A nonconvex variational approach for receptive field estimation

Audric Drogoul, Gilles Aubert, Bruno Cessac, Pierre Kornprobst

▶ **To cite this version:**

Audric Drogoul, Gilles Aubert, Bruno Cessac, Pierre Kornprobst. A nonconvex variational approach for receptive field estimation. [Research Report] 8837, Inria Sophia Antipolis. 2016, pp.42. hal-01279999v3

HAL Id: hal-01279999
https://inria.hal.science/hal-01279999v3

Submitted on 31 Aug 2016

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

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

A nonconvex variational approach for receptive field estimation

Audric Drogoul, Gilles Aubert, Bruno Cessac, Pierre Kornprobst

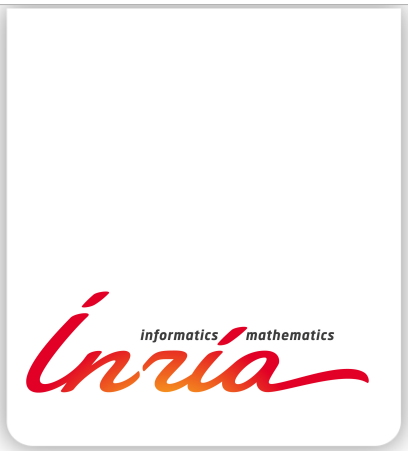
RESEARCH REPORT

N° 8837

September 2016

Project-Team Biovision

ISSN 0249-6399 ISRN INRIA/RR--8837--FR+ENG



A nonconvex variational approach for receptive field estimation

Audric Drogoul*, Gilles Aubert†, Bruno Cessac*, Pierre Kornprobst*

Project-Team Biovision

Research Report n° 8837 — September 2016 — 40 pages

Abstract: Determining the receptive field of a visual sensory neuron is a first but crucial step towards the characterization of neurons response to local spatio-temporal stimuli. Existing methods are based on convex optimization methods neglecting biophysical constraints of neurons (bounded firing rate), and they are relatively poor in terms of accuracy and running time. We propose a new method to estimate receptive fields by a nonconvex variational approach, thus relaxing the simplifying and unrealistic assumption of convexity made by standard approaches. The method consists in studying a relaxed discrete energy minimized by a proximal alternating minimization algorithm. We compare our approach with the classical spike-triggered-average technique on simulated data, considering a typical retinal ganglion cell. Results show a high improvement in term of accuracy and convergence with respect to the duration of the experiment.

Key-words: Neurosciences, inverse problem, receptive field estimation, variational approach, nonsmooth and nonconvex optimization, Kurdyka-Lojasiewicz

* Université Côte d'Azur, Inria, Biovision team, France
† Université Côte d'Azur, CNRS, LJAD, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Une approche variationelle non-convexe pour l'estimation des champs recepteurs

Résumé : Déterminer le champ récepteur d'un neurone sensoriel visuel est une première mais cruciale étape pour caractériser la réponse neuronale à des stimuli spatio-temporels locaux. Les méthodes existantes sont basées sur des approches d'optimisation convexes négligeant les contraintes biophysiques des neurones (taux de décharge borné), et elles sont relativement faibles en terme de précision et de temps de calcul. Nous proposons une nouvelle méthode pour estimer les champs récepteurs basée sur une approche variationnelle non convexe, relaxant ainsi l'hypothèse simplificatrice et irréaliste de convexité faite dans les approches standard. La méthode consiste à étudier une énergie discrète relaxée, minimisée par un algorithme proximal de minimisation alternée. Nous comparons notre approche avec la technique classique de *spike triggered averaged* sur des données simulées, en considérant une cellule ganglionnaire type de la rétine. Les résultats montrent une forte amélioration en terme de précision et de convergence par rapport à la durée de l'expérience.

Mots-clés : Neurosciences, problème inverse, estimation de champs récepteur, approche variationnelle, optimisation non régulière et non convexe, Kurdyka-Lojasiewicz

Contents

1 Introduction **4**

2 The discrete problem: well-posedness and relaxation **6**
2.1 Preliminaries . . . 6
2.1.1 Notions of subdifferentials . . . 6
2.1.2 Notion of proximal operator . . . 7
2.1.3 Notion of conjugate function . . . 7
2.2 Problem definition . . . 7
2.3 Well-posedness . . . 9
2.4 Main algorithm . . . 11
2.5 Solving problem (16a) . . . 12
2.5.1 Using proximal operator definition . . . 12
2.5.2 Case $\boldsymbol{\xi_i = 0}$ (no spike) . . . 12
2.5.3 Case $\boldsymbol{\xi_i > 0}$ (presence of spikes) . . . 14
2.5.4 Conclusion on a good choice of sigmoid . . . 15
2.6 Solving problem (16b) . . . 15
2.6.1 Proximal algorithm: Forward-backward algorithm with splitting . . . 15
2.6.2 Solving problem (16b) by approximation of the smoothness constraint . . . 16
2.6.3 Exact solving of problem (16b) . . . 18
2.7 Conclusion: the complete algorithm . . . 19

3 Numerical results **19**
3.1 Comparison with the state-of-the-art . . . 19
3.1.1 Variational approach with convex nonlinear function f . . . 20
3.1.2 Spike Triggered average (STA) . . . 20
3.2 Simulated spike train: the direct problem . . . 21
3.3 Reconstruction . . . 22

4 Conclusion **23**

Appendices **26**

A Bayesian approach **26**

B Study of the operator $\times$ **28**

C Around the Kurdyla-Lojasiewicz property **29**

D Solving problem (16a) for a piecewise linear sigmoid **32**
D.1 Case $\xi = 0$ . . . 32
D.2 Case $\xi > 0$ . . . 33

E Solving problem (16a) for a piecewise cubic sigmoid **35**
E.1 Computation of $y_1(x)$ . . . 36
E.2 Computation of $y_2(x)$ . . . 36
E.3 Expression of $prox_{\gamma f}(x)$ . . . 37

1 Introduction

This paper deals with the estimation of receptive fields of individual visual neurons (e.g., ganglion cells in the retina). Knowing the receptive field of a particular neuron allows to characterize the relation between stimulus (images) and the neuron response (action potentials, also called spikes). It is a first but crucial step to understand which region of the visual field and which stimuli a neuron is sensitive to. Thus, biologists always allow time for estimating receptive field in their experimental protocole.*

A classical view is that visual sensory neurons respond to specific local spatio-temporal patterns, and that their response can be characterized by a spatio-temporal convolution kernel called the receptive field, followed by a static nonlinearity and stochastic (Poisson-like) mechanisms of spikes generation. The static nonlinearity is in general non convex (e.g., sigmoid shape) since firing rate is bounded because neurons have a refractory period. Mathematically, this description corresponds to the so-called linear-nonlinear Poissonian (LNP) model [15, 40, 36, 32]. LNP models can simulate the spiking activity of ganglion cells (or cortical cells) in response to synthetic or natural images [13], and they voluntarily ignore the detailed neuronal mechanisms. These functional model are widely used by experimentalists to characterise the cells that they record, map their receptive field and characterise their spatiotemporal feature selectivities [15].

The LNP model is illustrated in Fig. 1: Given a visual stimulus $s : (x,t) \in \Omega \times [0,T] \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ is the spatial domain and $T > 0$ is the duration of the experiment, a LNP neuron generates a sequence of $n(T)$ spikes times $\{t_i\}_{1\leq i\leq n(T)}$ such that

$$\{t_i\}_{1\leq i\leq n(T)} \text{ is generated by a Poisson process of rate } r(t) = f\left((s \times u)(t)\right), \qquad (\mathcal{P})$$

where f is a nonlinear function and $u : \Omega_d := \Omega \times [-d, 0] \longrightarrow \mathbb{R}$ is the so-called receptive field which corresponds to the linear part of the processing, where $d > 0$ is the length of its temporal support: the response of the neuron at time t depends on the history of the stimulus up to time $t - d$ back in the past. The operator $\times$ is defined by

$$[s \times u](t) = \int_{-\infty}^{0} \int_{\Omega} s(x, t+\tau)u(x,\tau)dx\ d\tau. \qquad (1)$$

Assuming Hypothesis $(\mathcal{P})$, the problem of estimating the receptive field $u(x,t)$ is an inverse problem: given a stimulus $s(x,t)$ of duration T and the $n(T)$ spikes $\{t_i\}_{1\leq i\leq n(T)}$, recover the unknown receptive field $u(x,t)$.

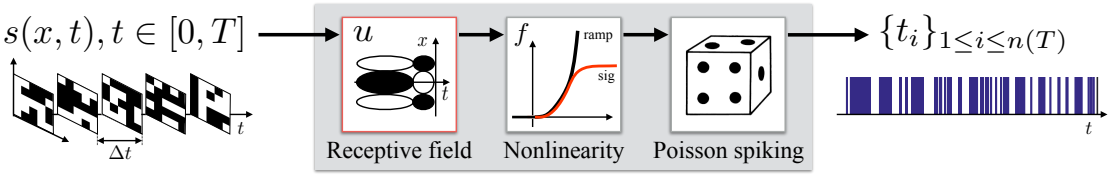


Figure 1: Illustration of the LNP model $(\mathcal{P})$ (gray box). Two kinds of nonlinear functions f are illustrated: ramp- or sigmoid-like nonlinearities. Here the stimulus is a sequence of white noise images, which is the classical stimulus used for the spike-triggered averaged approach (STA, see [15]).

*Experimentally, measuring retinal activity can be achieved by recording simultaneous light-evoked responses from hundreds of RGCs thanks to multielectrode arrays (MEA). One may use conventional 60-MEAs or, e.g., new generation of large-scale, high density MEA consisting of 4096 electrodes (APS CMOS-MEA, [5, 24]).

This inverse problem can be formulated using a Bayesian approach formulated in a discrete setting. Considering a discretization of the receptive field $u^N \in \mathbb{R}^N$, we search u^N as maximum likelihood or equivalently (see, e.g., [31, 35, 33]):

$$\begin{aligned}&\inf_{u^N} \big\{ -log\left(\rho(u^N|\{t_i\}_{1\leq i\leq n(T)})\right) \big\}\\ &= \inf_{u^N} \big\{ -log(\rho(\{t_i\}_{1\leq i\leq n(T)}|u^N)) - log(\rho(u^N))\big\},\end{aligned} \tag{2}$$

where $\rho(u^N|\{t_i\}_{1\leq i\leq n(T)})$ is the density probability of the random variable u^N given $\{t_i\}_{1\leq i\leq n(T)}$, $\rho(\{t_i\}_{1\leq i\leq n(T)}|u^N)$ is the density probability of the random variable $\{t_i\}_{1\leq i\leq n(T)}$ given u^N, and $\rho(u^N)$ is the density probability of the random variable u^N.

Then, assuming that $\rho(u^N)$ is a distribution of the form $\rho(u^N) = e^{-J(u^N)}$, and passing to the limit when $N \to \infty$ in (2) we get that the continuous problem is (see Appendix, Sect. A):

$$\inf_u \quad \mathcal{E}(u) = \psi(s \times u) + J(u), \tag{3}$$

where $\psi(s \times u)$ is the data fidelity term with ψ defined by

$$\psi(z) = \int_0^T f(z(\tau))d\tau - \sum_{i=1}^{n(T)} log(f(z(t_i))), \tag{4}$$

and $J(u)$ is the prior term to infer qualitative properties to the solution, according to what is known about the general shape of a receptive field.

To solve (3), it is classical to assume that the nonlinearity f is a ramp function or an exponential. In that case, one has to solve a convex problem which can be done by classical methods [31, 35, 26]. On the opposite, in this paper we consider the case when the nonlinearity f can be a sigmoid, which is more realistic from a physiological point of view: it models the fact that the neuron has a bounded firing rate (see [21], Sect.5.2.3). However, since f is nonconvex, the data fidelity term $\psi(s \times u)$ also becomes nonconvex.

The prior term $J(u)$ is chosen according to the properties we want to impose on the solution. The first property is that u should be localized in space and time since neurons are sensitive to a particular region of the visual field. This can be imposed by a sparsity constraint term. Here we choose to use a convex relaxation of the sparsity and we penalise the L^1–norm of u [12]. The second property is that u should be smooth. We propose that u belongs to the space BV_2 that contains piecewise linear functions. Note that in [6], this regularity constraint has been used for image restoration. It allows to recover functions with fast smooth variations more precisely than using a simple $\|\nabla u\|^2_{L^2}$–regularity constraint.

The space BV_2, is defined by

$$BV_2(\mathcal{O}) = \left\{ u \in W^{1,1}(\mathcal{O}),\ \frac{\partial u}{\partial x_i} \in BV(\mathcal{O}) \ \text{ for } i \in \{1, ..., N\} \right\},$$

where $\mathcal{O} \subset \mathbb{R}^N$ with $N \in \mathbb{N}^\star$ is a regular domain, $W^{1,1}(\mathcal{O})$ and $BV(\mathcal{O})$ stands respectively for the Sobolev space and the space of functions of bounded variations. On BV_2, the following norm can be defined:

$$|u|_{BV_2(\mathcal{O})} \equiv \int_{\mathcal{O}} |D^2 u| = sup \left\{ \int_{\mathcal{O}} < \nabla u, div(\varphi) >,\ \varphi \in C_0^1(\mathcal{O})^{N\times N},\ \|\varphi\|_{L^\infty(\mathcal{O})} \leq 1 \right\},$$

being the total variation of the gradient where $C_0^1(\mathcal{O})$ is the space of functions $C^1(\mathcal{O})$ with compact support in $\mathcal{O}$ [6, 17]. So the prior term will be defined by

$$J(u) = \lambda \|u\|_{L^1(\Omega_d)} + \mu |u|_{BV_2(\Omega_d)}, \tag{5}$$

where $\lambda, \mu > 0$ are weights associated respectively to the sparsity and to the regularity of u.

This paper is concerned with the study of problem (3) with ψ and J defined respectively by (4) and (5). In Sect. 2, we study the discrete version of problem (3). We introduce a relaxed problem which allows to compute an approximation of the solution. We make a theoretical study of the relaxed problem and propose and alternated minimizing algorithm converging toward a critical point of the relaxed energy. In Sect. 3, we test the approach on simulated data to provide a quantitative evaluation with comparisons to the classical spike triggered averaged technique (STA, see [15]) and we compare our approach with the classical variational approach using as non linearity a convex function close to the nonlinear sigmoid function describing the firing rate of a real neuron. Section 4 is a discussion and conclusion section. Let us also mention several appendices presenting more technical developments so that the paper be self-contained.

2 The discrete problem: well-posedness and relaxation

In this section we study the discrete problem. Section 2.1 gives some mathematical preliminaries. Section 2.3 is about the well-posedness. Section 2.4 presents the main algorithm which is an alternating minimization procedure followed by sections describing algorithms to compute solutions of each sub-problem.

2.1 Preliminaries

We first recall some classical concepts of variational analysis which will be useful in the sequel. Let $\phi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ a proper lower semicontinuous function. We denote as usual $\mathrm{dom}(\phi) = \{x \in \mathbb{R}^n,\ \phi(x) < +\infty\}$, $\mathrm{epi}(\phi) = \{(x, \lambda) \in \mathbb{R}^n,\ \phi(x) \geq \lambda\}$ and $\mathrm{graph}(\phi) = \{(x, \lambda) \in \mathbb{R}^n,\ \phi(x) = \lambda\}$ respectively the *domain*, the *epigraph* and the *graph* of ϕ.

2.1.1 Notions of subdifferentials

We recall that the *Fréchet subdifferential* of ϕ at $x \in \mathrm{dom}(f)$ is defined by (see [38], Definition 8.3)

$$\partial^F \phi(x) = \left\{ x^\star \in \mathbb{R}^n,\ \liminf_{\substack{y \to x \\ y \neq x}} \frac{1}{\|x - y\|} \left(\phi(y) - \phi(x) - \langle x^\star, y - x \rangle\right) \geq 0 \right\}.$$

The *limiting subdifferential* of ϕ at $x \in dom(\phi)$ is denoted and defined by

$$\widehat{\partial}\phi(x) = \{x^\star \in \mathbb{R}^n \ \ \exists x_n \to x,\ \phi(x_n) \to \phi(x), x_n^\star \in \partial^F \phi(x_n) \to x^\star\}.$$

Let us remark that

$$\partial^F \phi \subset \widehat{\partial}\phi,$$

where the first set is convex and closed while the second is closed (see [38], Theorem 8.6).

We say that ϕ is *convex* if its epigraph is convex, and *semiconvex* if there exists a constant $c \geq 0$, such that $\phi(z) + \frac{c}{2} z^2$ is convex. If ϕ is convex, for $x \in \mathrm{dom}(\phi)$ we have

$$\partial^F \phi(x) = \widehat{\partial}\phi(x) = \partial\phi(x) \stackrel{\text{def}}{:=} \{x^\star \in \mathbb{R}^n,\ \phi(x) - \phi(x^\star) \geq \langle x^\star, x - x^\star \rangle\ \forall x \in \mathbb{R}^n\}.$$

These equalities remains true for semiconvex functions (see [38], Example 8.8).
We say that x is a *critical point* of ϕ if and only $0 \in \partial\phi(x)$.
Finally, we say that $\phi : \mathbb{R}^n \to \mathbb{R}$ is *separable* if it can be written as $\phi(x) = \sum_{i=1}^{n} \phi_i(x_i)$ where ϕ_i are real functions.

