HARNESSING RANDOMIZATION IN LEARNING-AUGMENTED CONTRACT SCHEDULING

Anonymous authors
Paper under double-blind review

ABSTRACT

Contract scheduling is a fundamental framework for systems that trade computation time for solution quality when the available running time is unknown. While deterministic contract schedules have been extensively studied, the benefits of randomization remain largely unexplored. In this work, we study randomized contract scheduling both in the classical, agnostic setting and in the presence of predictions. First, we settle the agnostic setting by proving a tight lower bound on the best-possible performance of a randomized schedule. We then address *learning-augmented* contract scheduling, where the scheduler receives a prediction of the interruption time. We characterize the consistency–robustness tradeoff by providing a lower bound and a randomized schedule with nearly matching guarantees. We further mitigate the brittleness inherent in consistency-optimized algorithms via smoothing, obtaining provably gradual degradation under prediction error at only a small cost in consistency. Finally, we give the first analysis of a general sequential learning game that uses historical prediction data to calibrate trust in subsequent predictions, and determine the appropriate smoothness. Our theoretical and experimental results demonstrate that randomization can significantly improve the performance of interruptible systems based on contract scheduling.

1 INTRODUCTION

The ability to trade computation time for solution quality is central to intelligent systems operating under uncertainty. In many applications, including medical diagnosis, robotic planning, autonomous systems, and financial decision making, the available computation time is determined by external events and is therefore unknown to the algorithm. For example, a medical system must be capable of producing a diagnosis at any critical time, e.g., if the patient experiences a sudden complication. Similarly, the user of an LLM may interrupt the reasoning process and request a response if reasoning is time consuming. This raises the question of designing systems whose performance improves as computation progresses, while remaining efficient when execution is terminated unexpectedly.

A rich body of work has studied this challenge through the framework of *anytime* algorithms, which provide a principled way of trading computation time for solution quality. Within this framework, (Zilberstein & Russell, 1996) introduced the distinction between *interruptible* and *contract* algorithms. Interruptible algorithms can be queried at arbitrary times during their execution and return a progressively improving solution. Contract algorithms, in contrast, require the available computation time to be specified in advance: when allowed to run for at least this prescribed contract length, they guarantee a valid output, but they may fail to provide a meaningful solution if interrupted earlier. Despite this limitation, contract algorithms are often easier to design and implement, making them attractive building blocks for systems operating under computational constraints.

Contract scheduling provides a general methodology for building interruptible systems from contract algorithms by repeatedly executing the contract algorithm with increasing contract lengths (Russell & Zilberstein, 1991). More precisely, a schedule is specified by the sequence of lengths allocated to these executions, with the objective of guaranteeing that, at any possible interruption time, the system has completed a sufficiently long contract to provide a high-quality solution. For example, consider the doubling schedule, in which the i-th contract has length 2^i. If an interruption occurs at time T, then the longest completed contract has length at least $T/4$, regardless of the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

value of T. The factor 4 captures the overhead introduced by not knowing the interruption time in advance and by allocating computation to a sequence of contracts rather than to a single execution.

Formally, given a contract algorithm A, a contract schedule is defined as an increasing sequence $X = (x_i)_{i\geq 0}$, where x_i denotes the length of the i-th contract execution of A. The performance of a schedule is measured by its *acceleration ratio* (Russell & Zilberstein, 1991). Let $\ell(X, T)$ denote the length of the longest contract completed by time T. The acceleration ratio of X is defined as

$$R(X) = \sup_T \frac{T}{\ell(X, T)}. \tag{1}$$

This measure captures the worst-case loss incurred by using a schedule that does not know the interruption time in advance. The optimal deterministic acceleration ratio is equal to 4, achieved by the doubling schedule $X = (2^i)_{i\geq 0}$, and no deterministic schedule can achieve a smaller value.

Contract scheduling has an extensive literature spanning AI and real-time systems. Beyond the single-processor setting, where the doubling schedule achieves the optimal deterministic acceleration ratio of 4, a rich body of work has investigated more complex environments including multiprocessor systems (Bernstein et al., 2002), multiple concurrent problem instances (Zilberstein et al., 2003), and combinations of these settings (Bernstein et al., 2003; López-Ortiz et al., 2014). Other directions have considered relaxed interruption models (Angelopoulos et al., 2017), alternative performance measures (Angelopoulos & López-Ortiz, 2020), adaptive schedules (Angelopoulos & Panagiotou, 2023), end guarantees on completion time (Angelopoulos & Jin, 2019), and learning-augmented settings (Angelopoulos & Kamali, 2023; Angelopoulos et al., 2024).

Contract scheduling belongs in a broader class of *sequencing* problems under uncertainty, including problems such as *searching on the infinite line* (Beck & Newman, 1970) and *online bidding* (Chrobak et al., 2008; Chrobak & Kenyon-Mathieu, 2006). Such problems require committing to a sequence of actions (such as contract lengths, search lengths, or bids) without knowing the event determining its quality (i.e., the interruption time, or hider's position, or target value, respectively). The performance is evaluated against an optimal strategy that knows the relevant event in advance, typically using a worst-case metric such as the *competitive ratio* (Borodin & El-Yaniv, 1998).

Randomization has been a powerful tool in online search and online bidding, where randomized strategies are known to outperform deterministic ones and their optimal competitive ratios are well understood (Gal, 1980; Kao et al., 1996; Chrobak et al., 2008; Angelopoulos & Simon, 2025). This raises the natural question of whether randomization can similarly improve contract scheduling. However, randomized contract schedules remain largely unexplored: the only existing result provides an upper bound of 2.455 on the randomized acceleration ratio (Angelopoulos, 2015), without showing whether this bound is tight. Moreover, even in settings where the scheduler is provided with predictions about the interruption time, all existing approaches rely on deterministic strategies (Angelopoulos & Kamali, 2023; Angelopoulos et al., 2024), and the potential benefits of randomization in this prediction-based setting remain completely unknown.

1.1 Contribution

We investigate the power of randomization in contract scheduling, both in the classical, i.e., agnostic setting, and in the presence of predictions. We first settle the agnostic setting by proving a matching lower bound on the acceleration ratio (Theorem 2). Together with the upper bound of (Angelopoulos, 2015), this establishes that the optimal randomized acceleration ratio is equal to 2.455.

We then consider the learning-augmented setting, where the scheduler receives a prediction of the interruption time. We first study the *consistency–robustness* tradeoff, a standard performance criterion in learning-augmented algorithms. Here, consistency measures the performance under error-free prediction, whereas robustness is the worst-case acceleration ratio under arbitrary prediction error (Lykouris & Vassilvitskii, 2021). We derive a lower bound on the consistency–robustness tradeoff (Theorem 3). We then design and analyze a schedule whose consistency nearly matches this lower bound for every robustness $R \geq 2.65$ (Theorem 6). For $R \in [2.455, 2.65)$, the same performance analysis applies, however the feasibility is supported by computer-assisted verification.

While our learning-augmented schedule is near-optimal with respect to consistency and robustness, it exhibits a common limitation of algorithms optimizing this tradeoff: brittleness, whereby even small prediction errors can lead to significant performance degradation. We address this issue using

a smoothing approach, showing that smooth degradation can be enforced at only a small loss in consistency (Theorem 8). Furthermore, we introduce and study a general sequential game in which the algorithm uses the oracle's past performance to estimate its error and adapt the degree of smoothness so as to optimize overall performance (Theorem 9). This departs from previous work based on a single, static prediction by allowing past prediction performance to guide future decisions. An experimental evaluation demonstrates that our schedules consistently outperform the state of the art.

From a technical perspective, our approach relies on an abstraction that we call *contract function*, extending the framework recently developed for online bidding (Degryse et al., 2026; Lee et al., 2026). This abstraction replaces discrete randomized schedules with continuous objects, allowing for a more systematic analysis. Despite this conceptual similarity, contract scheduling presents fundamentally different challenges. The key distinction is that online bidding minimizes the cost appearing in the numerator of the competitive ratio, whereas contract scheduling maximizes the longest completed contract, which appears in the denominator of the acceleration ratio. Consequently, key tools for bidding no longer apply. In particular, local exchange arguments establishing the equivalence between bidding functions and strategies do not carry over to contract scheduling, making the correspondence between contract functions and schedules considerably more delicate.

Moreover, whereas optimal bidding functions can be characterized through *positive* delay differential equations, contract scheduling gives rise to a *mixed* delay differential equation, which is significantly harder to analyze. This loss of positivity complicates both the derivation of candidate schedules and the proof of their feasibility and performance guarantees. We overcome these obstacles using tools from complex analysis such as Laplace-transform methods, which provide the analytical and computational machinery required to certify the guarantees of our schedules.

1.2 Further Related Work

Algorithms augmented with a machine-learned prediction oracle have witnessed remarkable growth in recent years, starting with the seminal works (Lykouris & Vassilvitskii, 2021; Purohit et al., 2018). In particular, several works have studied the tradeoff between performance when predictions are accurate and worst-case guarantees when predictions are unreliable. This consistency-robustness perspective has been investigated in a variety of problems, including caching, covering, energy management, and search (Sun et al., 2021; Lee et al., 2024; Benomar et al., 2025; Bamas et al., 2020; Christianson et al., 2023; Almanza et al., 2021). For contract scheduling, (Angelopoulos & Kamali, 2023) gave optimal tradeoffs for deterministic algorithms. In contrast, the study of randomized consistency-robustness tradeoffs is more recent and has been explored for problems such as ski rental (Wei & Zhang, 2020), search games (Angelopoulos et al., 2026), and online bidding, for which the recent works (Degryse et al., 2026; Lee et al., 2026) gave optimal randomized strategies.

2 A Tight Bound for Agnostic Contract Scheduling

We begin with the study of the agnostic setting, in which the scheduler has no prediction of the interruption time. Let $X = (x_i)_{i\in\mathbb{Z}}$[1] denote a schedule in which the contract lengths x_i are random variables. Following (Angelopoulos, 2015), the acceleration ratio is defined as

$$R(X) = \sup_T \frac{T}{\mathbb{E}[\ell(X,T)]}. \tag{2}$$

Define

$$R^* = \inf_{b>1} \frac{b^2 \log b}{(b-1)^2} \approx 2.455. \tag{3}$$

The following upper bound, due to (Angelopoulos, 2015) shows that an acceleration ratio at most R^* can be achieved by schedules with geometrically increasing lengths, and a random shift in $[0,1]$.

Theorem 1. *Given $b > 1$ and $\lambda \sim \mathcal{U}[0,1]$, the schedule $X = (b^{i+\lambda})_{i\in\mathbb{Z}}$ has acceleration ratio equal to $\frac{b^2\log b}{(b-1)^2}$. Thus, there exists a schedule of acceleration ratio at most R^*.*

[1] As in previous work on contract scheduling and related incremental allocation problems, we consider bi-infinite sequences; e.g., (Angelopoulos et al., 2024). This convention simplifies exposition and avoids additional technical assumptions, such as requiring that the interruption occurs after at least a contract has been completed.

We show a tight lower bound that establishes the optimality of such geometric schedules.

Theorem 2. *For any schedule X and every $\varepsilon > 0$, it holds that $R(X) \geq R^* - \varepsilon$.*

Proof sketch. Using Yao's principle (Yao, 1977), it suffices to give a matching lower bound on the expected acceleration ratio of a deterministic schedule, for an appropriately chosen distribution for the interruption; specifically, we consider a log-uniform distribution. The technical part consists of decomposing the acceleration ratio over consecutive completion-time intervals, and obtaining a telescoping sum arbitrarily close to R^*, as the support of the distribution tends to infinity. □

3 The Learning-Augmented Setting

In this section, we study the learning-augmented setting and prove near-matching upper and lower bounds on the consistency-robustness tradeoff. We begin with some preliminary definitions and notation. We denote by $\widehat{T}$ the oracle's prediction of the interruption time. In our randomized setting, the consistency and robustness are defined as natural extensions of the deterministic counterpart (Angelopoulos & Kamali, 2023), namely:

$$\mathrm{cons}(X) = \mathrm{nc}(X, \widehat{T}), \quad \text{and} \quad \mathrm{rob}(X) = \sup_T \mathrm{nc}(X, T),$$

where $\mathrm{nc}(X, T) := T/\mathbb{E}[\ell(X, T)]$ denotes the *normalized cost* of X for interruption T.

Note that the normalized cost is invariant under scaling; i.e., for every $\delta > 0$, $\mathrm{nc}(X, \widehat{T}) = \mathrm{nc}(\delta X, \delta\widehat{T})$. This implies that for every prediction, there exists a schedule X such that

$$\mathrm{cons}(X) = \inf_T \mathrm{nc}(X, T) \text{ and } \mathrm{rob}(X) = \sup_T \mathrm{nc}(X, T), \tag{4}$$

which provides a convenient equivalent definition for our metrics.

We say that X is R-robust and C-consistent if it has robustness and consistency at most R and C, respectively. Note that R must be at least the optimal agnostic acceleration ratio R^* of Eq. (3).

3.1 Lower bound

The following result establishes a lower bound on the consistency of any R-robust schedule.

Theorem 3. *Any R-robust schedule X has consistency that satisfies* $\mathrm{cons}(X) \geq \left(1 - \frac{\ln(2)}{R}\right)^{-1}$.

Proof sketch. This proof relies on establishing that, for any schedule X and any T it holds that $\frac{\mathbb{E}[\ell(X,T)]}{T} + \int_{T/2}^{T} \frac{\mathbb{E}[\ell(X,t)]}{t^2}\, dt \leq 1$. When T is the predicted interruption time, this inequality captures both consistency and robustness. It follows that $\frac{1}{\mathrm{cons}(X)} + \frac{\ln 2}{\mathrm{rob}(X)} \leq 1$, yielding the proof. We refer to Appendix B.2 for the technical details. □

3.2 Contract Functions

We define formally the concept of the *contract function*, which is instrumental in our algorithm design and analysis. This notion is inspired by the concept of the bidding function, or *profile*, introduced recently in (Degryse et al., 2026; Lee et al., 2026) in the context of the online bidding problem. While the contract and bidding functions are conceptually similar, they yield much different analyses, as discussed in Section 1.1.

Definition 4. *A contract function is a nondecreasing function $B : \mathbb{R} \to \mathbb{R}^+$ such that $\lim_{t\to-\infty} B(t) = 0$, $\lim_{t\to+\infty} B(t) = +\infty$ and $B(0) = 1$. Its induced randomized schedule X_B is defined as $\big(B(i+\lambda) - B(i-1+\lambda)\big)_{i\in\mathbb{Z}}$, where $\lambda \sim U[0,1]$.*

For a fixed shift λ, contract i completes at time $B(i+\lambda)$. Its length is therefore $B(i+\lambda) - B(i+\lambda-1)$. This definition implies the following useful interpretation of consistency and robustness.

Proposition 5. *Let B be a contract function and X_B be its induced sequence. We define the transform $\widetilde{B}(t) := \int_{t-1}^{t} B(x) - B(x-1)\,\mathrm{d}x$. Then*

$$\mathrm{cons}(B) := \mathrm{cons}(X_B) \leq \inf_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}, \quad \textit{and} \quad \mathrm{rob}(B) := \mathrm{rob}(X_B) \leq \sup_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}.$$

Contract functions provide a principled framework based on continuous functions instead of discrete sequences. Proposition 5 reduces the search for good consistency-robustness guarantees to the following continuous optimization problem P. Refer to Appendix B.1 for the proof.

$$\max \quad O = \int_{-1}^{0} B(x) - B(x-1)\,\mathrm{d}x \tag{P}$$

$$\text{s.t.} \quad \forall t \in \mathbb{R}, R\int_{t-1}^{t} B(x) - B(x-1)\,\mathrm{d}x \geq B(t) \tag{5}$$

$$B(0) = 1.$$

3.3 Upper bound

In this section, we derive and analyze our algorithm, namely a schedule induced by an appropriate contract function, denoted by B. Ideally, B would be obtained from an optimal solution to P, which is difficult to characterize directly. Instead, discretizing and solving P experimentally reveals structural properties that guide our analytical construction. Specifically, for $t < -\tau$, for some τ, B is exponential and has normalized cost equal to the robustness requirement R; for $t \in [-\tau, 0]$, B is constant, with a jump at 0^+; and for $t > 0$, B satisfies constraint (5) with equality, giving rise to a Delay Differential Equation (DDE) formulation. These requirements can be formalized as follows:

$$B(t) = e^{\alpha(t+\tau)}, \qquad \forall t \leq -\tau \tag{6}$$

$$B(t) = 1, \qquad \forall -\tau \leq t \leq 0 \tag{7}$$

$$B(0^+) = R \cdot \widetilde{B}(0), \tag{8}$$

$$B'(t) = R\big(B(t) - 2B(t-1) + B(t-2)\big), \qquad \forall t > 0. \tag{9}$$

Two technical challenges remain: determining suitable values for the parameters τ and α, and ensuring that the solution to (9) indeed defines a feasible contract function according to Definition 4.

For the parameter α, we enforce (5) with equality by choosing α as the largest solution to

$$\Phi(x) := \frac{x}{(1-e^{-x})^2} = R.$$

Having fixed α, we choose τ to maximize O. Intuitively, τ controls the extent to which the schedule exploits the prediction, balancing premature termination of a recently started contract against leaving a long contract unfinished at the predicted time. This is optimized by $\tau = 1 - \ln(2 - e^{-\alpha})/\alpha$.

It remains to ensure that the solution to the DDE (9) defines a feasible contract function. For a broad range of robustness values, namely for $R \geq R_{\mathrm{th}} := 2.65$, we can in fact use an exponential function that achieves the same consistency bound as the DDE solution and yields a feasible contract function. We formalize this result in the following theorem.

Theorem 6. *For a given $R \geq R_{\mathrm{th}} := 2.65$, let $\beta < \alpha$ be the roots of $\Phi(x) = R$, and $\tau = 1 - \frac{\ln(2-e^{-\alpha})}{\alpha}$. Define the contract function* CF

$$\mathrm{CF}(t) = \begin{cases} e^{\alpha(t+\tau)} & \textit{if } t < -\tau, \\ 1 & \textit{if } -\tau \leq t \leq 0, \\ R\tau e^{\beta t} & \textit{if } 0 < t. \end{cases} \tag{10}$$

Then, CF is R-robust and has consistency $\mathrm{cons}(\mathrm{CF}) \leq \left(1 - \frac{\ln(2-e^{-\alpha})}{\alpha}\right)^{-1}$.

