# Temporally Admissible Belief-to-Policy Compilation

A Filtration-Constrained, Divergence-Driven, Convex-Projected
Architecture for Auditable Financial Decision Systems

Research Manuscript

Audience: PhD-level readers in Machine Learning, Quantitative Finance, and Stochastic Control

February 2026

### Abstract

We formalize a stochastic control architecture that maps market-implied priors, thesis posteriors, and a decomposed divergence object into auditable, deterministic policy artifacts over a temporally admissible filtration. The system enforces a strict information-measurability contract—no component of the decision pipeline may access data whose publication timestamp exceeds the current decision time—thereby eliminating look-ahead bias by construction. We provide (i) a measure-theoretic statement of temporal admissibility with an operational proof of filtration-measurability, (ii) a complete derivation of the Breeden–Litzenberger density-inversion bridge from option prices to risk-neutral priors, (iii) convex analysis of the constrained quadratic projection compiler with explicit Karush–Kuhn–Tucker conditions and economic shadow-price interpretation, (iv) Wasserstein distributionally robust extensions via Kantorovich duality with closed-form tractable reductions, (v) an implicit-function-theorem-based sensitivity calculus for differentiable bilevel decision-focused learning, and (vi) a two-timescale stochastic-approximation interpretation of the deep reinforcement learning policy split. All mathematical steps are derived in full, requiring consultation of external references only for foundational existence theorems.

## Contents

# 1 Introduction

## 1.1 Motivation

Financial decision systems operate under a compound challenge: they must integrate heterogeneous information sources—options markets, macroeconomic indicators, textual evidence, knowledge graphs—into executable portfolio allocations, and they must do so under strict temporal integrity constraints that prevent any form of future-information leakage. The failure to enforce temporal admissibility renders backtests meaningless and production decisions legally and financially hazardous.

Most practical systems either (a) use a single predictive distribution and optimize against it, ignoring the distinction between market-implied and privately-held beliefs, or (b) rely on opaque end-to-end models that resist audit. This paper formalizes an architecture that maintains an explicit dual-distribution structure—a market-implied prior $Q_t$ and a thesis posterior $P_t$—with decisions driven by their *geometric disagreement* $\Delta_t$, subject to convex feasibility constraints and full auditability.

## 1.2 Contributions

This manuscript makes six primary contributions, each developed with complete mathematical derivations:

(i) **Temporal admissibility as a measure-theoretic invariant.** We define the AS_OF($t$) operator on a bitemporal data model, construct the induced filtration, and prove that all emitted policy artifacts are measurable with respect to it (Section 2).

(ii) **Market prior via Breeden–Litzenberger inversion.** We derive, from first principles, the second-derivative identity connecting European call prices to risk-neutral state-price densities, then formulate the discrete constrained estimator as a convex quadratic program with explicit noise-bias trade-off analysis (Section 3).

(iii) **Thesis posterior as simplex-preserving belief operators.** We characterize the class of multiplicative tilting maps on the probability simplex, prove simplex invariance, establish local Lipschitz regularity, and connect calibration quality to Kullback–Leibler divergence via strict propriety of the logarithmic scoring rule (Section 4).

(iv) **Divergence geometry.** We define a composite divergence functional combining Kullback–Leibler, Wasserstein-1, tail-gap, and dependence-mismatch terms, derive its gradient structure, and justify the decomposition from an information-theoretic and optimal-transport perspective (Section 5).

(v) **Convex projection compiler with robust extensions.** We formulate the strategy compiler as a constrained quadratic program, derive its KKT system with economic interpretation of dual multipliers, and extend it to distributionally robust counterparts via Wasserstein ambiguity sets with full Kantorovich duality derivations (Sections 6 and 7).

(vi) **Differentiable bilevel extension and two-timescale DRL.** We derive exact implicit gradients through the KKT system under regularity conditions, present the matrix-calculus form via Schur complements, and formalize the reinforcement learning adaptation layer as a two-timescale stochastic approximation with convergence conditions (Sections 8 and 9).

### 1.3 Notation

We collect all notation used throughout the paper. Every symbol is defined at its first use and listed here for reference.

| **Symbol** | **Meaning** |
|---|---|
| $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^{T}, \mathbb{P})$ | Filtered probability space |
| $t \in \{0, 1, \ldots, T\}$ | Discrete decision times |
| $\mathcal{X}_t$ | Set of records admissible at time $t$ |
| $\mathcal{F}_t = \sigma(\mathcal{X}_0, \ldots, \mathcal{X}_t)$ | Information $\sigma$-algebra at time $t$ |
| $Q_t$ | Market-implied risk-neutral prior at time $t$ |
| $P_t$ | Thesis posterior (agent's subjective belief) at time $t$ |
| $\Delta_t$ | Divergence object: $\Delta_t = D(P_t \| Q_t)$ |
| $w_t \in \mathbb{R}^N$ | Portfolio weight vector at time $t$ |
| $w_t^{\text{raw}}$ | Unconstrained raw target weights |
| $w_t^{\text{prev}} = w_{t-1}$ | Previous period's weights |
| $\mathcal{C}_t$ | Feasible set (polyhedral) at time $t$ |
| $\Sigma_t$ | Covariance matrix of asset returns |
| $S_t$ | Underlying asset price at time $t$ |
| $C_t(K)$ | European call price at strike $K$, time $t$ |
| $r$ | Risk-free rate; $\tau$ time to expiration |
| $q_t(K)$ | Risk-neutral state-price density |
| $\Delta^{B-1}$ | Probability simplex in $\mathbb{R}^B$ |
| $\mathrm{KL}(P \| Q)$ | Kullback–Leibler divergence |
| $W_c(P, Q)$ | Optimal transport cost (Wasserstein distance) |
| $\mathrm{CVaR}_\alpha(L)$ | Conditional Value-at-Risk at level $\alpha$ |
| $\theta$ | Generic parameter vector for learnable models |
| $\pi_\theta$ | Parameterized policy function |

## 2 Mathematical Foundations: Temporal Admissibility

The entire architecture rests on a single invariant: no decision may depend on information unavailable at the time it is made. We formalize this as a measurability condition on a filtered probability space.

## 2.1 Bitemporal data model

Every data record $x$ carries two timestamps: a *knowledge time* $t_{\text{know}}(x)$ (when the event occurred) and a *publication time* $t_{\text{pub}}(x)$ (when the record became available to the system). The decision-relevant timestamp is always the publication time, because a fact that occurred but has not yet been published cannot inform a legitimate decision.

**Definition 2.1** (Admissible record set)**.** At decision time $t$, the set of *admissible records* is

$$\mathcal{X}_t \;=\; \big\{\, x \,:\, t_{\text{pub}}(x) \le t \,\big\}. \tag{1}$$

We write AS_OF$(t)$ for the operator that restricts any data-access query to return only records in $\mathcal{X}_t$.

**Definition 2.2** (Decision filtration)**.** The *decision filtration* is the increasing family of $\sigma$-algebras

$$\mathcal{F}_t \;=\; \sigma\big(\mathcal{X}_0,\, \mathcal{X}_1,\, \ldots,\, \mathcal{X}_t\big), \qquad t = 0, 1, \ldots, T, \tag{2}$$

where $\sigma(\cdot)$ denotes the $\sigma$-algebra generated by the indicated collection. Since $\mathcal{X}_s \subseteq \mathcal{X}_t$ for $s \le t$, the family $(\mathcal{F}_t)_{t=0}^{T}$ is indeed a filtration: $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \le t$.

## 2.2 Admissibility of the policy map

**Definition 2.3** (Policy artifact)**.** A *policy artifact* at time $t$ is a tuple $\text{StrategySpec}_t$ comprising target weights, constraints, rationale metadata, and execution parameters. The *policy map* is

$$\Phi_t \;:\; (\mathcal{F}_t,\, \mathcal{C}_t,\, w_{t-1}) \;\longmapsto\; \text{StrategySpec}_t, \tag{3}$$

where $\mathcal{C}_t$ is an $\mathcal{F}_t$-measurable feasible set encoding budget, leverage, turnover, and regulatory constraints.

**Lemma 2.4** (Filtration-measurability of emitted policies)**.** *Suppose every data-access function in the pipeline is* AS_OF$(t)$*-guarded, and all downstream computational modules consume only the outputs of guarded functions. Then the emitted* $\text{StrategySpec}_t$ *is* $\mathcal{F}_t$*-measurable.*

*Proof.* Let $A_t$ denote the event that all records accessed during the computation of $\text{StrategySpec}_t$ satisfy $t_{\text{pub}}(x) \le t$. By hypothesis, every data-access function $f$ satisfies $f(\cdot; t) \subseteq \mathcal{X}_t$. Since $\mathcal{X}_t$ is $\mathcal{F}_t$-measurable by construction (Definition 2.1), and compositions and pointwise operations of measurable functions are measurable, every intermediate quantity in the pipeline—$Q_t$, $P_t$, $\Delta_t$, $w_t^{\text{raw}}$, and the final projected $w_t$—is $\mathcal{F}_t$-measurable. In particular, $\mathbb{P}(A_t) = 1$ deterministically (it holds for every $\omega \in \Omega$ by the guard construction, not merely almost surely), so $\text{StrategySpec}_t$ is $\sigma(\mathcal{X}_0, \ldots, \mathcal{X}_t)$-measurable as claimed. □

*Remark* 2.5. This lemma is not a stylistic convention. It is the formal anti-leakage condition that makes replay verification—reconstructing past decisions from archived data and verifying they match historical outputs—logically meaningful. Without it, any observed backtest performance may be contaminated by future information, rendering statistical inference about system quality invalid (White, 2000; Hansen, 2005; López de Prado, 2018).