2.1.2 Notion of proximal operator

The following mapping introduced by Moreau [28] is very useful in non-smooth optimisation, and is a generalization of the notion of projector. The *proximal operator* of ϕ is denoted and defined by:

$$prox_\phi(x) = \arg\min_{y\in\mathbb{R}^n} \left(\phi(y) + \frac{1}{2}\|x-y\|_2^2\right) \quad \forall x \in \mathbb{R}^n, \tag{6}$$

and it can be shown that

$$prox_\phi = \left(I + \widehat{\partial}\phi\right)^{-1}.$$

Remarks 2.1. *Some properties of the proximal operator.*

(i) *If $\phi + \frac{1}{2}(.-x)^2$ is convex for all x, then $prox_\phi$ is univalued.*

(ii) *If ϕ is not convex, the proximal operator $prox_\phi$ can be multivalued.*

(iii) *Let $\lambda > 0$, the function $\phi_\lambda : x \mapsto \inf_y\{\phi(y) + \frac{1}{2\lambda}\|x-y\|_2^2\}$ is called the Moreau envelope or Moreau-Yoshida regularization of ϕ. The Moreau envelope can be considered as a smooth approximation of ϕ. Note that $\phi_\lambda(x) \leq \phi(x) \forall x$.*

(iv) *Let us consider a shifted and scaled proper function $\tilde{\phi}(x) = \phi(ax+b)$, then its proximal operator is given by $prox_{\tilde{\phi}}(x) = (prox_{a^2\phi}(ax+b) - b)/a$.*

2.1.3 Notion of conjugate function

The *Legendre-Fenchel transform* of ϕ is defined by (see [38], p. 473)

$$\phi^\star(x) = \sup_y \ \langle x, y\rangle - \phi(y) \quad \forall x \in \mathbb{R}^n. \tag{7}$$

Function $\phi^\star$ is called the conjugate of function ϕ, and $\phi^{\star\star} = (\phi^\star)^\star$ is the biconjugate of function ϕ. For a subset $K \subset \mathbb{R}^n$ we define the *indicator function* i_K and the *support function* δ_K by

$$i_K(x) = \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{otherwise,} \end{cases} \quad \text{and} \quad \delta_K(x) = \sup_{x\in K} \langle x, y\rangle. \tag{8}$$

The link between δ_K and i_K is that they are mutually conjugate when K is convex and closed. Let us recall that all functions ϕ lower semicontinous, convex and one-homogeneous (i.e., $\phi(\lambda x) = \lambda\phi(x)$ for $\lambda > 0$) verify

$$\phi^\star(x) = \delta_{K_\phi}(x) \ \text{ with } \ K_\phi = \{x \in \mathbb{R}^n, \ \forall y \in \mathbb{R}^n \ \langle y, x\rangle \leq \phi(y)\}. \tag{9}$$

2.2 Problem definition

In this section, we study a discrete version of (3). The stimulus is a sequence of N_t images, each one presented during a period of Δt so that the duration of the experiment is $T = N_t\Delta t$. Each image is of size $N_x \times N_y$ pixels. Receptive field is of size $N_x \times N_y \times D$ where D is a fixed depth in time. We introduce the real vector spaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ so that:

$$u \in \mathcal{X} = \mathbb{R}^{N_x\times N_y\times D}, \quad s \in \mathcal{Y} = \mathbb{R}^{N_x\times N_y\times N_t}, \quad z \in \mathcal{Z} = \mathbb{R}^{N_t}, \tag{10}$$

endowed with the scalar product $\langle ., .\rangle$ and the associated norm $\|.\|_2 = \sqrt{\langle ., .\rangle}$. We denote by $\|.\|_1$ the l_1–norm.

Given these notations, if $\xi = (\xi_i)_{1\leq i\leq N_t}$ is the number of spikes per time step (in $[i\Delta t, (i+1)\Delta t[$) and if we assume that $\Delta t = 1$ without loss of generality (up to change f by $\Delta t f$), then the data fidelity term (4) can be rewritten as

$$\psi_\xi(z) = \sum_{i=1}^{N_t} f(z_i) - \sum_{i=1}^{n(T)} log(f(z_{k_i})),$$

where k_i is the index of the small interval containing the spiking time t_i, so that

$$\psi_\xi(z) = \sum_{i=1}^{N_t} \psi_{\xi_i}(z_i), \quad \text{with} \quad \psi_{\xi_i}(z_i) = f(z_i) - \xi_i log(f(z_i)). \tag{11}$$

The function $\psi_{\xi_i}(z)$ has a limited domain of definition denoted by $\mathcal{Z}^+_{\xi_i}$ equal to $\mathbb{R}$ if $\xi_i = 0$ (no spikes), and $\{f > 0\}$ if $\xi_i > 0$.

We impose the following hypotheses on the sigmoid function f:

Hypothesis 1. *f is $C^0(\mathbb{R})$ and is bounded from below.*

Hypothesis 2. (i) *f is definable in the log-exp structure (see Appendix C for definition and properties around these notions).*

(ii) *f is semi-convex, i.e., there exists $\rho > 0$ such that $f(z) + \frac{\rho}{2}z^2$ is convex.*

(iii) *$z \mapsto -log(f(z))$ is convex on the set $\{f > 0\}$.*

Remark 2.1. *Hypothesis* 2(i) *is required by Theorem 2.1 concerning the convergence of the algorithm (see Sect. 2.4). Note that the assumption to be definable in the log-exp structure is not very restrictive. For example, the nonlinearities generally chosen are $C^0(\mathbb{R})$ and analytic on $\mathbb{R}$ except at a finite number of points (see, e.g., [31]), so that they verify this hypothesis. Concerning Hypothesis* 2(ii)*, remark that it is verified by convex functions, so that all results coming next will hold in the convex case too.*

In practice, we will choose f as a non decreasing piecewise cubic sigmoid function varying on $[\theta_1, \theta_2]$, equals to zeros for $z < \theta_1$ and equals to a constant $c > 0$ on $[\theta_2, +\infty[$. Figure 2 shows the shape of the data fidelity term $\psi_\xi(.)$ for this kind of sigmoid. One can easily verify that the data fidelity term is nonconvex when $0 \leq \xi < c$.

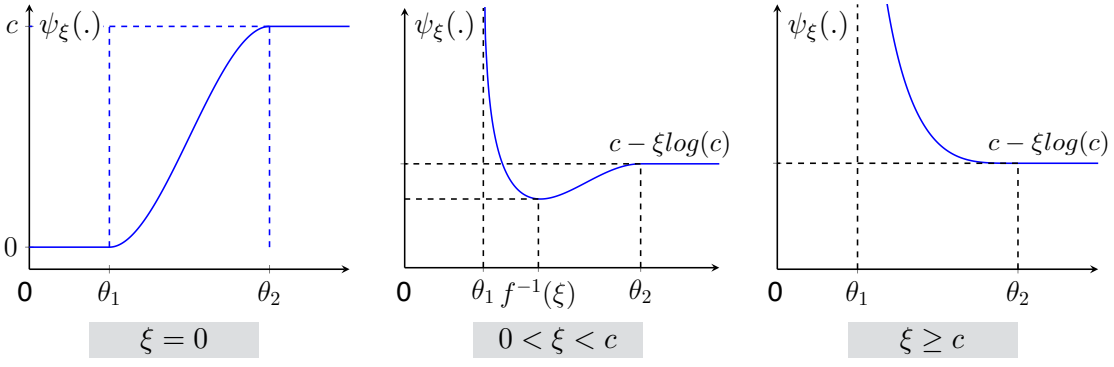


Figure 2: Graphs of the data fidelity term $\psi_\xi(.)$ for $\xi \in \mathbb{R}^+$, depending on values of ξ.

The prior term is defined by

$$J(u) = \lambda\|u\|_1 + \mu\|Hu\|_1, \tag{12}$$

where $H : \mathcal{X} \to \mathcal{X}^\nu$ is the Hessian matrix operator ($\nu = 9$ in 3D).

To discretize the Hessian operator, we use symmetric boundary conditions: $(Hu)_{i,j,k} = ((Hu)_{i,j,k})^{p,q}$ for $1 \leq p, q \leq 3$ with

$$(Hu)^{1,1}_{ijk} = \begin{cases} u_{i+1,j,k} - 2u_{i,j,k} + u_{i-1,j,k} & \text{if } 1 < i < N_x, \\ u_{i+1,j,k} - u_{i,j,k} & \text{if } i = 1, \\ -(u_{i,j,k} - u_{i-1,j,k}) & \text{if } i = N_x, \end{cases}$$

$$(Hu)^{1,2}_{ijk} = \begin{cases} u_{i,j+1,k} - u_{i-1,j+1,k} - u_{i,j,k} + u_{i-1,j,k} & \text{if } 1 < i < N_x \text{ and } 1 < j < N_y, \\ 0 & \text{if } i = 1, \\ 0 & \text{if } j = N_y, \end{cases}$$

Other derivatives can be obtained by permutation of the indices.

Denoting by $\times_d$ the discretization of $\times$ we get the discrete problem associated to (3):

$$\inf_{u\in\mathcal{X}} \mathcal{E}(u) = \psi_\xi(s \times_d u) + i_{\mathcal{Z}^+_\xi}(s \times_d u) + \lambda\|u\|_1 + \mu\|Hu\|_1, \tag{13}$$

where $i_{\mathcal{Z}^+_\xi}$ denotes the characteristic function associated to $\mathcal{Z}^+_\xi = \prod_{i=1}^{N_t} \mathcal{Z}^+_{\xi_i}$. In (13), the data fidelity term only depends on $s \times_d u$ and when minimizing it w.r.t. u, it is difficult to guarantee that $s \times_d u$ remains in the domain of definition of ψ_ξ. To overcome this problem, we propose to introduce an auxiliary variable $z \in \mathcal{Z}$ and solve the following relaxed formulation:

$$\inf_{z\in\mathcal{Z},u\in\mathcal{X}} \mathcal{E}_\alpha(z,u) = \psi_\xi(z) + i_{\mathcal{Z}^+_\xi}(z) + \frac{\alpha}{2}\|s \times_d u - z\|_2^2 + \lambda\|u\|_1 + \mu\|Hu\|_1, \tag{14}$$

where the term in α penalizes the difference between z and $s \times_d u$. Another interest of this relaxed problem is that now the problem in z containing the nonconvexity is separable (it leads to N_t one-dimensional independent problems). This relaxation is close to augmented Lagrangian for alternating direction method of multipliers (ADMM) algorithms [9] used in a convex context (this later just introduces another dual variable to guarantee the equality constraint at critical point and consists in two descents and one ascent of gradient). Let us notice that (14) gives an approximation of solutions of (13) when $\alpha \to \infty$ while ADMM algorithms with augmented Lagrangian give exact solutions for a convex energy $\mathcal{E}$. In this paper we consider the general case where $\mathcal{E}_\alpha$ is not convex and we will apply existing generic algorithms to compute critical points of $\mathcal{E}_\alpha$ which contain solutions of (14).

2.3 Well-posedness

This section is about existence and uniqueness of solutions of problems (13) and (14).

Proposition 2.1 (existence)**.** *Assuming Hypothesis 1, then*

1. *Let $\lambda > 0$, the energy $\mathcal{E}(u)$ (13) admits at least one minimizer.*
2. *Let $\alpha > 0$ and $\lambda > 0$, the energy $\mathcal{E}_\alpha(z,u)$ (14) admits at least one minimizer.*

Proof. We only prove that $\mathcal{E}_\alpha(z,u)$ admits a minimizer for $\alpha > 0$ and $\lambda > 0$, the existence of minimizer of $\mathcal{E}(u)$ being similar. Thanks to compactness of bounded sequences in $\mathbb{R}^n$ and lower semicontinuity of $\mathcal{E}_\alpha(z,u)$ ($\mathcal{E}_\alpha$ is even continuous), the existence question reduces to the sufficient question of coerciveness. Let $(z_n,u_n)_n$, a sequence such that $\mathcal{E}_\alpha(z_n,u_n)$ be bounded by a constant $M > 0$. As $\psi_\xi(z)$ is bounded from below, we deduce that $\|u_n\|_1$ and $\|z_n - s \times_d u_n\|$ are bounded. From that it is clear that z_n is also bounded. Thus by extracting subsequences and using the lower semicontinuity of $\mathcal{E}_\alpha$, we easily deduce that there exists $(\bar{z},\bar{u})$ such that $\mathcal{E}_\alpha(\bar{z},\bar{u}) = \inf \mathcal{E}_\alpha$. □

Proposition 2.2 (uniqueness)**.** *Assuming Hypothesis* 2 (ii)–(iii)*, if $\alpha > \rho$ then the energy $\mathcal{E}_\alpha(z,u)$* (14) *is strictly convex w.r.t. to z and convex w.r.t u separately (but not convex w.r.t. (z,u)). If moreover we assume that f is convex ($\rho = 0$ in Hypothesis* 2(ii)*), then* (13) *and* (14) *are convex and critical points are global minima.*

Proof. $u \mapsto \mathcal{E}_\alpha(z,u)$ is a sum of convex terms plus a quadratic one so it is strictly convex. Thanks to Hypothesis 2(ii), we easily deduce that $z_i \mapsto f(z_i) + \frac{\alpha}{2}(z_i - x)^2$ is strictly convex for each $x \in \mathbb{R}$ if $\alpha > \rho$. From the separability of $\mathcal{E}_\alpha(z,u)$ w.r.t z and Hypothesis 2(iii) we deduce that $z \mapsto \mathcal{E}_\alpha(z,u)$ is strictly convex. The first term $z \mapsto \Psi_\xi(z)$ is convex, the second term $(z,u) \mapsto \|s \times_d u - z\|_2^2$ also, and the third term $u \mapsto J(u)$ as well. Hence, we get that $\mathcal{E}_\alpha(z,u)$ is convex and we deduce that critical points are global minima. Indeed, we have

$$\mathcal{E}_\alpha(z,u) - \mathcal{E}_\alpha(\bar{z},\bar{u}) \geq \partial\mathcal{E}_\alpha(\bar{z},\bar{u}).(z-\bar{z},u-\bar{u}) = 0, \quad \forall\ (z,u) \in \mathcal{Z} \times \mathcal{X},$$

for each critical points $(\bar{z},\bar{u})$. □

Remark 2.2. *Even if we assume that f is strictly convex, the energy $\mathcal{E}_\alpha(z,u)$ is in general not strictly convex. The reason is that $u \mapsto \frac{\alpha}{2}\|z - s \times_d u\|_2^2 + \phi(Ku)$ with $K := I \times H : \mathcal{X} \to \mathcal{X} \times \mathcal{X}^\nu$ and $\phi(v) = \sum_i |v_{1i}| + |v_{2i}|$ is in general not strictly convex since ϕ is not strictly convex and $u \mapsto s \times_d u$ is in general not injective.*

The following proposition establishes the link between solutions of the relaxed problem and solutions of the initial one.

Proposition 2.3. *Let α be a positive integer and let $\{(z_\alpha,u_\alpha)\}_{\alpha\geq 1}$ be a family of solution of problem* (14) *then $\mathcal{E}_\alpha(z,u)$ is bounded independently of α and*

(i) *All its cluster points are couples $(s\times_d\bar{u},\bar{u})$ such that $\bar{u}$ is a solution of* (13)*.*

(ii) *The infimum converges:* $\inf_{(z,u)} \mathcal{E}_\alpha(z,u) = \mathcal{E}_\alpha(z_\alpha,u_\alpha) \underset{\alpha\to+\infty}{\longrightarrow} \inf_u \mathcal{E}(u)$*.*

(iii) *If $\mathcal{E}(u)$ admits a unique minimizer $\bar{u}$, then* $(z_\alpha,u_\alpha) \underset{\alpha\to+\infty}{\longrightarrow} (s \times_{\boldsymbol{d}} \bar{u},\bar{u})$*.*

Proof. Let be $\bar{u}$ a minimizer of the initial energy $\mathcal{E}(u)$ (13) and (z_α,u_α) a minimizer of the relaxed energy $\mathcal{E}_\alpha(z,u)$ (14). Let $\alpha_0 > 0$, and $\alpha \geq \alpha_0$, we have that

$$\mathcal{E}_{\alpha_0}(z_\alpha,u_\alpha) \leq \mathcal{E}_\alpha(z_\alpha,u_\alpha) \leq \mathcal{E}_\alpha(s\times_d\bar{u},\bar{u}) = \mathcal{E}(\bar{u}). \tag{15}$$

As in the proof of Prop. 2.1, we deduce that $\{(z_\alpha,u_\alpha)\}_{\alpha>0}$ is bounded independently of α and that it converges up to a subsequence to $(\tilde{z},\tilde{u}) \in \mathcal{Z}_\xi^+ \times \mathcal{X}$. From the definition of $\mathcal{E}_\alpha(z,u)$ and since $\psi_\xi(z)$ and $J(u)$ are bounded from below we deduce that $\|z_\alpha - s\times_d u_\alpha\| \underset{\alpha\to+\infty}{\longrightarrow} 0$ and that $\tilde{z} = s\times_d\tilde{u}$. Now let us prove that $\tilde{u}$ is a minimizer of $\mathcal{E}(u)$. It is easily seen that the quantity $m_\alpha = \mathcal{E}_\alpha(z_\alpha,u_\alpha)$ is non-decreasing and bounded from above by $\mathcal{E}(\bar{u})$ so it converges to its supremum. By uniqueness of the limit we have that $\lim_\alpha m_\alpha = \mathcal{E}(\tilde{u})$ and by passing to the limit into (15), we get that $\mathcal{E}(\tilde{u}) = \mathcal{E}(\bar{u})$. Hence $\tilde{u}$ is a minimiser of $\mathcal{E}(u)$. □

2.4 Main algorithm

To compute a solution of (14), we propose the proximal alternating minimization introduced in [1] in a more general context. Given an initial condition $u_0 \in \mathcal{X}$, the algorithm consists in the following two steps:

$$z^{(k+1)} \in \arg\min_z \ \mathcal{E}_\alpha(z, u^{(k)}) + \frac{1}{2\beta^{(k)}}\|z - z^{(k)}\|_2^2, \tag{16a}$$

$$u^{(k+1)} \in \arg\min_u \ \mathcal{E}_\alpha(z^{(k+1)}, u) + \frac{1}{2\gamma^{(k)}}\|u - u^{(k)}\|_2^2, \tag{16b}$$

where $\beta^{(k)}, \gamma^{(k)}$ are sequences of parameters belonging to $[r^-, r^+]$ with $0 < r^- < r^+$ for all $k \geq 0$ (note that this is the only condition on these parameters to obtain convergence). The quadratic terms in (16) are necessary to show the convergence and it can be shown that one minimization step leads to a decrease of the energy such that

$$\mathcal{E}_\alpha(z^{(k+1)}, u^{(k+1)}) + \frac{1}{2\beta^{(k)}}\|z^{(k+1)} - z^{(k)}\|_2^2 + \frac{1}{2\gamma^{(k)}}\|u^{(k+1)} - u^{(k)}\|_2^2 \leq \mathcal{E}_\alpha(z^{(k)}, u^{(k)}),$$

which gives an interpretation of the qualitative role of parameters $\beta^{(k)}, \gamma^{(k)}$ for convergence.