Proof sketch. The main technical challenge is to show that $R \cdot \widetilde{\mathrm{CF}}(t) \geq \mathrm{CF}(t)$ for all $t \in \mathbb{R}$. We reduce this claim to a collection of inequalities involving the parameters α, β, and τ proved in Appendix B.4 □

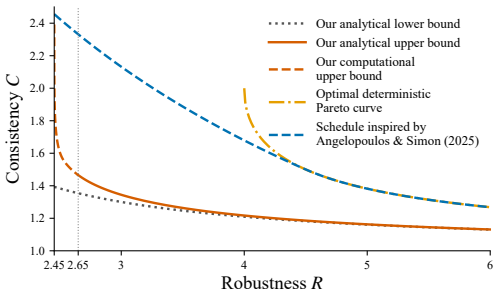


Figure 1: Consistency–robustness tradeoffs for various schedules and ranges of the robustness R. The blue dashed curve corresponds to an adaptation of the randomized online bidding algorithm of (Angelopoulos & Simon, 2025) to the contract scheduling problem. The optimal deterministic (yellow dashed-dotted) curve is due to (Angelopoulos & Kamali, 2023).

The remaining regime $R \in [2.455, 2.65)$ requires a considerably more technical approach. In this regime, we rely on explicit solutions to the DDE (9), since exponential approximations and other "simple" functions yield poor bounds. A natural approach is to solve the DDE computationally using the same choice of τ as above; however, the resulting function is infeasible for $R < 2.54$.

For smaller R, a simple exponential tail no longer gives a sufficiently strong guarantee, so we work directly with the solution of the DDE (9). Although this solution has no convenient closed form, its Laplace transform can be computed explicitly. We use this transform to choose τ. With the previous choice of τ, the coefficient of the $e^{\alpha t}$ mode becomes negative below a critical value R_c, rendering B infeasible. We therefore choose τ so that this mode is cancelled, i.e., $c_\alpha = 0$. See Appendix B.5 for the Laplace-transform derivation.

Theorem 7. *Let B be the unique solution of* (6)-(9) *with* $\tau = \begin{cases} 1 - \dfrac{\log(2 - e^{-\alpha})}{\alpha} & \text{if } R \geq R_c, \\ 1 + \dfrac{\log(p)}{\alpha} & \text{if } R \leq R_c, \end{cases}$

where, for $R \leq R_c$, p is the unique root of

$$\Xi(x) = \alpha + \log x + 1 + 2e^{-\alpha} - e^{-2\alpha} + x\big(\alpha e^{-\alpha} + (2 - e^{-\alpha}) \log x - 2(2 - e^{-\alpha})\big) = 0$$

in the interval $\left[e^{-\alpha}, \frac{1}{2-e^{-\alpha}}\right]$, *and* $R_c = \Phi(\alpha_c) \approx 2.538$ *where α_c is the unique solution of* $1 - \frac{\log(2-e^{-x})}{x} = \frac{1-e^{-x}}{2-e^{-x}} + \frac{(1-e^{-x})^2}{2x}$.

Assuming that B is feasible, for all $R \geq 2.455$, B is R-robust and $\chi(R)$-consistent, where:

$$\chi(R) = \begin{cases} \left(1 - \dfrac{\ln(2 - e^{-\alpha})}{\alpha}\right)^{-1} & \text{if } R \geq R_c, \\ \left(1 + \dfrac{\log(p) + 1 - (2 - e^{-\alpha})p}{\alpha}\right)^{-1} & \text{if } R < R_c. \end{cases}$$

Proof sketch. For $R < R_c$, cancelling the $e^{\alpha t}$ mode of the DDE solution reduces the choice of τ to $\Xi(p) = 0$, where $p = e^{\alpha(\tau-1)}$. Concavity and the endpoint signs of Ξ ensure a unique admissible p and well-defined parameters. Assuming feasibility, robustness follows from $B(t) \leq R\widetilde{B}(t)$, while evaluation at $t = 0$ gives $\mathrm{cons}(B) \leq \chi(R)$. □

Theorem 7 holds under the assumption that the resulting function B is feasible according to Definition 4. While establishing feasibility analytically is technically very challenging, it suffices to verify that $B'(t) > 0$ for all $t > 0$ and that $B(t)$ diverges to infinity. In Appendix B.7, we provide strong computational evidence that these properties hold for several values of $R \in [2.455, 2.65)$.

Figure 1 provides an illustration of our results, and a comparison to the state of the art.

4 Smoothness Guarantees

In this section, we go beyond the consistency–robustness tradeoffs, by extending the schedule of Section 3.3 so as to guarantee a *smooth* dependence of the performance on the prediction er-

ror. Smoothness is an important goal in learning-augmented algorithm design, ensuring graceful degradation as prediction error increases. This is particularly significant for algorithms optimizing consistency–robustness tradeoffs since they are often *brittle*, in that an arbitrarily small prediction error may lead to a significant degradation in performance, as noted in (Elenter et al., 2024).

To mitigate this phenomenon, we adopt the *smoothing mechanism* introduced by (Benomar & Perchet, 2025) in the context of deterministic learning-augmented algorithms. The key idea is to add a small random perturbation to the prediction so as to reduce the algorithm's sensitivity to small prediction errors. This is conceptually related to the paradigm of *smoothed analysis*, in which small random perturbations are used to alleviate worst-case phenomena (Spielman & Teng, 2004). In contrast to (Benomar & Perchet, 2025), however, our underlying schedules are themselves randomized, making the analysis of the smoothed schedules considerably more subtle.

In Section 4.1 we discuss the smoothness mechanism and prove its guarantees, namely the tradeoff between smoothness and consistency. In Section 4.2, we introduce a more general setting, in which the algorithm uses the oracle's past performance so as to choose the right amount of smoothness.

4.1 Smoothness-consistency tradeoff

Given a schedule X, a prediction $\widehat{T}$, and a *smoothing* parameter $\rho \in [0,1]$, we define by X_ρ the algorithm that relies on the perturbed prediction

$$\widetilde{T} = \frac{\widehat{T}}{1+\rho\xi}, \tag{11}$$

i.e., uses $\widetilde{T}$ as the prediction, instead of $\widehat{T}$. Here, ξ is a random perturbation defined as

$$\mathbb{P}(\xi \geq s) = \frac{1}{(1+s)^2}, \qquad s \geq 0. \tag{12}$$

The parameter ρ controls the magnitude of the perturbation, and hence the degree of smoothing, following (11). At the extreme value $\rho = 0$ there is no smoothing, hence $X_0 \equiv X$. As ρ increases, so does the smoothness, but this comes at the expense of an increased consistency.

Let us denote by CF_ρ the smoothed version of the contract-function based schedule CF we presented in Section 3.3. We define the prediction *error* η as $\eta = \max\{\frac{\widehat{T}}{T}, \frac{T}{\widehat{T}}\} - 1$, where T denotes the actual interruption time. Theorem 8 upper bounds the normalized cost of CF_ρ, as a function of the prediction error η and the parameter ρ, thus establishing smoothness. Define

$$L := \sup_{T > \widehat{T}} |T\,\mathrm{nc}'(\mathrm{CF}, T)|,\ C := \mathrm{cons}(\mathrm{CF}), \quad \Psi(\rho) := \mathbb{E}_\xi[\log(1+\rho\xi)] = \begin{cases} \frac{\rho(1-\rho+\rho\log\rho)}{(1-\rho)^2}, & \rho < 1, \\ \frac{1}{2}, & \rho = 1. \end{cases} \tag{13}$$

Theorem 8. *For any $R > R_{\mathrm{th}}$, $\rho \in (0,1]$ and prediction $\widehat{T}$, CF_ρ is R-robust and satisfies*

$$\mathrm{nc}\left(\mathit{CF}_\rho, T\right) \leq C + L\Psi(\rho) + \begin{cases} L\eta, & T \geq \widehat{T}, \\ \left(L + \dfrac{2(R-C)}{\rho}\right)\eta, & T < \widehat{T}. \end{cases} \tag{14}$$

Proof sketch. The proof first establishes the continuity of the smoothed normalized cost by dominated convergence. We then upper bound its harmonic representation by an arithmetic average, and apply logarithmic regularity to bound the loss in consistency. □

Let us explain the significance of Theorem 8. Since the term $2(R-C)/\rho$ is decreasing in ρ, the smoothness improves as ρ increases. On the other hand, as $\rho \to 0$, $\Psi(\rho) \to 0$, and thus $\mathrm{nc}(\mathrm{CF}_\rho, \widehat{T}) \to C$. Hence, CF_ρ recovers the consistency of CF as $\rho \to 0$. In addition, when $\eta = 0$, i.e., the prediction is perfect, the smoothed algorithm CF_ρ has consistency at most $C + L\Psi(\rho)$; here the term $L\Psi(\rho)$ quantifies the loss in consistency relative to CF incurred by smoothing.

4.2 Adaptive smoothness via sequential learning

While Theorem 8 establishes a useful tradeoff between smoothness and consistency, it assumes a single, static prediction $\widehat{T}$. A more challenging question is how to leverage historical data on the oracle's accuracy to learn the appropriate value of the smoothness parameter ρ.

We address this question through a novel sequential learning game. Consider a sequence of (past) interruption times T_i. On each instance i, the oracle provides a prediction $\widehat{T}_i$ before the algorithm executes its schedule, and the actual interruption time T_i is observed afterwards. We assume that the prediction–interruption pairs are i.i.d., in that $(\widehat{T}_i, T_i) \overset{\text{i.i.d.}}{\sim} P$, where P is unknown. Given the *history* $\mathcal{H}_n = ((\widehat{T}_i, T_i))_{i=1}^n$, the algorithm chooses a smoothing parameter ρ_{n+1} for the next instance. The objective is to minimize the conditional expected normalized cost, namely

$$\mathcal{L}_P(\rho_{n+1}) := \mathbb{E}_P\left[\mathrm{nc}\big(\mathrm{CF}_{\rho_{n+1}}(\widehat{T}_{n+1}), T_{n+1}\big) \,\middle|\, \mathcal{H}_n\right]. \tag{15}$$

Theorem 9 extends Theorem 8 to the adaptive setting. It provides an explicit estimator of ρ_{n+1}, and a corresponding upper bound on the expected normalized cost. We first introduce some notation. We define $\eta_- = (\widehat{T}/T - 1)_+$, $\mu = \mathbb{E}_P[\eta_-]$, and $\widehat{\mu}_n = n^{-1}\sum_{i=1}^n (\widehat{T}_i/T_i - 1)_+$, where $(x)_+ = \max(x, 0)$. Using the refined global bound (68), established as part of the proof of Theorem 8, and taking expectations over the joint distribution P of $(\widehat{T}, T)$, we obtain $\mathcal{L}_P(\rho) \le \min\{R, \mathcal{U}_P(\rho)\}$, where

$$\mathcal{U}_P(\rho) = C + L\mathbb{E}_P\left|\log(T/\widehat{T})\right| + L\Psi(\rho) + \frac{2(R-C)\mu}{\rho},$$

Define $G : (0, 1] \to (0, 1/3]$, with $G(\rho) = \rho^2\Psi'(\rho)$, and note that G is strictly increasing in $(0, 1)$. For any $\varepsilon \in (0, 1]$, define Q_ε as the inverse of G saturated to $[\varepsilon, 1]$, namely

$$Q_\varepsilon(z) = G^{-1}(\min\{G(1),\, \max\{G(\varepsilon), z\}\}).$$

Theorem 9. *Let $n \ge 1$. Suppose that $\eta < M$ almost surely for some $M > 0$. For $\varepsilon = n^{-1/4}$, consider the estimator $\rho_{n+1} = Q_\varepsilon\left(\frac{2(R-C)}{L}\widehat{\mu}_n\right)$, which minimizes the empirical estimation of $\mathcal{U}_P$ over $[\varepsilon, 1]$. Then, for any $\delta \in (0, 1)$, the following holds with probability at least $1 - \delta$:*

$$\mathcal{L}_P(\rho_{n+1}) \le \min\left\{R,\ \inf_{\rho\in[\,\varepsilon,\,1\,]} \mathcal{U}_P(\rho) + 4(R-C)M\sqrt{\frac{\log(2/\delta)}{2\sqrt{n}}}\right\}.$$

Let us explain the significance of Theorem 9. From its definition, $\hat{\mu}_n$ is the average overprediction error of the oracle over the n past samples. Since Q_ε is increasing, a larger empirical error leads to a larger smoothness parameter ρ_{n+1}. Thus, the algorithm adapts the amount of smoothing to the observed reliability of the oracle. In particular, when n is small, there is limited information on the reliability of the oracle, and the smoothness parameter is expected to be large. In contrast, as n grows, the bound converges to the tightest upper bound obtainable from Theorem 8 by optimizing the smoothing parameter ρ_{n+1} over its full range. Moreover, if $\hat{\mu}_n$ is small, i.e., the oracle is deemed reliable, then ρ_{n+1} is close to ε, and the schedule approaches the schedule CF of Theorem 6 that optimizes the consistency. We refer to Appendix C for the proofs of Theorems 8 and 9.

5 Experimental Evaluation

Benchmarks. For a given robustness $R \ge 2.65$, we evaluate our contract-function based schedule CF_ρ, defined in Section 4, with smoothness parameter $\rho \in (0, 1)$. We compare CF_ρ to AS, an adaptation to contract scheduling of the randomized online bidding algorithm of (Angelopoulos & Simon, 2025), as well as its smoothed variant AS_ρ, using the same smoothness parameter ρ as CF_ρ.

Performance evaluation. We consider an oracle that outputs a noisy prediction $\widehat{T}$ of the actual interruption time T, generated according to $\widehat{T} = Te^Z$, where $Z \sim \mathcal{N}(0, \sigma^2)$ and $\sigma^2 \in [0, 0.5]$. This means that with probability at least 0.95, $\widehat{T} \in [0.25T, 4T]$. By scaling invariance, the choice of T is immaterial, and we thus set $T = 1$.

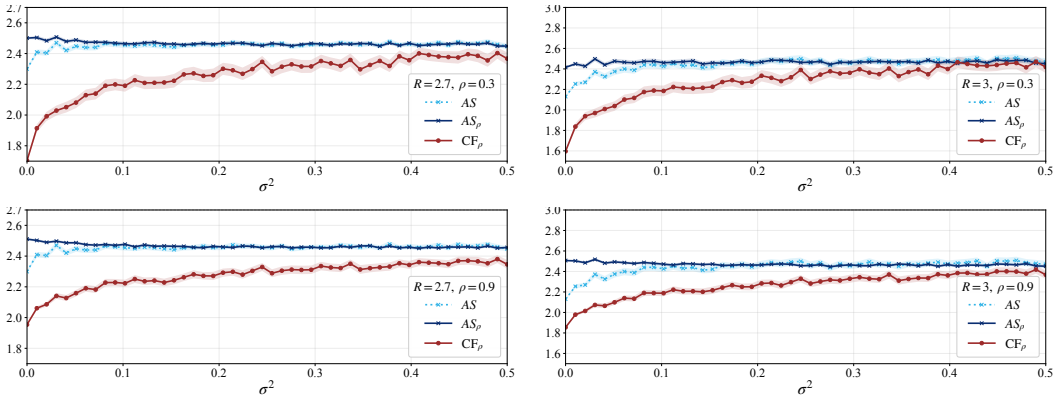


Figure 2: Empirical acceleration ratio as a function of σ^2 for $R \in \{2.7, 3\}$ and $\rho \in \{0.3, 0.9\}$.

We consider 51 values of $\sigma^2 \in [0, 0.5]$ with step size 0.01. For a given σ^2, we generate 200 samples of $\widehat{T}$. For each sample $\widehat{T}$, and every schedule X in our benchmarks, we compute the ratio $T/\widehat{\ell}(X, T)$, where $\widehat{\ell}(X, T)$ is the *average* contract length completed by T in X. Here, the average is computed over 300 independent samples for the parameters (ξ, λ) of Eq. (11) and Definition 4, respectively. The average, over the 200 samples of $\widehat{T}$, of this ratio is the average normalized cost of the schedule, to which we refer as the *empirical* acceleration ratio.

Results. Figure 2 depicts the results for $R \in \{2.7, 3\}$, and $\rho \in \{0.3, 0.9\}$. Across the entire range of prediction error considered, CF_ρ achieves a substantially better empirical acceleration ratio than AS and AS_ρ, with a particularly pronounced advantage at low and moderate error levels. The improvement is also more pronounced for smaller values of R, as expected, since this is precisely the regime in which the performance gap with prior approaches is largest. Both observations are consistent with our theoretical analysis, as summarized in Figure 1.

The experiments demonstrate that as ρ increases, the empirical acceleration ratio of CF_ρ grows more slowly with σ^2, albeit at the cost of worse empirical consistency. This behavior is consistent with the smoothing effect established in Section 4. In contrast, for AS_ρ, the improvement in smoothness does not compensate for the loss in consistency, and the original schedule AS tends to perform better. This is because smoothing is most beneficial for brittle algorithms with sharp consistency–robustness tradeoffs, whereas AS is not brittle. Supplementary experiments are given in Appendix D.

6 Conclusion

We gave the first systematic study of randomization in the design and analysis of anytime algorithms through contract scheduling. We proved a tight bound on the acceleration ratio of agnostic contract scheduling, settling an open question from (Angelopoulos, 2015), and initiated the study of learning-augmented randomized scheduling, obtaining non-trivial and near-tight consistency–robustness tradeoffs that significantly improve upon the deterministic schedules of (Angelopoulos & Kamali, 2023). We further provided the first analytical smoothness guarantees for randomized schedules and introduced a generalized prediction setting in which the scheduler uses historical information about the prediction oracle's reliability to learn the appropriate degree of smoothness.

Our results reveal that randomized contract scheduling is challenging even with a single algorithm and processor. A natural direction is to extend the contract-function framework to multiple algorithms and processors, settings previously studied only for deterministic, agnostic schedules; see (López-Ortiz et al., 2014) and many other works discussed in Section 1. More broadly, we would like to extend the ideas underlying contract functions to general randomized *sequencing* problems (Chrobak & Kenyon-Mathieu, 2006), obtaining systematic tools. Such tools are known in the deterministic setting, most notably Gal's *functional* theorem (Gal, 1980), which establishes the optimality of exponential strategies for a broad class of objectives. Developing a corresponding theory for randomized sequencing is an important broad direction for future work.

AI USE STATEMENT

In this work, we used generative AI tools for providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, for feedback on research methodology, and for implementing methods. We have not used generative AI tools for any other tasks with required disclosure. Additionally, we used such tools for editing software code and drafting parts of the paper, but for no other tasks with recommended disclosure. We carefully reviewed all AI-assisted work. Specifically: All mathematical claims were verified independently by at least three authors, and similarly for all code produced. Furthermore, the application of these tools was mainly limited to few technical parts, e.g., mainly in Lemma 12 for the concept of interval decomposition, and Lemma 14 for decomposing the result to a series of claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Matteo Almanza, Flavio Chierichetti, Silvio Lattanzi, Alessandro Panconesi, and Giuseppe Re. Online facility location with multiple advice. *Advances in Neural Information Processing Systems*, 34:4661–4673, 2021.