### 2.3 Constrained stochastic control formulation

At each decision time $t$, the system solves the constrained stochastic program:

$$w_t \;\in\; \operatorname*{arg\,min}_{w \in \mathcal{C}_t(\mathcal{F}_t)} \; \mathbb{E}\Big[ L_t\big(w;\, Q_t,\, P_t,\, \Delta_t,\, w_{t-1}\big) \;\Big|\; \mathcal{F}_t \Big], \tag{4}$$

where the feasible set $\mathcal{C}_t(\mathcal{F}_t)$ is $\mathcal{F}_t$-measurable and the loss function $L_t$ integrates contributions from market structure ($Q_t$), thesis beliefs ($P_t$), disagreement geometry ($\Delta_t$), and transaction costs (via $w_{t-1}$).

The measurable-selection requirement is central: no component of $L_t$ or $\mathcal{C}_t$ may depend on records with $t_{\text{pub}} > t$. This single constraint propagates through every module boundary in the architecture.

## 3 Market Prior $Q_t$: Option-Implied Density Recovery

The market prior $Q_t$ is a probability distribution over future asset states, extracted from observed option prices under the risk-neutral measure. The mathematical bridge is the Breeden–Litzenberger identity, which we derive from first principles.

### 3.1 Setup and risk-neutral pricing

We work on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$, where $\mathbb{Q}$ denotes a risk-neutral (equivalent martingale) measure. Under standard no-arbitrage conditions (Black and Scholes, 1973; Merton, 1973), the time-$t$ price of a European call option with strike $K$ and expiry $T = t + \tau$ is

$$C_t(K) \;=\; e^{-r\tau}\, \mathbb{E}_{\mathbb{Q}}\big[\, (S_T - K)^+ \;\big|\; \mathcal{F}_t \,\big], \tag{5}$$

where $S_T$ is the terminal asset price, $r$ is the continuously compounded risk-free rate, and $(x)^+ = \max(x, 0)$.

**Assumption 3.1.** The risk-neutral conditional distribution of $S_T$ given $\mathcal{F}_t$ admits a density $q_t : \mathbb{R}_+ \to \mathbb{R}_+$ with respect to Lebesgue measure, so that $\mathbb{Q}(S_T \in A \mid \mathcal{F}_t) = \int_A q_t(s)\, ds$ for Borel sets $A \subseteq \mathbb{R}_+$. Furthermore, $q_t$ is continuous and $\int_0^\infty s\, q_t(s)\, ds < \infty$.

### 3.2 The Breeden–Litzenberger identity

**Theorem 3.2** (Breeden–Litzenberger, 1978). *Under Assumption 3.1, the risk-neutral density is recovered from call prices by*

$$q_t(K) \;=\; e^{r\tau} \frac{\partial^2 C_t}{\partial K^2}(K) \tag{6}$$

*for all $K > 0$ at which $C_t(\cdot)$ is twice differentiable.*

*Proof.* We proceed in two differentiation steps.

**Step 1: First derivative.** Write the call payoff using the indicator representation:

$$(S_T - K)^+ \;=\; \int_K^\infty \mathbf{1}_{\{S_T>u\}}\, du. \tag{7}$$

To verify this identity, note that if $S_T \le K$ both sides equal zero, while if $S_T > K$ the right-hand side equals $\int_K^{S_T} 1\, du = S_T - K = (S_T - K)^+$.

Substituting (7) into (5) and applying Fubini's theorem (justified by the integrability condition in Assumption 3.1):

$$\begin{aligned} C_t(K) &= e^{-r\tau} \int_K^\infty \mathbb{E}_{\mathbb{Q}}\big[\mathbf{1}_{\{S_T>u\}} \mid \mathcal{F}_t\big]\, du \\ &= e^{-r\tau} \int_K^\infty \mathbb{Q}(S_T > u \mid \mathcal{F}_t)\, du \\ &= e^{-r\tau} \int_K^\infty \left( \int_u^\infty q_t(s)\, ds \right) du. \end{aligned} \tag{8}$$

Differentiating (8) with respect to $K$ using the Leibniz integral rule (the integrand is continuous by Assumption 3.1):

$$\frac{\partial C_t}{\partial K}(K) \;=\; -e^{-r\tau}\, \mathbb{Q}(S_T > K \mid \mathcal{F}_t) \;=\; -e^{-r\tau} \int_K^\infty q_t(s)\, ds. \tag{9}$$

This result has a direct economic interpretation: $-\partial C_t/\partial K$ is the discounted probability that the option finishes in the money.

**Step 2: Second derivative.** Differentiating (9) once more with respect to $K$:

$$\frac{\partial^2 C_t}{\partial K^2}(K) \;=\; -e^{-r\tau} \cdot \frac{d}{dK}\left( \int_K^\infty q_t(s)\, ds \right) \;=\; -e^{-r\tau} \cdot \big(-q_t(K)\big) \;=\; e^{-r\tau}\, q_t(K). \tag{10}$$

Rearranging gives the Breeden–Litzenberger identity:

$$q_t(K) = e^{r\tau}\, \frac{\partial^2 C_t}{\partial K^2}(K). \qquad \square$$

*Remark* 3.3 (Economic content). The identity states that the curvature of the call-price function in the strike dimension directly encodes the market's risk-neutral beliefs about future states. A higher second derivative at strike $K$ means the market assigns greater probability to the asset price landing near $K$. This is the mathematically rigorous bridge from traded prices to a probability distribution, and it is the foundation of the prior $Q_t$.

### 3.3 Discrete estimator on a finite strike grid

In practice, call prices are observed at finitely many strikes $K_1 < K_2 < \cdots < K_M$ with associated prices $C_1, C_2, \ldots, C_M$. We must approximate the second derivative on this nonuniform grid.

**Definition 3.4** (Nonuniform central second difference)**.** For interior point $i \in \{2, \ldots, M-1\}$, define the left and right slopes:

$$\text{slope}_i^L = \frac{C_i - C_{i-1}}{K_i - K_{i-1}}, \qquad \text{slope}_i^R = \frac{C_{i+1} - C_i}{K_{i+1} - K_i}, \tag{11}$$

and the central second difference:

$$\widehat{C}_i'' = \frac{2\big(\text{slope}_i^R - \text{slope}_i^L\big)}{K_{i+1} - K_{i-1}}. \tag{12}$$

The discrete density estimate is then:

$$\hat{f}_i = \max\!\big(e^{r\tau}\,\widehat{C}_i'',\ 0\big), \qquad \hat{p}_i = \frac{\max(\hat{f}_i \cdot \Delta K_i,\ \epsilon)}{\sum_{j=1}^{M} \max(\hat{f}_j \cdot \Delta K_j,\ \epsilon)}, \tag{13}$$

where $\Delta K_i$ is the width of the $i$-th bin and $\epsilon > 0$ is a small regularization constant ensuring full support.

### 3.4 Convex shape-projection for arbitrage-free smoothing

Raw market quotes may violate the no-arbitrage shape constraints (monotone decreasing calls, convex in strike). We enforce these via a convex quadratic pre-projection.

**Proposition 3.5** (Arbitrage-free call-price projection)**.** *Define the constrained least-squares problem:*

$$\min_{c \in \mathbb{R}^M} \frac{1}{2} \left\| c - C^{\text{obs}} \right\|_2^2 \tag{14}$$

*subject to the no-arbitrage constraints:*

$$c_{i+1} \le c_i, \qquad i = 1, \dots, M-1, \tag{15}$$

$$c_{i-1} - 2c_i + c_{i+1} \ge 0, \qquad i = 2, \dots, M-1. \tag{16}$$

*This is a convex QP with a strictly convex objective and linear constraints, hence it admits a unique global minimizer $c^*$.*

*Proof.* The objective $f(c) = \frac{1}{2}\left\| c - C^{\text{obs}} \right\|_2^2$ has Hessian $\nabla^2 f = I_M$ (the $M \times M$ identity matrix), which is positive definite. The constraints (15)–(16) are affine in $c$, so the feasible set is a polyhedron. Since the objective is strictly convex and coercive ($f(c) \to \infty$ as $\|c\| \to \infty$), a unique global minimizer exists provided the feasible set is nonempty. The feasible set is nonempty because, for instance, the constant function $c_i = \max_j C_j^{\text{obs}}$ for all $i$ satisfies both monotonicity and convexity (with equality throughout). □

### 3.5 Noise-bias trade-off in finite-difference inversion

The central second difference (12) involves a bias-variance trade-off that is fundamental to the density recovery problem.

**Proposition 3.6** (Noise amplification in second-difference estimation)**.** *Let the observed call prices be $C_i^{\text{obs}} = C_i^{\text{true}} + \varepsilon_i$, where $\varepsilon_i$ are independent noise terms with $\mathbb{E}[\varepsilon_i] = 0$ and $\text{Var}(\varepsilon_i) = \sigma_\mu^2$*

*(microstructure noise). For a locally uniform grid with spacing $h$, the bias and variance of $\widehat{C}_i''$ satisfy:*

$$\mathrm{Bias}(\widehat{C}_i'') = O(h^2), \qquad \mathrm{Var}(\widehat{C}_i'') = O\left(\frac{\sigma_\mu^2}{h^4}\right). \tag{17}$$

*Proof.* For the bias, Taylor-expand $C^{\mathrm{true}}$ around $K_i$ on a uniform grid:

$$\begin{aligned} C_{i+1}^{\mathrm{true}} &= C_i^{\mathrm{true}} + h\,C_i' + \tfrac{h^2}{2}\,C_i'' + \tfrac{h^3}{6}\,C_i''' + \tfrac{h^4}{24}\,C_i^{(4)} + O(h^5), \\ C_{i-1}^{\mathrm{true}} &= C_i^{\mathrm{true}} - h\,C_i' + \tfrac{h^2}{2}\,C_i'' - \tfrac{h^3}{6}\,C_i''' + \tfrac{h^4}{24}\,C_i^{(4)} + O(h^5). \end{aligned}$$

Summing: $C_{i+1}^{\mathrm{true}} + C_{i-1}^{\mathrm{true}} - 2C_i^{\mathrm{true}} = h^2 C_i'' + \frac{h^4}{12} C_i^{(4)} + O(h^6)$. Dividing by $h^2$:

$$\frac{C_{i+1}^{\mathrm{true}} + C_{i-1}^{\mathrm{true}} - 2C_i^{\mathrm{true}}}{h^2} = C_i'' + \frac{h^2}{12}\,C_i^{(4)} + O(h^4),$$

giving Bias $= O(h^2)$.

