Online Learning Approach for Survival Analysis

Camila Fernandez, Pierre Gaillard, Joseph de Vilmarest, Olivier Wintenberger

Online Learning Approach for Survival Analysis

Camila Fernandez Pierre Gaillard Joseph de Vilmarest
Olivier Wintenberger

Abstract

We introduce an online mathematical framework for survival analysis, allowing real time adaptation to dynamic environments and censored data. This framework enables the estimation of event time distributions through an optimal second order online convex optimization algorithm—Online Newton Step (ONS). This approach, previously unexplored, presents substantial advantages, including explicit algorithms with non-asymptotic convergence guarantees. Moreover, we analyze the selection of ONS hyperparameters, which depends on the exp-concavity property and has a significant influence on the regret bound. We propose a stochastic approach that guarantees logarithmic stochastic regret for ONS. Additionally, we introduce an adaptive aggregation method that ensures robustness in hyperparameter selection while maintaining fast regret bounds. The findings of this paper can extend beyond the survival analysis field, and are relevant for any case characterized by poor exp-concavity and unstable ONS. Finally, these assertions are illustrated by simulation experiments.

Keywords: Online learning, survival analysis, regret bounds, convex optimization, stochastic risk.

1 INTRODUCTION

On the one hand the primary objective of survival analysis is to estimate the time until a critical event occurs, often referred to as survival time or failure time. Examples of such events include customer churn [21], machine failures [3], and employees' attrition [23]. Survival analysis is particularly suited for scenarios where the occurrence of the event may not be observed for all individuals in the dataset. This phenomenon arises when data collection happened before the event occurred, or individuals left the study before experiencing the event, and is called right censoring. As neglecting the censored data is restrictive, it is essential to consider censorship in estimating event time distributions to avoid bias and underestimation. For each individual i with event time t_i, we define the survival probability function as

$$S_i(t) = \mathbb{P}(t_i \geq t), \qquad t \geq 0.$$

On the other hand convex optimization aims to find the minimum of a convex function over a convex set. It can be extended to an online approach in which the dataset becomes available in sequential order and is used to update the estimations of the algorithms at each step. This setting is suitable when the dataset is rapidly evolving over time, allowing for efficient processing of large volumes of data. Online convex optimization is a broad field with diverse applications such as online portfolio selection in finance, signal processing, communication, and machine learning algorithms; see Hazan [11] and references therein.

In this paper, we propose the application of online convex optimization algorithms to survival analysis. The combination of these two approaches has not been explored before. Our method offers significant advantages, including explicit algorithms with non-asymptotic convergence guarantees, making it a promising tool for the survival analysis field.

Specifically, we estimate a parametric survival probability function S_i using online convex optimization algorithms: let Θ be a non-empty, convex, compact set in $\mathbb{R}^d$, and ℓ_t the

negative log-likelihood of the individuals at risk during the interval $(t-1, t]$, $t \geq 1$. The performance of online convex optimization algorithms is measured with the regret

$$Regret_n := \sum_{t=1}^{n} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{n} \ell_t(\theta), \qquad n \geq 1,$$

which indicates how close the cumulative loss is to the optimal solution. A smaller regret implies better performance, and our objective is to bound its growth with respect to n as slowly as possible.

One of the most widely used algorithms in online convex optimization is the Online Newton Step (ONS) of Hazan et al. [12], renowned for its fast regret convergence rate for exp-concave loss functions. This second-order algorithm relies on a hyperparameter known as the learning rate, whose optimal selection is directly dependent on the exp-concavity properties of the loss functions. The exp-concavity constant plays a fundamental role in the theoretical regret analysis of ONS.

We give a detailed mathematical framework for online survival analysis data and we implement the ONS method to optimize the negative log-likelihood of the exponential model. We note that the ONS algorithm requires a careful selection of the learning rate to ensure robust performance. However, certain choices, such as the learning rate proposed by Hazan et al. [12], might lead to an explosive increase in regret, particularly when applied to the survival losses ℓ_t. Therefore, proper selection of the learning rate is essential in our application.

We discuss various strategies for selecting the learning rate hyperparameter. The first contribution involves applying the stochastic setting from Wintenberger [29] to the survival case. This setting enhances convex properties by assessing stochastic risks rather than cumulative losses, allowing us to attain theoretical guarantees for the stochastic regret that is strongly related to the exp-concavity properties on average. Consequently, this provides the convergence of the algorithm estimations to the real parameter under well-specification. Secondly, in the deterministic setting, we propose to apply ONS to an auxiliary function that recursively adapts the learning rate in response to updates

in the exp-concavity constant. We introduce the algorithm SurvONS, an aggregation procedure which ensures a logarithmic regret bound and robustness in hyperparameter selection over a fixed grid. This provides a new compromise in the context of second-order algorithms: the algorithm either performs well on average (as in the case of BOA [28]) or performs well for certain iterations (as in our case with SurvONS). It is important to emphasize that this algorithm is applicable not only to the survival case but also to any case where the exp-concavity properties are poor and the original versio of ONS is unstable. Finally, we conduct experiments using simulated data to examine our algorithm's behavior under different constraints. We discuss the choice of the grid, and we observe that the combination of multiple ONS allows us to use larger grids in SurvONS than in BOA-ONS [28].

The literature in survival analysis is considerable. The approaches range from non-parametric methods, such as the one proposed by Kaplan and Meier in 1958 [17], to semi-parametric methods like Cox proportional hazards [5], and more recent machine learning applications. For instance, Ishwaran proposed an adapted random forest for censored data inx [15]. Another example is DeepSurv, which was introduced by Katzman in [18]. DeepSurv utilizes deep learning techniques to estimate the log-risk function in the Cox model. From a theoretical perspective, Arjas and Haara [1] proposed a dynamic setting called discrete-time logistic regression. In this model, events are always treated in the order in which they occurred in real time. The authors provided an asymptotic normality result for the maximum likelihood estimator of the regression coefficients. The discrete model is a suitable choice when events are observed at discrete time points; see Tutz [26]. Building upon Arjas and Haara's framework, Fahrmeir [9] introduced a state-space approach for analyzing discrete-time survival data. This approach includes the estimation of time-varying covariate effects achieved by maximizing posterior densities through the use of Kalman Filter algorithms. Christoffersen [4] provided a method for discretising continuous event times when the instantaneous hazard follows an exponential shape. In a similar setting we provide adaptive estimators with non-asymptotic guarantees for the first time.

2 BACKGROUND ON PARAMETRIC INFERENCE

2.1 Notation

We consider a set of N individuals denoted by $i \in \{1, \dots, N\}$, each associated with an arrival time $\tau_i \geq 0$. Such time could represent when a patient enters the hospital, a client joins the company, or simply when an individual enrolls in the study. Every individual has a unique event time t_i, which is a positive random variable. By definition, we have $t_i \geq \tau_i$ almost surely (a.s). We also define c_i, which marks the cessation of observation for the individual i; this time is referred to as the censored time. For instance, this might be applicable in cases where the observation period has a predetermined ending. In a more general context, c_i can be a positive random variable satisfying $c_i \geq \tau_i$ a.s. Given that some individuals are censored before the event occurs, and vice versa, it is natural to define the observed time as $u_i := \min\{t_i, c_i\}$. We also define the event indicator $\delta_i := \mathbb{1}\{t_i \leq c_i\}$, which provides a way to discern whether an event has happened or if it is censored. For each individual $i \in \{1, \dots, N\}$, we obseve the random variables $(u_i, \delta_i) \in \mathbb{R}_+ \times \{0, 1\}$. Furthermore, we suppose that both t_i and c_i are independent across all individuals.

Explanatory variables are defined to give context through time to each of the individuals, and these will be represented by left continuous functions $x_i : \mathbb{R}_+ \to \mathbb{R}^d$. The explanatory variables $x_i(t) \in \mathbb{R}^d$ combine covariates of the individual $i \in \{1, \dots, N\}$ at time $t \geq 0$. It's important to note that we use the variable t to refer to time in general, while t_i represents the specific event time of individual i. We assume that given x_i, a short notation for $(x_i(t))_{t \geq 0}$, the times t_i and c_i are conditionally independent. Additionally, we suppose t_i follows a continuous distribution of density $f(t|x_i, \tau_i)$ and c_i a continuous distribution of density $g(t|x_i, \tau_i)$. We have $g(t|x_i, \tau_i) = f(t|x_i, \tau_i) = 0$ for all $t < \tau_i$ since $t_i, c_i \geq \tau_i$ a.s.

In addition, we suppose that g satisfies the following property:

$$\forall t \geq \varepsilon > 0 : g(t|x_i, \tau_i) = g(t - \varepsilon|x_i, \tau_i - \varepsilon)\,.$$

Note that this assumption is also necessary for the density function f. However, as we will know its specific shape, the property is inherently satisfied. Finally, we denote by I_d the identity matrix of dimension d.

2.2 Survival probability

The objective of survival analysis is to predict the length of time until a specified event occurs. Consequently, it is necessary to estimate the distribution of these events. We define the survival probability function of individual i to be the complement of the cumulative distribution, that is, $S(t|x_i, \tau_i) = 1 - \int_{\tau_i}^{t} f(s|x_i, \tau_i)ds$, which can also be expressed as the probability of surviving up to time t:

$$S(t|x_i, \tau_i) = \mathbb{P}(t_i \geq t|x_i, \tau_i)\,, \qquad t \geq 0\,.$$

To estimate this function, it is common to assume a particular shape for the hazard function. The hazard function is defined as:

$$H(t|x_i, \tau_i) = -\frac{\partial}{\partial t}\log(S(t|x_i, \tau_i))\,, \qquad t \geq 0\,,$$

which represents the instantaneous risk of the event occurring at time t. Notably, we can derive the survival function from the hazard function:

$$S(t|x_i, \tau_i) = \exp\left(-\int_0^t H(s|x_i, \tau_i)ds\right)\,, \qquad t \geq 0\,.$$

For more details on event times distributions, refer to Cox and Oakes [6].

2.3 Likelihood

In order to estimate the survival probability we suppose the hazard function is a function of a specified parametric family Θ given the explanatory variables. The parameters will be determined following the likelihood principle observing $(u_i, \delta_i) \in \mathbb{R}_+ \times \{0, 1\}$ and knowing (x_i, τ_i). As usual we implicitly make the assumption of non-informative censoring (see Kalbfleisch et al. [16]), which means that the censored distribution does not involve the parameter θ.

As mentioned earlier, some models assume a specific shape for the hazard function, such as additive, exponential, logistic or Weibull (see Cox and Oakes [6]). In this paper, we assume that the hazard function is exponential, and we detail this assumption below.

Definition 1 (Log-linear regression model for the Hazard function)**.** *We assume that there exist a vector $\theta \in \mathbb{R}^d$, such that the hazard function satisfies for all $t \geq 0$ and all $x_i : \mathbb{R}_+ \to \mathbb{R}^d$,*

$$H(t|x_i, \tau_i) := h(\theta^T x_i(t)) \mathbb{1}\{t \geq \tau_i\}\,, \qquad t \geq 0\,,$$

where $h : x \in \mathbb{R} \mapsto \exp(x)$ is the response function.

By using this exponential model we obtain a formula to compute the negative log-likelihood which is the function that we aim to minimize.

Proposition 1. *Under the exponential model from Definition 1 and omitting additional constants, the negative log-likelihood function $\ell : \Theta \to \mathbb{R}$ can be written in the following way:*

$$\ell(\theta) = \sum_{i=1}^{N} -\delta_i \theta^T x_i(u_i) + \int_{\tau_i}^{u_i} \exp(\theta^T x_i(s)) ds\,. \tag{1}$$

We call this function the complete log-likelihood and the proof of this proposition is detailed in Appendix A.

2.4 Sequential likelihood optimization

We consider a horizon time n and a time partition $(t-1, t]$ with discrete time $t = 1, 2, \ldots$ that is independent of the observations $(u_i, \delta_i)_{1 \leq i \leq N}$. In many real-life situations, data continues to evolve; new patients may arrive, some patients may leave, and the optimization algorithm may need to update its estimation as new information becomes available. This is the focus of our work: to update online convex optimization algorithms for sequential survival data.

For individual i we define $y_{it} := \delta_i \mathbb{1}\{t - 1 < u_i \leq t\}$ which indicates whether an event is observed for individual i during the interval $(t-1, t]$ or not. Additionally, we denote the

risk indicator as $r_{it} := \mathbb{1}\{\tau_i \leq t, u_i > t-1\}$ for event i in the interval $(t-1, t]$. Then, we define the log-likelihood on the interval $(t-1, t]$ by the expression

$$\ell_t(\theta) := \sum_{i=1}^{N} -y_{it}\theta^T x_i(u_i) + r_{it} \int_{\tau_i \vee (t-1)}^{u_i \wedge t} \exp(\theta^T x_i(s))ds\,, \qquad \theta \in \Theta,\ t = 1, 2, \dots\,, \quad (2)$$

where we remind $u_i \wedge t = \min\{u_i, t\}$ and $\tau_i \vee (t-1) = \max\{\tau_i, t-1\}$. Let us notice that, analogous to Equation (1), the contribution to the log-likelihood of an individual that experiences an event in the interval $(t-1, t]$ is given by $\theta^T x_i(u_i) + \int_{\tau_i \vee (t-1)}^{u_i} \exp(\theta^T x_i(s))ds$, and the contribution of an individual that is censored in the interval $(t-1, t]$—either by $u_i = c_i$ or by t—is $\int_{\tau_i \vee (t-1)}^{u_i \wedge t} \exp(\theta^T x_i(s))ds$. If an individual is not yet present in the interval, i.e., $\tau_i > t$, or its observed time has passed before the beginning of the interval $(u_i \leq t-1)$, its contribution to the log-likelihood is zero.

Finally, the log-likelihood up to time n is given by:

$$\ell^n(\theta) := \sum_{t=1}^{n} \ell_t(\theta), \qquad \theta \in \Theta.$$

It is important to notice that if n is sufficiently large, i.e., $n \geq u_i$ for every $1 \leq i \leq N$, and all the events have been observed, the complete log-likelihood of Equation (1) corresponds to the sum of all the interval contributions. Therefore, $\ell(\theta) = \ell^n(\theta)$, $\theta \in \Theta$, for n sufficiently large when N is finite.

3 ONLINE CONVEX OPTIMIZATION

3.1 Setting

A convex optimization problem consists of approximating the minimum of a convex function over a convex set. This problem can be extended to a recursive setting where, at each iteration t, a convex optimization algorithm predicts the parameter θ_t and incurs a loss of $\ell_t(\theta_t)$. This approach is particularly good in situations where the data evolves over time, requiring fast adaptation and decision making. We apply this methodology to survival analysis, introducing a novel perspective in a field traditionally dominated by batch processed data.