Spyros Angelopoulos. Further connections between contract-scheduling and ray-searching problems. In *Proc. International Joint Conference in Artificial Intelligence (IJCAI)*, pp. 1516–1522, 2015.

Spyros Angelopoulos and Shendan Jin. Earliest-completion scheduling of contract algorithms with end guarantees. In *Proc. International Joint Conference on Artificial Intelligence, (IJCAI)*, pp. 5493–5499, 2019.

Spyros Angelopoulos and Shahin Kamali. Contract scheduling with predictions. *J. Artif. Intell. Res.*, 77:395–426, 2023.

Spyros Angelopoulos and Alejandro López-Ortiz. Interruptible algorithms for multiproblem solving. *J. Sched.*, 23(4):451–464, 2020.

Spyros Angelopoulos and Konstantinos Panagiotou. Weighted online search. *J. Comput. Syst. Sci.*, 138:103457, 2023.

Spyros Angelopoulos and Bertrand Simon. Learning-augmented online bidding in stochastic settings. *Advances in Neural Information Processing Systems*, 38:67649–67679, 2025.

Spyros Angelopoulos, Alejandro López-Ortiz, and Angèle M. Hamel. Optimal scheduling of contract algorithms with soft deadlines. *J. Sched.*, 20(3):267–277, 2017.

Spyros Angelopoulos, Marcin Bienkowski, Christoph Dürr, and Bertrand Simon. Contract scheduling with distributional and multiple advice. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI*, pp. 3652–3660, 2024.

Spyros Angelopoulos, Thomas Lidbetter, and Konstantinos Panagiotou. Search games with predictions. *Operations Research*, 2026.

Etienne Bamas, Andreas Maggiori, and Ola Svensson. The primal-dual method for learning augmented algorithms. *Advances in Neural Information Processing Systems*, 33:20083–20094, 2020.

Anatole Beck and Donald J. Newman. Yet more on the linear search problem. *Israel Journal of Mathematics*, 8:419–429, 1970.

Ziyad Benomar and Vianney Perchet. On tradeoffs in learning-augmented algorithms. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Ziyad Benomar, Lorenzo Croissant, Vianney Perchet, and Spyros Angelopoulos. Pareto-optimality, smoothness, and stochasticity in learning-augmented one-max-search. In *Forty-second International Conference on Machine Learning, ICML*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025.

Daniel S. Bernstein, T. J. Perkins, Shlomo Zilberstein, and Lev Finkelstein. Scheduling contract algorithms on multiple processors. In *Proc. AAAI Conference on Artificial Intelligence*, pp. 702–706, 2002.

Daniel S. Bernstein, Lev Finkelstein, and Shlomo Zilberstein. Contract algorithms and robots on rays: Unifying two scheduling problems. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1211–1217, 2003.

Allan Borodin and Ran El-Yaniv. *Online computation and competitive analysis*. Cambridge University Press, 1998.

Nicolas Christianson, Junxuan Shen, and Adam Wierman. Optimal robustness-consistency tradeoffs for learning-augmented metrical task systems. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 9377–9399, 2023.

Marek Chrobak and Claire Kenyon-Mathieu. SIGACT news online algorithms column 10: Competitiveness via doubling. *SIGACT News*, 37(4):115–126, 2006.

Marek Chrobak, Claire Kenyon, John Noga, and Neal E. Young. Incremental medians via online bidding. *Algorithmica*, 50(4):455–478, 2008.

Mathis Degryse, Imrane Zaakour, Christoph Dürr, and Spyros Angelopoulos. The Pareto frontier of randomized learning-augmented online bidding. *arXiv preprint arXiv:2605.06106*, 2026.

Alex Elenter, Spyros Angelopoulos, Christoph Dürr, and Yanni Lefki. Overcoming brittleness in pareto-optimal learning augmented algorithms. *Advances in Neural Information Processing Systems*, 37:9329–9357, 2024.

Shmuel Gal. *Search Games*. Academic Press, 1980.

Ming-Yang Kao, John H Reif, and Stephen R Tate. Searching in an unknown environment: an optimal randomized algorithm for the cow-path problem. *Information and Computation*, 131(1): 63–80, 1996.

Changyeol Lee, Dahoon Lee, Jongseo Lee, Yongho Shin, and Changki Yun. Optimal learning-augmented algorithm for online bidding. *arXiv preprint arXiv:2605.07349*, 2026.

Russell Lee, Bo Sun, Mohammad Hajiesmaili, and John CS Lui. Online search with predictions: Pareto-optimal algorithm and its applications in energy markets. In *e-Energy*, pp. 50–71. ACM, 2024.

Alejandro López-Ortiz, Spyros Angelopoulos, and Angele Hamel. Optimal scheduling of contract algorithms for anytime problem-solving. *J. Artif. Intell. Res.*, pp. 533–554, 2014.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *J. ACM*, 68(4):24:1–24:25, 2021.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *Proc. Conference on Neural Information Processing Systems (NIPS)*, pp. 9661–9670, 2018.

Stuart J. Russell and Shlomo Zilberstein. Composing real-time systems. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 212–217, 1991.

Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004. doi: 10.1145/990308.990310.

Bo Sun, Russell Lee, Mohammad Hajiesmaili, Adam Wierman, and Danny Tsang. Pareto-Optimal Learning-Augmented Algorithms for Online Conversion Problems. In *Advances in Neural Information Processing Systems*, volume 34, pp. 10339–10350, 2021.

Alexander Wei and Fred Zhang. Optimal robustness-consistency trade-offs for learning-augmented online algorithms. *Advances in Neural Information Processing Systems*, 33:8042–8053, 2020.

Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pp. 222–227. IEEE Computer Society, 1977.

Shlomo Zilberstein and Stuart J. Russell. Optimal composition of real-time systems. *Artif. Intell.*, 82(1-2):181–213, 1996.

Shlomo Zilberstein, Francois Charpillet, and Philippe Chassaing. Optimal sequencing of contract algorithms. *Ann. Math. Artif. Intell.*, 39(1-2):1–18, 2003.

A SUPPLEMENTARY MATERIAL FOR SECTION 2

Proof of Theorem 2. The acceleration ratio of schedule X can be written equivalently as:

$$R(X) = \frac{1}{\inf_T \mathbb{E}\left[\frac{\ell(X,T)}{T}\right]}. \tag{16}$$

Let D denote a distribution over interruption times. From (Yao, 1977), it follows that:

$$\inf_T \ \mathbb{E}\left[\frac{\ell(X,T)}{T}\right] \leq \sup_A \ \mathbb{E}_{T\sim D}\left[\frac{\ell(A,T)}{T}\right], \tag{17}$$

where $A = (a_i)_{i\in\mathbb{Z}}$ denotes a *deterministic* schedule.

Consider the distribution D_ε, for any given $\varepsilon \in (0,1)$, as follows. Define

$$S = e^{\frac{1-\varepsilon}{\varepsilon}}. \tag{18}$$

D_ε has density ε/t on $[1,S)$, and an atom of mass equal to ε at S. It is easy to verify that D_ε is a valid distribution. For any deterministic schedule A, the definition of D_ε implies that

$$\mathbb{E}_{T\sim D_\varepsilon}\left[\frac{\ell(A,T)}{T}\right] = \varepsilon\left(\frac{\ell(A,S)}{S} + \int_1^S \frac{\ell(A,t)}{t^2}\,dt\right). \tag{19}$$

Let $y_i = \sum_{k\leq i} a_k$ denote the completion time of contract i in A. Denote also by p the first index such that $y_p \geq 1$ and by q the last index such that $y_q \leq S$. Two cases arise: either there exists an index i such that $y_i \in [1,S]$, or no contract is completed during $[1,S]$. In the first case, $y_p \leq y_q$, whereas in the second case, $y_p > y_q$. The latter case is straightforward and will be handled separately at the end of this proof, hence we can assume that $p \leq q$. Then, (19) gives

$$\mathbb{E}_{T\sim D_\varepsilon}\left[\frac{\ell(A,T)}{T}\right] \leq \varepsilon\left(1 + \int_1^{y_p} \frac{\ell(A,t)}{t^2}\,dt + \sum_{i=p}^{q-1}\int_{y_i}^{y_{i+1}} \frac{\ell(A,t)}{t^2}\,dt + \int_{y_q}^{S} \frac{\ell(A,t)}{t^2}\,dt\right), \tag{20}$$

For every $i \in \mathbb{Z}$, define

$$z_i := \frac{a_i}{y_i} = \frac{y_i - y_{i-1}}{y_i} = 1 - \frac{y_{i-1}}{y_i} \leq 1.$$

With this notation, we obtain

$$\int_{y_i}^{y_{i+1}} \frac{\ell(A,t)}{t^2}\,dt = a_i\left(\frac{1}{y_i} - \frac{1}{y_{i+1}}\right) = z_i z_{i+1} \leq 1. \tag{21}$$

Using (21), it follows that the first and last integrals of (20) are each at most 1, therefore

$$\mathbb{E}_{T\sim D_\varepsilon}\left[\frac{\ell(A,T)}{T}\right] \le \varepsilon\left(3+\sum_{i=p}^{q-1}\int_{y_i}^{y_{i+1}}\frac{\ell(A,t)}{t^2}\,dt\right). \tag{22}$$

Define $\alpha = 1/R^*$, then from (3) α can be written equivalently as

$$\alpha = \max_{z\in(0,1)}\frac{z^2}{-\log(1-z)}, \tag{23}$$

which also implies that for every $z \in (0,1)$

$$z^2 \le -\alpha(\log(1-z)). \tag{24}$$

Using $uv \le (u^2+v^2)/2$ and (24), we obtain, for all $u, v \in (0,1)$

$$uv \le \frac{u^2+v^2}{2} \le \frac{\alpha}{2}\big(-\log(1-u)-\log(1-v)\big). \tag{25}$$

Applying (25) with $u = z_i$ and $v = z_{i+1}$, implies after some steps that

$$\sum_{i=p}^{q-1} z_i z_{i+1} \le 1 + \sum_{i=p+1}^{q-1} z_i z_{i+1} \le 1 + \alpha\sum_{i=p+1}^{q} -\log(1-z_i). \tag{26}$$

If $p = q$, the sum above is empty. The right-hand side of (26) is a telescoping sum since $-\log(1-z_i) = \log\frac{y_i}{y_{i-1}}$. Namely

$$\sum_{i=p+1}^{q} -\log(1-z_i) = \log\left(\prod_{i=p+1}^{q}\frac{y_i}{y_{i-1}}\right) = \log\big(\frac{y_q}{y_p}\big) \le \log S.$$

Therefore (26) gives

$$\sum_{i=p}^{q-1} z_i z_{i+1} \le 1 + \alpha\log S. \tag{27}$$

Combining (21), (22) and (27) we obtain

$$\mathbb{E}_{T\sim D_\varepsilon}\left[\frac{\ell(A,T)}{T}\right] \le \varepsilon\,(4+\alpha\log S) = \frac{4+\alpha\log S}{1+\log S}, \tag{28}$$

where the last equation follows from (18). Last, note that for any $\delta > 0$ the right-hand side of (28) is at most $\alpha + \delta$, for sufficiently large S (i.e., for sufficiently small ε). The proof follows then from (17) and (16).

To conclude, we handle the remaining case where the algorithm does not complete any contract in $[1, S]$. In this case, $\ell(A,T)$ is constant for $T \in [1,S]$, and is equal to $a_q < 1$. Choose S sufficiently large so that $\varepsilon < \alpha + \delta$.

$$\begin{aligned}\mathbb{E}_{T\sim D_\varepsilon}\left[\frac{\ell(A,T)}{T}\right] &= \varepsilon\left(\frac{a_q}{S}+\int_1^S\frac{a_q}{t^2}\,dt\right)\\ &\le \varepsilon\left(\frac{1}{S}+a_q(1-\frac{1}{S})\right) \le \varepsilon < \alpha+\delta,\end{aligned}$$

and the analysis falls back to the general case. □

B SUPPLEMENTARY MATERIAL FOR SECTION 3

B.1 JUSTIFICATION OF THE OPTIMIZATION PROBLEM P

For a contract function B, recall the notation

$$\widetilde{B}(t) := \int_{t-1}^{t} \big(B(x) - B(x-1)\big)\, dx,$$

By Proposition 5, the randomized schedule X_B induced by B satisfies

$$\mathrm{cons}(X_B) \le \inf_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}, \qquad \mathrm{rob}(X_B) \le \sup_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}.$$

Fix a robustness requirement R. We show that optimizing the consistency guarantee supplied by Proposition 5 is equivalent to Problem P.

Proposition 10. *Let $\mathcal{B}$ be the set of contract functions satisfying*

$$B(t) \le R\widetilde{B}(t), \qquad t \in \mathbb{R}.$$

We have

$$\inf_{B\in\mathcal{B}} \inf_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)} = \frac{1}{O_R^\star},$$

where

$$O_R^\star := \sup\left\{ \int_{-1}^{0} \big(B(x) - B(x-1)\big)\, dx : B \in \mathcal{B},\ B(0) = 1 \right\}.$$

Consequently, every feasible solution of Problem P with objective value O induces an R-robust schedule whose consistency is at most $1/O$.

Proof. We first observe that the robustness constraint in Problem P is precisely

$$B(t) \le R\widetilde{B}(t), \qquad t \in \mathbb{R},$$

or equivalently $\forall t \in \mathbb{R}, B(t)/\widetilde{B}(t) \le R$. Hence, by Proposition 5, every feasible contract function induces an R-robust schedule. Moreover, since $B(0) = 1$, the objective of Problem P is

$$O = \int_{-1}^{0} \big(B(x) - B(x-1)\big)\, dx = \widetilde{B}(0),$$

and therefore $B(0)/\widetilde{B}(0) = 1/O$. Proposition 5 then immediately yields

$$\mathrm{cons}(X_B) \le \inf_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)} \le \frac{B(0)}{\widetilde{B}(0)} = \frac{1}{O}.$$

Thus maximizing O minimizes the consistency guarantee obtained at $t = 0$. It remains to show that fixing the distinguished point to be $t = 0$ causes no loss of generality. Let $B \in \mathcal{B}$ and fix any $s \in \mathbb{R}$. Consider the translated and normalized contract function

$$B_s(t) := \frac{B(t+s)}{B(s)}.$$

Clearly $B_s(0) = 1$, and B_s is again a contract function. Its transform is

$$\begin{aligned}
\widetilde{B}_s(t) &= \int_{t-1}^{t} \big(B_s(x) - B_s(x-1)\big)\, dx \\
&= \frac{1}{B(s)} \int_{t-1}^{t} \big(B(x+s) - B(x+s-1)\big)\, dx \\
&= \frac{\widetilde{B}(t+s)}{B(s)}.
\end{aligned}$$

Consequently $B_s(t)/\widetilde{B}_s(t) = B(t+s)/\widetilde{B}(t+s)$. In particular, the robustness certificate is invariant under this translation:

$$\sup_t \frac{B_s(t)}{\widetilde{B}_s(t)} = \sup_t \frac{B(t)}{\widetilde{B}(t)} \leq R.$$

Hence B_s is feasible whenever B is feasible. On the other hand, its objective value is $\widetilde{B}_s(0) = \frac{\widetilde{B}(s)}{B(s)}$. Thus, translating B by s makes the consistency certificate at s become the certificate at 0. Taking the supremum over all s gives

$$\sup_{s\in\mathbb{R}} \widetilde{B}_s(0) = \sup_{s\in\mathbb{R}} \frac{\widetilde{B}(s)}{B(s)} = \frac{1}{\inf_{s\in\mathbb{R}} \frac{B(s)}{\widetilde{B}(s)}}.$$

Finally, taking the supremum over all R-feasible contract functions gives

$$\begin{aligned} O_R^\star &= \sup_{B\in\mathcal{B}} \sup_{s\in\mathbb{R}} \frac{\widetilde{B}(s)}{B(s)} \\ &= \frac{1}{\inf\limits_{B\in\mathcal{B}} \inf\limits_{s\in\mathbb{R}} \dfrac{B(s)}{\widetilde{B}(s)}}, \end{aligned}$$

which concludes the proof. □

B.2 Proof of Theorem 3

Proof. Let A be a deterministic algorithm. We denote by $\ell(A,T)$ the largest completed contract by A by time T.

Let $T > 0$ and let b be the last completion time at or before T and a be the preceding completion time so that $\ell(A,T) = b - a$. We want to prove that

$$\frac{\ell(A,T)}{T} + \int_{T/2}^{T} \frac{\ell(A,t)}{t^2}\, dt \leq 1. \tag{29}$$

We first treat the case of T being a completion time, i.e. $b = T$. Let x and y be the last two completed contract lengths, with $x = b - a$. We distinguish two cases:

- If $a \leq T/2$, then

$$\frac{\ell(A,T)}{T} + \int_{T/2}^{T} \frac{\ell(A,t)}{t^2}\, dt = \frac{x}{T} + y\left(\frac{2}{T} - \frac{1}{T}\right) = \frac{x+y}{T} \leq 1.$$

- If $a > T/2$, then

$$\frac{\ell(A,T)}{T} + \int_{T/2}^{T} \frac{\ell(A,t)}{t^2}\, dt \leq \frac{x}{T} + x\left(\frac{2}{T} - \frac{1}{T}\right) = \frac{2x}{T} \leq 1,$$

  where the last inequality comes from $x = b - a \leq T - T/2 = T/2$.

So (29) holds when T is a completion time. We now consider an arbitrary T:

- If $b \leq T/2$, then for all $t \in [T/2, T]$, we have $\ell(A,t) = b - a$. Therefore

$$\begin{aligned} &\frac{\ell(A,T)}{T} + \int_{T/2}^{T} \frac{\ell(A,t)}{t^2}\, dt \\ &\qquad = \frac{b-a}{T} + (b-a)\left(\frac{2}{T} - \frac{1}{T}\right) \\ &\qquad = \frac{2(b-a)}{T} \leq \frac{2b}{T} \leq 1. \end{aligned}$$

- If $b > T/2$, we have

$$\begin{aligned}
&\frac{\ell(A,T)}{T} + \int_{T/2}^{T} \frac{\ell(A,t)}{t^2}\,dt \\
&= \frac{b-a}{T} + \int_{T/2}^{b} \frac{\ell(A,t)}{t^2}\,dt + \int_{b}^{T} \frac{\ell(A,t)}{t^2}\,dt \\
&= \frac{b-a}{T} + \int_{T/2}^{b} \frac{\ell(A,t)}{t^2}\,dt + (b-a)\left(\frac{1}{b} - \frac{1}{T}\right) \\
&\leq \frac{b-a}{b} + \int_{b/2}^{b} \frac{\ell(A,t)}{t^2}\,dt \\
&= \frac{\ell(A,b)}{b} + \int_{b/2}^{b} \frac{\ell(A,t)}{t^2}\,dt \leq 1,
\end{aligned}$$

  where the last inequality is given by (29) applied to b, which is a completion time.

