



# **Inverse problems in signal processing**

Functional optimization, parameter estimation and machine learning

POL DEL AGUILA PLA

Doctoral Thesis in Electrical Engineering  
Stockholm, Sweden, 2019

Division of Information Science and Engineering  
School of Electrical Engineering and Computer Science  
TRITA-EECS-AVL-2019:51 KTH Royal Institute of Technology  
ISBN 978-91-7873-213-5 Malvinas väg 10, 114 28 Stockholm, Sweden

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i Elektroteknik måndagen den 16 september 2019 kl. 09:00 i hörsal F3, Lindstedtsvägen 26, Stockholm.

Academic thesis which, with permission of the KTH Royal Institute of Technology, is submitted for public scrutiny for the completion of the Ph.D. in Electrical Engineering on Monday September 16, 2019 at 09:00 am in the lecture hall F3, Lindstedtsvägen 26, Stockholm.

© Pol del Aguila Pla, unless otherwise stated. September 16, 2019

Tryck: Universitetsservice US-AB

*To the early believers,  
Pels meus avis, que no ho han pogut veure,*



From left to right, Jordi Pla Sangenís, Eulàlia Bonaparte Torrents,  
Pol del Aguila Pla, Agustina Miquel Fité, and Manel del Aguila Rodríguez.  
Picture taken in 1991.



## Abstract

Inverse problems arise in any scientific endeavor. Indeed, it is seldom the case that our senses or basic instruments, i.e., the data, provide the answer we seek. It is only by using our understanding of how the world has generated the data, i.e., a model, that we can hope to infer what the data imply. Solving an inverse problem is, simply put, using a model to retrieve the information we seek from the data.

In signal processing, systems are engineered to generate, process, or transmit signals, i.e., indexed data, in order to achieve some goal. The goal of a specific system could be to use an observed signal and its model to solve an inverse problem. However, the goal could also be to generate a signal so that it reveals a parameter to investigation by inverse problems. Inverse problems and signal processing overlap substantially, and rely on the same set of concepts and tools. This thesis lies at the intersection between them, and presents results in modeling, optimization, statistics, machine learning, biomedical imaging and automatic control.

The novel scientific content of this thesis is contained in its seven composing publications, which are reproduced in Part II. In five of these, which are mostly motivated by a biomedical imaging application, a set of related optimization and machine learning approaches to source localization under diffusion and convolutional coding models are presented. These are included in Publications A, B, E, F and G, which also include contributions to the modeling and simulation of a specific family of image-based immunoassays. Publication C presents the analysis of a system for clock synchronization between two nodes connected by a channel, which is a problem of utmost relevance in automatic control. The system exploits a specific node design to generate a signal that enables the estimation of the synchronization parameters. In the analysis, substantial contributions to the identifiability of sawtooth signal models under different conditions are made. Finally, Publication D brings to light and proves results that have been largely overlooked by the signal processing community and characterize the information that quantized linear models contain about their location and scale parameters.



## Sammanfattning

Inversa problem uppstår vid alla vetenskapliga undersökningar. Våra sinnen och mätinstrument -rådata -ger faktiskt sällan svaren vi letar efter. Vi behöver då utveckla vår förståelse av hur data genererats, d.v.s., använda en modell, för att kunna dra korrekta slutsatser. Att lösa inversa problem är, enkelt uttryckt, att använda modeller för att få fram den information man vill ha från tillgängliga data.

Signalbehandling handlar om utveckling av system som skapar, behandlar eller överför signaler (d.v.s., indexerade data) för att nå ett visst mål. Ett exempel på mål för en sådant system är att lösa ett inverst problem utifrån den analyserade signalen med hjälp av en modell. Signalbehandling kan dock även handla om att skapa en signal, så att denna avslöjar en parameter för utredning genom ett inverst problem. Inversa problem och signalbehandling är två fält som överlappar i stor utsträckning, och som använder sig av samma koncept och verktyg. Denna avhandling utforskar gränslandet mellan dessa två fält, och presenterar resultat inom modellering, optimering, statistik, maskininlärning, biomedicinsk utbildning och automatisk kontroll.

Det nya vetenskapliga innehållet i den här avhandlingen är baserat på de sju artiklar som återges här i Del II. I fem av dessa artiklar beskrivs ett antal relaterade metoder för optimering och maskininlärning för källkalibrering med hjälp av diffusions- och konvolutionsmodellering, med tillämpningar framför allt inom biomedicinsk bildbehandling. Dessa inkluderas i Publikationer A, B, E, F och G, och behandlar också modellering och simulering av en familj av bildbaserade immunkemiska detektionsmetoder. Publikation C presenterar analys av ett system för klocksynchronisering mellan två noder förbundna med en kanal, vilket är ett problem med särskild relevans för automatisk kontroll. Systemet använder en specifik noddesign för att generera en signal som möjliggör skattning av synkroniseringsparametrarna. Analysen bidrar avsevärt till metodiken för att identifiera sågtandsmönstrande signalmodeller under olika förhållanden. Avslutningsvis presenteras i Publikation D resultat som tidigare i stora drag förbisetts inom signalbehandlingsfältet. Här karakteriseras även den information som kvantiserade linjära modeller innehåller om deras läges- och skalparametrar.

## Acknowledgements

My first words of thanks are for **Professor Joakim Jaldén**, my Ph.D. supervisor, mentor, colleague, and friend for these last five years. His technical acumen and intuition are only bound by interest and time. His integrity and dedication to research and education are inspirational. However, these are only part of the qualities that made him an ideal supervisor for me. His patience, empathy and involvement in my education and development went well beyond the terms of his employment, and for this, above all, I am grateful.

Breaking all protocols, my next words are for my wife **Doctor Celia García Pareja**: I am immensely proud to have had you as shadow supervisor and student, life partner, codebtor, sambo, wife, coparent-to-be, and friend for these past 5 years, and I am looking forward to a life of expanding our relationship to new categories. Thank you for your love and support, and thank you for reading through almost everything I have ever written professionally, including this thesis.

To my cosupervisor, **Professor Magnus Jansson**: thank you for your always encouraging and flattering words, your presence, your trust, and your patience. Having you close by and available to my random questions was a source of comfort I am not likely to forget.

To my mentor of years ago, **Professor Ferran Marqués**, who first taught me signal processing, eased me into research, and guided me in my early career decisions, my most heartfelt thank you and my sincere apology for not having kept in contact. I shall endeavor to improve.

To my opponent, **Professor Yonina Eldar**, the members of my examination committee, **Doctor Silvia Gazzola**, **Professor Andreas Jakobsson**, and **Assistant Professor Alexander Jung**, and my thesis' preliminary reviewer and substitute committee member, **Associate Professor Johan Karlsson**: thank you for agreeing to serve in my defense, reading through my thesis, and asking any questions that come to mind. I shall be happy to answer them.

To my coauthors, **Vidit Saxena**, **Lissy Pellaco**, **Doctor Satyam Divedi**, **Professor Peter Händel**, and **Professor Joakim Jaldén**, I give my thanks for all they have taught me, their patience, and above all, their tolerance towards my nitpicking tendencies. Having someone to work with has been one of the great remedies to the all-to-often isolating experience of getting this Ph.D. thesis ready, and nothing tickles the mind quite as much as lively academic disputes. I am looking forward to continue to collaborate with all of you.

I shall not forget also those coauthors of works that have not seen the light of day yet or may never see it, as their contribution to my knowledge of the topics covered here is still much valued, such as **Doctor Arun Venkataraman** and **Christophe Kervazo**.

To my academic siblings and friends **Vidit Saxena**, **Lissy Pellaco**, and **Xuechun Xu**, thank you for acting like a proper family at work. Our time together has been too short, and I hope we will get more opportunities to enjoy each other's company. Please, know and remember that I shall be available whenever you need me.



To my honorary academic sibling, **Doctor Michele Santacatterina**: I thank you for the struggles we shared, fought and won, for the learning I have had the pleasure to witness, and for your patience (and lack thereof) to solve statistical decision theory problems well into the morning hours.

Through my doctoral education at KTH I have taken courses for 80.5 ECTS, and so I have many great teachers to thank for the years of learning. Above all others, however, **Professor Krister Svanberg** from the Department of Mathematics at KTH deserves a mention. The absolute dedication he offered to his students in the course “SF3810 Convexity and optimization in linear spaces” is unparalleled, and I can only hope to one day know the details of a subject I teach as well as he did his.

During this time, I also had the opportunity to work with people from Mabtech AB. With them, I have had the luck to observe the design and launch of a product that brought some of my work to its actual intended use by biomedical scientists. For countless meetings, open minds, and a fantastic result, I want to thank **Doctor Christian Smedman**, **Associate Professor Saffan Paulie**, **Doctor Kajsa Prokopec** and **Tomas Dillenbeck**. Within this project, I had the pleasure to shortly collaborate and meet with **Doctor Haopeng Li**, now at Qamcom Research & Technology AB, which was one of the best teaching assistants I had the pleasure to be guided by during my time as an undergraduate student at KTH.

I would also like to thank **Professor Jean-Luc Starck** and **Dr. Samuel Farrens** for hosting me during January 2019 at Cosmostat in CEA Paris-Saclay. From the very warm and hospitable group of people I met there, I would like to thank **Vanshika Kansal** and **Virginia Ajani** for their friendliness and our unique trade of chocolates for coffee.

Similarly, I would like to thank **Professor Stephen Boyd** and his group for hosting me for a week in Stanford and sharing some of their research. Particular thanks go to **Doctor Enzo Bussetti** and **Qungyun Sun** for the discussions on academia and life.

To the heads of the departments I have belonged to, **Professor Peter Händel** for the now defunct Department of Signal Processing, and **Professor Mikael Skoglund** for the Department (now Division) of Information Science and Engineering: I thank you for exemplifying the lack of uniqueness of the solution to a complicated variational problem.

More importantly, to the wonderful administrators that kept those departments running from the inside, **Tove Schwartz** and **Raine Tiivel**, my thanks for their patience towards my nagging questions and requests. **Tove**, I have dearly missed our chats on assorted topics. **Raine**, I shall dearly miss your efficiency, regardless where my next steps take me.

I remember fondly the excitement of starting my doctoral education and sharing courses with **Doctor Ahti Aniomäe**, **Doctor Ehsan Olfat**, **Doctor Arun Venkitaraman**, and **Doctor Marie Maros**. Among them, sincere and special thanks go to **Doctor Marie Maros**, without whom it would have been incredibly harder to start my international academic adventure: Gràcies per ser-hi durant un temps, espero que a la llarga en recordem només les parts alegres.

