d(x, y, T)$, which evolves coupled to

A Physical Model for Biomedical Assays (Modeling I)

Relevant quantities for the assay are

- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\text{obs}}(x, y) = d(x, y, T)$, which evolves coupled to
- the 3D density of free particles $c(x, y, z, t) \geq 0$ on $z \geq 0$, and to

A Physical Model for Biomedical Assays (Modeling I)

Relevant quantities for the assay are

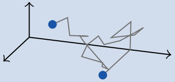
- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\mathrm{obs}}(x, y) = d(x, y, T)$, which evolves coupled to
- the 3D density of free particles $c(x, y, z, t) \geq 0$ on $z \geq 0$, and to
- the source density rate of new particles $s(x, y, t) \geq 0$, that is spatially sparse and reveals the cell locations and secretion over time.

A Physical Model for Biomedical Assays (Modeling I)

Relevant quantities for the assay are

- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\mathrm{obs}}(x, y) = d(x, y, T)$, which evolves coupled to
- the 3D density of free particles $c(x, y, z, t) \geq 0$ on $z \geq 0$, and to
- the source density rate of new particles $s(x, y, t) \geq 0$, that is spatially sparse and reveals the cell locations and secretion over time.

$$
\begin{aligned}
\frac{\partial}{\partial t} c &= D \Delta c \,, \\
\frac{\partial}{\partial t} d &= \kappa_{\mathrm{a}} c\big|_{z=0} - \kappa_{\mathrm{d}} d \,, \\
-D \frac{\partial}{\partial z} c\big|_{z=0} &= s - \frac{\partial d}{\partial t} \,.
\end{aligned}
$$



This physical model was presented before, also for ELISPOT and Fluorospot.

A Physical Model for Biomedical Assays (Modeling I)

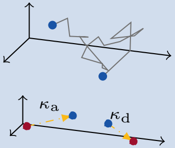
Relevant quantities for the assay are

- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\mathrm{obs}}(x, y) = d(x, y, T)$, which evolves coupled to
- the 3D density of free particles $c(x, y, z, t) \geq 0$ on $z \geq 0$, and to
- the source density rate of new particles $s(x, y, t) \geq 0$, that is spatially sparse and reveals the cell locations and secretion over time.

$$\frac{\partial}{\partial t} c = D \Delta c \,,$$

$$\frac{\partial}{\partial t} d = \kappa_{\mathrm{a}} c\Big|_{z=0} - \kappa_{\mathrm{d}} d \,,$$

$$-D \frac{\partial}{\partial z} c\Big|_{z=0} = s - \frac{\partial d}{\partial t} \,.$$



This physical model was presented before, also for ELISPOT and Fluorospot.

A Physical Model for Biomedical Assays (Modeling I)

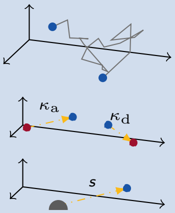
Relevant quantities for the assay are

- A density of bound particles $d(x, y, t) \geq 0$, where the image will be $d_{\mathrm{obs}}(x, y) = d(x, y, T)$, which evolves coupled to
- the 3D density of free particles $c(x, y, z, t) \geq 0$ on $z \geq 0$, and to
- the source density rate of new particles $s(x, y, t) \geq 0$, that is spatially sparse and reveals the cell locations and secretion over time.

$$\frac{\partial}{\partial t} c = D \Delta c \,,$$

$$\frac{\partial}{\partial t} d = \kappa_{\mathrm{a}} c\Big|_{z=0} - \kappa_{\mathrm{d}} d \,,$$

$$-D \frac{\partial}{\partial z} c\Big|_{z=0} = s - \frac{\partial d}{\partial t} \,.$$



This physical model was presented before, also for ELISPOT and Fluorospot.

An Observation Model for Biomedical Assays (I) (Modeling II)

We consider the image observation $d_{\mathrm{obs}} \in \mathcal{D}_+$, with $\mathcal{D} = \mathrm{L}^2\left(\mathbb{R}^2\right)$ and prove that

$$d_{\mathrm{obs}}(x, y) = \int_0^{\sigma_{\mathsf{max}}} \left(g_\sigma(\bar{x}, \bar{y}) * a(\bar{x}, \bar{y}, \sigma)\right)(x, y) \, \mathrm{d}\sigma \,,$$

with $a \in \mathcal{A}_+$ and $\mathcal{A} \subset \mathrm{L}^2\left(\mathbb{R}^2 \times \mathbb{R}_+\right)$ a space of functions with bounded spatial support, $\sigma_{\mathsf{max}} = \sqrt{2DT}$, and

An Observation Model for Biomedical Assays (I) (Modeling II)

We consider the image observation $d_{\mathrm{obs}} \in \mathcal{D}_+$, with $\mathcal{D} = \mathrm{L}^2\left(\mathbb{R}^2\right)$ and prove that

$$d_{\mathrm{obs}}(x, y) = \int_0^{\sigma_{\max}} G_\sigma a_\sigma \,\mathrm{d}\sigma, \left(\text{discr. } \tilde{d}_{\mathrm{obs}} = \sum_{k=1}^{K} \tilde{g}_k \circledast \tilde{a}_k\right),$$

with $a \in \mathcal{A}_+$ and $\mathcal{A} \subset \mathrm{L}^2\left(\mathbb{R}^2 \times \mathbb{R}_+\right)$ a space of functions with bounded spatial support, $\sigma_{\max} = \sqrt{2DT}$, and

An Observation Model for Biomedical Assays (I) (Modeling II)

We consider the image observation $d_{\mathrm{obs}} \in \mathcal{D}_+$, with $\mathcal{D} = \mathrm{L}^2\left(\mathbb{R}^2\right)$ and prove that

$$d_{\mathrm{obs}}(x, y) = Aa, \text{ we call } A \text{ the } \textit{diffusion} \text{ operator},$$

with $a \in \mathcal{A}_+$ and $\mathcal{A} \subset \mathrm{L}^2\left(\mathbb{R}^2 \times \mathbb{R}_+\right)$ a space of functions with bounded spatial support, $\sigma_{\mathsf{max}} = \sqrt{2DT}$, and

An Observation Model for Biomedical Assays (I) (Modeling II)

We consider the image observation $d_{\mathrm{obs}} \in \mathcal{D}_+$, with $\mathcal{D} = \mathrm{L}^2\left(\mathbb{R}^2\right)$ and prove that

$$d_{\mathrm{obs}}(x, y) = Aa, \text{ we call } A \text{ the } \textit{diffusion} \text{ operator}\,,$$

with $a \in \mathcal{A}_+$ and $\mathcal{A} \subset \mathrm{L}^2\left(\mathbb{R}^2 \times \mathbb{R}_+\right)$ a space of functions with bounded spatial support, $\sigma_{\mathsf{max}} = \sqrt{2DT}$, and

- $a(x, y, \sigma)$ is an equivalent of $s(x, y, t)$ where the effect of adsorption and desorption have been summarized.

$$a(x, y, \sigma) = \frac{\sigma}{D} \int_{\frac{\sigma^2}{2D}}^{T} s(x, y, T - \eta)\, \varphi\left(\frac{\sigma^2}{2D}, \eta\right) \mathrm{d}\eta\,.$$

- $a(x, y, \sigma)$ preserves all the spatial information in $s(x, y, t)$.

