Information-Constrained Prophet Inequalities via Kolmogorov-Wiener Prediction Theory

Ziran Liu *

March 15, 2026

Abstract

We study finite-horizon sequential selection and pricing under compressed past information. Our main idea is that the relevant online state is not the raw history itself but a canonical predictive state: a structured summary of the past that preserves the continuation values driving optimal decisions.

We first prove a general compression-to-decision principle: for threshold-based selection policies, the loss relative to the full-information online optimum is bounded by the cumulative distortion of the continuation values induced by state compression. This yields a decomposition of the prophet gap into a classical future-information term and a new past-compression penalty.

We then instantiate the predictive layer in a Gaussian observation model using Hardy-space spectral factorization. Canonical predictive states are defined by orthogonal projection onto past subspaces, and their finite-dimensional approximations are analyzed through the outer spectral factor. For general outer-factor classes, innovation-tail estimates yield explicit prophet and pricing guarantees. For rational causally invertible spectral factors, we prove exact finite-dimensional realizability: the predictive state admits a finite linear state-space realization, and the corresponding compressed-state policy coincides with the full-information online optimum.

These results provide a bridge between Dym–McKean style prediction theory and modern prophet-inequality / posted-pricing frameworks, and suggest a broader theory in which online approximation guarantees are controlled by the geometry of predictive-state compression.

Keywords. prophet inequality, posted pricing, optimal stopping, predictive state, Hardy space, spectral factorization, finite-past prediction, Gaussian process, state compression, online decision-making.

MSC 2020. Primary 60G40, 60G25, 62M20; Secondary 47B35, 30H10, 90B50, 91B26.

Contents

*Affiliation and email.

1 Introduction

Classical prediction theory studies how future-relevant quantities can be recovered from the past of a stochastic process. In the scalar stationary Gaussian setting, this subject is deeply shaped by the Kolmogorov–Wiener program and by the function-theoretic developments surrounding Dym–McKean, where Hardy-space factorization, past–future geometry, finite-past prediction, and interpolation are central themes [9, 10]. In a recent revisit, Bingham emphasized once again that the two basic classical problems are prediction from the whole past and prediction from a finite section of the past, and advocated short proofs that reveal the probabilistic meaning of the theory [2]. The multivariate extension developed by Arov and Dym places finite-past prediction in a vector-valued de Branges/Hankel framework and points toward structured finite-dimensional realizations [1].

A seemingly distant literature, originating in optimal stopping and now central in operations research, theoretical computer science, and mechanism design, studies *prophet inequalities.* Here an online decision maker is compared against an omniscient offline benchmark [11]. Over the past decade this area has grown rapidly, in part because of its close connection to posted-price mechanisms. In particular, prophet inequalities and posted pricing are now understood as two faces of a common sequential decision problem: thresholds on the stopping side become prices on the mechanism-design side, and conversely price-based mechanisms can often be analyzed through prophet-style guarantees [7, 6, 8].

This paper proposes a bridge between these two traditions. Our basic claim is that past information should not be viewed as raw history but as a compressible Hilbert-space object. The right online state is therefore not an arbitrary summary of the past, but a *canonical predictive state*: a structured compression of the past that preserves the continuation values relevant for decision-making.

1.1 The question and our contributions.

Once this predictive layer is isolated, one can ask a new class of questions: **how does finite-past or finite-state compression degrade a stopping policy, a prophet guarantee, or a posted-price mechanism?**

Our main contribution is to show that this question can be answered in a clean two-layer theory. The lower layer comes from prediction theory: Dym–McKean/Hardy-space methods define canonical predictive states and provide structured finite-dimensional approximations. The upper layer comes from sequential decision-making: continuation values, thresholds, and prices are shown to be stable under perturbations of these predictive states. The resulting framework splits online–offline loss into two distinct components:

$$\text{offline–online gap} = \text{future-information gap} + \text{past-compression penalty}.$$

The first term is the classical prophet loss. The second is new, and is controlled by prediction-theoretic approximation quantities.

Precisely, the paper makes the following contributions.

1. **Predictive-state formulation of sequential selection.** We formulate finite-horizon online selection with pre-decision information as a state-compression problem. The central object is the continuation value $C_t^\star$, and the key distortion variable is

$$\Delta_t = \widehat{C}_t - C_t^\star.$$

   This shifts the focus from reconstructing raw histories to preserving the decision-relevant geometry of the past.

2. **Compression-to-decision decomposition.** We prove a stability theorem showing that the loss of a compressed-state threshold policy is bounded by the sum of continuation distortions:

$$V_1^\star - V_1^{\widehat{\pi}} \le \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

As a corollary,

$$P_n - V_1^{\widehat{\pi}} = (P_n - V_1^\star) + (V_1^\star - V_1^{\widehat{\pi}}),$$

which gives a clean decomposition of prophet loss into a classical future-information term and a new past-compression term.

3. **Canonical predictive states from Hardy-space compression.** In a discrete-time Gaussian observation model, we define the canonical predictive state $S_t^\star$ as the orthogonal projection of a future-relevant feature U_t onto the closed past subspace. Under a decision-sufficiency assumption, the continuation value factors as

$$C_t^\star = \Theta_t(S_t^\star).$$

   This is the formal point at which Dym–McKean / Hardy-space prediction enters sequential decision-making.

4. **Quantitative approximation and exact finite-state realizations.** We derive explicit approximation bounds by expanding $S_t^\star$ in past innovations determined by the outer spectral factor. This yields decision-side rates from Hardy-side decay estimates. In the rational spectral-factor case, we prove exact finite-dimensional realizability: the canonical predictive state admits a finite linear state-space representation, and the corresponding compressed-state policy coincides exactly with the full-information online optimum.

5. **Prophet and posted-pricing transfers.** We show that any prophet inequality for the full-information benchmark transfers immediately to compressed predictive-state policies, with an additive penalty controlled by predictive-state approximation error. We then formulate a pricing analogue for adaptive single-item posted pricing, proving that predictive-state compression also induces additive revenue loss bounds. This identifies predictive-state compression as a common information-theoretic layer beneath both stopping rules and price-based mechanisms.

6. **A program for price-based online optimization.** Beyond the single-item setting, our framework suggests a general research direction: whenever a posted-price or prophet framework is driven by benchmark thresholds or prices that are measurable functions of a canonical predictive state, one should expect state-compression errors to propagate into approximation losses in a controlled way. This opens a path from classical prediction theory to modern online allocation and mechanism design.

1.2 Main results.

Our results are organized around four theorem packages.

(i) Compression-to-decision decomposition. For a compressed-state threshold policy $\widehat{\pi}$ driven by continuation surrogates $\widehat{C}_t$, let

$$\Delta_t := \widehat{C}_t - C_t^\star$$

denote the continuation-value distortion relative to the full-information online benchmark. We prove that

$$V_1^\star - V_1^{\widehat{\pi}} \le \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

As a consequence,

$$P_n - V_1^{\widehat{\pi}} = (P_n - V_1^\star) + (V_1^\star - V_1^{\widehat{\pi}}),$$

so the prophet loss decomposes into a classical future-information term and a new past-compression penalty.

(ii) Predictive-state transfer principle. If the continuation value factors through a canonical predictive state,

$$C_t^\star = \Theta_t(S_t^\star),$$

and Θ_t is Lipschitz, then any predictive-state approximation $\widehat{S}_t$ induces a decision bound of the form

$$P_n - V_1^{\widehat{\pi}} \le (P_n - V_1^\star) + \sum_{t=1}^{n} L_t \mathbb{E} \left\| \widehat{S}_t - S_t^\star \right\|_2 .$$

(iii) Quantitative Hardy-side approximation. In the Gaussian / Hardy-space observation model, the canonical predictive state admits an explicit innovation expansion

$$S_t^\star = \sum_{\ell \ge 1} A_{t,\ell} \varepsilon_{t-\ell}.$$

Truncation at lag L yields the exact approximation formula

$$\left\| S_t^\star - \widehat{S}_t^{(L)} \right\|_{L^2}^2 = \sum_{\ell > L} \|A_{t,\ell}\|_F^2 .$$

This transfers directly into prophet and pricing guarantees.

(iv) Rational exactness. When the outer spectral factor is rational and causally invertible, the canonical predictive state admits an exact finite-dimensional linear realization. In that case the compressed-state policy coincides with the full-information online optimum, and the additional compression penalty vanishes identically.

Methodological viewpoint. A conceptual point deserves emphasis. We do not import Hardy spaces into sequential decision-making as external decoration. Rather, we use them where they are intrinsically strongest: to describe the geometry of the past, to construct canonical predictive states, and to quantify finite-state or finite-past approximation. The decision-theoretic layer is then built on top of this predictive geometry. In this sense, the paper is neither a repackaging of classical prediction theory nor a purely algorithmic prophet-inequality paper; it is a coupling of the two.

Related literature. Our work sits at the intersection of three literatures.

First, it builds on classical prediction theory for stationary processes, especially the line running from Kolmogorov–Wiener prediction through Dym–McKean and its function-theoretic descendants. In the perspective emphasized by Bingham, the two basic classical problems are prediction from the whole past and prediction from a finite section of the past [2]. Our use of canonical predictive states is directly inspired by this viewpoint: the past is not treated as raw data, but as a Hilbert-space object that can be projected, compressed, and approximated. On the multivariate side, Arov and Dym developed a broad finite-past framework in which the relevant projections are expressed in vector-valued de Branges / Hankel language [1]. Our paper inherits this predictive geometry, but deploys it in a new decision-theoretic role.