I also want to mention the senior doctoral students I looked up to from the day of my arrival and until they moved on with their own adventures, such as **Doctor Efthymios Tsakonas**, **Doctor Rasmus Brandt**, **Doctor Du Liu** (yet another wonderful teaching assistant), **Doctor Vijaya Parampalli Yajnanarayana**, and **Assistant Professor Hadi Ghauch**.

Luckily, I occasionally used my time in the office smartly, and invested it in talking to wonderful people such as **Sahar Imtiaz**, **Dong Liu**, **Boules Atef Mouris**, or **Baptiste Cavarec**.

Among my administrative duties, I cherish having taken care of the *candy corner* for some time, and I thank **Doctor Zuxing Li** for his wonderful stories of his world back home, which made the drive to the supermarket that much more interesting.

My Ph.D. project would not have even started without the support of the institutions that funded it. Therefore, I would like to thank **Vetenskapsrådet** (the Swedish Research Council) for trusting Professor Joakim Jaldén with grant 2015-04026, as well as **Mabtech AB** for trusting us all to make the best algorithm we possibly could. Similarly, my career would not have developed as it has without the economic support to my trips by the **KTH Opportunities Fund**, the **Knut and Alice Wallenberg Jubilee appropriation**, the **Malme's foundation**, the **Åforsk Foundation**, **Kungliga Vetenskapsakademien** (the Royal Swedish Academy of Sciences, call ES2017-0011), and **Gälostiftelsen**.

From my family in Sweden, with which we have shared birthdays, fears, November tears, local celebrations, and a lot of good food, I want to thank **Doctor Adeline Rachalski**, **Assistant Professor Nasren Jaff**, **Doctor Paola Martínez Murillo**, **Associate Professor Paolo Frumento**, and **Assistant Professor Kristoffer Sahlholm** (who also joyfully accepted the task of translating the abstract of this thesis at hours' notice). **Adeline**, thank you for always being just a call away: your support is more appreciated than what any hug can show. Both in Stockholm and in Paris, **Doctor Ana Luísa Pinho** was the comfort of an old friend who was actually much newer.

My life in Stockholm these last three years would not have been the same without the weekly escape of **Sandra Schöning's** tap dance classes and our yearly performances. Thank you **Sandra**, I would have never guessed I could do the things that now I do every Tuesday.

I want to mention also my longest standing friends, **Sarah Moyà** and **Ahmed Amine Ramdani**. In their individual, unique ways, they are responsible for most of my extremely positive views of this world. Gràcies, **Sarah**, per la teva constància i persistència, after all this time.

To my in-laws: **Sapo Pareja Matute** i **Joan-Francesc Vidal Morist**, gràcies per acceptar-me i adoptar-me, pel vostre respecte i sinceritat, per la Celia, i per les ganes de parlar sobre tot, sempre. Per l'**Iris Vidal Pareja**, thank you for falling for me as fast and hard as I did for you, and for then seeing me as your awesome brother before your cautious self could even stand up again. A l'**Albert Viñas Villaró**, gràcies per la teva proximitat sense preàmbuls, per les hores que no dorms per parlar amb nosaltres, i per la germanor càlida que n'ha resultat.

To my family: **Georgina Pla Bonaparte**, **Aina del Aguila Pla**, i **Agustí del Aguila Miquel**, gràcies per mai deixar d'intentar entendre el que em passa pel cap, i perdó per sempre ser aquells a qui jutjo amb més duresa. No em puc ni imaginar com ha sigut per vosaltres veure'm créixer, però vull que sabeu que aquest llibre és el fruit, també, del vostre esforç i amor. Us estimo.

I gràcies a tu, **Celia**, altre cop: sempre una mica més a tu, sense excepció, voluntàriament, sempre tot per tu. Gràcies per ser-hi, gràcies pels mesos, durs, que han passat i passaran, i gràcies per la nostra vida, casa, i família.

Pol del Aguila Pla, Stockholm, September 2019



# Contents

<b>I</b>	<b>Summary</b>	<b>1</b>
<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Models . . . . .	4
1.2	Forward problems . . . . .	5
1.3	Inverse problems . . . . .	8
	Well-posedness and identifiability . . . . .	8
	Stochastic methods and logconcavity . . . . .	9
	Regularization methods . . . . .	11
	Iterative solvers for nonsmooth optimization problems . . . . .	13
<b>2</b>	<b>Overview of contributions</b>	<b>19</b>
	<b>Bibliography</b>	<b>27</b>
<b>II</b>	<b>Publications</b>	<b>31</b>
<b>A</b>	<b>Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse Problems</b>	<b>33</b>
	<b>Journal paper.</b> P. del Aguila Pla and J. Jaldén, <i>IEEE Transactions on Signal Processing</i> , vol. 66, no. 20, pp. 5407–5421, Oct. 2018.	
<b>B</b>	<b>Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation</b>	<b>51</b>
	<b>Journal paper.</b> P. del Aguila Pla and J. Jaldén, <i>IEEE Transactions on Signal Processing</i> , vol. 66, no. 20, pp. 5422–5437, Oct. 2018.	
<b>*</b>	<b>Cell detection by functional inverse diffusion and non-negative group sparsity — Supplementary material</b>	<b>69</b>
	<b>Supplementary material</b> to Publications A and B. P. del Aguila Pla and J. Jaldén, <i>IEEE Transactions on Signal Processing</i> , Oct. 2018.	

<b>C</b>	<b>Clock synchronization over networks — Identifiability of the sawtooth model</b>	<b>83</b>
	<b>Submitted manuscript.</b> P. del Aguila Pla, L. Pellaco, S. Dwivedi, P. Händel and J. Jaldén, submitted to <i>IEEE Transactions on Control Systems Technology</i> , 2019.	
<b>*</b>	<b>Clock synchronization over networks — Identifiability of the sawtooth model, Supplementary material</b>	<b>101</b>
	<b>Supplementary material</b> to Publication C.	
<b>D</b>	<b>Inferences from quantized data — Likelihood logconcavity</b>	<b>109</b>
	<b>Manuscript.</b> P. del Aguila Pla and J. Jaldén, <i>Work in progress</i> , 2019.	
<b>E</b>	<b>Cell detection on image-based immunoassays</b>	<b>115</b>
	<b>Conference paper.</b> P. del Aguila Pla and J. Jaldén, in <i>2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)</i> , Apr. 2018, pp. 431–435.	
<b>F</b>	<b>Convolutional group-sparse coding and source localization</b>	<b>123</b>
	<b>Conference paper.</b> P. del Aguila Pla and J. Jaldén, in <i>2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> , Apr. 2018, pp. 2776–2780.	
<b>G</b>	<b>SpotNet — Learned iterations for cell detection in image-based immunoassays</b>	<b>131</b>
	<b>Conference paper.</b> P. del Aguila Pla, V. Saxena and J. Jaldén, in <i>2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)</i> , Apr. 2019, pp. 1023–1027.	

# Part I

## Summary

*Cause I'm the greatest star, I am by far! But no one knows it.*

Lyrics by *Bob Merrill*, "Funny Girl", 1964.

Favourite version performed by *Barbra Streisand* in the 1968 movie of the same name.



# Chapter 1

## Introduction

*Many theses start with a few words on the hard work of research. This is not one of those theses, since I can think of few things equally rewarding.*

Modified, from *J. Jaldén's*, “Detection for multiple input multiple output channels: Analysis of sphere decoding and semidefinite relaxation”, Doctoral thesis at KTH Royal Institute of Technology, 2006.

While reading this thesis, one may have the impression that the chapters of Part II are disconnected investigations into different aspects of signal processing and inverse problems through different applications. This is true, because the directions of the research projects were mostly guided by circumstance, impact, funding, and the coauthors’ interests. Nonetheless, this thesis, as a collection of works, consistently revolves around the ideas of signals, modeling, and inverse problems, while all the considered applications address signal processing problems.

In particular, three different projects are considered in the thesis. The most extensive involves Publications A, B, E, F, and G. We will refer to it as *Cell detection*, even if the formulation in Publication F contemplates the more general problem of source localization in imaging under convolutional coding models. Furthermore, when discussing cell detection, the functional notation of Publications A and B will be preferred over the discretized notation of Publications E, F, and G. Further information on the transformation from one to the other can be found in Publication A. The second project delves on *Clock synchronization* over networks, and it is detailed in Publication C. Finally, the third project explores *Inferences from quantized data*, and it is explored in Publication D.

Each of the publications in Part II thoroughly motivates the corresponding project, refers to the relevant state of the art, and presents the proposed solutions. Consequently, here I introduce inverse problems in signal processing in a general setting, punctuating the explanations with examples from our work. Nonetheless, I do not aim to give a complete picture of the field of inverse problems, but simply to give the basics that are more relevant to the understanding of the results in Part II,

which I summarize in Chapter 2. For excellent descriptions of the field from a number of different perspectives and conceptual frameworks, see [1, 3, 14, 19, 20, 29, 33].

## 1.1 Models

Mathematical models are at the core of modern scientific reasoning. A model represents an understanding, however accurate, on the relation between some quantities, say, a measured signal  $y \in \mathcal{Y}$  and some parameter  $x \in \mathcal{X}$ . Here,  $\mathcal{Y}$  and  $\mathcal{X}$  are two generic topological spaces with arbitrary dimension and structure, and their properties will be determined in each specific problem.

A fully specified model is a mapping  $M : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ , where  $\mathcal{P}(\mathcal{Y})$  is the space of all probability measures on  $\mathcal{Y}$ . In this manner, for any given  $x \in \mathcal{X}$ ,  $M(x)$  is a probability distribution over  $\mathcal{Y}$  such that  $y \sim M(x)$ , and, for any set  $\mathcal{S} \subset \mathcal{Y}$ , we can measure its probability according to our model as  $\Pr_x[\mathcal{S}] = M(x)[\mathcal{S}]$ . This mirrors the physical reality that even when the relevant parameter  $x$  is known, a measurement is never completely determined. Consequently, we will refer to the fully specified models  $M$  as stochastic models. In many cases in science and engineering this full characterization is not available or is mathematically intractable, and a deterministic model  $D : \mathcal{X} \rightarrow \mathcal{Y}$  such that  $y = D(x)$  is used instead. A stochastic model often arises from a deterministic model  $D$  combined with a stochastic model on possible modeling errors. In the simplest cases, the deterministic model  $D$  characterizes the location parameter of the observed signal  $y$ , while the prediction error  $D(x) - y$  is modeled by a stochastic model. For example, the common additive white Gaussian noise model, where  $x \in \mathbb{R}^M$  and  $y \in \mathbb{R}^N$ , is clearly in this category, because  $y \sim M(x) = \mathcal{N}(D(x), \sigma^2 \mathbf{I}_N)$ . In other words, we have a deterministic model for  $y$ ,  $D : \mathbb{R}^M \rightarrow \mathbb{R}^N$ , and we model the prediction errors of the model as independent random samples from a fixed normal distribution. For the rest of this introduction we shall refer to models constructed in this manner as deterministic-and-error models

A further category that is often employed in the inverse problems, signal processing and statistical communities is that of linear models. However, the term is not well defined in the generic framework presented here, and, most importantly, does not correspond to a linear mapping between  $\mathcal{X}$  and  $\mathcal{P}(\mathcal{Y})$ . In the introduction to this thesis, I refer to models as linear in two cases. First, a deterministic model is linear if it is an affine continuous mapping between  $\mathcal{X}$  and  $\mathcal{Y}$ , i.e.,  $D(x) = Ax + b$  with  $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$  and  $b \in \mathcal{Y}$ , where  $\mathcal{L}(\mathcal{X}, \mathcal{Y})$  is the set of linear and continuous operators between  $\mathcal{X}$  and  $\mathcal{Y}$ . Second, a stochastic model is linear if it can be constructed from a linear deterministic model for the location parameter of the distribution and a stochastic model for the modeling error. Table 1.1 includes a classification of the models employed in the different applications explored in Part II with respect to the categorization presented here.