Let $\mathcal{A}$ be a randomized algorithm. Taking the expectation in the above inequality gives us, for all $T > 0$,

$$\frac{\mathbb{E}[\ell(\mathcal{A},T)]}{T} + \int_{T/2}^{T} \frac{\mathbb{E}[\ell(\mathcal{A},t)]}{t^2}\,dt \leq 1.$$

Choose T such that $\mathbb{E}[\ell(\mathcal{A},T)]/T \geq \frac{1}{C} - \varepsilon$. We have as well $\mathbb{E}[\ell(\mathcal{A},t)]/t \geq 1/R$ by definition of the robustness, for all $t \in [T/2, T]$. Hence we get

$$\frac{1}{C} - \varepsilon + \frac{1}{R}\big(\ln(T) - \ln(T/2)\big) = \frac{1}{C} + \frac{\ln(2)}{R} - \varepsilon \leq 1.$$

As ε tends to 0, we get the desired inequality. □

B.3 Proof of Proposition 5

We first prove that for every $s \in \mathbb{R}$ such that $B(s) \leq T$,

$$\mathbb{E}[\ell(X_B, T)] \geq \widetilde{B}(s). \tag{30}$$

We fix some interruption time T, $s \in \mathbb{R}$ with $B(s) \leq T$ and a random seed $\lambda \in [0,1)$. Let $j := \lfloor s - \lambda \rfloor$. By definition of a contract function, B is nondecreasing, so we have

$$B(j+\lambda) \leq B(s) \leq T.$$

Thus contract j is already completed at time T, and has length $D(j+\lambda) := B(j+\lambda) - B(j+\lambda-1)$, hence we get

$$\ell(X_B, T) \geq D(j+\lambda).$$

Now notice that $\sigma = s - \lambda - \lfloor s - \lambda \rfloor$ is also uniform on $[0,1)$, and $j + \lambda = s - \sigma$, so taking the expectation over σ instead of λ gives us (30):

$$\begin{aligned}
\mathbb{E}[\ell(X_B,T)] &\geq \int_0^1 D(j+\lambda)\,d\lambda \\
&= \int_0^1 D(s-\sigma)\,d\sigma \\
&= \widetilde{B}(s).
\end{aligned}$$

For the consistency inequality, fix $t \in \mathbb{R}$ and consider $T = B(t)$ and $s = t$. We obtain

$$\mathrm{nc}(X_B, B(t)) \leq \frac{B(t)}{\widetilde{B}(t)}.$$

Taking the infimum over $t \in \mathbb{R}$,

$$\mathrm{cons}(X_B) \leq \inf_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}.$$

For the robustness, we use the right limit

$$B(r^+) := \lim_{\substack{s\downarrow r \\ s>r}} B(s),$$

and the generalized inverse

$$r_B(T) = \sup\{s \in \mathbb{R} : B(s) \leq T\}.$$

Fix $T > 0$, for every $s < r_B(T)$, we have $B(s) \leq T$, therefore (30) gives

$$\mathbb{E}[\ell(X_B, T)] \geq \widetilde{B}(s).$$

By continuity of $\widetilde{B}$, as $s \uparrow r_B(T)$, we obtain

$$\mathbb{E}[\ell(X_B, T)] \geq \widetilde{B}(r_B(T)),$$

which means that

$$\begin{aligned}\mathrm{nc}(X_B, T) &\leq \frac{T}{\widetilde{B}(r_B(T))} \\ &\leq \frac{B(r_B(T)^+)}{\widetilde{B}(r_B(T))} \\ &\leq \sup_{t\in\mathbb{R}} \frac{B(t)}{\widetilde{B}(t)}.\end{aligned}$$

We finish the proof by taking the supremum over T, obtaining $\mathrm{rob}(X_B) \leq \sup_{t\in\mathbb{R}} B(t)/\widetilde{B}(t)$.

B.4 Proof of Theorem 6

We first establish in Lemmas 11-14 several inequalities on the parameters α, β and τ. These bounds are then used to verify the robustness requirement $\mathrm{CF}(t) \leq R\widetilde{\mathrm{CF}}(t)$ for every $t \in \mathbb{R}$. Since the expression of $\widetilde{\mathrm{CF}}(t)$ changes at the breakpoints of CF, the proof proceeds by considering the corresponding intervals separately.

Lemma 11. *Let $R \geq R_{\mathrm{th}}$, let $\beta < \alpha$ be the two positive roots of $\Phi(x) = R$, and $\tau = 1 - \frac{\ln(2-e^{-\alpha})}{\alpha}$. We have the following numerical bounds:*

$$1.948 \leq \alpha, \tag{31}$$
$$\beta \leq 0.789, \tag{32}$$
$$e^{-\alpha} \leq 0.145, \tag{33}$$
$$0.682 \leq \tau < 1, \tag{34}$$
$$(1 - e^{-\alpha})^2 > 0.73, \tag{35}$$
$$\beta e^{-\beta}(2 - e^{-\beta}) < 0.558, \tag{36}$$
$$2\beta e^{-\beta} < 0.719 \tag{37}$$

And for the specific case $R = R_{\mathrm{th}} = 2.65$, we have

$$1.948 \leq \alpha_{\mathrm{th}} \leq 1.949 \tag{38}$$
$$0.788 \leq \beta_{\mathrm{th}} \leq 0.789 \tag{39}$$
$$0.682 \leq \tau_{\mathrm{th}} \leq 0.683 \tag{40}$$

Proof. For every R, the values $\beta(R), \alpha(R) > 0$ are the two solutions of $R = \Phi(x)$, where $\Phi(x) = x/(1-e^{-x})^2$. Notice that Φ is decreasing on $(0, r^*)$ and increasing on $(r^*, +\infty)$, where r^* is around 1.256 and $\Phi(r^*) = R^*$. Hence $\beta(R)$ is decreasing, whereas $\alpha(R)$ is increasing.

We have $\Phi(1.948) < R_{\text{th}} < \Phi(1.949)$ and $\Phi(0.789) < R_{\text{th}} < \Phi(0.788)$, which gives us the two inequalities (38) and (39).

For (34) and (40), we prove that τ is an increasing function of α. Indeed,

$$\frac{\mathrm{d}\tau}{\mathrm{d}\alpha} = \frac{\ln(2 - e^{-\alpha}) - \frac{\alpha e^{-\alpha}}{2-e^{-\alpha}}}{\alpha^2}.$$

Let $h(\alpha) = \ln(2 - e^{-\alpha}) - \frac{\alpha e^{-\alpha}}{2-e^{-\alpha}}$, we have $h(0) = 0$ and $\frac{\mathrm{d}h}{\mathrm{d}\alpha} = \frac{2\alpha e^{-\alpha}}{(2-e^{-\alpha})^2} > 0$. Hence h is positive and so τ is increasing in α. Now recall that α is increasing in R, so $\frac{\mathrm{d}\tau}{\mathrm{d}R} > 0$. Computations then give us the inequalities.

Every other inequality follows from monotonicity, namely:

- $e^{-\alpha}$ is decreasing in α.
- $(1 - e^{-\alpha})^2$ is increasing in α.
- $\beta e^{-\beta}(2 - e^{-\beta})$ is increasing in β.
- $2\beta e^{-\beta}$ is increasing in β,

which concludes the proof. □

Lemma 12. *Let $R \geq R_{\text{th}}$, and let $\beta(R) < \alpha(R)$ be the two positive solutions of $x = R(1 - e^{-x})^2$. The function $R \mapsto \alpha(R)\beta(R)$ is decreasing.*

Proof. Since $R \mapsto \alpha(R)$ is strictly increasing, define b by

$$b(\alpha(R)) = \beta(R).$$

Thus $b(\alpha)$ is the small root corresponding to the large root α. Let $P(\alpha) = \alpha b(\alpha)$; it suffices to prove that $\mathrm{d}P/\mathrm{d}\alpha \leq 0$. We have $\Phi(x) = x/(1 - e^{-x})^2$, hence $\log \Phi(x) = \log x - 2\log(1 - e^{-x})$, which gives

$$\frac{x\Phi'(x)}{\Phi(x)} = 1 - \frac{2x}{e^x - 1}.$$

We know that $\Phi(b(\alpha)) = \Phi(\alpha)$. Differentiating with respect to α gives

$$\Phi'(b(\alpha))\frac{\mathrm{d}b}{\mathrm{d}\alpha} = \Phi'(\alpha).$$

Let $L(x) = x\Phi'(x)/\Phi(x)$. We have

$$\frac{\mathrm{d}b}{\mathrm{d}\alpha} = \frac{\Phi'(\alpha)}{\Phi'(b(\alpha))} = \frac{b(\alpha)L(\alpha)}{\alpha L(b(\alpha))},$$

and finally

$$\frac{\mathrm{d}P}{\mathrm{d}\alpha} = b(\alpha) + \alpha\frac{\mathrm{d}b}{\mathrm{d}\alpha} = \frac{b(\alpha)\big(L(\alpha) + L(b(\alpha))\big)}{L(b(\alpha))}.$$

Note that $L(x)$ is negative on the small-root branch, which is the range of b, and positive on the large-root branch, which is the domain of α. Hence

$$\frac{\mathrm{d}P}{\mathrm{d}\alpha} \leq 0 \quad \Longleftrightarrow \quad L(\alpha) + L(b(\alpha)) \geq 0,$$

which itself is equivalent to $\alpha/(e^\alpha - 1) + b(\alpha)/(e^{b(\alpha)} - 1) \leq 1$. In order to prove this inequality, we introduce $A(x) = x/(e^x - 1)$; we need $A(\alpha) + A(b(\alpha)) \leq 1$.

Since $R \geq R_{\text{th}}$, we have $\alpha \geq 1.948$, we start by proving this inequality when $1.948 \leq \alpha \leq 10$.

Note that A is decreasing on $\mathbb{R}^+$, indeed

$$A'(x) = \frac{e^x - 1 - xe^x}{(e^x - 1)^2} < 0.$$

The function b is decreasing. Choose a partition of $[1.948, 10]$ and consider one of its intervals $[u, v]$. For every $\alpha \in [u, v]$, we have

$$\alpha \geq u, \qquad b(\alpha) \geq b(v).$$

Since A is decreasing, we get that $A(\alpha) \leq A(u)$ and $A(b(\alpha)) \leq A(b(v))$. Hence

$$A(\alpha) + A(b(\alpha)) \leq A(u) + A(b(v)).$$

We will verify that $A(u) + A(b(v)) \leq 1$, for each $[u, v]$. The following nine intervals suffice; U denotes the certified upper bound $A(u) + A(b(v))$, and the last column is $1 - U$.

Table 1: Interval certificate for $1.948 \leq \alpha \leq 10$.

interval $[u, v]$	$b(v)$ enclosure	U	$1 - U$
$[1.948, 2.000]$	$[0.765860106, 0.765860107]$	0.989353357	0.010646643
$[2.000, 2.070]$	$[0.736546629, 0.736546630]$	0.989566832	0.010433168
$[2.070, 2.160]$	$[0.701526530, 0.701526531]$	0.988840464	0.011159536
$[2.160, 2.290]$	$[0.655639525, 0.655639526]$	0.989323050	0.010676950
$[2.290, 2.490]$	$[0.594226405, 0.594226406]$	0.990170469	0.009829531
$[2.490, 2.800]$	$[0.516405749, 0.516405750]$	0.989031536	0.010968464
$[2.800, 3.360]$	$[0.413313644, 0.413313645]$	0.988831081	0.011168919
$[3.360, 4.670]$	$[0.274885966, 0.274885967]$	0.989756281	0.010243719
$[4.670, 10.000]$	$[0.111690346, 0.111690347]$	0.989376762	0.010623238

It remains to prove the case $\alpha \geq 10$.

Since $\alpha \geq 10$, we must have $b(\alpha) \leq 1$ (in fact even close to 0.1). We have the inequality

$$\forall x \in (0, 1], \qquad A(x) \leq 1 - \frac{x}{3},$$

indeed, $e^x - 1 \geq x(1 + x/2)$ and so $A(x) \leq 1/(1 + x/2) \leq 1 - x/3$. This last inequality can be seen as $(1 - x/3)(1 + x/2) = 1 + x(1 - x)/6 \geq 1$.

Applying this new inequality to $b(\alpha)$ gives $A(b(\alpha)) \leq 1 - b(\alpha)/3$, so we only need to prove that $A(\alpha) \leq b(\alpha)/3$. We have, since $\alpha \geq 10$,

$$A(\alpha) = \frac{\alpha}{e^\alpha - 1} \leq 2\alpha e^{-\alpha} \leq \frac{1}{6\alpha},$$

where the last inequality uses that $12\alpha^2 e^{-\alpha}$ is decreasing for $\alpha \geq 10$ and $1200e^{-10} < 1$. We need a last relation between $b(\alpha)$ and α. We have $1 - e^{-b(\alpha)} \leq b(\alpha)$, so $\Phi(b(\alpha)) \geq 1/b(\alpha)$. Since $\Phi(b(\alpha)) = \Phi(\alpha)$, we have

$$b(\alpha) \geq \frac{1}{\Phi(\alpha)} = \frac{(1 - e^{-\alpha})^2}{\alpha} \geq \frac{1}{2\alpha}.$$

The last inequality also holds because $\alpha \geq 10$. This finally gives us $A(\alpha) \leq 1/(6\alpha) \leq b(\alpha)/3$, which means $\mathrm{d}P/\mathrm{d}\alpha \leq 0$. □

Lemma 13. *Let $R \geq R_{\text{th}}$, let $\beta < \alpha$ be the solutions of $\Phi(x) = R$, and let $\tau = 1 - \ln(2 - e^{-\alpha})/\alpha$. It holds that*

$$\tau R e^{-\beta\tau} > 1.04.$$

Proof. We make the dependence on R explicit and define

$$g(R) = \tau(R) R e^{-\beta(R)\tau(R)},$$
$$\tau(R) = 1 - \frac{\ln(2 - e^{-\alpha(R)})}{\alpha(R)}.$$

We show that $\mathrm{d}g/\mathrm{d}R \geq 0$. We have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\tau}{\alpha}\right) = \frac{2\ln(2-e^{-\alpha}) - \alpha - \frac{\alpha e^{-\alpha}}{2-e^{-\alpha}}}{\alpha^3} \leq \frac{2\ln(2)-\alpha}{\alpha^3} \leq 0 \tag{41}$$

using that $\alpha \geq 1.948$ from (31). Moreover,

$$\beta(R)\tau(R) = (\alpha(R)\beta(R)) \cdot \left(\frac{\tau(R)}{\alpha(R)}\right).$$

Both factors on the right hand side are decreasing in R, the first one by Lemma 12 and the second one by equation (41); hence $\beta(R)\tau(R)$ is decreasing. Since $\tau(R)$ is increasing, all three positive factors in $g(R)$ are increasing, and therefore $\mathrm{d}g/\mathrm{d}R \geq 0$. We obtain

$$g(R) \geq R_{\mathrm{th}}\tau(R_{\mathrm{th}})e^{-\beta(R_{\mathrm{th}})\tau(R_{\mathrm{th}})}.$$

Now recall that

$$\begin{aligned} 1.948 < \alpha(R_{\mathrm{th}}) < 1.949 \\ 0.788 < \beta(R_{\mathrm{th}}) < 0.789 \\ 0.682 < \tau(R_{\mathrm{th}}) < 0.683. \end{aligned}$$

This gives us that $\beta(R_{\mathrm{th}})\tau(R_{\mathrm{th}}) \leq 0.538887$ and thus

$$g(R) \geq 0.682 \cdot R_{\mathrm{th}} \cdot e^{-0.538887} > 1.04.$$

□

Lemma 14. *Let $R \geq R_{\mathrm{th}}$, and let $0 < \beta < \alpha$ be the two solutions of*

$$\Phi(x) = \frac{x}{(1-e^{-x})^2} = R.$$

Let $\tau = 1 - \frac{\ln(2-e^{-\alpha})}{\alpha}$, $K = \frac{R\tau}{\beta}$, $\delta = K - \tau$, and $m = \delta - 2Ke^{-\beta}$. Finally, define

$$\begin{aligned} F(t) &:= \delta e^{\beta t} - K - 2(t-1) - \tau + \frac{e^{\alpha(\tau+t-2)}-1}{\alpha}, \\ G(t) &:= m\beta e^{\beta t} + e^{\alpha(\tau+t-2)}. \end{aligned}$$

Then G is nondecreasing on $[1, 2-\tau]$ and:

$$G(2-\tau) < 0, \tag{42}$$
$$\delta\beta^2 \leq \alpha, \tag{43}$$
$$m < 0, \tag{44}$$
$$F(t) \geq 0, \qquad 1-\tau \leq t \leq 1. \tag{45}$$

Proof. Since $\Phi(\beta) = R$, we first have

$$K = \frac{R\tau}{\beta} = \frac{\tau}{(1-e^{-\beta})^2}. \tag{46}$$

Consequently,

$$\delta = \tau\frac{e^{-\beta}(2-e^{-\beta})}{(1-e^{-\beta})^2}, \qquad m = -\tau\frac{e^{-2\beta}}{(1-e^{-\beta})^2}. \tag{47}$$

We prove the four claims separately.

Proof of (42)**.** Differentiating gives

$$G'(t) = m\beta^2 e^{\beta t} + \alpha e^{\alpha(\tau+t-2)}.$$

Using (47) and $\Phi(\alpha) = \Phi(\beta)$, the inequality $G'(t) \geq 0$ is equivalent to

$$e^{\alpha(\tau+t-2)+\beta(2-t)} \geq \frac{\tau\beta}{(1-e^{-\alpha})^2}. \tag{48}$$

Since $\alpha > \beta$, the left-hand side is increasing in t. It is therefore enough to verify (48) at $t = 1$. Using $e^{\alpha(\tau-1)} = \frac{1}{2-e^{-\alpha}}$, the desired inequality becomes

$$\tau\beta e^{-\beta}(2 - e^{-\alpha}) \leq (1 - e^{-\alpha})^2.$$

By Lemma 11,

$$\tau\beta e^{-\beta}(2 - e^{-\alpha}) \leq 2\beta e^{-\beta} < 0.719 < 0.73 < (1 - e^{-\alpha})^2.$$

Thus $G'(t) \geq 0$ for every $t \in [1, 2 - \tau]$. At the right endpoint,

$$\begin{aligned} G(2-\tau) &= 1 + \beta m e^{\beta(2-\tau)} \\ &= 1 - \frac{\tau\beta e^{-\beta\tau}}{(1-e^{-\beta})^2} \\ &= 1 - \tau R e^{-\beta\tau}. \end{aligned}$$

Lemma 13 gives $\tau R e^{-\beta\tau} > 1.04$, and therefore $G(2 - \tau) < 0$.