An Observation Model for Biomedical Assays (II) (Modeling II)

The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\mathsf{max}}} G_\sigma a_\sigma \mathrm{d}\sigma \,.$$

How?

- Independence of Brownian motion in x, y and z.

An Observation Model for Biomedical Assays (II) (Modeling II)

The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\mathsf{max}}} G_\sigma a_\sigma \mathrm{d}\sigma \,.$$

How?

- Independence of Brownian motion in x, y and z.
- Adsorption (κ_{a}) and desorption (κ_{d}) only regulated by z-movement.

An Observation Model for Biomedical Assays (II) (Modeling II)

The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\max}} G_\sigma a_\sigma \mathrm{d}\sigma \,.$$

How?

- Independence of Brownian motion in x, y and z.
- Adsorption (κ_{a}) and desorption (κ_{d}) only regulated by z-movement.
- x- and y-movements only depend on τ, total time in Brownian motion. In particular, according to Green function for 2D diffusion, $g_{\sqrt{2D\tau}}(x, y)$.

An Observation Model for Biomedical Assays (II) (Modeling II)

The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\max}} G_\sigma a_\sigma \mathrm{d}\sigma \, .$$

How?

- Independence of Brownian motion in x, y and z.
- Adsorption (κ_{a}) and desorption (κ_{d}) only regulated by z-movement.
- x- and y-movements only depend on τ, total time in Brownian motion. In particular, according to Green function for 2D diffusion, $g_{\sqrt{2D\tau}}(x, y)$.
- $\varphi(\tau, t)$ summarizes the effect of adsorption and desorption onto the time in free motion τ for each time of final adsorption t.

An Observation Model for Biomedical Assays (II) (Modeling II)

The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\max}} G_\sigma a_\sigma \mathrm{d}\sigma \,.$$

How?

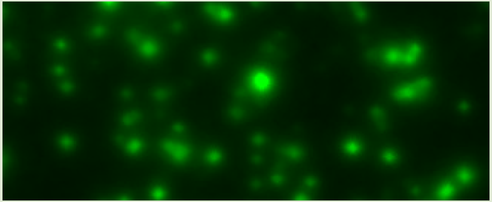
- Independence of Brownian motion in x, y and z.
- Adsorption (κ_{a}) and desorption (κ_{d}) only regulated by z-movement.
- x- and y-movements only depend on τ, total time in Brownian motion. In particular, according to Green function for 2D diffusion, $g_{\sqrt{2D\tau}}(x, y)$.
- $\varphi(\tau, t)$ summarizes the effect of adsorption and desorption onto the time in free motion τ for each time of final adsorption t.
- Change variables to those significative to x- and y-movement, $\sigma = \sqrt{2D\tau}$.

An Observation Model for Biomedical Assays (III) (Modeling II)

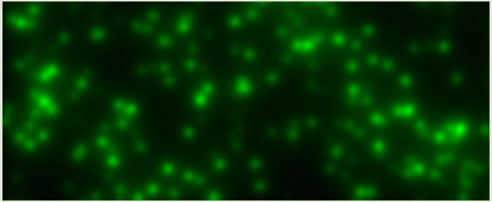
The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\max}} G_\sigma a_\sigma \mathrm{d}\sigma \, .$$

Consequences



Real observation (section)



Simulated observation (section)

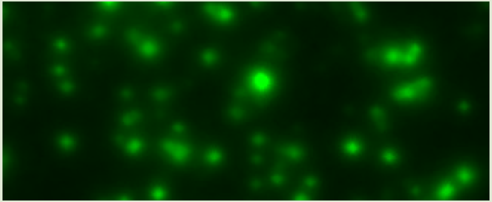
- Synthetic data

An Observation Model for Biomedical Assays (III) (Modeling II)

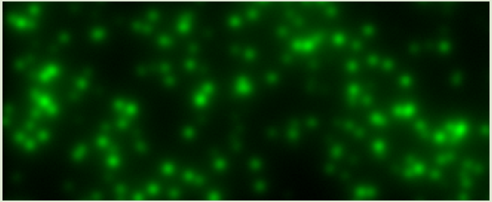
The modeling result: The image $d_{\mathrm{obs}} \in \mathcal{D}_+$ can be expressed as

$$d_{\mathrm{obs}} = \int_0^{\sigma_{\mathsf{max}}} G_\sigma a_\sigma \mathrm{d}\sigma \, .$$

Consequences



Real observation (section)



Simulated observation (section)

- Synthetic data
- An inverse problem

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \underbrace{\delta_{\mathcal{A}_+}(a)}_{\text{non-negative}} + \lambda \underbrace{\int_{\mathbb{R}^2} \left(\int_0^{\sigma_{\mathrm{max}}} \xi^2(\sigma) a^2(x, y, \sigma)\, \mathrm{d}\sigma \right)^{\frac{1}{2}} \mathrm{d}x\mathrm{d}y}_{\text{group-sparsity}} \right] ,$$

Proximal Optimization

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right],$$

with $\mathbf{r} = (x, y)$.

Proximal Optimization

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right] ,$$

with $\mathbf{r} = (x, y)$.

Proximal Optimization

- How do we solve this optimization problem? Can it be solved?

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a\in\mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right] ,$$

with $\mathbf{r} = (x, y)$.

Proximal Optimization

- How do we solve this optimization problem? Can it be solved?
- Convex problem, but the existance and unicity of the solution are not given (function spaces). Three terms, two non-smooth (with known prox), one smooth (with non-trivial but manageable gradient).

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a\in\mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right] ,$$

with $\mathbf{r} = (x, y)$.

Proximal Optimization

- How do we solve this optimization problem? Can it be solved?
- Convex problem, but the existance and unicity of the solution are not given (function spaces). Three terms, two non-smooth (with known prox), one smooth (with non-trivial but manageable gradient).
- Do we need forward-backward primal-dual splitting?

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|^2_{\mathcal{D}} + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right] ,$$

with $\mathbf{r} = (x, y)$.