The online convex optimization algorithm aims to minimize its regret at any horizon time $n \geq 1$:

$$Regret_n := \sum_{t=1}^{n} \ell_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{n} \ell_t(\theta).$$

In this paper, we aim to optimize the losses $\ell_t(\theta)$ from Equation (2). To apply online convex optimization algorithms, we must first assume that $\Theta \subseteq \mathbb{R}^d$ is a non-empty, convex, bounded, and closed set. Subsequently, we verify the convexity of the objective function. Here the choice of the response function h is crucial. For $h(x) = \exp(x)$ the cost function $\ell_t(\theta)$ is defined in Equation (2) for every iteration t. We derive its gradient and Hessian:

$$\nabla \ell_t(\theta) = \sum_{i=1}^{N} -y_{it} x_i(u_i) + r_{it} \int_{\tau_i \vee t-1}^{u_i \wedge t} \exp(\theta^T x_i(s)) x_i(s) ds, \tag{3}$$

and

$$\nabla^2 \ell_t(\theta) = \sum_{i=1}^{N} r_{it} \int_{\tau_i \vee t-1}^{u_i \wedge t} \exp(\theta^T x_i(s)) x_i(s) x_i(s)^T ds \succcurlyeq 0. \tag{4}$$

The positive semi-definite Hessian confirms the convexity of the losses. Additionally, we formalize the boundedness assumption.

Assumption 1 (Bounded domain and gradient)**.** *There exists $D, G > 0$ such that for all $t = 1, 2, \ldots$ and $\theta \in \Theta$, $\|\theta\| \leq D$ and $\|\nabla \ell_t(\theta)\| \leq G$.*

One of the most ancient algorithms for online convex optimization is named "follow the leader" (FTL), and it consists of choosing, at each iteration t, the point that optimizes the cumulative loss up to $t-1$. This algorithm does not satisfy any non-trivial regret guarantee for linear losses. However, under some modifications, like the randomized version proposed by Hannan [10], it can achieve an $\mathcal{O}(\sqrt{n})$ regret bound. Additionally, the approach from Cesa-Bianchi and Lugosi [2], where the losses are strongly convex, achieves a logarithmic regret in the number of iterations.

In 2003, Zinkevich [31] proposed a sequential version of the gradient descent algorithm (OGD), which satisfies a uniform regret bound of $\mathcal{O}(\sqrt{n})$ for an arbitrary sequence of

convex cost functions and under the previous conditions (bounded gradients and domain). Later, Hazan et al. [12] proved that Zinkevich's algorithm attains a $\mathcal{O}(\log(n))$ regret for an arbitrary sequence of strongly convex functions (with bounded first and second derivatives). They also introduced an online version of the Newton-Raphson method, which they named the Online Newton Step (ONS), and demonstrated that it also achieves logarithmic regret. More algorithms and details can be found in Hazan [11].

We implement the ONS algorithm to minimize the negative log-likelihood and study the selection of its hyperparameters along with its regret bounds.

3.2 Exp-concavity and directional derivative condition

To ensure a logarithmic regret bound, the loss function must satisfy specific conditions. First, we review the definition of exp-concavity.

Definition 2. *(Exp-concavity) A convex function $\ell : \Theta \to \mathbb{R}$ is μ-exp-concave iff the function $p(\theta) := \exp(-\mu\ell(\theta))$ is concave.*

This property is fundamental in the regret analysis and replaces the strong convexity condition required by the OGD algorithm. This means that the ONS algorithm requires a weaker hypothesis on the losses $(\ell_t)_{t=1,2,\dots}$, to achieve logarithmic regret. Furthermore, we introduce a study based on this weaker condition, which is essential to derive the regret bound described by Hazan [11] in survival analysis.

Definition 3. *(Directional derivative condition – DDC) We say a function $\ell : \Theta \to \mathbb{R}$ satisfy the directional derivative condition for a constant $\gamma > 0$ if for any pair $\theta_1, \theta_2 \in \Theta$*

$$\ell(\theta_2) \geq \ell(\theta_1) + \nabla\ell(\theta_1)(\theta_2 - \theta_1) + \frac{\gamma}{2}\left(\nabla\ell(\theta_1)(\theta_2 - \theta_1)\right)^2. \tag{DDC}$$

To determine the directional derivative constant γ, we must first compute the exp-concavity constant μ.

Lemma 1. *A twice differentiable function $\ell : \Theta \to \mathbb{R}$ is μ-exp-concave iff*

$$\nabla^2\ell(\theta) \succcurlyeq \mu\nabla\ell(\theta)\nabla\ell(\theta)^T, \qquad \theta \in \Theta. \tag{5}$$

This holds with

$$\mu \leq \min_{\theta \in \Theta} \frac{\nabla \ell(\theta)^T \nabla^2 \ell(\theta) \nabla \ell(\theta)}{||\nabla \ell(\theta)||^4}.$$

This lemma provides us a way for calculating the exp-concavity constant μ. The proof of Lemma 1 can be found in Appendix B. Given a μ-exp-concave function ℓ, we can also determine its directional derivative constant γ. We have the following bound:

Lemma 2. *A μ-exp-concave function $\ell : \Theta \to \mathbb{R}$, satisfying Assumption 1, admits a directional derivative constant $\gamma > 0$ satisfying*

$$\gamma \leq \min_{\theta \in \Theta} \frac{-\frac{2}{\mu} \log(1 + \mu ||\nabla \ell(\theta)|| D) + ||\nabla \ell(\theta)|| D}{(||\nabla \ell(\theta)|| D)^2}.$$

We note that this lower bound improves upon the upper bound provided by Hazan [11]:

$$\gamma \leq \frac{1}{2} \min \left\{ \frac{1}{GD}, \mu \right\},$$

and the proof of Lemma 2 can also be found in Appendix B.

3.3 Online Newton Step

The ONS algorithm is an online analogue of the Newton-Raphson method; see Ypma [30] for more details. The Newton-Raphson algorithm moves in the direction of the inverse of the Hessian multiplied by the gradient. For exp-concave loss functions ℓ_t with $t = 1, 2, \ldots$, we can replace the Hessian matrix with an approximation of it:

$$A_t = \sum_{k=1}^{t} \nabla \ell_k(\theta_k) \nabla \ell_k(\theta_k)^T.$$

At each iteration, the algorithm updates the estimation of the parameter as follows:

$$\theta_{t+1} = \theta_t - \frac{1}{\gamma} A_t^{-1} \nabla \ell_t(\theta_t),$$

where γ is an algorithm hyperparameter denoting the learning rate and its optimal selection aligns with the DDC constant. This might lead to a point outside the convex

set Θ and so we need to project it back. This projection is somewhat different than the standard projection as it is characterized by the norm defined by A_t instead of the Euclidean norm. The iteration step of the algorithm is:

$$\theta_{t+1} = \text{Proj}_t\Big(\theta_t - \frac{1}{\gamma} A_t^{-1} \nabla \ell_t(\theta_t)\Big)\,,$$

where $\text{Proj}_t(\theta^*) \in \arg\min_{\theta \in \Theta} (\theta - \theta^*)^T A_t (\theta - \theta^*)$.

Let us remark that ONS requires to invert a large matrix A_t, and in order to avoid expensive calculations, we consider the Sherman-Morrisson formula [25] which provides a recursion for A_t^{-1} from $A_0^{-1} := (1/\epsilon) I_d$:

$$A_t^{-1} = A_{t-1}^{-1} - \frac{A_{t-1}^{-1} \nabla \ell_t(\theta_t) \nabla \ell_t(\theta_t)^T A_{t-1}^{-1}}{1 + \nabla \ell_t(\theta_t) A_{t-1}^{-1} \nabla \ell_t(\theta_t)^T}\,, \qquad t = 1, 2, \ldots.$$

We formally describe the Online Newton Step algorithm 2 in Appendix B. Hazan [11] proved the following regret bound of ONS.

Theorem 1 (Hazan [11])**.** *Let us consider the losses $\ell_t : \Theta \to \mathbb{R}$ μ-exp-concave and satisfying Assumption 1. Then, Algorithm 2 with hyperparameters $\gamma = \frac{1}{2}\min\{\frac{1}{GD}, \mu\}$ and $\epsilon = (\gamma D)^{-2}$ satisfies $Regret_n \leq \gamma^{-1} d \log(2nG^2\gamma^2 D^2)$ for any $n \geq 4$.*

Let us remind that we want to apply ONS algorithm to the losses $(\ell_t)_{t=1,2,\ldots}$ described in Equation (2), where we assume the exponential model defined in 1. The exp-concavity property is fundamental in the regret analysis of ONS. We first notice that we can work under (DDC) rather than μ-exp-concavity, focusing our work on the study of the constant γ which is the hyperparameter of ONS, rather than on μ. Then we see that the choice of this constant is very sensitive to variations in the gradients, which depend on the number of people at risk at each time. If μ is small, which can happen when the gradient of the loss is small, the choice of $\gamma = \frac{1}{2}\min\{\frac{1}{GD}, \mu\}$ proposed by Hazan [11] will also be small, potentially exploding the regret bound and causing issues with the convergence. Additionally, to properly tune the hyperparameter γ we need to know the exp-concavity constant in advance, but this constant might depend on the gradient of the losses that are not known before running the algorithm. Adjusting γ is not trivial and we provide some insights in the following sections.

4 STOCHASTIC SETTING

The first solution we propose is to use a stochastic approach to bound the regret of ONS. We present the general stochastic setting introduced by Wintenberger [29] and apply one of its results to the survival case. The main difficulty in sequential survival analysis is the intrinsic time dependence in the loss functions $(\ell_t)_{t=1,2,\dots}$. Indeed, even if the individuals are iid, the log-likelihoods ℓ_t are dependent because of the individuals that are at risk during consecutive time intervals $(t-1,t]$ for $t = 1, 2, \dots$.

4.1 Stochastic Model

We model the arrival times $\tau_i \geq 0$ as a homogeneous Poisson process with intensity λ; see Kingman [20] for a reference textbook on the subject. For each $t > 0$, we define the count random variable $N_t := \sum_{i=1}^{\infty} \mathbb{1}\{\tau_i \leq t\}$ which represents the number of individuals that arrive before t, and τ_{N_t} represents the arrival time of the last individual arriving before t. We assume a constant rate λ, such that $\mathbb{E}[N_t] = \lambda t$, indicating the average number of individuals arriving at time t. Additionally, in this section, we consider the covariate functions to be constant, i.e., $x_i(t) = x_i$ for all $t > 0$, and that they follow, independently, the distribution of a random variable X. In this stochastic setting we rewrite the loss function:

$$\ell_t(\theta) = \sum_{i=1}^{N_t} -y_{it}\theta^T x_i + r_{it}\exp(\theta^T x_i)((u_i \wedge t) - (\tau_i \vee (t-1))), \tag{6}$$

where we replaced N by N_t in Equation (2). It is important to note that the derivation of this expression is based on the assumption of the exponential model 1. Now, we want to apply ONS to optimize this loss and study what happens with its stochastic regret.

For each iteration $t = 1, 2, \dots$, we consider the stochastic loss ℓ_t and the filtration $\mathcal{F}_t$ of σ-algebras such that the predictions of the online learning algorithm θ_t and the past losses $(\ell_s)_{s=1}^{t-1}$ are $\mathcal{F}_{t-1}$-measurable. To simplify notation, we use $\mathbb{E}_t[\cdot]$ to represent the conditional expectation given $\mathcal{F}_t$, denoted as $\mathbb{E}[\cdot|\mathcal{F}_t]$. In this context, our objective is to

minimize the stochastic regret at any horizon time $n \geq 1$:

$$Risk_n := \sum_{t=1}^{n} L_t(\theta_t) - \min_{\theta \in \Theta} \sum_{t=1}^{n} L_t(\theta)\,,$$

where $L_t(\theta_t)$ is the conditional risk, defined as $L_t(\theta_t) := \mathbb{E}_{t-1}[\ell_t(\theta_t)]$ for $t = 1, 2, \ldots$. Let us notice that in our case, where the stochastic losses ℓ_t are defined in Equation (6), the σ-algebra $\mathcal{F}_t$ is generated by y_{is}, x_i, τ_i, and $u_{is} = \min\{u_i, s\}$ for all $i = 1, \ldots, N_{t-1}$ and $s = 1, \ldots, t-1$.

The main difference with the setting presented in Section 3 is the use of the conditional risk L_t instead of the loss functions ℓ_t in the calculation of regret. This allows us to relax the convexity conditions imposed on ℓ_t and instead focus on the convexity properties of L_t.

4.2 Stochastically Exp-Concavity

It was proved in Wintenberger [29] that the ONS algorithm achieves a $\mathcal{O}(\log(n))$ stochastic regret bound under a stochastic exp-concavity condition for ℓ_t which is described below.

Definition 4 (Stochastic exp-concavity)**.** *A sequence of random functions $(\ell_t)_{t=1,2,\ldots}$ is said to be γ stochastically exp-concave with respect to a filtration $\mathcal{F}_t$ if for all $\theta_1, \theta_2 \in \Theta$ and $t = 1, 2, \ldots$*

$$L_t(\theta_1) \leq L_t(\theta_2) + \nabla L_t(\theta_1)^T(\theta_1 - \theta_2) - \frac{\gamma}{2}\mathbb{E}_{t-1}\left[(\nabla \ell_t(\theta_1)^T(\theta_1 - \theta_2))^2\right], \qquad a.s.$$

Let us note that this property corresponds to the stochastic counterpart of the directional derivative condition (DDC). This property plays a crucial role in the proof of Theorem 7 of Wintenberger [29], which establishes the logarithm stochastic regret bound. However, the losses ℓ_t defined in (6) do not satisfy this property. Nevertheless, we demonstrate that the events where this inequality is not fulfilled have a small probability and therefore, we can still bound the stochastic regret. In addition, we need to make the following design assumption.

Assumption 2. *There exist* $A > 0$ *such that* $\mathbb{E}[xx^\top \mathbb{1}\{T \leq C\}(1-T)_+ | \tau = 0] \succcurlyeq AI_d$.

This assumption is not trivial, and it is not always satisfied; however, when all the individuals experience an event and $T \leq 1$, it corresponds to a classical design. When $t \geq 1$ an alternative analyses is required.

4.3 Stochastic Regret

To apply Theorem 7 from Wintenberger [29], the losses need to satisfy certain hypothesis, among which are stochastic exp-concavity and a stochastic bound on the gradients of the losses. We prove that our losses ℓ_t, whose do not satisfy exactly the conditions of Theorem 7, still leads ONS algorithm to achieve a logarithmic stochastic regret. We present the result in the following theorem.

Theorem 2. *Given* $\varrho > 0, n \geq 1$ *and the stochastic losses* $(\ell_t)_{t=1,2,\dots}$ *from Equation* (6)*, then under Assumption 2, a bounded domain of diameter* D *and hyperparameter* γ*, the stochastic exp-concavity constant, the ONS algorithm has logarithmic stochastic regret with probability* $1 - 4\varrho$*. Specifically,* $Risk_n = \mathcal{O}(\log(n/\varrho)/\gamma)$, $n \geq 1$.