For the variance, the estimator is a linear combination of three noisy observations with coefficients $1/h^2, -2/h^2, 1/h^2$:

$$\mathrm{Var}(\widehat{C}_i'') = \frac{1}{h^4}\left(\sigma_\mu^2 + 4\sigma_\mu^2 + \sigma_\mu^2\right) = \frac{6\,\sigma_\mu^2}{h^4}. \qquad \square$$

*Remark* 3.7. This creates the fundamental inversion tension: coarse grids ($h$ large) reduce variance but increase bias, while fine grids ($h$ small) reduce bias but amplify microstructure noise. The convex shape projection of Section 3.4 acts as an estimator regularizer: it enforces economically valid curvature (nonnegative second derivatives) while damping high-frequency noise, analogous to constrained nonparametric regression (Aït-Sahalia and Lo, 1998; Fengler, 2009).

### 3.6 Marginals versus joint law for multiple assets

For a universe of $N > 1$ assets, single-name option surfaces identify only the marginal risk-neutral laws $Q_t^{(i)}$ for each asset $i$, not the joint law $Q_t^{\mathrm{joint}}$. Formally:

**Definition 3.8** (Joint law via coupling)**.** A *coupling* $\eta_t$ is a probability measure on $\mathbb{R}_+^N$ whose $i$-th marginal equals $Q_t^{(i)}$ for each $i = 1, \ldots, N$. The joint prior is:

$$Q_t^{\mathrm{joint}} = \mathrm{Compose}\big(Q_t^{(1)}, \ldots, Q_t^{(N)};\ \eta_t\big). \tag{18}$$

Dependence assumptions encoded in $\eta_t$ materially alter tail co-movement, joint risk measures (VaR, CVaR), and robust allocations. Without explicit dependence modeling, portfolio optimization implicitly assumes independence or uses historical correlation estimates, both of which can be severely misleading during stress episodes (Bliss and Panigirtzoglou, 2004).

# 4 Thesis Posterior $P_t$: Structured Belief Operators on the Simplex

While $Q_t$ reflects the market's collective pricing, the thesis posterior $P_t$ encodes the agent's subjective beliefs about future states, constructed from macroeconomic analysis, textual evidence, and structured knowledge. The key mathematical requirement is that all transformations preserve probabilistic validity.

## 4.1 The probability simplex

**Definition 4.1** (Probability simplex). For $B \geq 2$ discrete outcome bins, the $(B-1)$-*dimensional probability simplex* is

$$\Delta^{B-1} \;=\; \left\{ p \in \mathbb{R}^B \;:\; p_j \geq 0 \text{ for all } j,\; \sum_{j=1}^{B} p_j = 1 \right\}. \tag{19}$$

Its *relative interior* is $\mathrm{ri}(\Delta^{B-1}) = \{p \in \Delta^{B-1} : p_j > 0 \text{ for all } j\}$.

In the implemented system, $B = 3$ corresponding to outcomes $\{\mathrm{down}, \mathrm{flat}, \mathrm{up}\}$, but all results hold for general $B$.

## 4.2 Multiplicative tilting operators

The posterior is formed by applying a sequence of belief-shaping operators to a base distribution.

**Definition 4.2** (Multiplicative tilting operator). Given a strictly positive *modifier vector* $m \in \mathbb{R}^B_{++}$ (each component strictly positive), define the operator $T_m : \Delta^{B-1} \to \Delta^{B-1}$ by

$$T_m(p) \;=\; \mathrm{Normalize}\big(p \odot m\big), \qquad \text{where} \quad \big[\mathrm{Normalize}(v)\big]_j = \frac{v_j}{\sum_{k=1}^{B} v_k}, \tag{20}$$

and $\odot$ denotes the Hadamard (elementwise) product.

**Proposition 4.3** (Simplex invariance). *If $p \in \Delta^{B-1}$ and $m \in \mathbb{R}^B_{++}$, then $T_m(p) \in \Delta^{B-1}$. Moreover, if $p \in \mathrm{ri}(\Delta^{B-1})$, then $T_m(p) \in \mathrm{ri}(\Delta^{B-1})$.*

*Proof.* Set $v_j = p_j \cdot m_j$. Since $p_j \geq 0$ and $m_j > 0$, we have $v_j \geq 0$. The normalizing constant is $Z = \sum_{k=1}^{B} v_k = \sum_{k=1}^{B} p_k m_k > 0$ (since $p \in \Delta^{B-1}$ implies at least one $p_k > 0$, and all $m_k > 0$). Therefore $[T_m(p)]_j = v_j/Z \geq 0$ and $\sum_j v_j/Z = 1$. If additionally $p_j > 0$ for all $j$, then $v_j = p_j m_j > 0$ for all $j$, so $T_m(p)$ lies in the relative interior. □

**Proposition 4.4** (Local Lipschitz continuity of tilting). *For bounded modifier vectors satisfying $0 < m_{\min} \leq m_j \leq m_{\max}$ for all $j$, the map $p \mapsto T_m(p)$ is locally Lipschitz on the relative interior of $\Delta^{B-1}$. Specifically, for any compact subset $K \subset \mathrm{ri}(\Delta^{B-1})$, there exists $L = L(K, m_{\min}, m_{\max}) < \infty$ such that*

$$\big\|T_m(p) - T_m(p')\big\|_1 \;\leq\; L\,\big\|p - p'\big\|_1 \qquad \textit{for all } p, p' \in K. \tag{21}$$

*Proof.* Write $T_m(p) = g(p)/h(p)$ componentwise, where $g_j(p) = p_j m_j$ and $h(p) = \sum_k p_k m_k$. On $K \subset \mathrm{ri}(\Delta^{B-1})$, we have $h(p) \geq m_{\min} \cdot \min_{p \in K} \min_j p_j > 0$. Both $g_j$ and $h$ are affine (hence smooth) in $p$, and the quotient of smooth functions with bounded-away-from-zero denominator is Lipschitz on compact sets. The Lipschitz constant can be bounded explicitly via $\|\nabla(g_j/h)\|_\infty$ over $K$, which involves $m_{\max}$, $m_{\min}$, and the minimum component value of $p$ on $K$. $\square$

*Remark* 4.5. This Lipschitz property matters for bilevel training (Section 8): it bounds the gradient magnitudes transmitted from the decision layer back into the posterior parameters, preventing gradient explosion during end-to-end optimization.

## 4.3 Exponential-family interpretation

The multiplicative tilting admits an equivalent exponential-family view. Consider a one-parameter family indexed by intensity $\theta \in \mathbb{R}$:

$$p'_j(\theta) \;=\; \frac{p_j \cdot \exp\bigl(\theta \cdot \mathrm{sgn}(x_j)\bigr)}{Z(\theta)}, \qquad Z(\theta) = \sum_{k=1}^{B} p_k \cdot \exp\bigl(\theta \cdot \mathrm{sgn}(x_k)\bigr), \tag{22}$$

where $x_j$ is a signed feature associated with bin $j$ (e.g., the bin's midpoint return) and $\mathrm{sgn}(\cdot)$ is the signum function. The modifier $m_j(\theta) = \exp(\theta \cdot \mathrm{sgn}(x_j))$ is strictly positive for all $\theta$.

*Remark* 4.6. The implemented system uses linearized multipliers of the form $m_j = 1 + \theta \cdot \mathrm{sgn}(x_j)$ (clamped to ensure positivity). This is the first-order Taylor approximation of the exponential form: $\exp(\theta \cdot \mathrm{sgn}(x_j)) \approx 1 + \theta \cdot \mathrm{sgn}(x_j)$ for $|\theta|$ small. The two forms coincide to first order near zero tilt, and the linear version avoids numerical overflow for large $\theta$.

## 4.4 Channel fusion for multi-source beliefs

The thesis posterior integrates multiple evidence channels: macroeconomic indicators, retrieval-augmented generation (RAG) from textual sources, and knowledge-graph (KG) structural signals.

**Definition 4.7** (Channel fusion)**.** Let $z_t^{(k)} \in \mathbb{R}^d$ be the feature vector from channel $k \in \{1, \ldots, K\}$ at time $t$, and let $\alpha_k \geq 0$ with $\sum_k \alpha_k = 1$ be channel weights. The *fused representation* is

$$z_t \;=\; \sum_{k=1}^{K} \alpha_k \, z_t^{(k)}. \tag{23}$$

The fused representation is then mapped to a modifier vector $m(z_t) \in \mathbb{R}^B_{++}$ via a strictly positive link function (e.g., softplus or exponentiation), and the posterior is

$$P_t \;=\; T_{m(z_t)}(P_t^{\mathrm{base}}), \tag{24}$$

where $P_t^{\mathrm{base}}$ is a base distribution (e.g., uniform or derived from macro scores).