Second, our work is related to the prophet-inequality literature. Prophet inequalities compare an online decision maker against an offline benchmark with full knowledge of realized future values, and by now form a central toolkit in online decision-making, stochastic optimization, and mechanism design [11, 5]. We do not seek new classical prophet constants in their usual information model. Instead, we ask how prophet-style guarantees degrade when the online algorithm has access only to a compressed predictive state rather than the full past. In this sense, our contribution is orthogonal to much of the traditional prophet-inequality literature: the new issue is not how to choose the best threshold under a given information structure, but how the information structure itself may be compressed.

Third, our work is closely connected to price-based online algorithms. A number of works have shown that threshold strategies and posted-price mechanisms are deeply related, and that prophet inequalities can often be proved by constructing suitable price rules [7, 6, 8]. Our pricing section should be read as a predictive-state counterpart to this line of work: we show that if benchmark thresholds or prices factor through canonical predictive states, then approximation of those states induces additive losses in both stopping and pricing performance.

Finally, our framework is also complementary to recent prophet-inequality extensions that enrich the information model, including settings with unknown distributions, noisy observations, side predictions, and correlated values [4, 12, 3, 13]. These works modify the information available to the online algorithm in different ways. Our approach is different: we assume the online algorithm has access to the past only through a compressed predictive representation, and ask how much of the full benchmark can still be recovered from that state.

Organization. Section 2 introduces the sequential selection model and the full-information benchmark. Section 3 proves the compression-to-decision decomposition. Section 4 constructs canonical predictive states from Gaussian / Hardy-space data. Section 5 develops quantitative approximation bounds, including exact finite-state realizability in the rational case. Section 6 derives prophet-inequality and posted-pricing consequences. We conclude with directions toward richer price-based prophet frameworks and learned predictive states.

2 Model and Benchmarks

2.1 Standing assumptions and notation

Throughout the paper, the online decision maker acts at times $t = 1, \dots, n$. The pre-decision information at time t is denoted by $\mathcal{G}_t$, while post-observation information is denoted by $\mathcal{F}_t$.

In the predictive-state layer, we assume that the pre-decision information is generated by the past of an auxiliary observation process

$$\mathcal{G}_t = \sigma(Y_s : \; s \leq t-1),$$

where Y is a centered weakly stationary Gaussian process.

For each t, we fix a future-relevant feature

$$U_t \in L^2(\Omega; \mathbb{R}^{q_t}),$$

and assume that the decision benchmark depends on the past only through the conditional law of U_t given $\mathcal{G}_t$. In the stopping model this will be encoded through the continuation values $C_t^\star$; in the pricing model it will be encoded through benchmark price maps factoring through canonical predictive states.

2.2 Sequential selection with pre-decision information

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let

$$X_1, \ldots, X_n$$

be nonnegative integrable random variables. At each time $t \in \{1, \ldots, n\}$, the decision maker first has access to a pre-decision information σ-field

$$\mathcal{G}_t \subseteq \mathcal{F},$$

then the reward X_t is revealed, and an irrevocable stop/continue decision must be made.

We write

$$\mathcal{F}_t := \mathcal{G}_t \vee \sigma(X_t)$$

for the post-observation information at time t. An admissible stopping rule is an $\{\mathcal{F}_t\}$-stopping time τ with values in $\{1, \ldots, n, \infty\}$. We adopt the convention $X_\infty := 0$.

The offline benchmark is the prophet value

$$P_n := \mathbb{E}\Big[\max_{1 \le t \le n} X_t \Big].$$

Our goal is to compare the performance of online policies that act using compressed summaries of the past with both the full-information online optimum and the prophet benchmark.

2.3 The full-information online optimum

We first define the pre-decision value process of the full-information online optimum. Set

$$V^\star_{n+1} := 0.$$

For $t = n, n-1, \ldots, 1$, define the continuation value

$$C^\star_t := \mathbb{E}\big[V^\star_{t+1} \mid \mathcal{G}_t\big],$$

and the pre-decision value

$$V^\star_t := \mathbb{E}[\max\{X_t, C^\star_t\} \mid \mathcal{G}_t].$$

Proposition 2.1. *The threshold rule*

$$\pi^\star : \qquad \textit{stop at time } t \textit{ if and only if } X_t \ge C^\star_t$$

is optimal among all admissible stopping rules. In particular,

$$V^\star_1 = \sup_\tau \mathbb{E}[X_\tau].$$

Proof. This follows by backward induction. At time t, conditional on $\mathcal{G}_t$, the decision maker either stops and receives X_t, or continues and receives conditional value $C^\star_t = \mathbb{E}[V^\star_{t+1} \mid \mathcal{G}_t]$. Therefore the optimal post-observation action is to stop if and only if $X_t \ge C^\star_t$. □

2.4 Compressed-state threshold policies

We now introduce a compressed state representation of the past. Let $d \in \mathbb{N}$, and let

$$Z_t = \Phi_t(\mathcal{G}_t) \in \mathbb{R}^d$$

be a $\mathcal{G}_t$-measurable state, where Φ_t is a measurable compression map.

A compressed-state continuation surrogate is any $\mathcal{G}_t$-measurable random variable of the form

$$\widehat{C}_t = \Gamma_t(Z_t),$$

for some measurable map $\Gamma_t : \mathbb{R}^d \to \mathbb{R}$.

The associated compressed threshold policy is

$$\widehat{\pi} : \qquad \text{stop at time } t \text{ if and only if } X_t \geq \widehat{C}_t.$$

Its pre-decision value process is defined recursively by

$$V^{\widehat{\pi}}_{n+1} := 0,$$

$$C^{\widehat{\pi}}_t := \mathbb{E}\left[V^{\widehat{\pi}}_{t+1} \mid \mathcal{G}_t\right],$$

and

$$V^{\widehat{\pi}}_t := \mathbb{E}\left[X_t \, \mathbf{1}_{\{X_t \geq \widehat{C}_t\}} + V^{\widehat{\pi}}_{t+1} \, \mathbf{1}_{\{X_t < \widehat{C}_t\}} \;\middle|\; \mathcal{G}_t\right].$$

The central distortion quantity in this paper is the *continuation-value compression error*

$$\Delta_t := \widehat{C}_t - C^{\star}_t.$$

Remark 2.2. The quantity Δ_t is decision-theoretic rather than purely predictive. It measures not how well the past itself is reconstructed, but how well the compressed state preserves the continuation value that drives the optimal stopping decision.

3 Stability Under Predictive-State Compression

This section is the decision-theoretic core of the paper. No Gaussianity, stationarity, Hardy-space structure, or analytic machinery is used here. All such structure will enter later, through the construction of the compressed states themselves.

3.1 A threshold perturbation lemma

For $a, c \in \mathbb{R}$, define the one-step threshold payoff map

$$g_a(x, c) := x \, \mathbf{1}_{\{x \geq a\}} + c \, \mathbf{1}_{\{x < a\}}.$$

Note that

$$g_c(x, c) = \max\{x, c\}.$$

Lemma 3.1 (Threshold perturbation)**.** *For all $x, a, c, c' \in \mathbb{R}$,*

$$0 \leq \max\{x, c\} - g_a(x, c) \leq |a - c|,$$

and

$$|g_a(x, c) - g_a(x, c')| \leq |c - c'|.$$

Proof. We first prove

$$0 \le \max\{x, c\} - g_a(x, c) \le |a - c|.$$

If $a \ge c$, then $g_a(x, c) = \max\{x, c\}$ unless $x \in [c, a)$, in which case

$$\max\{x, c\} - g_a(x, c) = x - c \le a - c = |a - c|.$$

If $a < c$, then $g_a(x, c) = \max\{x, c\}$ unless $x \in [a, c)$, in which case

$$\max\{x, c\} - g_a(x, c) = c - x \le c - a = |a - c|.$$

This proves the first inequality.

For the second inequality, fix a and x. If $x \ge a$, then

$$g_a(x, c) = x = g_a(x, c'),$$

so the difference is zero. If $x < a$, then

$$g_a(x, c) = c, \qquad g_a(x, c') = c',$$

hence

$$|g_a(x, c) - g_a(x, c')| = |c - c'|.$$

□

3.2 Compression-to-decision decomposition

We now show that continuation-value distortion controls the loss of the compressed-state policy.