The proof of Theorem 2 can be found in Appendix C and the explicit regret bound in Equation (11). To finish, we prove the following corollary.

Corollary 1. *Given* $\varrho > 0, n \geq 1$ *and the stochastic losses* $(\ell_t)_{t=1,2,\dots}$ *from Equation* (6)*, we consider* θ_t *the ONS prediction at time* t *and* $\bar{\theta}_n$ *the average prediction* $\bar{\theta}_n = \frac{1}{n}\sum_{t=1}^n \theta_t$*. Defining the optimal parameter*

$$\theta^* = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} L_t(\theta),$$

then, under Assumption 2, a bounded domain of diameter D *and hyperparameter* γ*, with probability* $1 - 4\varrho$ *we have:*

$$||\bar{\theta}_n - \theta^*||^2 \leq \mathcal{O}\left(\frac{\log(n/\varrho)}{\gamma n}\right), n \geq 1.$$

This corollary ensures the convergence of the algorithm predictions to the real parameter, which is possible thanks to the strong convexity of the risk functions L_t. It is important to remark that this does not hold in the deterministic setting. The proof can be found in Appendix C and the explicit bound in Equation (12).

5 SURVIVAL ONS ALGORITHM

As mentioned earlier, the choice of γ has a significant influence on the algorithm's performance, particularly regarding the regret bound. To avoid convergence issues and address the challenge posed by the small optimal constant proposed by Hazan [11], we propose an adaptive setting that allows us to select the most suitable learning rate at each step while maintaining control over the regret bound. We introduce SurvONS (Algorithm 1), a survival version of MetaGrad from van Erven et al. [27], that uses Bernstein Online Aggregation (BOA, introduced in Wintenberger [28]) to aggregate multiple ONS applied to an adaptive auxiliary function. SurvONS strategically selects larger learning rates to handle sub-optimal parameters. The key difference between our algorithm and MetaGrad lies in the approach to updating the adaptive learning rate. We explain this algorithm in detail throughout this section.

5.1 Recursive adaptation to the constants

We present first the recursive adaptation of the constants μ and γ. Lemma 1 provides a bound for the exp-concavity constant μ, and Lemma 2 offers a bound for the directional derivative constant γ based on μ. We aim to apply this approach to $\ell = \ell_t$ for all $t = 1, 2, \ldots$, and recursively obtain μ_t and $\gamma_t(\mu_t)$ such that they satisfy the bounds of Lemma 1 and Lemma 2.

In Hazan's approach, as described in [11], the idea is to select a universal constant μ that renders all the functions $(\ell_t)_{t=1,2,\ldots,}$ μ-exp-concave. The natural choice would be to take:

$$\mu := \min_{t\in\{1,\ldots,n\}} \mu_t^*, \qquad \text{where} \qquad \mu_t^* := \min_{\theta\in\Theta} \frac{\nabla\ell_t(\theta)^\top \nabla^2\ell_t(\theta)\nabla\ell_t(\theta)}{||\nabla\ell_t(\theta)||^4},$$

is the bound given by Lemma 1. With this configuration, we guarantee exp-concavity for every function. However, the challenge of minimizing over the parameter set Θ in the definition of μ_t^* might be more intricate than minimizing the loss function ℓ_t. In addition, we can not know the constant in advance because ℓ_t is revealed at the t-th iteration only in our online setting.

To solve this problem, we define at each time $t = 1, 2, \ldots$ an adaptive estimation of the exp-concavity constant:

$$\mu_t := \frac{\nabla \ell_t(\theta_t)^\top \nabla^2 \ell_t(\theta_t) \nabla \ell_t(\theta_t)}{||\nabla \ell_t(\theta_t)||^4},$$

and, similarly from Lemma 2,

$$\gamma_t(\mu_t) := \frac{-\frac{2}{\mu_t} \log(1 + \mu_t ||\nabla \ell_t(\theta_t)|| D) + 2||\nabla \ell_t(\theta_t)|| D}{(||\nabla \ell_t(\theta_t)|| D)^2},$$

where θ_t is the parameter predicted by the algorithm at time t. Let us notice that this choice of $\mu_t \geq \mu$ and $\gamma_t(\mu) \geq \gamma(\mu)$ assures the exp-concavity and the directional derivative condition for ℓ_t close to θ_t at time t. We sometimes refer to $\gamma_t(\mu_t)$ as γ_t when the specification is not necessary.

5.2 SurvONS

Now, we have an adaptive way to choose μ_t and γ_t that preserves the exp-concavity properties at each iteration. However, this choice might not be optimal, in some iterations the gradient $\nabla \ell_t$ can be close to zero due to the lack of individuals at risk, and this might lead to numeric problems setting μ_t and γ_t. Thus we propose an intermediate choice of the learning rate. Given a user specified constant $\gamma > 0$, we define for each time $t = 1, 2 \ldots$:

$$\tilde{\gamma}_t := \max\{\gamma_t(\mu_t)/4, \gamma\},$$

a value that chooses a portion of the optimal directional derivative condition constant $\gamma_t/4$ when it is not too small, and the user specified constant γ when the quarter of the

optimal constant decreases under γ. This choice $\tilde{\gamma}_t$ is a trade off in between choosing the optimal directional derivative condition constant and a worse constant when the optimal one is susceptible to bring convergence problems.

In order to keep the logarithmic regret bound we cannot directly use the adaptive choice of the constant as the algorithm's learning rate. Instead, it was proposed by van Erven et al. [27] to optimize an adaptive auxiliary function. Let us consider $\hat{\theta}$ such that $\ell_t(\hat{\theta})$ and $\nabla\ell_t(\hat{\theta})$ have been observed and $\gamma > 0$, we define the directional derivative function:

$$\hat{\ell}_{t,\gamma}(\theta) := \ell_t(\hat{\theta}) + \nabla\ell_t(\hat{\theta})(\theta - \hat{\theta}) + \frac{\gamma}{2}\left(\nabla\ell_t(\hat{\theta})(\theta - \hat{\theta})\right)^2, \qquad \theta \in \Theta\,, \qquad t = 1, 2, \dots\,. \tag{7}$$

We prove that this function satisfies the directional derivative condition for a different constant $\hat{\gamma}$.

Lemma 3. *Let $\gamma > 0$, $\Theta \subseteq \mathbb{R}^d$ of diameter $D > 0$, $\hat{\theta} \in \Theta$ and $\ell_t : \Theta \to \mathbb{R}$ the log-likelihood defined in Equation (2). Then, the function $\hat{\ell}_{t,\gamma}$ from (7) satisfies for every $\theta_1, \theta_2 \in \Theta$:*

$$\hat{\ell}_{t,\gamma}(\theta_2) \geq \hat{\ell}_{t,\gamma}(\theta_1) + \nabla\hat{\ell}_{t,\gamma}(\theta_1)(\theta_2 - \theta_1) + \frac{\gamma}{2(1 + \tilde{\gamma}\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}))^2}\left(\nabla\hat{\ell}_{t,\gamma}(\theta_1)(\theta_2 - \theta_1)\right)^2,$$

and thus, the function $\hat{\ell}_{t,\gamma}$ has directional derivative constant $\hat{\gamma}$ with $\hat{\gamma} := \frac{\gamma}{2(1+\gamma D||\ell_t(\hat{\theta})||)^2}$.

The proof of Lemma 3 is presented in Appendix D. The idea of the algorithm is to use ONS routine to optimize the functions $\hat{\ell}_{t,\gamma} = \hat{\ell}_{t,\tilde{\gamma}_t}$, i.e., the auxiliary function with $\gamma = \tilde{\gamma}_t$, which adapt at each step according to the current optimal γ_t and the algorithm predictions θ_t.

In addition, to obtain an algorithm that is robust for the choice of the learning rate, we propose an aggregation procedure which applies ONS and combines it with multiple choices of the learning rate γ. To formalize this idea, we consider a grid $\Gamma = \{\gamma_i\}_{i=1,\dots,K}$ and $\mathcal{E} = \{\epsilon_i\}_{i=1,\dots,K}$ such that $\epsilon_i = \frac{1}{(\gamma_i D)^2}$ for all $i = 1, \cdots, K$. Then, at each iteration $t = 1, \dots, n$, and for each $i = 1, \dots, K$, we define $\tilde{\gamma}_{it} = \max\{\gamma_t/4, \gamma_i\}$ and we aggregate

ONS applied to $(\hat{\ell}_{t,\tilde{\gamma}_{it}})_{t=1,2,\dots}$. The aggregation is held by BOA algorithm of Wintenberger [28], which is a recursive procedure that considers exponential weights with a second order refinement. The algorithm SurvONS is described in Algorithm 1 and it is important to notice that the difference between SurvONS and MetaGrad is the choice of the constant $\tilde{\gamma}$.

Aggregation methods allow us to avoid bad choices of γ and therefore, the convergency issues. Let us remind that we consider the exponential model 1. We prove that the regret of Algorithm 1 is bounded.

Theorem 3. *Let $n \geq 1$, $(\ell_t)_{t=1,\dots,n}$ be the sequence of losses defined in* (2)*, that are assumed to satisfy Assumption 1 and* (DDC) *with constants $\gamma_t \in (0, 1/GD)$. Let $K \geq 1$ and $\Gamma \in (0, 1/(4GD))^K$. Then, Algorithm 1 with hyperparameters Γ and $\mathcal{E} = 1/(\Gamma D)^2$, satisfies the regret upper-bound:*

$$Regret_n \leq \min_{\gamma \in \Gamma} \left\{ \frac{2\log(K) + 5d\log(n)}{\gamma} + \gamma G^2 D^2 n_\gamma \right\},$$

where $n_\gamma := \sum_{t=1}^{n} \mathbb{1}\{\gamma_t < \gamma\}$, $\gamma > 0$.

This theorem provides a regret bound that proposes a trade-off between the bad choices of γ and the frequency with which the algorithm selects γ over γ_t, thereby compensating for the regret increment. The proof of Theorem 3 can be found in Appendix D. Let us notice that this analysis is also valid for MetaGrad algorithm [27], and Theorem 3, which was developed for the survival losses (2), holds equally true for any loss satisfying Assumption 1 and (DDC).

5.3 Theoretical regret bounds comparison

We show in Figure 1 the differences between the regret bound orders, in order to illustrate the importance of the constant adaptation $\tilde{\gamma}_t$ in SurvONS, and the interest of the stochastic setting. We compare the theoretical regret bound orders of ONS [12] with the optimal hyperparameter γ_t, OGD [31], SurvONS 1, and ONS with an average hyperparameter $\bar{\gamma}_t = \sum_{s=1}^{t} \gamma_s$, representing the stochastic approach. The bounds are detailed in Table 1.

Table 1: Regret bound order after n iterations (up to logarithmic factors)

	OGD	ONS	SurvONS	ONS($\bar{\gamma}$)
Regret bound	$\sqrt{n}$	$\frac{1}{\min_{1\leq t\leq n}\gamma_t}$	$\min_\gamma\{\frac{1}{\gamma}+n_\gamma\}$	$\frac{1}{\bar{\gamma}_n}$

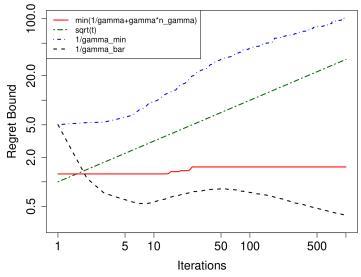

Figure 1: Regret bound orders (up to logarithmic factors)

In this comparison, we omit constants and logarithmic terms. We estimate γ_t with SurvONS, and we use these estimations to construct the bounds. The simulation framework for this experiment is detailed in Section 6. The graph is presented in log-log scale.

Figure 1 traces the regret behavior of the different algorithms (see Table 1). Without an explicit calculation of the stochastic constant, we show the interest of considering an average case through plotting the average constant $\bar{\gamma}_t$. We observe that although in theory, the bound of ONS appears better than the bound of OGD ($\mathcal{O}(\log(n)/\gamma)$ v/s $\mathcal{O}(\sqrt{n})$), when γ_t goes to 0, the bound of ONS is not $\mathcal{O}(\log(n))$, but $\mathcal{O}(\log(n)/\min_t\gamma_t)$. A similar finding in logistic regression has been made rigorous by Hazan et al. [13] with the help of lower bounds matching $\mathcal{O}(\log(n)/\min_t\gamma_t)$. In practical applications, it is essential to consider more detailed analyses that remain robust in scenarios where $\min_t\gamma_t$ goes to 0, which is what we propose with SurvONS and the stochastic approach.

6 SIMULATION EXPERIMENTS

In this section we present simulation results of our method. We considered a number of individuals $N = 10\,000$ and a number of iterations $n = 1\,000$. Then we sample a

multivariate random normal of dimension $(N, d-1)$ with $d = 4$ and mean vector and covariance matrix:

$$\eta := \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \Sigma := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We add an intersect column that transforms the matrix into one of dimension (N, d). This matrix corresponds to the covariates information $\{x_i\}_{i=1}^N$, which does not depend on time. The real parameter θ^* is set randomly following a $\mathcal{N}(0, I_d)$ distribution. We sample the arrival times τ_i as a uniform between 0 and n and we simulate T_i and C_i following an exponential distribution of rate $\exp(\theta^{*T} x_i)$,

$$T_i \sim \tau_i + \exp(\exp(\theta^{*T} x_i)), \qquad C_i \sim \tau_i + \exp(\exp(\theta^{*T} x_i)).$$

For more details on the common use of exponential distributions in survival analysis we refer to Selvin [24]. We repeat this procedure 100 times, and the results are the average curves over the 100 data simulations. Additionally, we consider two exponential grids for the aggregation methods of size $K = 10$. First, we test a random grid, consider the SurvONS predictions θ_t for $t = 1, \ldots, n$, and define:

$$G := \max_{t=1,\ldots,n} ||\nabla \ell_t(\theta_t)|| \,.$$

This process is repeated multiple times to ensure the stability of G estimation. Second, we choose 10 equidistant points and then we generate the grids by considering the exponential of each point, such that:

$$\Gamma_1 := (1/\sqrt{n}, \ldots, 1/4GD) \,, \qquad \Gamma_2 := (1/GD, \ldots, 10/GD) \,,$$

where D is adjusted a posteriori such as $D := 1.1||\theta^*||$. Throughout this section we compare the results of the two choices of grid Γ_1 and Γ_2.

