Portfolio Optimization and the Successive Convex Approximation Framework

Ahmer Nadeem Khan

Numerical Optimization Fall 2025

December 11, 2025

Table of Contents

1 Portfolio Optimization
- Introduction
- Financial Data and Modeling
- Mean-Variance Optimization (MVO)

2 Generalizations of MVO
- Mean-Variance Framework
- General-Purpose Algorithm
- Higher Order Portfolios

3 Iterative Algorithms for Nonconvex Optimization
- Majorization-Minimization (MM) Method
- Successive Convex Approximation (SCA) Framework
- Applications to Portfolio Optimization

4 Conclusions

5 References

Table of Contents

1. Portfolio Optimization
   - Introduction
   - Financial Data and Modeling
   - Mean-Variance Optimization (MVO)
2. Generalizations of MVO
   - Mean-Variance Framework
   - General-Purpose Algorithm
   - Higher Order Portfolios
3. Iterative Algorithms for Nonconvex Optimization
   - Majorization-Minimization (MM) Method
   - Successive Convex Approximation (SCA) Framework
   - Applications to Portfolio Optimization
4. Conclusions
5. References

Introduction

- Modern portfolio theory started with Harry Markowitz's 1952 seminal paper "Portfolio Selection" [1].
- Received the **Nobel Prize** in Economics in 1990.
- Roughly **$75 trillion** of global assets under management employ portfolio optimization frameworks, broadly descended from Markowitz mean-variance optimization.

Financial Data

- Financial data is difficult to model because it exhibits non-stationarity, volatility clustering, heavy tails, gain/loss asymmetry etc. at different frequencies [2].
- We want to model returns on a particular asset or a collection of assets (a portfolio).

$$r_{t,i} = \frac{p_{t,i} - p_{t-1,i}}{p_{t-1,i}}$$

is the linear return on the i-th security from time $t-1$ to t.

Modeling Financial Data

- Collect data of prices for N securities and compute returns over the time periods.
- Generate the vector of expected returns $\boldsymbol{\mu}_t \in \mathbb{R}^N$ and the covariance matrix $\boldsymbol{\Sigma}_t \in \mathbb{R}^{N \times N}$.
- Noisy estimation leads to poor performance; lots of improvements to estimation and time-series modeling, e.g. Kalman filer, learning graphs etc [3].

Mean-Variance Optimization (MVO)

- A portfolio is a vector $\boldsymbol{w_t} \in \mathbb{R}^N$ with w_i denoting the proportion (weight) of wealth invested in asset i at time t.
- Thus we get that $\mathbf{1}^\top \boldsymbol{w} = 1$ as an initial constraint (fully invested i.e. no cash positions).
- The main quantities of interest are the return and risk of the portfolio

$$
\begin{aligned}
f_t^P &= \mathbb{E}\left[r_t^{\mathrm{P}}\right] = \boldsymbol{w}_{t-1}^\top \boldsymbol{\mu}, \\
\sigma_t^P &= \mathrm{Var}\left[r_t^{\mathrm{P}}\right] = \boldsymbol{w}_{t-1}^\top \boldsymbol{\Sigma}\, \boldsymbol{w}_{t-1}.
\end{aligned}
$$

where the portfolio P is defined by $\boldsymbol{w}$ and

$$
r_t^P = \boldsymbol{w}_{t-1}^\top \boldsymbol{r_t}
$$

Mean-Variance Optimization

The Markowitz MVO problem can be formulated as the following quadratic program (QP):

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \lambda\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} \\ \text{subject to} \quad & \mathbf{1}^\top \boldsymbol{w} = 1 \text{ (fully invested)} \end{aligned}$$

where λ is the risk-aversion parameter and the objective is a utility function. This can also be reformulated as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} \geq \beta, \\ & \mathbf{1}^\top \boldsymbol{w} = 1 \end{aligned}$$

Constrained Portfolio Optimization

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w} \geq \mathbf{0} \qquad\qquad\qquad \text{(long-only)} \end{aligned}$$

Constrained Portfolio Optimization

$$\begin{aligned}
\max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} & \\
\text{subject to} \quad & \boldsymbol{w} \geq \mathbf{0} & \text{(long-only)} \\
& \|\boldsymbol{w}\|_1 \leq \gamma & \text{(leverage)}
\end{aligned}$$