## 4.5 Calibration and strict propriety of the log score

The quality of $P_t$ as a probabilistic forecast is measured by a *scoring rule.*

**Definition 4.8** (Scoring rule and strict propriety)**.** A *scoring rule* $S : \Delta^{B-1} \times \{1, \ldots, B\} \to \mathbb{R}$ maps a reported distribution $q$ and a realized outcome $j$ to a reward $S(q, j)$. The *expected score* under true distribution $p$ is $\bar{S}(q; p) = \sum_{j=1}^{B} p_j \, S(q, j)$. The rule is *strictly proper* if $\bar{S}(q; p) < \bar{S}(p; p)$ for all $q \neq p$ in $\Delta^{B-1}$.

**Theorem 4.9** (Strict propriety of the logarithmic score)**.** *The* logarithmic scoring rule $S_{\log}(q, j) = \log q_j$ *is strictly proper. Moreover, the gap between truthful and misreported expected scores equals the Kullback–Leibler divergence:*

$$\bar{S}_{\log}(p; p) - \bar{S}_{\log}(q; p) \;=\; \mathrm{KL}(p \| q). \tag{25}$$

*Proof.* **Step 1: Expected score computation.** For the binary case ($B = 2$, with true parameter $p = p_1$ and $p_2 = 1 - p$), the expected log score of report $q = (q_1, q_2)$ is:

$$\bar{S}_{\log}(q; p) = p \, \log q_1 + (1 - p) \, \log q_2 = p \, \log q_1 + (1 - p) \, \log(1 - q_1). \tag{26}$$

For general $B$: $\bar{S}_{\log}(q; p) = \sum_{j=1}^{B} p_j \log q_j$.

**Step 2: Optimality at truthful report.** Consider maximizing $\bar{S}_{\log}(q; p)$ over $q \in \Delta^{B-1}$. This is the problem:

$$\max_{q \in \Delta^{B-1}} \sum_{j=1}^{B} p_j \log q_j.$$

Form the Lagrangian with multiplier $\lambda$ for the constraint $\sum_j q_j = 1$:

$$\mathcal{L}(q, \lambda) = \sum_{j=1}^{B} p_j \log q_j - \lambda \left( \sum_{j=1}^{B} q_j - 1 \right).$$

First-order conditions: $\partial \mathcal{L} / \partial q_j = p_j / q_j - \lambda = 0$, giving $q_j = p_j / \lambda$. Substituting into the constraint: $\sum_j p_j / \lambda = 1$, so $\lambda = 1$ (since $\sum_j p_j = 1$). Therefore $q_j^* = p_j$.

**Step 3: Strict maximum (second-order condition).** The Hessian of $\bar{S}_{\log}$ with respect to $q$ is $\nabla^2_{qq} \bar{S}_{\log} = -\mathrm{diag}(p_1/q_1^2, \ldots, p_B/q_B^2)$, which is negative definite on the interior of the simplex. Hence $q^* = p$ is the unique global maximizer.

**Step 4: Gap identity.** The score gap is:

$$\begin{aligned} \bar{S}_{\log}(p; p) - \bar{S}_{\log}(q; p) &= \sum_{j=1}^{B} p_j \log p_j - \sum_{j=1}^{B} p_j \log q_j \\ &= \sum_{j=1}^{B} p_j \log \frac{p_j}{q_j} \;=\; \mathrm{KL}(p \| q). \end{aligned}$$

□

*Remark* 4.10 (Architectural significance). This theorem links three conceptually distinct components of the architecture: (i) the forecast calibration quality of $P_t$, (ii) the reward shaping in the reinforcement learning layer (Section 9), and (iii) the divergence diagnostics $\Delta_t$ (Section 5). Specifically, incentivizing accurate probability reports via a log-score reward is mathematically equivalent to minimizing KL divergence from the truth.

# 5 Divergence Geometry $\Delta_t$

The divergence object $\Delta_t = D(P_t \| Q_t)$ quantifies the geometric disagreement between the thesis posterior and the market prior. We define a composite functional that captures complementary aspects of distributional discrepancy.

## 5.1 Kullback–Leibler divergence

**Definition 5.1** (KL divergence for discrete distributions)**.** For distributions $p, q \in \mathrm{ri}(\Delta^{B-1})$ (both with full support), the *Kullback–Leibler divergence* is

$$\mathrm{KL}(p \| q) \;=\; \sum_{j=1}^{B} p_j \log \frac{p_j}{q_j}. \tag{27}$$

**Proposition 5.2** (Properties of KL divergence)**.** *The KL divergence satisfies:*

*(a)* ***Non-negativity (Gibbs' inequality):*** $\mathrm{KL}(p\|q) \geq 0$, *with equality if and only if* $p = q$.

*(b)* ***Asymmetry:*** $\mathrm{KL}(p\|q) \neq \mathrm{KL}(q\|p)$ *in general.*

*(c)* ***Convexity:*** $p \mapsto \mathrm{KL}(p\|q)$ *is strictly convex for fixed* $q$.

*Proof of (a).* Apply Jensen's inequality to the concave function $\log(\cdot)$:

$$-\mathrm{KL}(p\|q) = \sum_j p_j \log \frac{q_j}{p_j} \leq \log\left(\sum_j p_j \cdot \frac{q_j}{p_j}\right) = \log\left(\sum_j q_j\right) = \log 1 = 0.$$

Equality holds if and only if $q_j/p_j$ is constant for all $j$ with $p_j > 0$, which combined with $\sum_j p_j = \sum_j q_j = 1$ forces $p = q$. □

## 5.2 Wasserstein-1 distance

KL divergence is insensitive to the *geometry* of the outcome space: it treats bins as unordered categories. For ordered outcomes (e.g., return bins), we supplement KL with a transport-based metric.

**Definition 5.3** (Wasserstein-1 distance)**.** Let $c : \{1, \ldots, B\} \times \{1, \ldots, B\} \to \mathbb{R}_+$ be a ground cost, typically $c(i,j) = |x_i - x_j|$ where $x_i$ is the midpoint of bin $i$. Let $\Pi(p,q)$ denote the set of all couplings (joint distributions on $\{1, \ldots, B\}^2$ with marginals $p$ and $q$). The *Wasserstein-1 distance* is

$$W_1(p,q) \;=\; \inf_{\gamma \in \Pi(p,q)} \sum_{i,j} \gamma_{ij}\, c(i,j). \tag{28}$$

**Proposition 5.4** (CDF representation of $W_1$ for ordered bins)**.** *When the ground cost is* $c(i,j) = |x_i - x_j|$ *with* $x_1 < x_2 < \cdots < x_B$*, the Wasserstein-1 distance admits the closed-form representation:*

$$W_1(p,q) \;=\; \sum_{j=1}^{B-1} |F_p(j) - F_q(j)| \cdot (x_{j+1} - x_j), \tag{29}$$

*where* $F_p(j) = \sum_{k=1}^{j} p_k$ *and* $F_q(j) = \sum_{k=1}^{j} q_k$ *are cumulative distribution functions.*

*Proof.* This is the discrete analogue of the Kantorovich–Rubinstein duality $W_1(\mu, \nu) = \int |F_\mu(x) - F_\nu(x)|\, dx$ for measures on $\mathbb{R}$, applied to the piecewise-constant CDFs on the ordered grid. The optimal coupling transports mass monotonically (from left to right or right to left) between the two CDFs, and the total cost is the $L^1$ distance between CDFs weighted by bin widths. ☐

*Remark* 5.5. Wasserstein distance captures location/spread shifts that KL can understate. For example, if $p$ and $q$ have the same shape but are translated by one bin, KL may be moderate (depending on overlap), while $W_1$ directly measures the displacement. Conversely, KL is more sensitive to multiplicative changes in bin probabilities. This complementarity motivates using both.

## 5.3 Composite divergence functional

For each asset $i = 1, \ldots, N$, we define marginal divergence components. We also define a cross-asset dependence divergence.

**Definition 5.6** (Composite divergence)**.** The divergence object is the weighted composite:

$$\Delta_t \;=\; \underbrace{a \cdot \sum_{i=1}^{N} \mathrm{KL}\big(P_t^{(i)} \| Q_t^{(i)}\big)}_{\text{directional belief shift}} \;+\; b \cdot \underbrace{\|\mathrm{Corr}_{P_t} - \mathrm{Corr}_{Q_t}\|_F}_{\text{dependence mismatch}} \;+\; \underbrace{c \cdot \sum_{i=1}^{N} \mathrm{Tail}_i(P_t, Q_t)}_{\text{tail disagreement}}, \tag{30}$$

where $a, b, c > 0$ are fixed weights, $\|\cdot\|_F$ is the Frobenius norm, and

$$\mathrm{Tail}_i(P, Q) \;=\; \left| L_i^P - L_i^Q \right| + \left| R_i^P - R_i^Q \right|, \tag{31}$$

with $L_i^P, R_i^P$ denoting left-tail and right-tail probability masses under $P$ for asset $i$, and similarly for $Q$.

*Remark* 5.7 (Why three components)*.* No single divergence captures all economically relevant discrepancies in finite samples:

(i) **KL** responds strongly to interior-bin miscalibration (relative likelihood ratios).

(ii) **Wasserstein/tail terms** respond to location and shape transport, especially for extreme outcomes.

(iii) **Dependence mismatch** responds to joint-structure misspecification, which marginal divergences cannot detect.

This tri-part decomposition prevents any one failure mode from dominating the signal.

## 5.4 Gradient structure for differentiable integration

For downstream optimization and bilevel training (Section 8), we need the gradients of $\Delta_t$ with respect to belief parameters.