We observe in Figure 2 the distribution of the average γ_t estimations that we obtained from SurvONS. The average for Γ_1 is 1.24 and 1.64 for Γ_2. The similarity between both estimations elucidates the proximity of the graphs in Figure 1, which is unsurprising

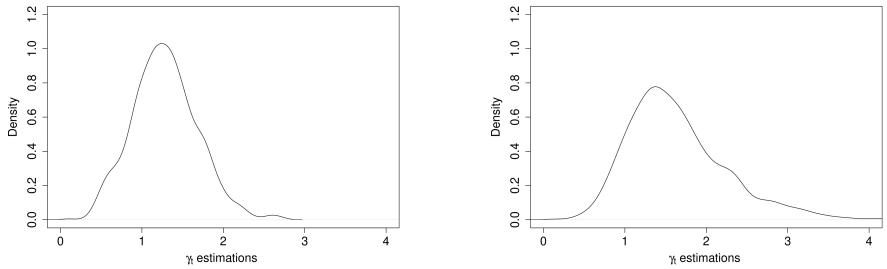

Figure 2: Density of γ_t estimation obtained by Algorithm 1, with Γ_1 [left] and Γ_2 [right]

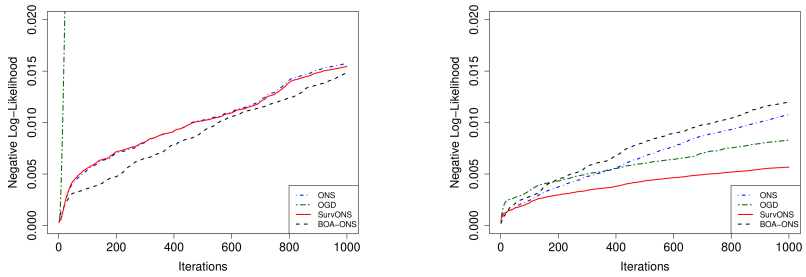

Figure 3: Cumulative negative log-likelihood with hyperparameters in grid Γ_1 [left] and Γ_2 [right]

given that the directional derivative constant is inherent to the loss function and does not depend on the algorithm or the selected grids.

We compare SurvONS, described in Algorithm 1, with the BOA-ONS proposed by Wintenberger [28]. Additionally, we fit several ONS and OGD with constant learning rate equal to each γ in the grid, and then we select the one that performs better to include in the comparison. Remark that this procedure overestimates the performances of ONS and OGD. We show the average cumulative difference between the negative log-likelihood of the estimations and the real parameters in Figure 3.

In Figure 3, we observe that SurvONS (in purple) does not outperform BOA-ONS (in black) with the Γ_1 grid. However, the scenario changes with the second grid, Γ_2, where SurvONS proves to be more effective than the other methods. This unexpected result

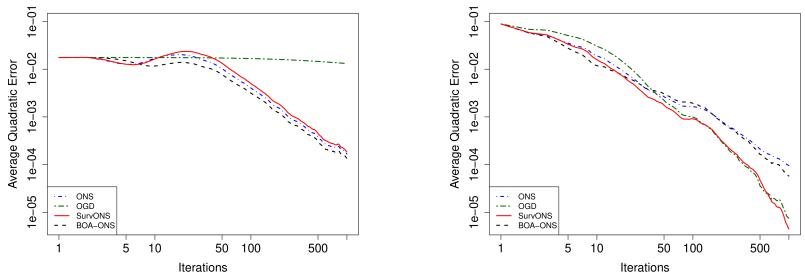

Figure 4: Average quadratic error with hyperparameters in Γ_1 [left] and Γ_2 [right]

arises from the fact that the Γ_1 grid falls within the theoretical limits. Nevertheless, we observe a consistent improvement in performance for all algorithms when considering a larger grid. This discrepancy could arise from either an overestimation of the constant G or the presence of outlier points exhibiting extremely large gradients. Nonetheless, given the similarity in the constant γ_t estimation across the two grids, shown in Figure 2, we recommend opting for larger grids, ranging from 4 to 40 times the theoretical bound of $1/4GD$.

In addition, Figure 4 presents the quadratic error, where we consider the cumulative average of the estimations. Specifically, given a sequence of algorithm predictions $(\theta_s)_{s=1}^t$, the cumulative average is defined as $\bar{\theta}_t := t^{-1}\sum_{s=1}^t \theta_s$. Let us remind that the curves depicted represent the average of 100 instances obtained from simulating 100 datasets. The figure is in log-log scale.

Figure 4 corroborates the result of Corollary 1, which establishes the convergence of the estimations θ_t to the real parameter θ^* when using the ONS algorithm. The findings of Wintenberger in [29], which demonstrate the $\mathcal{O}(\log(n))$ stochastic regret of BOA-ONS, together with the insights from Figure 4, suggest the potential to extend a similar corollary to both BOA-ONS and SurvONS. Furthermore, Corollary 1 can be easily extended to BOA-ONS by replacing the application of Theorem 2 with Theorem 4 from [29].

7 CONCLUSIONS

In this paper, we presented a detailed mathematical framework for online survival data, analyzing the regret of Online Newton Step and its sensitivity to the learning rate. Notably, we found that tuning this parameter is challenging, and the regret bound is highly sensitive to its adjustment. Our first contributions is introducing a stochastic setting to ensure that ONS achieves logarithmic stochastic regret in the survival context. Additionally, we proposed an adaptive method, SurvONS, which aggregates ONS with different learning rates. Adaptive methods, commonly used in first-order algorithms like AdaGrad [8] or Adam [19], offer a promising avenue for enhancing second-order algorithms. Our approach leverages adaptive strategies to improve efficiency and convergence, extending its applicability beyond the online survival domain. The regret analysis of SurvONS strategically selects larger learning rates to address sub-optimal parameters. In conclusion, aggregation methods enhance robustness in selecting algorithm hyperparameters; however, achieving and maintaining fast rates remains a non-trivial task.

Finally, in the simulation experiments, we compared two grid choices. Figure 2 shows that γ_t estimations closely align within the grids, and the second grid produces values that do not approach zero to the same extent as the first grid. Additionally, Figure 3 indicates that choosing larger values for the learning rate grid accelerates convergence, suggesting the preference for larger grids.

Algorithm 1 SurvONS

Input: $(\ell_t)_{t=1,2,\ldots}, D > 0, n \geq 1$, grids $\Gamma, \mathcal{E}$

Initialization: for each γ_k in Γ: $\theta_0(\gamma_k) \in \Theta$, $\pi_{0,k} = \frac{1}{K}$, $\hat{\theta}_0 \in \Theta$, $A_0^{-1} = \mathcal{E}^{-1}\mathbb{1}_d$

for iteration $t = 1, \ldots, n$ **do**

Update: $\hat{\theta}_t = \sum_{k=1}^{K} \pi_{t,k}\, \theta_t(\gamma_k)$

Observe:
$$\nabla \ell_t(\hat{\theta}_t) = \sum_{i=1}^{N} -y_{it} x_i(u_i) + r_{it} \int_{\tau_i \vee t-1}^{u_i \wedge t} \exp(\hat{\theta}_t^T x_i(s)) x_i(s) ds$$
$$\nabla^2 \ell_t(\hat{\theta}_t) = \sum_{i=1}^{N} r_{it} \int_{\tau_i \vee t-1}^{u_i \wedge t} \exp(\hat{\theta}_t^T x_i(s)) x_i(s) x_i(s)^T ds$$
$$\mu_t = \frac{\nabla \ell_t(\hat{\theta}_t)^T \nabla^2 \ell_t(\hat{\theta}_t) \nabla \ell_t(\hat{\theta}_t)}{||\nabla \ell_t(\hat{\theta}_t)||^4}$$
$$\gamma_t = 2\frac{-\frac{1}{\mu_t}\log(1+\mu_t||\nabla \ell_t(\hat{\theta}_t)||D) + ||\nabla \ell_t(\hat{\theta}_t)||D}{(||\nabla \ell_t(\hat{\theta}_t)||D)^2}$$

for $\gamma_k \in \Gamma$ **do**

Observe:
$$\tilde{\gamma}_t = \max\{\gamma_t/4, \gamma_k\}$$
$$\nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k)) = \nabla \ell_t(\hat{\theta}_t)(1 + \tilde{\gamma}_t \nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma_k) - \hat{\theta}_t))$$

Recursion:

$$A_t^{-1} = A_{t-1}^{-1} - \frac{A_{t-1}^{-1} \nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k)) \nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k))^T A_{t-1}^{-1}}{1 + \nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k)) A_{t-1}^{-1} \nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k))^T}$$

$$\theta_{t+1}(\gamma_k) = \mathrm{Proj}_t\left(\theta_t(\gamma_k) - \frac{1}{\gamma_k} A_t^{-1} \nabla \hat{\ell}_{t,\tilde{\gamma}_t}(\theta_t(\gamma_k))\right)$$

end for

Update: $\pi_{t+1,\cdot} = \pi_t \exp\left(-\Gamma \nabla \ell_t(\hat{\theta}_t)^T(\hat{\theta}_t - \theta_t(\Gamma)) - \Gamma^2(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\Gamma)))^2\right)$

end for

return $\hat{\theta}_n$

REFERENCES

[1] Elja Arjas and Pentti Haara. A logistic regression model for hazard: asymptotic results. *Scandinavian Journal of Statistics*, pages 1–18, 1987.

[2] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge university press, 2006.

[3] Chong Chen, Ying Liu, Shixuan Wang, Xianfang Sun, Carla Di Cairano-Gilfedder, Scott Titmus, and Aris A Syntetos. Predictive maintenance using cox proportional hazard deep learning. *Advanced Engineering Informatics*, 44:101054, 2020.

[4] Benjamin Christoffersen. Dynamichazard: Dynamic hazard models using state space models. *Journal of Statistical Software*, 99:1–38, 2021.

[5] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[6] David Roxbee Cox and David Oakes. *Analysis of Survival Data*, volume 21. CRC press, 1984.

[7] Daryl J Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. I.* Springer-Verlag, 2003.

[8] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[9] Ludwig Fahrmeir. Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika*, 81(2):317–330, 1994.

[10] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3(2):97–139, 1957.

[11] Elad Hazan. *Introduction to Online Convex Optimization*. MIT Press, 2022.

[12] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

[13] Elad Hazan, Tomer Koren, and Kfir Y Levy. Logistic regression: Tight bounds for stochastic and online optimization. In *Conference on Learning Theory*, pages 197–209. PMLR, 2014.

[14] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

[15] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.

[16] John D Kalbfleisch and Ross L Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2011.

[17] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[18] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):1–12, 2018.

[19] Diederik P Kingma and Ba J Adam. A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 1412, 2014.

[20] John F Kingman. *Poisson Processes*, volume 3. Clarendon Press, 1992.

[21] Junxiang Lu. Predicting customer churn in the telecommunications industry: An application of survival analysis modeling using sas. *SAS User Group International (SUGI27) Online Proceedings*, 114:27, 2002.

[22] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge university press, 2017.

[23] Krishna Kumar Mohbey. Employee's attrition prediction using survival analysis and cox proportional hazard model. *International Journal of Digital Enterprise Technology*, 2(1):27–37, 2022.

[24] Steve Selvin. *Survival Analysis for Epidemiologic and Medical Research*. Cambridge University Press, 2008.

[25] Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[26] Gerhard Tutz and Matthias Schmid. *Modeling Discrete Time-to-Event Data*. Springer, 2016.

[27] Tim Van Erven, Wouter M Koolen, and Dirk Van der Hoeven. Metagrad: Adaptation using multiple learning rates in online learning. *arXiv preprint arXiv:2102.06622*, 2021.

[28] Olivier Wintenberger. Optimal learning with bernstein online aggregation. *Machine Learning*, 106(1):119–141, 2017.

[29] Olivier Wintenberger. Stochastic online convex optimization; application to probabilistic time series forecasting. *arXiv preprint arXiv:2102.00729*, 2021.

[30] Tjalling J Ypma. Historical development of the newton–raphson method. *SIAM review*, 37(4):531–551, 1995.

[31] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936, 2003.

A BACKGROUND ON PARAMETRIC INFERENCE

A.1 Proof of Proposition 1

Proof. We define the equivalent of the survival probability for the censored distribution

$$G(t|x_i, \tau_i) = \mathbb{P}(c_i \geq t|x_i, \tau_i)\,.$$

Given $\theta \in \Theta$, we write the density of u_i distinguishing two cases:

$$\mathbb{P}(u_i \in [t, t+h), \delta_i = 1|\theta, x_i, \tau_i) = \mathbb{P}(t_i \in [t, t+h), c_i \geq t|x_i, \tau_i, \theta),$$

and

$$\mathbb{P}(u_i \in [t, t+h), \delta_i = 0|\theta, x_i, \tau_i) = \mathbb{P}(c_i \in [t, t+h), t_i \geq t|\theta, x_i, \tau_i).$$

By conditional independence we obtain

$$\begin{aligned}
\mathbb{P}(u_i \in [t, t+h), \delta_i = 1|\theta, x_i, \tau_i) &= \mathbb{P}(t_i \in [t, t+h)|\theta, x_i, \tau_i)\mathbb{P}(c_i \geq t+h|\theta, x_i, \tau_i),\\
\mathbb{P}(u_i \in [t, t+h), \delta_i = 0|\theta, x_i, \tau_i) &= \mathbb{P}(c_i \in [t, t+h)|\theta, x_i, \tau_i)S(t+h|\theta, x_i, \tau_i).
\end{aligned}$$

When h goes to zero, it tends respectively to

$$G(t|\theta, x_i, \tau_i)f(t|\theta, x_i, \tau_i),$$

and

$$g(t|\theta, x_i, \tau_i)\mathbb{P}(t_i \geq t|\theta, x_i, \tau_i).$$

Therefore, by the independence of the random variables (t_i, c_i) among the events $i \in \{1, \ldots, N\}$ we obtain the density

$$f_{(u_i,\delta_i)_{1\leq i\leq N}}((u_i, \delta_i)_{1\leq i\leq N} \mid \theta, x_i, \tau_i) = \prod_{i=1}^{N} g(u_i|\theta, x_i, \tau_i)^{1-\delta_i} G(u_i|\theta, x_i, \tau_i)^{\delta_i} \prod_{i=1}^{N} f(u_i|\theta, x_i, \tau_i)^{\delta_i} S(u_i|\theta, x_i, \tau_i)^{1-\delta_i}\,.$$

Here we use the assumption of non-informative censoring (see Kalbfleisch et al. [16]), which means that the censored distribution does not involve the parameter θ. Then we obtain a simplified version of the likelihood, up to a multiplicative constant

$$\ell(\theta) \propto \prod_{i=1}^{N} f(u_i|\theta, x_i, \tau_i)^{\delta_i} S(u_i|\theta, x_i, \tau_i)^{1-\delta_i}\,.$$

Omitting an additional constant, we can equivalently write the log-likelihood to be

$$\log(\ell(\theta)) = \sum_{i=1}^{N} \delta_i \log(f(u_i|\theta, x_i, \tau_i)) + (1 - \delta_i) \log(S(u_i|\theta, x_i, \tau_i)).$$