Constrained Portfolio Optimization

$$
\begin{aligned}
\max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} & \\
\text{subject to} \quad & \boldsymbol{w} \geq \mathbf{0} & \text{(long-only)} \\
& \|\boldsymbol{w}\|_1 \leq \gamma & \text{(leverage)} \\
& \|\boldsymbol{w} - \boldsymbol{w}_0\|_1 \leq \tau & \text{(turnover)}
\end{aligned}
$$

Constrained Portfolio Optimization

$$
\begin{aligned}
\max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} & \\
\text{subject to} \quad & \boldsymbol{w} \geq \mathbf{0} & \text{(long-only)} \\
& \|\boldsymbol{w}\|_1 \leq \gamma & \text{(leverage)} \\
& \|\boldsymbol{w} - \boldsymbol{w}_0\|_1 \leq \tau & \text{(turnover)} \\
& |\boldsymbol{w}| \leq \boldsymbol{u} & \text{(max position size)}
\end{aligned}
$$

Constrained Portfolio Optimization

$$
\begin{array}{rll}
\max\limits_{\boldsymbol{w}} & \boldsymbol{w}^\top \boldsymbol{\mu} - \dfrac{\lambda}{2}\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} & \\
\text{subject to} & \boldsymbol{w} \geq \mathbf{0} & \text{(long-only)} \\
& \|\boldsymbol{w}\|_1 \leq \gamma & \text{(leverage)} \\
& \|\boldsymbol{w} - \boldsymbol{w}_0\|_1 \leq \tau & \text{(turnover)} \\
& |\boldsymbol{w}| \leq \boldsymbol{u} & \text{(max position size)} \\
& \boldsymbol{\beta}^\top \boldsymbol{w} = 0 & \text{(market neutral)}
\end{array}
$$

Constrained Portfolio Optimization

$$
\begin{aligned}
\max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma}\,\boldsymbol{w} & \\
\text{subject to} \quad & \boldsymbol{w} \geq \mathbf{0} & \text{(long-only)} \\
& \|\boldsymbol{w}\|_1 \leq \gamma & \text{(leverage)} \\
& \|\boldsymbol{w} - \boldsymbol{w}_0\|_1 \leq \tau & \text{(turnover)} \\
& |\boldsymbol{w}| \leq \boldsymbol{u} & \text{(max position size)} \\
& \boldsymbol{\beta}^\top \boldsymbol{w} = 0 & \text{(market neutral)} \\
& \|\boldsymbol{w}\|_0 \leq K & \textit{(sparsity)}
\end{aligned}
$$

Convex Optimization

In general, we have

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \, \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w} \in \mathcal{K} \end{aligned}$$

where $\mathcal{K}$ is a convex feasible set. This defines a convex optimization problem with a quadratic objective function, and is thus readily solved by algorithms discussed in class.

- High quality implementations of recent interior-point methods for convex optimization are available in software packages, e.g. MOSEK [4] [5].

Table of Contents

1 Portfolio Optimization
- Introduction
- Financial Data and Modeling
- Mean-Variance Optimization (MVO)

2 Generalizations of MVO
- Mean-Variance Framework
- General-Purpose Algorithm
- Higher Order Portfolios

3 Iterative Algorithms for Nonconvex Optimization
- Majorization-Minimization (MM) Method
- Successive Convex Approximation (SCA) Framework
- Applications to Portfolio Optimization

4 Conclusions

5 References

Maximum Sharpe Ratio Portfolio

The MSRP is an example of a concave-convex fractional program

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \frac{\boldsymbol{w}^\top \boldsymbol{\mu} - r_f}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w}}} \\ \text{subject to} \quad & \mathbf{1}^\top \boldsymbol{w} = 1, \\ & \boldsymbol{w} \geq \mathbf{0}, \end{aligned}$$

where r_f is the risk-free rate.

- Concave-convex FPs may be turned into a series of convex problems using various transform methods [3].
- Iterative algorithms - convergence?

Kelly Criterion Portfolio

The Kelly Criterion maximizes long-term wealth by maximizing the growth rate, and can be expressed as

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \mathbb{E}\Big[\log\Big(1+\boldsymbol{w}^\top \boldsymbol{r}\Big)\Big] \\ \text{subject to} \quad & \mathbf{1}^\top \boldsymbol{w} = 1, \\ & \boldsymbol{w} \geq \mathbf{0}. \end{aligned}$$

- This is a convex problem since the log is concave.
- Generalized to maximizing expected utility via mean-value approximations or exponential cone programming (SOCP).