The framework introduced here does not entail neither a frequentist nor a

Project	Cell detection	Clock synchronization
Parameter $x$	Density rate (PSDR) $a$	Parameter vector $\theta = [\rho, f_d, \phi_S]$
" space $\mathcal{X}$	$\mathcal{A} = \text{L}_2(\text{B} \times [0, \sigma_{\max}))$	$\mathbb{R}_+ \times \mathbb{R} \times [0, 2\pi)$
Signal $y$	Image $d_{\text{obs}}$	Measured RTTs $y$
" space $\mathcal{Y}$	$\mathcal{D} = \text{L}_2(\mathbb{R}^2)$	$\mathbb{R}^N$
Model type	Deterministic	Stochastic
Linear	✓	✗
Project	<b>Inferences from quantized data</b>	
Parameter $x$	Location and scale parameters $(\mathbf{x}, \Psi)$	
" space $\mathcal{X}$	$\mathbb{R}^m \times \mathcal{M}_n^+(\mathbb{R})$	
Signal $y$	Quantized observation $z$	
" space $\mathcal{Y}$	Countable set $\mathcal{Z}$	
Model type	Stochastic	
Linear	✗	

Table 1.1: Signal and parameter spaces, as well as model classification, in the different applications explored in Part II. Here, we use  $\text{B}$  as a generic bounded set in  $\mathbb{R}^2$  and  $\mathcal{M}_n^+(\mathbb{R})$  as the set of symmetric positive definite matrices of size  $n \times n$ .

Bayesian treatment of statistics. The frequentist view is that there is a true value  $x \in \mathcal{X}$ , while the Bayesian treatment is to propose that the parameter is sampled from a distribution over the parameter space, i.e.,  $x \sim \Pi \in \mathcal{P}(\mathcal{X})$ . Simply stated, a stochastic model  $\text{M}$  is equivalent to the likelihood function in statistics (see Section 1.3 for more details).

To conclude this section, I include a word of warning for those interested in working on mathematical modeling. Like animal models in biology, mathematical models only represent the phenomena we are interested in up to a certain extent. Although we use the precise language of mathematics in statements as  $y \sim \text{M}(x)$  or  $y = \text{D}(x)$ ,  $y$  and  $x$  do not refer to the real quantities we assign them to, but only to our understanding of them, i.e., to their modeled versions. As a consequence, the use of mathematical modeling in science and engineering should always be accompanied by honest and dedicated empirical validation. This is illustrated in Fig. 1.1, with a side-to-side comparison of animal models used to study human behavior and the mathematical model we introduced in Publication A for cell detection.

## 1.2 Forward problems

Mathematical models subject to the due empirical scrutiny and limited to their verified application range remain extremely useful. Their most direct application is forward evaluation (solving the forward problem), i.e., for a given parameter  $x \in \mathcal{X}$  or a given distribution  $\Pi \in \mathcal{P}(\mathcal{X})$ , generating a signal  $y \in \mathcal{Y}$  according

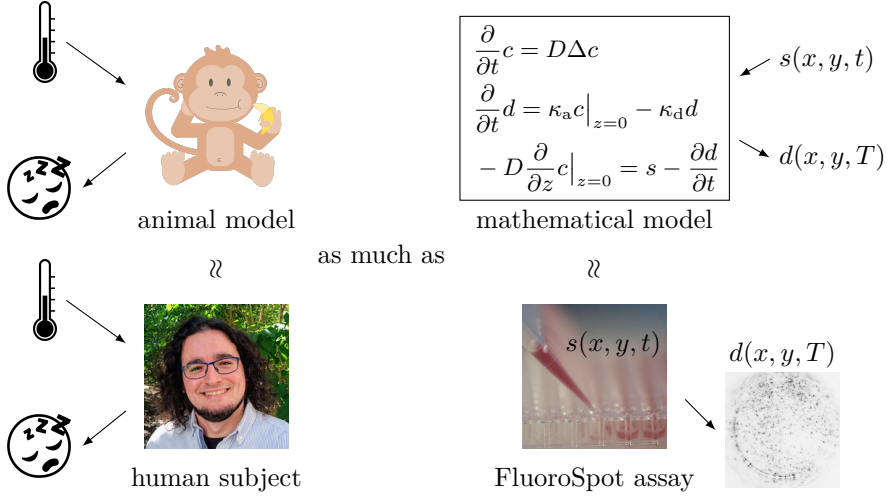


Figure 1.1: Visual reminder of the disclaimer at the end of Section 1.1. Mathematical models are only a specific type of models. All models represent reality only up to a certain extent. Real FluoroSpot data (displayed with inverted colors) provided by Mabtech AB. Assay picture by Kristoffer Hellman and Mabtech AB. Picture of the author, a human subject, by Dr. Celia García-Pareja.

to the model. In the context of stochastic models, this implies sampling from the distribution  $M(x)$ , while in the context of deterministic models, it implies evaluating the mapping  $D(x)$ .

Depending on the formulation of the model and the characteristics of  $\mathcal{X}$  and  $\mathcal{Y}$ , forward evaluation can either be trivial or extremely challenging to do exactly. In fact, even deterministic linear models, e.g.,  $D(x) = Ax + b$ , may be expressed in manners that make evaluation laborious. For example, the physical partial-differential equation (PDE) model for image-based immunoassays used in Publication A could not be evaluated easily without the parametrization of the solution we developed in Theorem 1 therein. Indeed, other known methods to evaluate the model are to either employ numerical solvers for the PDE or run non-exact particle-by-particle stochastic simulations. Both these methods are approximated and computationally expensive. While our solution still requires numerical approximations, it is much more efficient, and it enabled the numerical results in Publications B, E, and G. For details on our approach and the approximations involved, see the sketch in Fig. 1.2, and the explanations in the supplementary material to Publications A and B. What prevents exact evaluation in this case is that the model operates on (infinite-dimensional) function spaces  $\mathcal{X}$  and  $\mathcal{Y}$ . Infinite-dimensional models often result in the impossibility of obtaining closed-form expressions for evaluating the model for a generic parameter value, even when one only aims to

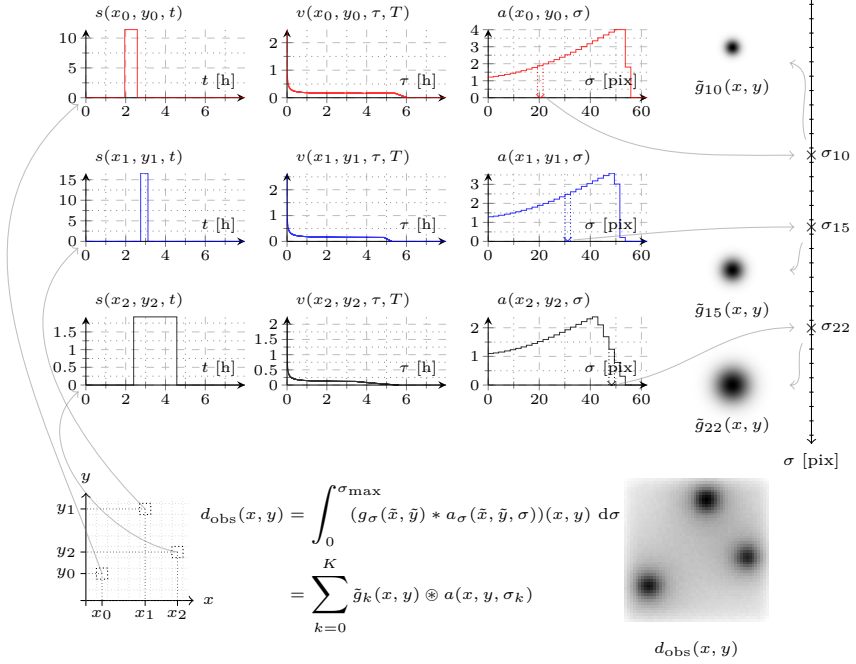


Figure 1.2: Diagram for the forward model evaluation for cell detection in a toy example. The random locations and random square-pulse secretion rate of each cell are sampled independently. Each of the source density rates  $s(x_i, y_i, t)$  are translated to their approximated (step-constant) equivalent PSDRs  $a(x_i, y_i, \sigma)$  (see Publication A and its supplementary material). Then, the model is evaluated from this approximated PSDR by using discrete convolution ( $\otimes$ ) with the integrated kernels  $\tilde{g}_k$ . The data and the kernels are displayed with inverted colors for visualization clarity. In connection with Publication G, an implementation of this forward evaluation procedure is publicly available in [9]. Note that in Publications A and B, we only studied the inversion of the last part of the model, i.e., we obtain PSDRs  $a$  from an image  $d_{\text{obs}}$ .

evaluate a specific discretization of the corresponding signal  $y \in \mathcal{Y}$ . Similar issues arise in the evaluation of infinite-dimensional stochastic models. For example, in stochastic models of continuous-time processes, e.g., diffusions, exact simulation is a novel and exciting topic of active research [15].

In the previous section, we introduced models as relationships between a parameter  $x$  and a signal  $y$ , but we never attributed any meaning to these terms, i.e., to the choice of one name or the other for a given quantity in an application. Consequently, whether a problem is a *forward problem* (obtaining  $y$  from  $x$ ) or an *inverse problem* (obtaining  $x$  from  $y$ ) is simply a matter of choice, convenience or convention. Although traditionally the forward relation was assumed to be causal, the mathematics employed to solve either of these problems do not depend on or assess the causal structure of the model. Indeed, we will see that while the causal relation between parameter and signal is quite clear in our three applications, it plays no role in how we formulate and solve the different inverse problems. Techniques for causal discovery and inference do exist, however, and are an exciting topic of active research in statistics and machine learning [24, 26, 27]. In this thesis, I assume that the forward problem can be (maybe approximately) solved with less computations than the inverse problem. This is important because, as we will see in the next section, solving the inverse problem will often involve solving a collection of forward problems under different conditions.