Thus, the algorithm (16) consists at each iteration to compute the proximal mapping (up to a multiplicative constant) of $z \mapsto \mathcal{E}_\alpha(z, u^{(k)})$ and of $u \mapsto \mathcal{E}_\alpha(z^{(k+1)}, u)$ successively. From a numerical point of view, this algorithm is easy to handle:

- If f is a piecewise cubic function verifying Hypotheses 1 and 2, when $\xi_i = 0$, then problem (16a) can be solved analytically thanks to its separability: it is equivalent to compute the proximal operator of f up to a multiplicative constant depending on α and $\beta^{(k)}$. When $\alpha + \frac{1}{\beta^{(k)}} > \rho$ the problem is strictly convex so that the proximal operator is univalued. Otherwise, in the general case set by Hypotheses 1 and 2, we use a Newton algorithm.
- Problem (16b) is strictly convex. It can be solved by regularising the non differentiable and non separable term $\|Hu\|_1$ and by using a standard forward-backward proximal algorithm of Nesterov type [2]. Another possibility is to rewrite the non differentiable term of (16b) by using the Fenchel transform (7) and to introduce a dual variable, inspired by [3].

Theorem 2.1. (convergence) *Assume Hypotheses* 1 *and* 2(i)*, Algorithm* (16a)–(16b) *generates a sequence* $(z^{(k)}, u^{(k)})_k$ *which converges to a critical point* $(\bar{z}, \bar{u})$ *of* $\mathcal{E}_\alpha(z, u)$ *(i.e.,* $0 \in \partial\mathcal{E}_\alpha(\bar{z}, \bar{u})$*) and the sequence* $(z^{(k)}, u^{(k)})$ *verifies*

$$\sum_{k=1}^{\infty} \|z^{(k)} - z^{(k-1)}\| + \|u^{(k)} - u^{(k-1)}\| < \infty.$$

Proof. To prove convergence, we apply Theorem 9 from [1]. The difficult point is to check that $\mathcal{E}_\alpha$ verifies the Kurdyka-Lojasiewicz (KL) property which is a notion introduced in 1963 by Lojasiewicz for analytical functions and extended by Kurdyka in 1998 to functions definable in an o-minimal structure. This is detailed in Appendix C. □

Remark 2.3. *The symbol* $\in$ *can be replaced by* $=$ *in* (16b)*, since the subproblem is strictly convex for* $\alpha > 0$*. If we furthermore assume Hypothesis 2(ii)-(iii) then we can do as well in* (16a)*. What is remarkable is that the sequence* $(z^{(k)}, u^{(k)})$ *converges to a critical point of* $\mathcal{E}_\alpha$ *even if each subproblem of* (16) *has a non unique solution, and from our knowledge this theorem is the only one that establishes the convergence of the whole sequence to a critical point, just assuming that the energy verifies the KL property (Definition C.1) and some classical assumptions to guarantee existence of minimizer(s).*

Corollary 2.1. *Assume that f verifies Hypothesis 1 and 2 and moreover that f is convex, then the sequence $(z^{(k)}, u^{(k)})_k$ converges to a global minimum of $\mathcal{E}_\alpha$.*

Proof. Under these assumptions and thanks to Prop. 2.2, we get that $\mathcal{E}_\alpha$ is convex so critical points are global minima which ends the proofs by applying Th. 2.1. □

2.5 Solving problem (16a)

2.5.1 Using proximal operator definition

We remark that $\mathcal{E}_\alpha(z, u)$ is separable w.r.t z. Hence minimizing it w.r.t z at fixed u is equivalent to minimize N_t one dimensional energies:

$$z_i^{(k+1)} = \arg\min_{z_i \in \mathbb{R}} \ \psi_{\xi_i}(z_i) + i_{\mathcal{Z}^+_{\xi_i}}(z_i) + \frac{\beta^{(k)}\alpha + 1}{2\beta^{(k)}} \left(z_i - \frac{\alpha\beta^{(k)}(s \times_d u^{(k)})_i + z_i^{(k)}}{\beta^{(k)}\alpha + 1} \right)^2,$$

where ξ_i is the number of spikes in the i-th time interval and ψ_{ξ_i} defined in (11) is the data fidelity term. We rewrite $z_i^{(k+1)}$ by using the proximal operator (6) associated to the i-th component of the data fidelity term:

$$z_i^{(k+1)} = prox_{h^{(k)}\psi_{\xi_i}} \left(\frac{\alpha\beta^{(k)}(s \times_d u^{(k)})_i + z_i^{(k)}}{\beta^{(k)}\alpha + 1} \right), \tag{17}$$

where $h^{(k)} = \frac{\beta^{(k)}}{\beta^{(k)}\alpha+1}$ and where we shorten the notation $\psi_{\xi_i} + i_{\mathcal{Z}^+_{\xi_i}}$ by ψ_{ξ_i}. The function ψ_{ξ_i} depends on the number of spikes ξ_i and is defined from the sigmoid function. Generally the sigmoid function is approximated by an exponential based function (see Hypothesis 2(i) and Fig. 3(a)). The problem is that finding an analytical expression of $z_i^{(k+1)}$ in that case is a hard task. As a consequence, we propose two approximations for which analytical expressions can be derived. The first approximation is defined by a piecewise linear approximation (see Fig. 3(b) and Appendix D) but in this case problem (16a) is nonconvex independently of α and $\beta^{(k)}$ (the proximal operator is multivalued). The second approximation is defined by a piecewise cubic sigmoid function (see Fig. 3(c) and Appendix E). It is the simplest choice verifying Hypotheses 1 and 2 which enables to compute $z_i^{(k+1)}$ (17) analytically when there is no spike ($\xi_i = 0$) and numerically otherwise. This fact is interesting because spikes are sparsely distributed.

In the two following sections we study (17) for $\xi_i = 0$ (no spike in the i-th interval) and $\xi_i > 0$ (at least one spike).

2.5.2 Case $\xi_i = 0$ (no spike)

In this case, (17) reduces to the computation of the proximal operator of γf with $\gamma = h^{(k)}$. Thanks to Remark 2.1-(iv), without loss of generality, we focus on a normalized sigmoid function. Considering different nonlinear functions f, Table 1 summarizes their properties, if they verify Hypotheses 1 and/or 2 and their proximal operator. We give results for the three nonconvex functions described in Fig. 3 but also for convex functions for which the same method applies (see numerical results in Sect. 3). Details about how to estimate proximal operators are given in Appendices D and E for piecewise linear and piecewise cubic sigmoid cases respectively.

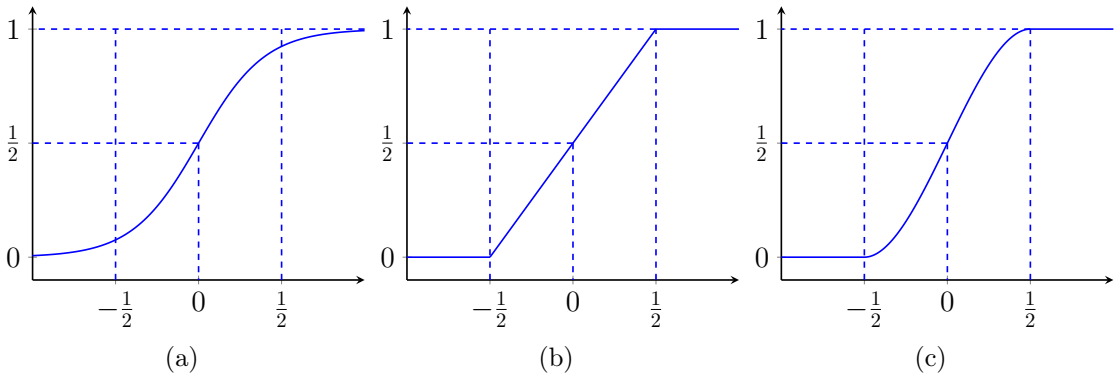


Figure 3: Sigmoid-like functions. (a) Exponential based function $\frac{1}{1+e^{-5x}}$. (b) Piecewise linear approximation (38). (c) Piecewise cubic approximation (41).

Nonlinear function $f(x)$	**Properties**	**Proximal operator** $prox_{\gamma f}(x)$
$\begin{cases} 0 & \text{if } x < -\frac{1}{2} \\ \frac{1}{2}+x & \text{if } \lvert x\rvert \le \frac{1}{2} \\ 1 & \text{if } x > \frac{1}{2} \end{cases}$	Hyp. 1 Hyp. 2(i)-(iii) nonconvex	$\begin{cases} x & \text{if } x \le -\frac{1}{2} \\ -\frac{1}{2} & \text{if } -\frac{1}{2} \le x \le \gamma \\ x-\gamma & \text{if } \frac{-1+2\gamma}{2} \le x \le \frac{1+\gamma}{2} \\ x & \text{if } x \ge \frac{1}{2} \end{cases}$
$\begin{cases} 0 & \text{if } x \le -\frac{1}{2} \\ \frac{1}{2}+\frac{3}{2}x-2x^3 & \text{if } \lvert x\rvert \le \frac{1}{2} \\ 1 & \text{if } x \ge \frac{1}{2} \end{cases}$	Hyp. 1 and 2 nonconvex	$\begin{cases} x & \text{if } x \le -\frac{1}{2} \\ y_-(x) & \text{if } \lvert x\rvert \le \frac{1}{2} \\ x & \text{if } x \ge \frac{1}{2} \end{cases}$ $y_-(x)$ given by (42) multivalued if $\gamma > \frac{1}{6}$
$\frac{1}{1+e^{-x}}$	Hyp. 1 and 2 nonconvex	No analytical expression Use a Newton algorithm
$\begin{cases} 0 & \text{if } x \le -\frac{1}{2} \\ \frac{1}{2}+x & \text{if } x \ge -\frac{1}{2} \end{cases}$	Hyp. 1 and 2 convex	$\begin{cases} x & \text{if } x \le -\frac{1}{2} \\ -\frac{1}{2} & \text{if } -\frac{1}{2} \le x \le \frac{-1+2\gamma}{2} \\ x-2\gamma & \text{if } x \ge \frac{-1+2\gamma}{2} \end{cases}$
$\begin{cases} 0 & \text{if } x \le -\frac{1}{2} \\ \frac{1}{2}+2x+2x^2 & \text{if } x \ge -\frac{1}{2} \end{cases}$	Hyp. 1 and 2 convex	$\begin{cases} x & \text{if } x \le -\frac{1}{2} \\ \frac{x-2\gamma}{4\gamma+1} & \text{if } x \ge -\frac{1}{2} \end{cases}$
$log(1+e^x)$	Hyp. 1 and 2 convex	No analytical expression Use a Newton algorithm

Table 1: Proximal operators of γf for different nonlinear functions f (see Sect. 2.5.2)

Nonlinear function $f(x)$	**Properties**	**Proximal operator** $prox_{\gamma\psi_\xi}(x)$
$\begin{cases} 0 & \text{if } x < -\frac{1}{2} \\ \frac{1}{2} + x & \text{if } \lvert x\rvert \leq \frac{1}{2} \\ 1 & \text{if } x > \frac{1}{2} \end{cases}$	Hyp. 1 Hyp. 2(i)-(iii) nonconvex	$\begin{cases} y_+(x) & \text{if } x \leq x_1^\star \\ \frac{1}{2} & \text{if } x_1^\star \leq x \leq \frac{1}{2} \\ x & \text{if } x \geq \frac{1}{2} \end{cases}$ $y_+(x)$ given by (40) $x_1^\star = \gamma(1-\xi) + \frac{1}{2}$ multivalued if $\xi < 1$
Non decreasing non negative and C^0, piecewise C^1	–	$\begin{cases} \in]-\frac{1}{2}, max(x, f^{(-1)}(\xi)] & \\ \quad \text{if } f^{(-1)}(\xi) \neq \emptyset & \\ x \quad \text{otherwise} & \end{cases}$ Use a Newton algorithm introducing a regularizing parameter[†] or use a Golden search algorithm.

Table 2: Proximal operators of $\gamma\psi_\xi$ (with $\psi_\xi(x) = f(x) - \xi log(f(x))$) for different nonlinear functions f (see Sect. 2.5.3).

2.5.3 Case $\xi_i > 0$ (presence of spikes)

In this case, (17) expands as

$$z_i^{(k+1)} = \arg\min_{y \in Z_{\xi_i}^+} \left\{ f(y) - \xi_i log(f(y)) + \frac{1}{2h^{(k)}} \left(y - a_{i,k}\right)^2 \right\}, \tag{18}$$

with $a_{i,k} = \frac{\alpha\beta^{(k)}(s \times_d u^{(k)})_i + z_i^{(k)}}{\beta^{(k)}\alpha + 1}$ and $h^{(k)} = \frac{\beta^{(k)}}{\alpha\beta^{(k)}+1}$. As stated in Prop. 2.2, assuming Hypothesis 2(i)-(ii), problem (18) is strictly convex for $\alpha > \rho - \frac{1}{\beta^{(k)}}$, hence it admits a unique solution that is the proximal operator of $h^{(k)}\psi_\xi$ evaluated at $a_{i,k}$. If we model the nonlinearity by a piecewise linear sigmoid function (which does not satisfy Hyp. 2(ii)), we get an analytical expression of the proximal operator of $\gamma\psi_\xi$ for some $\gamma > 0$ which is given in Table 2. Otherwise, to compute efficiently an approximation of $z_i^{(k+1)}$, either we use a Newton algorithm by regularizing the term $-log(f(y))$ by $-log(f(y) + \eta)$ with $\eta > 0$ small enough or we use a Golden search algorithm by using the fact that the solution belongs to $]\theta_1, z_{max}[$ with $z_{max} = max(a_{i,k}, f^{-1}(\xi))$ with the convention $f^{-1}(x) = \emptyset$ if x is not in the range of f as summed up in Table 2.

Remark 2.4. *Let us remark that if we know the proximal operator associated to* $\psi_\xi^f = f - \xi log(f)$ *for a given non linearity* f *we can deduce the one associated to* ψ_ξ^g *for whatever* $g(x) = cf(ax+b)$ *thanks to this relation*

$$\psi_\xi^g(x) = c\psi_{\frac{\xi}{c}}^f(ax+b) - \xi log(c),$$

and Remark 2.1(iv).

[†]In Newton algorithm, if f can be null, one can regularize $log(f)$ by changing it to $log(f + \eta)$ with $\eta > 0$ small enough, so that $\text{dom}(\psi_\xi) = \mathbb{R}$.

2.5.4 Conclusion on a good choice of sigmoid

By modelling the sigmoid function by a $C^1(\mathbb{R})$ piecewise cubic function, Hypotheses 1 and 2 are verified and problem (14) is convex for α large enough (Proposition 2.2). In this case we also get an analytical expression of the proximal operator associated to the data fidelity term (11) in the most frequent case which is the case of no spike has been emitted in the small i-th temporal bin ($\xi_i = 0$). It remains true for some other nonlinearities enumerated in Table 1. The choice of a piecewise linear sigmoid function enables to have an analytical expression of the proximal operator associated to the data fidelity term $\gamma\psi_\xi$ (11) for some $\gamma > 0$ and even when $\xi > 0$. Otherwise for other choices of nonlinearity we can get some information about the location of the solution (see Table 2).

2.6 Solving problem (16b)

Let us remind sightly rewrite problem (16b) as

$$\begin{aligned} u^{(k+1)} &= \arg\min_u \ \mathcal{E}_\alpha(z^{(k+1)}, u) + \frac{1}{2\gamma^{(k)}}\|u - u^{(k)}\|_2^2, \\ &= \arg\min_u \ \frac{\alpha'}{2}\|s\times_d u - z^{(k+1)}\|_2^2 + \frac{1}{2\gamma'^{(k)}}\|u - u^{(k)}\|_2^2 + \|u\|_1 + \mu'\|Hu\|_1, \end{aligned} \tag{19}$$

where parameters α', μ' and $\gamma'^{(k)}$ are normalized by λ. The difficulty in (19) comes from the prior term which the sum of a non differentiable separable term (the sparsity constraint, $\|u\|_1$), and non differentiable and non separable term (the smoothness constraint, $\mu'\|Hu\|_1$). Because of this non differentiability, one needs algorithms taken from non-smooth convex optimization. In this paper we use a proximal algorithm described in Sect. 2.6.1.