Mean-Variance Framework

- It was recently shown in Xiu et al. (2023) that these problems can be formulated by a particular choice of the hyperparameter λ, i.e. a general function $f(x, y)$ which is a tradeoff between the return and the risk.
- A general-purpose algorithm was developed based on the Successive Convex Approximation (SCA) method using quadratic approximations - a sequential QP [6].

Higher Order Portfolios

- The MVO formulation has certain drawbacks; one of them is only second-order information.
- We can include higher moments, in particular, skewness and kurtosis.
- Parameters: $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$, with complexities $O(N), O(N^2), O(N^3), O(N^4)$ for storage and computation.

MVSK Portfolio

The Mean-Variance-Skewness-Kurtosis portfolio is formulated as

$$\begin{array}{ll}\underset{\boldsymbol{w}}{\text{min}} & -\lambda_1\, \boldsymbol{w}^\top \boldsymbol{\mu} \;+\; \lambda_2\, \boldsymbol{w}^\top \boldsymbol{\Sigma}\, \boldsymbol{w} \\ & \lambda_3\, \boldsymbol{w}^\top \boldsymbol{\Phi}\,(\boldsymbol{w} \otimes \boldsymbol{w}) \;+\; \lambda_4\, \boldsymbol{w}^\top \boldsymbol{\Psi}\,(\boldsymbol{w} \otimes \boldsymbol{w} \otimes \boldsymbol{w}) \\ \text{subject to} & \boldsymbol{w} \in \mathcal{W}.\end{array}$$

- This is non-convex due to the third moment.
- Structure and estimation methods need to be exploited.
- Efficient algorithms involve the MM and SCA frameworks [7]

Table of Contents

1 Portfolio Optimization
- Introduction
- Financial Data and Modeling
- Mean-Variance Optimization (MVO)

2 Generalizations of MVO
- Mean-Variance Framework
- General-Purpose Algorithm
- Higher Order Portfolios

3 Iterative Algorithms for Nonconvex Optimization
- Majorization-Minimization (MM) Method
- Successive Convex Approximation (SCA) Framework
- Applications to Portfolio Optimization

4 Conclusions

5 References

Nonconvex Optimization

Suppose we have have a well-defined minimization problem (1)

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & V(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned}$$

where V is nonconvex and $\mathcal{X}$ is a convex set.

- The MM method is an iterative algorithm producing $\boldsymbol{x}^0, \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots$ converging to the optimal $\boldsymbol{x}^*$.
- At each iteration k, a subproblem is solved of the type

  $$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \ \widetilde{f}(\boldsymbol{x} \mid \boldsymbol{x}^k).$$

  where $\widetilde{f}(\cdot \mid \boldsymbol{x}^k)$ is a surrogate function that upper-bounds f globally.

Majorization-Minimization Method

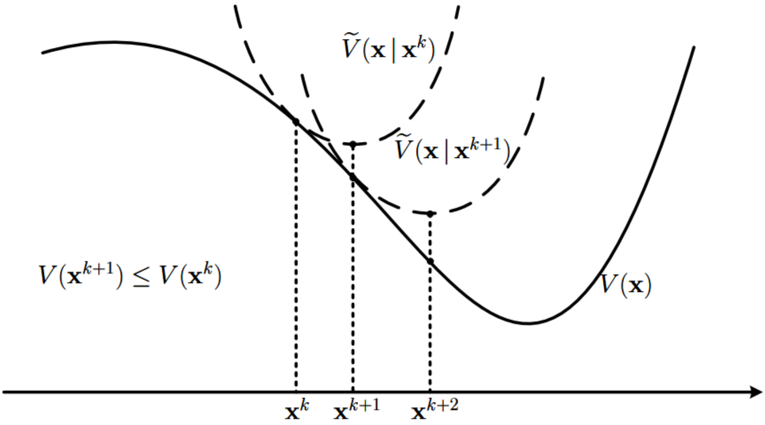


Figure: Pictorial Representation of MM [8].