### 1.3 Inverse problems

*Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases,... the latter is called the inverse problem.*

From Joseph B. Keller's, "Inverse problems", *The American Mathematical Monthly*, vol. 83, no. 2, pp. 107–118, Feb. 1976.

Another application of mathematical models is inverse evaluation, i.e., using an observed signal  $y \in \mathcal{Y}$  to make statements about the parameter. In particular, one generally aims to obtain either a distribution over the parameter space, i.e.,  $\Pi_y \in \mathcal{P}(\mathcal{X})$ , or an approximation of the parameter value  $\hat{x}(y) \in \mathcal{X}$ . To obtain the former, one needs to have a fully specified stochastic model  $M(x)$ , and one employs techniques rooted in probability, Bayesian statistics, and measure theory (see [1, 29, 33]). To obtain the latter, one can use a more diverse collection of tools from either frequentist and Bayesian statistics or variational problems' theory.

### Well-posedness and identifiability

In this thesis, I only consider problems in which the aim is to recover an approximation or estimate of  $x$ ,  $\hat{x}(y)$ . In these cases, a relevant gold standard was introduced by Jacques Hadamard for deterministic models  $D$ , the well-posed problem. A well-posed problem is an inverse problem for which  $\forall y \in \mathcal{Y}$ , i)  $\exists \hat{x}(y) \in \mathcal{X}$

satisfying  $D(\hat{x}(y)) = y$ , ii)  $\hat{x}(y)$  is unique, and iii) the mapping that inverts  $D$ , i.e.,  $\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$ , is continuous. Intuitively, i) and ii) guarantee that what the solution to the inverse problem is well defined, while iii) ensures that random deviations of the signal  $y$ , which are not accounted for in deterministic models, do not affect the solution  $\hat{x}(y)$  wildly. In many practical cases, an inverse problem is not well-posed (known as ill-posed), or establishing whether it is or not is challenging. This does not mean a solution can not be found, and in fact, most of the active research addresses precisely these cases. For example, in Publications A and B it is not established whether a solution always exists, and when it does, it is definitely not unique. Nonetheless, some of the methods I introduce below allow one to incorporate additional information into the problem and design an algorithm to provide an estimate for the solution.

A notion similar that is similar to well-posedness but applies to stochastic models is identifiability. A stochastic model  $M$  is identifiable if for any two parameters  $x_1, x_2 \in \mathcal{X}$ ,  $M(x_1) = M(x_2)$  is equivalent to  $x_1 = x_2$ . Here, the equality between two probability distributions should be interpreted in terms of agreeing measures, i.e., that for any  $\mathcal{S} \subset \mathcal{Y}$ ,  $M(x_1)[\mathcal{S}] = M(x_2)[\mathcal{S}]$ . The notation introduced in Section 1.1 reveals that identifiability is simply the parallel of condition ii) of well-posedness for stochastic models. Indeed, in simple cases such as the example of additive white Gaussian noise in Section 1.1, identifiability is simply a relaxed version of well-posedness. In general cases, however, the difference between ii) in well-posedness and identifiability is that now one requires knowledge on the entire distribution  $M(x)$  to uniquely map it back to  $x$ . This links the difficulty of an identifiable problem with the degree of structure in  $M(x)$ . On one hand, if  $M(x)$  is, for example, a product measure with many identical factors, i.e.,  $y$  is a large vector of independent and identically distributed samples, it may be straightforward to estimate  $x$  accurately. On the other hand, if  $M(x)$  does not have much structure, it may be demanding to obtain a good estimate of  $x$  from  $y$ . Similarly to well-posedness, identifiability may be challenging to verify, and its absence does not necessarily preclude consistent estimators. In Publication C, we study the identifiability of a deterministic-and-error stochastic model constructed by first deriving a non-linear deterministic model  $D$  and then adding stochastic terms to represent the known sources of error. There, we obtain a rather unexpected result: the model is identifiable when all sources of error are considered, but not when some are disregarded.

### Stochastic methods and logconcavity

A common technique to estimate the parameter  $x$  when one has a stochastic model  $M$  is maximum likelihood. In this methodology, one estimates  $x$  by maximizing the Radon-Nikodym derivative evaluated at the observed signal  $y$ , i.e., the likelihood,

$$\hat{x}(y) = \arg \max_{x \in \mathcal{X}} \{ \log [\mathcal{L}(x; y)] \}, \text{ where } \mathcal{L}(x; y) = \frac{dM(x)}{d\mu}(y). \quad (1.1)$$

The Radon-Nikodym derivative should be interpreted as an infinitesimal increment of probability when measuring  $y$  with the measure assigned to the parameter  $x$  by the stochastic model  $M$ . Here, the increment is measured with a reference measure  $\mu$  of  $\mathcal{Y}$  that should satisfy certain technical conditions, the most intuitive being absolute continuity, i.e., that for any  $\mathcal{S} \subset \mathcal{Y}$ ,  $\mu[\mathcal{S}] = 0$  implies that  $M(x)[\mathcal{S}] = 0$  for any  $x \in \mathcal{X}$ . This set-up accommodates a wide variety of signal spaces  $\mathcal{Y}$ , as long as one can find the right reference measure  $\mu$ . For example, in Publication D, we choose the counting measure on  $\mathcal{Y} = \mathcal{Z}$  (a countable set, see Fig. 1.1) as a reference to define a likelihood  $\mathcal{L}(x; y)$ , while in Publication C we choose the Lebesgue measure on  $\mathcal{Y} = \mathbb{R}^N$ . The maximum likelihood approach also has many benefits when  $y$  is composed of many independent and identically distributed replicates (large sample properties) [28, section 7.3.2]. However, there are many conditions that have to be fulfilled for (1.1) to be a valid definition of an estimator  $\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$ . Some of these are rather technical and guarantee the existence of a maximum in (1.1), while others contribute to its unicity. For example, if the model  $M$  is not identifiable, there may be multiple maxima of the likelihood for some observed signals  $y$ . Note here that for deterministic-and-error stochastic models, evaluating the likelihood involves evaluating the deterministic model to determine the parameters of the distribution. Thereby, each step in any numerical optimization technique to solve (1.1) comes at least at the computational cost of evaluating the deterministic forward model. Finally, in certain infinite-dimensional parameter spaces  $\mathcal{X}$ , direct maximum likelihood is known to exhibit theoretical problems (see [18]).

A property that guarantees that the multiple maxima of (1.1) (if any) are together in a convex set and can be found numerically is likelihood logconcavity. Basically, this property ensures that the cost functional in (1.1) is concave with respect to  $x$ , and thus, the optimization problem is convex if  $\mathcal{X}$  is convex. Many of the most common stochastic models have a parametrization with logconcave likelihood. Besides the advantages for maximum likelihood estimation, likelihood logconcavity also provides many benefits for other techniques, such as uncertainty quantification, hypothesis testing and the Bayes filter. In Publication D, we prove that a broad range of quantized linear models driven by continuous noises have logconcave likelihood with respect to both location and scale parameters. This result was initially stated (without explicit proof) in the statistical literature in the 1980s [5], but seems to have been overlooked by the signal processing community.

Nonetheless, practical problems in which a stochastic model is not identifiable or has a likelihood that is not logconcave are commonplace. Within the Bayesian community, there is an obvious way to proceed: if one first infers a probability distribution over the parameter space,  $\Pi_y \in \mathcal{P}(\mathcal{X})$ , that incorporates all the knowledge available, i.e., i) a prior distribution  $\Pi \in \mathcal{P}(\mathcal{X})$ , ii) the stochastic model  $M(x)$ , and iii) the fact that the signal  $y$  has been observed, one has all the information to build a good estimator of  $x$ . Then, according to Bayesian decision theory [28, ch. 5], one should choose a loss functional  $\ell : \mathcal{X} \times \mathcal{X} \rightarrow \bar{\mathbb{R}}$  such that  $\ell(x, \tilde{x})$  reflects the practical relevance of estimating  $\hat{x}(y) = \tilde{x} \neq x$ . Then, one should find the estimator



$\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$  that minimizes the posterior risk

$$r(\hat{x}|y) = \int_{\mathcal{X}} \ell(x, \hat{x}(y)) \, d\Pi_y, \quad (1.2)$$

for each possible observed  $y \in \mathcal{Y}$ . Bayesian estimation approaches have even been successful in infinite-dimensional parameter spaces, in what is known as the promising (and wrongly-named [22]) field of Bayesian nonparametric inference [16, 17].

Minimizing (1.2) often requires large amounts of computations. Most often, a simpler technique to build an estimate from the posterior distribution  $\Pi_y$  is used, i.e., maximum-a-posteriori estimation. In this technique, the Radon-Nikodym derivative of  $\Pi_y$  with respect to a measure  $\nu$  on  $\mathcal{X}$  is maximized. In particular, if the prior accepts a Radon-Nikodym derivative  $\pi(x) = d\Pi/d\nu(x)$  and the likelihood  $\mathcal{L}(x; y)$  is well defined, the maximum-a-posteriori estimate is

$$\hat{x} = \arg \max_{x \in \mathcal{X}} \{ \log [\mathcal{L}(x; y)] + \log [\pi(x)] \}. \quad (1.3)$$

In (1.3), the conditions for the optimization problem to be convex, and thus efficiently solvable numerically, are more relaxed. Indeed, even if the likelihood is not logconcave, the addition of the logarithm of the prior in (1.3) may still make the overall cost functional concave. In fact, when (1.3) is a convex problem, [25] shows that the maximum-a-posteriori estimator is actually a formal Bayes rule, i.e., it minimizes (1.2) for each observed signal  $y \in \mathcal{Y}$ , for a specific loss functional  $l(\cdot, \cdot)$  induced by the geometric structure of the posterior  $\Pi_y$  over  $\mathcal{X}$ .