We give here two ways to compute an approximation of (16b). The first is based on a regularization of the smoothness term and needs to introduce a small positive parameter ε but it is more computationally fast (Sect. 2.6.2). The second relies on duality and is largely inspired from algorithm used in image processing [3, 14] (Sect. 2.6.3).

2.6.1 Proximal algorithm: Forward-backward algorithm with splitting

Proximal algorithms generalize descent algorithm with projection. For example, in the particular case of an indicator function of a convex set $\mathcal{C}$, the proximal operator reduces to the projector on the set $\mathcal{C}$ denoted $P_\mathcal{C}$. Generally, to solve $\min_\mathcal{C} g = \min(g_1 + i_\mathcal{C})$ where g_1 is a differentiable function and $i_\mathcal{C}$ the characteristic function associated to $\mathcal{C}$, we use the iteration $x^{(n+1)} = P_\mathcal{C}(x^{(n)} - \delta_n \nabla g_1(x^n))$. The idea of the proximal operator is to handle a more general problem by replacing the characteristic function by a general proper function g_2 and the projector $P_\mathcal{C}$ by the proximal mapping $prox_{\gamma g_2}$ with γ small enough. In [16] a collection of proximal splitting methods is presented.

Among this class of proximal algorithms, we focused on the so-called *forward-backward* algorithm which consists in minimizing a non differentiable convex and proper function g which decomposes as

$$g = g_1 + g_2, \tag{20}$$

with g_1 differentiable with a L-Lipschitz continuous gradient on its domain and g_2 a *simple*[‡] convex proper function. It can be decomposed in two steps: a forward (explicit) gradient step using the function g_1 and a backward (implicit) step using the function g_2. The forward-backward

[‡]We say that a function is simple if its proximal operator can be easily computed.

algorithm can be interpreted as a combination of the gradient method scheme $x^{(n+1)} = x^{(n)} - \delta_n \nabla g_1(x^{(n)})$ and of the proximal point iteration $x^{(n+1)} = prox_{\delta_n g_2}(x^{(n)})$.

Another point of view can be obtained by discretizing and searching stationary points of the dynamical system $\frac{dx}{dt} = -\left(\partial g_1(x) + \nabla g_2(x)\right)$. If we take an implicit scheme in ∂f_2 and an explicit one in ∇g_1 we obtain

$$x^{(n+1)} = x^{(n)} - \delta_n \Big(\partial g_1(x^{(n+1)}) + \nabla g_2(x^{(n)})\Big),$$

and from (6) this rewrites as

$$x^{(n+1)} = prox_{\delta_n g_2}(x^{(n)} - \delta_n \nabla g_1(x^{(n)})). \tag{21}$$

Algorithm (21) is known to converge if $g_1 \in C^{1,1}$ with gradient L-Lipschitz and $\delta_n \leq \frac{1}{L}$. After this short recalling on proximal algorithms, we present a variant introduced by Nesterov and revised by Beck and Teboule called Fast Iterate Shrinkage-Thresholding Algorithm (FISTA) [2]:

Algorithm 1 FISTA algorithm [2].

1. Initialization: $x_0 \in dom(g_1)$, $y^{(1)} = x_0$, $t_1 = 1$, $0 < h \leq \frac{1}{L}$
2. $x^{(n)} = prox_{hg_2}\left(y^{(n)} - h\nabla g_1(y^{(n)})\right)$
3. $t_{n+1} = \frac{t_n + \sqrt{t_n^2 + 4t_n^2}}{2}$
4. $y^{(n+1)} = x^{(n)} + \frac{t_n - 1}{t_{n+1}}(x^{(n)} - x^{(n-1)})$

Algorithm 1 will be used in the two subsequent sections. The interest of Algorithm 1 compared to the original original formulation (21) is convergence speed. Algorithm (21) is known to converge in terms of the objective function in $O(1/n)$ [29]. Nesterov showed that this speed can be improved and reach $O(1/n^2)$ by choosing an appropriate steps size $(\delta_n)_n$ and a weighted average on the previous iterate $x^{(n)}$ and the candidate given by (21). Besides Nesterov demonstrated that it was not possible to obtain a better convergence speed by using a first order algorithm (i.e., which only use the first order derivatives of the function). More precisely, if $x^\star$ is the minimum of g, given an initial data $x_0 \in dom(g_1)$ algorithms defined by Nesterov [29] ensure that

$$g(x^{(n)}) - g(x^\star) \leq \frac{L\|x^\star - x^0\|^2}{n^2}.$$

We refer the reader to [41] for a justification of the convergence of this algorithm by using dynamical system tools.

2.6.2 Solving problem (16b) by approximation of the smoothness constraint

As stated before, the non smooth part should be simple in the sense that its proximal operator is easily computable. On that point, since $\|Hu\|_1$ is not separable, it is difficult to compute its proximal operator while for the term $\|u\|_1$ it is easy to do. A solution is to introduce a parametrized regularization $\|Hu\|_{1,\varepsilon}$ with $\varepsilon > 0$ small enough such that

$$\|Hu\|_{1,\varepsilon} = \sum_{i,j,k} \sqrt{\varepsilon^2 + |(Hu)_{i,j,k}|^2}$$

Problem (19) is then replaced by

$$\min_{u \in \mathcal{X}} \left(g(u) := \frac{\alpha'}{2}\|s \times_d u - z\|_2^2 + \mu'\|Hu\|_{1,\varepsilon} + \|u\|_1 + \frac{1}{2\gamma'^{(k)}}\|u - u^{(k)}\|_2^2 \right). \tag{22}$$

To use Algorithm 1, let us split g as in (20), with

$$\begin{cases} g_1(u) = \dfrac{\alpha'}{2}\|s \times_d u - z^{(k+1)}\|_2^2 + \mu'\|Hu\|_{1,\varepsilon} + \dfrac{1}{2\gamma'^{(k)}}\|u - u^{(k)}\|_2^2, \\ g_2(u) = \|u\|_1, \end{cases} \tag{23}$$

where g_1 is differentiable on $\mathcal{X}$ and its gradient is given by

$$\nabla g_1(u) = \alpha' s \times_d^\star (s \times_d u - z^{(k+1)}) + \mu' H^T \left(\frac{Hu}{\sqrt{\varepsilon^2 + |Hu|^2}} \right) + \frac{1}{\gamma'^{(k)}}(u - u^{(k)}), \tag{24}$$

where $\times_d^\star$ is the adjoint operator of $\times_d$ (see Appendix B for definition and properties). From Lemma B.1 and Lemma B.2, we have $\|s \times_d .\|_2 \leq \|s\|_1$. One can easily verify that in 3D, $\|H\|_2 \leq 12$. Hence the Lipschitz constant of ∇g_1 is bounded by

$$L = \frac{144\mu'}{\varepsilon} + \alpha'\|s\|_1^2 + \frac{1}{\gamma'^{(k)}}. \tag{25}$$

The proximal function associated to γg_2 for some constant $\gamma > 0$ is defined by the element-wise soft threshold operator:

$$prox_{\gamma g_2}(u) = prox_{\gamma\|.\|_1}(u) = (prox_{\gamma|.|}(u_1), ..., prox_{\gamma|.|}(u_N)),$$

where N denotes the number of elements in u and with $prox_{\gamma|.|}$ defined by (see Fig. 4)

$$prox_{\gamma|.|}(x) = \begin{cases} x - \gamma & \text{if } x \geq \gamma, \\ x + \gamma & \text{if } x \leq -\gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

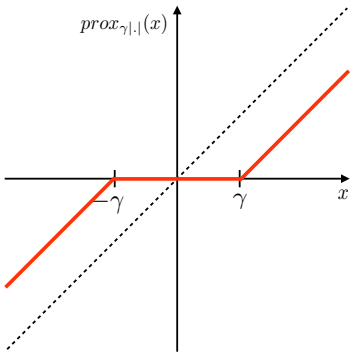


Figure 4: Graph of function $prox_{\gamma|.|}(x)$ defined by (26) which is a soft thresholding

Remark 2.5. *More generally, if we replace the regularizing term $\|Hu\|_1$ by, e.g., $\phi(Mu) = \sum_{i=1}^N \phi((Mu)_i)$ with ϕ a $C^{1,1}$ convex function whose its gradient is Lispschitz, and M a matrix, we can apply Algorithm 1 by putting this term into the regular part (g_1).*

2.6.3 Exact solving of problem (16b)

One problem of the previous approach is that it needs another parameter ε that cannot be fixed *a priori* before knowing the range of values of the unknown. To avoid this, we propose another standard approach in non-smooth optimization which consists in introducing a dual variable using the Fenchel transform (7).

Let us first rewrite problem (19) as

$$u^{(k+1)} = \arg\min_u \ \frac{1}{2}\langle A^{(k)}u, u\rangle - \langle u, b^{(k)}\rangle + \|Ku\|_{\mu'}, \tag{27}$$

with $K = I \times H$ mapping $\mathcal{X}$ into $\mathcal{X} \times \mathcal{X}^\nu$ with $\nu = 9$ for the Hessian operator in 3D ($\nu = 3$ for the gradient in 3D), $\|v\|_{\mu'} = \|v_1\|_1 + \mu'\|v_2\|_1$, $Au = \alpha' s \times_d^\star s \times_d u + \frac{u}{\gamma'^{(k)}}$ and $b = \alpha' s \times_d^\star z^{(k+1)} + \frac{u^{(k)}}{\gamma'^{(k)}}$. To shorten notations, we will skip the dependence on the parameter k in the sequel.

Lemma 2.1 (Bect et al. [3])**.** *If B is a symmetric positive matrix such that $\|B\|_2 < 1$ then we have*

$$\langle Bu, u\rangle = \inf_w \|u - w\|_2^2 + \langle Cw, w\rangle,$$

with $C = B(I - B)^{-1}$. Moreover, the maximum is reached at $w = (I + C)^{-1}u = (I - B)u$.

We set $\rho > 0$ such that $\rho\|A\|_2 < 1$. Thanks to Lemma 2.1 we get that (27) rewrites as

$$\inf_{u,w} \frac{1}{2\rho}\left(\|u - w\|_2^2 + \langle Cw, w\rangle\right) - \langle u, b\rangle + \|Ku\|_{\mu'},$$

where $C = (I - B)^{-1}B$ with $B = \rho A$. We define

$$F(u, w) = H(u, w) + \|Ku\|_{\mu'} \quad \text{with} \quad H(u, w) = \frac{1}{2\rho}\left(\|u - w\|_2^2 + \langle Cw, w\rangle\right) - \langle u, b\rangle.$$

Thanks to the convexity of $F(u, w)$, we have the relation

$$(u, w) \text{ minimizes } F(u, w) \Leftrightarrow \begin{cases} u \text{ minimizes } F(., w), \\ w \text{ minimizes } F(u, .). \end{cases}$$

Thanks to Lemma 2.1, $w_u = (I - B)u$ minimizes $F(u, .)$. Now let us minimize $F(w, .)$ at fixed $w \in \mathcal{X}$. We define the two following convex sets:

$$\mathcal{H}_K = \{K^T r,\ r \in \mathcal{H}\} \quad \text{with} \quad \mathcal{H} = \{(p, q) \in \mathcal{X} \times \mathcal{X}^\nu,\ |p_i| \le 1,\ |q_i| \le \mu'\}.$$

Lemma 2.2. *The problem*

$$\inf_u \frac{1}{2\rho}\|u - c\|_2^2 + \|Ku\|_{\mu'}, \tag{28}$$

has the unique solution $u = (I - \Pi_{\rho\mathcal{H}_K})(c)$ where $\Pi_{\mathcal{H}_K}$ is the orthogonal projection on the set $\mathcal{H}_K = \{K^T p,\ p \in \mathcal{H}\}$. Moreover let v be such that $K^T v = \frac{1}{\rho}\Pi_{\rho\mathcal{H}_K}(w)$, then v is solution of

$$\inf_{v \in \mathcal{H}} \|\rho K^T v - c\|_2^2. \tag{29}$$

Proof. By writing the dual problem of (28) and by using relation (9), we get:

$$\inf_u \frac{1}{2\rho}\|u - c\|_2^2 + \delta_{\mathcal{H}_K}(u),$$

where $\delta_{\mathcal{H}_K}$ is the support function given by (8). Interchanging sup and inf by using minimax theorem [19] (VI Prop. 2.1), we get the result. For more details see [14]. □

Thanks to Lemma 2.2, we deduce that $u_w = (I - \Pi_{\rho\mathcal{H}_K})(w + \rho b)$ minimizes $F(., w)$. Now we can derive the following algorithm to compute the solution of (19):

Proposition 2.4. *Problem* (27) *can be solved with the following iterative algorithm:*

$$c_n = (I - B)u_n + \rho b, \tag{30a}$$

$$v_n = arg\min_{v\in\mathcal{H}} \|c_n - \rho K^T v\|_2^2 = \Pi_{\rho\mathcal{H}_K}(c_n), \tag{30b}$$

$$u_{n+1} = c_n - \rho K^T v_n. \tag{30c}$$

Remark that to compute the projection $\Pi_{\mathcal{H}_{\rho K}}$ one needs to find the solution of (29) and for that we use Algorithm 1 with the expression of the projector $\Pi_{\mathcal{H}} : \mathcal{X} \times \mathcal{X}^\nu \to \mathcal{H}$ given by

$$(\Pi_{\mathcal{H}}((p, q)))_i = \left(\frac{p_i}{\max(1, |p_i|)}, \frac{\mu' q_i}{\max(\mu', |q_i|)}\right). \tag{31}$$

Theorem 2.2. *If ρ is such that*

$$\rho\|A^T A\|_2 < 1,$$

then all the sequence $(u_n)_n$ converge to the solution of (19).

Proof. Algorithm (30) can be rewritten as

$$u_{n+1} = (I - \Pi_{\rho\mathcal{H}_K})\left((I - B)u_n + \rho b\right), \tag{32}$$

which is a contraction on $\mathcal{X}$ (see [3] for more details). □

2.7 Conclusion: the complete algorithm

In this section we recaptulate the complete algorithm we use to solve problem (13). It is presented in Algo. 2. The block dedicated to the computation of $u^{(k+1)}$ can be changed using Sect. 2.6.2 (introducing another parameter $\epsilon > 0$ close to 0 to regularize the term $\|Hu\|_1$) but all results presented here are obtained from Algo. 2.

3 Numerical results

In this section we show results using a simulated spike train, that is obtained thanks to an artificial neuron following here the LNP model ($\mathcal{P}$). Section 3.1 presents two approaches from the state-of-the-art that we will use for comparison. Section 3.2 describes the experimental protocol and in particular how simulated spike train is generated. Section 3.3 gives results concerning both estimation quality and convergence.

3.1 Comparison with the state-of-the-art

In this section we remind two state-of-the-art methods used in Sect. 3.3 to make comparisons with our variational nonconvex approach (referred by VarNCvx in this section).

[§]With $g_1(v) = \|\rho K^T v - c\|_2^2$ and $g_2(v) = i_{\mathcal{H}}(v)$ (of proximal operator given by (31)).

Algorithm 2 Final algorithm.

Require: Parameters of the functional (14): α, λ, μ, f verifying Hypotheses 1 and 2(i).
Require: Parameters of the algorithm: N_{alt}, N_{nest}, N_{proj}, $(\gamma^{(k)})_{1\leq k\leq N_{alt}}$, $(\beta^{(k)})_{1\leq k\leq N_{alt}}$,
Require: Initialization: $u^{(0)} \in \mathcal{X}$
  for $k = 1 : N_{alt}$ **do**
    for $i = 1 : N_t$ **do**
      $z_i^{(k+1)}$ defined by (17) and computed thanks to Table 1 and 2.
    end for
    Set $u_0 = u^{(k)}$
    for $n = 1 : N_{nest}$ **do**
      Compute c_n (30a)
      Compute v_n (30b) with N_{proj} iterations of Algorithm 1[§]
      Compute u_n (30c)
    end for
    Set $u^{(k+1)} = u_{N_{nest}}$
  end for

3.1.1 Variational approach with convex nonlinear function f

As explained in the introduction which has motivated all this work, in state-of-the-art variational approaches, convex nonlinear functions f are used [31, 26, 35]. To show what is the interest of using a non convex function f, the same Algorithm (16a)–(16b) will be applied but with a convex function instead. This method will be referred to as VarCvx and based on the following function f_c (see Fig. 5(e)):

$$f_c(x) = \begin{cases} 0 & \text{if } x < -\frac{1}{2}, \\ \frac{1}{2} + 2x + 2x^2 & \text{otherwise.} \end{cases} \tag{33}$$

3.1.2 Spike Triggered average (STA)

The STA approach is a very classical method used in the neuroscience community to obtain a discrete approximation of the receptive field of sensory neurons [15].

In the case of a visual neuron stimulated by a visual stimulus s, STA estimator is defined by

$$\mathrm{STA}(x,\tau) = \frac{1}{n(T)} \sum_{i=1}^{n(T)} s(x, t_i - \tau),$$

where $\{t_i\}_{1\leq i\leq n(T)}$ is the sequence of $n(T)$ spikes times generated by the neuron.