Proximal Optimization

- How do we solve this optimization problem? Can it be solved?
- Convex problem, but the existance and unicity of the solution are not given (function spaces). Three terms, two non-smooth (with known prox), one smooth (with non-trivial but manageable gradient).
- Do we need forward-backward primal-dual splitting? No. Not if we can find the prox of the sum of the two non-smooth terms. It is faster (Pustelnik and Condat, 2017).

Functional Inverse Diffusion (Optimization I)

We have $d_{\mathrm{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\mathrm{obs}}\|_{\mathcal{D}}^2 + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right] ,$$

with $\mathbf{r} = (x, y)$.

Diffusion Operator, $a \mapsto \int_0^{\sigma_{\max}} G_\sigma a \, \mathrm{d}\sigma$

i) Bound on its operator norm. Then, using Jensen's inequality and that $\|G_\sigma\|_{\mathcal{L}(\mathrm{L}^2(\mathbb{R}^2),\mathrm{L}^2(\mathbb{R}^2))} = 1$,

$$\|A\|_{\mathcal{L}(\mathcal{A},\mathcal{D})} \leq \sqrt{\sigma_{\max}} \,.$$

Functional Inverse Diffusion (Optimization I)

We have $d_{\text{obs}} \in \mathcal{D}_+$ and want to recover $a \in \mathcal{A}_+$. We propose the (non-smooth, constrained) convex problem

$$\min_{a \in \mathcal{A}} \left[\|Aa - d_{\text{obs}}\|^2_{\mathcal{D}} + \delta_{\mathcal{A}_+}(a) + \lambda \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)} \right],$$

with $\mathbf{r} = (x, y)$.

Diffusion Operator, $a \mapsto \int_0^{\sigma_{\max}} G_\sigma a \,\mathrm{d}\sigma$

i) Bound on its operator norm. Then, using Jensen's inequality and that $\|G_\sigma\|_{\mathcal{L}(\mathrm{L}^2(\mathbb{R}^2),\mathrm{L}^2(\mathbb{R}^2))} = 1$,

$$\|A\|_{\mathcal{L}(\mathcal{A},\mathcal{D})} \leq \sqrt{\sigma_{\max}}\,.$$

ii) Adjoint operator. We use that $G^*_\sigma = G_\sigma$,

$$(A^* d)(x, y, \sigma) = G_\sigma\{d(x, y)\}\,.$$

Proximal Op. for the Non-negative Group-Sparsity Regularizer (Optimization II)

$$\gamma\mathcal{R}(a) = \delta_{\mathcal{A}_+}(a) + \lambda\gamma \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)}$$

- The proximal operator of $\gamma\mathcal{R}(a)$ for $\gamma > 0$ does not follow easily

Proximal Op. for the Non-negative Group-Sparsity Regularizer (Optimization II)

$$\gamma\mathcal{R}(a) = \delta_{\mathcal{A}_+}(a) + \lambda\gamma \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)}$$

- The proximal operator of $\gamma\mathcal{R}(a)$ for $\gamma > 0$ does not follow easily
- The separable sum property tells us that it would follow easily from $\mathrm{prox}_{\gamma\vartheta}(x)$ for $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, which is still hard because there is no good calculus for prox operators

Proximal Op. for the Non-negative Group-Sparsity Regularizer (Optimization II)

$$\gamma\mathcal{R}(a) = \delta_{\mathcal{A}_+}(a) + \lambda\gamma \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)}$$

- The proximal operator of $\gamma\mathcal{R}(a)$ for $\gamma > 0$ does not follow easily
- The separable sum property tells us that it would follow easily from $\mathrm{prox}_{\gamma\vartheta}(x)$ for $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, which is still hard because there is no good calculus for prox operators
- We show that this is one of the "good sums", i.e., $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_{\xi}(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$, where $\mathrm{P}_{\mathcal{Z}}$ are projections on $\mathcal{Z} \subset \mathcal{X}$, $\bar{\mathcal{B}}_{\xi}(\gamma) = \{x \in \mathcal{X} : \left\|\xi^{-1}x\right\|_{\mathcal{X}} \leq \gamma\}$ and $\circ$ represents composition

Proximal Op. for the Non-negative Group-Sparsity Regularizer (Optimization II)

$$\gamma\mathcal{R}(a) = \delta_{\mathcal{A}_+}(a) + \lambda\gamma \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)}$$

- The proximal operator of $\gamma\mathcal{R}(a)$ for $\gamma > 0$ does not follow easily
- The separable sum property tells us that it would follow easily from $\mathrm{prox}_{\gamma\vartheta}(x)$ for $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, which is still hard because there is no good calculus for prox operators
- We show that this is one of the "good sums", i.e., $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_\xi(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$, where $\mathrm{P}_{\mathcal{Z}}$ are projections on $\mathcal{Z} \subset \mathcal{X}$, $\bar{\mathcal{B}}_\xi(\gamma) = \{x \in \mathcal{X} : \left\|\xi^{-1}x\right\|_{\mathcal{X}} \leq \gamma\}$ and $\circ$ represents composition
- Then, in the simple case $\xi(\sigma) = 1$ if $\sigma \in \aleph \subset [0, \sigma_{\mathsf{max}}]$ and 0 otherwise, if $p = \mathrm{prox}_{\gamma\mathcal{R}}(a)$, and we decompose $a = a_\aleph + a_{\aleph^c}$,

$$p_{\mathbf{r}} = [a_{\aleph^c,\mathbf{r}}]_+ + [a_{\aleph,\mathbf{r}}]_+ \left(1 - \frac{\gamma\lambda}{\left\|[a_{\aleph,\mathbf{r}}]_+\right\|_{\mathrm{L}^2(\aleph)}}\right)_+ .$$

Proximal Op. for the Non-negative Group-Sparsity Regularizer (Optimization II)

$$\gamma \mathcal{R}(a) = \delta_{\mathcal{A}_+}(a) + \lambda\gamma \left\| \|\xi a_{\mathbf{r}}\|_{\mathrm{L}^2(\mathbb{R}_+)} \right\|_{\mathrm{L}^1(\mathbb{R}^2)}$$

- The proximal operator of $\gamma\mathcal{R}(a)$ for $\gamma > 0$ does not follow easily
- The separable sum property tells us that it would follow easily from $\mathrm{prox}_{\gamma\vartheta}(x)$ for $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, which is still hard because there is no good calculus for prox operators
- We show that this is one of the "good sums", i.e., $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_\xi(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$, where $\mathrm{P}_{\mathcal{Z}}$ are projections on $\mathcal{Z} \subset \mathcal{X}$, $\bar{\mathcal{B}}_\xi(\gamma) = \{x \in \mathcal{X} : \left\|\xi^{-1} x\right\|_{\mathcal{X}} \leq \gamma\}$ and $\circ$ represents composition
- Then, in the simple case $\xi(\sigma) = 1$ if $\sigma \in \aleph \subset [0, \sigma_{\mathsf{max}}]$ and 0 otherwise, if $p = \mathrm{prox}_{\gamma\mathcal{R}}(a)$, and we decompose $a = a_\aleph + a_{\aleph^c}$,

$$p_{\mathbf{r}} = [a_{\aleph^c,\mathbf{r}}]_+ + [a_{\aleph,\mathbf{r}}]_+ \left(1 - \frac{\gamma\lambda}{\left\|[a_{\aleph,\mathbf{r}}]_+\right\|_{\mathrm{L}^2(\aleph)}}\right)_+ .$$

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathsf{P}_{\bar{\mathcal{B}}_{\xi}(\gamma)}) \circ \mathsf{P}_{\mathcal{X}_+}$.