## Regularization methods

Let now  $\mathcal{L}(x; y) = \exp(-g_y(x))$  and  $\pi(x) = \exp(-f(x))$ , with  $g_y$  and  $f$  proper, lower semi-continuous functionals. Then, we see that logconcavity of  $\mathcal{L}(x; y)$  and  $\pi(x)$  corresponds to convexity of  $g_y(x)$  and  $f(x)$ , and that we may write (1.3) as

$$\hat{x}(y) = \arg \min_{x \in \mathcal{X}} \{ g_y(x) + f(x) \}. \quad (1.4)$$

The minimization in (1.4) of the sum of a signal-dependent cost function  $g_y : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  and a signal-independent functional  $f : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  that promotes features of the solution that are known apriori or desired (also known as a regularizer) is much more general than Bayesian statistics. In fact, it is one of the most representative among regularization methods [14], which are techniques designed specifically to solve challenging inverse problems by incorporating additional (prior) information to the solution. Other regularization methods (see [14]) are i) projection or discretization methods that solve the inverse problem in a lower-dimensional space, and ii) iterative optimization techniques designed to be stopped after a number of steps. As an example of i), in [18], maximum likelihood methods are extended to otherwise problematic infinite-dimensional parameter spaces by introducing the

“method of sieves”, in which the parameter is estimated on a finite-dimensional subspace of the parameter space, with dimension that grows according to the amount of structure in  $M(x)$ .

In this thesis, I only study regularization through the variational formulation in (1.4) for ill-posed inverse problems that arise from deterministic models. In my exposition, I assume that both  $\mathcal{X}$  and  $\mathcal{Y}$  are Hilbert spaces, and therefore i) are equipped with norms  $\|\cdot\|_{\mathcal{X}}$  and  $\|\cdot\|_{\mathcal{Y}}$ , respectively, which allow one to measure distances between any two elements of the space, or to easily define neighbourhoods, ii) their norms are coherent to their respective inner products, i.e.,  $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$ , which enables interpretable Fréchet derivatives through their Riesz representation, and iii) are complete, which implies that they may have compact subsets, in which the extrema of continuous functions are met. The variational formulation of regularization is the most studied due to its simple interpretation and flexibility. In short, by choosing the functionals  $g_y(x)$  and  $f(x)$ , we are specifying, respectively, in exactly which sense do we want the solution to relate to the observed signal, and exactly which features we want to promote in it. Furthermore, the variational approach is attractive due to its connection to Tikhonov regularization, the most classic technique in the field. In Tikhonov regularization theory, a linear continuous operator  $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$  is considered as a deterministic linear model, i.e.,  $y = D(x) = Ax$ , where  $A$  has a non-empty nullspace. Thus, estimating  $x$  from  $y$  is an ill-posed inverse problem. Then, the Tikhonov technique is to estimate the parameter as  $\hat{x}(y) = (A^*A + \lambda I_d)^{-1} A^*y$ , where  $\lambda \geq 0$  is a regularization parameter and  $A^* \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$  is the adjoint to  $A$ . In fact, the Tikhonov estimate  $\hat{x}(y)$  is also the solution to (1.4) for  $g_y(x) = \|Ax - y\|_{\mathcal{Y}}^2$  and  $f(x) = \lambda \|x\|_{\mathcal{X}}^2$ . Here, the intuitive understanding is that the regularizer promotes solutions that are small in norm, so that the resulting inverse mapping,  $\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$ , is bounded, i.e., continuous, recovering the gold standard of point iii) in the definition of well-posedness (see above). Through this variational formulation of Tikhonov regularization, one can easily extend the approach to non-linear models by simply choosing  $g_y(x) = \|D(x) - y\|^2$ , which in general will not yield a closed form solution for (1.4). A common fact in variational regularization is that there are a number of criteria to select the “right” regularization parameter  $\lambda \geq 0$ , depending on the theoretical guarantees one aims to obtain. For instance, if we use the relation between (1.4) with maximum-a-posteriori estimation (see (1.3)) in Euclidean spaces, Tikhonov regularization corresponds to an isotropic normal stochastic model around  $D(x)$  for the signal  $y$  with isotropic normal prior around 0 for the parameter  $x$ , and  $\lambda$  is a ratio between the variances of the likelihood and the prior.

Despite its attractive closed-form solution, Tikhonov regularization only incorporates the loose prior knowledge that a solution that is “small” is preferred. Indeed, in linear Tikhonov regularization the operator  $A$  completely determines the parametric form of the solution, while  $f(x) = \lambda \|x\|^2$  only contributes to its specific coefficients. In contrast, the most common family of regularizers in current use, sparsity-promoting regularizers, fully characterize the parametric form of the

solution independently of  $A$  when  $g_y(x) = \|Ax - y\|_{\mathcal{Y}}^2$ , as shown by the representer theorems in [30, 31]. In these techniques, the idea is to identify some feature of the parameter that is known to be sparse, say  $R(x)$ , and to use as a regularizer a functional  $f$  that promotes zeroes in this feature. Here,  $R : \mathcal{X} \rightarrow \mathcal{R}$ , where  $\mathcal{R}$  is some Banach space, and one generally selects the regularizer as  $f(x) = \lambda \|R(x)\|_{\mathcal{R}}$ , i.e., proportional to the norm in  $\mathcal{R}$ , which should be a sparsity-promoting functional. Most commonly,  $\mathcal{R}$  will correspond to i)  $\ell_1$  or its subspaces in countable or finite-dimensional spaces, ii)  $L_1$  in function spaces, or iii) the space of signed Radon measures  $\mathcal{M}$ , i.e., the continuous dual of the space of continuous functions imbued with the  $L_1$  norm, in measure spaces [10, 30]. A number of particular cases of this approach have been extensively investigated, and theoretical properties and intuitive explanations can be found in many sources [10–13, 21, 30–32]. For example, sparsity-promoting regularizers have been linked back through the maximum-a-posteriori interpretation of (1.4) to the theory of sparse stochastic processes [32]. Finally, sparsity has had much success within the field of compressed sensing [13], in which the focus is on designing a methodology to represent accurately and with the least samples possible continuous (infinite-dimensional) signals that are known to have some underlying structure [12].

As an example, in Publication A, we have  $\mathcal{X} = L_2(B \times [0, \sigma_{\max}))$ ,  $\mathcal{Y} = L_2(\mathbb{R}^2)$  and  $\mathcal{R} = L_1(B)$  (for  $B \subset \mathbb{R}^2$  a bounded set, see Table 1.1 in Section 1.1 and Publication A) with  $R : L_2(B \times [0, \sigma_{\max})) \rightarrow L_1(B)$  such that  $R(a)(\mathbf{r}) = \|a(\mathbf{r}, \cdot)\|_{L_2([0, \sigma_{\max}))}$ , where  $x = a$  is a generic point in  $\mathcal{X}$  and  $\mathbf{r} \in \mathbb{R}^2$ . This selection is a group sparsity regularizer, in which one promotes sparsity on an object constructed by taking  $L_2/\ell_2$ -norms of subsets of the parameter. The aim in selecting this regularizer is to i) induce joint behavior in each of these subsets, i.e., either all elements in a subset become zero or all become non-zero, ii) promote boundedness in the mapping  $\hat{x} : \mathcal{Y} \rightarrow \mathcal{X}$ , and iii) promote sparsity in the number of subsets that are non-zero. This concept matches excellently with our cell detection application, in which the parameter  $a \in L_2(B \times [0, \sigma_{\max}))$  is expected to be sparse in its spatial dimensions,  $\mathbf{r} \in B$ , which represent cell locations, while the third dimension characterizes the scale description of the spots generated by those cells, which are supposed to stay in the same location throughout the experiment. Note here that the regularizer proposed in Publication A is slightly more complicated, including i) a term to impose non-negativity on  $a$ , and ii) a weighting function  $\xi$  over the domain  $[0, \sigma_{\max})$ , which can be used to incorporate further prior information.

### Iterative solvers for nonsmooth optimization problems

Problems of the form (1.4) do not generally have closed-form solutions. Consequently, estimators based on this optimization problem have to be obtained from iterative algorithms that converge to one of its solutions (if any). In this thesis, I discuss a specific first-order method [2] known as the accelerated proximal gradient (APG) algorithm or the “fast iterative shrinkage-thresholding algorithm” (FISTA), on which we based the results of our cell detection publications. First-

order methods are techniques that only require information on the first derivative (or subdifferential) of the functions to optimize. These methods are generally preferred when the parameter  $x$  (or its discrete representation) is high-dimensional and the forward evaluation of the deterministic model  $D(x)$  has a high computational cost. This is because each iteration comes at a cheaper cost in memory and computation compared to alternative approaches. Notwithstanding, for the common choice  $g_y(x) = \|D(x) - y\|_y^2$  each such iteration comes at a cost proportional to the evaluation of the forward model  $D(\cdot)$ . In the most common example of a linear model  $D(x) = Ax$ , for example, the major cost of an iteration is dominated by that of evaluating the gradient, and thus, evaluating either  $A$  and  $A^*$  once, or  $A^*A$  once.

The APG algorithm is particularly tailored to problems in which  $g_y$  is smooth (differentiable and with a Lipschitz gradient) and  $f$  is non-smooth, such as the regularizers discussed above to promote sparsity. In particular, the algorithm is described by the iterations

$$x^{(i)} \leftarrow \text{prox}_{\gamma f} \left[ \tilde{x}^{(i-1)} - \gamma \nabla g_y \left( \tilde{x}^{(i-1)} \right) \right], \quad (1.5)$$

$$\tilde{x}^{(i)} \leftarrow x^{(i)} + \alpha^{(i)} \left( x^{(i)} - x^{(i-1)} \right). \quad (1.6)$$

Here,  $\gamma \geq 0$  depends on the smoothness properties of  $g_y$ , while  $\alpha^{(i)}$  is a sequence that regulates the momentum term  $x^{(i)} - x^{(i-1)}$ , and thereby, the speed of convergence (see Publication B for some details and [2] for a comprehensive overview). Additionally, the proximal operator of the functional  $f$  is defined as

$$\text{prox}_{\gamma f}(x) = \arg \min_{\tilde{x} \in \mathcal{X}} \left\{ \|\tilde{x} - x\|_{\mathcal{X}}^2 + 2\gamma f(\tilde{x}) \right\}. \quad (1.7)$$

Intuitively, the prox operator can be seen as a bridge between the two extremes of a projection and a gradient step. Indeed, on one extreme, if we consider the  $(\infty, 0)$ -indicator of a convex set  $C$ , i.e., the functional  $\delta_C : \mathcal{X} \rightarrow \mathbb{R}$  such that  $\delta_C(x) = \infty$  if  $x \notin C$  and  $\delta_C(x) = 0$  if  $x \in C$ , we see that  $\text{prox}_{\gamma \delta_C}(x) = P_C[x] = \arg \min \{ \|\tilde{x} - x\|^2 : \tilde{x} \in C \}$ , i.e., the proximal operator is simply a projection onto  $C$ . On the other extreme, if  $f$  is convex and differentiable and  $p = \text{prox}_{\gamma f}(x)$ , we have that  $x = p + \gamma \nabla f(p)$ , i.e., ascending one gradient step from  $p$  would lead us to  $x$ .