Preliminaries

Definition (Directional derivative)

A function $f : \mathbb{R}^m \to (-\infty, \infty]$ is directionally differentiable at $\boldsymbol{x} \in \operatorname{dom} f \triangleq \{\boldsymbol{x} \in \mathbb{R}^m : f(\boldsymbol{x}) < \infty\}$ along a direction $\boldsymbol{d} \in \mathbb{R}^m$ if the limit

$$f'(\boldsymbol{x}; \boldsymbol{d}) \; = \; \lim_{\lambda \downarrow 0} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}$$

exists. This value $f'(\boldsymbol{x}; \boldsymbol{d})$ is called the *directional derivative* of f at $\boldsymbol{x}$ along $\boldsymbol{d}$. If f is directionally differentiable at $\boldsymbol{x}$ along all directions, then f is said to be directionally differentiable at $\boldsymbol{x}$.

If f is differentiable then $f'(\boldsymbol{x}; \boldsymbol{d}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$. For convex functions, the directional derivative exists and is given by

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \inf_{\lambda > 0} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}.$$

Preliminaries (Cont.)

Since f is convex, it is also locally Lipschitz continuous. This can be used to show that if $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} f)$, there exists a finite constant $L > 0$ such that $|f'(\boldsymbol{x}; \boldsymbol{d})| \leq L\|\boldsymbol{d}\|_2$ for all $\boldsymbol{d} \in \mathbb{R}^m$.

Definition (Subgradient and Subdifferential)

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ and let $\boldsymbol{x} \in \operatorname{dom} f$. A vector $\boldsymbol{\xi} \in \mathbb{R}^m$ is called a *subgradient* of f at $\boldsymbol{x}$ if

$$f(\boldsymbol{x} + \boldsymbol{d}) \;\geq\; f(\boldsymbol{x}) + \boldsymbol{\xi}^\top \boldsymbol{d}, \qquad \forall\, \boldsymbol{d} \in \mathbb{R}^m.$$

The *subdifferential* of f at $\boldsymbol{x} \in \operatorname{dom} f$ is defined as the set

$$\partial f(\boldsymbol{x}) \;=\; \left\{\boldsymbol{\xi} \in \mathbb{R}^m : f(\boldsymbol{x} + \boldsymbol{d}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^\top \boldsymbol{d},\ \forall\, \boldsymbol{d} \in \mathbb{R}^m\right\}.$$

Preliminaries (Cont.)

Theorem

Let $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} f)$*. Then the subdifferential* $\partial f(\boldsymbol{x})$ *is nonempty, compact, and convex.*

Lemma

The subdifferential at $\boldsymbol{x} \in \operatorname{dom} f$ *can be equivalently written as*

$$\partial f(\boldsymbol{x}) \triangleq \left\{\boldsymbol{\xi} \in \mathbb{R}^m : f'(\boldsymbol{x}; \boldsymbol{d}) \geq \boldsymbol{\xi}^\top \boldsymbol{d}, \ \forall \, \boldsymbol{d} \in \mathbb{R}^m\right\}.$$

Lemma (Max formula)

At any $\boldsymbol{x} \in \operatorname{int}(\operatorname{dom} f)$ *and all* $\boldsymbol{d} \in \mathbb{R}^m$*, it holds that*

$$f'(\boldsymbol{x}; \boldsymbol{d}) = \sup_{\boldsymbol{\xi} \in \partial f(\boldsymbol{x})} \boldsymbol{\xi}^\top \boldsymbol{d}.$$

Optimality

If f is differentiable and convex, then $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$ using the lemma above. We now define directional-stationarity.

Definition (d-stationarity)

Given Problem (1) in the above setting, a point $\boldsymbol{x}^* \in \mathcal{X}$ is a *d-stationary solution* of (1) if

$$V'(\boldsymbol{x}^*; \boldsymbol{y} - \boldsymbol{x}^*) \geq 0, \qquad \forall\, \boldsymbol{y} \in \mathcal{X}.$$

- Finding a global optimal solution is in general not possible.
- We have to settle for stationarity, as defined above.

Optimality (Cont.)