Theorem 3.2 (Compression-to-decision decomposition)**.** *Define*

$$D_t := V_t^\star - V_t^{\widehat{\pi}}, \qquad t = 1, \dots, n+1.$$

Then $D_{n+1} = 0$, $D_t \ge 0$ for all t, and for each $t = 1, \dots, n$,

$$D_t \le |\Delta_t| + \mathbb{E}[D_{t+1} \mid \mathcal{G}_t] \qquad a.s.$$

Consequently,

$$V_1^\star - V_1^{\widehat{\pi}} \le \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

Proof. We argue by backward induction that $D_t \ge 0$ for all t. Since $D_{n+1} = 0$, the claim is trivial at $t = n+1$. Suppose $D_{t+1} \ge 0$ a.s. Then

$$C_t^\star - C_t^{\widehat{\pi}} = \mathbb{E}[V_{t+1}^\star - V_{t+1}^{\widehat{\pi}} \mid \mathcal{G}_t] = \mathbb{E}[D_{t+1} \mid \mathcal{G}_t] \ge 0.$$

By definition,

$$V_t^\star = \mathbb{E}\big[g_{C_t^\star}(X_t, C_t^\star) \mid \mathcal{G}_t\big] = \mathbb{E}[\max\{X_t, C_t^\star\} \mid \mathcal{G}_t],$$

while

$$V_t^{\widehat{\pi}} = \mathbb{E}\Big[g_{\widehat{C}_t}(X_t, C_t^{\widehat{\pi}}) \mid \mathcal{G}_t\Big].$$

Hence

$$D_t = \mathbb{E}\Big[\max\{X_t, C_t^\star\} - g_{\widehat{C}_t}(X_t, C_t^{\widehat{\pi}}) \;\Big|\; \mathcal{G}_t\Big].$$

Add and subtract $g_{\widehat{C}_t}(X_t, C_t^\star)$:

$$D_t = \mathbb{E}\Big[\max\{X_t, C_t^\star\} - g_{\widehat{C}_t}(X_t, C_t^\star) \;\Big|\; \mathcal{G}_t\Big] + \mathbb{E}\Big[g_{\widehat{C}_t}(X_t, C_t^\star) - g_{\widehat{C}_t}(X_t, C_t^{\widehat{\pi}}) \;\Big|\; \mathcal{G}_t\Big].$$

By Lemma 3.1,

$$\max\{X_t, C_t^\star\} - g_{\widehat{C}_t}(X_t, C_t^\star) \le |\widehat{C}_t - C_t^\star| = |\Delta_t|,$$

so the first term is bounded above by $|\Delta_t|$.

Again by Lemma 3.1,

$$\Big|g_{\widehat{C}_t}(X_t, C_t^\star) - g_{\widehat{C}_t}(X_t, C_t^{\widehat{\pi}})\Big| \le |C_t^\star - C_t^{\widehat{\pi}}| = \mathbb{E}[D_{t+1} \mid \mathcal{G}_t].$$

Therefore

$$D_t \le |\Delta_t| + \mathbb{E}[D_{t+1} \mid \mathcal{G}_t].$$

This also implies $D_t \ge 0$, completing the induction.

Taking expectations and iterating backward from $t = n$ to $t = 1$, using $D_{n+1} = 0$, yields

$$\mathbb{E}[D_1] \le \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

Since $D_1 = V_1^\star - V_1^{\widehat{\pi}}$, the claim follows. □

Corollary 3.3 (Online–offline gap decomposition)**.** *The loss of the compressed-state policy relative to the prophet benchmark decomposes as*

$$P_n - V_1^{\widehat{\pi}} = \underbrace{(P_n - V_1^\star)}_{\textit{future-information gap}} + \underbrace{(V_1^\star - V_1^{\widehat{\pi}})}_{\textit{past-compression penalty}}.$$

In particular,

$$P_n - V_1^{\widehat{\pi}} \le (P_n - V_1^\star) + \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

Proof. This is immediate from

$$P_n - V_1^{\widehat{\pi}} = (P_n - V_1^\star) + (V_1^\star - V_1^{\widehat{\pi}})$$

and Theorem 3.2. □

Corollary 3.4 (Compressed-state prophet bound)**.** *Suppose a classical prophet inequality yields*

$$V_1^\star \ge \alpha P_n$$

for some $\alpha \in (0, 1]$. Then

$$V_1^{\widehat{\pi}} \ge \alpha P_n - \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

Proof. By Corollary 3.3,

$$V_1^{\widehat{\pi}} \ge P_n - (P_n - V_1^\star) - \sum_{t=1}^{n} \mathbb{E}|\Delta_t| = V_1^\star - \sum_{t=1}^{n} \mathbb{E}|\Delta_t|.$$

The stated bound follows from $V_1^\star \ge \alpha P_n$. □

3.3 Interface with predictive states

We now isolate the point at which predictive-state constructions enter the theory.

Assumption 3.5 (Decision sufficiency)**.** For each t, there exists an L^2 random vector $S_t^\star$ and a measurable map

$$\Theta_t : \mathbb{R}^{m_t} \to \mathbb{R}$$

such that

$$C_t^\star = \Theta_t(S_t^\star).$$

Assumption 3.6 (Lipschitz continuation map)**.** For each t, the map Θ_t is Lipschitz with constant $L_t < \infty$, namely

$$|\Theta_t(u) - \Theta_t(v)| \le L_t \, \|u - v\|_2 \qquad \text{for all } u, v \in \mathbb{R}^{m_t}.$$

Let $\widehat{S}_t$ be any compressed approximation to $S_t^\star$, and define

$$\widehat{C}_t := \Theta_t(\widehat{S}_t).$$

Proposition 3.7 (Predictive-state interface)**.** *Under Assumptions 3.5 and 3.6,*

$$\mathbb{E}|\Delta_t| = \mathbb{E}\big|\Theta_t(\widehat{S}_t) - \Theta_t(S_t^\star)\big| \le L_t \, \mathbb{E} \left\| \widehat{S}_t - S_t^\star \right\|_2 .$$

Consequently,

$$P_n - V_1^{\widehat{\pi}} \le (P_n - V_1^\star) + \sum_{t=1}^{n} L_t \, \mathbb{E} \left\| \widehat{S}_t - S_t^\star \right\|_2 .$$

Proof. The first bound follows directly from Assumption 3.6:

$$|\Delta_t| = |\Theta_t(\widehat{S}_t) - \Theta_t(S_t^\star)| \le L_t \left\| \widehat{S}_t - S_t^\star \right\|_2 .$$

Taking expectations and substituting into Corollary 3.3 gives the result. □

Remark 3.8 (Role of the predictive layer)**.** Section 3 reduces the sequential decision problem to the task of constructing compressed predictive states $\widehat{S}_t$ that approximate the canonical states $S_t^\star$. The next section instantiates this interface using Dym–McKean/Hardy-space machinery, finite-past approximation, and structured state compression.

4 Canonical Predictive States from Hardy-Space Compression

In this section we instantiate the abstract interface of Proposition 3.7 using a discrete-time Gaussian observation layer with Hardy-space spectral factorization. The role of this section is not yet to prove quantitative approximation rates, but to identify the canonical predictive state and to define the finite-dimensional compression families that will later be analyzed.

4.1 Gaussian observation process and past subspaces

We work with a centered p-variate weakly stationary Gaussian process

$$Y = (Y_k)_{k\in\mathbb{Z}}, \qquad Y_k \in \mathbb{R}^p,$$

with matrix spectral density

$$W(e^{i\theta}) \in \mathbb{C}^{p\times p}, \qquad \theta \in [-\pi, \pi].$$

Assumption 4.1 (Szegő outer factorization)**.** The spectral density W is positive definite a.e. and satisfies

$$\log\det W \in L^1([-\pi,\pi]).$$

Hence there exists an outer matrix function

$$H \in H^2(\mathbb{D}; \mathbb{C}^{p\times p})$$

such that

$$W(e^{i\theta}) = H(e^{i\theta})H(e^{i\theta})^* \qquad \text{for a.e. } \theta \in [-\pi,\pi].$$

For each decision time $t \in \{1,\dots,n\}$, we identify the pre-decision information with the observation past

$$\mathcal{G}_t := \sigma(Y_s : \ s \le t-1).$$

Let

$$\mathcal{H}_t^- := \overline{\operatorname{span}}\{u^\top Y_s : \ u \in \mathbb{R}^p, \ s \le t-1\} \subset L^2(\Omega).$$

For $q \in \mathbb{N}$, define the vector-valued past space

$$\mathcal{H}_t^-(q) := \left(\mathcal{H}_t^-\right)^q \subset L^2(\Omega;\mathbb{R}^q).$$

4.2 Future-relevant linear features and canonical predictive states

For each t, let

$$U_t \in L^2(\Omega; \mathbb{R}^{q_t})$$

be a future-relevant linear feature of the form

$$U_t = \sum_{j=0}^{r_t} B_{t,j} Y_{t+j}, \qquad B_{t,j} \in \mathbb{R}^{q_t\times p}.$$

The integer $r_t \ge 0$ is allowed to depend on t.

Definition 4.2 (Canonical predictive state)**.** The canonical predictive state associated with U_t is

$$S_t^\star := \Pi_{\mathcal{H}_t^-(q_t)} U_t,$$

the orthogonal projection of U_t onto the vector-valued past space.

Proposition 4.3 (Gaussian conditional law is encoded by the canonical state)**.** *For each t:*

1. $$S_t^\star = \mathbb{E}[U_t \mid \mathcal{G}_t] \qquad a.s.$$

2. *The residual*

$$R_t^\star := U_t - S_t^\star$$

is independent of $\mathcal{G}_t$.