Let us remark that $f(t|\theta, x_i, \tau_i) = H(t|x_i, \tau_i)S(t|x_i, \tau_i)$ and from the definition of $H(t|x_i, \tau_i)$ we can write the log-likelihood as

$$\log(\ell(\theta)) = \sum_{i=1}^{N} \delta_i \log(H(u_i|x_i, \tau_i)) - \int_{\tau_i}^{u_i} H(s|x_i, \tau_i)ds.$$

Following the exponential model of Definition 1 we replace $H(t|x_i, \tau_i)$ in the previous equation to get

$$\log(\ell(\theta)) = \sum_{i=1}^{N} \delta_i \theta^T x_i(u_i) - \int_{\tau_i}^{u_i} \exp(\theta^T x_i(s))ds\,,$$

We write the negative log-likelihood:

$$\ell(\theta) = -\log(\ell(\theta)) = \sum_{i=1}^{N} -\delta_i \theta^T x_i(u_i) + \int_{\tau_i}^{u_i} \exp(\theta^T x_i(s))ds.$$

□

B ONLINE CONVEX OPTIMIZATION

B.1 Proof of Lemma 1

Proof. Only the second assertion needs to be proven, the first one being Lemma 4.2.1 from Hazan [11] is already showed. To prove the second assertion we first see that Equation (5) means that for all $\theta \in \Theta$

$$\nabla^2 \ell(\theta) \succcurlyeq \mu \nabla \ell(\theta) \nabla \ell(\theta)^\top,$$

which implies that for all vector $\nu \in \mathbb{R}^d$

$$\nu^\top \nabla^2 \ell(\theta) \nu \geq \mu\ \nu^\top \nabla \ell(\theta) \nabla \ell(\theta)^\top \nu.$$

Since $\nabla\ell(\theta)\nabla\ell(\theta)^\top$ is a rank one matrix and $\nu = \nabla\ell(\theta)$ is an eigenvector associated to the unique non-null eigenvalue, we can replace ν in the previous equation to get

$$\nabla\ell(\theta)^\top\nabla^2\ell(\theta)\nabla\ell(\theta) \geq \mu\nabla\ell(\theta)^\top\nabla\ell(\theta)\nabla\ell(\theta)^\top\nabla\ell(\theta).$$

When $\nabla\ell(\theta) \neq 0$ we can write

$$\mu \leq \frac{\nabla\ell(\theta)^\top\nabla^2\ell(\theta)\nabla\ell(\theta)}{||\nabla\ell(\theta)||^4},$$

and as this is true for every $\theta \in \Theta$ we have

$$\mu \leq \min_{\theta\in\Theta} \frac{\nabla\ell(\theta)^\top\nabla^2\ell(\theta)\nabla\ell(\theta)}{||\nabla\ell(\theta)||^4}.$$

□

B.2 Proof of Lemma 2

Proof. The proof starts similarly than the one of Lemma 4.2.2 of Hazan [11]. We consider the concave function $p(\theta) = \exp(-\mu\ell(\theta))$. We derive that for $\theta_1, \theta_2 \in \Theta$:

$$\begin{aligned}
\ell(\theta_2) &\geq \ell(\theta_1) - \frac{1}{\mu}\log(1 - \mu(\nabla\ell(\theta_1)^T(\theta_2 - \theta_1))) \\
&\geq \ell(\theta_1) + \nabla\ell(\theta_1)^T(\theta_2 - \theta_1) \\
&\qquad - \left(\frac{1}{\mu}\log(1 - \mu(\nabla\ell(\theta_1)^T(\theta_2 - \theta_1))) + \nabla\ell(\theta_1)^T(\theta_2 - \theta_1)\right).
\end{aligned}$$

Using the Cauchy-Schwarz inequality we upper bound $|\nabla\ell(\theta_1)^T(\theta_2 - \theta_1)| \leq \|\nabla\ell(\theta_1)\|D$ for any $\theta_2 \in \Theta$. Combined with the monotonicity of the function $\mu^{-1}\log(1 - \mu z) + z$ which is decreasing for any $-\|\nabla\ell(\theta_1)\|D \leq z \leq \|\nabla\ell(\theta_1)\|D$ we obtain:

$$\ell(\theta_2) \geq \ell(\theta_1) + \nabla\ell(\theta_1)^T(\theta_2 - \theta_1) - \frac{1}{\mu}\log(1 + \mu\|\nabla\ell(\theta_1)\|D) + \|\nabla\ell(\theta_1)\|D\,.$$

By definition of the directional derivative constant, we thus can estimate:

$$\begin{aligned}
\gamma &\leq \min_{\theta_1,\theta_2\in\Theta} 2\frac{-\frac{1}{\mu}\log(1 + \mu\|\nabla\ell(\theta_1)\|D) + \|\nabla\ell(\theta_1)\|D}{(\nabla\ell(\theta_1)(\theta_2 - \theta_1))^2}\,, \\
&\leq \min_{\theta_1\in\Theta} 2\frac{-\frac{1}{\mu}\log(1 + \mu\|\nabla\ell(\theta_1)\|D) + \|\nabla\ell(\theta_1)\|D}{(\|\nabla\ell(\theta_1)\|D)^2}
\end{aligned}$$

by another application of the Cauchy-Schwarz inequality. □

B.3 The Online Newton Step algorithm

Algorithm 2 Online Newton Step [12]

Input: $(\ell_t)_{t=1,2,\dots}, \gamma > 0, n \geq 1, \epsilon > 0$

Initialization: $\theta_0 \in \Theta,\ A_0^{-1} = (1/\epsilon)\mathbb{1}_d$

for iteration $t = 1, \dots, n$ **do**

Observe: $\nabla \ell_t(\theta_t)$

Recursion:

$$A_t^{-1} = A_{t-1}^{-1} - \frac{A_{t-1}^{-1} \nabla \ell_t(\theta_t) \nabla \ell_t(\theta_t)^T A_{t-1}^{-1}}{1 + \nabla \ell_t(\theta_t) A_{t-1}^{-1} \nabla \ell_t(\theta_t)^T}$$

$$\theta_{t+1} = \mathrm{Proj}_t\Big(\theta_t - \frac{1}{\gamma} A_t^{-1} \nabla \ell_t(\theta_t)\Big)$$

where $\mathrm{Proj}_t(\theta^*) \in \arg\min_{\theta \in \Theta} (\theta - \theta^*)^T A_t (\theta - \theta^*)$.

end for

return θ_n

C STOCHASTIC SETTING

In this section we prove Theorem 2, and for this we need to recall the hypothesis of Theorem 7 from [29].

(H1) The diameter of Θ is D and the loss functions ℓ_t are continuously differentiable over Θ a.s. with integrable gradients.

(H2) The random loss functions $(\ell_t)_{t=1,2,\dots}$ are stochastically exp-concave 4 for some $\gamma \geq 0$.

(H3) The gradients $(\nabla \ell_t(\theta_t))_{t=1,2,\dots}$, satisfy for $G_1, G_2 > 0$ and all $k \geq 1, t = 1, 2, \dots,$

and $\theta \in \Theta$:

$$\begin{aligned}
\mathbb{E}_{t-1}[(\nabla \ell_t(\theta_t)^\top(\theta_t - \theta))^{2k}] &\leq k!\,(G_1 D)^{2(k-1)}\mathbb{E}_{t-1}[(\nabla \ell_t(\theta_t)^\top(\theta_t - \theta))^2] \qquad a.s., \\
\mathbb{E}_{t-1}[||\nabla \ell_t(\theta)||^{2k}] &\leq k!\,G_1^{2(k-1)}\mathbb{E}_{t-1}[||\ell_t(\theta_t)||^2] \qquad a.s., \\
\mathbb{E}_{t-1}[||\nabla \ell_t(\theta)||^2] &\leq G_2^2 \qquad\qquad a.s.
\end{aligned}$$

Let us notice that condition **(H3)** is satisfied in the bounded cases $||\nabla \ell_t(\theta_t)||^2 \leq G^2,\ t = 1, 2, \ldots$ with $G_1 := G_2 := G$. Condition **(H3)** is independent on the risk $L_t(\theta_t) = \mathbb{E}_{t-1}[\ell_t(\theta_t)],\ t = 1, 2, \ldots$, and thus, it does not interfere with condition **(H2)**. Additionally, we notice that in our setting where we consider the stochastic losses ℓ_t defined in (6), the hypothesis **(H1)** is already satisfied. Now, we recall the stochastic regret bound theorem.

Theorem 4 (Wintenberger [29])**.** *Under **(H1)**, **(H2)** with constant γ and **(H3)**, for $\varrho > 0$ and $n \geq 1$ the ONS algorithm 2 with learning rate $\gamma/3$ satisfies with probability $1 - 3\varrho$ the stochastic regret bound:*

$$\begin{aligned}
Risk_n \leq \frac{3}{2\gamma}\left(1 + d\log\left(1 + \frac{2(\gamma D)^2\left(nG_2^2 + G_1^2\log(\varrho^{-1})\right)}{9}\right)\right) \\
+ \left(\frac{4\gamma(G_1 D)^2}{9} + \frac{18}{\gamma}\right)\log(\varrho^{-1})\,.
\end{aligned}$$

In order to simplify notation we refer to the right-hand-side bound as $\mathcal{B}(n)$. In addition we need a proposition presented in [29] that gives us a constant γ that assures the stochastic exp-concavity of the losses.

Proposition 2 (Wintenberger [29])**.** *If L_t is μ-strongly convex and there exists $G > 0$ such that*

$$G^2 I_d \succcurlyeq \mathbb{E}_{t-1}[\nabla \ell_t(\theta)\nabla \ell_t(\theta)^T], \qquad \forall \theta \in \Theta, a.s., t = 1, 2, \ldots\,,$$

then Definition 4 holds with $\gamma := \mu/G^2$.

In the ideal case, we would like to prove that ℓ_t satisfies the conditions **(H2)** and **(H3)**. To prove **(H2)** we can use Proposition 2 if we find a constant such that the loss is strongly convex and a constant that bounds the expectation of the gradients. Unfortunately, we are not able to find this last constant a.s. but, proving a weaker version of **(H3)** we can define an auxiliary loss function that satisfies all the hypothesis and allows us to prove Theorem 2.

First, we prove that with high probability there is a constant G that upper bounds the norm of the gradients of $(\ell_t)_{t=1,2,\dots}$, this corresponds to the weaker **(H3)**. Secondly, we prove that the conditional risks $(L_t)_{t=1,2,\dots}$ are strongly convex for some constant μ, which consists of finding a lower bound of $\nabla^2 L_t(\theta)$ that does not depend on θ and t. This corresponds to only one of the conditions of Proposition 2, necessary to prove **(H2)**. Finally, we show how to use weak **(H3)** and half of **(H2)** to prove Theorem 2.

C.1 Upper bound (H3)

We want to find an upper bound for $||\nabla \ell_t(\theta)||^2$ and for this we first define for all $t = 1, 2, \dots$

$$R_t = \sum_{i=1}^{N_t} r_{it} \qquad \text{where} \qquad r_{it} = \mathbb{1}\{\tau_i \leq t, u_i > t-1\}\,,$$

and where N_t is the count function of the Poisson process defined in Section 4. Following, we prove that for all $t = 1, 2, \dots$, R_t is upper bounded with high probability.

Lemma 4. *Let $\varrho > 0$. Then, with probability at least $1 - \varrho$, for all $t = 1, 2, \dots$, we have*

$$R_t \leq 32 e^{Dx_\infty}(4\lambda + 1 + \log(2/\varrho))\,.$$

Proof. Since $u_i = \min\{c_i, t_i\} \leq t_i$ we can upper bound $R_t \leq \sum_{i=1}^{\infty} \mathbb{1}\{t_i \geq t-1\}\mathbb{1}\{\tau_i \leq t\}$. Then, we define $A_t = \{i : \tau_i \leq t\}$ and $Z_t = \sum_{i \in A_t} \mathbb{1}\{t_i \geq t-1\}$ and therefore, it will be enough to find a bound to Z_t to conclude. Given a constant $z > 0$ and $m \geq 1$, we fix

$N_t \leq m$ and we first upper bound the conditional probability

$$\mathbb{P}(Z_t \geq z | N_t = m) = \mathbb{P}\left(\sum_{i \in A_t} \mathbb{1}\{t_i \geq t-1\} \geq z \Big| N_t = m \right).$$

Let us notice that $N_t = |A_t|$. We would like to apply the concentration inequality of Chernoff-Hoeffding to the sum of Bernouilli random variables $\mathbb{1}\{t_i \geq t-1\}$ (see Hoeffding [14]), and for this we need to upper bound

$$\begin{aligned}
&\mathbb{P}(t_i \geq t-1 | i \in A_t, x_i) \\
&\qquad = \sum_{s=1}^{t} \mathbb{P}(t_i \geq t-1 | s-1 \leq \tau_i \leq s, x_i) \mathbb{P}(s-1 \leq \tau_i \leq s | 0 \leq \tau_i \leq t) \\
&\qquad = \frac{1}{t} \sum_{s=1}^{t} \mathbb{P}(t_i \geq t-1 | s-1 \leq \tau_i \leq s, x_i)\,,
\end{aligned}$$

where we use the uniform distribution of the Poisson process points given an interval (for more details on Poisson processes, see Daley and Vere-Jones [7]). Then, by the definition of the survival function (see Section 2) we get

$$\begin{aligned}
\mathbb{P}(t_i \geq t-1 | i \in A_t, x_i) &\leq \frac{1}{t} \sum_{s=1}^{t} S(t-1|s, x_i) \wedge 1 \\
&= \frac{1}{t} \sum_{s=1}^{t} \exp\left(-(t-1-s) e^{\theta^{*\top} x_i} \right) \wedge 1 \\
&\leq \frac{1}{t} \sum_{s=1}^{t} \exp\left(-(t-1-s) e^{-D x_\infty} \right) \wedge 1 \\
&\leq \frac{1}{t} \sum_{s=-1}^{\infty} \exp\left(-s e^{-D x_\infty} \right) \wedge 1 \\
&= \frac{2 - \exp(-e^{-D x_\infty})}{t\left(1 - \exp(-e^{-D x_\infty})\right)} \\
&\leq \frac{4\, e^{D x_\infty}}{t}\,.
\end{aligned}$$

In the last line we use that $1 - \exp(-x) \geq x/2$ for $0 \leq x \leq 1$.