- When V is convex, the max formula implies that $\boldsymbol{x}^*$ is a d-stationary (and thus a global optimal) solution of Problem (1) if there exists $\boldsymbol{\xi} \in \partial f(\boldsymbol{x}^*)$ such that

$$\boldsymbol{\xi}^\top(\boldsymbol{y} - \boldsymbol{x}^*) \geq 0, \qquad \forall\, \boldsymbol{y} \in \mathcal{X}.$$

- If V is differentiable, then since $V'(\boldsymbol{x}; \boldsymbol{d}) = \nabla V(\boldsymbol{x})^\top \boldsymbol{d}$, condition the the condition becomes

$$(\boldsymbol{y} - \boldsymbol{x}^*)^\top \nabla V(\boldsymbol{x}^*) \geq 0, \qquad \forall\, \boldsymbol{y} \in \mathcal{X}.$$

Problem Assumptions

Given Problem (1), we assume the following technical conditions:

1. $\mathcal{X} \neq \varnothing$ is a closed and convex set in $\mathbb{R}^m$;
2. $V : O \to \mathbb{R}$ is continuous on the open set $O \supseteq \mathcal{X}$;
3. $V'(\boldsymbol{x}; \boldsymbol{d})$ exists at any $\boldsymbol{x} \in \mathcal{X}$ and for all feasible directions $\boldsymbol{d} \in \mathbb{R}^m$ at $\boldsymbol{x}$;
4. V is bounded from below.

Surrogate Function

The surrogate function $\widetilde{V} : O \times O \to \mathbb{R}$ satisfies the following conditions:

1. $\widetilde{V}(\cdot \mid \cdot)$ is continuous on $\mathcal{X} \times \mathcal{X}$;
2. $\boldsymbol{y} \in \arg\min_{\boldsymbol{x} \in \mathcal{X}} \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{y}) - V(\boldsymbol{x})$;
3. The directional derivative of $\widetilde{V}$ satisfies

$$\widetilde{V}'(\boldsymbol{x}; \boldsymbol{d} \mid \boldsymbol{x}) = V'(\boldsymbol{x}; \boldsymbol{d}),$$

for all $\boldsymbol{x} \in \mathcal{X}$ and feasible directions $\boldsymbol{d} \in \mathbb{R}^m$ at $\boldsymbol{x}$.

The MM Algorithm

Algorithm The Majorization–Minimization (MM) Algorithm

1: Choose initial point $\boldsymbol{x}^0 \in \mathcal{X}$

2: $k \leftarrow 0$

3: **repeat**

4: Construct the surrogate (majorizer) $\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k)$ of V around $\boldsymbol{x}^k$

5: Compute next iterate:

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k)$$

6: $k \leftarrow k + 1$

7: **until** convergence

Convergence

Theorem (Convergence of MM)

Let $\{\boldsymbol{x}^k\}_{k \in \mathbb{N}_+}$ be the sequence generated by the MM algorithm under the assumptions stated before. Then any limit point of $\{\boldsymbol{x}^k\}$ (if it exists) is a d-stationary solution of Problem (1).

Proof of Convergence (1/3)

Proof. The MM algorithm generates a nonincreasing sequence $\{V(\boldsymbol{x}^k)\}$. Indeed, by the majorization property, for each k we have

$$V(\boldsymbol{x}^{k+1}) \;\leq\; \widetilde{V}(\boldsymbol{x}^{k+1} \mid \boldsymbol{x}^k) - c^k \;\leq\; \widetilde{V}(\boldsymbol{x}^k \mid \boldsymbol{x}^k) - c^k \;=\; V(\boldsymbol{x}^k),$$

where

$$c^k \triangleq \widetilde{V}(\boldsymbol{x}^k \mid \boldsymbol{x}^k) - V(\boldsymbol{x}^k),$$

and the last equality follows from the fact that the majorization inequality is achieved with equality at $\boldsymbol{x} = \boldsymbol{x}^k$.
Thus $\{V(\boldsymbol{x}^k)\}$ is nonincreasing and bounded from below.