Proof of (43)**.** Using $\Phi(\alpha) = \Phi(\beta)$ and (47), we obtain

$$\begin{aligned} \delta\beta^2 &= \tau\beta^2 \frac{e^{-\beta}(2 - e^{-\beta})}{(1-e^{-\beta})^2} \\ &= \tau\alpha \frac{\beta e^{-\beta}(2 - e^{-\beta})}{(1-e^{-\alpha})^2}. \end{aligned}$$

By Lemma 11, we have $\tau < 1$ as well as $\beta e^{-\beta}(2 - e^{-\beta}) < 0.558$ and $(1 - e^{-\alpha})^2 > 0.73$. Hence

$$\frac{\delta\beta^2}{\alpha} < \frac{0.558}{0.73} < 1,$$

which proves (43).

Proof of (44)**.** The second identity in (47) gives directly

$$m = -\tau \frac{e^{-2\beta}}{(1-e^{-\beta})^2} < 0.$$

Proof of (45)**.** We have

$$\begin{aligned} F'(t) &= \delta\beta e^{\beta t} - 2 + e^{\alpha(\tau+t-2)}, \\ F''(t) &= \delta\beta^2 e^{\beta t} + \alpha e^{\alpha(\tau+t-2)} > 0, \\ F'''(t) &= \delta\beta^3 e^{\beta t} + \alpha^2 e^{\alpha(\tau+t-2)} > 0, \end{aligned}$$

so F'' is positive and increasing on $[1 - \tau, 1]$. Suppose that, on a given range of α, we have certified bounds

$$F(1-\tau) \geq A, \qquad F'(1-\tau) \geq B, \qquad F''(1-\tau) \geq \mu > 0.$$

For $s = t - 1 + \tau \geq 0$, Taylor's theorem gives

$$F(1 - \tau + s) \geq A + Bs + \frac{\mu}{2}s^2.$$

Therefore

$$F(t) \geq A - \frac{\max\{0, -B\}^2}{2\mu}. \tag{49}$$

We have the following equalities

$$F(1-\tau) = \delta\big(e^{\beta(1-\tau)} - 1\big) + \frac{e^{-\alpha} - 1}{\alpha},$$
$$F'(1-\tau) = \delta\beta e^{\beta(1-\tau)} - 2 + e^{-\alpha},$$
$$F''(1-\tau) = \delta\beta^2 e^{\beta(1-\tau)} + \alpha e^{-\alpha}.$$

Using the monotonicity of β and τ with respect to α, together with outward-rounded interval arithmetic, gives the following certificate.

α-interval	A_{low}	B_{low}	μ_{low}	margin
$[\alpha(R_{\text{th}}), 1.980]$	0.018486955	−0.226181399	1.554393652	0.002031013
$[1.980, 2.000]$	0.025606903	−0.207977098	1.548438669	0.011639778
$[2.000, 2.200]$	0.029655186	−0.197371832	1.490417639	0.016586486
$[2.200, 2.350]$	0.071979140	−0.073220859	1.452074242	0.070133059
$[2.350, 3.000]$	0.099677848	0.023156089	1.315131951	0.099677848
$[3.000, +\infty)$	0.103691952	0.213651136	0.602011391	0.103691952

Table 2: Certified lower bounds for Equation (49).

In each interval the last column, namely

$$A_{\text{low}} - \frac{\max\{0, -B_{\text{low}}\}^2}{2\mu_{\text{low}}},$$

is strictly positive. Hence

$$F(t) \geq 0 \qquad \text{for every } t \in [1-\tau, 1],$$

which proves (45). □

We now use Lemma 14 to prove that CF satisfies the robustness requirement. We also check that CF is indeed a contract function, and that its consistency matches the one stated in the theorem.

Proof of Theorem 6. We first verify that CF is a contract function. Recall that

$$\mathrm{CF}(t) = \begin{cases} e^{\alpha(t+\tau)}, & t < -\tau, \\ 1, & -\tau \leq t \leq 0, \\ R\tau e^{\beta t}, & t > 0. \end{cases}$$

The function is clearly nondecreasing except possibly at $t = 0$. By Lemma 13, $R\tau e^{-\beta\tau} > 1$. In particular, $R\tau > 1$, so the jump at $t = 0$ is also nondecreasing. Moreover,

$$\lim_{t\to-\infty} \mathrm{CF}(t) = 0, \qquad \lim_{t\to+\infty} \mathrm{CF}(t) = +\infty, \qquad \mathrm{CF}(0) = 1.$$

Thus CF is indeed a contract function. It remains to prove the robustness condition

$$\mathrm{CF}(t) \leq R\widetilde{\mathrm{CF}}(t), \qquad t \in \mathbb{R}. \tag{50}$$

We define $J(x) := \int_0^x \mathrm{CF}(s)\, ds$, which gives

$$J(x) = \begin{cases} \dfrac{e^{\alpha(x+\tau)} - 1}{\alpha} - \tau, & x \leq -\tau, \\ x, & -\tau \leq x \leq 0, \\ K(e^{\beta x} - 1), & x \geq 0, \end{cases}$$

where $K := \frac{R\tau}{\beta}$. Moreover,

$$\widetilde{\mathrm{CF}}(t) = J(t) - 2J(t-1) + J(t-2). \tag{51}$$

We set

$$\Delta(t) := \widetilde{\mathrm{CF}}(t) - \frac{\mathrm{CF}(t)}{R},$$

so that the robustness criterion (50) reduces to the following property

$$\Delta(t) \geq 0 \qquad \text{for every } t \in \mathbb{R}.$$

The expression of Δ changes whenever one of the arguments t, $t-1$, or $t-2$ crosses one of the breakpoints $-\tau$ and 0 of CF. We therefore verify the inequality successively on the following intervals:

$$(-\infty, -\tau), [-\tau, 0], [0, 1-\tau], [1-\tau, 1], [1, 2-\tau], [2-\tau, 2], [2, +\infty).$$

Case 1: $t < -\tau$**.**

In this case, all three arguments t, $t-1$, and $t-2$ in (51) lie in the exponential part of CF. Hence

$$\widetilde{\mathrm{CF}}(t) = e^{\alpha(t+\tau)} \frac{(1-e^{-\alpha})^2}{\alpha}.$$

Since $\Phi(\alpha) = R$, we have

$$\frac{(1-e^{-\alpha})^2}{\alpha} = \frac{1}{R},$$

and therefore $\widetilde{\mathrm{CF}}(t) = \mathrm{CF}(t)/R$. Thus

$$\Delta(t) = 0, \qquad t < -\tau.$$

Case 2: $-\tau \leq t \leq 0$**.**

Direct substitution into (51) gives

$$\widetilde{\mathrm{CF}}(t) = t + \tau + \frac{1}{\alpha} + \frac{e^{\alpha(\tau+t-1)}(e^{-\alpha} - 2)}{\alpha}.$$

Using $(2 - e^{-\alpha})e^{\alpha(\tau-1)} = 1$, which follows directly from the definition of τ, we obtain

$$\widetilde{\mathrm{CF}}'(t) = 1 - e^{\alpha t} \geq 0.$$

Hence $\widetilde{\mathrm{CF}}$ is nondecreasing on $[-\tau, 0]$. At the left endpoint,

$$\widetilde{\mathrm{CF}}(-\tau) = \frac{(1-e^{-\alpha})^2}{\alpha} = \frac{1}{R}.$$

Since $\mathrm{CF}(t) = 1$ on $[-\tau, 0]$, it follows that $\widetilde{\mathrm{CF}}(t) \geq 1/R = \mathrm{CF}(t)/R$. Therefore

$$\Delta(t) \geq 0, \qquad -\tau \leq t \leq 0.$$

Case 3: $0 < t \leq 1 - \tau$**.**

On this interval, direct substitution in (51) gives

$$\Delta(t) = \delta(e^{\beta t} - 1) + \frac{1 - e^{\alpha t}}{\alpha}.$$

Consequently,

$$\Delta''(t) = \delta\beta^2 e^{\beta t} - \alpha e^{\alpha t}.$$

By Lemma 14, $\delta\beta^2 \leq \alpha$. Since $\beta < \alpha$ and $t \geq 0$, we also have $e^{\beta t} \leq e^{\alpha t}$. Thus

$$\Delta''(t) \leq 0,$$

so Δ is concave on $(0, 1-\tau]$. Moreover, $\Delta(0^+) = 0$, while, by the definition of F and Lemma 14,

$$\Delta(1-\tau) = F(1-\tau) \geq 0$$

By concavity, Δ lies above the chord joining its endpoint values. Since both endpoint values are nonnegative,

$$\Delta(t) \geq 0, \qquad 0 < t \leq 1 - \tau.$$

Case 4: $1 - \tau \leq t \leq 1$**.**

On this interval,

$$\Delta(t) = \delta e^{\beta t} - K - 2(t-1) - \tau + \frac{e^{\alpha(\tau+t-2)} - 1}{\alpha} = F(t).$$

Therefore the condition $\Delta(t) \geq 0$ follows directly from Lemma 14.

Case 5: $1 \leq t \leq 2 - \tau$**.**

Here,

$$\Delta(t) = \delta e^{\beta t} - 2Ke^{\beta(t-1)} + K - \tau + \frac{e^{\alpha(\tau+t-2)} - 1}{\alpha}.$$

Differentiating yields

$$\Delta'(t) = m\beta e^{\beta t} + e^{\alpha(\tau+t-2)} = G(t).$$

By Lemma 14, G is nondecreasing on $[1, 2-\tau]$ and $G(2-\tau) < 0$. Hence, for every $t \leq 2 - \tau$ we have

$$G(t) \leq G(2-\tau) < 0.$$

Thus $\Delta'(t) < 0$ throughout $[1, 2-\tau]$, and so Δ is decreasing on this interval. It will be proved in the next case that

$$\Delta(2-\tau) \geq 0.$$

Since Δ is decreasing, this implies

$$\Delta(t) \geq \Delta(2-\tau) \geq 0, \qquad 1 \leq t \leq 2 - \tau.$$

Case 6: $2 - \tau \leq t \leq 2$**.**

Equation (51) gives

$$\Delta(t) = me^{\beta t} + K + t - 2,$$
$$\Delta'(t) = m\beta e^{\beta t} + 1.$$

Since $m < 0$ by Lemma 14, Δ' is decreasing. Furthermore,

$$\Delta'(2-\tau) = 1 + m\beta e^{\beta(2-\tau)} = G(2-\tau) < 0.$$

Therefore

$$\Delta'(t) < 0, \qquad 2 - \tau \leq t \leq 2,$$

so Δ is decreasing on this interval as well. At the right endpoint,

$$\begin{aligned}\Delta(2) &= me^{2\beta} + K \\ &= K(e^{\beta} - 1)^2 - \tau e^{2\beta} \\ &= 0,\end{aligned}$$

where the last equality follows from the definition of K and β, given as (46). Since Δ is decreasing and $\Delta(2) = 0$, we obtain

$$\Delta(t) \geq 0, \qquad 2 - \tau \leq t \leq 2.$$

In particular, we get the inequality needed for the previous case, i.e.

$$\Delta(2-\tau) \geq 0.$$

Case 7: $t \geq 2$**.**

For $t \geq 2$, all three arguments t, $t-1$, and $t-2$ in (51) belong to the exponential tail. Therefore

$$\begin{aligned}\widetilde{\mathrm{CF}}(t) &= Ke^{\beta t}(1 - e^{-\beta})^2 \\ &= \tau e^{\beta t} \\ &= \frac{\mathrm{CF}(t)}{R}.\end{aligned}$$

Thus

$$\Delta(t) = 0, \qquad t \geq 2.$$

Combining the seven cases, we have proved that $\Delta(t) \geq 0 \qquad$ for every $t \in \mathbb{R}$. Equivalently,

$$\mathrm{CF}(t) \leq R\widetilde{\mathrm{CF}}(t) \qquad \text{for every } t \in \mathbb{R}.$$

Hence CF is R-robust by Proposition 5. It remains to compute its consistency guarantee. At $t = 0$,

$$\begin{aligned}\widetilde{\mathrm{CF}}(0) &= \tau + \frac{1 - 2e^{\alpha(\tau-1)} + e^{\alpha(\tau-2)}}{\alpha} \\ &= \tau + \frac{1 - e^{\alpha(\tau-1)}(2 - e^{-\alpha})}{\alpha} \\ &= \tau,\end{aligned}$$

again by the definition of τ. Proposition 5 therefore gives

$$\mathrm{cons}(\mathrm{CF}) \leq \frac{\mathrm{CF}(0)}{\widetilde{\mathrm{CF}}(0)} = \frac{1}{\tau} = \left(1 - \frac{\ln(2 - e^{-\alpha})}{\alpha}\right)^{-1}.$$

Thus CF is R-robust and

$$\left(1 - \frac{\ln(2 - e^{-\alpha})}{\alpha}\right)^{-1}\text{-consistent.}$$

□

B.5 Laplace Transform motivation for the choice of τ

The purpose of this section is to motivate the choice of the parameter τ used in Theorem 7. We first study general solutions of the DDE (9), independently of the particular construction of Theorem 7. In particular, we establish an exponential growth bound, which guarantees that their Laplace transform is well defined in a right half-plane. We then specialize to a contract function which verifies (7)-(8), derive an explicit expression for the Laplace transform and examine the contribution associated with the characteristic root α. This suggests choosing τ so that the corresponding coefficient is nonnegative, and, in the low-robustness regime, setting it to zero. This analysis is used only to guide the construction; feasibility of the resulting function is not treated.

Lemma 15. *Let $R > 0$, and let B be any contract function satisfying* (9)*. Then*

$$|B(t)| = O(e^{4Rt}) \qquad (t \to +\infty).$$

Proof. Let us define $M(t) = \max\{|B(0^+)|, \sup_{-2 \leq s \leq t} |B(s)|\}$. Since B is bounded on $[-2, 0]$ we get $M(0) < +\infty$. From (9),

$$\begin{aligned}|B'(t)| &\leq R\big(|B(t)| + 2|B(t-1)| + |B(t-2)|\big) \\ &\leq 4RM(t).\end{aligned}$$

Thus for all $t \geq 0$,

$$\begin{aligned}|B(t)| &\leq |B(0^+)| + \int_0^t |B'(s)|\, ds \\ &\leq M(0) + 4R \int_0^t M(s)\, ds.\end{aligned}$$

This gives $M(t) \leq M(0) + 4R \int_0^t M(s)\, ds$ and so by Grönwall's Lemma, $|B(t)| \leq M(t) \leq M(0)e^{4Rt}$. □

We now specialize to contract functions which satisfy (7)-(9). Let us fix $R > R^*$ and let

$$\widehat{B}(z) := \int_0^{+\infty} e^{-zt} B(t)\, dt$$

be the Laplace Transform of B. Note that by Lemma 15, this converges on the half-plane $\Re z > 4R$. Recall that $B(0^+) = R\widetilde{B}(0)$ and for all $t > 0$, $B'(t) = R(B(t)-2B(t-1)+B(t-2))$, multiplying this equality by e^{-zt} and integrating from 0 to $+\infty$ gives us

$$\int_0^{+\infty} e^{-zt}B'(t)\,dt = R\left(\widehat{B}(z) - 2e^{-z}A_1(z) - 2e^{-z}\widehat{B}(z) + e^{-2z}A_2(z) + e^{-2z}\widehat{B}(z)\right),$$

$$A_1(z) := \int_{-1}^{0} e^{-zt}B(t)\,dt = \frac{e^{\tau z} - e^{\alpha(\tau-1)+z}}{\alpha - z} + \frac{e^{\tau z} - 1}{z},$$

$$A_2(z) := \int_{-2}^{0} e^{-zt}B(t)\,dt = \frac{e^{\tau z} - e^{\alpha(\tau-2)+2z}}{\alpha - z} + \frac{e^{\tau z} - 1}{z}.$$

Integration by part gives, for all $z \in \mathbb{C}$ such that $\Re z > 4R$:

$$\begin{aligned}\int_0^{+\infty} e^{-zt}B'(t)\,dt &= \left[e^{-zt}B(t)\right]_0^{+\infty} + z\widehat{B}(z)\\ &= z\widehat{B}(z) - B(0^+)\\ &= z\widehat{B}(z) - R\widetilde{B}(0).\end{aligned}$$

Hence we get that for sufficiently large $\Re z$,

$$\widehat{B}(z) = \frac{R\big(e^{-2z}A_2(z) - 2e^{-z}A_1(z) + \widetilde{B}(0)\big)}{z - R(1-e^{-z})^2}.$$

Hence we define $\mathcal{N}(z) = R(e^{-2z}A_2(z) - 2e^{-z}A_1(z) + \widetilde{B}(0))$ and $\mathcal{H}(z) = z - R(1-e^{-z})^2$. We get that $\mathcal{N}/\mathcal{H}$ is the meromorphic continuation of $\widehat{B}$ on $\mathbb{C}$.

To motivate the choice of the parameters, we only compute the coefficient associated with the characteristic root α. Since α is a simple root of $\mathcal{H}$, which we can check by $\mathcal{H}'(\alpha) \neq 0$, we have around $z = \alpha$:

$$\mathcal{H}(z) = \mathcal{H}'(\alpha)(z-\alpha) + O((z-\alpha)^2).$$

Adding that $\mathcal{N}(z) = \mathcal{N}(\alpha) + O(z-\alpha)$, we get that

$$\frac{\mathcal{N}(z)}{\mathcal{H}(z)} = \frac{\mathcal{N}(\alpha)}{\mathcal{H}'(\alpha)}\frac{1}{z-\alpha} + F(z),$$

where F is a function holomorphic near α. Under the standard inverse-Laplace correspondence, a pole with residue c_α contributes a term $c_\alpha e^{\alpha t}$, hence the coefficient associated with the mode $e^{\alpha t}$ is

$$c_\alpha = \frac{\mathcal{N}(\alpha)}{\mathcal{H}'(\alpha)}.$$

The residue c_α can be interpreted as the coefficient associated with the $e^{\alpha t}$ mode. A negative value of c_α is therefore an indication that the resulting function will fail to remain feasible. This motivates requiring $c_\alpha \geq 0$.