The Chernoff-Hoeffding's inequality gives us for any sequence $X_1, \ldots, X_m$ with $\mathbb{E}[X_i] \leq p$ and any $\varepsilon > 0$

$$\mathbb{P}\left(\sum_{i=1}^{m} X_i \geq pm + \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2}{2mp(1-p)}\right) \leq \exp\left(-\frac{\varepsilon^2}{2mp}\right).$$

Applying this to the sum of the $\mathbb{1}\{t_i \geq t-1\}$ with $\mathbb{E}[\mathbb{1}\{t_i \geq t-1\}|i \in A_t] \leq \frac{e^{2+Dx_\infty}}{t}$ given $|A_t| = m$ and using the conditional independence of the Poisson process points, we obtain

$$\mathbb{P}\left(Z_t \geq \frac{m}{t} 4e^{Dx_\infty} + \varepsilon \middle| |A_t| = m\right) \leq \exp\left(-\frac{\varepsilon^2 t}{8me^{Dx_\infty}}\right).$$

Therefore, replacing $|A_t|$ by N_t

$$\mathbb{P}\left(Z_t \geq \frac{m}{t} 4e^{Dx_\infty} + \varepsilon \middle| N_t = m\right) \leq \exp\left(-\frac{\varepsilon^2 t}{8me^{Dx_\infty}}\right).$$

We set $z = \frac{m}{t} 4e^{Dx_\infty} + \varepsilon$ with which we get $\varepsilon = z - \frac{m}{t} 4e^{Dx_\infty}$ and

$$\mathbb{P}\left(Z_t \geq z \middle| N_t = m\right) \leq \exp\left(-\frac{\left(z - \frac{m}{t} 4e^{Dx_\infty}\right)^2 t}{8me^{Dx_\infty}}\right).$$

If we suppose $\frac{n}{t} 4e^{Dx_\infty} \leq \frac{z}{2}$ we obtain

$$\mathbb{P}\left(Z_t \geq z \middle| N_t = m\right) \leq \exp\left(-\frac{z^2 t}{32me^{Dx_\infty}}\right).$$

With this we found a bound for the conditional probability of Z_t being bigger than a certain constant. The next step is to bound $\mathbb{P}(Z_t \geq z)$, and for this we need to upper bound the probability of N_t being large. Let $M > 0$, since N_t follows a Poisson distribution of intensity λt, we can apply a Chernoff bound argument (more details in Mitzenmacher and Upfal [22]) obtaining

$$\mathbb{P}(N_t > M) \leq \left(\frac{e\lambda t}{M}\right)^M e^{-\lambda t} \qquad \text{for} \qquad M > \lambda t\,,$$

and

$$\mathbb{P}(N_t > M) \leq e^{-M-\lambda t} \qquad \text{when} \qquad M > e^2 \lambda t \,.$$

Now, we compute

$$\begin{aligned}
\mathbb{P}(Z_t \geq z) &= \sum_{m=1}^{M} \mathbb{P}(Z_t \geq z | N_t = m)\mathbb{P}(N_t = m) + \mathbb{P}(Z_t \geq z | N_t > M)\mathbb{P}(N_t > M) \\
&\leq \sum_{m=1}^{M} \exp\left(-\frac{z^2 t}{8me^{2+Dx_\infty}}\right) \mathbb{P}(N_t = m) \\
&\qquad\qquad + \mathbb{P}(Z_t \geq z | N_t > M)\mathbb{P}(N_t > M) \\
&\leq \exp\left(-\frac{z^2 t}{8Me^{2+Dx_\infty}}\right) + \exp(-M - \lambda t)\,,
\end{aligned}$$

where we use the bounds we previously found for $\mathbb{P}(Z_t \geq z | N_t = m)$ and $\mathbb{P}(N_t > M)$. Finally, we need to choose z and M such that $\mathbb{P}(Z_t \geq z) \leq \varrho/t^2$. Reminding the constrain $M > e^2 \lambda t$, we want

$$\exp(-M - \lambda t) \leq \frac{\varrho}{2t^2}\,,$$

which is true if

$$M \geq \log\left(\frac{2t^2}{\varrho}\right) - \lambda t$$

and then we can choose $M = e^2 \lambda t + \log(2t^2/\varrho)$ that satisfies both conditions. Similarly, we want

$$\exp\left(-\frac{z^2 t}{32Me^{Dx_\infty}}\right) \leq \frac{\varrho}{2t^2}\,,$$

which is true if

$$z \geq \sqrt{\frac{32Me^{Dx_\infty}}{t} \log(2t^2/\varrho)}\,,$$

and reminding the constrain $z \geq \frac{8M}{t} e^{Dx_\infty}$ we choose z such that

$$z \geq \frac{8M}{t} e^{Dx_\infty} + 2\sqrt{\frac{8Me^{Dx_\infty}}{t} \log(2t^2/\varrho)}\,.$$

Due to Young's inequality $a + 2\sqrt{ab} \leq 2a + b$, we can also choose

$$z \geq \frac{16M}{t} e^{Dx_\infty} + \log(2t^2/\varrho)\,,$$

which replacing M becomes

$$\begin{aligned}\frac{16}{t}e^{Dx_\infty}(e^2\lambda t + \log(2t^2/\varrho)) + \log(2t^2/\varrho) \\ &= 16\lambda e^{Dx_\infty} + \left(1 + \frac{16}{t}e^{Dx_\infty}\right)\log(2t^2/\varrho) \\ &\leq 32e^{Dx_\infty}(4\lambda + 1 + \log(1/\varrho))\,.\end{aligned}$$

In conclusion, we choose $z = 32e^{Dx_\infty}(4\lambda + 1 + \log(1/\varrho))$ and we get

$$\begin{aligned}\mathbb{P}(Z_t \geq z) &= \mathbb{P}(Z_t \geq 32e^{Dx_\infty}(4\lambda + 1 + \log(1/\varrho))) \\ &\leq \exp\left(-\frac{z^2 t}{32Me^{Dx_\infty}}\right) + \exp(-M - \lambda t) \\ &\leq \frac{\varrho}{2t^2} + \frac{\varrho}{2t^2} \\ &= \frac{\varrho}{t^2}\,.\end{aligned}$$

Using an upper-bound over t

$$\mathbb{P}(\forall t = 1, 2, \ldots \qquad Z_t \geq 32e^{Dx_\infty}(4\lambda + 1 + \log(1/\varrho))) \leq \sum_{t=1,2,\ldots} \frac{\varrho}{t^2} = \varrho\frac{\pi^2}{6} \leq 2\varrho\,,$$

which concludes the proof. □

Finally, we are now ready to give the desired upper bound for $||\nabla\ell_t(\theta)||^2$ in the following proposition

Proposition 3. *Let $\varrho > 0$. Then, with probability $1 - \varrho$ we have*

$$||\nabla\ell_t(\theta)||^2 \leq G^2 \qquad \forall\theta \in \Theta, t = 1, 2, \ldots\,,$$

with $G := 32e^{Dx_\infty}(4\lambda + 1 + \log(2/\varrho))(1 + e^{Dx_\infty})x_\infty$.

Proof. Let us notice that $\nabla\ell_t(\theta) \in \mathbb{R}^d$ and recall

$$\nabla\ell_t(\theta) = \sum_{i=1}^{N_t} -y_{it}x_i + r_{it}\exp(\theta^\top x_i)x_i\,(u_i \wedge t - \tau_i \vee 0)_+\,.$$

Then, we have

$$||\nabla \ell_t(\theta)|| \leq \sum_{i=1}^{N_t} ||y_{it} x_i|| + ||r_{it} \exp(\theta^\top x_i) x_i \left(u_i \wedge t - \tau_i \vee (t-1)\right)_+ ||\,,$$

noticing that $y_{it} \leq r_{it}$, $x_i \leq x_\infty$, $\exp(\theta^\top x_i) \leq \exp(D x_\infty)$
and $\left(u_i \wedge t - \tau_i \vee (t-1)\right)_+ \leq 1$,

$$\begin{aligned}
||\nabla \ell_t(\theta)|| &\leq \sum_{i=1}^{N_t} ||r_{it} x_\infty|| + ||r_{it} \exp(D x_\infty) x_\infty|| \\
&\leq \sum_{i=1}^{N_t} r_{it} \cdot \Big(1 + \exp(D x_\infty)\Big) x_\infty \\
&\leq R_t \Big(1 + \exp(D x_\infty)\Big) x_\infty \,,
\end{aligned}$$

by definition of $R_t = \sum_{i=1}^{N_t} r_{it}$. In consequence,

$$||\nabla \ell_t(\theta)||^2 \leq \left(32 e^{D x_\infty} (4\lambda + 1 + \log(2/\varrho))(1 + \exp(D x_\infty)) x_\infty\right)^2 ,$$

with probability $1 - \varrho$ and where the last inequality is due to Lemma 4. This conclude the proof. □

C.2 Strong convexity (H2)

Before showing the strong convexity let us remark that we can write $S(t|x_i, \tau_i) = \exp\left(-\int_{\tau_i}^{t} H(s|x_i, \tau_i) ds\right)$ and because $f(t|x_i, \tau_i) = H(t|x_i, \tau_i) S(t|x_i, \tau_i)$ the density of t_i is given by

$$f(t|x_i, \tau_i) = H(t|x_i, \tau_i) \exp\left(-\int_{\tau_i}^{t} H(s|x_i, \tau_i) ds\right).$$

Given $\theta^* \in \Theta$, the real parameter and replacing by our parametric model $h(t|x_i, \theta^*, \tau_i) = \exp(\theta^{*T} x_i)$ we have

$$f(t|x_i, \tau_i) := \exp(\theta^{*T} x_i) \exp\left(-(t - \tau_i) \exp(\theta^{*T} x_i)\right) \mathbb{1}\{t \geq \tau_i\}. \tag{8}$$

We also denote

$$\ell_t(\theta; s, c, x, \tau) := \Big(- \mathbb{1}\{t - 1 < s \leq t \wedge c\} \theta^T x + \exp(\theta^T x)((c \wedge s \wedge t) - (\tau \vee (t-1)))_+ \Big)$$

and recalling that $y_{it} = \mathbb{1}\{t-1 < t_i \leq t \wedge c_i\}$, $u_i = t_i \wedge c_i$ and $r_{it} = \mathbb{1}\{\tau_i < t, u_i \geq t-1\}$, we have

$$\begin{aligned}\ell_t(\theta) &= \sum_{i=1}^{N_t} -y_{it}\theta^T x_i + r_{it}\exp(\theta^T x_i)((u_i \wedge t) - (\tau_i \vee (t-1))) \\ &= \sum_{i=1}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\,.\end{aligned}$$

In addition, as $(t_i)_{i\geq 1}$ and $(c_i)_{i\geq 1}$ are i.i.d. we name T and C random variables that are distributed as t_i and c_i, respectively. We first prove the following Lemma that gives us an explicit expression of the risk function $L_t(\theta) := \mathbb{E}_{t-1}[\ell_t(\theta)]$, $\theta \in \Theta$, $t = 1, 2, \ldots$.

Lemma 5. *For every $t = 1, 2, \ldots$ and every $\theta \in \Theta$ the risk function is given by*

$$\begin{aligned}L_t(\theta) &= \lambda\mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x)\mathbb{1}\{T \leq C\}(1-T)_+ \mid \tau = 0\Big] \\ &+ \sum_{\substack{i:\{u_i > t-1\} \\ i:\{\tau_i \leq t-1\}}} \Big(e^{(\theta-\theta^*)^\top x_i} - \theta^\top x_i\Big)\,\mathbb{P}(\mathbb{1}\{t-1+\tau_i < T \leq \tau_i + t \wedge C\}|x_i, \tau_i, \tau = 0)\,.\end{aligned}$$

Proof. The expected value is

$$\mathbb{E}_{t-1}[\ell_t(\theta)] = \mathbb{E}_{t-1}\left[\sum_{i=1}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right],$$

which we separate in two terms

$$\mathbb{E}_{t-1}[\ell_t(\theta)] = \mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right] + \sum_{i=1}^{N_{t-1}} \mathbb{E}_{t-1}\left[\ell_t(\theta; t_i, c_i, x_i, \tau_i)|x_i, \tau_i\right]. \tag{9}$$

Now, recalling that g and f respectively denote the conditional densities of t_i and c_i given (τ_i, x_i) and, because c_i and t_i are independent given (τ_i, x_i), we first calculate the

first term

$$\mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right]$$
$$= \mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \mathbb{E}\left[\ell_t(\theta; t_i, c_i, x_i, \tau_i)|x_i, \tau_i\right]\right]$$
$$= \mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s, c, x_i, \tau_i) g(c|x_i, \tau_i) f(s|x_i, \tau_i) ds dc\right].$$

Now because x_i are i.i.d. and independent from τ_i and c_i, denoting by x a random variable with the same distribution we have

$$\mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right]$$
$$= \mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s, c, x, \tau_i) g(c|x, \tau_i) f(s|x, \tau_i) ds dc\right]$$

which can be written as the stochastic integral

$$\mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right]$$
$$= \mathbb{E}\left[\int_{t-1}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s, c, x, v) g(c|x, v) f(s|x, v) ds dc dN(v)\right]$$
$$= \lambda\mathbb{E}\left[\int_{t-1}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s, c, x, v) g(c|x, v) f(s|x, v) ds\, dc\, dv\right]$$
$$= \lambda\mathbb{E}\left[\int_{t-1}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s+v, c+v, x, v) g(c+v|x, v) f(s+v|x, v) ds\, dc\, dv\right]$$
$$= \lambda\mathbb{E}\left[\int_{t-1}^{t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ell_t(\theta; s+v, c+v, x, v) g(c|x, 0) f(s|x, 0) ds\, dc\, dv\right]$$
$$= \lambda\mathbb{E}\left[\int_{-\infty}^{\infty}\int_{t-1}^{t}\int_{-\infty}^{\infty} \ell_t(\theta; s+v, c+v, x, v)\, f(s|x, 0)\, ds\, dv\, g(c|x, 0) dc\right].$$

We do not know $g(c|x,v)$ but we know that $g(c|x,v) = g(c-\epsilon|x,v-\epsilon)$ for all $\epsilon \in \mathbb{R}$. For instance, $g(c|x,v) = g(c-v|x,0)$ and, the same is satisfied for f. Then, we change the variable $v \in [0,t]$ in $w = v-(t-1)$:

$$\int_{t-1}^{t} \ell_t(\theta; s+v, c+v, x, v)\, dv = \int_0^1 \ell_1(\theta; s+w, c+w, x, w)\, dw\,.$$

We obtain

$$\mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right]$$
$$= \lambda \mathbb{E}\left[\int_{-\infty}^{\infty}\int_0^1\int_{-\infty}^{\infty} \ell_1(\theta; s+w, c+w, x, w) f(s|x,0) ds\, dw\, g(c|x,0) dc\right].$$