Proof of Convergence (2/3)

Let $\boldsymbol{x}^*$ be a limit point of $\{\boldsymbol{x}^k\}$, i.e., there exists a subsequence $\{\boldsymbol{x}^{k_t}\}$ such that $\boldsymbol{x}^{k_t} \to \boldsymbol{x}^* \in \mathcal{X}$ as $t \to \infty$.
For all $\boldsymbol{x} \in \mathcal{X}$ we have

$$\begin{aligned}
\widetilde{V}(\boldsymbol{x}^{k_{t+1}} \mid \boldsymbol{x}^{k_{t+1}}) - c^{k_{t+1}} &= V(\boldsymbol{x}^{k_{t+1}}) \\
&\leq V(\boldsymbol{x}^{k_t+1}) \\
&\leq \widetilde{V}(\boldsymbol{x}^{k_t+1} \mid \boldsymbol{x}^{k_t}) - c^{k_t} \\
&\leq \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^{k_t}) - c^{k_t}.
\end{aligned}$$

Letting $t \to \infty$ and using the continuity of $\widetilde{V}(\cdot \mid \cdot)$ and $V(\cdot)$, we obtain that $\{c^k\}$ converges (to a finite value) and

$$\widetilde{V}(\boldsymbol{x}^* \mid \boldsymbol{x}^*) \ \leq \ \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^*), \qquad \forall\, \boldsymbol{x} \in \mathcal{X}.$$

Hence $\boldsymbol{x}^*$ is a minimizer of the surrogate $\widetilde{V}(\cdot \mid \boldsymbol{x}^*)$ over $\mathcal{X}$.

Proof of Convergence 3/3)

Since $\boldsymbol{x}^*$ minimizes $\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^*)$ over $\mathcal{X}$, its directional derivative in any feasible direction is nonnegative:

$$0 \;\leq\; \widetilde{V}'(\boldsymbol{x}^*; \boldsymbol{d} \mid \boldsymbol{x}^*), \qquad \forall\, \boldsymbol{d} \in \mathbb{R}^m \text{ such that } \boldsymbol{x}^* + \boldsymbol{d} \in \mathcal{X}.$$

By assumption (directional derivative matching),

$$\widetilde{V}'(\boldsymbol{x}^*; \boldsymbol{d} \mid \boldsymbol{x}^*) = V'(\boldsymbol{x}^*; \boldsymbol{d}),$$

and hence

$$0 \;\leq\; V'(\boldsymbol{x}^*; \boldsymbol{d}), \qquad \forall\, \boldsymbol{d} \in \mathbb{R}^m \text{ such that } \boldsymbol{x}^* + \boldsymbol{d} \in \mathcal{X}.$$

This is exactly the definition of a d-stationary solution of Problem (1), so every limit point $\boldsymbol{x}^*$ of $\{\boldsymbol{x}^k\}$ is d-stationary. $\square$

MM Algorithm Implementation

- We have choose a termination criteria; merit function

$$f = ||\boldsymbol{x}^{k+1} - \boldsymbol{x}^{k}||.$$

- The major hurdle is choosing the surrogate functions.
- Examples include **first order Taylor expansion** (concave V) and **second order Taylor expansion**:

$$\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{y}) = V(\boldsymbol{y}) + \nabla V(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}).$$

$$\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{y}) = V(\boldsymbol{y}) + \nabla V(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^\top \mathbf{M} (\boldsymbol{x} - \boldsymbol{y}).$$

where $\mathbf{M}$ satisfies $\mathbf{M} - \nabla^2 V(\boldsymbol{x}) \succeq 0$ for all $\boldsymbol{x} \in \mathcal{X}$.

Successive Convex Approximation Method

In SCA, we again use a surrogate convex function, but with different assumptions. It need not be a global upper-bound, and

- $\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k)$ must be strongly convex on the feasible set $\mathcal{X}$; and
- $\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k)$ must be differentiable with

$$\nabla_{\boldsymbol{x}} \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k) = \nabla V(\boldsymbol{x}).$$

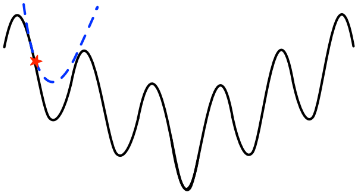
There is also a smoothing step

$$\widehat{\boldsymbol{x}}^{k+1} = \arg\min_{\boldsymbol{x} \in \mathcal{X}} \ \widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k),$$

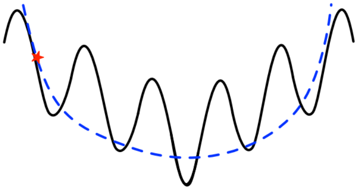
$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \gamma^k \left(\widehat{\boldsymbol{x}}^{k+1} - \boldsymbol{x}^k \right),$$

where $\gamma^k \in (0, 1]$ is a properly designed stepsize sequence.