With the original choice $\tau = 1 - \frac{\ln(2-e^{-\alpha})}{\alpha}$, the coefficient c_α becomes negative below the critical value R_c. This motivates modifying τ in the low-robustness regime. We choose τ so that $c_\alpha = 0$; writing $p = e^{\alpha(\tau-1)}$, this condition is equivalent to $\Xi(p,\alpha) = 0$, which yields the definition used in Theorem 7.

B.6 Proof of Theorem 7

We first verify that the parameters α_c and p used in the construction are well defined. Let α^* denote $\alpha(R^*)$, i.e. the minimizer of Φ. Recall that

$$\Phi'(\alpha) = \frac{h(\alpha)}{(1-e^{-\alpha})^3}, \quad h(\alpha) := 1 - (2\alpha+1)e^{-\alpha},$$

and therefore that $h(\alpha^*) = 0$, and $h(\alpha) > 0$ for $\alpha > \alpha^*$. For convenience, we redefine Ξ as a two-variable function, i.e.

$$\Xi(x, \alpha) := \alpha + \log x + 1 + 2e^{-\alpha} - e^{-2\alpha} + x(\alpha e^{-\alpha} + (2 - e^{-\alpha}) \log x - 2(2 - e^{-\alpha}))$$

which is defined for $\alpha \geq \alpha^*$ and $x \in I_\alpha = [e^{-\alpha}, (2 - e^{-\alpha})^{-1}]$.

Proof that α_c is well defined Differentiating Ξ twice with respect to the first variable gives

$$\frac{\partial \Xi(x, \alpha)}{\partial x^2} = -\frac{1}{x^2} + \frac{2 - e^{-\alpha}}{x} = \frac{x(2 - e^{-\alpha}) - 1}{x^2} \leq 0,$$

with strict inequality in the interior of I_α. Hence $\Xi(\cdot, \alpha)$ is strictly concave on I_α. At the left endpoint, we have

$$\Xi(e^{-\alpha}, \alpha) = (1 - e^{-\alpha})h(\alpha) \geq 0,$$

with equality only for $\alpha = \alpha^*$. We now study the value at the right endpoint. Define

$$\psi(\alpha) = \Xi((2 - e^{-\alpha})^{-1}, \alpha).$$

A direct computation gives us

$$\psi(\alpha) = \frac{2\alpha}{2 - e^{-\alpha}} - 2 \log(2 - e^{-\alpha}) - (1 - e^{-\alpha})^2.$$

Notice that the equation $\psi(\alpha) = 0$ is the defining equation of α_c in Theorem 7. We next show that this equation has a unique solution for $\alpha \geq \alpha^*$. Differentiating ψ yields

$$\psi'(\alpha) = \frac{2}{(2 - e^{-\alpha})^2}((1 - e^{-\alpha})^3(2 - e^{-\alpha}) - \alpha e^{-\alpha}).$$

Since $h(\alpha) \geq 0$ for $\alpha \geq \alpha^*$, we get $(2\alpha + 1)e^{-\alpha} \leq 1$ and thus $\alpha e^{-\alpha} \leq \frac{1 - e^{-\alpha}}{2}$. Moreover, $\alpha^* > 1$, and therefore $e^{-\alpha} \leq e^{-1} < 0.4$. Consequently,

$$(1 - e^{-\alpha})^2(2 - e^{-\alpha}) > (0.6)^2 \cdot \frac{8}{5} > \frac{1}{2},$$

and thus $(1 - e^{-\alpha})^3(2 - e^{-\alpha}) > \frac{1 - e^{-\alpha}}{2} \geq \alpha e^{-\alpha}$, which gives

$$\psi'(\alpha) > 0, \quad \forall \alpha \geq \alpha^*.$$

We now consider $\alpha = \alpha^*$, we have the equalities

$$\Xi(e^{-\alpha^*}, \alpha^*) = 0$$
$$\frac{\partial \Xi}{\partial x}(e^{-\alpha^*}, \alpha^*) = 0.$$

Since $\Xi(\cdot, \alpha^*)$ is strictly concave, it follows that $\psi(\alpha^*) = \Xi((2 - e^{-\alpha^*})^{-1}, \alpha^*) < 0$.

On the other hand, as $\alpha \to +\infty$, we get $\psi(\alpha) = \alpha - 2 \log 2 - 1 + o(1)$, which gives us by continuity and monotonicity of ψ that there exists a unique $\alpha_c > \alpha^*$ such that $\psi(\alpha_c) = 0$.

Proof that p is well defined It only remains to show that p is well defined when $R^* \leq R \leq R_c$, or equivalently when $\alpha^* \leq \alpha \leq \alpha_c$. From the previous calculations, we have

$$\Xi(e^{-\alpha}, \alpha) \geq 0, \quad \Xi((2 - e^{-\alpha})^{-1}, \alpha) = \psi(\alpha) \leq 0.$$

Hence, by continuity, $\Xi(\cdot, \alpha)$ has a zero in I_α, which we call p. Strict concavity shows that this zero is unique.

Proof that τ is well defined If $R \geq R_c$, then $\tau > 0$ is equivalent to $2 \leq e^{\alpha} + e^{-\alpha}$, which is true since $\alpha > 0$.

Else, if $R < R_c$ we use that $p \in I_\alpha$, which means that $p \geq e^{-\alpha}$ and that $\log(p) \geq -\alpha$, hence $\tau \geq 0$.

Robustness requirement of B Let us assume, as stated in the theorem, that the obtained function B is indeed a contract function (it is nondecreasing on $\mathbb{R}$). We start by verifying that the construction is R-robust, i.e. that $B(t) \leq R\widetilde{B}(t)$ for all $t \in \mathbb{R}$. We consider the following three cases:

- If $t < -\tau$, then notice that since α is a root of $\Phi(x) = R$, we get $B(t) = R\widetilde{B}(t)$ on the whole interval.
- If $-\tau \leq t \leq 0$, then we use the same argument as for the interval $[-\tau, 0]$ of Theorem 6, replacing the equality $(2 - e^{-\alpha})e^{\alpha(\tau-1)} = 1$ by the inequality $(2 - e^{-\alpha})e^{\alpha(\tau-1)} \leq 1$.
- If $t \geq 0$, then we have by construction
$$B'(t) = R\big(B(t) - 2B(t-1) + B(t-2)\big),$$
which is equivalent to $(B(t) - R\widetilde{B}(t))' = 0$ on the same interval. Hence for all $t > 0$, we have
$$B(t) - R\widetilde{B}(t) = B(0^+) - R\widetilde{B}(0^+).$$
We also have by definition $B(0^+) = R\widetilde{B}(0)$ and by continuity of $\widetilde{B}$ that $B(0^+) = R\widetilde{B}(0^+)$. This gives us that $B(t) = R\widetilde{B}(t)$ for all $t > 0$.

Consistency of B We have in general that
$$\widetilde{B}(0) = \tau + \frac{1 - (2 - e^{-\alpha})e^{\alpha(\tau-1)}}{\alpha}.$$
If $R \geq R_c$, then $e^{\alpha(\tau-1)} = \frac{1}{2-e^{-\alpha}}$, thus
$$\widetilde{B}(0) = \tau = 1 - \frac{\log(2 - e^{-\alpha})}{\alpha}.$$
If $R < R_c$, then $e^{\alpha(\tau-1)} = p$, so
$$\widetilde{B}(0) = 1 + \frac{\log p + 1 - (2 - e^{-\alpha})p}{\alpha}$$
In both cases, we have $B(0)/\widetilde{B}(0) = \chi(R)$, thus the consistency of B is at most $\chi(R)$.

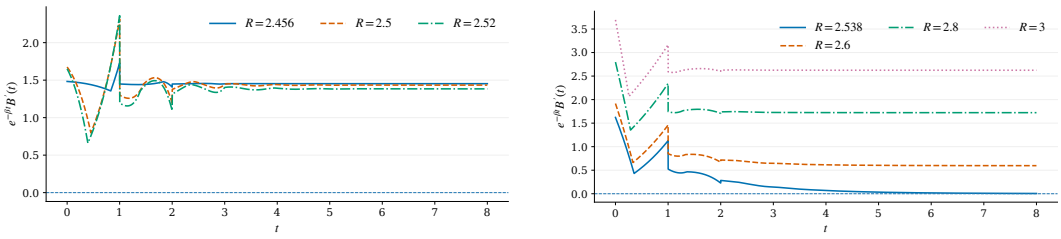
B.7 Numerical evaluation of the DDE solution

We provide numerical evidence that the solution B in Theorem 7 is a contract function. For some given R, we compute B iteratively using the method of steps: starting from $t = 0$, we solve the DDE successively on intervals where the delayed terms $B(t-1)$ and $B(t-2)$ are already known. This yields explicit exponential-polynomial expressions for B on each interval, which can be evaluated together with their derivatives. Our implementation follows this interval-by-interval construction rather than using a standard time-stepping DDE solver.

Since B is already nondecreasing on $(-\infty, 0)$ and the jump at $t = 0$ is upward, one only needs to check that $B'(t) > 0$ for $t > 0$ in order to verify monotonicity of B.

Moreover, the Laplace-transform analysis of Appendix B.5 suggests that the large-t behavior of B is governed by its dominant positive characteristic mode. In particular, we expect
$$B'(t) \sim_{t\to+\infty} \begin{cases} \beta c_\beta e^{\beta t}, & R \leq R_c \\ \alpha c_\alpha e^{\alpha t}, & R > R_c. \end{cases}$$
This motivates plotting the normalized derivative $B'(t)e^{-\gamma t}$, where γ is equal to β or α depending on R, as seen above. If the expected behavior holds, this function should converge to a positive constant that should approach 0 when R is slightly above R_c as c_α tends to 0. The implementation directly computes and plots this normalized derivative.

Figure 3 shows $B'(t)e^{-\gamma t}$ for representative values of R. In all tested cases, the normalized derivative remains positive and stabilizes away from zero, providing numerical evidence that B is nondecreasing and diverges to $+\infty$. Moreover, as expected, the limiting value is close to 0 when R is slightly above $R_c \approx 2.538$.



(a) Normalized derivative $e^{-\beta t}B'(t)$.

(b) Normalized derivative $e^{-\alpha t}B'(t)$.

Figure 3: Numerical evaluation of the DDE solution for several values of R.

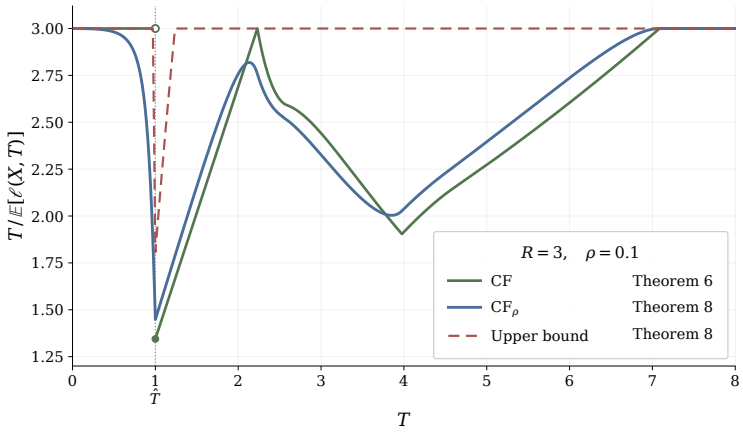


Figure 4: Normalized cost of CF and CF_ρ as a function of the interruption time, together with the upper bound of Theorem 8.

C Supplementary material for Section 4

We first illustrate in Figure 4 the smoothness improvement achieved by CF_ρ, introduced in Theorem 8, in comparison to the schedule CF, from Theorem 6. The figure also depicts the upper bound of Theorem 8 established below.

C.1 Proof of Theorem 8

We first give a high-level description of the proof. We begin with a lemma establishing regularity of the normalized-cost function of CF, including local absolute continuity when $T > \widehat{T}$ and a uniform bound on its logarithmic derivative. We then prove Theorem 8, fixing a prediction $\widehat{T} > 0$. First, we establish continuity of the smoothed normalized cost, showing that CF_ρ is non-brittle. Next, we bound its harmonic representation by an arithmetic average and use the regularity lemma to quantify the consistency loss :

$$\mathrm{nc}(\mathrm{CF}_\rho, \widehat{T}) - \mathrm{nc}(\mathrm{CF}, \widehat{T}) \le L\Psi(\rho).$$

Finally, writing $\eta = \max\{\widehat{T}/T, T/\widehat{T}\} - 1$, we analyze separately the cases $T \ge \widehat{T}$ and $T < \widehat{T}$. Logarithmic variation bounds, together with a bound on the probability of crossing the discontinuity, yield global estimates valid for every $T > 0$. Simplifying these estimates gives the stated guarantee, without an asymptotic expansion in η. We first establish a regularity property of the normalized-cost function of CF.

Lemma 16. *Let $R > R_{\mathrm{th}}$, fix a prediction $\widehat{T} > 0$, and let* CF *be the randomized schedule of Theorem 6. Define its normalized-cost function by*

$$f(T) := \mathrm{nc}(\mathrm{CF}, T), \qquad T > 0.$$

Then

$$1 \le f(T) \le R \qquad \textit{for every } T > 0,$$

and

$$f(T) = R \qquad \textit{for } 0 < T < \widehat{T}.$$

Moreover, f is locally absolutely continuous on $[\widehat{T}, \infty)$. In particular, f is continuous on $(0, \infty) \setminus \{\widehat{T}\}$ and right-continuous at $\widehat{T}$.

Let $\beta > 0$ denote the exponent of the right exponential tail of the contract function defined in Theorem 6. Then the log-slope of f is uniformly bounded:

$$L := \sup_{\substack{T > \widehat{T} \\ f'(T) \textit{ exists}}} |T f'(T)| \le R + \frac{R^2}{\beta} < \infty.$$

Consequently, for every $\widehat{T} \le x \le y$,

$$|f(y) - f(x)| \le L \log \frac{y}{x}.$$

Proof of Lemma 16. By scaling invariance, it suffices to consider $\widehat{T} = 1$. This change of variables preserves local absolute continuity. Moreover, at every corresponding differentiability point,

$$T f'(T) = \frac{T}{\widehat{T}} f_1'\left(\frac{T}{\widehat{T}}\right),$$

so it also preserves the bound on the logarithmic derivative. For the remainder of the proof, we work with prediction 1 and write $f = f_1$.

Let X_{CF} denote the randomized schedule induced by the contract function CF of Theorem 6. By Definition 4, for a uniform random shift $\lambda \in [0, 1)$, contract $i \in \mathbb{Z}$ has length

$$x_i = \mathrm{CF}(i + \lambda) - \mathrm{CF}(i - 1 + \lambda)$$

and completes at time $\mathrm{CF}(i + \lambda)$. Since every contract completed by time T has length at most T, we have $\mathbb{E}[\ell(X_{\mathrm{CF}}, T)] \le T$. Together with the R-robustness established in Theorem 6, this gives

$$1 \le f(T) \le R \qquad \text{for every } T > 0.$$

We next prove that $f(T) = R$ for every $0 < T < 1$. By (10),

$$\mathrm{CF}(t) = e^{\alpha(t+\tau)}, \qquad t < -\tau,$$

where α is the larger of the two positive solutions of $\Phi(x) = R$. On $(-\infty, -\tau)$, CF is strictly increasing and maps this interval onto $(0, 1)$. Thus, for any $T \in (0, 1)$, there is a unique

$$t = \frac{\log T}{\alpha} - \tau < -\tau \qquad \text{such that } \mathrm{CF}(t) = T.$$

Fix a shift $\lambda \in [0, 1)$ and set $j = \lfloor t - \lambda \rfloor$. The contracts completed by time T are exactly those with indices $i \le j$, since CF is strictly increasing on $(-\infty, -\tau)$ and $\mathrm{CF}(u) \ge 1 > T$ for every $u \ge -\tau$. Moreover, for each $i \le j$, both arguments defining x_i belong to $(-\infty, -\tau)$, where CF is exponential. Hence,

$$x_i = (1 - e^{-\alpha}) e^{\alpha(i+\lambda+\tau)}.$$

These lengths are strictly increasing in i. Consequently, contract j is the longest completed contract, and $\ell(X_{\mathrm{CF}}, T) = x_j$.

As in the proof of Proposition 5, the fractional part $\sigma = t - \lambda - \lfloor t - \lambda \rfloor$ is uniform on $[0, 1)$, and $j + \lambda = t - \sigma$. Therefore,

$$\ell(X_{\mathrm{CF}}, T) = T(1 - e^{-\alpha}) e^{-\alpha\sigma},$$

and hence

$$\mathbb{E}[\ell(X_{\mathrm{CF}}, T)] = T(1 - e^{-\alpha}) \int_0^1 e^{-\alpha s}\, ds = T \frac{(1 - e^{-\alpha})^2}{\alpha} = \frac{T}{R},$$

where the last equality follows from $\Phi(\alpha) = R$. Thus $f(T) = R$ for every $0 < T < 1$.

We now establish regularity on $[1, \infty)$. Define

$$m(T) := \mathbb{E}[\ell(X_{\text{CF}}, T)] = \frac{T}{f(T)}.$$

Recall $R\tau > 1$, as established in the proof of Theorem 6. For $t > 0$, we have

$$\text{CF}(t) = R\tau e^{\beta t},$$

where β is the smaller of the two positive solutions of $\Phi(x) = R$. The range of CF over this interval is $(R\tau, \infty)$. Hence, above time 1 completion times can only occur in $(R\tau, \infty)$. To quantify these completions, recall that for every nonnegative measurable function h,

$$\mathbb{E}_\lambda\left[\sum_{i\in\mathbb{Z}} h(i+\lambda)\right] = \sum_{i\in\mathbb{Z}} \int_0^1 h(i+\lambda)\, d\lambda = \int_{\mathbb{R}} h(v)\, dv,$$

by Tonelli's theorem. Applying this identity to CF over $(0, \infty)$ and making the substitution $u = R\tau e^{\beta v}$ shows that, for every nonnegative measurable function q,

$$\mathbb{E}_\lambda\left[\sum_{\substack{i\in\mathbb{Z}\\ i+\lambda>0}} q\big(\text{CF}(i+\lambda)\big)\right] = \int_0^\infty q\left(R\tau e^{\beta v}\right) dv = \int_{R\tau}^\infty q(u)\frac{du}{\beta u}.$$

Thus, the expected density of completion times is $1/(\beta u)$ on $(R\tau, \infty)$ and zero on $(1, R\tau]$. More precisely, for any interval $I \subset (R\tau, \infty)$, the expected number of contracts completing in I is

$$\int_I \frac{du}{\beta u},$$

whereas no contract can complete in $(1, R\tau]$.