Considering the integral on the time s

$$\int_{-\infty}^{\infty} \ell_1(\theta; s+w, c+w, x, w) f(s|x,0) ds$$
$$= \int_0^{\infty} \ell_1(\theta; s+w, c+w, x, w) e^{\theta^{*T}x} \exp(-se^{\theta^{*T}x}) ds\,,$$

we obtain

$$-\int_{(-w)_+}^{(1-w)\wedge c} \theta^T x e^{\theta^{*T}x} \exp(-se^{\theta^{*T}x}) ds$$
$$= -\theta^T x \mathbb{P}((-w)_+ < T \le (1-w)\wedge c | \tau = 0)\,,$$

and

$$\int_{(1-w)\wedge c}^{\infty} ((c+w)\wedge(s+w)\wedge 1 - w_+)_+ e^{\theta^T x} e^{\theta^{*T}x} \exp(-se^{\theta^{*T}x}) ds$$
$$= e^{\theta^T x}((c+w)\wedge 1 - w_+)_+ \mathbb{P}(T \ge (1-w)\wedge c | \tau = 0)\,,$$

and

$$\int_{(-w)_+}^{(1-w)\wedge c} ((s+w) - w_+) e^{\theta^T x} e^{\theta^{*T}x} \exp(-se^{\theta^{*T}x}) ds$$
$$= e^{\theta^T x}\Big(-((s+w)-w_+)\exp(-se^{\theta^{*T}x})\,|_{(-w)_+}^{(1-w)\wedge c} + \int_{(-w)_+}^{(1-w)\wedge c} \exp(-se^{\theta^{*T}x}) ds\Big)$$
$$= -e^{\theta^T x}((c+w)\wedge 1 - w_+)_+ \mathbb{P}(T \ge (1-w)\wedge c | \tau = 0)$$
$$+ \exp((\theta-\theta^*)^T x) \mathbb{P}((-w)_+ < T \le (1-w)\wedge c | \tau = 0)\,.$$

All in all we obtain

$$\int_0^\infty \ell_1(\theta; s+w, c+w, x, w) e^{\theta^{*T} x} \exp(-s e^{\theta^{*T} x}) ds$$
$$= (e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{P}((-w)_+ < T \leq (1-w) \wedge c | \tau = 0),$$

thus we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=N_{t-1}}^{N_t} \ell_t(\theta; t_i, c_i, x_i, \tau_i)\right] \\
&= \lambda \mathbb{E}\left[(e^{(\theta-\theta^*)^T x} - \theta^T x) \int_{-\infty}^{\infty} \int_0^1 \mathbb{P}((-w)_+ < T \leq (1-w) \wedge c | \tau = 0) dw \, g(c|x, 0) dc\right] \\
&= \lambda \mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{E}\Big[\int_0^1 \mathbb{1}\{(-w)_+ < T \leq (1-w) \wedge C\} dw \Big| x, \tau = 0\Big]\Big] \\
&= \lambda \mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{E}\Big[\int_{0 \vee -T}^{1-T} \mathbb{1}\{T \leq C\} dw \Big| x, \tau = 0\Big]\Big] \\
&= \lambda \mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{E}\Big[\mathbb{1}\{T \leq C\}((1-T) - 0 \vee -T)_+ \Big| x, \tau = 0\Big]\Big] \\
&= \lambda \mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{E}\Big[\mathbb{1}\{T \leq C\}(1-T)_+ \Big| x, \tau = 0\Big]\Big].
\end{aligned}$$

Replacing in Equation (9) we have

$$\begin{aligned}
\mathbb{E}_{t-1}[\ell_t(\theta)] = \lambda \mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x) \mathbb{E}\Big[\mathbb{1}\{T \leq C\}(1-T)_+ \Big| x, \tau = 0\Big]\Big] \\
+ \sum_{i=1}^{N_{t-1}} \mathbb{E}_{t-1}\left[\ell_t(\theta; t_i, c_i, x_i, \tau_i) | x_i, \tau_i\right].
\end{aligned}$$

To calculate the second term, we note that we know u_i if $u_i \leq t-1$ and in this case $\ell_t(\theta; t_i, c_i, x_i, \tau_i) = 0$, therefore, we consider only the individuals i such that $u_i > t-1$. The sum becomes

$$\begin{aligned}
&\sum_{i=1}^{N_{t-1}} \mathbb{E}_{t-1}\left[\ell_t(\theta; t_i, c_i, x_i, \tau_i) | x_i, \tau_i\right] \\
&\qquad = \sum_{\substack{i:\{u_i > t-1\} \\ i:\{\tau_i \leq t-1\}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ell_t(\theta; s, c, x_i, \tau_i) g(c|x_i, \tau_i) f(s|x_i, \tau_i) ds dc,
\end{aligned}$$

which, following the calculations of the first term we obtain

$$\sum_{i=1}^{N_{t-1}} \mathbb{E}_{t-1}\left[\ell_t(\theta; t_i, c_i, x_i, \tau_i)|x_i, \tau_i\right]$$
$$= \sum_{\substack{i:\{u_i>t-1\}\\ i:\{\tau_i\le t-1\}}} \left(e^{(\theta-\theta^*)^\top x_i} - \theta^\top x_i\right) \mathbb{P}(t-1+\tau_i < T \le \tau_i + t \wedge C | x = x_i, \tau_i, \tau = 0)\,,$$

where (T, C, x, τ) is independent of (t_i, c_i, x_i, τ_i) for every $i \ge 1$. Let us notice that τ_i and x_i, which we suppose are observed at the same time as τ_i, are known at time $t-1$. Replacing this term in Equation (9) leads to

$$\mathbb{E}_{t-1}[\ell_t(\theta)]$$
$$= \lambda\mathbb{E}\Big[(e^{(\theta-\theta^*)^T x} - \theta^T x)\mathbb{1}\{T \le C\}(1-T)_+ \mid \tau = 0\Big]$$
$$+ \sum_{\substack{i:\{u_i>t-1\}\\ i:\{\tau_i\le t-1\}}} \left(e^{(\theta-\theta^*)^\top x_i} - \theta^\top x_i\right) \mathbb{P}(t-1+\tau_i < T \le \tau_i + t \wedge C | x_i, \tau_i, \tau = 0),$$

that finalizes the proof. □

We define

$$J(\theta) := \lambda\mathbb{E}\Big[(e^{(\theta-\theta^*)^\top x} - \theta^T x)\mathbb{1}\{T \le C\}(1-T)_+ \Big| x, \tau = 0\Big]\,, \tag{10}$$

and we are ready to show the strong convexity of the risk function that we give in the following proposition.

Proposition 4. *The risk function satisfies*

$$\nabla^2 L_t(\theta) \succcurlyeq \lambda e^{-Dx_\infty}\mathbb{E}[xx^\top \mathbb{1}\{T \le C\}(1-T)_+ | \tau = 0], \qquad \forall \theta \in \Theta, t = 1, \dots, n\,.$$

Therefore, under Assumption 2 L_t is μ-strongly convex for $\mu = \lambda e^{-Dx_\infty} A$.

Proof. Lemma 5 gives us an expression of the risk $L_t =: J + R_t$ with R_t some random convex function. By convexity $\nabla^2 L_t(\theta) \succcurlyeq \nabla^2 J(\theta)$, $\theta \in \Theta$, and therefore, it is enough to bound the hessian of the first term J. We calculate

$$\nabla J(\theta) = \lambda\mathbb{E}\Big[(e^{(\theta-\theta^*)^\top x} - 1)x\mathbb{1}\{T \le C\}(1-T)_+ | x, \tau = 0\Big]$$

and

$$\nabla^2 J(\theta) = \lambda \mathbb{E}\Big[e^{(\theta-\theta^*)^\top x} x x^\top \mathbb{1}\{T \leq C\}(1-T)_+ | x, \tau = 0\Big] .$$

Let us notice that $e^{(\theta-\theta^*)^\top x} \geq e^{-D\|x\|_\infty}$ and then

$$\nabla^2 J(\theta) \succcurlyeq \lambda e^{-D\|x\|_\infty} \mathbb{E}[x x^\top \mathbb{1}\{T \leq C\}(1-T)_+ | x, \tau = 0] ,$$

which due to Assumption 2 concludes the proof. □

C.3 Proof of Theorem 2

Proof. First of all, we remind that Proposition 3 implies **(H3)** with $G = G_1 = G_2$, but this bound for the gradients is satisfied with probability $1-\varrho$ instead of almost surely and therefore we cannot claim that **(H3)** is always fulfilled. But there is a problem with this definition because **(H3)** considers all $t = 1, 2, \ldots$ and, in order to have a $\mathcal{F}_t$-measurable function we need to define a time dependent $\textbf{(H3)}_t$:

$\textbf{(H3)}_t$ For $t+1 \geq s \geq 1$ the gradients $\nabla \ell_s(\theta_s)$, satisfy for $G > 0$ and all $k \geq 1$, and $\theta \in \Theta$:

$$\begin{aligned}
\mathbb{E}_{s-1}[(\nabla \ell_s(\theta_s)^\top (\theta_s - \theta))^{2k}] &\leq k! \, (GD)^{2(k-1)} \mathbb{E}_{s-1}[(\nabla \ell_s(\theta_s)^\top (\theta_s - \theta))^2] , \\
\mathbb{E}_{s-1}[||\nabla \ell_s(\theta)||^{2k}] &\leq k! \, G^{2(k-1)} \mathbb{E}_{s-1}[||\ell_s(\theta_s)||^2] , \\
\mathbb{E}_{s-1}[||\nabla \ell_t(\theta)||^2] &\leq G^2 .
\end{aligned}$$

We define $\Omega_t = \{(y_{is}, x_i, \tau_i, u_i)_{s \leq t}$ such that $\textbf{(H3)}_t$ is satisfied$\}$ for all $t = 1, 2, \ldots$ sand we check that Ω_t is $\mathcal{F}_t$-measurable. Next, for all $t = 1, 2, \ldots$ we define the auxiliary loss function

$$\hat{\ell}_t(\theta) = \ell_t(\theta) \mathbb{1}\{\Omega_{t-1}\} ,$$

which is $\mathcal{F}_t$-measurable. Let us notice that we need to define Ω_t using $\textbf{(H3)}_t$ instead of the inequality of Proposition 3 to preserve the past dependency and the measurability. We prove that the function $\hat{\ell}_t$ satisfies the conditions **(H1)**, **(H2)** and **(H3)**.

First of all, **(H1)** is still verified because the indicator function does not depend on θ. Secondly, if $\mathbf{(H3)}_t$ is not realized then the function $\hat{\ell}_t$ is zero and all the bounds hold. Thirdly, if $\mathbf{(H3)}_t$ is realized, ℓ_t satisfies the inequalities of **(H3)** and $\hat{\ell}_t = \ell_t$ by definition. Then the bounds in **(H3)** are also true for $\hat{\ell}_t$, concluding that $\hat{\ell}_t$ satisfies the inequalities of **(H3)** for all $t = 1, 2, \ldots$. Finally, it remains to prove **(H2)**.

By $\mathcal{F}_{t-1}$-measurability of Ω_{t-1} we calculate for $\theta \in \Theta$:

$$\mathbb{E}_{t-1}[\nabla\hat{\ell}_t(\theta)\nabla\hat{\ell}_t(\theta)^\top] = \mathbb{1}\{\Omega_{t-1}\}\mathbb{E}_{t-1}[\nabla\ell_t(\theta)\nabla\ell_t(\theta)^\top]\,.$$

If $\mathbf{(H3)}_t$ is not realized, $\mathbb{1}\{\Omega_{t-1}\} = 0$ and so (4) is true for any constant $\gamma \geq 0$. If $\mathbf{(H3)}_t$ is realized, $\mathbb{1}\{\Omega_{t-1}\} = 1$ and there exist $G > 0$ such that:

$$\mathbb{E}_{t-1}[\nabla\hat{\ell}_t(\theta)\nabla\hat{\ell}_t(\theta)^\top] = \mathbb{E}_{t-1}[\nabla\ell_t(\theta)\nabla\ell_t(\theta)^\top] \preccurlyeq G^2\mathcal{I}_d\,.$$

This, together with the strong convexity of Proposition 4 give us the hypothesis of Proposition 2 assuring the stochastic exp-concavity for $\gamma = \lambda e^{-Dx_\infty}A/G^2$ and concluding **(H2)**. Now, we have that $\hat{\ell}_t$ satisfies all the conditions of Theorem 4 assuring the logarithmic stochastic regret bound of ONS.

To study the stochastic regret bound we need also to define for all $t = 1, 2, \ldots$ the risk function $\hat{L}_t(\theta) = \mathbb{E}_{t-1}[\hat{\ell}_t(\theta)]$ and we notice that as $\mathbb{1}\{\Omega_{t-1}\}$ is $\mathcal{F}_{t-1}$-measurable:

$$\hat{L}_t(\theta) = \mathbb{1}\{\Omega_{t-1}\}\mathbb{E}_{t-1}[\ell_t(\theta)] = \mathbb{1}\{\Omega_{t-1}\}L_t(\theta)\,, \qquad \theta \in \Theta.$$

Now, it remains to prove that ONS has logarithmic stochastic regret also for L_t and therefore, we calculate for every $n \geq 1, t = 1, \ldots, n$,

$\theta^* \in \arg\min_{\theta\in\Theta} \sum_{t=1}^n L_t(\theta)$ and θ_t the prediction of ONS at time t:

$$\begin{aligned}\mathbb{P}\left[\sum_{t=1}^n L_t(\theta_t) - L_t(\theta) > \mathcal{B}(n)\right] &= \mathbb{P}\left[\sum_{t=1}^n L_t(\theta_t) - L_t(\theta^*) > \mathcal{B}(n), \bigcap_{t\geq 2}\Omega_{t-1}\right] \\ &\quad + \mathbb{P}\left[\sum_{t=1}^n L_t(\theta_t) - L_t(\theta^*) > \mathcal{B}(n), \left(\bigcap_{t\geq 2}\Omega_{t-1}\right)^c\right] \\ &\leq \mathbb{P}\left[\sum_{t=1}^n \left(L_t(\theta_t) - L_t(\theta^*)\right)\mathbb{1}\{\Omega_{t-1}\} > \mathcal{B}(n)\right] \\ &\quad + \mathbb{P}\left[\left(\bigcap_{t\geq 2}\Omega_{t-1}\right)^c\right],\end{aligned}$$

where $\mathcal{B}(n)$ is the stochastic regret bound for $\hat{L}(\theta)$ of Theorem 4 which we remind:

$$\begin{aligned}\mathcal{B}(n) = \frac{3}{2\gamma}\left(1 + d\log\left(1 + \frac{2(\gamma D)^2 G^2\left(n + \log(\varrho^{-1})\right)}{9}\right)\right) \\ + \left(\frac{4\gamma(GD)^2}{9} + \frac{18}{\gamma}\right)\log(\varrho^{-1})\,.\end{aligned}$$

Plugging in $\mathcal{B}(n)$ the specific values of γ, G and μ found in Propositions 2, 3, and 4, respectively,

$$\gamma = \frac{\mu}{G^2} = \frac{\lambda e^{-Dx_\infty} A}{\left(32e^{Dx_\infty}(4\lambda + 1 + \log(2/\varrho))(1 + e^{Dx_\infty})x_\infty\right)^2}\,,$$

we obtain the regret bound

$$\begin{aligned}Risk_n \leq \frac{3G^2 e^{Dx_\infty}}{2\lambda A}\left(1 + d\log\left(1 + \frac{2(\lambda AD)^2\left(n + \log(\varrho^{-1})\right)}{9G^2 e^{2Dx_\infty}}\right)\right) \\ + \left(\frac{4\lambda AD^2}{9e^{Dx_\infty}} + \frac{18Ge^{Dx_\infty}}{\lambda A}\right)\log(\varrho^{-1})\,. \qquad (11)\end{aligned}$$