3. *There exists a deterministic covariance matrix*

$$\Sigma_t^\perp \in \mathbb{R}^{q_t \times q_t}$$

such that

$$\mathcal{L}(U_t \mid \mathcal{G}_t) = \mathcal{N}(S_t^\star, \Sigma_t^\perp) \qquad a.s.$$

Proof. Since U_t is a finite linear combination of Gaussian coordinates $(Y_s)_{s\in\mathbb{Z}}$, it belongs to the Gaussian Hilbert space generated by Y. The conditional expectation onto $\mathcal{G}_t$ agrees with orthogonal projection onto the closed linear span $\mathcal{H}_t^-(q_t)$, which proves (1).

By construction, $R_t^\star$ is orthogonal to $\mathcal{H}_t^-(q_t)$. Since the joint law is Gaussian, orthogonality implies independence; hence $R_t^\star$ is independent of $\mathcal{G}_t$. This proves (2).

Finally, by (1) and (2),

$$U_t = S_t^\star + R_t^\star,$$

where $S_t^\star$ is $\mathcal{G}_t$-measurable and $R_t^\star$ is independent of $\mathcal{G}_t$. Therefore the conditional law of U_t given $\mathcal{G}_t$ is Gaussian with mean $S_t^\star$ and covariance

$$\Sigma_t^\perp := \mathrm{Cov}(R_t^\star),$$

which is deterministic. □

4.3 Decision sufficiency via conditional-law reduction

The previous proposition shows that, in the Gaussian linear setting, the entire conditional law of U_t given the past is encoded by the single random vector $S_t^\star$ together with a deterministic covariance matrix.

This yields a concrete sufficient condition for Assumption 3.5 from Section 3.

Proposition 4.4 (Decision sufficiency from Gaussian reduction)**.** *Assume that for each t there exists a measurable functional*

$$\Lambda_t : \mathcal{P}_2(\mathbb{R}^{q_t}) \to \mathbb{R}$$

such that

$$C_t^\star = \Lambda_t\big(\mathcal{L}(U_t \mid \mathcal{G}_t)\big) \qquad a.s.$$

Then there exists a measurable map

$$\Theta_t : \mathbb{R}^{q_t} \to \mathbb{R}$$

such that

$$C_t^\star = \Theta_t(S_t^\star) \qquad a.s.$$

In particular, Assumption 3.5 holds with state dimension $m_t = q_t$.

Proof. By Proposition 4.3,

$$\mathcal{L}(U_t \mid \mathcal{G}_t) = \mathcal{N}(S_t^\star, \Sigma_t^\perp) \qquad \text{a.s.}$$

Define

$$\Theta_t(s) := \Lambda_t\big(\mathcal{N}(s, \Sigma_t^\perp)\big), \qquad s \in \mathbb{R}^{q_t}.$$

Then

$$C_t^\star = \Lambda_t\big(\mathcal{L}(U_t \mid \mathcal{G}_t)\big) = \Lambda_t\big(\mathcal{N}(S_t^\star, \Sigma_t^\perp)\big) = \Theta_t(S_t^\star).$$

□

Remark 4.5. Proposition 4.4 is the formal place where prediction enters decision-making. Once the continuation value depends on the past only through the conditional law of a future-relevant Gaussian feature, that law collapses to the canonical predictive state $S_t^\star$.

Remark 4.6 (Observable versus innovation coordinates)**.** Although several constructions below are expressed in innovation coordinates, all predictive states and compression subspaces are elements of the observable past space $\mathcal{H}_t^-$. Thus the resulting compressed states should be interpreted as causal functions of past observations, not as quantities that require direct access to the innovation sequence.

4.4 Finite-past Hardy compression families

For $L \in \mathbb{N}$, define the finite-window past space

$$\mathcal{H}_{t,L}^- := \operatorname{span}\{u^\top Y_s : \ u \in \mathbb{R}^p, \ t-L \le s \le t-1\} \subset \mathcal{H}_t^-,$$

and again let

$$\mathcal{H}_{t,L}^-(q) := \left(\mathcal{H}_{t,L}^-\right)^q.$$

Definition 4.7 (Hardy compression family)**.** For each (m, L, t), let

$$\mathcal{K}_{m,L,t} \subseteq \mathcal{H}_{t,L}^-(q_t)$$

be a finite-dimensional subspace. We call

$$\{\mathcal{K}_{m,L,t}\}_{m,L,t}$$

a Hardy compression family if:

1. $$\mathcal{K}_{m,L,t} \subseteq \mathcal{K}_{m+1,L,t}, \qquad \mathcal{K}_{m,L,t} \subseteq \mathcal{K}_{m,L+1,t},$$

   for all admissible (m, L, t);

2. each $\mathcal{K}_{m,L,t}$ is generated from a structured basis adapted to the outer factor H (for example, a Hardy, model-space, or de Branges basis associated with the past-to-future prediction problem).

Definition 4.8 (Compressed predictive state)**.** Given a Hardy compression family, define the compressed predictive state by

$$\widehat{S}_t^{(m,L)} := \Pi_{\mathcal{K}_{m,L,t}} U_t.$$

Its approximation defect is

$$\varepsilon_t(m, L) := \left\| S_t^\star - \widehat{S}_t^{(m,L)} \right\|_{L^2(\Omega;\mathbb{R}^{q_t})}.$$

Whenever Proposition 4.4 applies, define

$$\widehat{C}_t^{(m,L)} := \Theta_t(\widehat{S}_t^{(m,L)}),$$

and let $\pi_{m,L}$ denote the corresponding compressed threshold policy.

4.5 From Hardy compression to prophet guarantees

Theorem 4.9 (Predictive-state compression bound)**.** *Assume the hypotheses of Proposition 4.4, and suppose in addition that* Θ_t *is Lipschitz with constant* L_t *for each* t*:*

$$|\Theta_t(u) - \Theta_t(v)| \le L_t \|u - v\|_2 \qquad \text{for all } u, v \in \mathbb{R}^{q_t}.$$

Then

$$\mathbb{E}\big|\, \widehat{C}_t^{(m,L)} - C_t^\star \,\big| \le L_t\, \varepsilon_t(m, L),$$

and therefore

$$P_n - V_1^{\pi_{m,L}} \le (P_n - V_1^\star) + \sum_{t=1}^{n} L_t\, \varepsilon_t(m, L).$$

Equivalently,

$$V_1^{\pi_{m,L}} \ge V_1^\star - \sum_{t=1}^{n} L_t\, \varepsilon_t(m, L).$$

Proof. By definition,

$$C_t^\star = \Theta_t(S_t^\star), \qquad \widehat{C}_t^{(m,L)} = \Theta_t(\widehat{S}_t^{(m,L)}).$$

Hence

$$\big|\widehat{C}_t^{(m,L)} - C_t^\star\big| = \big|\Theta_t(\widehat{S}_t^{(m,L)}) - \Theta_t(S_t^\star)\big| \le L_t \left\|\widehat{S}_t^{(m,L)} - S_t^\star\right\|_2.$$

Taking expectations and using the definition of $\varepsilon_t(m, L)$ yields

$$\mathbb{E}\big|\, \widehat{C}_t^{(m,L)} - C_t^\star \,\big| \le L_t\, \varepsilon_t(m, L).$$

The decision bound then follows from Corollary 3.3 in Section 3. □

Proposition 4.10 (Density implies asymptotic optimality)**.** *Fix* t*. Assume that*

$$\overline{\bigcup_{m,L\ge 1} \mathcal{K}_{m,L,t}} = \mathcal{H}_t^-(q_t).$$

Then

$$\varepsilon_t(m, L) \to 0 \qquad \text{as } m, L \to \infty,$$

along any cofinal sequence.

Consequently, if

$$\sum_{t=1}^{n} L_t\, \varepsilon_t(m, L) \to 0,$$

then

$$V_1^{\pi_{m,L}} \to V_1^\star.$$

Moreover,

$$P_n - V_1^{\pi_{m,L}} \to P_n - V_1^\star.$$

Proof. Since

$$S_t^\star = \Pi_{\mathcal{H}_t^-(q_t)} U_t, \qquad \widehat{S}_t^{(m,L)} = \Pi_{\mathcal{K}_{m,L,t}} U_t,$$

and the subspaces $\mathcal{K}_{m,L,t}$ are increasing with dense union in $\mathcal{H}_t^-(q_t)$, the Hilbert projection theorem implies

$$\left\|\widehat{S}_t^{(m,L)} - S_t^\star\right\|_{L^2} \to 0.$$

Thus $\varepsilon_t(m, L) \to 0$.

The convergence of values follows from Theorem 4.9. □

5 Quantitative Approximation: Innovation Tails and Rational Exactness

In this section we develop the first two theorem packages of the paper.

- For general outer-factor classes, we derive explicit approximation bounds from the innovation expansion of the canonical predictive state.
- For rational spectral factors, we show that the canonical predictive state admits an exact finite-dimensional realization, yielding exact compressed-state optimality.

The first part gives robust quantitative control with minimal structural assumptions; the second part identifies the exact finite-state regime that should serve as the principal model class for the first paper.

5.1 Innovation representation of the canonical predictive state

We retain the setup of Section 4. In addition to Assumption 4.1, we assume that the process admits the corresponding Wold–Hardy innovation representation

$$Y_t = \sum_{\ell=0}^{\infty} h_\ell \varepsilon_{t-\ell}, \qquad t \in \mathbb{Z},$$

where

$$H(z) = \sum_{\ell=0}^{\infty} h_\ell z^\ell$$

is the outer spectral factor of W, and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. centered Gaussian innovation sequence in $\mathbb{R}^p$ with covariance matrix I_p.