**Proposition 5.8** (Gradients of the composite divergence). *For strictly positive bins ($p_{i,j} > 0$ and $q_{i,j} > 0$ for all $i, j$):*

*(a)* ***KL gradient:***

$$\frac{\partial}{\partial p_{i,j}} \mathrm{KL}\big(P_t^{(i)} \| Q_t^{(i)}\big) \;=\; \log \frac{p_{i,j}}{q_{i,j}} + 1. \tag{32}$$

*(b)* ***Dependence-mismatch gradient (Frobenius):***

$$\frac{\partial}{\partial \mathrm{Corr}_P} \, \|\mathrm{Corr}_P - \mathrm{Corr}_Q\|_F^2 \;=\; 2\big(\mathrm{Corr}_P - \mathrm{Corr}_Q\big). \tag{33}$$

*(c)* ***Tail and Wasserstein components*** *are piecewise linear in cumulative probabilities, hence their gradients are piecewise constant (subgradients at non-differentiable points).*

*Proof.* Part (a): $\mathrm{KL} = \sum_j p_j \log(p_j/q_j)$. Differentiating with respect to $p_j$ (treating $q$ as fixed):

$$\frac{\partial}{\partial p_j}\left(p_j \log \frac{p_j}{q_j}\right) = \log \frac{p_j}{q_j} + p_j \cdot \frac{1}{p_j} = \log \frac{p_j}{q_j} + 1.$$

Part (b) follows from $\frac{d}{dA} \|A - B\|_F^2 = 2(A - B)$ for matrices.

Part (c): The $W_1$ representation (29) involves $|F_p(j) - F_q(j)|$, which is piecewise linear in $F_p(j)$, and $F_p(j) = \sum_{k \le j} p_k$ is linear in $p$. The composition of a piecewise-linear function with a linear function is piecewise linear, with subgradients at kink points (where $F_p(j) = F_q(j)$). ☐

# 6 Strategy Compilation as Convex Projection

The compiler transforms the raw signal $(w_t^{\mathrm{raw}}, w_t^{\mathrm{prev}}, \mathcal{C}_t)$ into a feasible portfolio $w_t$ via a constrained convex optimization. This is the decision-making core of the architecture.

## 6.1 Problem formulation

**Definition 6.1** (Projection compiler). The *projection compiler* solves the convex program:

$$\min_{w \in \mathbb{R}^N} \; \underbrace{\frac{1}{2}\,\|w - w^{\mathrm{raw}}\|_2^2}_{\text{tracking}} \;+\; \underbrace{\lambda_{\mathrm{tc}}\,\|w - w^{\mathrm{prev}}\|_1}_{\text{transaction cost}} \tag{34}$$

subject to the constraints:

$$\begin{aligned}
&\text{Budget:} && \mathbf{1}^\top w = \beta, && && (35)\\
&\text{Box bounds:} && \ell_i \le w_i \le u_i, && i = 1, \dots, N, && (36)\\
&\text{Turnover:} && \|w - w^{\mathrm{prev}}\|_1 \le \tau_{\max}, && && (37)\\
&\text{Sector caps:} && S_k^\top w \le s_k^{\max}, && k = 1, \dots, K_s, && (38)\\
&\text{Leverage:} && \|w\|_1 \le L_{\max}, && && (39)\\
&\text{Auxiliary:} && A_{\mathrm{aux}}\, w \le b_{\mathrm{aux}}, && && (40)
\end{aligned}$$

where $\beta$ is the target budget (typically 1), $\lambda_{\text{tc}} \geq 0$ is the transaction-cost penalty, $\tau_{\max}$ is the maximum turnover, $S_k$ are sector-membership indicator vectors, and $A_{\text{aux}}, b_{\text{aux}}$ encode additional linear constraints from persona or governance rules.

## 6.2 Linearization of $L^1$ terms

The $L^1$-norm terms in (34)–(37)–(39) are not smooth but can be exactly linearized.

**Proposition 6.2** ($L^1$ linearization)**.** *The optimization problem* (34) *is equivalent to a convex QP after introducing auxiliary variables. For the turnover term, define* $d_i^+, d_i^- \geq 0$ *with* $w_i - w_i^{\text{prev}} = d_i^+ - d_i^-$*. Then:*

$$\|w - w^{\text{prev}}\|_1 = \sum_{i=1}^{N} (d_i^+ + d_i^-), \qquad d_i^+, d_i^- \geq 0. \tag{41}$$

*The leverage constraint* $\|w\|_1 \leq L_{\max}$ *is handled analogously by introducing* $w_i = w_i^+ - w_i^-$ *with* $w_i^+, w_i^- \geq 0$*.*

*Proof.* For any real $x$, we have $|x| = \min_{d^+, d^- \geq 0}\{d^+ + d^- : x = d^+ - d^-\}$. At the optimum of the minimization, either $d^+ = 0$ or $d^- = 0$ (but not both when $x \neq 0$), so $d^+ + d^- = |x|$. This is because adding equal amounts to both $d^+$ and $d^-$ would increase the objective while maintaining the equality constraint, which contradicts optimality. $\square$

After this transformation, the full problem becomes:

$$\min_{w, d^+, d^-} \; \frac{1}{2}(w - w^{\text{raw}})^\top (w - w^{\text{raw}}) + \lambda_{\text{tc}} \mathbf{1}^\top (d^+ + d^-) \tag{42}$$

subject to $w - w^{\text{prev}} = d^+ - d^-$, $d^+ \geq 0$, $d^- \geq 0$, and all constraints (35)–(40). This is a convex QP with a positive definite Hessian (in the $w$-block) and a polyhedral feasible set.

## 6.3 Existence and uniqueness

**Theorem 6.3** (Well-posedness of the compiler)**.** *If the feasible set defined by constraints* (35)–(40) *is nonempty, then the compiler* (42) *has a unique optimal solution* $w^*$*.*

*Proof.* The objective function $f(w, d^+, d^-) = \frac{1}{2}\|w - w^{\text{raw}}\|_2^2 + \lambda_{\text{tc}} \mathbf{1}^\top (d^+ + d^-)$ is continuous, convex, and coercive (it tends to $+\infty$ as $\|(w, d^+, d^-)\| \to \infty$). The feasible set is a closed polyhedron (intersection of finitely many closed half-spaces and hyperplanes). By the Weierstrass extreme value theorem, a minimizer exists.

For uniqueness in $w$: the Hessian of $f$ with respect to $w$ is $I_N$ (positive definite), so $f$ is strictly convex in $w$. If $(w_1, d_1^+, d_1^-)$ and $(w_2, d_2^+, d_2^-)$ are both optimal with $w_1 \neq w_2$, then by strict convexity of the $w$-component, their midpoint achieves a strictly lower objective value, contradicting optimality. $\square$

## 6.4 KKT conditions and economic interpretation

The Karush–Kuhn–Tucker (KKT) conditions provide both the optimality certificate and an economic interpretation of the solution.

**Definition 6.4** (KKT conditions)**.** The KKT conditions for a convex program $\min_x f(x)$ subject to $g_i(x) \le 0$ $(i = 1, \dots, m)$ and $h_j(x) = 0$ $(j = 1, \dots, p)$ consist of four requirements:

(i) **Stationarity:** $\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{p} \nu_j \nabla h_j(x^*) = 0$.

(ii) **Primal feasibility:** $g_i(x^*) \le 0$ and $h_j(x^*) = 0$ for all $i, j$.

(iii) **Dual feasibility:** $\mu_i \ge 0$ for all $i$.

(iv) **Complementary slackness:** $\mu_i \cdot g_i(x^*) = 0$ for all $i$.

Here $\mu_i$ are dual variables (Lagrange multipliers) for inequality constraints and $\nu_j$ for equality constraints.

**Theorem 6.5** (KKT sufficiency for convex programs)**.** *For a convex program with convex $f$, convex $g_i$, and affine $h_j$, any point $(x^*, \mu^*, \nu^*)$ satisfying the KKT conditions is a global optimum. If additionally $f$ is strictly convex, $x^*$ is the unique global optimum.*

*Proof.* By stationarity, $x^*$ minimizes the Lagrangian $\mathcal{L}(x, \mu^*, \nu^*) = f(x) + \sum_i \mu_i^* g_i(x) + \sum_j \nu_j^* h_j(x)$ over $x$ (since $\mathcal{L}$ is convex in $x$ and its gradient vanishes at $x^*$). For any feasible $x$: $g_i(x) \le 0$ and $h_j(x) = 0$, so $\mathcal{L}(x, \mu^*, \nu^*) \le f(x)$ (using $\mu_i^* \ge 0$). By complementary slackness, $\mathcal{L}(x^*, \mu^*, \nu^*) = f(x^*)$. Therefore $f(x^*) = \mathcal{L}(x^*, \mu^*, \nu^*) \le \mathcal{L}(x, \mu^*, \nu^*) \le f(x)$ for all feasible $x$. $\square$

For the compiler (42), the stationarity condition yields:

$$(w^* - w^{\text{raw}}) + \nu_{\text{budget}} \mathbf{1} + \sum_i \mu_i^{\text{box}} e_i + \mu_{\text{turn}} \partial \left\| w^* - w^{\text{prev}} \right\|_1 + \sum_k \mu_k^{\text{sector}} S_k + A_{\text{aux}}^\top \mu_{\text{aux}} = 0, \tag{43}$$

where $e_i$ are standard basis vectors and $\partial \left\| \cdot \right\|_1$ denotes the subdifferential of the $L^1$ norm.

