High-Dimensional Portfolio Selection with Cardinality Constraints

Presenter: Ben-Zheng Li

December 7, 2025

Reference: Jin-Hong Du, Yifeng Guo, Xueqin Wang, *High-Dimensional Portfolio Selection with Cardinality Constraints*, *Journal of the American Statistical Association*, 118(542): 779–791.

Background: High-Dimensional Portfolio Selection

- Modern equity universes are **high-dimensional**: hundreds to thousands of stocks (d large), while historical samples are limited (n moderate).

Background: High-Dimensional Portfolio Selection

- Modern equity universes are **high-dimensional**: hundreds to thousands of stocks (d large), while historical samples are limited (n moderate).
- Two practical challenges:
  - **Estimation error**: sample mean/covariance are noisy when $d \approx n$ or $d > n$.
  - **Unrealistic portfolios**: classical optimizers often output **dense portfolios** with many tiny positions $\Rightarrow$ high transaction/management costs.

Background: High-Dimensional Portfolio Selection

- Modern equity universes are **high-dimensional**: hundreds to thousands of stocks (d large), while historical samples are limited (n moderate).
- Two practical challenges:
  - **Estimation error**: sample mean/covariance are noisy when $d \approx n$ or $d > n$.
  - **Unrealistic portfolios**: classical optimizers often output **dense portfolios** with many tiny positions $\Rightarrow$ high transaction/management costs.
- Real investors prefer **sparse portfolios** with at most s active stocks (cardinality constraint).
- Key question: *How to construct sparse portfolios that remain efficient out-of-sample in high dimensions?*

Goal: robust risk–return tradeoff and controllable number of holdings.

Baseline I: Markowitz Mean–Variance Model

Step 1: Portfolio return is a weighted sum.

Let $X \in \mathbb{R}^d$ be the random return vector of d assets. Portfolio return:

$$R_p = w^\top X.$$

Baseline I: Markowitz Mean–Variance Model

Step 1: Portfolio return is a weighted sum.

Let $X \in \mathbb{R}^d$ be the random return vector of d assets. Portfolio return:

$$R_p = w^\top X.$$

Step 2: Expected return is linear in weights.

Let $\mu = \mathbb{E}[X]$. Then

$$\mathbb{E}[R_p] = \mathbb{E}[w^\top X] = w^\top \mu.$$

Baseline I: Markowitz Mean–Variance Model

Step 1: Portfolio return is a weighted sum.

Let $X \in \mathbb{R}^d$ be the random return vector of d assets. Portfolio return:

$$R_p = w^\top X.$$

Step 2: Expected return is linear in weights.

Let $\mu = \mathbb{E}[X]$. Then

$$\mathbb{E}[R_p] = \mathbb{E}[w^\top X] = w^\top \mu.$$

Step 3: Risk is measured by variance.

Let $\Sigma = \mathrm{Cov}(X)$. Then

$$\mathrm{Var}(R_p) = \mathrm{Var}(w^\top X) = w^\top \Sigma w.$$

Thus the risk–return tradeoff becomes a clean optimization: **small variance** vs. **large expected return**.

Baseline I: Markowitz Mean–Variance Model

Mean–variance portfolio selection:

$$\begin{aligned} \min_{w} \quad & w^\top \Sigma w \\ \text{s.t.} \quad & w^\top \mu \ge r, \\ & \mathbf{1}^\top w = 1, \\ & w \ge 0. \end{aligned}$$

Interpretation of constraints

- $w^\top \mu \ge r$: achieve at least a target expected return r.
- $\mathbf{1}^\top w = 1$: fully invested (weights sum to 100%).
- $w \ge 0$: long-only (no short selling).

> Among all portfolios with expected return $\ge r$, choose the one with minimal variance (risk).

Baseline II: Why It Fails in High Dimensions

1) Estimation error dominates.

- Mean–variance relies on estimating

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\mu})(X_t - \hat{\mu})^\top.$$

- In high dimensions:
  - $\hat{\mu}$ is extremely noisy $\Rightarrow$ weights overfit sample returns.
  - $\hat{\Sigma}$ ill-conditioned / singular $\Rightarrow$ optimization unstable.

Baseline II: Why It Fails in High Dimensions

1) Estimation error dominates.

- Mean–variance relies on estimating

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\mu})(X_t - \hat{\mu})^\top.$$

- In high dimensions:
  - $\hat{\mu}$ is extremely noisy $\Rightarrow$ weights overfit sample returns.
  - $\hat{\Sigma}$ ill-conditioned / singular $\Rightarrow$ optimization unstable.

2) Dense solutions.

- MV typically outputs **dense portfolios** with many tiny positions $\Rightarrow$ high turnover and transaction/management costs.
- A natural fix is to enforce a cardinality constraint:

$$\|w\|_0 \le s,$$

  but this yields an **NP-hard** combinatorial problem.