Recall that

$$U_t = \sum_{j=0}^{r_t} B_{t,j} Y_{t+j}, \qquad B_{t,j} \in \mathbb{R}^{q_t \times p}.$$

Proposition 5.1 (Innovation expansion of the canonical predictive state)**.** *Define matrices*

$$A_{t,\ell} := \sum_{j=0}^{r_t} B_{t,j} h_{j+\ell}, \qquad \ell \geq 1.$$

Then the canonical predictive state admits the exact expansion

$$S_t^\star = \sum_{\ell=1}^{\infty} A_{t,\ell}\, \varepsilon_{t-\ell} \qquad \text{in } L^2(\Omega; \mathbb{R}^{q_t}).$$

Equivalently,

$$U_t = S_t^\star + R_t^\star,$$

where

$$R_t^\star = \sum_{j=0}^{r_t} \sum_{u=0}^{j} B_{t,j} h_u\, \varepsilon_{t+j-u},$$

and $R_t^\star$ is independent of $\mathcal{G}_t$.

Proof. Expanding each Y_{t+j} in innovations gives

$$U_t = \sum_{j=0}^{r_t} B_{t,j} \sum_{\ell=0}^{\infty} h_\ell \varepsilon_{t+j-\ell}.$$

Split the sum according to whether $t+j-\ell \le t-1$ or $t+j-\ell \ge t$.

The past-measurable part corresponds to $\ell \ge j+1$. Writing

$$\ell = j + r, \qquad r \ge 1,$$

we obtain

$$\sum_{j=0}^{r_t} B_{t,j} \sum_{r=1}^{\infty} h_{j+r}\, \varepsilon_{t-r} = \sum_{r=1}^{\infty} \left(\sum_{j=0}^{r_t} B_{t,j} h_{j+r} \right) \varepsilon_{t-r} = \sum_{r=1}^{\infty} A_{t,r} \varepsilon_{t-r}.$$

This belongs to $\mathcal{H}_t^-(q_t)$, hence must coincide with $S_t^\star = \Pi_{\mathcal{H}_t^-(q_t)} U_t$.

The remaining terms are precisely those with $\ell \le j$, i.e.

$$R_t^\star = \sum_{j=0}^{r_t} \sum_{u=0}^{j} B_{t,j} h_u\, \varepsilon_{t+j-u}.$$

These depend only on innovations $\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+r_t}$ and are therefore independent of $\mathcal{G}_t$. □

Remark 5.2. Proposition 5.1 is the first place where the Hardy factor H enters quantitatively. The approximation problem for $S_t^\star$ is now reduced to approximating an explicit innovation series with coefficients $A_{t,\ell}$ determined by the spectral factor.

5.2 Finite-window compression and exact tail formula

The simplest structured compression family is the innovation-window family. For $L \in \mathbb{N}$, define

$$\mathcal{K}_{L,t}^{\mathrm{inn}} := \left\{ \sum_{\ell=1}^{L} M_\ell \varepsilon_{t-\ell} : M_\ell \in \mathbb{R}^{q_t \times p} \right\} \subset \mathcal{H}_t^-(q_t).$$

Define the truncated predictive state

$$\widehat{S}_t^{(L)} := \Pi_{\mathcal{K}_{L,t}^{\mathrm{inn}}} U_t = \sum_{\ell=1}^{L} A_{t,\ell} \varepsilon_{t-\ell}.$$

Theorem 5.3 (Exact innovation-tail formula)**.** *For every t and $L \ge 1$,*

$$\varepsilon_t(L)^2 := \left\| S_t^\star - \widehat{S}_t^{(L)} \right\|_{L^2(\Omega;\mathbb{R}^{q_t})}^2 = \sum_{\ell > L} \|A_{t,\ell}\|_F^2 .$$

Consequently, if $C_t^\star = \Theta_t(S_t^\star)$ and Θ_t is Lipschitz with constant L_t, then

$$P_n - V_1^{\pi_L} \le (P_n - V_1^\star) + \sum_{t=1}^{n} L_t \left(\sum_{\ell > L} \|A_{t,\ell}\|_F^2 \right)^{1/2},$$

where π_L denotes the threshold policy driven by

$$\widehat{C}_t^{(L)} := \Theta_t(\widehat{S}_t^{(L)}).$$

Proof. By Proposition 5.1,

$$S_t^\star - \widehat{S}_t^{(L)} = \sum_{\ell > L} A_{t,\ell}\varepsilon_{t-\ell}.$$

Since the innovations are independent with covariance I_p,

$$\mathbb{E}\left\|\sum_{\ell > L} A_{t,\ell}\varepsilon_{t-\ell}\right\|_2^2 = \sum_{\ell > L} \operatorname{tr}(A_{t,\ell}A_{t,\ell}^\top) = \sum_{\ell > L} \|A_{t,\ell}\|_F^2 .$$

This proves the exact tail formula.

The decision bound follows from Theorem 4.9 in Section 4. □

Corollary 5.4 (Geometric rate)**.** *Assume that for each t there exist constants $C_t > 0$ and $\rho_t \in (0,1)$ such that*

$$\|A_{t,\ell}\|_F \le C_t \rho_t^\ell \qquad \textit{for all } \ell \ge 1.$$

Then

$$\varepsilon_t(L) \le \frac{C_t\, \rho_t^{L+1}}{\sqrt{1-\rho_t^2}},$$

and hence

$$P_n - V_1^{\pi_L} \le (P_n - V_1^\star) + \sum_{t=1}^n \frac{L_t C_t\, \rho_t^{L+1}}{\sqrt{1-\rho_t^2}}.$$

Proof. Apply Theorem 5.3 and sum the geometric series:

$$\varepsilon_t(L)^2 \le C_t^2 \sum_{\ell > L} \rho_t^{2\ell} = C_t^2 \frac{\rho_t^{2L+2}}{1-\rho_t^2}.$$

Taking square roots gives the claim. □

Corollary 5.5 (Polynomial rate)**.** *Assume that for each t there exist constants $C_t > 0$ and $\beta_t > 0$ such that*

$$\|A_{t,\ell}\|_F \le C_t(1+\ell)^{-(\beta_t+1)} \qquad \textit{for all } \ell \ge 1.$$

Then there exists a constant $\widetilde{C}_t > 0$ such that

$$\varepsilon_t(L) \le \widetilde{C}_t\, L^{-(\beta_t + \frac{1}{2})},$$

and therefore

$$P_n - V_1^{\pi_L} \le (P_n - V_1^\star) + \sum_{t=1}^n L_t\, \widetilde{C}_t\, L^{-(\beta_t+\frac{1}{2})}.$$

Proof. By Theorem 5.3,

$$\varepsilon_t(L)^2 \le C_t^2 \sum_{\ell > L} (1+\ell)^{-2\beta_t - 2}.$$

Since $2\beta_t + 2 > 1$, the tail of this series is bounded by a constant times $L^{-(2\beta_t+1)}$. Taking square roots yields the result. □

Remark 5.6 (From spectral-factor decay to decision rates)**.** Because

$$A_{t,\ell} = \sum_{j=0}^{r_t} B_{t,j} h_{j+\ell},$$

any decay estimate on the spectral-factor coefficients h_ℓ transfers directly to the predictive coefficients $A_{t,\ell}$. Thus geometric or polynomial approximation rates for the compressed policy can be read off from Hardy-side regularity of the outer factor.

5.3 Exact finite-dimensional realization for rational spectral factors

We now identify the exact finite-state regime. This is the principal exact model class for the first paper.

Assumption 5.7 (Rational causally invertible outer factor)**.** The outer factor H is rational and admits a stable state-space realization

$$H(z) = D + z\,C(I - zA)^{-1}B,$$

with

$$A \in \mathbb{R}^{d\times d}, \qquad B \in \mathbb{R}^{d\times p}, \qquad C \in \mathbb{R}^{p\times d}, \qquad D \in \mathbb{R}^{p\times p},$$

and spectral radius $\rho(A) < 1$. In addition,

$$H^{-1} \in H^\infty(\mathbb{D}; \mathbb{C}^{p\times p}).$$

Under Assumption 5.7,

$$h_0 = D, \qquad h_\ell = CA^{\ell-1}B, \qquad \ell \geq 1.$$

Definition 5.8 (Rational predictive state)**.** Define the d-dimensional state

$$\xi_t := \sum_{\ell=1}^{\infty} A^{\ell-1} B\, \varepsilon_{t-\ell}.$$

Proposition 5.9 (Finite-dimensional realization of the canonical predictive state)**.** *Under Assumption 5.7, the process (ξ_t) is $\mathcal{G}_t$-measurable and obeys the linear recursion*

$$\xi_{t+1} = A\xi_t + B\varepsilon_t.$$

Moreover, for each t,

$$S_t^\star = M_t \xi_t, \qquad M_t := \sum_{j=0}^{r_t} B_{t,j} C A^j.$$

Proof. By Proposition C.1 in Appendix C, the innovation sequence is recoverable from the observable past by a stable causal filter. Hence the state

$$\xi_t = \sum_{\ell=1}^{\infty} A^{\ell-1} B\, \varepsilon_{t-\ell}$$

is $\mathcal{G}_t$-measurable.