*Remark* 6.6 (Economic shadow prices)*.* Each dual variable has an economic interpretation:

- $\nu_{\text{budget}}$: marginal cost of the budget constraint (capital scarcity).
- $\mu_i^{\text{box}}$: marginal cost of position limits for asset $i$.
- $\mu_{\text{turn}}$: marginal cost of turnover capacity.
- $\mu_k^{\text{sector}}$: marginal cost of sector concentration limits.
- $\mu_{\text{aux}}$: marginal costs of governance/persona constraints.

A large dual variable indicates a binding constraint that materially shapes the allocation, enabling model-risk dialogue: analysts can identify which constraints most affect the policy.

## 7 Distributionally Robust Extensions

The compiler of Section 6 assumes a point estimate for expected returns (the "edge" vector embedded in $w^{\text{raw}}$). In practice, this estimate is uncertain. We formalize robust counterparts that remain convex and tractable.

### 7.1 Uncertainty-set robust optimization

**Definition 7.1** (Robust linear objective)**.** Given an uncertain edge vector $e \in \mathcal{U}$ and a baseline estimate $\hat{e}$, the *robust linear objective* is

$$\max_{w} \min_{e \in \mathcal{U}} e^\top w - \frac{\beta_r}{2} w^\top \Sigma w - \gamma \left\| w - w^{\text{prev}} \right\|_1 . \tag{44}$$

The inner minimization has a closed form depending on the geometry of $\mathcal{U}$.

**Proposition 7.2** (Support-function identities)**.** *Let* $\sigma_{\mathcal{U}}(w) = \sup_{e \in \mathcal{U}} e^\top w$ *denote the support function of* $\mathcal{U}$*. Then* $\min_{e \in \mathcal{U}} e^\top w = \hat{e}^\top w - \sigma_{\mathcal{U} - \hat{e}}(w)$ *where* $\mathcal{U} - \hat{e} = \{e - \hat{e} : e \in \mathcal{U}\}$*.*

*(a)* ***Box uncertainty*** *(*$\mathcal{U} = \{e : \|e - \hat{e}\|_\infty \leq \kappa\}$*):*

$$\min_{e \in \mathcal{U}} e^\top w = \hat{e}^\top w - \kappa \|w\|_1 . \tag{45}$$

*(b)* ***Ellipsoidal uncertainty*** *(*$\mathcal{U} = \{e : (e - \hat{e})^\top M^{-1} (e - \hat{e}) \leq \kappa^2\}$ *with* $M \succ 0$*):*

$$\min_{e \in \mathcal{U}} e^\top w = \hat{e}^\top w - \kappa \left\| M^{1/2} w \right\|_2 . \tag{46}$$

*Proof.* We derive each case using support-function duality.

**(a) Box case.** Write $e = \hat{e} + \delta$ with $\|\delta\|_\infty \leq \kappa$. Then:

$$\min_{\|\delta\|_\infty \leq \kappa} (\hat{e} + \delta)^\top w = \hat{e}^\top w + \min_{\|\delta\|_\infty \leq \kappa} \delta^\top w = \hat{e}^\top w - \max_{\|\delta\|_\infty \leq \kappa} (-\delta)^\top w.$$

Now $\max_{\|\delta\|_\infty \leq \kappa} \delta^\top w = \kappa \|w\|_1$ by the duality $\|\cdot\|_1 = (\|\cdot\|_\infty)^*$, i.e., the dual norm of $\ell^\infty$ is $\ell^1$. To verify directly: for each component, the maximizer sets $\delta_i = \kappa \cdot \text{sgn}(w_i)$, achieving $\sum_i \kappa |w_i| = \kappa \|w\|_1$. By symmetry ($-\delta$ ranges over the same set as $\delta$), $\min_{\|\delta\|_\infty \leq \kappa} \delta^\top w = -\kappa \|w\|_1$.

**(b) Ellipsoidal case.** Write $e = \hat{e} + \delta$ with $\delta^\top M^{-1} \delta \leq \kappa^2$. Substituting $\delta = M^{1/2} u$ with $\|u\|_2 \leq \kappa$:

$$\min_{\delta^\top M^{-1} \delta \leq \kappa^2} \delta^\top w = \min_{\|u\|_2 \leq \kappa} (M^{1/2} u)^\top w = \min_{\|u\|_2 \leq \kappa} u^\top M^{1/2} w.$$

By Cauchy–Schwarz, $\min_{\|u\|_2 \leq \kappa} u^\top v = -\kappa \|v\|_2$ (achieved at $u = -\kappa v / \|v\|_2$ when $v \neq 0$). Setting $v = M^{1/2} w$ gives the result. ☐

*Remark* 7.3. Both robust counterparts preserve convexity: subtracting $\kappa \|w\|_1$ (convex in $w$) from the linear objective makes the overall objective concave in $w$, and maximizing a concave function over a convex set is a convex problem. This means the robust compiler remains solvable by standard QP/SOCP solvers (Ben-Tal and Nemirovski, 1998; Goldfarb and Iyengar, 2003; Bertsimas and Sim, 2004).

### 7.2 Wasserstein distributionally robust optimization

A more principled approach models distributional ambiguity directly. Rather than perturbing the edge vector, we consider the worst-case expectation over a ball of distributions centered at the empirical measure.

**Definition 7.4** (Wasserstein ambiguity set)**.** Given empirical observations $r_1, \ldots, r_n \in \mathbb{R}^N$ (historical return vectors) with empirical measure $\hat{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{r_i}$, and a transport cost $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+$, the *Wasserstein ambiguity set* of radius $\rho > 0$ is

$$\mathcal{U}_\rho \;=\; \big\{ P \;:\; W_c(P, \hat{P}_n) \le \rho \big\}, \tag{47}$$

where $W_c(P, P')$ denotes the optimal transport cost between $P$ and $P'$ under cost function $c$.

**Definition 7.5** (Distributionally robust expectation)**.** The *worst-case expected loss* under the ambiguity set $\mathcal{U}_\rho$ is

$$\Phi(w) \;=\; \sup_{P \in \mathcal{U}_\rho} \mathbb{E}_P\big[\ell(r; w)\big], \tag{48}$$

where $\ell(r; w)$ is a loss function (e.g., negative portfolio return, variance contribution).

The key to tractability is Kantorovich duality, which transforms the infinite-dimensional worst-case over distributions into a finite-dimensional optimization.

**Theorem 7.6** (Kantorovich dual of Wasserstein DRO)**.** *Under mild regularity conditions (continuous loss $\ell$, compact support or sufficient integrability), the worst-case expectation* (48) *admits the dual representation:*

$$\Phi(w) \;=\; \inf_{\lambda \ge 0} \left\{ \lambda\rho \;+\; \frac{1}{n} \sum_{i=1}^{n} \sup_{r \in \mathbb{R}^N} \big[\, \ell(r; w) - \lambda\, c(r, r_i) \,\big] \right\}. \tag{49}$$

*Proof.* We derive this in four steps.

**Step 1: Primal formulation as transport problem.** The worst-case expectation is:

$$\Phi(w) = \sup_P \left\{ \int \ell(r; w)\, dP(r) \;:\; W_c(P, \hat{P}_n) \le \rho \right\}.$$

The Wasserstein constraint can be written via couplings: $W_c(P, \hat{P}_n) = \inf_{\pi \in \Pi(P, \hat{P}_n)} \int c(r, r')\, d\pi(r, r')$. Reformulating:

$$\Phi(w) = \sup_\pi \left\{ \int \ell(r; w)\, d[\pi_1](r) \;:\; \pi \in \Pi(\cdot, \hat{P}_n),\; \int c(r, r')\, d\pi(r, r') \le \rho \right\},$$

where $\pi_1$ denotes the first marginal of $\pi$.

**Step 2: Lagrange relaxation of the transport budget.** Introduce multiplier $\lambda \ge 0$ for the budget constraint:

$$\begin{aligned}
\Phi(w) &\le \inf_{\lambda \ge 0} \sup_\pi \left\{ \int \ell(r; w)\, d[\pi_1](r) + \lambda\Big(\rho - \int c(r, r')\, d\pi\Big) \right\} \\
&= \inf_{\lambda \ge 0} \left\{ \lambda\rho + \sup_\pi \int \big[\ell(r; w) - \lambda c(r, r')\big]\, d\pi(r, r') \right\}.
\end{aligned} \tag{50}$$

Under strong duality conditions (which hold here by the finite support of $\hat{P}_n$ and convexity structure), the inequality is tight.

**Step 3: Decomposition over empirical atoms.** Since $\hat{P}_n = \frac{1}{n}\sum_i \delta_{r_i}$, any coupling $\pi$ with second marginal $\hat{P}_n$ decomposes as $\pi = \frac{1}{n}\sum_i \pi_i$, where each $\pi_i$ is a probability measure on $\mathbb{R}^N$ (the conditional distribution given the atom $r_i$). Substituting:

$$\begin{aligned}\sup_{\pi} \int \big[\ell(r;w) - \lambda c(r,r')\big]\, d\pi &= \frac{1}{n}\sum_{i=1}^{n} \sup_{\pi_i} \int \big[\ell(r;w) - \lambda c(r,r_i)\big]\, d\pi_i(r) \\ &= \frac{1}{n}\sum_{i=1}^{n} \sup_{r} \big[\ell(r;w) - \lambda c(r,r_i)\big], \end{aligned} \tag{51}$$

where the last equality uses the fact that the supremum over all probability measures $\pi_i$ of a linear functional $\int f\, d\pi_i$ equals the essential supremum of $f$, which for continuous $f$ on $\mathbb{R}^N$ is $\sup_r f(r)$.