A common visual stimulus used in experiments is a sequence of white noise images (see Fig. 5(a) and also Fig. 1 for illustrations), i.e., images whose the power spectrum is constant. That particular case is interesting since it can be shown that the STA estimator at time $n(T)$ fixed converges in law to $\mathcal{N}(u, \frac{\sigma^2}{n(T)})$ when $n(T)$ is large and where u is the receptive field to recover. Even if the process describing the stimulus does not generate exactly a white noise but only an independent discrete uniform binary random signal, by an heuristic way, thanks to the central limit and Bussgang's theorems, convergence remains true when $n(T) \to +\infty$.

Remark 3.1. *An important parameter of the STA is the spatial resolution of the estimation, which corresponds to the size of the blocks of the white noise stimulus images. In general, it is experimentally fixed to reach a compromise to have a neuron sufficiently responding: indeed, for*

STA to be precise, one needs a sufficient number of spikes, which can be obtained both by increasing the experiment duration (still limited by physiological constrains) or choosing an "optimal" block size: If too small, neuronal responses might be too weak since receptive field would have to integrate too small details, thus leading to receptive field with low Signal-to-Noise-Ratio; on the contrary, if too large, smallest receptive field will be lost, or not described with enough details, because of the coarse approximation.

3.2 Simulated spike train: the direct problem

In this section we describe how simulated spike train is generated, following the direct problem described in Fig. 1 and 5 and assuming that we consider classical ON type retinal ganglion cell which can be approximated by a LNP model.¶

Stimulus We choose as visual stimulus a sequence of N_t binary‖ white noise images of size 20×20 pixels with block size 4×4 pixels, thus allowing direct comparisons with STA in the sequel. One sample image is shown in Fig. 5(a). Each images is fed as input to the neuron during a time period of length Δt. Following notations of Sect. 2, up to changing the maximum of the sigmoid c by $c\Delta t$, we choose $\Delta t = 1$.

Receptive field of the artificial neuron The linear part of the response of classical retinal ganglion cells is generally modeled by a nonseparable spatio-temporal filtering (see, e.g., [44]). A common approximation is to consider that it is separable in space and time. We make this assumption here. Note that this is only to define our ground truth more simply, but of course this is not an hypothesis needed for our approach to work.

We assume that the receptive field u of the ON retinal ganglion cell is separable in space and time, so that

$$u(x,t) = v(x)w(t).$$

The spatial part can be described by a difference of Gaussian functions (DoG) (Fig. 5(b)–(c)):

$$v(x) = \omega_C G_{\sigma_C}(x) - \omega_S G_{\sigma_S}(x),$$

where

$$G_\sigma(x) = e^{-\frac{|x|^2}{2\sigma^2}}.$$

This center-surround behavior which has been found in the 60's [39, 20] corresponds to a measure of the local contrast. This DoG is driven by four parameters which can be all interpreted functionally. σ_C is a measure of the blur applied to the image hitting the retina originating in particular from the sampling frequency of photoreceptors. In some sense it defines the precision of the retina. Ratio σ_S/σ_C defines the relative surround. ω_C is a linear gain which gives the orders of magnitude for retina amplification. Ratio ω_S/ω_C defines the relative surround weight. In our simulations, we chose $\sigma_C = 2.2$, $\sigma_S = 3$, $\omega_C = 100$ and $\omega_S/\omega_C = 0.9$.

The temporal part can be described by a difference of Exponential functions (DoE) (Fig. 5(d)):

$$w(t) = \omega_A E(t, n_A, \tau_A) - \omega_B E(t, n_B, \tau_B), \quad \text{with } \tau_A < \tau_B,$$

¶Remark that primate retina has around 20 types of ganglion cells which interrogates any point in visual space about a number of distinct qualities [43, 22, 25]. This is achieved via parallel complex retinal circuitry which cannot be all approximated accurately by LNP models.

‖We choose as convention that s takes values in $\{-1, 1\}$.

where

$$E(t,n,\tau) = (nt)^n \frac{exp(-\frac{nt}{\tau})}{(n-1)!\tau^{n+1}}.$$

This band-pass temporal filter has a strong functional similarity with the DoG model. The first part $E(t,n_A,\tau_A)$ corresponds to the low-pass properties while the second part $E(t,n_B,\tau_B)$ represents the delayed inhibition that makes the response transient.** Ratio ω_B/ω_A defines the strength of the transient. In our simulations we chose $\omega_A = \omega_B = 1$, $\tau_A = 5$, $\tau_B = 7$, $n_A = 5$, $n_B = 7$

Receptive field is then discretized as a spatio-temporal volume of size $20 \times 20 \times 30$ pixels. It defines our ground truth (GT). Some temporal slices are shown in Fig. 6 (first row). So, given u, one can already compute the linear response $s \times_d u$.

Nonlinearity of the artificial neuron Given $s \times_d u$, one has to set the nonlinearity f to compute the rate of the Poisson process generating spikes ($\mathcal{P}$). To do so, we choose to take $f(x) = cf_0(ax+b)$ where f_0 is the piecewise cubic function (41), so that the distribution of $s \times_d u$ values mainly falls outside saturation regime, so that neuron response present sufficient variability to recover receptive field (see Fig. 5(e) for illustration). This is done empirically. Parameters are $a = 0.167$, $b = 0.1$, $c = 0.8$.

Simulated spike train Since we use as temporal discretization step the inter-frame Δt, we simulate $\xi = (\xi_i)_{1\leq i\leq N_t}$, the vector associated to the number of spikes occurring during each bin $[(i-1)\Delta t, i\Delta t[$. From Lemma A.1, ξ follows a Poisson distribution of parameter $f(s \times_d u)$. Figure 1 was giving an example of simulated spike train with bars width proportional to the number of spikes by bin. Here the number of spikes is around 500.

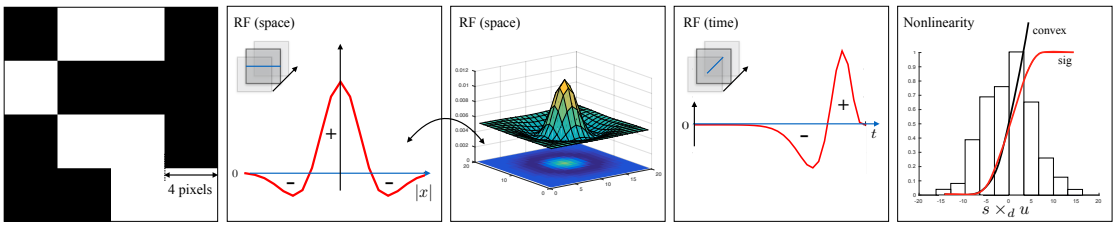


Figure 5: How simulated spike trains are generated (see Sect. 3.2)? (a) One image of the stimulus consisting in a sequence of images with white noise. This kind of stimulus is commonly used for STA estimator, and it is also used here in our approach for comparison. (b) and (c) Spatial profile of the receptive field used to generate the simulated spike train. (d) Temporal profile of the receptive field. (e) Nonlinearities super-imposed on the histogram of $s \times_d u$ for two cases: there is the non convex one which is used to generate the spike train and for our approach, and the convex one used for comparion (Sect. 3.1.1).

3.3 Reconstruction

This section presents results obtained with our approach, using Algorithm 2 described in Sect. 2.7. Parameters of the approach were chosen as $\alpha = 1000$, $\gamma^{(k)} = \beta^{(k)} = 10$, $N_{alt} = 300$, $N_{nest} = 40$,

**In response to an input step function, DoE temporal filtering part makes induces a peal of activity, and then a decrease of activity as the inhibition (of time scale τ_B) builds up and compete with the direct signal (of time scale τ_A).

$N_{proj} = 20$ and $u_0 = 0$.

Figure 6, shows comparisons of receptive field estimation with different approaches. Second row shows STA estimation. Result is noisy and resolution is constrained by the block size of the stimulus (by definition). Remark that the block size of the stimulus should depend on the receptive field to reconstruct (in general, it cannot be of size 1 pixel) because of the operator $\times_d$ which sums spatially on the stimuli. Hence if the support of the receptive field is composed of several pixels, the linear response $s \times_d u$ leads to a gray value constant in time, so that receptive field cannot be recovered (see also Remark 3.1). Third row shows a linear interpolation of receptive field to have a 20×20 pixels spatial resolution. This is given to compare with results from variational approaches which have that resolution. Fourth and fifth rows show results using the variational approach, with the convex nonlinearity f_c (VarCvx) and the nonconvex nonlinearity f (VarNCvx) respectively. Qualitatively, variational approach enables to reconstruct very accurately the receptive field compared to STA. Note that significant qualitative differences between VarCvx and VarNCvx appear. In particular, VarCvx does not enable to reconstruct the surrond of the receptive field while VarNCvx does (see Fig. 8 for a more quantitative comparison).

Figure 7 shows the influence of weights λ and μ defining the prior in energy (14). Qualitatively, as expected, the more the sparsity weight λ is large the more the receptive field is localized, and the more the smoothness constraint is large, the more the receptive field is smooth. In both cases, we find one optimal value of parameters to reach the lowest covariance error.

Figure 8 compares the temporal profiles of receptive field obtained with VarCvx and VarNCvx. Results show that VarCvx case is always biased while VarNCvx converges to the ground truth when the duration of the experiment increases (equivalently when the number of spikes increases).

Figure 9(a)–(b) compare the rate of convergence w.r.t the number of spikes observed. Our approach converges faster in the sense of both the covariance error $\mathcal{E}_{cov}(u, u_{\mathrm{GT}})$ and the l_2-norm $\mathcal{E}_{l_2}(u, u_{\mathrm{GT}})$. The three methods converge in the sense of the covariance error but we see that in the sense of the l_2 error the variational approach with the convex approach does not converge. Another observation is that error decays linearly (in log scale) for both the STA approach and the nonconvex variational approach with a slope of the convergence equal for both methods which can be explained by the central limit Theorem.

Figure 9(c) illustrates the speed of convergence of the algorithm: we compare the error decaying for our approach VarNCvx and for VarCvx, using a convex approximation f_c (33) as usually done in the literature.

Finally, Fig. 10 shows the execution time in seconds of Algorithm 2 implemented in Matlab and run on a GPU Quadro K4000.

4 Conclusion

Following the classical model of visual sensory neuron that is the linear-nonlinear Poissonian (LNP) model, we focused on the case when the nonlinearity is a sigmoid which is more realistic from a physiological point of view. This work presents for the first time, up to our knowledge, an approach to solve the nonconvex variational formulation for receptive fields estimation. Compared to the state-of-the-art, results on synthetic data are very promising. Our approach allows to estimate receptive field with great accuracy and it converges faster w.r.t. stimulus duration, which makes it a good candidate for experimentalists who need to estimate receptive fields in a limited amount of time[††].

[††]Let us emphasize again that estimating receptive fields is an important step done in almost every recording session. However, it is generally not the main goal of the experimentalist who want to investigate other properties of cells. Thus, because of physiological constraints such as the limited life duration of the tissue, keeping the receptive field estimation step short will let more time to experimentalists for the main purpose of their main

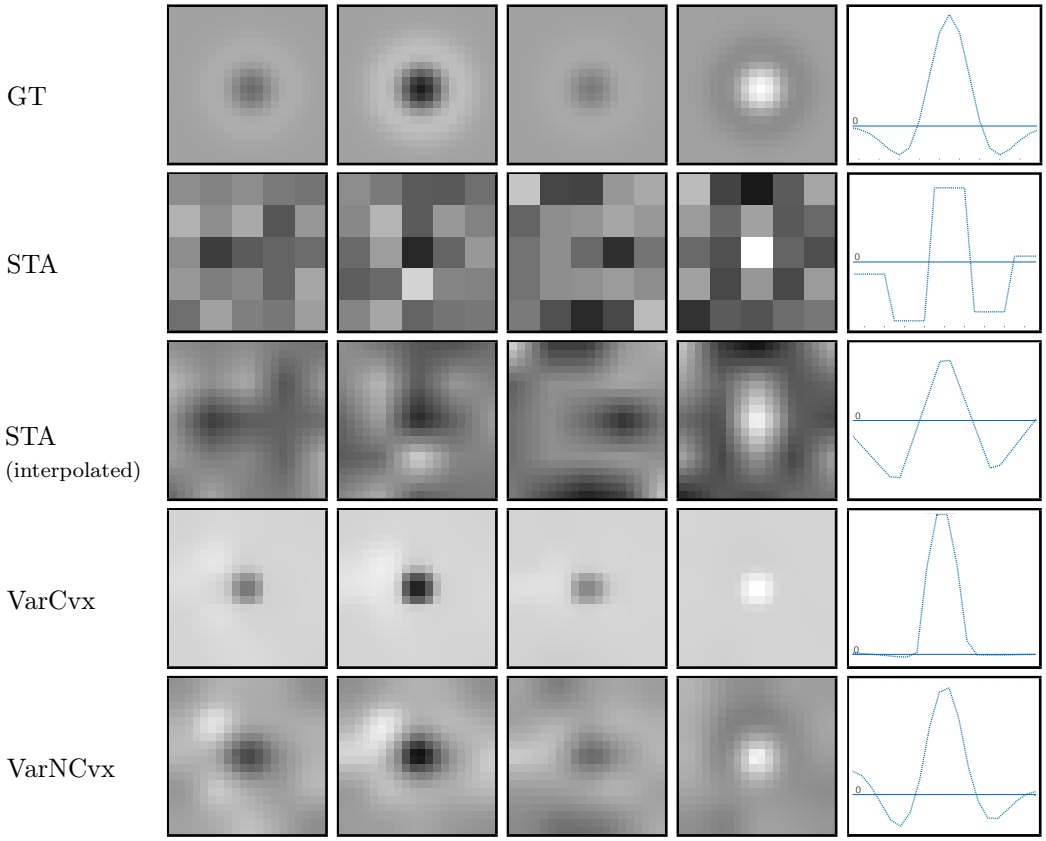


Figure 6: Qualitative comparison of receptive field reconstructions with respect to ground truth (first row, GT). Columns 1–4 show a selection of temporal slices and column 5 shows a horizontal cut passing through the center of receptive field(at temporal depth represented in column 4). Rows 2 and 3 show STA results. Rows 4 and 5 show results obtained using the variational approach, with the convex nonlinearity f_c (VarCvx) and the nonconvex nonlinearity f (VarNCvx) respectively (for $\lambda = 9$ and $\mu = 11$).

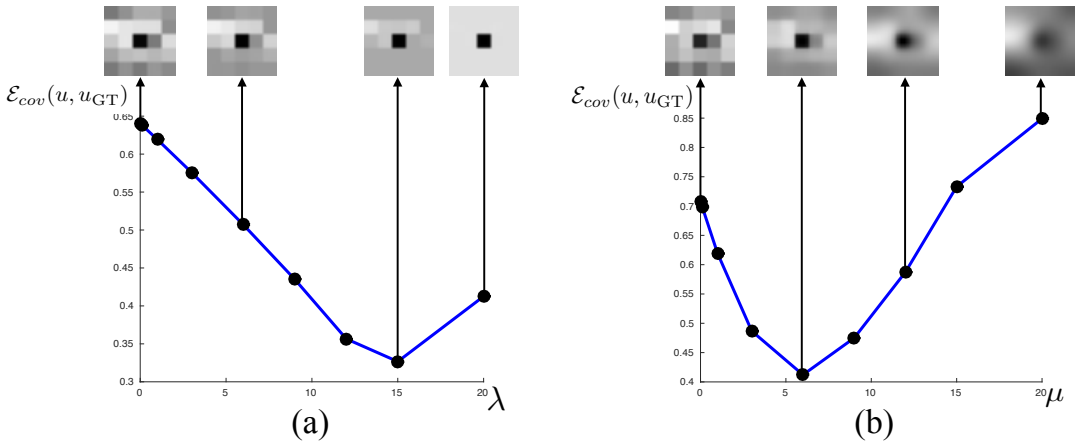


Figure 7: Quantitative influence of prior term parameters on receptive field estimation, using VarNCvx. We show covariance error ($\mathcal{E}_{cov}(u, u_{\mathrm{GT}}) = 1 - cov(u, u_{\mathrm{GT}})/\sigma_u \sigma_{u_{\mathrm{GT}}}$) as a function of (a) the sparsity constraint parameter λ (for $\mu = 1$ fixed) and (b) the regularity constraint parameter μ (for $\lambda = 1$ fixed).

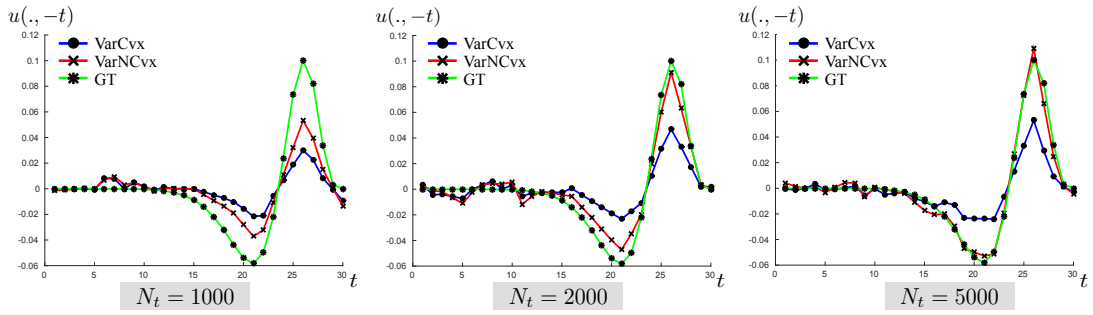


Figure 8: Comparison of the temporal profile of the receptive field estimates as function of the duration of the experiment N_t. Results show that VarNCvx gives a non-biaised estimated better than VarCvx.