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_{\xi}(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$.
- Consider first the well-known case of $g(x)$, the scaled norm, and replace step by step.

$$\begin{tikzcd}
g(x) = \gamma \|x\|_{\mathcal{X}} \arrow[r, "\mathrm{prox}"'] \arrow[d, leftrightarrow, "\text{Fenchel conj.}"'] & \mathrm{prox}_g(x) = x - \mathrm{P}_{\bar{\mathcal{B}}(\gamma)}[x] \arrow[d, leftrightarrow, "\text{Moreau's Id.}"] \\
g^*(x^*) = \delta_{\bar{\mathcal{B}}^*(\gamma)}(x^*) \arrow[r, "\mathrm{prox}"'] & \mathrm{prox}_{g^*}(x^*) = \mathrm{P}_{\bar{\mathcal{B}}^*(\gamma)}[x^*]
\end{tikzcd}$$

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_\xi(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$.
- Consider first the well-known case of $g(x)$, the scaled norm, and replace step by step.

$$\gamma\vartheta(x) = \gamma\,\|\xi x\|_{\mathcal{X}} + \delta_{\mathcal{X}_+}(x)$$

$$\begin{array}{ccc}
g(x) = \gamma\,\|x\|_{\mathcal{X}} & \xrightarrow[\text{prox}]{} & \mathrm{prox}_g(x) = x - \mathrm{P}_{\bar{\mathcal{B}}(\gamma)}[x] \\
{\scriptstyle \text{Fenchel conj.}}\updownarrow & & \updownarrow{\scriptstyle \text{Moreau's Id.}} \\
g^*(x^*) = \delta_{\bar{\mathcal{B}}^*(\gamma)}(x^*) & \xrightarrow[\text{prox}]{} & \mathrm{prox}_{g^*}(x^*) = \mathrm{P}_{\bar{\mathcal{B}}^*(\gamma)}[x^*]
\end{array}$$

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_{\xi}(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$.
- Consider first the well-known case of $g(x)$, the scaled norm, and replace step by step.

$$\gamma\vartheta(x) = \gamma\,\|\xi x\|_{\mathcal{X}} + \delta_{\mathcal{X}_+}(x)$$

$$\begin{tikzcd}
g(x) = \gamma\,\|x\|_{\mathcal{X}} \arrow[r, "\mathrm{prox}"'] \arrow[d, leftrightarrow, "\text{Fenchel conj.}"'] & \mathrm{prox}_g(x) = x - \mathrm{P}_{\bar{\mathcal{B}}(\gamma)}[x] \arrow[d, leftrightarrow, "\text{Moreau's Id.}"] \\
g^*(x^*) = \delta_{\bar{\mathcal{B}}^*(\gamma)}(x^*) \arrow[r, "\mathrm{prox}"'] & \mathrm{prox}_{g^*}(x^*) = \mathrm{P}_{\bar{\mathcal{B}}^*(\gamma)}[x^*]
\end{tikzcd}$$

$$(\gamma\vartheta)^*(x^*) = \delta_{\bar{\mathcal{B}}^*_{\xi}(\gamma)}(x^*_{\mathrm{p}})$$

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_{\xi}(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$.
- Consider first the well-known case of $g(x)$, the scaled norm, and replace step by step.

$$\gamma\vartheta(x) = \gamma\,\|\xi x\|_{\mathcal{X}} + \delta_{\mathcal{X}_+}(x)$$

$$\begin{tikzcd}
g(x) = \gamma\,\|x\|_{\mathcal{X}} \arrow[r, "\mathrm{prox}"'] \arrow[d, leftrightarrow, "\text{Fenchel conj.}"'] & \mathrm{prox}_g(x) = x - \mathrm{P}_{\bar{\mathcal{B}}(\gamma)}[x] \arrow[d, leftrightarrow, "\text{Moreau's Id.}"] \\
g^*(x^*) = \delta_{\bar{\mathcal{B}}^*(\gamma)}(x^*) \arrow[r, "\mathrm{prox}"'] & \mathrm{prox}_{g^*}(x^*) = \mathrm{P}_{\bar{\mathcal{B}}^*(\gamma)}[x^*]
\end{tikzcd}$$

$$(\gamma\vartheta)^*(x^*) = \delta_{\bar{\mathcal{B}}^*_{\xi}(\gamma)}(x^*_{\mathrm{p}}) \qquad \mathrm{prox}_{(\gamma\vartheta)^*}(x^*) = x^*_{\mathrm{n}} + \mathrm{P}_{\bar{\mathcal{B}}^*_{\xi}(\gamma)}\left[x^*_{\mathrm{p}}\right]$$

Proximal Operator for the Non-negative Weigthed Norm (Optimization III)

- We show that if $\vartheta(x) = \delta_{\mathcal{X}_+}(x) + \|\xi x\|_{\mathcal{X}}$ for $\mathcal{X} = \mathrm{L}^2(\aleph)$, $x \in \mathcal{X}$, and $\aleph = \{\sigma : \xi(\sigma) > 0\}$, $\mathrm{prox}_{\gamma\vartheta} = (\mathrm{Id} - \mathrm{P}_{\bar{\mathcal{B}}_\xi(\gamma)}) \circ \mathrm{P}_{\mathcal{X}_+}$.
- Consider first the well-known case of $g(x)$, the scaled norm, and replace step by step.