The recursion

$$\xi_{t+1} = \sum_{\ell=1}^{\infty} A^{\ell-1} B\, \varepsilon_{t+1-\ell} = B\varepsilon_t + \sum_{\ell=2}^{\infty} A^{\ell-1} B\, \varepsilon_{t+1-\ell} = B\varepsilon_t + A\xi_t$$

is immediate from the definition.

Next, by Proposition 5.1 and the rational form of h_ℓ,

$$A_{t,\ell} = \sum_{j=0}^{r_t} B_{t,j} h_{j+\ell} = \sum_{j=0}^{r_t} B_{t,j} C A^{j+\ell-1} B = \left(\sum_{j=0}^{r_t} B_{t,j} C A^j \right) A^{\ell-1} B = M_t A^{\ell-1} B.$$

Therefore

$$S_t^\star = \sum_{\ell=1}^{\infty} A_{t,\ell} \varepsilon_{t-\ell} = \sum_{\ell=1}^{\infty} M_t A^{\ell-1} B\, \varepsilon_{t-\ell} = M_t \xi_t.$$

□

Theorem 5.10 (Exact finite-state compression in the rational case)**.** *Assume the hypotheses of Proposition 4.4, Assumption 3.6, and Assumption 5.7. Let the compressed state be*

$$Z_t := \xi_t \in \mathbb{R}^d.$$

Define

$$\widehat{C}_t := \Theta_t(M_t Z_t).$$

Then

$$\widehat{C}_t = C_t^\star \qquad \textit{a.s. for every } t,$$

and hence the compressed threshold policy π_{rat} *coincides with the full-information online optimum:*

$$V_1^{\pi_{\mathrm{rat}}} = V_1^\star.$$

Consequently,

$$P_n - V_1^{\pi_{\mathrm{rat}}} = P_n - V_1^\star.$$

Proof. By Proposition 5.9,

$$S_t^\star = M_t \xi_t = M_t Z_t.$$

By decision sufficiency,

$$C_t^\star = \Theta_t(S_t^\star) = \Theta_t(M_t Z_t) = \widehat{C}_t.$$

Therefore the threshold rule generated by $\widehat{C}_t$ is exactly the optimal rule $\pi^\star$ from Proposition 2.1. Hence

$$V_1^{\pi_{\mathrm{rat}}} = V_1^\star.$$

The final identity is immediate. □

Corollary 5.11 (Exact transfer of prophet guarantees in the rational case)**.** *If the underlying model class admits a prophet inequality*

$$V_1^\star \geq \alpha P_n$$

for some $\alpha \in (0, 1]$*, then under Assumption 5.7,*

$$V_1^{\pi_{\mathrm{rat}}} \geq \alpha P_n.$$

Proof. Combine Theorem 5.10 with the displayed bound for $V_1^\star$. □

Remark 5.12 (AAK/Hankel refinement)**.** Theorem 5.3 suggests a direct but possibly inefficient compression family. A sharper alternative is to construct $\mathcal{K}_{m,L,t}$ from a finite-rank approximation of the relevant predictive Hankel operator. This is natural for at least two reasons:

1. finite-rank Hankel structure is the analytic signature of rationality;
2. AAK theory identifies best rank-m Hankel approximation with rational approximation in Hardy-space language.

In the rational case, the Hankel rank is finite and exactness should occur once m exceeds that rank. Developing this sharper compression theorem is a natural goal for a strengthened final version or a second paper.

6 Prophet and Posted-Pricing Consequences

This section records the main consequences of the predictive-state framework for prophet inequalities and posted-price mechanisms.

6.1 Prophet-inequality transfers

Corollary 6.1 (General prophet transfer)**.** *Assume that the full-information online benchmark satisfies*

$$V_1^\star \geq \alpha P_n$$

for some $\alpha \in (0,1]$. Under the hypotheses of Theorem 4.9,

$$V_1^{\pi_{m,L}} \geq \alpha P_n - \sum_{t=1}^{n} L_t\, \varepsilon_t(m,L).$$

Equivalently,

$$P_n - V_1^{\pi_{m,L}} \leq (1-\alpha)P_n + \sum_{t=1}^{n} L_t\, \varepsilon_t(m,L).$$

Proof. By Theorem 4.9,

$$V_1^{\pi_{m,L}} \geq V_1^\star - \sum_{t=1}^{n} L_t\, \varepsilon_t(m,L).$$

Combining this with $V_1^\star \geq \alpha P_n$ yields the claim. □

Corollary 6.2 (Innovation-tail prophet bound)**.** *Under the hypotheses of Theorem 5.3, if*

$$V_1^\star \geq \alpha P_n,$$

then

$$V_1^{\pi_L} \geq \alpha P_n - \sum_{t=1}^{n} L_t \left(\sum_{\ell > L} \|A_{t,\ell}\|_F^2 \right)^{1/2}.$$

Proof. Combine Corollary 6.1 with Theorem 5.3. □

Corollary 6.3 (Geometric prophet transfer)**.** *Under the hypotheses of Corollary 5.4, if*

$$V_1^\star \geq \alpha P_n,$$

then

$$V_1^{\pi_L} \geq \alpha P_n - \sum_{t=1}^{n} \frac{L_t C_t \, \rho_t^{L+1}}{\sqrt{1-\rho_t^2}}.$$

Proof. Immediate from Corollary 6.2. ☐

Corollary 6.4 (Polynomial prophet transfer)**.** *Under the hypotheses of Corollary 5.5, if*

$$V_1^\star \geq \alpha P_n,$$

then

$$V_1^{\pi_L} \geq \alpha P_n - \sum_{t=1}^{n} L_t \, \widetilde{C}_t \, L^{-(\beta_t + \frac{1}{2})}.$$

Proof. Immediate from Corollary 6.1 and Corollary 5.5. ☐

Corollary 6.5 (Rational exact transfer)**.** *Assume the hypotheses of Theorem 5.10. If the full-information online benchmark satisfies*

$$V_1^\star \geq \alpha P_n,$$

then the rational predictive-state policy satisfies the same guarantee:

$$V_1^{\pi_{\mathrm{rat}}} \geq \alpha P_n.$$

In particular,

$$P_n - V_1^{\pi_{\mathrm{rat}}} = P_n - V_1^\star.$$

Proof. This follows immediately from Theorem 5.10. ☐

Remark 6.6 (Interpretation)**.** The preceding corollaries show that prophet inequalities survive predictive-state compression with an additive penalty that is entirely prediction-theoretic. Thus the online approximation loss of a compressed policy is the sum of a classical prophet term and a Hardy-side state-compression term. In the rational case, the latter vanishes exactly.

6.2 Adaptive posted pricing with predictive states

We now formulate a pricing analogue of the compression principle. We consider the single-item sequential posted-pricing problem.

At each time $t \in \{1, \ldots, n\}$, provided the item has not yet been sold, a buyer arrives with random valuation

$$B_t \geq 0.$$

Before posting a price, the seller observes pre-price information

$$\mathcal{G}_t.$$

Let

$$\overline{F}_t(p) := \mathbb{P}(B_t \geq p \mid \mathcal{G}_t)$$

denote the conditional sale probability at price p.

If the seller posts price p_t, then the item is sold immediately if $B_t \geq p_t$, yielding revenue p_t; otherwise the seller continues to the next period. Thus, given a continuation value $c \geq 0$, the one-step conditional revenue functional is

$$J_t(p, c) := p\,\overline{F}_t(p) + \bigl(1 - \overline{F}_t(p)\bigr)c.$$

Assumption 6.7 (Price regularity)**.** For each t, the admissible price set $\mathcal{P}_t$ is a nonempty compact subset of $\mathbb{R}_+$, and for every fixed continuation value c, the map

$$p \mapsto J_t(p, c)$$

is continuous on $\mathcal{P}_t$.

Lemma 6.8 (Existence of measurable benchmark prices)**.** *Under Assumption 6.7, there exists a $\mathcal{G}_t$-measurable selector*

$$p_t^\star \in \arg\max_{p \in \mathcal{P}_t} J_t(p, K_t^\star).$$

Proof. This is a standard measurable maximum argument on a compact action set. □

Definition 6.9 (Full-information posted-pricing benchmark)**.** Set

$$R_{n+1}^\star := 0.$$

For $t = n, n-1, \ldots, 1$, define

$$K_t^\star := \mathbb{E}[R_{t+1}^\star \mid \mathcal{G}_t],$$

choose a measurable maximizer

$$p_t^\star \in \arg\max_{p \in \mathcal{P}_t} J_t(p, K_t^\star),$$

and set

$$R_t^\star := J_t(p_t^\star, K_t^\star).$$

Definition 6.10 (Compressed predictive-state pricing policy)**.** Let $\widehat{S}_t$ be a compressed predictive state. A predictive-state price map is a measurable function

$$\Psi_t : \mathbb{R}^{m_t} \to \mathbb{R}_+.$$

We define the compressed posted price by

$$\widehat{p}_t := \Psi_t(\widehat{S}_t).$$

The induced revenue-to-go process is

$$R_{n+1}^{\widehat{\mu}} := 0, \qquad K_t^{\widehat{\mu}} := \mathbb{E}[R_{t+1}^{\widehat{\mu}} \mid \mathcal{G}_t],$$

and

$$R_t^{\widehat{\mu}} := J_t(\widehat{p}_t, K_t^{\widehat{\mu}}).$$

Assumption 6.11 (Price-response Lipschitzness)**.** For each t, there exists $\kappa_t < \infty$ such that

$$|J_t(p, c) - J_t(p', c)| \leq \kappa_t |p - p'|$$

for all $p, p' \in \mathcal{P}_t$ and all admissible continuation values c.