Then, because of Theorem 4, this bound holds with probability 3ϱ and as

$$\mathbb{P}\left[\left(\bigcap_{t\geq 2}\Omega_{t-1}\right)^c\right] \leq \varrho\,,$$

we have:

$$\mathbb{P}\left[\sum_{t=1}^{n} L_t(\theta_t) - L_t(\theta) > \mathcal{O}(\log(n/\varrho))\right] \leq 4\varrho\,,$$

and thus, with probability $1-4\varrho$, ONS algorithm has logarithmic stochastic regret. □

C.4 Proof of Corollary 1

Proof. Due to the μ-strong convexity of $L_t(\theta)$ proved in Proposition 4 we have for all $t = 1, 2, \ldots$:

$$\mu||\theta_t - \theta^*||^2 \leq \nabla L_t(\theta^*)^\top(\theta_t - \theta^*) + \mu||\theta_t - \theta^*||^2 \leq L_t(\theta_t) - L_t(\theta^*)\,,$$

where the first inequality is true because $\nabla L_t(\theta^*)^\top(\theta_t - \theta^*) \geq 0$. Then, because of Theorem 2:

$$\sum_{t=1}^{n}||\theta_t - \theta^*||^2 \leq \frac{1}{\mu}\sum_{t=1}^{n} L_t(\theta_t) - L_t(\theta^*) \leq \frac{1}{\mu}\mathcal{B}(n)\,,$$

and remembering that $\mu = \lambda e^{-Dx_\infty}A$, the bound is:

$$\frac{1}{\mu}\mathcal{B}(n) = \frac{3G^2e^{2Dx_\infty}}{2\lambda^2A^2}\left(1 + d\log\left(1 + \frac{2(\lambda AD)^2\left(n + \log(\varrho^{-1})\right)}{9G^2e^{2Dx_\infty}}\right)\right) + \left(\frac{4\lambda AD^2}{9e^{Dx_\infty}} + \frac{18Ge^{Dx_\infty}}{\lambda A}\right)\log(\varrho^{-1})\,, \tag{12}$$

which is $\mathcal{O}(\log(n/\varrho))$. We conclude the convergency of θ_t to θ^* and then:

$$||\bar{\theta}_n - \theta^*||^2 \leq \frac{1}{n}\sum_{t=1}^{n}||\theta_t - \theta^*||^2 - \frac{1}{n}\sum_{t=1}^{n}||\theta_t - \bar{\theta}_n||^2 \leq \frac{1}{\mu}\mathcal{O}(\log(n/\varrho)/n)\,,$$

concluding the convergency of $\bar{\theta}_n$ to θ^*. □

D SURVONS

D.1 Proof of Lemma 3

Proof. We first compute

$$\begin{aligned}\nabla\hat{\ell}_{t,\gamma}(\theta_1) &= \nabla\ell_t(\hat{\theta}) + \gamma\left(\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\right)\nabla\ell_t(\hat{\theta}) \\ &= \left(1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\right)\nabla\ell_t(\hat{\theta})\,.\end{aligned}$$

We need to show that there exists $\hat{\gamma} > 0$ such that

$$\hat{\ell}_{t,\gamma}(\theta_2) \geq \hat{\ell}_{t,\gamma}(\theta_1) + \nabla\hat{\ell}_{t,\gamma}(\theta_1)(\theta_2 - \theta_1) + \frac{\hat{\gamma}}{2}\left(\nabla\hat{\ell}_{t,\gamma}(\theta_1)(\theta_2 - \theta_1)\right)^2$$

and if we replace $\hat{\ell}_{t,\gamma}$ this inequality is equivalent to

$$\begin{aligned}
&\ell_t(\hat{\theta}) + \nabla\ell_t(\hat{\theta})(\theta_2 - \hat{\theta}) + \frac{\gamma}{2}\left(\nabla\ell_t(\hat{\theta})(\theta_2 - \hat{\theta})\right)^2 \\
&\geq \ell_t(\hat{\theta}) + \nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}) + \frac{\gamma}{2}\left(\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\right)^2 \\
&\quad + \left(1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\right)\nabla\ell_t(\hat{\theta})(\theta_2 - \theta_1) \\
&\quad + \frac{\hat{\gamma}}{2}\left((1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}))\nabla\ell_t(\hat{\theta})(\theta_2 - \theta_1)\right)^2.
\end{aligned}$$

Grouping, this requirement becomes

$$\begin{aligned}
\frac{\gamma}{2}\left(\nabla\ell_t(\hat{\theta})(\theta_2 - \hat{\theta})\right)^2 \geq \frac{\gamma}{2}&\left(\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\right)^2 \\
&+ \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta})\nabla\ell_t(\hat{\theta})(\theta_2 - \theta_1) \\
&+ \frac{\hat{\gamma}}{2}\left((1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}))\nabla\ell_t(\hat{\theta})(\theta_2 - \theta_1)\right)^2,
\end{aligned}$$

which is

$$0 \geq \left(\frac{\hat{\gamma}}{2}(1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}))^2 + \frac{\gamma}{2}\right)\left(\nabla\ell_t(\hat{\theta})(\theta_2 - \theta_1)\right)^2.$$

To satisfy this inequality we need

$$\hat{\gamma} \leq \frac{\gamma}{(1 + \gamma\nabla\ell_t(\hat{\theta})(\theta_1 - \hat{\theta}))^2},$$

which is true for the choice $\hat{\gamma} = \frac{\gamma}{(1+\gamma D||\nabla\ell_t(\hat{\theta})||)^2}$ and this concludes the proof. □

D.2 Proof of Theorem 3

Proof. At each iteration t we consider $\hat{\theta}_t$ the prediction of SurvONS and $\theta_t(\gamma)$ the prediction of ONS with $\gamma \in \Gamma$. We define the directional derivative lower bound function as in Equation (7)

$$\hat{\ell}_{t,\gamma_t}(\theta) = \ell_t(\hat{\theta}_t) + \nabla\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t) + \frac{\gamma_t}{2}\left(\nabla\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t)\right)^2.$$

Let us notice that $\gamma \leq \frac{1}{4GD}$ and $\hat{\ell}_{t,\gamma_t}(\theta) \leq \ell_t(\theta)$ for all θ.

We take $\theta^* \in \arg\min_{\theta\in\Theta} \sum_{t=1}^n \ell_t(\theta)$ and we can upper-bound the regret for any $\gamma \in \Gamma$

$$\begin{aligned}
Regret_n &= \sum_{t=1}^n \ell_t(\hat{\theta}_t) - \ell_t(\theta^*) \\
&\leq \sum_{t=1}^n \hat{\ell}_{t,\gamma_t}(\hat{\theta}_t) - \hat{\ell}_{t,\gamma_t}(\theta^*) \\
&= \sum_{t=1}^n \hat{\ell}_{t,\gamma_t}(\hat{\theta}_t) - \hat{\ell}_{t,\gamma_t}(\theta_t(\gamma)) + \hat{\ell}_{t,\gamma_t}(\theta_t(\gamma)) - \hat{\ell}_{t,\gamma_t}(\theta^*) \\
&= \sum_{t=1}^n \nabla\ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)) - \sum_{t=1}^n \frac{\gamma_t}{2}\left(\nabla\ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma))\right)^2 \\
&\qquad + \sum_{t=1}^n \hat{\ell}_{t,\gamma_t}(\theta_t(\gamma)) - \hat{\ell}_{t,\gamma_t}(\theta^*)\,.
\end{aligned}$$

We upper-bound the first term using the regret-bound of BOA [28] which works for $\gamma \leq \frac{1}{4GD}$

$$\sum_{t=1}^n \nabla\ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)) \leq \frac{\log(K)}{\gamma} + 2\gamma \sum_{t=1}^n \left(\nabla\ell(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma))\right)^2 .$$

Therefore, the regret is bounded by

$$\begin{aligned}
Regret_n \leq \frac{\log(K)}{\gamma} + \sum_{t=1}^n \left(\frac{4\gamma - \gamma_t}{2}\right)\left(\nabla\ell(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma))\right)^2 \\
+ \sum_{t=1}^n \hat{\ell}_{t,\gamma_t}(\theta_t(\gamma)) - \hat{\ell}_{t,\gamma_t}(\theta^*).
\end{aligned}$$

We consider the surrogate losses $\hat{\ell}_{t,\hat{\gamma}_t}$ for $t = 1, 2, \ldots$ and $\hat{\gamma}_t = 4\max\{\gamma, \gamma_t/4\}$

$$\hat{\ell}_{t,\hat{\gamma}_t}(\theta) = \ell_t(\hat{\theta}_t) + \nabla\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t) + 2\max\{\gamma, \frac{\gamma_t}{4}\}\left(\nabla\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t)\right)^2 ,$$

and we write the last term of the regret bound

$$\begin{aligned}
&\sum_{t=1}^{n} \hat{\ell}_{t,\gamma_t}(\theta_t(\gamma)) - \hat{\ell}_{t,\gamma_t}(\theta^*) \\
&\quad = \sum_{t=1}^{n} \left(\hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) \right) + \sum_{t=1}^{n} \left(\hat{\ell}(\theta_t(\gamma)) - \hat{\ell}_t(\theta_t(\gamma);\gamma) \right) \\
&\qquad - \sum_{t=1}^{n} \left(\hat{\ell}_t(\theta^*) - \hat{\ell}_t(\theta^*;\gamma) \right) \\
&\quad = \sum_{t=1}^{n} \left(\hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) \right) \\
&\qquad - \sum_{t=1}^{n} \frac{(4\gamma - \gamma_t)_+}{2} \left(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)) \right)^2 \\
&\qquad\qquad + \sum_{t=1}^{n} \frac{(4\gamma - \gamma_t)_+}{2} \left(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*) \right)^2 .
\end{aligned}$$

We substitute this expression in the regret bound

$$\begin{aligned}
Regret_n \leq \frac{\log(K)}{\gamma} &+ \sum_{t=1}^{n} \left(\hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) \right) \\
&- \sum_{t=1}^{n} \frac{(\gamma_t - 4\gamma)_+}{2} \left(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)) \right)^2 \\
&+ \sum_{t=1}^{n} \frac{(4\gamma - \gamma_t)_+}{2} \left(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*) \right)^2 .
\end{aligned}$$

Now, we note that by Lemma 3 we have

$$\begin{aligned}
\hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) \leq \nabla \hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma) - \theta^*) \\
- \frac{\hat{\gamma}_t}{2} \left(\nabla \hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma) - \theta^*) \right)^2 ,
\end{aligned}$$

where we can write $\max\{\gamma, \gamma_t/4\} = \gamma + (\gamma_t/4 - \gamma)_+$ and get

$$\hat{\gamma}_t = \frac{4(\gamma + (\gamma_t/4 - \gamma)_+)}{(1 + 4(\gamma + (\gamma_t/4 - \gamma)_+)(\nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma) - \hat{\theta}_t)))^2} \geq \gamma .$$

Therefore, we can apply the regret bound of ONS which yields to

$$\sum_{t=1}^{n} \hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) \leq \frac{5d\log(n)}{\gamma} + \frac{\gamma}{2}\sum_{t=1}^{n}\left(\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma)-\theta^*)\right)^2 - \sum_{t=1}^{n}\frac{\hat{\gamma}_t}{2}\left(\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma)-\theta^*)\right)^2. \quad (13)$$

But, since

$$\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma) = \left(1 + 4(\gamma + (\gamma_t/4-\gamma)_+)\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\hat{\theta}_t)\right)\nabla\ell_t(\hat{\theta}_t)\,,$$

we can write

$$\left(\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma)-\theta^*)\right)^2 = \left(1 + 4(\gamma + (\gamma_t/4-\gamma)_+)\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\hat{\theta}_t)\right)^2\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2,$$

which yields to

$$\hat{\gamma}_t\left(\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma)-\theta^*)\right)^2 = 4(\gamma + (\gamma_t/4-\gamma)_+)\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2. \quad (14)$$

Using the assumption $4(\gamma + (\gamma_t/4-\gamma)_+) \leq 1/GD$ we can also get

$$\left(\nabla\hat{\ell}_t(\theta_t(\gamma);\gamma)(\theta_t(\gamma)-\theta^*)\right)^2 \leq 4\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2. \quad (15)$$

Therefore, plugging (14) and (15) in (13) we get

$$\begin{aligned}\sum_{t=1}^{n} \hat{\ell}_t(\theta_t(\gamma);\gamma) - \hat{\ell}_t(\theta^*;\gamma) &\leq \frac{5d\log(n)}{\gamma} + \frac{4\gamma}{2}\sum_{t=1}^{n}\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2 \\ &\quad - 2\sum_{t=1}^{n}(\gamma + (\gamma_t/4-\gamma)_+)\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2 \\ &= \frac{5d\log(n)}{\gamma} \\ &\quad - 2\sum_{t=1}^{n}(\gamma_t/4-\gamma)_+\left(\nabla\ell_t(\hat{\theta}_t)(\theta_t(\gamma)-\theta^*)\right)^2.\end{aligned}$$

Thus, the regret bound becomes

$$
\begin{aligned}
Regret_n \leq & \frac{2\log(K) + 5d\log(n)}{\gamma} \\
& - 2\sum_{t=1}^{n}(\gamma_t/4 - \gamma)_+ \left((\nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma) - \theta^*))^2 + (\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)))^2\right) \\
& + \sum_{t=1}^{n}\frac{(4\gamma - \gamma_t)_+}{2}(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*))^2,
\end{aligned}
$$

and as $(4\gamma - \gamma_t)_+ = 4\gamma + 4(\gamma_t/4 - \gamma)_+ - \gamma_t$, we can regroup and get

$$
\begin{aligned}
Regret_n \leq & \frac{2\log(K) + 5d\log(n)}{\gamma} \\
& + 2\sum_{t=1}^{n}(\gamma_t/4 - \gamma)_+ \Big((\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*))^2 - (\nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma) - \theta^*))^2 \\
& \qquad - (\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta_t(\gamma)))^2\Big) \\
& + \sum_{t=1}^{n}\frac{4\gamma - \gamma_t}{2}(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*))^2 \\
= & \frac{2\log(K) + 5d\log(n)}{\gamma} \\
& - 4\sum_{t=1}^{n}(\gamma_t/4 - \gamma)_+ \Big((\nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma) - \theta^*))(\nabla \ell_t(\hat{\theta}_t)(\theta_t(\gamma) - \hat{\theta}_t))\Big) \\
& + \sum_{t=1}^{n}\frac{4\gamma - \gamma_t}{2}(\nabla \ell_t(\hat{\theta}_t)(\hat{\theta}_t - \theta^*))^2
\end{aligned}
$$

□