Baseline II: Why It Fails in High Dimensions

1) Estimation error dominates.

- Mean–variance relies on estimating

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\mu})(X_t - \hat{\mu})^\top.$$

- In high dimensions:
  - $\hat{\mu}$ is extremely noisy $\Rightarrow$ weights overfit sample returns.
  - $\hat{\Sigma}$ ill-conditioned / singular $\Rightarrow$ optimization unstable.

2) Dense solutions.

- MV typically outputs **dense portfolios** with many tiny positions $\Rightarrow$ high turnover and transaction/management costs.
- A natural fix is to enforce a cardinality constraint:

$$\|w\|_0 \leq s,$$

  but this yields an **NP-hard** combinatorial problem.

Motivation: need a sparse and scalable method that avoids directly estimating (μ, Σ) in high dimensions.

Proposed Method

- **Problem 1: Unstable estimation of (μ, Σ) in high dimensions**
  $\Rightarrow$ **Switch objective: Expected Utility Maximization (EUM)**

$$\max_{w \geq 0} \ \mathbb{E}\big[u(w^\top X)\big]$$

  and approximate expectation by **Sample Average Approximation (SAA)**:

$$\max_{w \geq 0} \ \frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i)$$

Proposed Method

- **Problem 1: Unstable estimation of (μ, Σ) in high dimensions**
  $\Rightarrow$ **Switch objective: Expected Utility Maximization (EUM)**

$$\max_{w \geq 0} \ \mathbb{E}\big[u(w^\top X)\big]$$

  and approximate expectation by **Sample Average Approximation (SAA)**:

$$\max_{w \geq 0} \ \frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i)$$

- **Problem 2: Need sparsity but ℓ_0 is NP-hard**
  $\Rightarrow$ enforce sparsity via a **convex ℓ_1-regularized SAA**:

$$\min_{w \geq 0} \ -\frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i) + \lambda \|w\|_1$$

  (this paper proves ℓ_1-SAA is equivalent to ℓ_0 sparsity after normalization).

Roadmap: EUM $\Rightarrow$ SAA $\Rightarrow$ ℓ_1 equivalence $\Rightarrow$ scalable algorithm.

Proposed Method I: Expected Utility Maximization (EUM)

EUM formulation.

$$\begin{aligned} \max_{w \geq 0} \quad & \mathbb{E}\big[u(w^\top X)\big] \\ \text{s.t.} \quad & \mathbf{1}^\top w = 1. \end{aligned}$$

Notation and intuition

- $X \in \mathbb{R}^d$: random return vector of d assets.
- $w^\top X$: portfolio return (random variable).
- $u(\cdot)$: **concave utility** encoding risk aversion:
  - more concave $\Rightarrow$ more risk-averse;
  - utility evaluates the whole return distribution, not only mean/variance.

Risk–return tradeoff is handled implicitly through $u(\cdot)$.

Proposed Method I: Expected Utility Maximization (EUM)

Utility choices used in the paper

- Log utility:
$$u(z) = \log(z)$$
(growth-oriented; strongly penalizes downside).
- Exponential utility:
$$u(z) = -\exp(-az)$$
(CARA risk aversion; $a > 0$ controls risk aversion).

Proposed Method I: Expected Utility Maximization (EUM)

Utility choices used in the paper

- Log utility:
$$u(z) = \log(z)$$
(growth-oriented; strongly penalizes downside).
- Exponential utility:
$$u(z) = -\exp(-az)$$
(CARA risk aversion; $a > 0$ controls risk aversion).

Connection to Mean–Variance (comparability)

- Under standard conditions (e.g., normal returns or 2nd-order approximation), EUM reduces to a mean–variance tradeoff:
$$\max_{w} \; w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w.$$
- Hence EUM is an **upper-level framework** that includes MV as a special case.

Takeaway: EUM generalizes MV while remaining comparable.

Proposed Method II: Sample Average Approximation (SAA)

Challenge: expectation is unknown.

- EUM objective depends on the true return distribution:

$$\max_{w \geq 0} \mathbb{E}\big[u(w^\top X)\big].$$

- In practice we only observe historical samples $X_1, \dots, X_n$.

Proposed Method II: Sample Average Approximation (SAA)

Challenge: expectation is unknown.

- EUM objective depends on the true return distribution:

$$\max_{w \geq 0} \ \mathbb{E}\big[u(w^\top X)\big].$$

- In practice we only observe historical samples $X_1, \ldots, X_n$.

SAA replaces expectation by a sample average:

$$\max_{w \geq 0} \ \frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i) \quad \text{s.t. } \mathbf{1}^\top w = 1.$$

- Directly optimizes utility **on data** $\Rightarrow$ avoids estimating (μ, Σ).
- Still need to enforce sparsity (next step).