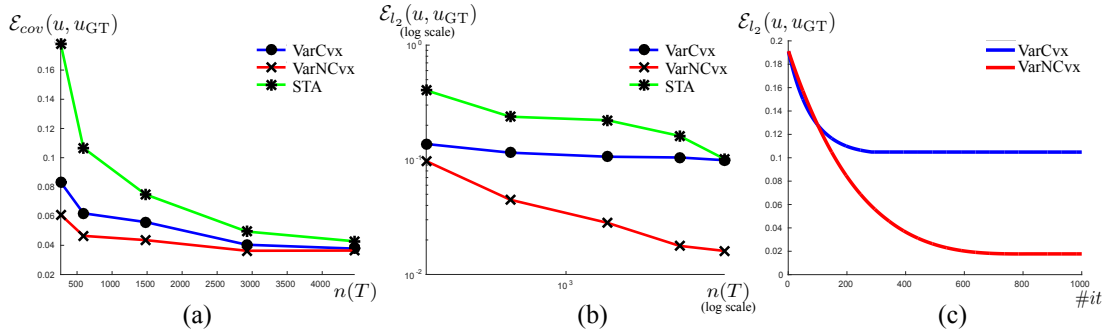


Figure 9: Study of the convergence of the VarNCvx approach. (a) Covariance error ($\mathcal{E}_{cov}(u, u_{\mathrm{GT}})$) as a function of the number of spikes for the convex and non convex variational approaches and the STA. (b) l_2-norm error ($\mathcal{E}_{l_2}(u, u_{\mathrm{GT}}) = \|u - u_{\mathrm{GT}}\|_2$) as a function of the number of spikes in log-log scale. (c) l_2-norm error as a function of the number of iterations for $N_t = 1000$.

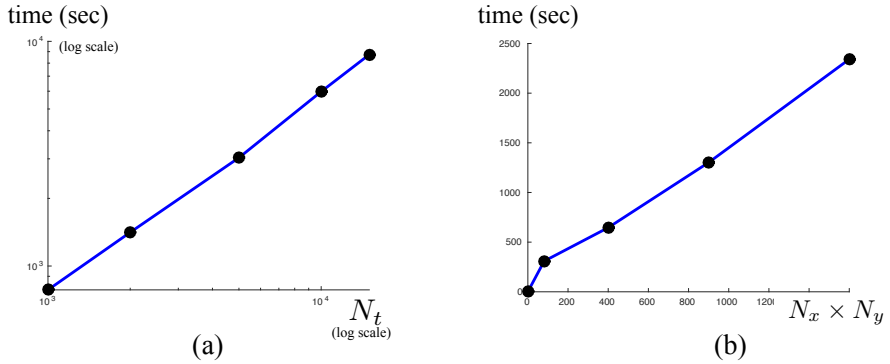


Figure 10: Execution time of Algorithm 2 implemented in Matlab on GPU Quadro K4000 depending on two parameters: (a) Influence of the temporal dimension of the stimulus N_t: we consider here a 1D receptive field, i.e., $N_x = N_y = 1$ and $D = 30$. (b) Influence of the spatial resolution $N_x \times N_y$ of the receptive field (and stimulus), with $D = 30$ and $N_t = 1000$.

Future work will focus on validating our approach on real cell recordings. Our plan is to study recordings coming from a new generation of large-scale, high density MEA consisting of 4096 electrodes (APS CMOS-MEA, [5, 24]. The main interest of this technology will be to provide recordings of thousands of cells, allowing to benchmark our approach w.r.t. the state of the art based on a large population of cells.

This work also opens new avenues to further improve it. Our approach works for any kind of stimulus, contrarily to classical approaches such as STA which assume that stimulus should be sequences of white noise images. So this allows to explore and optimize stimulus itself to improve reconstructions and convergence (see, e.g., [30, 23] for an example of alternative stimulus used to increase resolution). Here, we focused on LNP model, however, it is well known that some sub-population of cells cannot be modelled with such model (e.g., ON-OFF cells which respond to both onset and offset of a stimulus). So it will be interesting to investigate how our formulation can be extended using models generalizing LNP, e.g., by adding a non-linear feedback on spikes [37], adding a suppresive LNP cascade [11, 27]. Finally, when applying our approach to large populations of cells, one may need to have faster execution time which should be possible thanks to stochastic algorithm with asychronous parallelization [34].

Acknowledgements We thank Daniela Pamplona (Inria, Biovision team) as well as Evelyne Sernagor and Gerrit Hilgen (Institute of Neuroscience, Newcastle University, UK) and Matthias Hennig (Institute for Adaptive and Neural Computation, University of Edinburgh) for their helpful insights.

Appendices

A Bayesian approach

This section justifies the formulation (3) derived from $(\mathcal{P})$ using a Bayesian approach. Before giving the proof, some definitions and properties are reminded.

Definition A.1. *Let $\lambda(t) \in C^0(\mathbb{R}^+, \mathbb{R}^+)$. A process that produces random points in time is a Poisson random process of rate $\lambda(t)$ if the counting process $n(t)$ (number of events at t) satisfies the following properties $\forall t \geq 0$:*

$$P(n(t+dt) - n(t) = 1) = \lambda(t)dt + o(dt), \tag{34a}$$

$$\lim_{dt \to 0} \frac{P\left(n(t+dt) - n(t) > 1\right)}{dt} = 0. \tag{34b}$$

The arrival time $t^{(i)}$ for $i \geq 1$ are defined by

$$t^{(i)} = \inf\{t \in \mathbb{R}^+, n(t) \geq i\}.$$

Lemma A.1. *Let $n(t)$ be a Poisson random process of rate $\lambda(t)$, $t_0 > 0$ and $t \geq t_0$, the probability that n events have occurred from time t_0 to time t is*

$$P(n(t) - n(t_0) = n) = \frac{\Lambda(t_0, t)^n}{n!} e^{-\Lambda(t_0, t)},$$

where

$$\Lambda(t_0, t) = \int_{t_0}^{t} \lambda(s)ds.$$

study.

Proof. Let dt small enough, we have by denoting $g_k(t) = P(n(t) - n(t_0) = k)$:

$$\begin{aligned} g_n(t) &= \sum_{k=0}^{n} g_{n-k}(t) P(n(t+dt) - n(t) = k) \\ &= g_n(t) P(n(t+dt) - n(t) = 0) + g_{n-1}(t) P(n(t+dt) - n(t) = 1) + R_n(t, dt) \\ &= g_n(t)(1 - \lambda(t)dt + o(dt)) + g_{n-1}(t)\left(\lambda(t)dt + o(dt)\right) + R_n(t, dt), \end{aligned} \tag{35}$$

with

$$R_n(t, dt) = \sum_{k=2}^{n} g_{n-k}(t) P(n(t+dt) - n(t) = k) = o(dt).$$

By passing to the limit in (35) when $dt \to 0$ we get

$$\frac{dg_n}{dt} + \lambda g_n = \lambda g_{n-1}. \tag{36}$$

By using that $g_0(0) = 1$, the solution of this differential equation is

$$g_n(t) = \frac{\Lambda(t_0, t)^n}{n!} e^{-\Lambda(t_0, t)},$$

with $\Lambda(t_0, t) = \int_{t_0}^{t} \lambda(s) ds$.

□

Proof of formula (3). Let us first define a discretized approximation of $u(x, t)$. To do so, we split $\Omega \times [0, T[$ in N small voxels and we denote u_i^N the constant approximation of $u(x, t)$ in the i-th voxel. Similarly, we split $[0, T[$ in N_t small intervals of length $\Delta t = \frac{T}{N_t}$ and we denote by $\xi^{N_t}(t)$ the function equal to $\xi_i^{N_t}$ for t in $[(i-1)\Delta t, i\Delta t[$, where $\xi_i^{N_t}$ is the number of spikes in that interval. We have that

$$\begin{aligned} \xi^{N_t}(t) &\underset{N_t \to \infty}{\longrightarrow} \sum_{i=1}^{n(T)} \mathbb{1}_{t=t_i}, \\ u^N(x, t) &\underset{N \to \infty}{\longrightarrow} u(x, t), \end{aligned}$$

in the simple sense. Now we denote by $\xi_i^{N_t}$ the value of $\xi^{N_t}(t)$ when t belongs to the i-th bin and by u_i^N the value of $u^N(x, t)$ on the i-th voxel. We denote by u^N, s^N λ^{N_t} and ξ^{N_t} the vectors of component respectively u_i^N, s_i^N, $\lambda_i^{N_t} = f([s^N \times_d u^N]_i)$ and $\xi_i^{N_t}$. From Lemma A.1 and independence of $\xi_i^{N_t}$, we have

$$P(\xi^{N_t} | u^N) = \prod_{i=1}^{N_t} P(\xi_i^{N_t} | u^N) = e^{-\Delta t \sum_k \lambda_k^{N_t}} \prod_{i=1}^{N_t} \frac{(\lambda_i^{N_t} \Delta t)^{\xi_i^{N_t}}}{\xi_i^{N_t}!}.$$

Let be N_t tending to infinity, we get that

$$\rho(\{t_i\}_{1 \le i \le n(T)} | u^N) = e^{-\int_0^T \lambda^N} \prod_{i=1}^{n(T)} \lambda^N(t_i),$$

where $\lambda^N(t) = f([s \times u^N](t))$. Let be $N \to \infty$ we get the well known formula [31]. Passing to the $-log$ we get

$$-log(\rho(\{t_i\}_{1 \le i \le n(T)} | u)) = \int_0^T \lambda(s) ds - \sum_{i=1}^{n(T)} \lambda(t_i). \tag{37}$$

□

B Study of the operator $\times$

In this appendix we study properties of the operator $\times$ in the continuous case. Note that same estimations can be obtained in discrete case. To simplify notations, up to extend the temporal functions by 0 outside their support, we assume that the temporal support is $\mathbb{R}$ and we denote by $\Omega \subset \mathbb{R}^n$ the spatial domain. Notations which are used here are specific to this appendix (f denotes whatever function and is not linked the nonlinearity used in our model ($\mathcal{P}$)). The following lemma is an adaptation of the Young inequality proof (see [10], Th. 4.15) to the infinite dimensional case (here the state space is $L^q(\Omega)$ for some $q \geq 1$).

Lemma B.1. *Let be $s \in L^1(\mathbb{R}, L^q(\Omega))$ and $u \in L^1(\mathbb{R}, L^{q'}(\Omega))$ with q' such that $\frac{1}{q} + \frac{1}{q'} = 1$, then*

$$\|s \times u\|_{L^1(\mathbb{R})} \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \|u\|_{L^1(\mathbb{R}, L^{q'}(\Omega))}.$$

Proof. We set $F(t,\tau) = \int_\Omega s(x, t-\tau)u(x,\tau)dx$, by Holder inequality we have

$$|F(t,\tau)| \leq \|s(t-\tau)\|_{L^q(\Omega)} \|u(\tau)\|_{L^{q'}(\Omega)}.$$

By summing over t we get

$$\int_{\mathbb{R}} |F(t,\tau)|dt \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \|u(\tau)\|_{L^{q'}(\Omega)},$$

and by summing over τ we obtain the result. □

Corollary B.1. *Let $s \in L^1(\mathbb{R}, L^q(\Omega))$ and $u \in L^p(\mathbb{R}, L^{q'}(\Omega))$ with q' such that $\frac{1}{q} + \frac{1}{q'} = 1$ and $1 \leq p \leq \infty$, then*

$$\|s \times u\|_{L^p(\mathbb{R})} \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \|u\|_{L^p(\mathbb{R}, L^{q'}(\Omega))}.$$

Proof. Let us assume that $1 < p < \infty$ and let p' be such that $\frac{1}{p} + \frac{1}{p'} = 1$. Keeping the same notations as in the proof of Lemma B.1, one can rewrite the upper bound of $|F(t,\tau)|$ as

$$|F(t,\tau)| \leq \underbrace{\|s(t-\tau)\|_{L^q(\Omega)}^{\frac{1}{p'}}}_{\in L^{p'} \text{ w.r.t. } \tau} \underbrace{\|s(t-\tau)\|_{L^q(\Omega)}^{\frac{1}{p}} \|u(\tau)\|_{L^{q'}(\Omega)}}_{\in L^p \text{ w.r.t. } \tau}.$$

By summing over τ and thanks to Holder inequality we have

$$\int_{\mathbb{R}} |F(t,\tau)d\tau| \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))}^{\frac{1}{p'}} \left(\left(\|s\|_{L^q(\Omega)} \star \|u\|_{L^{q'}(\Omega)}^p \right)(t) \right)^{\frac{1}{p}},$$

where $\star$ denotes the usual convolution operator. Taking the power p we get

$$|(s \times u)(t)|^p \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))}^{\frac{p}{p'}} \left(\|s\|_{L^q(\Omega)} \star \|u\|_{L^{q'}(\Omega)}^p \right)(t).$$

By using Lemma B.1 we have

$$\begin{aligned} \left\| \|s\|_{L^q(\Omega)} \star \|u\|_{L^{q'}(\Omega)}^p \right\|_{L^1(\mathbb{R})} &\leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \||u|^p\|_{L^1(\mathbb{R}, L^{q'}(\Omega))}, \\ &= \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \|u\|_{L^p(\mathbb{R}, L^{q'}(\Omega))}^p. \end{aligned}$$

Summing over t and taking the p-th root lead to the result. □

Lemma B.2. *Let $s \in L^p(\mathbb{R}, L^q(\Omega))$ and $z \in L^{p'}(\mathbb{R})$ with $\frac{1}{p} + \frac{1}{p'} = 1$, then the adjoint operator of $\times$ is given by*

$$(s \times^\star z)(x,t) = \int_{\mathbb{R}} s(x, \tau - t) z(\tau),$$

and $s \times^\star z \in L^{p'}(\mathbb{R}, L^q(\Omega))$ with the estimation

$$\|s \times^\star z\|_{L^{p'}(\mathbb{R}, L^q(\Omega))} \leq \|s\|_{L^1(\mathbb{R}, L^q(\Omega))} \|z\|_{L^{p'}(\mathbb{R})}.$$

Proof. Computing the scalar product of $s \times u$ and z gives the expression of $\times^\star$. Then the estimation of the norm is similar to the proof of Cor. B.1. □

Remark B.1. *When $q = p = 2$ and the sequence $s \in L^1(\mathbb{R}, L^2(\Omega))$ is finite, from Lemma B.2 and Corollary B.1, the Lipschitz constant of $s \times .$ and $s \times^\star .$ in norm $L^2(\mathbb{R})$ and $L^2(\mathbb{R}, L^2(\Omega))$ is bounded by $\|s\|_{L^1(\mathbb{R}, L^2(\Omega))}$. Hence the Lipschitz constant of the operator $g \mapsto s \times^\star s \times g$ from $L^2(\Omega)$ into itself is bounded by $\|s\|^2_{L^1(\mathbb{R}, L^2(\Omega))}$.*

C Around the Kurdyla-Lojasiewicz property

To prove convergence, we apply Th. 9 from [1]. The difficult point is to check that $\mathcal{E}_\alpha$ verifies the Kurdyka-Lojasiewicz (KL) property which is a notion introduced in 1963 by Lojasiewicz for analytical functions and extended by Kurdyka in 1998 to functions definable in an o-minimal structure. To be self-content, let us first recall some definitions and notions associated to this property.

Definition C.1 (KL property)**.** *We say that a proper function ϕ has the Kurdyka-Lojasiewicz (KL) property at $\bar{x} \in dom(\partial\phi)$ if there exist a neighborhood $\mathcal{V}$ of $\bar{x}$, $\nu \in]0, \infty]$ and a continuous concave function $\psi : [0, \nu[\to \mathbb{R}^+$ with $\psi(0) = 0$ and such that:*

(i) *ψ is continuously differentiable on $(0, \nu)$ with $\psi' > 0$.*

(ii) *For all $x \in \mathcal{V}$ with $\phi(\bar{x}) \leq \phi(x) \leq \phi(\bar{x}) + \nu$, one has*

$$\psi'(\phi(x) - \phi(\bar{x})) dist(0, \partial\phi(x)) \geq 1. \tag{KL}$$

A proper closed function f satisfying the KL property at all points in $dom(\partial\phi)$ is called a KL function.

Definition C.2 (KL exponent)**.** *For a proper function ϕ satisfying the KL property at $\bar{x} \in dom(\phi)$, if the corresponding function ψ can be chosen as $\psi(s) = cs^{1-\alpha}$ for some $c > 0$ and $\alpha \in [0, 1[$, i.e., there exists c, $\varepsilon > 0$ and $\nu \in]0, \infty]$ such that*

$$dist(0, \partial\phi(x)) \geq c(\phi(x) - \phi(\bar{x}))^\alpha,$$

whenever $\|x - \bar{x}\|_2 \leq \varepsilon$ and $\phi(\bar{x}) \leq \phi(x) \leq \phi(\bar{x}) + \nu$, then we say that ϕ has the KL property at $\bar{x}$ with an exponent α. If ϕ is a KL function and has the same exponent α at $\bar{x} \in dom(\partial\phi)$, then we say that ϕ is a KL function with an exponent of α.