$$\gamma\vartheta(x) = \gamma\|\xi x\|_{\mathcal{X}} + \delta_{\mathcal{X}_+}(x) \qquad \mathrm{prox}_{\gamma\vartheta}(x) = x_+ - \mathrm{P}_{\bar{\mathcal{B}}_\xi(\gamma)}[x_+]$$

$$\begin{array}{ccc}
g(x) = \gamma\|x\|_{\mathcal{X}} & \xrightarrow[\text{prox}]{\hspace{6em}} & \mathrm{prox}_g(x) = x - \mathrm{P}_{\bar{\mathcal{B}}(\gamma)}[x] \\
\Big\updownarrow{\scriptstyle \text{Fenchel conj.}} & & \Big\updownarrow{\scriptstyle \text{Moreau's Id.}} \\
g^*(x^*) = \delta_{\bar{\mathcal{B}}^*(\gamma)}(x^*) & \xrightarrow[\text{prox}]{\hspace{6em}} & \mathrm{prox}_{g^*}(x^*) = \mathrm{P}_{\bar{\mathcal{B}}^*(\gamma)}[x^*]
\end{array}$$

$$(\gamma\vartheta)^*(x^*) = \delta_{\bar{\mathcal{B}}^*_\xi(\gamma)}(x^*_{\mathrm{p}}) \qquad \mathrm{prox}_{(\gamma\vartheta)^*}(x^*) = x^*_{\mathrm{n}} + \mathrm{P}_{\bar{\mathcal{B}}^*_\xi(\gamma)}\left[x^*_{\mathrm{p}}\right]$$

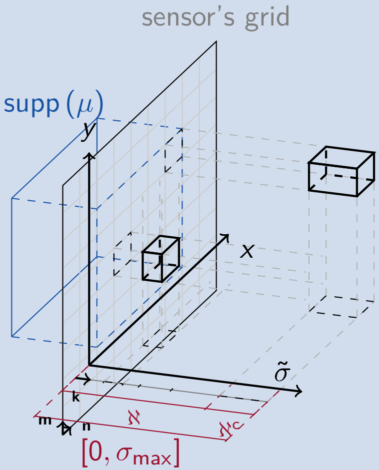
Functional Inverse Diffusion - APG algorithm (Optimization IV)

Require: Initial $a^{(0)} \in \mathcal{A}_+$, image observation $d_{\mathrm{obs}} \in \mathcal{D}_+$
Ensure: A solution $a_{\mathrm{opt}} \in \mathcal{A}_+$

$b^{(0)} \leftarrow a^{(0)}$, $i \leftarrow 0$
repeat
  $i \leftarrow i + 1$, $\alpha \leftarrow \frac{t(i-1)-1}{t(i)}$
  $a^{(i)} \leftarrow b^{(i-1)} - \sigma_{\mathrm{max}}^{-1} A^* \left(A b^{(i-1)} - d_{\mathrm{obs}} \right)$
  for all $\mathbf{r} \in \mathbb{R}^2$ **do**
    $a_{\mathbf{r}}^{(i)} \leftarrow \left[a_{\mathbf{r}}^{(i)} \right]_+ \left(1 - \frac{(2\sigma_{\mathrm{max}})^{-1} \lambda}{\left\| \left[a_{\mathbf{r}}^{(i)} \right]_+ \right\|_{\mathrm{L}^2([0,\sigma_{\mathrm{max}}])}} \right)_+$
  end for
  $b^{(i)} \leftarrow a^{(i)} + \alpha \left(a^{(i)} - a^{(i-1)} \right)$
until convergence
$a_{\mathrm{opt}} \leftarrow a^{(i)}$

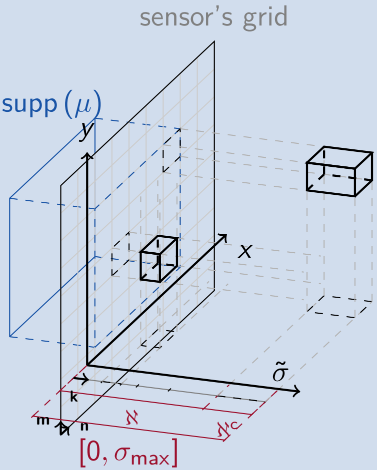
Sequences of $t(i)$ can be chosen as (Bech and Teboulle, 2009) or as (Chambolle and Dossal, 2015).

Discretization



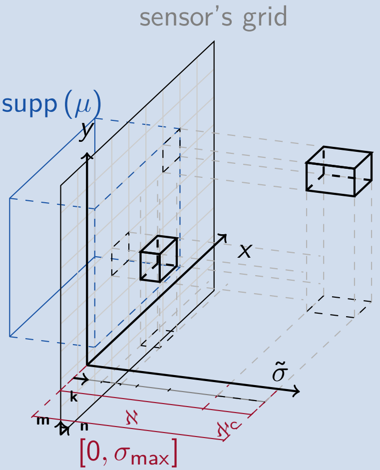
- Spatial grid given by camera sensor

Discretization



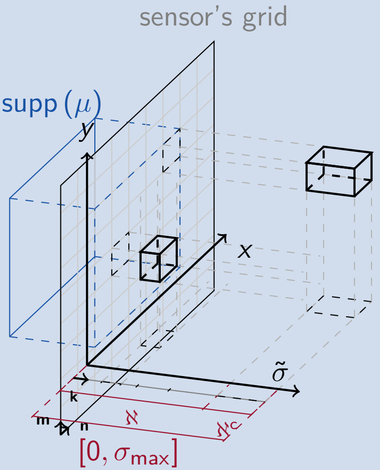
- Spatial grid given by camera sensor
- σ-grid with different levels of detail

Discretization



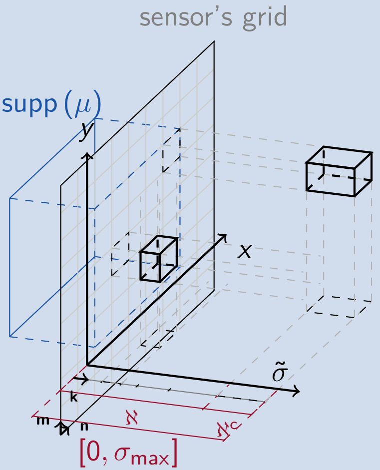
- Spatial grid given by camera sensor
- σ-grid with different levels of detail
- Inner approximation paradigm (step-constant functions)

Discretization



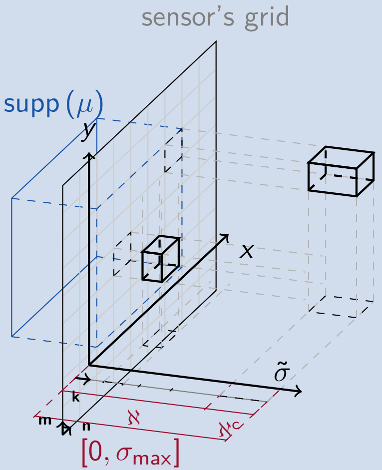
- Spatial grid given by camera sensor
- σ-grid with different levels of detail
- Inner approximation paradigm (step-constant functions)
- Choice of normalization in restriction and extension operators