**Step 4: Combining.** Substituting (51) into (50):

$$\Phi(w) = \inf_{\lambda \geq 0} \left\{ \lambda\rho + \frac{1}{n}\sum_{i=1}^{n} \sup_{r} \big[\ell(r;w) - \lambda c(r,r_i)\big] \right\}. \qquad \square$$

*Remark* 7.7. The dual representation (49) is computationally powerful because it replaces an optimization over an infinite-dimensional space of distributions with a one-dimensional optimization over $\lambda \geq 0$ combined with $n$ independent finite-dimensional maximization problems (one per data point). For specific loss functions (quadratic, linear, piecewise-linear), these inner suprema often admit closed forms (Esfahani and Kuhn, 2018; Blanchet et al., 2022).

### 7.3 Application: Wasserstein robust mean-variance

Following the framework of Blanchet et al. (2022), consider the robust mean-variance program:

$$\min_{w \in \mathcal{C}} \sup_{P \in \mathcal{U}_\rho} \mathrm{Var}_P(w^\top r) \qquad \text{subject to} \qquad \inf_{P \in \mathcal{U}_\rho} \mathbb{E}_P[w^\top r] \geq \mu_{\text{target}}. \tag{52}$$

**Proposition 7.8** (Tractable reduction)**.** *Under quadratic transport cost $c(r,r') = \|r - r'\|_2^2$ and the Wasserstein ambiguity set $\mathcal{U}_\rho$, the robust mean-variance program (52) reduces to:*

$$\min_{w \in \mathcal{C}} w^\top \hat{\Sigma}_n w + R_{\text{var}}(w,\rho) \qquad \textit{subject to} \qquad \hat{\mu}_n^\top w - R_{\text{ret}}(w,\rho) \geq \mu_{\text{target}}, \tag{53}$$

*where $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are the empirical mean and covariance, and $R_{\text{var}}(w,\rho)$ and $R_{\text{ret}}(w,\rho)$ are explicit regularization terms depending on $\rho$, $w$, and sample statistics.*

The precise form of the regularization terms follows from the Kantorovich dual applied to quadratic and linear loss functions, resulting in variance inflation and return shrinkage that grow with the ambiguity radius $\rho$. This has three key implications:

(i) **Convexity preservation:** the regularized problem remains convex under the original linear constraints, so it integrates directly into the projection compiler.

(ii) **Auditability:** distributional uncertainty is translated into explicit, interpretable penalty terms rather than hidden in ad-hoc parameter choices.

(iii) **Data-driven calibration:** the radius $\rho$ can be selected to satisfy a finite-sample coverage condition $\mathbb{P}(P_0 \in \mathcal{U}_{\rho_n}) \geq 1-\alpha$, where $P_0$ is the true data-generating distribution, providing a principled alternative to arbitrary robustness parameters (Blanchet et al., 2022).

# 8 Bilevel Decision-Focused Learning

This section develops the mathematical machinery for training belief models $P_t(\theta)$ by the quality of downstream decisions, rather than by prediction accuracy alone.

## 8.1 Bilevel structure

**Definition 8.1** (Bilevel decision-focused program)**.** Let $\theta$ parameterize the belief model(s). The *bilevel program* is:

$$\begin{aligned} &\text{(Upper level)} \qquad \min_{\theta} \; J(\theta) = \mathbb{E}\big[\, L_{\text{pred}}(\theta) + \alpha\, L_{\text{cal}}(\theta) + \beta\, L_{\text{dec}}\big(w^*(\theta)\big) \,\big], \\ &\text{(Lower level)} \qquad w^*(\theta) \;\in\; \operatorname*{arg\,min}_{w \in \mathcal{C}} \; \phi(w, \theta), \end{aligned} \tag{54}$$

where $L_{\text{pred}}$ is a prediction loss (e.g., cross-entropy), $L_{\text{cal}}$ is a calibration loss, $L_{\text{dec}}$ is a decision quality loss (e.g., realized negative Sharpe ratio), and $\phi$ is the compiler objective.

The challenge is computing $dJ/d\theta$, which requires differentiating through the arg min operation in the lower level.

## 8.2 Implicit differentiation through the KKT system

We now derive the machinery for exact gradient computation through the optimizer.

**Assumption 8.2** (Regularity conditions for implicit differentiation)**.** At the optimal solution $(w^*, \mu^*, \nu^*)$ of the lower-level problem with parameter $\theta$:

(a) **LICQ (Linear Independence Constraint Qualification):** The gradients of active inequality constraints and all equality constraints are linearly independent.

(b) **SSOSC (Strong Second-Order Sufficient Condition):** The Hessian of the Lagrangian is positive definite on the critical cone (the set of directions tangent to active constraints that are orthogonal to the objective gradient).

(c) **Strict complementarity:** For each active inequality constraint $i$, the corresponding multiplier satisfies $\mu_i^* > 0$.

**Definition 8.3** (KKT residual map)**.** Let $u = (w, \mu, \nu)$ collect primal and dual variables. The *KKT residual map* $F : \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_u}$ encodes the stationarity, complementarity, and feasibility conditions:

$$F(u, \theta) \;=\; \begin{pmatrix} \nabla_w \phi(w, \theta) + A(\theta)^\top \mu + E(\theta)^\top \nu \\ \operatorname{diag}(\mu)\, g(w, \theta) \\ h(w, \theta) \end{pmatrix} \;=\; 0, \tag{55}$$

where $g(w, \theta) \leq 0$ are active inequality constraints (with Jacobian $A$) and $h(w, \theta) = 0$ are equality constraints (with Jacobian $E$).

**Theorem 8.4** (Implicit gradient through the optimizer)**.** *Under Assumption 8.2, the implicit function theorem guarantees that in a neighborhood of $\theta$, the optimal solution $u^*(\theta)$ is a smooth function of $\theta$, with Jacobian:*

$$\frac{du^*}{d\theta} = -\left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}. \tag{56}$$

*The upper-level gradient is then:*

$$\frac{dJ}{d\theta} = \left.\frac{\partial J}{\partial \theta}\right|_{\text{explicit}} + \frac{\partial J}{\partial u} \cdot \frac{du^*}{d\theta}. \tag{57}$$

*Proof.* At the optimal $(u^*, \theta)$, we have $F(u^*(\theta), \theta) = 0$. By the implicit function theorem, if $\partial F/\partial u$ is nonsingular at $(u^*, \theta)$—which is guaranteed by LICQ, SSOSC, and strict complementarity together (Bonnans and Shapiro, 2000)—then $u^*$ is locally a smooth function of $\theta$ and differentiation of $F(u^*(\theta), \theta) = 0$ by the chain rule gives:

$$\frac{\partial F}{\partial u}\frac{du^*}{d\theta} + \frac{\partial F}{\partial \theta} = 0 \qquad \Longrightarrow \qquad \frac{du^*}{d\theta} = -\left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}.$$

The upper-level gradient follows by the total derivative: $dJ/d\theta = \partial J/\partial\theta + (\partial J/\partial u)(du^*/d\theta)$. ☐

### 8.3 Explicit matrix form via Schur complement

For practical computation, we work with the KKT matrix directly.

**Proposition 8.5** (KKT matrix system)**.** *When the active set is fixed locally, the linearized KKT system takes the saddle-point form:*

$$\underbrace{\begin{pmatrix} H & A^\top & E^\top \\ A & 0 & 0 \\ E & 0 & 0 \end{pmatrix}}_{K} \begin{pmatrix} dw \\ d\mu \\ d\nu \end{pmatrix} = -\underbrace{\begin{pmatrix} g_\theta \\ a_\theta \\ e_\theta \end{pmatrix}}_{b_\theta} d\theta, \tag{58}$$

*where $H = \nabla^2_{ww}\phi$ is the Hessian of the lower-level objective, and $g_\theta, a_\theta, e_\theta$ are the parameter-Jacobians of the stationarity, inequality, and equality conditions respectively.*

*Proof.* This is obtained by differentiating each block of the KKT residual (55) with respect to $(u, \theta)$ and evaluating at the optimal point. The stationarity block $\nabla_w \phi + A^\top \mu + E^\top \nu = 0$ differentiates to $H\,dw + A^\top d\mu + E^\top d\nu + g_\theta\, d\theta = 0$. The complementarity block, restricted to active constraints where $\mu > 0$ (strict complementarity), gives $A\,dw + a_\theta\, d\theta = 0$. The equality block gives $E\,dw + e_\theta\, d\theta = 0$. ☐

The solution can be computed efficiently via the Schur complement:

$$S = \begin{pmatrix} A \\ E \end{pmatrix} H^{-1} \begin{pmatrix} A^\top & E^\top \end{pmatrix}. \tag{59}$$

One first solves for the multiplier perturbations $(d\mu, d\nu)$ using $S$, then recovers $dw = -H^{-1}(A^\top d\mu + E^\top d\nu + g_\theta\, d\theta)$. The upper-level gradient is:

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial w}\frac{dw}{d\theta} + \left.\frac{\partial J}{\partial \theta}\right|_{\text{explicit}}. \tag{60}$$

*Remark* 8.6 (Numerical considerations). Two implementation issues arise: (i) the KKT matrix $K$ can be ill-conditioned near active-set transitions, and (ii) the $L^1$ terms in the compiler introduce non-smoothness. Standard remedies include Hessian regularization ($H \leftarrow H + \epsilon I$ for small $\epsilon > 0$), smooth approximations to absolute values (e.g., $|x| \approx \sqrt{x^2 + \delta}$), and active-set-stable minibatching that avoids parameter regions near constraint boundaries.

# 9 Deep Reinforcement Learning as Two-Timescale Stochastic Approximation

The architecture includes an online adaptation layer parameterized as a deep RL policy. We formalize this using the theory of multi-timescale stochastic approximation.