Remark C.1. *A proper lower semicontinuous function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has the KL property at every non critical point (see Lemma 2 in [1]) for whatever exponent $\alpha \in [0, 1]$.*

The exponent of the function involved in the KL inequality gives information on the rate of convergence toward critical points for descent algorithms used in optimisation [1] and toward stationary points for subgradient trajectories [7] from a dynamical systems point of view.

In order to determine which functions satisfy the KL property we introduce the notion of semialgebraic sets. A subset of $\mathbb{R}^n$ is called *semialgebraic* if it can be written as a finite union of sets of the form

$$\{x \in \mathbb{R}^n,\ p_i(x) = 0,\ q_i(x) < 0, i = 1, ..., p\},\}$$

where p_i, q_i are real polynomials. A proper function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is semialgebraic if its graph is a semialgebraic subset of $\mathbb{R}^{n+1}$. It is known that semialgebraic functions verify the KL property.

We recall the definition of o-minimal structure introduced in [18] which can be seen as an axiomatization of properties of semialgebraic sets.

Definition C.3. *Let $\mathcal{O} = \{\mathcal{O}_n\}_{n\in\mathbb{N}}$ be such that each $\mathcal{O}_n$ is a collection of subsets of $\mathbb{R}^n$. The family $\mathcal{O}$ is an o-minimal structure over $\mathbb{R}$ if it satisfies the following axioms:*

(i) *Each $\mathcal{O}_n$ is a boolean algebra. Namely $\emptyset \in \mathcal{O}_n$ and for each A, B in $\mathcal{O}_n$, $A \cup B$, $A \cap B$ and $\mathbb{R}^n \backslash A$ belong to $\mathcal{O}_n$.*

(ii) *For all A in $\mathcal{O}_n$, $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$.*

(iii) *For all A in $\mathcal{O}_{n+1}$, $\Pi(A) := \{(x_1, ..., x_n,) \in \mathbb{R}^n,\ (x_1, ..., x_n, x_{n+1}) \in A\}$ belongs to $\mathcal{O}_n$.*

(iv) *For all $i \neq j$ in $\{1, ..., n\}$, $\{(x_1, ..., x_n) \in \mathbb{R}^n,\ x_i = x_j\} \in \mathcal{O}_n$.*

(v) *The set $\{(x_1, x_2) \in \mathbb{R}^2,\ x_1 < x_2\}$ belongs to $\mathcal{O}_n$.*

(vi) *The elements of $\mathcal{O}_1$ are exactly finite unions of intervals.*

Given an o-minimal structure $\mathcal{O}$, a set A is said to be *definable* (in $\mathcal{O}$), if A belongs to $\mathcal{O}$. A function $\phi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ is said to be definable if its graph is a definable subset of $\mathbb{R}^n \times \mathbb{R}$.

o-minimal structures have interesting properties:

- Finite sums of definable functions are definable.
- Indicator functions of definable sets are definable.
- Compositions of definable functions are definable.
- Generalized inverses of definable functions are definable.
- Functions of the type $\mathbb{R}^n \ni x \to f(x) = \sup_{y\in C} g(x, y)$ or $\mathbb{R}^n \ni x \to f(x) = \inf_{y\in C} g(x, y)$ where g and C are definable, are definable.

The class of semialgebraic sets is an o-minimal structure [4] that we denote *semi-alg.*
A subset $X \subset \mathbb{R}^n$ is *semianalytic* if at each point $a \in X$ there exists a neighborhood U such that $X \cap U$ can be written as finite union and intersection of analytic equalities and inequalities.
A subset $X \subset \mathbb{R}^n$ is *subanalytic* if each point of $\mathbb{R}^n$ admits a neighborhood U such that $X \cap U$ can be written as the projection of a bounded semianalytic subset of $\mathbb{R}^n \times \mathbb{R}^m$ for some $m \geq 1$.
Notice that image and preimage of a subanalytic set are not in general a subanalytic set. Let τ_n defined by

$$\tau_n(x_1, ..., x_n) = \left(\frac{x_1}{1 + x_1^2}, ..., \frac{x_n}{1 + x_n^2}\right)$$

A subset $X \subset \mathbb{R}^n$ is *globally subanalytic* if its image under τ_n is a subanalytic subset of $\mathbb{R}^n$. Globally subanalytic sets are subanalytic and reciprocally any bounded subanalytic sets are globally subanalytic. The class of globally subanalytic sets is an o-minimal structure that we denote *global-subanal*. There exists an o-minimal structure denoted *log-exp* [42, 18] which contains *global-subanal* and the graph of $exp : \mathbb{R} \to \mathbb{R}$. This huge o-minimal structure contains all the aforementioned structures.

The following theorem enables to characterize KL functions by using the notion of o-minimal structure.

Theorem C.1 (from [8])**.** *Any proper lower semicontinuous function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ which is definable in an o-minimal structure $\mathcal{O}$ has the KL property at each point of $dom(\partial\phi)$. Moreover the function ψ of the* (KL) *inequality is definable in $\mathcal{O}$.*

Proof of Theorem 2.1
In order to apply Th. 9 of [1], the only point to check is if $\mathcal{E}_\alpha$ verifies the KL property (see Def. C.1). Thanks to Th. C.1, a sufficient condition is that $\mathcal{E}_\alpha$ be definable in an o-minimal structure. Hence we need to find the smallest structure in which our functional is defined. Our energy is composed of three kinds of terms:

$$\begin{aligned} \varphi_1(z) &= \psi_\xi(z), \\ \varphi_2(z,u) &= \|z - s \times_d u\|_2^2, \\ \varphi_3(u) &= \lambda\|u\|_1 + \mu\|Hu\|_1. \end{aligned}$$

For each term, one can show that it is definable:

- With Hypothesis 2, in the case $\xi = 0$ the first term $\varphi_1(z)$ is definable in the log-exp structure. For $\xi > 0$ by Hypothesis 2 and stability by composition with *log* and linear transform, $\varphi_1(z)$ is still be definable in this structure.
- The second term is semialgebraic so it is definable in the log-exp structure.
- The third term is the sum of terms of the form $|(Hu)_i|_{\mathbb{R}^\nu}$ where the norm $|v|_{\mathbb{R}^\nu} = \sqrt{\sum_{i=1}^\nu v_i^2}$ is semialgebraic (its graph can be written as $\{(v,z),\ \sum_{i=1}^\nu v_i^2 = z^2,\ z \geq 0\}$. Hence thanks to stability by sum and composition with linear operator, we deduce that $\varphi_3(u)$ is definable in the log-exp structure.

With the stability by the $\times$ operator (axiom (ii) of Def. C.3), we deduce that $(z,u) \mapsto \varphi_1(z)$ and $(z,u) \mapsto \varphi_3(u)$ are also in the log-exp structure. In conclusion, the energy $\mathcal{E}_\alpha(z,u) = \sum_i \psi_{\xi_i}(z) + \frac{\alpha}{2}\varphi_2(z,u) + \varphi_3(u)$ is definable in the log-exp structure. Theorem C.1 gives that $\mathcal{E}_\alpha$ verifies the KL property with a function ψ in (KL) which is definable in the log-exp structure. By applying Theorem 9 of [1] we get the result. □

Remark C.2. *Let us notice that to show that $\mathcal{E}_\alpha$ verifies KL we have shown that $\mathcal{E}_\alpha$ is definable in the log-exp structure. Hence we have no more information on the function ψ in* (KL) *than it is definable in this later structure. In general, f is piecewise defined by analytic expression, but if we can show that it admits locally an analytic extension then $\mathcal{E}_\alpha$ is subanalytic and verifies* (KL) *locally with ψ a power function* [7]*. This fact is important because according to the power of the function ψ, the rate of convergence is different (see [1], Th. 11).*

D Solving problem (16a) for a piecewise linear sigmoid

This appendix gives the expression of the solution of (16a) in the case of a piecewise linear continuous sigmoid-like non linearity (see Figure 3(b)). We will see that in this case (16a) cannot be reduced to an equality. The normalized piecewise linear sigmoid function that we study is

$$f(x) = \begin{cases} 0, \text{ for } x < -\dfrac{1}{2}, \\ \dfrac{1}{2} + x, \text{ for } |x| \leq \dfrac{1}{2}, \\ 1, \text{ for } x > \dfrac{1}{2} \end{cases} \tag{38}$$

Remark D.1. *The non linearity f (38) does not verify Hypothesis 2(ii): at point $x = \frac{1}{2}$, f cannot be locally split into the difference of a convex function and a quadratic one. Hence the proximal operator can be multivalued (which will be the case here, see below). However, convergence of* (16) *is still guaranteed thanks to Th. 2.1.*

D.1 Case $\xi = 0$

Let $\gamma > 0$, we are interested in computing the proximal operator associated to γf. Since $f(y)$ is constant for $|y| \geq \frac{1}{2}$, we split $\mathbb{R}$ into the union of two disjoint sets: $I_1 = \left[-\frac{1}{2}, \frac{1}{2}\right]$ and $I_2 = \left]-\infty, +\frac{1}{2}\right[\cup \left]\frac{1}{2}, +\infty\right[$, and we compute:

$$g_1(x) = \min_{|y| \leq \frac{1}{2}} \varphi_{x,\gamma}(y) \quad \text{with} \quad y_1(x) = \arg\min_{|y| \leq \frac{1}{2}} \varphi_{x,\gamma}(y),$$
$$g_2(x) = \min_{|y| \geq \frac{1}{2}} \varphi_{x,\gamma}(y) \quad \text{with} \quad y_2(x) = \arg\min_{|y| > \frac{1}{2}} \varphi_{x,\gamma}(y),$$

where $\varphi_{x,\gamma}(y) = f(y) + \frac{1}{2\gamma}(x-y)^2$. Standard computations lead to:

$$y_1(x) = \begin{cases} -\frac{1}{2} & \text{if } x \leq \frac{-1+2\gamma}{2}, \\ x - \gamma & \text{if } \frac{-1+2\gamma}{2} \leq x \leq \frac{1+2\gamma}{2}, \\ \frac{1}{2} & \text{if } x \geq \frac{1+2\gamma}{2}, \end{cases}$$

and

$$y_2(x) = \begin{cases} x & \text{if } x \leq -\frac{1}{2}, \\ -\frac{1}{2} & \text{if } -\frac{1}{2} \leq x \leq \gamma, \\ \frac{1}{2} & \text{if } \gamma \leq x \leq \frac{1}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases}$$

Computing $g(x) = min(g_1(x), g_2(x))$ we get the expression of the proximal operator of γf (see Figure 11)

$$prox_{\gamma f}(x) = \begin{cases} x & \text{if } x \leq -\frac{1}{2}, \\ -\frac{1}{2} & \text{if } -\frac{1}{2} \leq x \leq \gamma, \\ x - \gamma & \text{if } \frac{-1+2\gamma}{2} + \gamma \leq x \leq \frac{1+\gamma}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases} \tag{39}$$

Let us underline that the proximal function of γf is multivalued at $x = \frac{1+\gamma}{2}$ as we can see on Fig. 11.

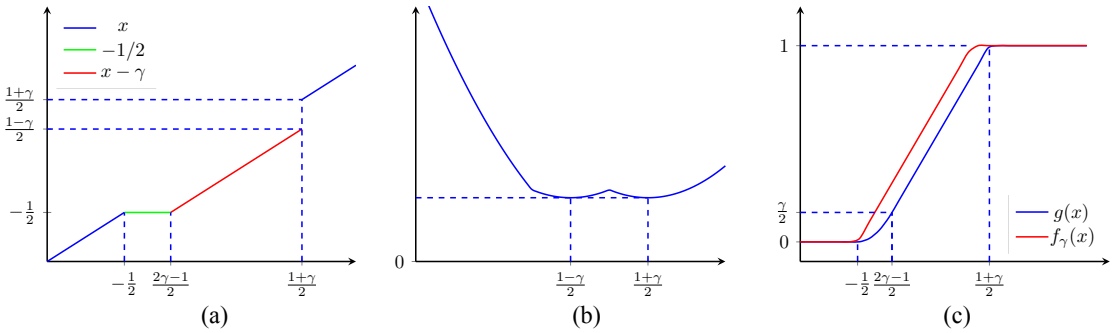


Figure 11: Plots of functions used in the piecewise linear sigmoid case. (a) $prox_{\gamma f}(x)$. (b) $\varphi_\gamma(x) = f(x) + \frac{1}{2\gamma}(x - \frac{1+\gamma}{2})^2$. (c) $f_\gamma(x) = \inf_y \; f(y) + \frac{1}{2\gamma}(x-y)^2$ the Moreau envelop of f compared to f.

D.2 Case $\xi > 0$

We recall the expression of the function ψ_ξ (11) in this case, extended on $\{f \leq 0\}$ by the value $+\infty$ (Fig. 12):

$$\psi_\xi(x) = \begin{cases} \frac{1}{2} + x - \xi log\left(\frac{1}{2} + x\right) & \text{if } |x| < \frac{1}{2}, \\ 1 & \text{if } x \geq \frac{1}{2}. \end{cases}$$

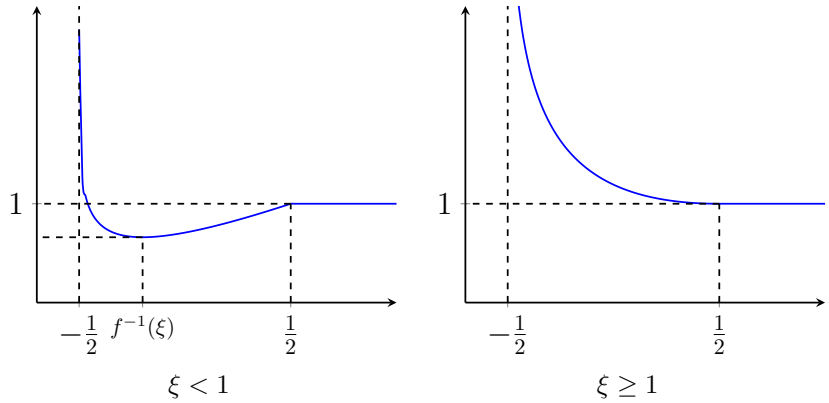


Figure 12: Shape of the data fidelity term ψ_ξ in the case of nonlinear sigmoid-like piecewise linear function f (38).

Remark D.2. *ψ_ξ is convex as soon as $\xi \geq 1$. This fact comes from the non increasing of ψ_ξ for $\xi \geq 1$.*

Let $\gamma > 0$, in order to compute the proximal operator of $\gamma\psi_\xi$, we compute

$$h_1(x) = \min_{|y|\leq\frac{1}{2}} \varphi_{x,\gamma}(y), \qquad z_1(x) = \arg\min_{|y|\leq\frac{1}{2}} \varphi_{x,\gamma}(y),$$
$$h_2(x) = \min_{|y|>\frac{1}{2}} \varphi_{x,\gamma}(y), \qquad z_2(x) = \arg\min_{|y|>\frac{1}{2}} \varphi_{x,\gamma}(y),$$

where $\varphi_{x,\gamma}(y) = f(y) - \xi_i log(f(y)) + \frac{1}{2\gamma}(x-y)^2$. It is easily seen that

$$z_2(x) = \begin{cases} \frac{1}{2} & \text{if } x \leq \frac{1}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases}$$

We set $\phi_{x,\gamma}(y) = \frac{1}{2} + y - \xi log(\frac{1}{2}+y) + \frac{1}{2\gamma}(x-y)^2$ for $y > -\frac{1}{2}$ then

$$\phi'_{x,\gamma}(y) = 0 \Leftrightarrow y^2 + \left(\gamma + \frac{1}{2} - x\right) y + \frac{\gamma}{2} - \frac{x}{2} - \gamma\xi = 0 \ , \ \ \xi > 0.$$

We search roots of $\phi'_{x,\gamma}(y) = 0$ belonging to $[-\frac{1}{2}, \frac{1}{2}]$.
We compute $\Delta = \left(x - (\gamma - \frac{1}{2})\right)^2 + 4\gamma\xi > 0$, there are two distinct roots

$$y_\pm(x) = \frac{-(\gamma + \frac{1}{2} - x) \pm \sqrt{\Delta}}{2}. \tag{40}$$

We check that $y_+ \geq -\frac{1}{2}$ and $y_- \leq -\frac{1}{2}$ for all $x \in \mathbb{R}$. Besides

$$y_+ \leq \frac{1}{2} \Leftrightarrow x \leq x_1^\star := \gamma(1-\xi) + \frac{1}{2}.$$

By setting $x_1^\star = \gamma(1-\xi) + \frac{1}{2}$, we obtain the following tables of variations:

y	$-\frac{1}{2}$		$\frac{1}{2}$		y_+		$+\infty$
$\phi'_{x,\gamma}$	$\Vert$		$-$		0		$+$
$\phi_{x,\gamma}$	$\Vert$	$\searrow$	$\phi_{x,\gamma}(\frac{1}{2})$	$\searrow$	$\phi_{x,\gamma}(y_+)$	$\nearrow$	

case $(x \geq x_1^\star)$

y	$-\frac{1}{2}$		y_+		$\frac{1}{2}$	$+\infty$
$\phi'_{x,\gamma}$	$\Vert^{+\infty}$	$-$	0	$+$		
$\phi_{x,\gamma}$	$\Vert^{+\infty}$	$\searrow$	$\phi_{x,\gamma}(y_+)$	$\nearrow$		

case $(x \leq x_1^\star)$

We can show that $x \mapsto y_+(x)$ is not decreasing and we have $y_+(x_1^\star) = \frac{1}{2}$ and $\lim_{x\to-\infty} y_+(x) = -\frac{1}{2}$. We get that:

$$z_1(x) = \begin{cases} y_+(x) & \text{if } x \leq x_1^\star, \\ \frac{1}{2} & \text{if } x \geq x_1^\star, \end{cases}$$

and by computing $h = min(h_1, h_2)$ we get the expression of $prox_{\gamma\psi_\xi}$ (Fig. 13):

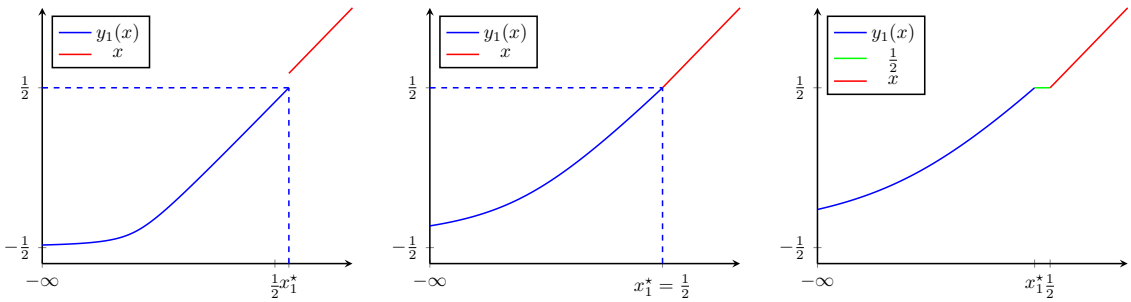
- Case $\xi \leq 1$

$$prox_{\gamma\psi_\xi}(x) = \begin{cases} y_+(x) & \text{if } x \leq x_1^\star, \\ x & \text{if } x \geq x_1^\star. \end{cases}$$

- Case $\xi > 1$

$$prox_{\gamma\psi_\xi}(x) = \begin{cases} y_+ & \text{if } x \leq x_1^\star, \\ \frac{1}{2} & \text{if } x_1^\star \leq x \leq \frac{1}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases}$$

We conclude this study by the observation that for $\xi < 1$ the function $\psi_\xi(z)$ is not semiconvex (even not lower C^2), that leads to the fact that its proximal operator is multivalued at $x = x_1^\star$ and is given at this point by the set $\{\frac{1}{2}, x_1^\star\}$.