The APG algorithm, then, is readily implementable for a large collection of problems, and provides some form of convergence guarantees regardless of the convexity assumptions on  $g_y$  and  $f$  [2]. Of particular interest is the worst-case function-value convergence rate of  $\mathcal{O}(1/k^2)$  when both  $g_y$  and  $f$  are convex and respect the conditions above. Given a new problem, one simply needs to compute, bound above, or approximate  $\gamma$ , and have routines to evaluate the gradient of  $g_y$  and the prox operator for  $f$ . In Publication B, we derive a number of results in order to employ the APG algorithm. On one hand, we have a linear model with squared norm cost,

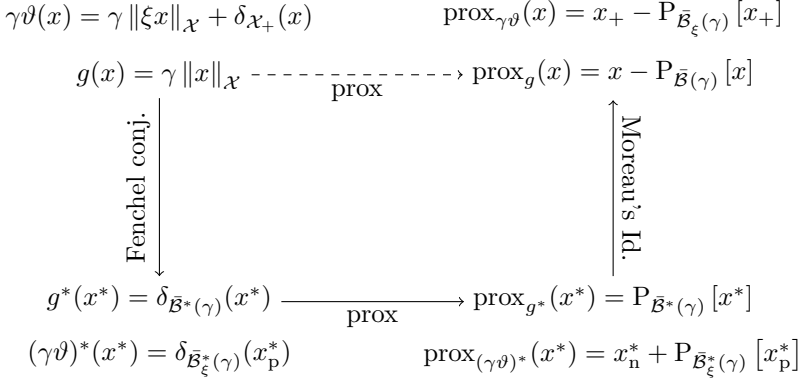


Figure 1.3: Scheme followed in the Appendix of Publication B to prove the expression for the proximal operator of the non-negative weighted norm in a Hilbert space  $\mathcal{X}$ . Here,  $\xi \in \mathcal{X}$ ,  $P_C : \mathcal{X} \rightarrow C$  is the projection onto a convex set  $C$ ,  $\bar{B}(\gamma)$  is the closed ball with norm bounded by  $\gamma$ ,  $\bar{B}_\xi(\gamma)$  is a closed ellipsoid with  $1/\xi$ -weighted norm bounded by  $\gamma$ , and  $\bar{B}^*(\gamma)$  and  $\bar{B}_\xi^*(\gamma)$  are their dual equivalents. The result indicated by the dashed arrow is obtained by following the solid arrows, using the results in the lower corners of the square as stepping stones to achieve it. In the inner square, the classical proof of the prox of a norm in a Hilbert space. In the outer square, the results of our generalization to the weighted, non-negatively constrained norm.

and so we obtain  $\nabla g_y(x)$  and  $\gamma$  by i) deriving the adjoint to what we call the diffusion operator, the mapping,  $a \in L_2(B \times [0, \sigma_{\max})) \mapsto \int_0^{\sigma_{\max}} G_\sigma a_\sigma d\sigma$ , where  $G_\sigma$  is a Gaussian blur with scale  $\sigma \geq 0$ , and ii) bounding the norm of this same operator to quantify the smoothness of  $\nabla g_y$ . On the other hand, we use a non-negative group-sparsity regularizer, i.e.,

$$f(a) = \lambda \left\| \|a(\mathbf{r}, \cdot)\|_{L_2([0, \sigma_{\max}))} \right\|_{L_1(B)} + \delta_{L_{2,+}(B \times [0, \sigma_{\max}))}(a). \quad (1.8)$$

A major technical result in Publication B is the derivation of the prox operator for (1.8) in closed form for any Hilbert space. The most important step towards that result is obtaining the prox operator of a non-negative weighted norm in any Hilbert space, a process which is summarized in Fig. 1.3. As it turned out, the result depicted in Fig. 1.3 was encompassed by a previous and broader one [4, Proposition 2.2]. Furthermore, the full expression for the prox of (1.8) was simultaneously derived in a broader setting in [7, Lemma 2.2] and in a more restricted setting in [34, Lemma 2], which definitely highlights the interest and timeliness of the result. Regardless, the result proved useful in the APG algorithm we obtained in Publication B and exploited in Publications E and F.

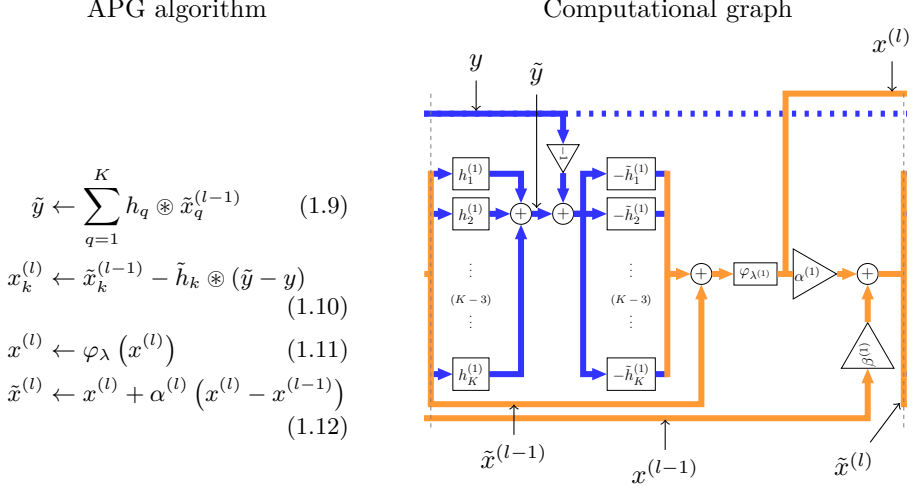


Figure 1.4: Steps and computational graph of a generic iteration of the APG algorithm for  $g_y(x) = \|Ax - y\|^2$  with  $\mathcal{X} = \mathbb{R}^{M,N,K}$ ,  $\mathcal{Y} = \mathbb{R}^{M,N}$ , and  $A$  such that  $Ax = \sum_{k=1}^K h_k \otimes x_k$ , with  $\{h_k\}_1^K$  a set of arbitrary convolutional kernels. Here,  $\tilde{h}_k$  refers to the matched filter for the corresponding  $h_k$ , and the regularizer is left unspecified under the assumption that  $\varphi_\lambda(x) = \text{prox}_{\lambda f}(x)$ . Here, simplifying assumptions on the norms of the  $\{h_k\}_1^K$  have been made with no loss of generality (see Publication G), and the computational graph incorporates some more degrees of freedom, e.g.,  $\alpha^{(l)} + \beta^{(l)} \neq 0$ . In a learned iterations framework, one trains a selection of the algorithm parameters above, i.e.,  $\alpha^{(l)}$ ,  $\beta^{(l)}$ ,  $\lambda^{(l)}$ , and the  $h_k^{(l)}$ s, independently for each layer.

Although first-order methods are computationally attractive, their rate of convergence is often too slow for practical applications. The advent of deep learning techniques, however, has generated technology such as differentiable programming frameworks. In these frameworks, any algorithm can be adapted through the optimization of a loss function on a collection of examples. This has led to the novel research topic known as unfolded algorithms, learning to learn, loop unrolling, or simply learned iterations. In this field, one typically implements a given number of iterations of a first-order method to solve a specific optimization problem within a differentiable programming framework. The number of iterations is either as small as possible or adapted to the computational requirements of the end application. Then, one gathers pairs of signals  $y$  and desired solutions  $\hat{x}(y)$ , and adapts a selection of parameters of the given steps so that the output of the resulting graph is as close as  $\hat{x}(y)$  as possible, in a precise sense defined by a given loss function. These pairs may be extracted from running the original first-order method on a platform with higher computational capabilities, or be artificially generated by exploiting the

forward model. Although simple, it has been shown that this technique can result on very fast algorithms [6] when it is trained adequately. Furthermore, even if these techniques lack many of the theoretical guarantees of more traditional approaches, they may be used as heuristic starting points for conventional iterative methods that have strong guarantees and convergence criteria, and thus still reduce the computational burden significantly. To the best of our knowledge, Publication G is the first to apply this approach to the APG algorithm, as described in Fig. 1.4.





## Chapter 2

# Overview of contributions

In this chapter, I highlight the contributions of each publication included in the thesis in the light of the context presented in Chapter 1. Furthermore, I attempt to duly acknowledge the work of all my co-authors, collaborators, reviewers, and other agents without whom these would not have been possible.

### **Publication A, Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse Problems, *Journal paper***

- Authors: Pol del Aguila Pla (**PdAP**) and Prof. Joakim Jaldén (**JJ**).
- Influential figures: Prof. Mark A. Davenport, the team of four anonymous reviewers, Prof. Lars Jonsson, Dr. Axel Ringh, Assoc. Prof. Johan Karlsson, Assoc. Prof. Ozan Öktem, Dr. Holger Kohr, Dr. Christian Smedman and Dr. Celia García-Pareja.

### **Contributions**

In this paper, we start by introducing the closed-form solution for the forward problem of a known deterministic linear model for certain image-based immunoassays (Theorem 1). In particular, we reparametrize the model so that forward evaluation can be approximated efficiently by two stages composed of simple (finite, by Lemmas 1 and 2) sums and convolutions. Then, we pose the inverse problem corresponding to the second stage of this new formulation of the model (Section III), which allows us to recover the cell detection information we seek in a manner that is independent of the unknown physical characteristics of the assays. We propose to solve this inverse problem using a regularized variational approach in function spaces (13), and we provide ingredients towards showing its good condition (Lemmas 3 and 6). Furthermore, we provide the adjoint and norm of the linear continuous operator at the core of the linear model, which are fundamental results for

any gradient-based iterative scheme to solve variational problems that involve this model (Lemmas 4 and 5). We then propose a specific discretization scheme and provide empirical cell detection results on real data that suggest human-like performance. These results are obtained using the algorithms we derived in Publication B for the specific variational problem proposed here.

## Division of work

The structure of the paper was designed jointly by both authors, with very substantial contributions from the fantastic team of reviewers and Associate Editor Prof. Mark A. Davenport. The writing of the paper was mostly **PdAP**'s responsibility, but some specific sections were heavily based on **JJ**'s initial drafts. **JJ** provided extensive feedback in several internal rounds of review, and occasionally polished some sections. **PdAP** did the literature review, classified the relevant state of the art and wrote the introduction. The paper was reviewed extensively by both authors.

**JJ** identified the physical model and linked it to the observation model (Theorem 1, Lemma 1, and Theorem 2 together with their proofs in Appendix A), where valuable insights and helpful discussions with Prof. Lars Jonsson were much appreciated. Nonetheless, **PdAP** decided on the final structure of the corresponding sections and was mostly responsible for their presentation and technical correctness. **PdAP** developed Lemma 2 and its proof in Appendix A to strengthen the results of Theorem 2.