Discretization



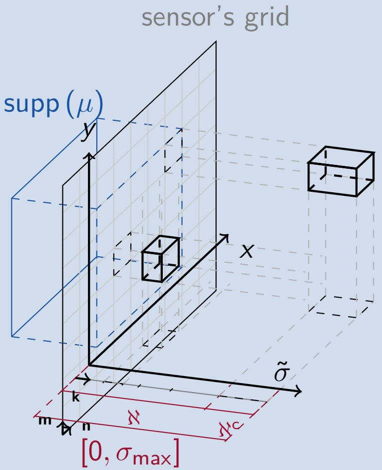
- Spatial grid given by camera sensor
- σ-grid with different levels of detail
- Inner approximation paradigm (step-constant functions)
- Choice of normalization in restriction and extension operators
- Resulting algorithm can be reasoned as discrete APG

Discretization



- Spatial grid given by camera sensor
- σ-grid with different levels of detail
- Inner approximation paradigm (step-constant functions)
- Choice of normalization in restriction and extension operators
- Resulting algorithm can be reasoned as discrete APG
- The typical size of the variable $a[m, n, k]$ to recover will be $2048^2 \times 6 = 25 \cdot 10^6$

Discretization



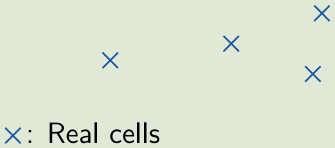
- Spatial grid given by camera sensor
- σ-grid with different levels of detail
- Inner approximation paradigm (step-constant functions)
- Choice of normalization in restriction and extension operators
- Resulting algorithm can be reasoned as discrete APG
- The typical size of the variable $a[m, n, k]$ to recover will be $2048^2 \times 6 = 25 \cdot 10^6$
- Different kernel approximations are considered

Evaluation on Synthetic Data

Besides thorough human testing on real data, we can evaluate our approach on synthetic data. To evaluate the location accuracy, we run 10000 iterations of the algorithm, find spatial maxima and threshold them optimally, and, defining a tolerance of $\Delta = 3\ \mathrm{pix}$ we compute the detection metrics

$$\mathrm{pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \ \mathrm{rec} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \text{ and } \mathrm{F1} = \frac{2\,\mathrm{pre}\cdot\mathrm{rec}}{\mathrm{pre}+\mathrm{rec}}.$$

Example

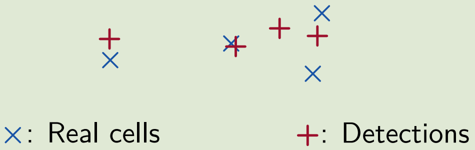


Evaluation on Synthetic Data

Besides thorough human testing on real data, we can evaluate our approach on synthetic data. To evaluate the location accuracy, we run 10000 iterations of the algorithm, find spatial maxima and threshold them optimally, and, defining a tolerance of $\Delta = 3\ \mathrm{pix}$ we compute the detection metrics

$$\mathrm{pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},\ \mathrm{rec} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},\ \text{and}\ \mathrm{F1} = \frac{2\,\mathrm{pre}\cdot\mathrm{rec}}{\mathrm{pre}+\mathrm{rec}}\,.$$

Example

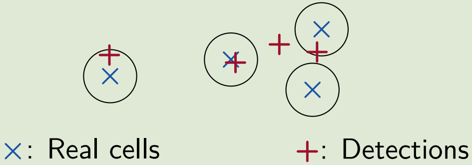


Evaluation on Synthetic Data

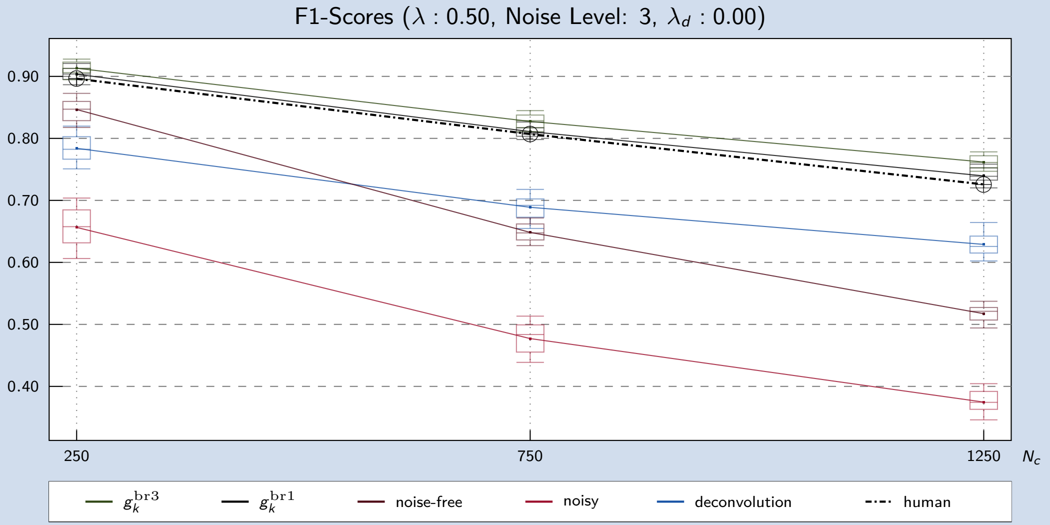
Besides thorough human testing on real data, we can evaluate our approach on synthetic data. To evaluate the location accuracy, we run 10000 iterations of the algorithm, find spatial maxima and threshold them optimally, and, defining a tolerance of $\Delta = 3\ \mathrm{pix}$ we compute the detection metrics

$$\mathrm{pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},\ \mathrm{rec} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},\ \text{and } \mathrm{F1} = \frac{2\,\mathrm{pre}\cdot\mathrm{rec}}{\mathrm{pre}+\mathrm{rec}}.$$

Example

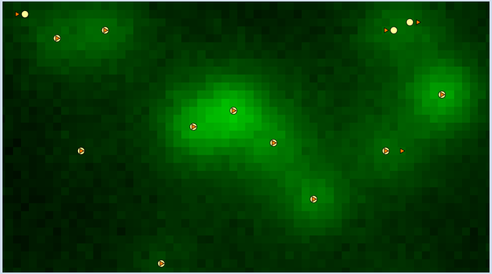


Results on Synthetic Data (I)

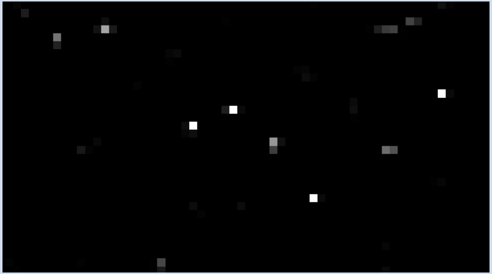


512×512 noisy images with noise equivalent to 6-bit quantization.