Theorem 6.12 (Price perturbation decomposition)**.** *Let*

$$D_t^{\mathrm{pp}} := R_t^\star - R_t^{\widehat{\mu}}.$$

Under Assumption 6.11,

$$D_t^{\mathrm{pp}} \le \kappa_t |p_t^\star - \widehat{p}_t| + \mathbb{E}[D_{t+1}^{\mathrm{pp}} \mid \mathcal{G}_t] \qquad a.s.$$

for every $t = 1, \dots, n$. Consequently,

$$R_1^\star - R_1^{\widehat{\mu}} \le \sum_{t=1}^{n} \kappa_t \, \mathbb{E}|p_t^\star - \widehat{p}_t|.$$

Proof. By definition of $R_t^\star$ and $R_t^{\widehat{\mu}}$,

$$D_t^{\mathrm{pp}} = J_t(p_t^\star, K_t^\star) - J_t(\widehat{p}_t, K_t^{\widehat{\mu}}).$$

Add and subtract $J_t(\widehat{p}_t, K_t^\star)$:

$$D_t^{\mathrm{pp}} = \big(J_t(p_t^\star, K_t^\star) - J_t(\widehat{p}_t, K_t^\star)\big) + \big(J_t(\widehat{p}_t, K_t^\star) - J_t(\widehat{p}_t, K_t^{\widehat{\mu}})\big).$$

The first term is bounded by

$$\kappa_t |p_t^\star - \widehat{p}_t|$$

by Assumption 6.11.

For the second term, since

$$J_t(p, c) = p \, \overline{F}_t(p) + (1 - \overline{F}_t(p))c,$$

we have

$$|J_t(\widehat{p}_t, K_t^\star) - J_t(\widehat{p}_t, K_t^{\widehat{\mu}})| = \big(1 - \overline{F}_t(\widehat{p}_t)\big)|K_t^\star - K_t^{\widehat{\mu}}| \le |K_t^\star - K_t^{\widehat{\mu}}|.$$

Finally,

$$K_t^\star - K_t^{\widehat{\mu}} = \mathbb{E}[R_{t+1}^\star - R_{t+1}^{\widehat{\mu}} \mid \mathcal{G}_t] = \mathbb{E}[D_{t+1}^{\mathrm{pp}} \mid \mathcal{G}_t].$$

Combining the bounds yields

$$D_t^{\mathrm{pp}} \le \kappa_t |p_t^\star - \widehat{p}_t| + \mathbb{E}[D_{t+1}^{\mathrm{pp}} \mid \mathcal{G}_t].$$

Taking expectations and iterating backward from $t = n$ proves the claim. □

Corollary 6.13 (Predictive-state pricing transfer)**.** *Assume that the benchmark prices are functions of the canonical predictive states:*

$$p_t^\star = \Psi_t(S_t^\star),$$

where Ψ_t is Lipschitz with constant M_t:

$$|\Psi_t(u) - \Psi_t(v)| \le M_t \left\| u - v \right\|_2.$$

Let

$$\widehat{p}_t := \Psi_t(\widehat{S}_t).$$

Then, under Assumption 6.11,

$$R_1^\star - R_1^{\widehat{\mu}} \le \sum_{t=1}^{n} \kappa_t M_t \, \mathbb{E} \left\| \widehat{S}_t - S_t^\star \right\|_2.$$

In particular, for the Hardy-compression family of Definition 4.8,

$$R_1^\star - R_1^{\widehat{\mu}_{m,L}} \le \sum_{t=1}^{n} \kappa_t M_t \, \varepsilon_t(m, L).$$

Proof. By Lipschitz continuity of Ψ_t,

$$|p_t^\star - \widehat{p}_t| = |\Psi_t(S_t^\star) - \Psi_t(\widehat{S}_t)| \le M_t \left\| \widehat{S}_t - S_t^\star \right\|_2 .$$

Substitute this into Theorem 6.12. □

Corollary 6.14 (Rational exact pricing transfer)**.** *Assume the rational setup of Theorem 5.10, and suppose the benchmark price rule factors through the canonical predictive state:*

$$p_t^\star = \Psi_t(S_t^\star).$$

If we use the rational predictive state

$$Z_t = \xi_t$$

from Definition 5.8, then the compressed pricing policy satisfies

$$\widehat{p}_t = p_t^\star \qquad \text{a.s. for every } t,$$

and hence

$$R_1^{\widehat{\mu}_{\mathrm{rat}}} = R_1^\star.$$

Proof. By Proposition 5.9,

$$S_t^\star = M_t \xi_t.$$

Thus the rational predictive state recovers the exact canonical state, and so

$$\widehat{p}_t = \Psi_t(S_t^\star) = p_t^\star.$$

The equality of revenue follows from Theorem 6.12. □

Remark 6.15 (How this connects to prophet-based pricing frameworks)**.** Theorem 6.12 is the single-item pricing analogue of the compression-to-decision decomposition for stopping rules. It should be viewed as the local building block for more general price-based prophet frameworks. In settings where a prophet inequality is established through benchmark prices (e.g. balanced or smooth posted-price rules), one expects the same predictive-state compression principle to propagate through the proof once the benchmark prices are measurable functions of the canonical predictive state and the price-based inequalities are stable under perturbation.

7 Discussion and Open Directions

The results of this paper identify predictive-state compression as a new information-theoretic layer beneath prophet inequalities and posted-price mechanisms. At a technical level, the paper isolates a canonical predictive state, proves that continuation values and prices are stable under perturbations of this state, and then transfers approximation bounds from the prediction layer to the decision layer. At a conceptual level, it suggests a broader program: online approximation guarantees should be analyzable not only in terms of the future information that is missing, but also in terms of how the available past is represented and compressed.

We conclude by outlining several directions in which this program can be pushed further.

7.1 Beyond single-item selection and pricing

The most immediate extension is from single-choice stopping and single-item posted pricing to richer feasibility environments. The price-based prophet framework suggests that many stochastic online maximization problems can be analyzed by benchmark prices rather than by raw threshold rules. This raises the following question.

> *When a benchmark price system is measurable with respect to a canonical predictive state, can predictive-state compression be propagated through the entire price-based proof in order to obtain additive or multiplicative approximation guarantees under compressed information?*

A positive answer would extend the present paper from binary stopping and single-item pricing to settings such as packing, matroids, matching, and more general online allocation problems. Conceptually, this would turn predictive compression into a reusable module inside price-based prophet frameworks.

7.2 Noisy observations, side predictions, and learning-augmented prophet inequalities

Recent work has moved prophet inequalities toward richer information models, including noisy observations and external predictions [4, 3, 13]. This suggests a natural next step for our framework. Instead of assuming that a compressed predictive state is computed from an ideal past, one may let the state be learned from noisy or partially observed historical signals. The resulting theory would have to combine two types of error:

$$\text{decision loss} \lesssim \text{predictive-state approximation error} + \text{observation / learning error}.$$

This leads to a concrete research agenda:

1. derive predictive-state compression bounds when the observation layer is noisy or misspecified;
2. combine those bounds with recent prophet-inequality models with noisy observations or side predictions;
3. study robustness/consistency tradeoffs when a learned predictor or side signal is accurate on average but unreliable in the worst case.

7.3 Correlations and richer dependence structures

Our present Gaussian observation layer is highly structured and is chosen because it gives a transparent bridge to Hardy-space prediction. But sequential decision problems often involve dependence structures that are not captured by independent or conditionally independent value models. This motivates a broader question:

> *Can predictive-state compression be used to organize prophet inequalities under dependence, by separating the effect of dependence itself from the effect of finite-memory or compressed-state representation?*

The recent literature on prophet inequalities with correlated values suggests that threshold-based methods become substantially more delicate once one leaves the product setting [12]. From our perspective, this is precisely where a predictive state may matter most: rather than reacting to raw correlated histories, an online policy could act on a canonical state that captures only the future-relevant dependence.

7.4 Continuous-time prediction and canonical systems

Although the present paper is written in discrete time, the predictive layer is clearly not limited to that setting. One of the motivations from the very beginning was the Dym–McKean/Bingham perspective on whole-past and finite-past prediction in continuous time [2]. This suggests a second line of future work: replace the discrete Hardy-space model on the disk by a continuous-time upper-half-plane or canonical-system model, and then transport the resulting predictive-state approximation theory back into sequential decision problems through discretization, windowing, or event-driven stopping models.

7.5 Low-rank Hardy compression and model reduction

The innovation-window truncation theorem in this paper is robust but not necessarily optimal in state dimension. The next natural improvement is therefore low-rank predictive compression. The guiding idea is to replace raw window truncation by a structured finite-rank approximation of the relevant predictive Hankel operator. This should produce substantially sharper guarantees, especially when the predictive layer is close to rational or nearly rational. In the rational case, the present paper already proves exact finite-state realizability. The next step is to understand the *near-rational* regime.