**JJ** had the original idea to pose an inverse problem to recover a 3-dimensional characterization of the assay and implemented solvers for the naive least squares problem (12). The addition of a group-sparsity regularizer emerged from a discussion between **JJ** and **PdAP** after **PdAP** implemented and compared empirically 4 different proximal-optimization solvers for 3 different regularized versions of (12). **PdAP** formulated the inverse problem in function spaces at **JJ**'s suggestion and as a reaction to Theorem 1 not requiring any discretization. Lemmas 3 to 5 and their proofs in Appendix B were derived by **PdAP**. However, substantial contributions from a reviewer heavily improved the proof of Property 1 therein. Lemma 6 and its proof in Appendix B were formulated and proved by **PdAP** at the suggestion of a reviewer and with some guidance from **JJ**.

The discretization in Section IV was heavily based on Dr. Axel Ringh's and Assoc. Prof. Johan Karlsson's valuable advice, as well as on an unpublished document by Assoc. Prof. Ozan Öktem and Dr. Holger Kohr that relates to [23]. The real-data example was implemented by **PdAP** using data provided by Mabtech AB and annotations by Dr. Christian Smedman. Finally, the choice of the random seeds to obtain realistic paths in Fig. 1, as well as the choice of the specific section where detection results on real data are shown in Fig. 4, was done by Dr. Celia García-Pareja, who also gave extensive feedback on earlier versions of Section I.

## Publication B, Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation, *Journal paper*

- Authors: Pol del Aguila Pla (**PdAP**) and Prof. Joakim Jaldén (**JJ**).
- Influential figures: Prof. Mark A. Davenport, the team of four anonymous reviewers, Prof. Mário A. T. Figueiredo, Prof. Radu Ioan Bot, Prof. Krister Svanberg, and Dr. Celia García-Pareja.

### Contributions

In this paper, we provide an iterative solver for the regularized variational problem we proposed in Publication A to solve an inverse problem in image-based immunoassays. For that, we provide a closed-form expression for the proximal operator of the non-negative group-sparsity regularizer (Theorems 1 and 2, along with Lemmas 1 to 4 as part of their extensive proof in the Appendix). Lemma 4, the proximal operator of the non-negative weighted norm in a Hilbert space, is included in previous work that we did not know at the time of acceptance [4, Proposition 2.2]. Theorems 1 and 2 are novel results, but were simultaneously derived in other works (see [7, Lemma 2.2] and [34, Lemma 2]), albeit in slightly different settings. After this result, we proceed to the thorough empirical evaluation of the proposed methodology on synthetic data under different experimental conditions. We evaluate the performance of our methodology both in terms of optimal-transport metrics that evaluate the quality of our inversion and in terms of purely operational metrics for cell detection. Yet again, the empirical results suggest human-like performance. The synthetic observations used in the empirical results are generated using the forward evaluation procedure presented in Publication A.

### Division of work

Similarly to Part I, the structure of the paper was designed jointly by both authors, with very substantial contributions from the fantastic team of reviewers and Associate Editor Prof. Mark A. Davenport. **PdAP** wrote the paper, with extensive feedback in several rounds of internal review by **JJ**. **PdAP** did the literature review and classified the relevant state of the art. The original idea to bring forward the contribution in Theorems 1 and 2 by writing Section I.A. can be attributed to one of the reviewers, and was heavily influenced by Prof. Mário A. T. Figueiredo, who, at the workshop *Generative models, parameter learning and sparsity* at the Isaac Newton Institute for Mathematical Sciences in 2017, also pointed out the importance of the result.

The introduction to proximal optimization, Theorem 1, and Theorem 2, as well as their proofs in the Appendix (and Lemmas 1, 2, 3, and 4) were developed by **PdAP** and reviewed by **JJ**. This was done in function spaces at **JJ**'s suggestion

and as a reaction to the model formulation in Publication A not requiring any discretization. To that respect, Prof. Radu Ioan Bot’s lecture series *Recent Advances in Numerical Algorithms for Convex Optimization*, given at KTH in May 2016, and Prof. Krister Svanberg’s course *SF3810 - Convexity and optimization in linear spaces*, given at KTH in 2015, were instrumental. Furthermore, the proof technique employed to show Lemma 1 was suggested by one of the anonymous reviewers.

The decision to implement the algorithm on graphical processing units and perform the filtering in the sample domain by approximating the kernels with low-rank or separable filters came after much discussion and empirical testing by both **JJ** and **PdAP**. **PdAP** investigated the different low-rank kernel approximations and discussed the results with **JJ**. **PdAP** implemented and designed the numerical results section, with the help of occasional discussions with **JJ**. In particular, **PdAP** almost exclusively developed and implemented the data-generation scheme detailed in the supplementary material. The optimal transport distance referred to as earth mover’s distance (EMD) was included by **PdAP** at the suggestion of one of the reviewers after discussions with **JJ**. Finally, the choice of the specific section where detection results on synthetic data are shown in Fig. 4 was done by Dr. Celia García-Pareja, who also gave extensive feedback on earlier versions of the Appendix.

### Publication C, Clock synchronization over networks — Identifiability of the sawtooth model, *Submitted manuscript*

- Authors: Pol del Aguila Pla (**PdAP**), Lissy Pellaco (**LP**), Dr. Satyam Dwivedi (**SD**), Prof. Peter Händel (**PH**) and Prof. Joakim Jaldén (**JJ**).
- Influential figures: Dr. Hugo Tullberg and Gerard Farré.

### Contributions

In this paper, we start by introducing novel non-linear deterministic and stochastic models for round-trip time measurements for short pulses between networked nodes under specific conditions (Theorem 1 and (9), under the protocol in Fig. 2). We validate the deterministic model by simulating the behavior of two such nodes and visually assessing the similarity of the predicted signal and the observed signal. We then proceed to provide identifiability results (negative in Lem. 1 and positive in Theorem 2) for a broader class of stochastic models, i.e., sawtooth signals, under different assumptions. In particular, we show that the addition of noise that affects the observation in a non-linear manner makes the model identifiable. We then propose estimation performance lower bounds derived from a linear approximation of our stochastic model, and present simple estimation techniques. We conclude by exemplifying these performance references through extensive numerical results and identifying future lines of research for the estimation of clock synchronization pa-

rameters using our stochastic model. Furthermore, we provide the implementation of all and any computations in a publicly accessible repository in [8].

## Division of work

The overall structure of the paper was designed by **PdAP** and discussed with **JJ**. The structure of Sections IV and V was designed by **PdAP** and discussed with **LP**. **PdAP** wrote the paper, although Section IV.B was based on an initial draft by **LP**, and both **JJ** and **LP** provided extensive feedback in several rounds of internal review. **PH** provided extensive and critical feedback in one final round of internal review.

The original research idea of exploring the theoretical condition of the sawtooth model was proposed by **SD**, who also introduced **JJ** and **PdAP** to clock synchronization over networks. **PdAP** derived the deterministic model in Theorem 1, formulated the stochastic model (9) and designed the standard parametrization in (10). Fig. 4 was designed by **PdAP** based on data obtained by **SD** and **PH** for a previous publication. Fig. 5 was designed by **LP** and **PdAP**, and implemented by **LP**. **PdAP** developed Lem. 1, Theorem 2 and their proofs, which were extremely improved by discussions with **JJ**. Furthermore, the proof of Theorem 2 was only completed due to Dr. Hugo Tullberg, who first directed the authors to the existing results on circular statistics. Similarly, the discussion of Lem. 1 was only complete after Gerard Farré suggested previous results from dynamical systems on the orbit of a rotation of the circle. **PdAP** developed the graphical representation of the proofs of Lemma 1 and Theorem 2 in Figs. 6, 7 and 8, which improved substantially due to extensive feedback from **JJ** and minor feedback from **LP**. **PdAP** also developed the Cramér-Rao lower bounds in Section IV.A (and their proofs in the Appendix and the supplementary material) at the suggestion of **SD**, **JJ** and **PH**, and with some feedback from **LP**.

The basic ideas for the estimation strategies were suggested by **PH** and **SD**, and were implemented and improved upon by **LP** and **PdAP**. All the empirical results were generated by **LP** and **PdAP**, with the exception of the simulated protocol that resulted in Fig. 3, which was entirely implemented and generated by **PdAP**.

## Publication D, Inferences from quantized data — Likelihood logconcavity, *Manuscript*

- Authors: Pol del Aguila Pla (**PdAP**) and Prof. Joakim Jaldén (**JJ**).
- Influential figures: Dr. Celia García-Pareja.

## Contributions

In this paper, we present to the signal processing community the most general likelihood logconcavity statement for quantized data to date (previously stated

in [5]), together with its proof, which, to the best of our knowledge, has never been published. In particular, we show how Prékopa’s theorem can be used to show that the likelihood for quantized linear models is jointly logconcave with respect to both its location and scale parameter in a broad range of cases (Theorem 1, including Lemmas 2 and 3 as partial steps). In order to show this result and explain the limitations of the proof technique, in Lemmas 4 and 5 we study sets generated by the generalization of the coefficients of a convex combination to positive semi-definite matrices (i.e., matrices whose sum is the identity). Finally, by the time of submission the manuscript will also include examples from a number of signal processing applications in which this result is useful.

### Division of work

The overall structure of the paper was designed by **PdAP** and discussed with **JJ**. **PdAP** wrote the paper, and **JJ** provided feedback in several rounds of internal review.

The original research idea of attempting to show likelihood logconcavity for quantized data originated in discussions between both authors. Both authors collaborated in proofs for specific subcases such as when the data were normal and identically distributed, or when the quantizers were composed of simple monotonic analog-to-digital converters for each dimension. Motivated by **JJ**, **PdAP** generalized the results to the current form of Theorem 1 and developed Lemmas 1 to 4 after discovering Prékopa’s theorem in the literature. For Lemma 5, **PdAP** proved that points generated by the generalization of convex combinations to positive semidefinite matrices would lie inside the ball and conjectured the converse, which **JJ** proved. **JJ** provided considerable feedback and advice on the first versions of the proofs of Lemmas 1 to 4. After having completed the proof of Theorem 1, **PdAP** discovered that [5] had presented a similar result without explicit proof that did refer to Prékopa’s theorem. The illustrations of Lemmas 4 and 5 in Fig. 4 were suggested by **JJ** and implemented by **PdAP**.

The signal processing applications used as examples will mostly be suggested by **JJ**, discussed between both authors, and implemented by **PdAP**. Finally, Dr. Celia García Pareja provided feedback on writing and clarity in several versions of the work, and found a major flaw in a previous version of the proofs.