Results on Synthetic Data (II)

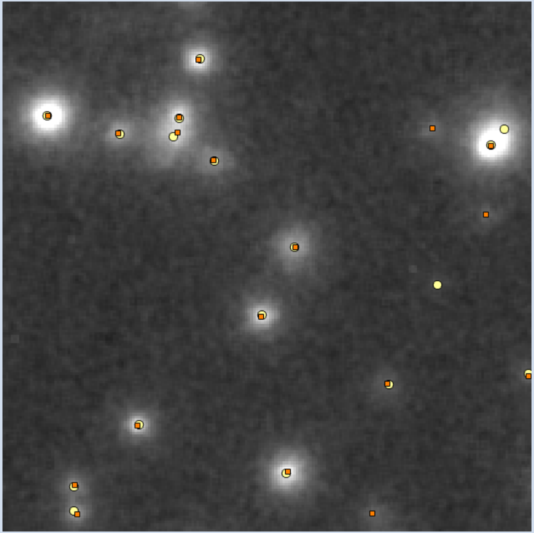


True positions (orange triangles) and detections (yellow circles).



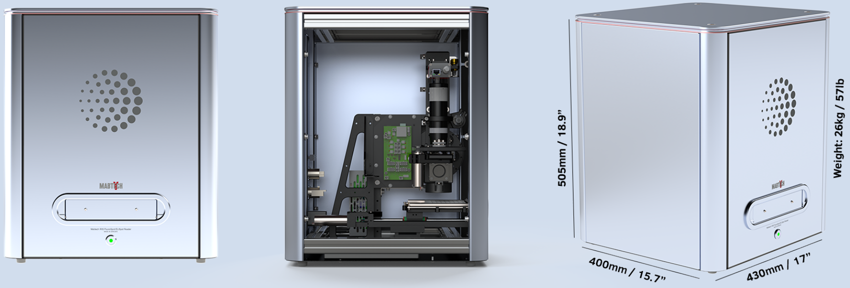
Pixels' contr. to the regularizer, i.e., $\sqrt{\int a^2(x, y, \sigma)\mathrm{d}\sigma}$.

Results on Real Data

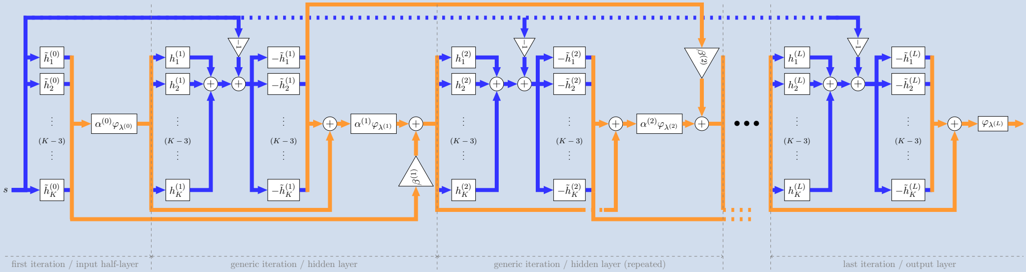


Detection results (yellow circles) and human labeling (orange squares). F1-Score relative to human, 0.9 (whole image).

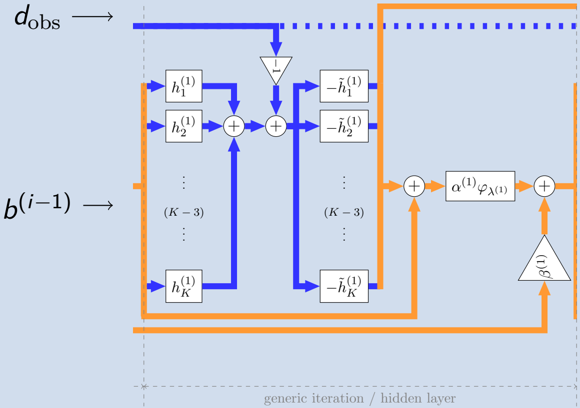
MabTech IRIS^{TM}



SpotNet - Learned iterations for faster inverse problems

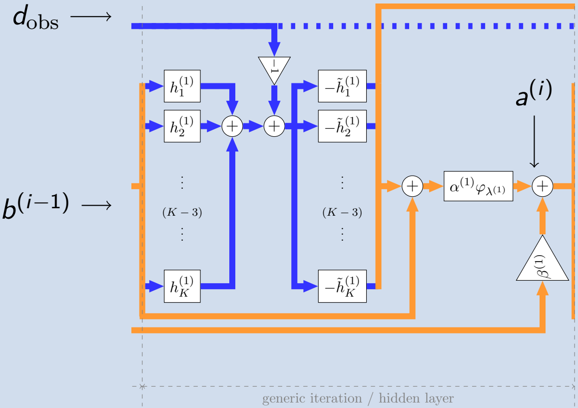


SpotNet - Learned iterations for faster inverse problems



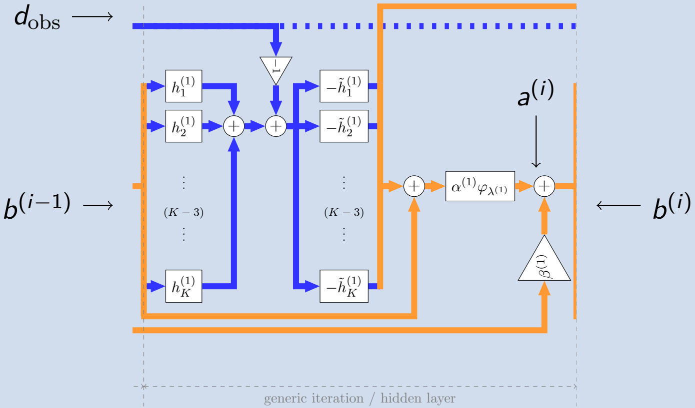
- $a^{(i)} \leftarrow \left[b^{(i-1)} - \sigma_{\text{max}}^{-1} A^* \left(A b^{(i-1)} - d_{\text{obs}} \right) \right]$