Next: add sparsity (ℓ_0) and show an equivalent convex ℓ_1-regularized SAA.

Proposed Method III: Sparse SAA with Cardinality (ℓ_0)

Goal: enforce sparsity in the SAA problem.

Cardinality-constrained SAA:

$$\begin{aligned} \max_{w \geq 0} \quad & \frac{1}{n} \sum_{i=1}^{n} u(w^\top X_i) \\ \text{s.t.} \quad & \mathbf{1}^\top w = 1, \\ & \|w\|_0 \leq s. \end{aligned}$$

- $\|w\|_0 \leq s$: hold at most s assets (true sparsity).
- This directly matches investor preference, but ℓ_0 makes the problem **NP-hard**.

Key challenge: solve sparse SAA **scalably** in high dimensions.

Proposed Method IV: ℓ_1–ℓ_0 Equivalence

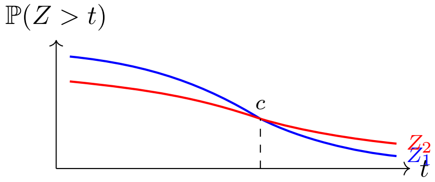
Theorem. ℓ_1–ℓ_0 equivalence after normalization

Assume utility $u(\cdot)$ is *strictly increasing* and *strictly concave*. If portfolio returns from any two feasible weights satisfy the ***single-crossing*** property (their survival functions cross at most once), then there exists $\lambda > 0$ such that for any minimizer w_λ of $(\mathcal{P}_1)$, its normalized version

$$\tilde{w}_\lambda = \frac{w_\lambda}{\mathbf{1}^\top w_\lambda}$$

is an optimal solution of $(\mathcal{P}_0)$.

Single-crossing schematic (one intersection)



Proposed Method IV: Equivalent Convex ℓ_1-Regularized SAA

Equivalent convex formulation:

$$\min_{w \geq 0} \quad -\frac{1}{n}\sum_{i=1}^{n} u(w^\top X_i) + \lambda \|w\|_1.$$

- $\|w\|_1 = \sum_i |w_i|$ induces sparsity **and keeps convexity**.
- λ controls sparsity level (larger $\lambda \Rightarrow$ fewer active stocks).
- Normalize the solution to satisfy $\mathbf{1}^\top w = 1$:

$$\tilde{w} = \frac{w}{\mathbf{1}^\top w}.$$

- The normalized $\tilde{w}$ matches the ℓ_0-constrained optimum.

Result: **true sparsity** + **global optimum** + **scalability**.

Algorithm Overview

Algorithm: Safe Screening + Proximal Gradient (Part I)

Input: Samples $\{X_i\}_{i=1}^n$, utility $u(\cdot)$, regularization λ, step size rule, tolerance ε.

Output: Sparse portfolio $\tilde{w}$.

Pre-loop initialization. Choose initial weights $w^{(0)} \geq 0$ (e.g., uniform on a small subset). Construct an initial dual feasible point $\alpha^{(0)}$ and compute

$$G^{(0)} = F(w^{(0)}) - D(\alpha^{(0)}), \quad F(w) = -\frac{1}{n}\sum_{i=1}^{n} u(w^\top X_i) + \lambda \|w\|_1.$$

Repeat for $t = 0, 1, 2, \ldots$ **until convergence:**

1. Safe screening (dimension reduction). Using $(\alpha^{(t)}, G^{(t)})$, build a safe region $\mathcal{R}^{(t)}$ guaranteed to contain the dual optimum $\alpha^\star$. For each asset j, compute

$$U_j^{(t)} = \max_{\alpha \in \mathcal{R}^{(t)}} |c_j(\alpha)|.$$

If $U_j^{(t)} \leq \lambda$, set $w_j^\star = 0$ and remove j from the active set. Denote remaining indices by $\mathcal{A}^{(t)}$.

Algorithm Overview

Algorithm: Safe Screening + Proximal Gradient (Part II)

2. Proximal gradient step on reduced problem. Let $f(w) = -\frac{1}{n}\sum_{i=1}^{n} u(w^\top X_i)$. Compute gradient on active set:

$$\nabla f(w^{(t)}) = -\frac{1}{n}\sum_{i=1}^{n} u'(w^{(t)\top} X_i) X_i, \quad \text{restricted to } \mathcal{A}^{(t)}.$$

Update:

$$v^{(t)} = w^{(t)} - \eta^{(t)} \nabla f(w^{(t)}), \qquad w_j^{(t+1)} = \max\{\, v_j^{(t)} - \eta^{(t)}\lambda,\ 0\,\}.$$