### Publication E, Cell detection on image-based immunoassays, *Conference paper*

- Authors: Pol del Aguila Pla (**PdAP**) and Prof. Joakim Jaldén (**JJ**).
- Influential figures: The team of four anonymous reviewers.

## Contributions

In this paper, we present a summary of the work in Publications A and B in a timely manner and in a language accessible to a broader community. The variational problem and the iterative solver are presented directly in finite-dimensional spaces, and the focus is placed on the application, operational evaluation metrics, and empirical results on real data.

## Division of work

The structure and content of the paper was decided jointly between both authors. **PdAP** wrote the article and **JJ** provided feedback in several rounds of internal review. The anonymous team of reviewers contributed to making the exposition of the results more accessible.

## Publication F, Convolutional group-sparse coding and source localization, *Conference paper*

- Authors: Pol del Aguila Pla (**PdAP**) and Prof. Joakim Jaldén (**JJ**).

## Contributions

In this paper, we draw the connection between problems that are very popular in different fields, i.e., convolutional sparse coding in machine learning, spatially variant deconvolution in mathematical imaging, and source-localization problems in many scientific-imaging application domains. Then, we demonstrate the strength of this connection by employing iterative solvers very similar to those developed in Publication B for cell detection for foreground recovery in astronomical data.

## Division of work

The structure of the paper was decided jointly between both authors. **PdAP** developed most of the content and wrote the paper, and **JJ** provided feedback in several rounds of internal review. **PdAP** chose and implemented the real-data example using data captured by the Hubble telescope.

## Publication G, SpotNet — Learned iterations for cell detection in image-based immunoassays, *Conference paper*

- Authors: Pol del Aguila Pla (**PdAP**), Vidit Saxena (**VS**) and Prof. Joakim Jaldén (**JJ**).

## Contributions

In this paper, we expand on the available empirical basis for the neural network design strategy known as learned iterations, in which the neural network is designed to replace an iterative solver to solve a variational problem by borrowing its computational graph. In this case, we used the structure of the solver developed in Publication B to design a neural network for cell detection, and provided ample evidence of the advantages of this approach when compared to another popular neural network design. Furthermore, we provide the implementation for any and all computations involved in the paper in a publicly available repository, including the synthetic data generation procedure proposed in Publication B.

## Division of work

The structure of the paper was decided by **VS** and **PdAP**, who also wrote the paper with some feedback from **JJ**. The idea of the paper was proposed and developed by **PdAP**. **VS** and **PdAP** implemented all the necessary code, and **PdAP** designed Figs. 1 to 4, with extensive feedback by **VS**.



# Bibliography

- [1] Richard C. Aster, Brian Borchers, and Clifford H. Thurber. *Parameter estimation and inverse problems*. Elsevier, 2013.
- [2] Amir Beck. *First-order methods in optimization*. SIAM & MOS, 2017.
- [3] Mario Bertero and Patrizia Boccacci. *Introduction to inverse problems in imaging*. IOP Publishing, 1998.
- [4] Luis M. Briceño-Arias and Patrick L. Combettes. Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery. *Numerical Mathematics: Theory, Methods & Applications*, 2(4):485–508, 2009.
- [5] J. Burridge. Some unimodality properties of likelihoods derived from grouped data. *Biometrika*, 69(1):145–151, 1982.
- [6] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS’18, pages 9079–9089, 2018.
- [7] Patrick L. Combettes and Christian L. Müller. Perspective functions: Proximal calculus and applications in high-dimensional statistics. *Journal of Mathematical Analysis and Applications, Special Issue on Convex Analysis and Optimization: New Trends in Theory and Applications*, 457(2):1283–1306, 2018.
- [8] Pol del Aguila Pla and Lissy Pellaco. Clock sync and range. GitHub repository, 2019. URL [https://github.com/poldap/clock\\_sync\\_and\\_range](https://github.com/poldap/clock_sync_and_range).
- [9] Pol del Aguila Pla and Vidit Saxena. SpotNet. GitHub repository, 2018. URL <https://github.com/poldap/SpotNet>.
- [10] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, October 2017.

- [11] Michael Elad. *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer, 2010.
- [12] Yonina C. Eldar. *Sampling theory beyond bandlimited systems*. Cambridge University Press, 2015.
- [13] Yonina C. Eldar and Gitta Kutyniok, editors. *Compressed sensing*. Cambridge University Press, 2012.
- [14] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, 1996.
- [15] Celia García-Pareja. *Topics on mathematical statistics for medical applications: Summary measures and exact simulation of diffusions*. PhD thesis, Karolinska Institutet, 2019.
- [16] Subhashis Ghosal and Aad van der Vaart. *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press, 2017.
- [17] Jayanta Kumar Ghosh and R. V. Ramamoorthi. *Bayesian nonparametrics*. Springer, 2003.
- [18] Ulf Grenander. *Abstract inference*. John Wiley & Sons, 1981.
- [19] Jari P. Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Springer, 2005.
- [20] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*. Springer, 2011.
- [21] Stéphane Mallat. *A wavelet tour of signal processing: The sparse way*. Elsevier, 2009.
- [22] Peter Müller and Fernando A. Quintana. Nonparametric Bayesian data analysis. *Statistical Science*, 19(1):95–110, February 2004.
- [23] Ozan Öktem, Jonas Adler, and Holger Kohr. Operator discretization library (ODL). GitHub repository, 2018. URL <https://github.com/odlgroup/odl>.
- [24] Judea Pearl. *Causality: Models, reasoning and inference*. Cambridge University Press, 2009.
- [25] Marcelo Pereyra. Revisiting maximum-a-posteriori estimation in log-concave models. *SIAM Journal on Imaging Sciences*, 12(1):650–670, 2019.
- [26] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. MIT Press, 2017.
- [27] Michele Santacatterina. *Mathematical programming for optimal probability weighting*. PhD thesis, Karolinska Institutet, 2018.

- [28] Mark J. Schervish. *Theory of statistics*. Springer, 1995.
- [29] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- [30] Michael Unser. A unifying representer theorem for inverse problems and machine learning. Preprint: arXiv:1903.00687, March 2019.
- [31] Michael Unser, Julien Fageot, and Harshit Gupta. Representer theorems for sparsity-promoting  $\ell_1$  regularization. *IEEE Transactions on Information Theory*, 62(9):5167–5180, September 2016.
- [32] Michael Unser and Pouya D. Tafti. *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.
- [33] Curtis R. Vogel. *Computational methods for inverse problems*. SIAM, 2002.
- [34] Masahiro Yukawa and Hideaki Kagami. Supervised nonnegative matrix factorization via minimization of regularized Moreau-envelope of divergence function with application to music transcription. *Journal of the Franklin Institute, Special issue on recent advances in machine learning for signal analysis and processing*, 355(4):2041–2066, 2018.



# Part II

# Publications

*Have ya guessed yet!? Who's the best yet? If ya ain't, I'll tell ya one more time.  
You'll bet your last dime. In all of the world so far I am the greatest, greatest star!*

Lyrics by *Bob Merrill*, "Funny Girl", 1964.

Favourite version performed by *Barbra Streisand* in the 1968 movie of the same name.

## Publication A

# Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse Problems

Journal paper. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla and J. Jaldén, *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5407–5421, Oct. 2018.

*A number of idealizations and assumptions were made when deriving the model. In practice these may be more or less invalid.*

Modified, from *M. Jansson's*, “On subspace methods in system identification and sensor array signal processing”, Doctoral thesis at KTH Royal Institute of Technology, 1997.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-233824> for the full text of the paper.





## Publication B

# Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation

Journal paper. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla and J. Jaldén, *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5422–5437, Oct. 2018.

*One may go quite far using the functional formulation, even if, at the end, some sort of discretization is used for the actual computations.*

From A. Tarantola's "Inverse Problem Theory and Methods for Model Parameter Estimation", SIAM, 2012.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-233827> for the full text of the paper.



# Cell detection by functional inverse diffusion and non-negative group sparsity — Supplementary material

Supplementary material to Publications A and B. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla and J. Jaldén, *IEEE Transactions on Signal Processing*, Oct. 2018.

*... cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.*

From  
*Pierre de Fermat's* notes on his copy of *Diophantus' "Arithmetica"*, 1637.

See [https://ieeexplore.ieee.org/ielx7/78/8464033/8453854/supp\\_mat.pdf?tp=&arnumber=8453854](https://ieeexplore.ieee.org/ielx7/78/8464033/8453854/supp_mat.pdf?tp=&arnumber=8453854) for the complete supplementary material.



## Publication C

# Clock synchronization over networks — Identifiability of the sawtooth model

Submitted manuscript. P. del Aguila Pla, L. Pellaco, S. Dwivedi, P. Händel and J. Jaldén, submitted to *IEEE Transactions on Control Systems Technology*, 2019.

*A consistent estimator of  $\Theta$  fails to exist if the parameter is not identifiable through the sequence of data values.*

From Mark J. Schervish's, "Theory of Statistics", Springer, 1995.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-256072> for the full text of the paper.



# Clock synchronization over networks — Identifiability of the sawtooth model, Supplementary material

Supplementary material to Publication C.

The supplementary material will be available online after the publication of the paper.





## Publication D

# Inferences from quantized data — Likelihood logconcavity

Manuscript. P. del Aguila Pla and J. Jaldén, *Work in progress*, 2019.

*At the time, most of these techniques were based on digitally approximating various well-known analog methods.*

From A. V. Oppenheim's,  
“Algorithm kings — The birth of digital signal processing”, *IEEE Solid-State Circuits Magazine*, vol. 4, no. 2, pp. 34–37, Spring 2012.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-256078> for the full text of the paper.



## Publication E

# Cell detection on image-based immunoassays

Conference paper. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla and J. Jaldén, *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr. 2018, pp. 431–435.

*Tidigare har analysen av inspelad data ofta gjorts manuellt, men detta är mycket tidskrävande och resultaten blir ofta subjektiva och svåra att reproducera.*

From *K. E. G. Magnusson's*, “Segmentation and tracking of cells and particles in time-lapse microscopy”, Doctoral thesis at KTH Royal Institute of Technology, 2016.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-223933> for the full text of the paper.



## Publication F

# Convolutional group-sparse coding and source localization

Conference paper. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla and J. Jaldén, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 2776–2780.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-224253> for the full text of the paper.



## Publication G

# SpotNet — Learned iterations for cell detection in image-based immunoassays

Conference paper. © 2018 IEEE. Reprinted, with permission, from P. del Aguila Pla, V. Saxena and J. Jaldén, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 1023–1027.

See <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-250464> for the full text of the paper.