For a fixed shift, the longest completed contract is nondecreasing in time and changes only at completion times. At a completion time u, its increase is at most the length of the newly completed contract, which is itself at most u. Consequently, for every $1 \le x \le y$,

$$0 \le m(y) - m(x) \le \mathbb{E}_\lambda\left[\sum_{\substack{i\in\mathbb{Z}\\ x<\text{CF}(i+\lambda)\le y}} \text{CF}(i+\lambda)\right] = \int_{(x,y]} \frac{u\mathbf{1}_{\{u>R\tau\}}}{\beta u} du \le \frac{y-x}{\beta}. \tag{52}$$

In particular, m is nondecreasing and Lipschitz continuous on $[1, \infty)$. Since $m(T) = T/f(T) \ge T/R > 0$, its reciprocal is Lipschitz continuous on every compact subinterval of $[1, \infty)$. It follows that $f(T) = T/m(T)$ is locally absolutely continuous on $[1, \infty)$. In particular, f is continuous on $(1, \infty)$ and right-continuous at 1. Together with $f(T) = R$ on $(0, 1)$, this proves continuity on $(0, \infty) \setminus \{1\}$.

At every $T > 1$ where f is differentiable, the identity $m(T) = T/f(T)$ implies that m is differentiable at T, with

$$m'(T) = \frac{f(T) - Tf'(T)}{f(T)^2}.$$

The preceding Lipschitz and monotonicity bounds imply $0 \le m'(T) \le 1/\beta$ at every such point. Hence

$$Tf'(T) = f(T) - f(T)^2 m'(T),$$

and therefore

$$|Tf'(T)| \le f(T) + f(T)^2 m'(T) \le R + \frac{R^2}{\beta}.$$

Taking the supremum over all differentiability points $T > 1$ gives

$$L \le R + \frac{R^2}{\beta} < \infty.$$

Finally, local absolute continuity implies that, for every $1 \le x \le y$,

$$|f(y) - f(x)| \le \int_x^y |f'(u)|\, du \le L \int_x^y \frac{du}{u} = L \log \frac{y}{x}.$$

Scaling back from prediction 1 to $\widehat{T}$ yields all the claimed properties. □

Proof of Theorem 8. Let $R > R_{\mathrm{th}}$, $\rho \in (0, 1]$ and $\widehat{T} > 0$. We denote by $\mathrm{CF}(\widehat{T})$ the learning-augmented schedule that observes the prediction $\widehat{T}$, and simply by CF when the prediction is clear from the context. Let CF denote the schedule from Theorem 6, and CF_ρ for its smoothed version discussed in Theorem 8.

A key property of the construction is its scaling invariance. Consider two predictions $\widehat{T}, \widehat{T}' > 0$, and their respective induced schedules $\mathrm{CF}(\widehat{T})$ and $\mathrm{CF}(\widehat{T}')$. Hence, for every $T > 0$, the normalized-cost of $\mathrm{CF}(\widehat{T})$ and $\mathrm{CF}(\widehat{T}')$ coincide up to a horizontal scaling by $\widehat{T}/\widehat{T}'$. More precisely,

$$\mathrm{nc}(\mathrm{CF}(\widehat{T}'), T) = \mathrm{nc}\left(\mathrm{CF}(\widehat{T}), T\frac{\widehat{T}}{\widehat{T}'}\right). \tag{53}$$

Recall that for any randomized schedule X, $\mathrm{nc}(X, T) = T/\mathbb{E}[\ell(X, T)]$. Conditional on the smoothing noise ξ, scaling invariance therefore gives expected reward $T/f(T(1+\rho\xi))$. Considering 53, with $\widehat{T}' = \widetilde{T} = \widehat{T}/(1+\rho\xi)$:

$$\frac{T}{\mathbb{E}[\ell(\mathrm{CF}_\rho(\widehat{T}), T) \mid \xi]} = \mathrm{nc}(\mathrm{CF}(\widetilde{T}), T) = \mathrm{nc}\left(\mathrm{CF}(\widehat{T}), T\frac{\widehat{T}}{\widetilde{T}}\right).$$

Then,

$$\begin{aligned}
\mathrm{nc}(\mathrm{CF}_\rho(\widehat{T}), T) &= \frac{T}{\mathbb{E}[\ell(\mathrm{CF}_\rho(\widehat{T}), T)]} \\
&= \frac{T}{\mathbb{E}_\xi\Big[\mathbb{E}[\ell(\mathrm{CF}(\widetilde{T}), T) \mid \xi]\Big]} = \frac{T}{\mathbb{E}_\xi\left[\frac{T}{f(T(1+\rho\xi))}\right]} \\
&= \frac{T}{T\,\mathbb{E}_\xi\left[\frac{1}{f(T(1+\rho\xi))}\right]} = \left(\mathbb{E}_\xi\left[\frac{1}{f(T(1+\rho\xi))}\right]\right)^{-1}.
\end{aligned}$$

For simplicity we write $f(T) = \mathrm{nc}(\mathrm{CF}, T)$, $g(T) = \mathrm{nc}(\mathrm{CF}_\rho, T)$, and $C = f(\widehat{T})$. Hence,

$$g(T) = \left(\mathbb{E}_\xi\left[\frac{1}{f(T(1+\rho\xi))}\right]\right)^{-1} \le \mathbb{E}_\xi\left[f(T(1+\rho\xi))\right] \tag{54}$$

$$= \int_0^\infty f(T(1+\rho s))\frac{2}{(1+s)^3}\, ds := \bar{g}(T), \tag{55}$$

where the inequality is the harmonic–arithmetic mean inequality. Moreover, since $f(T) \le R$ for every $T > 0$ by the R-robustness of CF, we have $g(T) \le \bar{g}(T) \le R$.

By Lemma 16, the normalized-cost function f of CF is continuous except possibly at $T = \widehat{T}$, and satisfies $1 \le f \le R$. We first use these properties to show that the normalized-cost function g of the smoothed schedule CF_ρ is continuous in T. First, we verify that the function

$$s \longmapsto \frac{1}{f(T(1+\rho s))}\frac{2}{(1+s)^3}. \tag{56}$$

satisfies the assumptions of the *Dominated Convergence Theorem* (DCT). For any $T > 0$, consider a sequence $T_n \to T > 0$. The only possible discontinuity of f is at $\widehat{T}$, which corresponds to at most one value of $s \ge 0$ such that $T(1+\rho s) = \widehat{T}$. Outside this set of Lebesgue measure zero,

$$\frac{1}{f(T_n(1+\rho s))}\frac{2}{(1+s)^3} \longrightarrow \frac{1}{f(T(1+\rho s))}\frac{2}{(1+s)^3}.$$

Moreover, since $f \geq 1$,

$$\left| \frac{1}{f(T_n(1+\rho s))} \frac{2}{(1+s)^3} \right| \leq \frac{2}{(1+s)^3}, \tag{57}$$

and the right-hand side is integrable on $[0, \infty)$ since it is the density of ξ. The DCT gives

$$\int_0^\infty \frac{1}{f(T_n(1+\rho s))} \frac{2}{(1+s)^3}\, ds \longrightarrow \int_0^\infty \frac{1}{f(T(1+\rho s))} \frac{2}{(1+s)^3}\, ds.$$

Both integrals are at least $1/R$, since $f \leq R$. Since $x \mapsto 1/x$ is continuous on $(0, \infty)$, Eq. (55) therefore gives $g(T_n) \to g(T)$.

The normalized cost $T \to \mathrm{nc}(\mathrm{CF}_\rho, T)$ is continuous and CF_ρ is non-brittle.

We now bound the arithmetic upper bound $\bar{g}$ for arbitrary prediction errors. Since f has a discontinuity at $\widehat{T}$, we analyze separately the cases $T \geq \widehat{T}$ and $T < \widehat{T}$. We combine the resulting estimates into a global bound to recover the guarantees of Theorem 8.

By Lemma 16, L is finite and, for every $\widehat{T} \leq x \leq y$,

$$|f(y) - f(x)| \leq L \log \frac{y}{x}. \tag{58}$$

In particular, taking $x = \widehat{T}$ and $y = T$, and recalling that $C = f(\widehat{T})$, we obtain

$$f(T) \leq C + L \log \frac{T}{\widehat{T}}, \qquad T \geq \widehat{T}.$$

At perfect prediction, $T = \widehat{T}$, the arithmetic upper bound satisfies

$$\begin{aligned} \bar{g}(\widehat{T}) &= \mathbb{E}_\xi \left[f(\widehat{T}(1+\rho\xi)) \right] \\ &\leq C + L\mathbb{E}_\xi \left[\log(1+\rho\xi) \right] \\ &= C + L\Psi(\rho). \end{aligned} \tag{59}$$

Hence, $\mathrm{cons}(\mathrm{CF}_\rho) = g(\widehat{T}) \leq \bar{g}(\widehat{T}) \leq C + L\Psi(\rho)$. Thus, the gain in smoothness comes at a controlled consistency cost, upper bounded by $L\Psi(\rho)$. Since $\Psi(\rho) \leq 1/2$, the dependence on the smoothing parameter remains controlled. The quantity L provides a uniform global control on the log-slope of f, and therefore need not tightly capture the local consistency loss induced by smoothing. Nevertheless, it admits the upper bound $L \leq R + R^2/\beta$, where β is defined in Theorem 6.

Therefore, an explicit expression for $\Psi(\rho)$ is obtained by integration by parts followed by the substitution $x = (1+s)^{-1}$:

$$\Psi(\rho) = \int_0^\infty \frac{\rho}{(1+\rho s)(1+s)^2}\, ds = \int_0^1 \frac{\rho x}{\rho + (1-\rho)x}\, dx = \begin{cases} \dfrac{\rho(1-\rho+\rho\log\rho)}{(1-\rho)^2}, & 0 < \rho < 1, \\ \dfrac{1}{2}, & \rho = 1. \end{cases}$$

Having bounded the consistency loss induced by smoothing, we now study the dependence of the smoothed normalized cost on the prediction error. We first treat the non-brittle case of CF. Let $T \geq \widehat{T}$, and write $T = \widehat{T}(1+\eta)$ with $\eta \geq 0$. By (55),

$$\bar{g}(T) - \bar{g}(\widehat{T}) = \int_0^\infty \left[f\big(\widehat{T}(1+\eta)(1+\rho s)\big) - f\big(\widehat{T}(1+\rho s)\big) \right] \frac{2}{(1+s)^3}\, ds.$$

Both arguments of f belong to $[\widehat{T}, \infty)$, and their ratio is $1+\eta$. Thus (58) bounds the absolute value of their difference by $L\log(1+\eta)$, uniformly in s. Integration considering the probability density gives

$$|\bar{g}(T) - \bar{g}(\widehat{T})| \leq L \log(1+\eta).$$

Combining this with (59) and $g \le \bar{g}$, we obtain

$$g(T) \le \bar{g}(T) \le C + L\Psi(\rho) + L\log(1+\eta), \qquad T \ge \widehat{T}. \tag{60}$$

This estimate holds for every $\eta \ge 0$.

Now consider the brittle case of CF, when $T < \widehat{T} = T(1+\eta)$ with $\eta > 0$. The smoothing mechanism is efficient when the perturbed prediction satisfies $\widetilde{T} \le T$ if and only if $\xi \ge \eta/\rho$, since

$$\widetilde{T} = \frac{\widehat{T}}{1+\rho\xi} = T\frac{1+\eta}{1+\rho\xi}. \tag{61}$$

Hence by Lemma 16 for any $s < \eta/\rho$, $f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right) = R$ therefore

$$\begin{aligned}
\bar{g}(T) &= \int_0^\infty f(T(1+\rho s))\frac{2}{(1+s)^3}\,ds, \\
&= \int_0^{\eta/\rho} f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right)\frac{2}{(1+s)^3}\,ds + \int_{\eta/\rho}^\infty f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right)\frac{2}{(1+s)^3}\,ds, \\
&= R\int_0^{\eta/\rho}\frac{2}{(1+s)^3}\,ds + \int_{\eta/\rho}^\infty f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right)\frac{2}{(1+s)^3}\,ds,
\end{aligned} \tag{62}$$

To compare $\bar{g}(T)$ with the perfect-prediction case $\bar{g}(\widehat{T})$, we use the same decomposition and obtain

$$\bar{g}(T) - \bar{g}(\widehat{T}) = \int_0^{\eta/\rho}\left[R - f\big(\widehat{T}(1+\rho s)\big)\right]\frac{2}{(1+s)^3}\,ds \tag{63}$$

$$+ \int_{\eta/\rho}^\infty\left[f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right) - f\big(\widehat{T}(1+\rho s)\big)\right]\frac{2}{(1+s)^3}\,ds. \tag{64}$$

The two terms capture two distinct effects. The first is due to the probability mass crossing the discontinuity of f at $\widehat{T}$, while the second measures the variation of f between two arguments in $[\widehat{T},\infty)$. We combine the first term with $\bar{g}(\widehat{T})$ and bound the second term separately.

Combining $\bar{g}(\widehat{T})$ with the first integral in (63), and using $f(\widehat{T}(1+\rho s)) \le C + L\log(1+\rho s)$, gives

$$\begin{aligned}
&\bar{g}(\widehat{T}) + \int_0^{\eta/\rho}\left[R - f\big(\widehat{T}(1+\rho s)\big)\right]\frac{2}{(1+s)^3}\,ds \\
&= R\left[1 - \left(1+\frac{\eta}{\rho}\right)^{-2}\right] + \int_{\eta/\rho}^\infty f\big(\widehat{T}(1+\rho s)\big)\frac{2}{(1+s)^3}\,ds \\
&\le C + (R-C)\left[1 - \left(1+\frac{\eta}{\rho}\right)^{-2}\right] + L\int_{\eta/\rho}^\infty \log(1+\rho s)\frac{2}{(1+s)^3}\,ds \\
&\le C + L\Psi(\rho) + (R-C)\left[1 - \left(1+\frac{\eta}{\rho}\right)^{-2}\right].
\end{aligned} \tag{65}$$

We now turn to (64). For $s \ge \eta/\rho$, we have

$$\widehat{T} \le \widehat{T}\frac{1+\rho s}{1+\eta} \le \widehat{T}(1+\rho s).$$

Both arguments belong to $[\widehat{T},\infty)$ and their ratio is $1+\eta$. Applying (58) gives

$$\left|f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right) - f\big(\widehat{T}(1+\rho s)\big)\right| \le L\log(1+\eta).$$

Since $\int_{\eta/\rho}^{\infty} 2(1+s)^{-3}\,ds = (1+\eta/\rho)^{-2}$, we obtain

$$\int_{\eta/\rho}^{\infty} \left[f\left(\widehat{T}\frac{1+\rho s}{1+\eta}\right) - f\big(\widehat{T}(1+\rho s)\big)\right] \frac{2}{(1+s)^3}\,ds \\ \leq L\log(1+\eta)\left(1+\frac{\eta}{\rho}\right)^{-2}. \tag{66}$$

Combining (63)–(64) with (65) and (66), and using $g \leq \bar{g}$, yields

$$g(T) \leq C + L\Psi(\rho) + L\log(1+\eta)\left(1+\frac{\eta}{\rho}\right)^{-2} + (R-C)\left[1-\left(1+\frac{\eta}{\rho}\right)^{-2}\right], \quad T < \widehat{T}. \tag{67}$$

This estimate holds for every $\eta > 0$.

We can now combine the two cases to obtain a global bound valid for any prediction error. From Eqs. (60) and (67), we obtain

$$\mathrm{nc}(\mathrm{CF}_\rho, T) \leq C + L\Psi(\rho) \\ + \begin{cases} L\log(1+\eta), & T \geq \widehat{T}, \\ (R-C)\left[1-\left(1+\frac{\eta}{\rho}\right)^{-2}\right] + L\log(1+\eta)\left(1+\frac{\eta}{\rho}\right)^{-2}, & T < \widehat{T}. \end{cases} \tag{68}$$

This refined bound can be simplified using, for every $x \geq 0$,

$$\log(1+x) \leq x, \qquad 1-(1+x)^{-2} = \int_0^x \frac{2}{(1+s)^3}\,ds \leq 2x, \qquad (1+x)^{-2} \leq 1.$$

Applying these inequalities to (68) yields the following simpler global bound:

$$\mathrm{nc}(\mathrm{CF}_\rho, T) \leq C + L\Psi(\rho) + \begin{cases} L\eta, & T \geq \widehat{T}, \\ \left(L + \frac{2(R-C)}{\rho}\right)\eta, & T < \widehat{T}. \end{cases} \tag{69}$$

This holds for any prediction error η. The constants C and L depend only on R. Hence, Eq. (69) establishes the claimed smoothness guarantee and completes the proof. □

C.2 Proof of Theorem 9

We prove Theorem 9 using the global upper bound established in the proof of Theorem 8. Since Eq. (68) holds for any fixed pair $(\widehat{T}, T)$, taking expectations with respect to their joint distribution P yields a population upper bound. This upper bound depends on the smoothing parameter ρ. We then characterize the minimizer of its empirical counterpart over n observations, denoted by ρ_{n+1}, and uniformly control the resulting estimation error over the smoothing parameter, which yields the statistical guarantee.

Throughout the proof, R is fixed within the range for which the construction CF is valid, $C = \mathrm{cons}(\mathrm{CF})$, $R > 2.65 > C$, and $L > 0$ is the finite logarithmic derivative bound introduced in the smoothness analysis of Section 4.1. All these quantities are fixed independently of the n observations used in the learning procedure of Section 4.2.

Statistical setting

We consider independent instances of the same fixed oracle: the pairs $(\widehat{T}_i, T_i)$, $i \geq 1$, are independent and identically distributed, take values in $(0,\infty)^2$, and have common joint distribution P.

This does not assume that $\widehat{T}_i$ and T_i are independent within an instance. For example, a fixed deterministic oracle applied to independent identically distributed instances satisfies this model. The dependence on the fixed oracle is suppressed in the notation $\mathcal{L}_P$ and in the population expectations. Independence across instances is a statistical assumption in addition to the oracle using the same prediction mechanism.

Let $\mathcal{H}_n = ((\widehat{T}_i, T_i))_{i=1}^n$ denote the observed history. We select a smoothness parameter ρ_{n+1}, measurable with respect to $\mathcal{H}_n$. Although the next prediction $\widehat{T}_{n+1}$ is available before execution, its value is not used to select ρ_{n+1}. Hence, ρ_{n+1} captures the historical accuracy of the oracle globally, independently of the particular value of the next prediction.