3. Update dual/gap. Construct $\alpha^{(t+1)}$ and

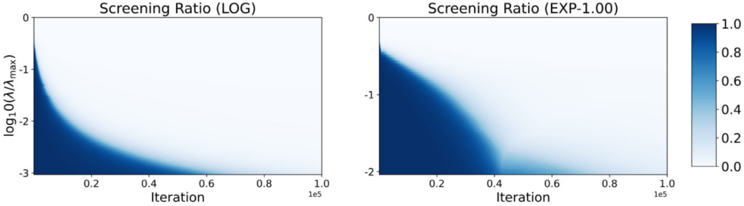
$$G^{(t+1)} = F(w^{(t+1)}) - D(\alpha^{(t+1)}).$$

4. Check convergence. Stop if $\|w^{(t+1)} - w^{(t)}\|_2 / \|w^{(t)}\|_2 \le \varepsilon$ (or $G^{(t+1)} \le \varepsilon$).

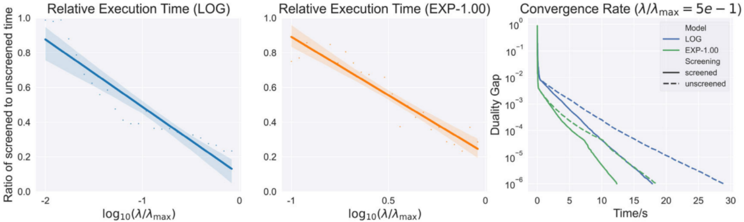
End repeat.

5. Post-loop normalization. $\tilde{w} = \frac{w^{(t+1)}}{\mathbf{1}^\top w^{(t+1)}}$.

Experiments I: Safe Screening Improves Scalability



(a) Screening ratio against iterations for $\log_{10}(\lambda/\lambda_{\max}) \in [-3, 0]$ for LOG and EXP-1.00 utility functions (The lighter, the more screened features).



(b) The relative execution time (the ratio of screened time to unscreened time) against $\log_{10}(\lambda/\lambda_{\max})$, and the convergence rate (duality gap) against execution time when $\lambda/\lambda_{\max} = 5e-1$, for LOG and EXP-1.00 utility functions.

Experiments II: Out-of-Sample Performance and Sparsity

Table 2. Out-of-sample results (without transaction fees) on S&P 500 from 2011 to 2020.

Method	Return	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Num. of assets
Benchmark					
EW	2.3231	0.3941	0.7084	0.9848	437
GMV-P	2.1061	0.2483	0.9487	1.4924	281
GMV-LS	2.4796	0.2929	0.9800	1.4648	258
GMV-NLS	2.3775	0.3223	0.9410	1.3577	323
MV-P	1.5481	0.2343	0.5287	0.7654	211
MV-LS	1.9540	0.2363	0.5981	0.8735	170
MV-NLS	4.0390	0.2341	0.8681	1.2980	118
Our methods					
LOG-1.00	8.9922	0.3194	0.9953	1.5743	20
EXP-0.05	6.2313	0.3504	0.8310	1.4563	37
EXP-0.10	5.7691	0.3027	0.8729	1.3894	23
EXP-0.50	8.5086	0.3558	0.9509	1.5421	7
EXP-1.00	8.7293	0.3558	0.9777	1.5679	6
EXP-1.50	8.1633	0.3558	0.9637	1.5077	5

Experiments III: Generalization to Russell 2000

Table 4. Out-of-sample results (without transaction fees) on Russell 2000 from 2005 to 2020.

Method	Return	Maximum drawdown	Sharpe ratio	Sortino ratio	Avg. Num. of assets
Benchmark					
EW	3.1023	0.6125	0.4176	0.5948	1640
GMV-P	1.2725	0.2677	0.4122	0.5948	838
GMV-LS	1.7997	0.3725	0.4402	0.6251	877
GMV-NLS	1.5940	0.4502	0.3687	0.5247	820
MV-P	−0.7531	0.9398	−0.0408	−0.0554	302
MV-LS	−0.8415	0.9530	−0.0879	−0.1195	351
MV-NLS	−0.8987	0.9722	−0.1180	−0.1590	406
Our methods					
LOG	3.4483	0.5473	0.4514	0.6330	160
EXP-0.05	1.8137	0.4942	0.3738	0.5135	307
EXP-0.10	1.9293	0.4940	0.3866	0.5313	332
EXP-0.50	1.9791	0.4868	0.3953	0.5432	314
EXP-1.00	1.9514	0.4830	0.3929	0.5397	308
EXP-1.50	2.0456	0.4753	0.4076	0.5602	255

Experiments IV: Return Stability Over Time

Figure 5. Annual and monthly returns on S&P 500 from 2011 to 2020 with $n = 120$, $d \leq 454$, $n_{\text{hold}} = 63$ for MV-NLS, LOG and EXP-1.00. The dash lines denote the mean annual returns.