SpotNet - Learned iterations for faster inverse problems



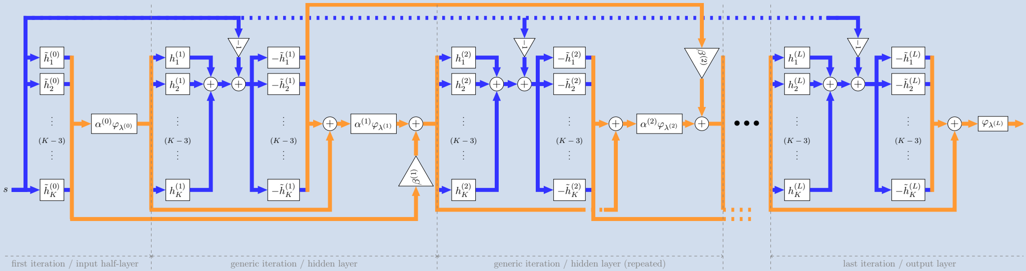
- $a^{(i)} \leftarrow \left[b^{(i-1)} - \sigma_{\max}^{-1} A^* \left(A b^{(i-1)} - d_{\text{obs}} \right) \right]$
- $a^{(i)} \leftarrow \varphi_\lambda \left(a^{(i)} \right)$

SpotNet - Learned iterations for faster inverse problems



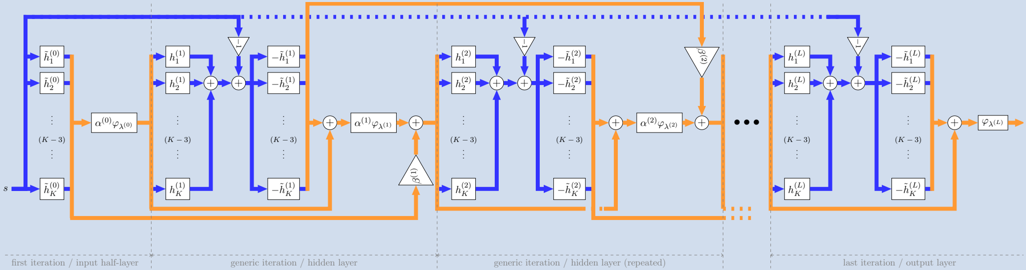
- $a^{(i)} \leftarrow \left[b^{(i-1)} - \sigma_{\text{max}}^{-1} A^* \left(Ab^{(i-1)} - d_{\text{obs}}\right)\right]$
- $a^{(i)} \leftarrow \varphi_\lambda \left(a^{(i)}\right)$
- $b^{(i)} \leftarrow a^{(i)} + \alpha \left(a^{(i)} - a^{(i-1)}\right)$

SpotNet - Learned iterations for faster inverse problems



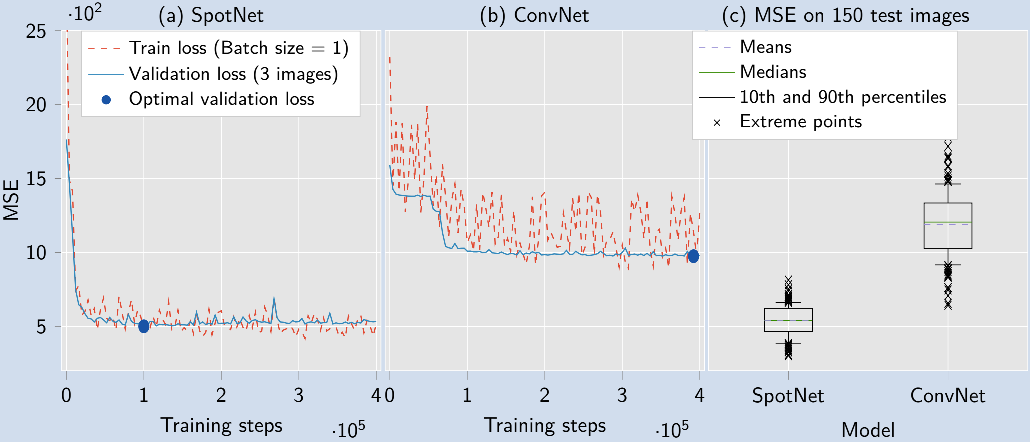
- $a^{(i)} \leftarrow \left[b^{(i-1)} - \sigma_{\max}^{-1} A^* \left(A b^{(i-1)} - d_{\mathrm{obs}} \right) \right]$
- $a^{(i)} \leftarrow \varphi_\lambda \left(a^{(i)} \right)$
- $b^{(i)} \leftarrow a^{(i)} + \alpha \left(a^{(i)} - a^{(i-1)} \right)$
- Based on the learned gradient descent of (Gregor and LeCun, 2010), recently explored by (Giryes, Eldar et al., 2018).

SpotNet - Learned iterations for faster inverse problems



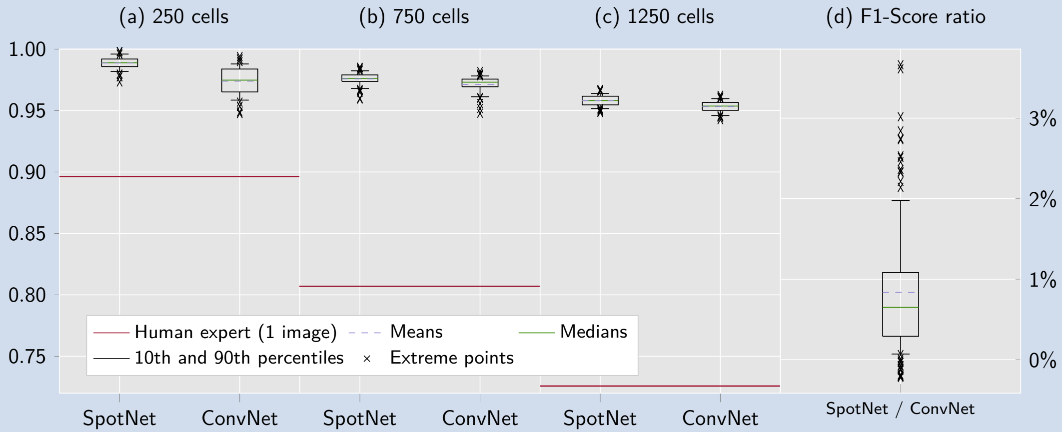
- Based on the learned gradient descent of (Gregor and LeCun, 2010), recently explored by (Giryes, Eldar et al., 2018).
- See all details at `https://github.com/poldap/SpotNet`.

Results for SpotNet with $L = 3$ and smaller kernels



- Evaluation of SpotNet and a generic ConvNet on $\mathrm{MSE}\{\hat{a}\}$.
- Training on 7 synthetic images with 1250 cells, validation on 3. Testing on 150 images containing 250, 750 or 1250 cells.

Results for SpotNet with $L = 3$ and smaller kernels



- Evaluation of SpotNet and a generic ConvNet on $\mathrm{F1}$ score as above.
- Trained on 7 images with 1250 cells.

Thank you

Please, feel free to ask questions.

January 28, 2019 at PARIETAL