The estimator ρ_{n+1} should adapt to the historical prediction error of the oracle. Due to the asymmetry of the problem, with brittleness occurring when $T < \widehat{T}$, the estimation of ρ_{n+1} focuses on overprediction errors, namely when $T < \widehat{T}$. The relevant statistic is the unconditional mean overprediction error, $\mu = \mathbb{E}_P[(\widehat{T}/T - 1)_+]$. When $T \geq \widehat{T}$, the potential benefit of the smoothing mechanism is limited.

For a fixed $\rho \in (0, 1]$, the expected normalized cost is

$$\mathcal{L}_P(\rho) := \mathbb{E}_{(\widehat{T},T)\sim P}\Big[\mathrm{nc}(\mathbf{CF}_\rho(\widehat{T}), T)\Big]. \tag{70}$$

In particular, independence of the next pair from $\mathcal{H}_n$ gives

$$\mathbb{E}_{(\widehat{T}_{n+1},T_{n+1})\sim P}\Big[\mathrm{nc}(\mathbf{CF}_{\rho_{n+1}}(\widehat{T}_{n+1}), T_{n+1}) \mid \mathcal{H}_n\Big] = \mathcal{L}_P(\rho_{n+1}). \tag{71}$$

This averages over the next instance, including its prediction; it is not a guarantee conditional on a particular value of $\widehat{T}_{n+1}$.

For any $\rho \in (0, 1]$, we derive from (68) an upper bound for $\mathcal{L}_P(\rho)$. (68) holds for every $T, \widehat{T} > 0$ and $\rho \in (0, 1]$, with $\eta_- = (\widehat{T}/T - 1)_+$,

$$\mathrm{nc}(\mathbf{CF}_\rho(\widehat{T}), T) \leq \min\left\{R,\, C + L\Psi(\rho) + L|\log(T/\widehat{T})| + \frac{2(R-C)\eta_-}{\rho}\right\}. \tag{72}$$

Indeed, for $\eta = \max\{\widehat{T}/T, T/\widehat{T}\} - 1$, $\log(1+\eta) = |\log(T/\widehat{T})|$, $(1+\eta/\rho)^{-2} \leq 1$, and $1 - (1+\eta/\rho)^{-2} \leq 2\eta/\rho$. The bound by R follows from the same theorem. Now let $\eta_- = (\widehat{T}/T - 1)_+$, $\mu = \mathbb{E}[\eta_-]$, and $m_P = \mathbb{E}[|\log(T/\widehat{T})|]$. Taking expectations in (72) gives

$$\mathcal{L}_P(\rho) \leq \min\{R, \mathcal{U}_P(\rho)\}, \qquad \mathcal{U}_P(\rho) = C + Lm_P + J_P(\rho), \quad J_P(\rho) = L\Psi(\rho) + \frac{2(R-C)\mu}{\rho}. \tag{73}$$

The quantity $\mathcal{U}_P$ is an upper bound on the expected normalized cost when $(\widehat{T}, T)$ is drawn from the distribution P. We choose the smoothing parameter to obtain the tightest such guarantee based on the observed history $\mathcal{H}_n$.

Let $\varepsilon \in (0, 1]$, we choose the value of $\rho \in [\varepsilon, 1]$ that minimizes $J(\rho)$. Since this objective depends on the joint distribution P, its minimizer, denoted by ρ_P, also depends on P. We thus associate with each distribution P a smoothing parameter ρ_P. Since P is unknown, we use the observed history $\mathcal{H}_n$ to construct the empirical distribution P_n and define the estimator

$$\rho_{n+1} := \rho_{P_n}.$$

In the following, we characterize this estimator through the first-order optimality condition for the empirical objective J. This amounts to finding ρ such that

$$G(\rho) = \rho^2\Psi'(\rho) = \frac{2(R-C)}{L}\mu.$$

We next show that G is invertible on its domain, which allows us to characterize the solution explicitly.

Lemma 17. *The function $G(\rho) = \rho^2\Psi'(\rho)$ is continuously differentiable and strictly increasing on $(0, 1)$, with*

$$G(0^+) = 0, \qquad G(1^-) = \frac{1}{3}, \qquad G(\rho) \sim \rho^2 \quad (\rho \to 0).$$

It extends continuously to $(0, 1]$ with $G(1) = 1/3$ and has a continuous inverse from $(0, 1/3]$ to $(0, 1]$, with $G^{-1}(1/3) = 1$.

Proof. Differentiation under the expectation is justified by the dominated convergence theorem. Indeed,

$$0 \le \frac{\xi}{1+\rho\xi} \le \xi,$$

with $\mathbb{E}[\xi] = 1$, and the corresponding difference quotients are locally dominated by an integrable multiple of ξ.

Hence,

$$\Psi'(\rho) = \mathbb{E}\left[\frac{\xi}{1+\rho\xi}\right], \qquad G(\rho) = \mathbb{E}\left[\frac{\rho^2\xi}{1+\rho\xi}\right]. \tag{74}$$

The derivative of the integrand in the latter expectation is

$$\frac{\rho\xi(2+\rho\xi)}{(1+\rho\xi)^2} = 1 - \frac{1}{(1+\rho\xi)^2},$$

which lies in $[0, 1]$. This provides an integrable bound for a second differentiation under the expectation and for continuity of the derivative. Since $\xi > 0$ almost surely,

$$G'(\rho) = \mathbb{E}\big[1 - (1+\rho\xi)^{-2}\big] > 0.$$

At zero, dominated convergence in (74), with dominating variable ξ, gives $\Psi'(\rho) \to \mathbb{E}[\xi] = 1$. Thus $G(\rho)/\rho^2 \to 1$. At one, the integrand defining G is bounded by ξ for $0 < \rho < 1$, so dominated convergence again gives

$$G(1^-) = \mathbb{E}\left[\frac{\xi}{1+\xi}\right] = \int_0^\infty \frac{2s}{(1+s)^4}\, ds = 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{3}.$$

Continuity, strict increase and these endpoint limits prove the stated range and the existence of the continuous inverse. □

We have established that we would ideally choose ρ_{n+1} such that

$$G(\rho_{n+1}) = \frac{2(R-C)}{L}\mu_n.$$

However, the right-hand side may lie outside the range of G. First, if

$$\frac{2(R-C)}{L}\mu_n \ge \frac{1}{3},$$

we set $\rho_{n+1} = 1$, corresponding to the maximum amount of smoothing allowed in our setting.

The critical case occurs when

$$\frac{2(R-C)}{L}\mu_n = 0,$$

or equivalently when $\mu_n = 0$, meaning that the oracle makes no overprediction in the observed history. Since we require $\rho > 0$ to avoid the brittleness of the original algorithm, we cannot set $\rho_{n+1} = 0$. To overcome this issue, we introduce a lower cutoff $\varepsilon \in (0, 1]$ and set $\rho_{n+1} = \varepsilon$ whenever

$$\frac{2(R-C)}{L}\mu_n \le G(\varepsilon).$$

This cutoff allows the amount of smoothing to remain arbitrarily small when the empirical prediction error is negligible, while still providing protection when the empirical error is not representative of the expected prediction error, as may occur when only a small number of observations is available. Therefore, the cutoff should decrease with the sample size n.

We set

$$\varepsilon = n^{-1/4}.$$

This choice balances the convergence rate of the final statistical guarantee with the restriction of the estimator to the interval $[\varepsilon, 1]$. Let $1 \geq \varepsilon > 0$, we define the saturated inverse explicitly by

$$Q_\varepsilon(z) = \begin{cases} \varepsilon, & z \leq G(\varepsilon), \\ G^{-1}(z), & G(\varepsilon) < z < 1/3, \\ 1, & z \geq 1/3. \end{cases} \tag{75}$$

The function Q_ε is continuous, including at both thresholds. The following lemma characterizes the optimal smoothing parameter.

Lemma 18. *For any $u \geq 0$, consider*

$$J_u(\rho) := L\Psi(\rho) + \frac{2(R-C)u}{\rho}.$$

Its unique minimizer over $[\varepsilon, 1]$ is

$$\rho_u = Q_\varepsilon\left(\frac{2(R-C)u}{L}\right).$$

Proof. The function is continuous on the compact interval $[\varepsilon, 1]$, so it attains a minimum. Set $z = 2(R-C)u/L$. Direct differentiation gives

$$J_u'(\rho) = \frac{L}{\rho^2}\big(G(\rho) - z\big). \tag{76}$$

Because $L/\rho^2 > 0$, the sign is exactly that of $G(\rho) - z$. If $z \leq G(\varepsilon)$, the derivative is strictly positive on $(\varepsilon, 1)$, so the unique minimum is ε; this includes $z = G(\varepsilon)$. If $z \geq G(1)$, it is strictly negative on $(\varepsilon, 1)$, so the unique minimum is 1, including $z = G(1)$. Otherwise, the strictly increasing function G crosses z once, at $\rho = G^{-1}(z)$.The derivative is negative for $\rho < G^{-1}(z)$ and positive for $\rho > G^{-1}(z)$. The function therefore strictly decreases and then strictly increases, proving that $G^{-1}(z)$ is the unique minimizer in this case. These cases give (75). □

We first define the estimator based on the observed history $\mathcal{H}_n$. We then derive an upper bound on its expected normalized cost $\mathcal{L}_P(\rho_{n+1})$ from the smoothness guarantee of Theorem 8. Assume the independent-instance setting for the observed data, with $n \geq 1$ and $\eta < M$ almost surely for some $M > 0$. Since $\eta_- \leq \eta < M$ almost surely and $|\log(T/\widehat{T})| = \log(1+\eta)$, we have $m_P = \mathbb{E}|\log(T/\widehat{T})| \leq \log(1+M) < \infty$..

Define

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{T}_i}{T_i} - 1\right)_+, \qquad \widehat{m}_n = \frac{1}{n}\sum_{i=1}^{n}\left|\log\left(\frac{T_i}{\widehat{T}_i}\right)\right|.$$

We use these two empirical quantities to construct the empirical counterpart of $\mathcal{U}_P$ defined in Eq. (73). For any ρ, this gives

$$\widehat{\mathcal{U}}_n(\rho) = C + L\widehat{m}_n + J_{P_n}(\rho), \qquad J_{P_n}(\rho) = L\Psi(\rho) + \frac{2(R-C)\mu_n}{\rho},$$

where the first two terms of $\widehat{\mathcal{U}}_n(\rho)$ do not depend on ρ.

Applying Lemma 18 with $u = \mu_n$ therefore shows that the unique minimizer of $\widehat{\mathcal{U}}_n$ over $[\varepsilon, 1]$ is

$$\rho_{n+1} = Q_\varepsilon\left(\frac{2(R-C)}{L}\mu_n\right). \tag{77}$$

Since μ_n is measurable with respect to the observed history $\mathcal{H}_n$ and Q_ε is continuous, the estimator ρ_{n+1} is also measurable with respect to $\mathcal{H}_n$. Likewise, the population minimizer over $[\varepsilon, 1]$ is

$$\rho_\varepsilon^\star = Q_\varepsilon\left(\frac{2(R-C)}{L}\mu\right).$$

We control the difference between $\mathcal{U}_P(\rho_{n+1})$ and $\mathcal{U}_P(\rho_\varepsilon^\star)$. Applying Hoeffding's inequality to the empirical mean μ_n yields

$$\mathbb{P}(|\mu_n - \mu| > x) \le 2\exp\left(-\frac{2nx^2}{M^2}\right), \qquad x > 0.$$

Consequently, for any $\delta \in (0,1)$, the event

$$\mathcal{E}_n = \{|\mu_n - \mu| \le \sigma_n\}, \qquad \text{where } \sigma_n = M\sqrt{\frac{\log(2/\delta)}{2n}}, \tag{78}$$

has probability at least $1-\delta$.

Recall that P_n denotes the empirical distribution of the observed sample $\mathcal{H}_n$. On $\mathcal{E}_n$, for every $\rho \in [\varepsilon, 1]$, we have

$$|J_{P_n}(\rho) - J_P(\rho)| = \frac{2(R-C)}{\rho}\,|\mu_n - \mu| \le \frac{2(R-C)}{\varepsilon}\sigma_n =: d_n. \tag{79}$$

This single event controls all ρ simultaneously, since the only random deviation is the scalar error in Eq. (78). Hence, no union bound over the parameter space is needed, and the inequality can in particular be evaluated at ρ_{n+1}.

Using Eq. (79) at ρ_{n+1} and $\rho_\varepsilon^\star$, together with the empirical optimality of ρ_{n+1}, gives

$$\begin{aligned} J_P(\rho_{n+1}) &\le J_{P_n}(\rho_{n+1}) + d_n \\ &\le J_{P_n}(\rho_\varepsilon^\star) + d_n \\ &\le J_P(\rho_\varepsilon^\star) + 2d_n. \end{aligned} \tag{80}$$

Adding $C + Lm_P$ to both sides and substituting the value of d_n yields

$$\mathcal{U}_P(\rho_{n+1}) \le \inf_{\rho\in[\varepsilon,1]} \mathcal{U}_P(\rho) + \frac{4(R-C)M}{\varepsilon}\sqrt{\frac{\log(2/\delta)}{2n}}. \tag{81}$$

The deterministic bound in Eq. (73) holds for every $\rho \in [\varepsilon, 1]$, and hence also for the realized ρ_{n+1}. Combining it with Eq. (81), on $\mathcal{E}_n$ we obtain

$$\mathcal{L}_P(\rho_{n+1}) \le \min\left\{R,\ \inf_{\rho\in[\varepsilon,1]} \mathcal{U}_P(\rho) + \frac{4(R-C)M}{\varepsilon}\sqrt{\frac{\log(2/\delta)}{2n}}\right\} \tag{82}$$

$$= \min\left\{R,\ \inf_{\rho\in[\varepsilon,1]} \mathcal{U}_P(\rho) + 4(R-C)M\sqrt{\frac{\log(2/\delta)}{2\sqrt{n}}}\right\}, \tag{83}$$

where the last equality follows from $\varepsilon = n^{-1/4}$. The resulting algorithm remains R-robust, and the upper bound by R holds even outside $\mathcal{E}_n$. For $n = 1$, the admissible interval is $[\varepsilon, 1] = \{1\}$, so the empirical and population minimizers both equal 1.

Finally, we characterize the estimator in the small-error regime before saturation. By Lemma 17,

$$G(\rho) \sim \rho^2 \qquad (\rho \to 0).$$

Since G is strictly increasing with $G(0^+) = 0$, this implies

$$G^{-1}(u) \sim \sqrt{u} \qquad (u \to 0).$$

Therefore, in the unsaturated regime,

$$G^{-1}\left(\frac{2(R-C)}{L}\mu_n\right) \sim \sqrt{\frac{2(R-C)}{L}\mu_n} \qquad (\mu_n \to 0).$$

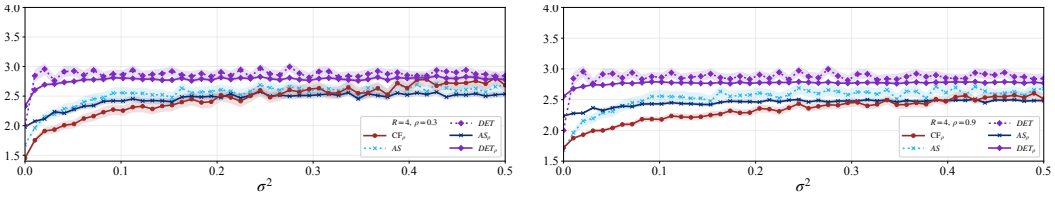


Figure 5: Empirical acceleration ratio as a function of σ^2 for $R = 4$.

D EXPERIMENTAL EVALUATION

We report supplemental experiments for Section 5. We first describe the schedule denoted by AS, which is inspired by the randomized geometric construction of Angelopoulos & Simon (2025). Their work studies the closely related online bidding problem, where an algorithm submits a sequence of bids until the first one that exceeds an unknown target. In this setting, there is no interruption time: the algorithm pays the total length of all contracts it has started, including those that fail to reach the target and the first one that succeeds.

We adapt this construction to our contract-scheduling setting. For $\alpha > 1$ and $\beta \in (0, 1]$, define the contract function

$$B_{\alpha,\beta}(t) = \widehat{T}\alpha^{\lfloor t \rfloor + \beta\{t\} - \beta},$$

where $\{t\} = t - \lfloor t \rfloor$. the scalar $\alpha^{-\beta}$ is chosen such that the left limit of the function at time 1 is equal to the prediction $\widehat{T}$:

$$B_{\alpha,\beta}(1^-) = \widehat{T}.$$

Given a prediction $\widehat{T}$, the schedule induced by $B_{\alpha,\beta}$ is the schedule $(X_i)_i$, where for $\lambda \sim \mathrm{Unif}[0, 1]$, we have

$$x_i = B_{\alpha,\beta}(i + \lambda) - B_{\alpha,\beta}(i - 1 + \lambda) = \frac{\alpha - 1}{\alpha} B_{\alpha,\beta}(i + \lambda), \qquad i \in \mathbb{Z}.$$

Equivalently, since $\{i + \lambda\} = \lambda$, this can be written as

$$x_i = \widehat{T}\frac{\alpha - 1}{\alpha}\alpha^{\,i+\beta\lambda-\beta}, \qquad i \in \mathbb{Z}.$$

Thus, AS is a geometric contract schedule with base α. Following an analysis along the lines of Angelopoulos & Simon (2025), for every α and β, we obtain the following expressions for the consistency and the robustness of the schedule.

$$\mathrm{rob}(B_{\alpha,\beta}) = \frac{\alpha^2 \beta \log(\alpha)}{(\alpha - 1)(\alpha^\beta - 1)}. \qquad \mathrm{cons}(B_{\alpha,\beta}) = \frac{\alpha^{\beta+1} \beta \log(\alpha)}{(\alpha - 1)(\alpha^\beta - 1)}.$$

For any given R, AS is defined as the schedule for which α, β are chosen so that $\mathrm{rob}(X) \le R$ and $\mathrm{cons}(X)$ is minimized.

We also consider the optimal deterministic schedule DET from Angelopoulos & Kamali (2023), which optimizes the deterministic consistency-robustness tradeoff. Note that this schedule can only be evaluated for $R \ge 4$, since no deterministic schedule can have robustness smaller than 4. DET is a brittle schedule; we define its smoothed variant by DET_ρ, analogously to CF_ρ and AS_ρ.

The experimental results, shown in Figure 5, confirm the qualitative behavior previously observed for $R \in \{2.7, 3\}$. For small prediction errors, CF_ρ achieves the lowest empirical acceleration ratio. As the prediction error increases, its performance gradually degrades, while AS and AS_ρ remain stable, reflecting the intrinsic smoothness of the randomized geometric schedule. In addition, CF_ρ performs significantly better than DET and DET_ρ, whose performances are very similar. The similarity between DET and DET_ρ can be explained by the periodic structure of the doubling schedule on the logarithmic scale.