SCA Example



(a) MM approach: Upper approximation (dotted blue line) of the original function (solid black line) at the base point (red star).



(b) SCA approach: Local approximation (dotted blue line) of the original function (solid black line) at the base point (red star).

Special Cases of Unconstrained SCA

- Using the surrogate

$$\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k) = V(\boldsymbol{x}^k) + \nabla V(\boldsymbol{x}^k)^\top (\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2\alpha^k}\|\boldsymbol{x} - \boldsymbol{x}^k\|^2.$$

  Minimizing this quadratic surrogate yields the update

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha^k \nabla V(\boldsymbol{x}^k),$$

  which is the **Gradient Descent method**.

- Using a second-order surrogate that includes the Hessian,

$$\widetilde{V}(\boldsymbol{x} \mid \boldsymbol{x}^k) = V(\boldsymbol{x}^k) + \nabla V(\boldsymbol{x}^k)^\top \Delta\boldsymbol{x} + \frac{1}{2\alpha^k}\Delta\boldsymbol{x}^\top \nabla^2 V(\boldsymbol{x}^k)\Delta\boldsymbol{x},$$

  where $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^k$. Minimizing this yields the **Newton method**

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha^k \left(\nabla^2 V(\boldsymbol{x}^k)\right)^{-1} \nabla V(\boldsymbol{x}^k).$$

SCA Applied to MSVK Optimization

We first split the objective function in the convex and non-convex parts as follows

$$\begin{aligned} f(\boldsymbol{w}) &= f_{\mathrm{cvx}}(\boldsymbol{w}) + f_{\mathrm{ncvx}}(\boldsymbol{w}), \\ f_{\mathrm{cvx}}(\boldsymbol{w}) &= -\lambda_1\,\phi_1(\boldsymbol{w}) + \lambda_2\,\phi_2(\boldsymbol{w}), \\ f_{\mathrm{ncvx}}(\boldsymbol{w}) &= -\lambda_3\,\phi_3(\boldsymbol{w}) + \lambda_4\,\phi_4(\boldsymbol{w}), \end{aligned}$$

where

$$\begin{aligned} \phi_1(\boldsymbol{w}) &= \mathbb{E}[\boldsymbol{w}^\top \boldsymbol{r}] = \boldsymbol{w}^\top \boldsymbol{\mu}, \\ \phi_2(\boldsymbol{w}) &= \mathbb{E}\left[(\boldsymbol{w}^\top \bar{\boldsymbol{r}})^2\right] = \boldsymbol{w}^\top \boldsymbol{\Sigma}\,\boldsymbol{w}, \\ \phi_3(\boldsymbol{w}) &= \mathbb{E}\left[(\boldsymbol{w}^\top \bar{\boldsymbol{r}})^3\right] = \boldsymbol{w}^\top \boldsymbol{\Phi}(\boldsymbol{w} \otimes \boldsymbol{w}), \\ \phi_4(\boldsymbol{w}) &= \mathbb{E}\left[(\boldsymbol{w}^\top \bar{\boldsymbol{r}})^4\right] = \boldsymbol{w}^\top \boldsymbol{\Psi}(\boldsymbol{w} \otimes \boldsymbol{w} \otimes \boldsymbol{w}). \end{aligned}$$

MVSK (Cont.)

We find a convex, quadratic approximation function for f_{ncvx} around the point $\boldsymbol{w}^k$ as

$$\widetilde{f}_{\mathrm{ncvx}}(\boldsymbol{w} \mid \boldsymbol{w}^k) = f_{\mathrm{ncvx}}(\boldsymbol{w}^k) + \nabla f_{\mathrm{ncvx}}(\boldsymbol{w}^k)^\top (\boldsymbol{w} - \boldsymbol{w}^k) + \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}^k)^\top \left[\nabla^2 f_{\mathrm{ncvx}}(\boldsymbol{w}^k)\right]_{\mathrm{PSD}} (\boldsymbol{w} - \boldsymbol{w}^k),$$

where $\left[\boldsymbol{\Xi}\right]_{\mathrm{PSD}}$ denotes the projection of the matrix $\boldsymbol{\Xi}$ onto the positive semidefinite cone.