$\xi \leq 1$ $\xi = 1$ $\xi > 1$

Figure 13: Plots of $prox_{\gamma\psi_\xi}(x)$ for different values of $\xi > 0$.

E Solving problem (16a) for a piecewise cubic sigmoid

The cubic approximation that we choose is:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \leq -\frac{1}{2}, \\ \frac{1}{2} + \frac{3}{2}x - 2x^3 & \text{if} \quad |x| \leq \frac{1}{2}, \\ 1 & \text{if} \quad x \geq \frac{1}{2}. \end{cases} \tag{41}$$

Remark E.1. *The sigmoid f given by* (41) *verifies Hypothesis 1 and 2:*

(i) *f is semialgebraic (see Appendix C) since its graph can be written as a finite union of polynomial inequalities.*

(ii) *The function $y \mapsto -log(f(y))$ is convex on $]-\frac{1}{2}, +\infty[$.*

(iii) *The function $f(y) + 3y^2$ is convex.*

We denote by $\varphi_{x,\gamma}$ the function to minimize at fixed x:

$$\varphi_{x,\gamma}(y) = f(y) + \frac{1}{2\gamma}(x-y)^2,$$

and we set $g(x) = \min_y \varphi_{x,\gamma}(y)$.

In practice γ will be close to 0 since α will be chosen large (see (17)). Hence, we assume that $\gamma \leq \frac{1}{6}$ so that $\varphi_{x,\gamma}$ be convex for all x.

Following the same reasoning as for the computation of the proximal operator in the case of a piecewise linear sigmoid, we compute

$$\begin{aligned} g_1(x) &= \min_{|y| \leq \frac{1}{2}} \varphi_{x,\gamma}(y), \quad y_1(x) = \arg\min_{|y| \leq \frac{1}{2}} \varphi_{x,\gamma}(y), \\ g_2(x) &= \min_{|y| \geq \frac{1}{2}} \varphi_{x,\gamma}(y), \quad y_2(x) = \arg\min_{|y| \geq \frac{1}{2}} \varphi_{x,\gamma}(y), \end{aligned}$$

and $g(x) = \min(g_1(x), g_2(x))$ and the proximal operator is defined by

$$prox_{\gamma f}(x) = \begin{cases} y_1(x) & \text{if } g(x) = g_1(x), \\ y_2(x) & \text{otherwise.} \end{cases}$$

E.1 Computation of $y_1(x)$

For $|y| \leq \frac{1}{2}$, $\varphi_{x,\gamma}(y) = \frac{1}{2} + \frac{3}{2}y - 2y^3 + \frac{1}{2\gamma}(x-y)^2$ and its derivative is $\varphi'_{x,\gamma}(y) = -6y^2 + \frac{y}{\gamma} + \frac{3}{2} - \frac{x}{\gamma}$. The problem is now to search roots of $\varphi'_{x,\gamma}(y) = 0$ belonging to $[-\frac{1}{2}, \frac{1}{2}]$. The discriminant of the equation $\varphi'_{x,\gamma}(y) = 0$ is $\Delta = \frac{1}{\gamma^2} + 36 - \frac{24x}{\gamma} = \frac{\Delta'}{\gamma^2}$ with $\Delta' = 1 + 36\gamma^2 - 24x\gamma$. There exists real roots if

$$\Delta \geq 0 \Leftrightarrow x \leq x_0^\star := \frac{3}{2}\gamma + \frac{1}{24\gamma}.$$

Case 1: $\Delta \geq 0$ (i.e. $x \leq x_0^\star$)
The equation has two solutions which are:

$$y_\pm = \frac{1 \pm \sqrt{\Delta'}}{12\gamma}, \tag{42}$$

and $\varphi_{x,\gamma}$ varies as

y	$-\infty$		y_-		y_+		$+\infty$
$\varphi'_{x,\gamma}$	$-$		0	$+$	0	$-$	
$\varphi_{x,\gamma}$	$\searrow$			$\nearrow$		$\searrow$	

(43)

We search values of x for which $y_\pm$ are in $[-\frac{1}{2}, \frac{1}{2}]$:

$$-\frac{1}{2} \leq y_- \leq \frac{1}{2} \Leftrightarrow -6\gamma + 1 \leq \sqrt{\Delta'} \leq 6\gamma + 1,$$
$$-\frac{1}{2} \leq y_+ \leq \frac{1}{2} \Leftrightarrow -6\gamma - 1 \leq \sqrt{\Delta'} \leq 6\gamma - 1.$$

Since $\gamma \leq \frac{1}{6}$, we easily check that $y_+ \geq \frac{1}{2}$ for all x and that $x_0^\star \leq \frac{1}{2}$. For y_- there are three cases:

$$\begin{cases} -\frac{1}{2} \leq y_-(x) \leq \frac{1}{2} & \text{if } -\frac{1}{2} \leq x \leq \frac{1}{2}, \\ y_-(x) \geq \frac{1}{2} & \text{if } x \geq \frac{1}{2}, \\ y_-(x) \leq -\frac{1}{2} & \text{if } x \leq -\frac{1}{2}, \end{cases}$$

and so $y_1(x)$ is given by

$$y_1(x) = \begin{cases} -\frac{1}{2} & \text{if } x \leq -\frac{1}{2}, \\ y_-(x) & \text{if } |x| \leq \frac{1}{2}, \\ \frac{1}{2} & \text{if } x \geq \frac{1}{2}, \end{cases}$$

where y_- is given by (42).

Case 2: $\Delta < 0$ (i.e. $x > x_0^\star$)
In this case, $\varphi'_{x,\gamma} \leq 0$ i.e. $\varphi_{x,\gamma}$ is non increasing and $y_1(x) = \frac{1}{2}$.

E.2 Computation of $y_2(x)$

By a similar reasoning as for the computation of g_1 and y_1, we get

$$y_2(x) = \begin{cases} x & \text{if } x \leq -\frac{1}{2}, \\ -\frac{1}{2} & \text{if } -\frac{1}{2} \leq x \leq \gamma, \\ \frac{1}{2} & \text{if } \gamma \leq x \leq \frac{1}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases}$$

E.3 Expression of $prox_{\gamma f}(x)$

(i) If $x \in [-\frac{1}{2}, \frac{1}{2}]$, $y_1(x) = y_-(x)$ and from (43), we deduce that $\varphi_{x,\gamma}(y)$ is minimal at $y = y_1(x)$ and its value is $g_1(x)$.
Hence $g_1(x) \leq min(\varphi_{x,\gamma}(-\frac{1}{2}), \varphi_{x,\gamma}(\frac{1}{2})) = g_2(x)$ and $g(x) = g_1(x)$ and $prox_{\gamma f}(x) = y_1(x)$.

(ii) If $x \geq \frac{1}{2}$ and $x \leq -\frac{1}{2}$ it is easily seen that $g(x) = g_2(x)$ and $prox_{\gamma f}(x) = y_2(x) = x$.

To summarize, $prox_{\gamma f}(x)$ is given by (Fig. 14):

$$prox_{\gamma f}(x) = \begin{cases} x & \text{if } x \leq -\frac{1}{2}, \\ y_-(x) & \text{if } |x| \leq \frac{1}{2}, \\ x & \text{if } x \geq \frac{1}{2}. \end{cases}$$

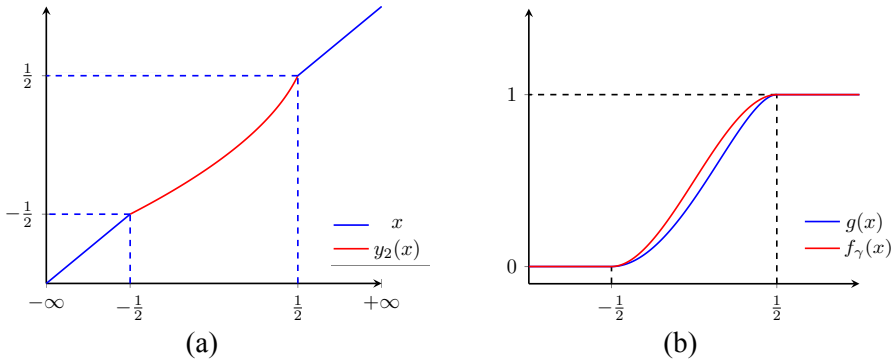


Figure 14: Plots of functions used in the piecewise cubic sigmoid case. (a) $prox_{\gamma f}(x)$ for $\gamma \leq \frac{1}{6}$. (b) $g(x) = \inf_y f(y) + \frac{1}{2\gamma}(y - x)^2$ the Moreau envelope of f.

Remark E.2. *We remind that the case $\gamma \geq \frac{1}{6}$ never happens in our applications since γ is small since it is a $O\left(\frac{1}{\alpha}\right)$ and α is large. However in this case we can show that the proximal operator is multivalued. More precisely, there exists $x_1^\star \in [\frac{1}{2}, x_0^\star]$ such that $y_-(x_1^\star) < x_1^\star$ and the proximal operator is for $\gamma \geq \frac{1}{6}$ of the form:*

$$prox_{\gamma f}(x) = \begin{cases} x & \text{if } x \leq -\frac{1}{2}, \\ y_-(x) & \text{if } -\frac{1}{2} \leq x \leq x_1^\star, \\ x & \text{if } x \geq x_1^\star, \end{cases}$$

and so $prox_{\gamma f}(x_1^\star) = \{y_-(x_1^\star), x_1^\star\}$ i.e. the proximal operator is multivalued.

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, May 2010.

[2] A. Beck and M. Teboulle. Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, November 2009.

[3] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle. A l1-unified variational framework for image restoration. In *Proceedings of the 8th European Conference on Computer Vision*, page 1–13, 2004.

[4] R. Benedetti and J. Risler. *Real algebraic and semi-algebraic sets*. Actualités mathématiques. Hermann, 1990.

[5] L. Berdondini, K. Imfeld, A. Maccione, M. Tedesco, S. Neukom, M. Koudelka-Hep, and S. Martinoia. Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large scale neuronal networks. *Lab on a Chip*, 9(18):2644–2651, 2009.

[6] M. Bergounioux and L. Piffet. A Second-Order Model for Image Denoising. *Set-Valued and Variational Analysis*, 18(3-4):277–306, December 2010.

[7] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz Inequality for Nonsmooth Subanalytic Functions with Applications to Subgradient Dynamical Systems. *SIAM Journal on Optimization*, 17(4):1205–1223, January 2007.

[8] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke Subgradients of Stratifiable Functions. *SIAM Journal on Optimization*, 18(2):556–572, January 2007.

[9] S. Boyd. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

[10] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.

[11] D. A. Butts, C. Weng, J. Jin, J.-M. Alonso, and L. Paninski. Temporal Precision in the Visual Pathway through the Interplay of Excitation and Stimulus-Driven Suppression. *Journal of Neuroscience*, 31(31):11313–11327, August 2011.

[12] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[13] M. Carandini, J. B. Demb, V. Mante, D. J. Tollhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597, November 2005.

[14] A. Chambolle. An Algorithm for Total Variation Minimization and Applications. *J. Math. Imaging Vis.*, 20(1-2):89–97, January 2004.

[15] E. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2):199–213, 2001.

[16] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 185–212. Springer New York, 2011.

[17] F. Demengel. Fonctions à hessien borné. *Annales de l'institut Fourier*, 34(2):155–190, 1984.

[18] L. Dries. *Tame topology and o-minimal structures*. Number 248 in London Mathematical Society lecture note series. Cambridge Univ. Press, Cambridge, digital print edition, 2003. OCLC: 550530719.

[19] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels*. Etudes mathématiques. Dunod, 1974.

[20] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol*, 187:517–552, 1966.

[21] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

[22] T. Gollisch and M. Meister. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–164, January 2010.

[23] G. Hilgen, S. Pirmoradian, D. Pamplona, P. Kornprobst, B. Cessac, M. H. Hennig, and E. Sernagor. Pan-retinal characterization of light responses from ganglion cells in the developing mouse retina. *bioRxiv*, 2016.

[24] A. Maccione, M. H. Hennig, M. Gandolfo, O. Muthmann, J. Coppenhagen, S. J. Eglen, L. Berdondini, and E. Sernagor. Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. *The Journal of physiology*, 592(7):1545–1563, 2014.

[25] R. H. Masland. Cell populations of the retina: The proctor lecture. *Investigative Ophthalmology and Visual Science*, 52(7):4581–4591, June 2011.

[26] J. McFarland, Y. Cui, and D. A. Butts. Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. *PLoS Computational Biology*, 9(7):e1003143, July 2013.

[27] J. McFarland, Y. Cui, and D. Butts. Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Computational Biology*, 9(7), 2013.

[28] J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[29] Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, August 2013.

[30] D. Pamplona, G. Hilgen, S. Pirmoradian, M. Hennig, B. Cessac, E. Sernagor, and P. Kornprobst. A super-resolution approach for receptive fields estimation of neuronal ensembles. Computational Neuroscience (CNS), July 2015.

[31] L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, November 2004.

[32] L. Paninski, J. Pillow, and J. Lewi. Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog Brain Res.*, 165:493–507, 2007.

[33] M. Park and J. W. Pillow. Receptive Field Inference with Localized Priors. *PLoS Computational Biology*, 7(10):e1002219, October 2011.

[34] Z. Peng, Y. Xu, M. Yan, and W. Yin. ARock: an Algorithmic Framework for Async-Parallel Coordinate Updates. *SIAM Journal on Scientific Computing*, 2016.

[35] J. Pillow. Likelihood-based approaches to modeling the neural code. *Bayesian brain: Probabilistic approaches to neural coding*, pages 53–70, 2007.

[36] J. Pillow, L. Paninski, V. Uzzell, E. Simoncelli, and E. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25(47):11003–11013, 2005.

[37] J. Pillow, J. Shlens, L. Paninski, A. Sher, A. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[38] R. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[39] R. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12):583–601, December 1965.

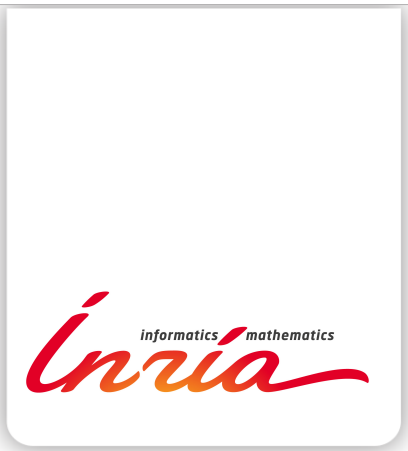
[40] E. Simoncelli, L. Paninski, J. Pillow, and O. Schwartz. *The Cognitive Neurosciences, 3rd edition*, chapter Characterization of neural responses with stochastic stimuli. MIT Press, 2004.

[41] W. Su, S. Boyd, and E. Candes. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2510–2518. Curran Associates, Inc., 2014.

[42] A. Wilkie. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *Journal of the American Mathematical Society*, 9(4):1051–1094, 1996.

[43] A. Wohrer. *Model and large-scale simulator of a biological retina with contrast gain control*. PhD thesis, University of Nice Sophia-Antipolis, 2008.

[44] A. Wohrer and P. Kornprobst. Virtual Retina: A biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, 26(2):219–249, April 2009.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399