7.6 Predictive states as a bridge to learned sequential decision models

Finally, the present framework suggests a more ambitious connection to modern AI and sequential learning. A recurrent or state-space decision model can be viewed as a compressed representation of the past. From this viewpoint, the canonical predictive state becomes a normative benchmark: it is the structured state that preserves the continuation values relevant for downstream decisions.

This suggests several longer-term directions:

1. define representation error for sequential decision models in terms of distance to a canonical predictive state;
2. relate state dimension, memory length, and online performance through predictive-state approximation quantities;
3. study whether learned recurrent or state-space policies can be regularized toward causally stable predictive-state architectures.

Final perspective. The main message of the paper is that missing future information is not the only source of online suboptimality. Even when the relevant history is available in principle, the online policy may only access it through a compressed state. Once this distinction is made explicit, prediction theory becomes directly relevant to online decision-making. In this sense, the present work should be read not as an endpoint, but as a first step toward a broader theory of information-constrained sequential optimization.

A Auxiliary Hilbert-space facts

This appendix records two standard Hilbert-space facts used implicitly in the main text.

Lemma A.1 (Projection onto increasing closed subspaces)**.** *Let $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \cdots$ be closed subspaces of a Hilbert space $\mathcal{H}$, and let*

$$\mathcal{M}_\infty := \overline{\bigcup_{m\geq 1} \mathcal{M}_m}.$$

If P_m and P_∞ denote the orthogonal projections onto $\mathcal{M}_m$ and $\mathcal{M}_\infty$, respectively, then for every $x \in \mathcal{H}$,

$$\|P_m x - P_\infty x\| \to 0 \qquad \text{as } m \to \infty.$$

Proof. Set $y = P_\infty x$. Since $y \in \mathcal{M}_\infty$, there exists a sequence $y_m \in \mathcal{M}_m$ such that $\|y_m - y\| \to 0$. By the minimizing property of orthogonal projection,

$$\|x - P_m x\| \leq \|x - y_m\|.$$

Passing to the limit gives

$$\limsup_{m\to\infty} \|x - P_m x\| \leq \|x - y\| = \|x - P_\infty x\|.$$

On the other hand, because $\mathcal{M}_m \subseteq \mathcal{M}_\infty$,

$$\|x - P_\infty x\| \leq \|x - P_m x\|.$$

Hence $\|x - P_m x\| \to \|x - P_\infty x\|$. Using the Pythagorean identity,

$$\|P_\infty x - P_m x\|^2 = \|x - P_m x\|^2 - \|x - P_\infty x\|^2 \to 0.$$

□

Lemma A.2 (Vector-valued projection convergence)**.** *Let $\mathcal{M}_m \subseteq \mathcal{M}_\infty$ be as in Lemma A.1. Then for any finite q and any $x \in \mathcal{H}^q$,*

$$\|P_m^{(q)} x - P_\infty^{(q)} x\|_{\mathcal{H}^q} \to 0,$$

where $P_m^{(q)}$ and $P_\infty^{(q)}$ denote the coordinatewise orthogonal projections onto $\mathcal{M}_m^q$ and $\mathcal{M}_\infty^q$.

Proof. Apply Lemma A.1 coordinatewise and sum over the q coordinates. □

B Measurability and selection details

We give a precise measurable-maximum statement underlying Lemma 6.8.

Proposition B.1 (Measurable maximum on a compact action set)**.** *Let $(\Omega, \mathcal{G})$ be a measurable space, let $P \subset \mathbb{R}$ be a nonempty compact set, and let $f : \Omega \times P \to \mathbb{R}$ satisfy:*

1. *for every $p \in P$, the map $\omega \mapsto f(\omega, p)$ is $\mathcal{G}$-measurable;*
2. *for every $\omega \in \Omega$, the map $p \mapsto f(\omega, p)$ is continuous.*

Then the argmax correspondence

$$\Gamma(\omega) := \arg\max_{p\in P} f(\omega, p)$$

has nonempty compact values and admits a $\mathcal{G}$-measurable selector.

Proof. By continuity and compactness, $\Gamma(\omega)$ is nonempty and compact for all ω. Since f is a Carathéodory integrand, the measurable maximum theorem implies that the value function

$$\omega \mapsto \max_{p \in P} f(\omega, p)$$

is measurable and that Γ has measurable graph. Because P is a compact metric space, the Kuratowski–Ryll-Nardzewski measurable selection theorem provides a measurable selector. $\square$

Proof of Lemma 6.8. Apply Proposition B.1 with

$$f(\omega, p) = J_t\big(p, K_t^\star(\omega)\big).$$

The assumed continuity in p and compactness of $\mathcal{P}_t$ give a nonempty compact argmax set, and measurability follows because $K_t^\star$ is $\mathcal{G}_t$-measurable. $\square$

C State-space realization details in the rational case

The purpose of this appendix is to justify the measurability claim used in Proposition 5.9.

Proposition C.1 (Causal invertibility and recovery of innovations)**.** *Assume Assumption 5.7. Write*

$$G(z) := H(z)^{-1} = \sum_{m=0}^{\infty} g_m z^m,$$

where $G \in H^\infty(\mathbb{D}; \mathbb{C}^{p \times p})$. Then the innovation sequence admits the causal representation

$$\varepsilon_t = \sum_{m=0}^{\infty} g_m Y_{t-m} \qquad \text{in } L^2(\Omega; \mathbb{R}^p).$$

In particular, ε_t is measurable with respect to $\sigma(Y_s : s \le t)$.

Proof. Since

$$Y_t = \sum_{\ell=0}^{\infty} h_\ell \varepsilon_{t-\ell},$$

we may write formally $Y = H(L)\varepsilon$, where L denotes the backward shift. Because H is causally invertible and $G = H^{-1} \in H^\infty$, we have $\varepsilon = G(L)Y$. Expanding G gives

$$\varepsilon_t = \sum_{m=0}^{\infty} g_m Y_{t-m}.$$

To justify convergence, note that $G \in H^\infty$ defines a bounded causal convolution operator on square-integrable one-sided sequences; applying this operator to the Wold representation of Y yields the identity in L^2. Since all finite partial sums are measurable with respect to $\sigma(Y_s : s \le t)$ and the limit holds in L^2, the limit is also measurable with respect to that σ-field. $\square$

Corollary C.2. *Under the hypotheses of Proposition C.1, for every t the state*

$$\xi_t = \sum_{\ell=1}^{\infty} A^{\ell-1} B\, \varepsilon_{t-\ell}$$

is measurable with respect to $\mathcal{G}_t = \sigma(Y_s : s \le t-1)$.

Proof. Each $\varepsilon_{t-\ell}$ is measurable with respect to $\sigma(Y_s : s \le t-\ell) \subseteq \mathcal{G}_t$. Hence every finite partial sum is $\mathcal{G}_t$-measurable. Since $\rho(A) < 1$, the series converges in $L^2(\Omega; \mathbb{R}^d)$, and therefore its limit is also $\mathcal{G}_t$-measurable. $\square$

References

[1] Damir Z. Arov and Harry Dym. *Multivariate Prediction, de Branges Spaces, and Related Extension and Inverse Problems*. Birkhäuser, Cham, 2018.

[2] N. H. Bingham. Prediction theory in continuous time. *arXiv preprint arXiv:2111.08560*, 2021.

[3] Johannes Brüstle, Ilan Reuven Cohen, and Stefano Leonardi. Prophet inequality with conservative prediction. *arXiv preprint arXiv:2602.17358*, 2026.

[4] José Correa, Paul Dütting, Felix Fischer, and Kevin Schewior. Prophet inequalities for i.i.d. random variables from an unknown distribution. *Mathematics of Operations Research*, 46(4):1247–1279, 2021.

[5] José Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Recent developments in prophet inequalities. *ACM SIGecom Exchanges*, 17(1):61–70, 2019.

[6] José Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Posted price mechanisms and optimal threshold strategies for random arrivals. *Mathematics of Operations Research*, 46(4):1452–1478, 2021.

[7] José Correa, Patricio Foncea, Dana Pizarro, and Victor Verdugo. From pricing to prophets, and back! *Operations Research Letters*, 47(1):25–29, 2019.

[8] Paul Dütting, Michal Feldman, Thomas Kesselheim, and Brendan Lucier. Prophet inequalities made easy: Stochastic optimization by pricing nonstochastic inputs. *SIAM Journal on Computing*, 49(3):540–582, 2020.

[9] Harry Dym and Henry P. McKean. *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Academic Press, New York, 1976.

[10] Harry Dym and Henry P. McKean. *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Dover, Mineola, NY, 2008.

[11] Theodore P. Hill and Robert P. Kertz. A survey of prophet inequalities in optimal stopping theory. In *Strategies for Sequential Search and Selection in Real Time*, Contemporary Mathematics 125, pages 191–207. American Mathematical Society, 1992.

[12] Nicole Immorlica, Sahil Singla, and Bo Waggoner. Prophet inequalities with linear correlations and augmentations. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 159–185, 2020.

[13] Jung-hun Kim and Vianney Perchet. Learning in prophet inequalities with noisy observations. In *International Conference on Learning Representations (ICLR)*, 2026.