- This must be evaluated efficiently in practice without computing the full eigenspectrum of the matrix.
- If the spectrum is available, we can achieve this by projecting the eigenvalues to the non-negative orthant. In this case, the matrix has to be made positive definite to preserve strong convexity [7].

MSVK (Cont.)

A quadratic convex surrogate of $f(\boldsymbol{w})$ at $\boldsymbol{w}^k$ is

$$\begin{aligned}\widetilde{f}(\boldsymbol{w} \mid \boldsymbol{w}^k) &= f_{\mathrm{cvx}}(\boldsymbol{w}) + \widetilde{f}_{\mathrm{ncvx}}(\boldsymbol{w} \mid \boldsymbol{w}^k) \\ &= \frac{1}{2}\,\boldsymbol{w}^\top Q^k \boldsymbol{w} \;+\; \boldsymbol{w}^\top q^k \;+\; \text{constant},\end{aligned}$$

where

$$\begin{aligned}Q^k &= \lambda_2\, \nabla^2 \phi_2(\boldsymbol{w}) + \left[\nabla^2 f_{\mathrm{ncvx}}(\boldsymbol{w}^k)\right]_{\mathrm{PSD}}, \\ q^k &= -\lambda_1\, \nabla \phi_1(\boldsymbol{w}) + \nabla f_{\mathrm{ncvx}}(\boldsymbol{w}^k) - \left[\nabla^2 f_{\mathrm{ncvx}}(\boldsymbol{w}^k)\right]_{\mathrm{PSD}}\, \boldsymbol{w}^k.\end{aligned}$$

- The gradients and the hessians can be computed using non-parametric or parametric methods.
- This has to be achieved efficiently in practice. See discussion in Chapter 9 of *Portfolio Optimization* [3].
- A similar approach can be applied using MM.

Table of Contents

1 Portfolio Optimization
- Introduction
- Financial Data and Modeling
- Mean-Variance Optimization (MVO)

2 Generalizations of MVO
- Mean-Variance Framework
- General-Purpose Algorithm
- Higher Order Portfolios

3 Iterative Algorithms for Nonconvex Optimization
- Majorization-Minimization (MM) Method
- Successive Convex Approximation (SCA) Framework
- Applications to Portfolio Optimization

4 Conclusions

5 References

Conclusions

- Further considerations include accommodating nonconvex sets, block-parallel algorithms and implementations, sparsity and regularization etc.
- SCA is an active area of research, with applications to optimization problems in many fields, including modern Portfolio Optimization.

Table of Contents

1 Portfolio Optimization
- Introduction
- Financial Data and Modeling
- Mean-Variance Optimization (MVO)

2 Generalizations of MVO
- Mean-Variance Framework
- General-Purpose Algorithm
- Higher Order Portfolios

3 Iterative Algorithms for Nonconvex Optimization
- Majorization-Minimization (MM) Method
- Successive Convex Approximation (SCA) Framework
- Applications to Portfolio Optimization

4 Conclusions

5 References

References I

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

Svitlana Vyetrenko, David Byrd, Nick Petosa, Mahmoud Mahfouz, Danial Dervovic, Manuela Veloso, and Tucker Balch.
Get real: realism metrics for robust limit order book market simulations.
In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2021. Association for Computing Machinery.

References II

Daniel P. Palomar.
Portfolio Optimization: Theory and Application.
Cambridge University Press, 2025.
Available online at
https://portfoliooptimizationbook.com/.

Stephen Boyd and Lieven Vandenberghe.
Convex Optimization.
Cambridge University Press, 2004.
Available online at
https://stanford.edu/~boyd/cvxbook/.

References III

MOSEK ApS.
Portfolio optimization.
https://www.mosek.com/resources/portfolio-optimization/, 2025.
Accessed: 2025-12-07.

Shengjie Xiu, Xiwen Wang, and Daniel P. Palomar.
A fast successive qp algorithm for general mean-variance portfolio optimization.
IEEE Transactions on Signal Processing, 71:2713–2727, 2023.

Rui Zhou and Daniel P. Palomar.
Solving high-order portfolios via successive convex approximation algorithms.
IEEE Transactions on Signal Processing, 69:892–904, 2021.

References IV

Gesualdo Scutari and Ying Sun.
Parallel and distributed successive convex approximation methods for big-data optimization, 2018.