## 9.1 Policy decomposition

The policy network is split into two components:

$$\pi_\theta = \pi_{\theta_{\text{struct}}, \theta_{\text{on}}}, \tag{61}$$

where $\theta_{\text{struct}}$ is a *structural encoder* (frozen during evaluation and replay) and $\theta_{\text{on}}$ is an *adaptive head* (updated online).

## 9.2 Reward structure

The per-step reward decomposes as:

$$r_{t+1} = \underbrace{R_{\text{net}}(w_t)}_{\text{net return}} + \underbrace{\alpha \cdot S_{t+1}}_{\text{calibration score}} - \underbrace{\beta \cdot \kappa_t}_{\text{prediction frequency penalty}}, \tag{62}$$

where $R_{\text{net}}$ is the transaction-cost-adjusted portfolio return, $S_{t+1}$ is the log-score of the realized outcome under the issued forecast (connecting to Theorem 4.9), and $\kappa_t$ is a count-based penalty discouraging excessive prediction activity.

## 9.3 Stochastic approximation framework

The online head update follows a stochastic approximation recursion:

$$\theta_{\text{on},t+1} = \theta_{\text{on},t} + a_t\, g_t(\theta_{\text{on},t}, \theta_{\text{struct}}), \tag{63}$$

where $g_t$ is a noisy gradient (e.g., from policy gradient, actor-critic, or PPO-style updates) and $a_t$ is the learning rate.

**Theorem 9.1** (Convergence conditions (Robbins–Monro))**.** *The recursion* (63) *converges to a stationary point of the expected objective provided the step sizes satisfy:*

$$\sum_{t=0}^{\infty} a_t = \infty, \qquad \sum_{t=0}^{\infty} a_t^2 < \infty, \tag{64}$$

*the noise* $g_t - \mathbb{E}[g_t \mid \mathcal{F}_t]$ *has bounded conditional variance, and the mean field* $h(\theta) = \mathbb{E}[g_t \mid \theta_{\mathrm{on},t} = \theta]$ *has suitable stability properties (e.g., the associated ODE* $d\theta/dt = h(\theta)$ *has a globally asymptotically stable equilibrium).*

*Proof sketch.* The proof follows the ODE method of Borkar and Meyn (2000). Under the step-size conditions (64), the discrete iterates $\theta_{\mathrm{on},t}$ track the solution of the limiting ODE $d\theta/dt = h(\theta)$ on compact time intervals with high probability. The first condition ($\sum a_t = \infty$) ensures the iterates have enough “fuel” to reach any region of the parameter space. The second condition ($\sum a_t^2 < \infty$) ensures that the accumulated noise variance is finite, so the iterates do not diffuse away from the ODE trajectory. If the ODE has a globally attracting equilibrium $\theta^*$, the iterates converge to $\theta^*$ almost surely. □

### 9.4 Two-timescale extension

If structural parameters are also updated (during training, not evaluation), two-timescale separation is required.

**Definition 9.2** (Two-timescale stochastic approximation)**.** The *fast process* (online head) and *slow process* (structural encoder) are:

$$\theta_{\mathrm{on},t+1} = \theta_{\mathrm{on},t} + a_t\, g_t(\theta_{\mathrm{on},t},\, \theta_{\mathrm{struct},t}), \tag{65}$$

$$\theta_{\mathrm{struct},t+1} = \theta_{\mathrm{struct},t} + b_t\, h_t(\theta_{\mathrm{on},t},\, \theta_{\mathrm{struct},t}), \tag{66}$$

where both step-size sequences satisfy the Robbins–Monro conditions, and additionally:

$$\lim_{t\to\infty} \frac{b_t}{a_t} = 0. \tag{67}$$

**Proposition 9.3** (Quasi-static separation)**.** *Under the timescale condition* (67)*:*

*(i) The fast process sees* $\theta_{\mathrm{struct}}$ *as quasi-static and converges to its equilibrium* $\theta^*_{\mathrm{on}}(\theta_{\mathrm{struct}})$ *on the fast timescale.*

*(ii) The slow process then evolves on its natural timescale, with the fast variables effectively at their equilibrium for each value of the slow variables.*

*This decoupling enables stable training even when the two parameter blocks have very different loss landscapes.*

### 9.5 Constraint enforcement via projection

Regardless of the RL policy output, the final portfolio is post-processed by the convex projection of Section 6. This provides a hard feasibility guarantee:

$$w_t^{\text{final}} = \mathrm{Proj}_{\mathcal{C}_t}\big(\pi_\theta(s_t)\big) \ \in \ \mathcal{C}_t, \tag{68}$$

where $\mathcal{C}_t$ is the feasible set and $\mathrm{Proj}_{\mathcal{C}_t}$ is the Euclidean projection. This is consistent with constrained RL formulations (Achiam et al., 2017) where policy optimization is combined with explicit feasibility operators, ensuring that even poorly trained or adversarial policy outputs cannot violate hard constraints.

# 10 Security and Governance as Mathematical Constraints

Security controls are formalized as domain restrictions on the information and action spaces.

**Definition 10.1** (Sanitization operator)**.** Let $\mathcal{S} : \mathcal{X}_t \to \mathcal{X}_t^{\text{safe}}$ be a deterministic operator that:

(i) removes or neutralizes adversarial content (injection attacks, malformed inputs),

(ii) validates structural conformity of all inputs against a type schema,

(iii) restricts tool invocations to a pre-approved allowlist.

The effective decision information is $\mathcal{X}_t^{\text{safe}} = \mathcal{S}(\mathcal{X}_t) \subseteq \mathcal{X}_t$.

All optimization and policy maps are composed with $\mathcal{S}$:

$$\Phi_t^{\text{safe}} \ = \ \Phi_t \circ \mathcal{S}. \tag{69}$$

This ensures that unsafe or adversarial inputs are removed *before* they can influence the optimization. Mathematically, $\mathcal{S}$ restricts the reachable state-action space to a policy-safe subspace.

# 11 Conclusion and Open Problems

## 11.1 Summary

We have formalized a belief-to-policy compilation architecture as a stochastic control system over a temporally admissible filtration. The architecture maintains an explicit dual-distribution structure ($Q_t$ from market prices, $P_t$ from structured thesis operators), quantifies their disagreement through a composite divergence $\Delta_t$, and maps these objects into feasible portfolio allocations via a convex projection compiler with full KKT-based sensitivity analysis.

The mathematical coherence of the system rests on several interlocking properties:

(i) Filtration-measurability ensures no leakage (Lemma 2.4).

(ii) The Breeden–Litzenberger identity provides a theorem-backed bridge from prices to priors (Theorem 3.2).

(iii) Simplex-preserving operators maintain probabilistic validity of posteriors (Proposition 4.3).

(iv) Strict propriety of the log score aligns calibration incentives with information-theoretic divergence (Theorem 4.9).

(v) Convex well-posedness guarantees unique, globally optimal compilations (Theorem 6.3).

(vi) Kantorovich duality provides tractable distributionally robust extensions (Theorem 7.6).

(vii) The implicit function theorem enables end-to-end gradient flow through the optimizer (Theorem 8.4).

(viii) Two-timescale stochastic approximation provides convergence guarantees for online adaptation (Theorem 9.1).

## 11.2 Open mathematical problems

Several questions of significant theoretical and practical interest remain:

**1. Joint-law identifiability.** Derive dependence models $\eta_t$ from tradeable multi-asset instruments (e.g., basket options, correlation swaps) while preserving tractable inference for the joint prior $Q_t^{\text{joint}}$.

**2. Active-set stability of implicit gradients.** Characterize the parameter regions where the KKT Jacobian $\partial F/\partial u$ becomes ill-conditioned due to active-set transitions, and derive smoothing procedures that preserve gradient quality without destroying the underlying optimization structure.

**3. Finite-sample robust calibration.** Couple Wasserstein-DRO certificates with proper-score calibration guarantees under temporal dependence, providing joint coverage and calibration bounds for the posterior $P_t$.

**4. Non-asymptotic regret bounds.** Derive finite-time regret bounds for the frozen-structural/adaptive-head policy decomposition under projection constraints, extending existing two-timescale theory beyond asymptotic convergence.

**5. Admissibility-preserving representation learning.** Ensure that learned latent representations remain $\mathcal{F}_t$-measurable under all feature-engineering transformations, including those involving lookback windows and rolling statistics.

# References

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.

Achiam, J. et al. (2017). Constrained policy optimization. *ICML (PMLR 70)*, pages 22–31.

Amos, B. and Kolter, J. Z. (2017). OptNet: Differentiable optimization as a layer in neural networks. *ICML (PMLR 70)*, pages 136–145.

Agrawal, A. et al. (2019). Differentiable convex optimization layers. *NeurIPS 2019*.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Blanchet, J., Chen, L., and Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science*, 68(9):6382–6410.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *Journal of Finance*, 59(1):407–446.

Bonnans, J. F. and Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems*. Springer.

Borkar, V. S. and Meyn, S. P. (2000). The ODE method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Elmachtoub, A. N. and Grigas, P. (2022). Smart "Predict, then Optimize". *Management Science*, 68(1):9–26.

Mohajerin Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric. *Mathematical Programming*, 171:115–166.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9(4):417–428.

Gao, R. and Kleywegt, A. J. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *JASA*, 102(477):359–378.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

Jackwerth, J. C. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, 51(5):1611–1632.

Konda, V. R. and Tsitsiklis, J. N. (2003). On actor-critic algorithms. *SIAM Journal on Control and Optimization*, 42(4):1143–1166.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

López de Prado, M. (2018). *Advances in Financial Machine Learning.* Wiley.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41.

Villani, C. (2009). *Optimal Transport: Old and New.* Springer.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.